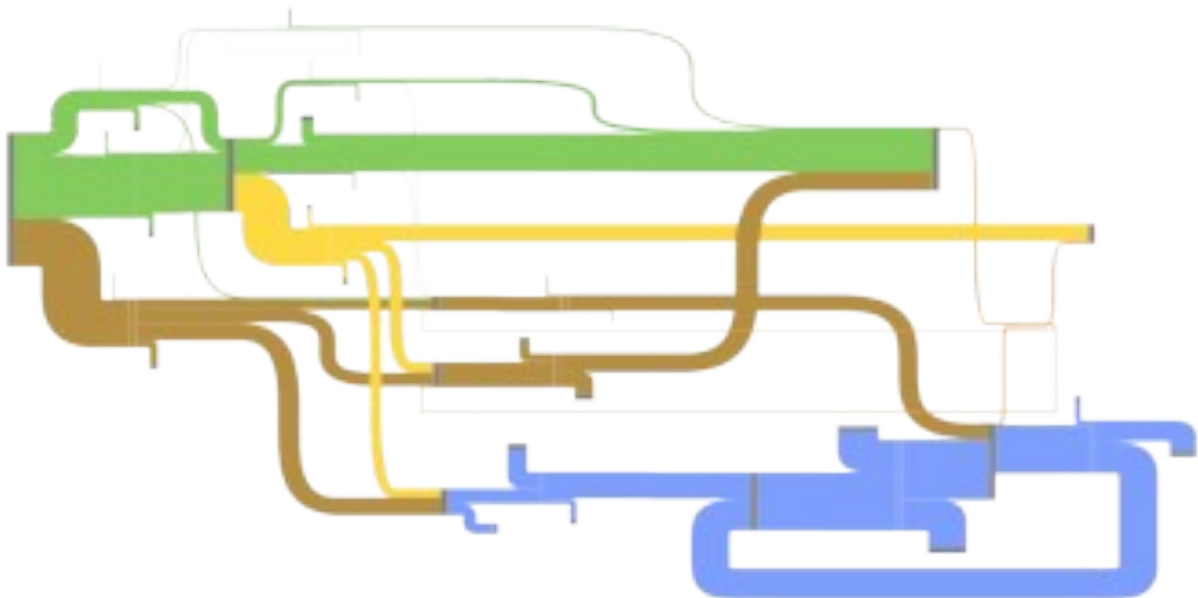


Rapport de stage de fin d'étude

Analyses de Flux de Matières de filières agricoles à l'échelle des régions françaises

Maxime Latgé



- Encadrant INRIA, STEEP *Jean-Yves Courtonne*
- Encadrant Terriflux *Julien Alapetite*
- Encadrant Ecole *Xavier Goaoc*

Résumé du rapport :

Mon stage s'est déroulé à l'INRIA de Grenoble, dans l'équipe STEEP en association avec la start-up Terriflux pour l'analyse de flux de matières de filières agricoles à l'échelle nationale et régionale. Ce stage s'inclut dans un projet financé par l'ADEME intitulé Scalable. Les raisons d'être de ce projet sont les suivantes : d'une part les systèmes agro-alimentaires qui produisent et transforment les biomasses d'origines agricoles sont de plus en plus globalisés et d'autre part ces mêmes systèmes sont exposés à des perturbations et stress divers (économiques, géopolitiques, sanitaires, climatiques). Il apparaît donc nécessaire, et c'est l'objectif de ce projet, d'être en mesure d'évaluer les vulnérabilités des territoires en terme de dépendance à d'autres territoires, d'empreinte écologique ou encore de concurrences d'usage pour toutes les filières de biomasses d'origine agricole (viande, céréales, industrie laitière, légumes, ...). Ainsi les analyses de flux de matière permettent d'obtenir une modélisation des différentes filières agro-alimentaires, une quantification de leurs flux et une analyse de leurs impacts environnementaux et socio-économiques.

Ces analyses de flux de matières se réalisent en différentes étapes successives que l'on peut réitérer : une définition de la structure de la filière, une collecte des données d'intérêt, un processus de réconciliation des données et d'estimation des incertitudes, une représentation graphique en diagramme de Sankey et une évaluation des résultats avec des experts de la filière. On peut également régionaliser un modèle de filière national déjà existant pour plus de précision géographique. Enfin, on peut mener un travail d'analyse pour évaluer les impacts socio-économiques et environnementaux de ces filières. Ainsi, ce travail de recherche permet d'obtenir une meilleure vision des systèmes de production, transformation et distribution des filières de biomasses d'origine agricole et de leurs impacts. Cela est primordial pour répondre de manière pertinente à des enjeux de relocalisation, résilience ou réduction des impacts environnementaux.

Remerciements

Je tiens tout d'abord à remercier Jean-Yves et Julien pour m'avoir encadré tout au long de ce stage. Ils ont su rester disponibles et à l'écoute malgré la distance. Merci à Mr Goacoc pour sa sérieuse implication en tant que tuteur de l'école des Mines. Un grand merci général à toute l'équipe STEEP qui pendant nos quelques moments partagés ont été très sympathiques et accueillants. Merci à Bernard pour m'avoir hébergé sur Grenoble quand cela a été nécessaire et merci à mes parents pour avoir hébergé une antenne de l'INRIA à la maison !

Sommaire

1) Contexte du stage.....	5
1.a) INRIA.....	5
1.b) STEEP.....	5
1.c) Terriflux.....	6
1.d) Projet SCALABLE.....	7
1.e) Le télétravail.....	8
2) Méthodologie.....	11
2.a) Cycle de travail pour une analyse de flux de matière.....	11
- La définition de la structure de la filière.....	11
- La collecte des données.....	12
- La réconciliation des données et l'estimation des incertitudes.....	12
- La représentation graphique en diagramme de Sankey.....	14
- Analyse et confrontation des résultats aux experts de la filière.....	14
2.b) Architecture logicielle.....	15
2.c) Exemple d'analyse.....	15
3) Mes Études.....	17
3.a) Filière Viande.....	17
3.b) Filière Lait.....	23
3.c) Filière Paille.....	28
3.d) Codes Pythons.....	32
4) Conclusion.....	34
5) Bibliographie.....	35
6) Annexes.....	36
6.a) Liste des abréviations.....	36
6.b) Archives Numériques.....	38
- Fichiers Excels.....	38
- Diagramme de Sankey.....	38

1) Contexte du stage

Mon stage a eu lieu au centre de l'INRIA de Grenoble. J'ai été intégré à l'équipe STEEP dans laquelle mon tuteur de stage Jean-Yves Courtonne travaille en tant que chercheur, ainsi qu'à la start-up Terriflux incubée à l'INRIA et dirigée par Julien Alapetite. L'essentiel du travail réalisé lors de mon stage s'inscrit dans un projet nommé SCALABLE, financé par l'ADEME.

1.a) INRIA

L'Institut National de Recherche en Informatique et en Automatique a été créé en 1967. Il a pour objectifs d'entreprendre des recherches fondamentales et appliquées mais aussi de soutenir l'innovation technologique. Le centre de Grenoble Rhône-Alpes a été créé en 1992. Il compte aujourd'hui 35 équipes de recherche et 25 start-ups. L'idée de travailler dans un centre de recherche m'a séduit car j'accorde beaucoup d'importance à la finalité de mon travail. Œuvrer pour produire du savoir à destination d'autres scientifiques, de politiques, d'industriels ou du grand public me semble être d'une grande utilité sociale.

1.b) STEEP

L'équipe STEEP, pour Soutenabilité, Transition, Environnement, Économie biophysique et Politiques locales, est une équipe de recherche interdisciplinaire qui se consacre à la modélisation systémique et à la simulation des interactions entre les facteurs environnementaux, économiques et sociaux. Cette équipe comprend 9 membres permanents, 7 thésards et 4 stagiaires. Le choix de ce stage résulte d'un intérêt marqué pour les sujets de recherche de cette équipe. En effet, étant sensible au dérèglement climatique et à ses conséquences, questionner les origines d'un tel changement implique de questionner les fondements socio-économiques de nos sociétés. Pour cela l'équipe STEEP se divise en 2 axes :

L'axe Alternatives sociotechniques (AST) a pour objectif d'étudier la base matérielle de l'économie et ses impacts environnementaux mais aussi de proposer des méthodes participatives permettant à des acteurs d'imaginer leurs propres alternatives en termes de modes de production et de consommation. C'est suivant cet axe que j'ai travaillé.

L'axe Risques systémiques globaux qui a pour objectif de comprendre les trajectoires biophysiques et les alternatives sociotechniques de nos sociétés. STEEP étudie deux grandes familles de risques systémiques : des risques tendanciels de long terme (décennies) et des risques de contagion systémique de court terme (mois ou années), de nature plus aléatoire que les précédents. L'objectif est de décrire certains risques actuels mais aussi d'évaluer les risques liés aux alternatives décrites dans l'axe AST.

Concernant l'axe AST, Il est conduit par Jean-Yves Courtonne et il se compose de deux sous-axes : les Analyses de Flux de Matières (AFM) pour des filières particulières et les analyses systémiques d'alternatives sociotechniques de l'ensemble de l'économie prenant en compte les interactions entre secteurs. J'ai pour ma part travaillé à l'analyse des flux de matière pour les filières de biomasse d'origine agricole. Afin de réaliser ces analyses de flux de matière, STEEP collabore étroitement avec l'entreprise Terriflux sur cette thématique.

1.c) Terriflux

Terriflux est une start-up essaimée de l'équipe STEEP, portée par Julien Alapetite. Le caractère innovant de cette start-up, se base sur le développement de la méthodologie AFM Filières développée par Jean-Yves Courtonne et Julien Alapetite. Cette méthodologie permet la modélisation des chaînes d'approvisionnement de filières économiques aux échelles nationales et infra-nationales. La base de cette méthodologie est décrite dans une thèse [1] ainsi que dans des articles scientifiques [2]. Pour mettre en pratique cette méthodologie, deux outils logiciels permettant la réalisation effective d'études de filière ont été développés de 2017 à fin 2019, dans le cadre d'un précédent projet de recherche ADEME sur la modélisation des flux de biomasse. Un premier outil réalise la réconciliation des données et un deuxième outil permet la visualisation sous la forme de diagramme de Sankey. Nous y reviendrons plus tard. In fine, ces études permettent de nourrir des réflexions autour des enjeux économiques, sociaux et environnementaux des filières considérées. À terme, l'objectif de Terriflux est de pouvoir proposer au grand public ses outils de réconciliation des données et de représentation graphique. Cela permettra à différents acteurs¹ de pouvoir expertiser des filières d'intérêt. Voici ci dessous, un schéma synthétisant la méthodologie AFM Filière.



Figure 1: Méthodologie AFM Filière.

¹ Acteurs tels que des collectivités, des interprofessions, des agences régionales ou nationales.

1.d) Projet Scalable

Le projet sur lequel j'ai été affecté s'intitule « Scalable ». Issu d'un appel à projet Graine, ce projet est financé par l'Agence de la Transition Écologique (ADEME). Les appels à projet Graine cherchent à favoriser le développement d'une bioéconomie au service de la transition écologique et énergétique des territoires. La bioéconomie est une économie « décarbonée », qui ne reposerait plus sur les énergies fossiles. Plus concrètement, il s'agit d'articuler accompagnement des filières de production et de valorisation des biomasses avec la réduction des impacts environnementaux et la prise en compte des services rendus par les écosystèmes. Il est intéressant de noter que ce projet s'inscrit dans la continuité de 2 anciens projets Graine : le projet Boat et le projet AF-Filière [3] sur lequel Jean-Yves Courtonne a travaillé.

Le projet Scalable, est en collaboration avec l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), l'agence régionale Auvergne-Rhône-Alpes Énergie Environnement (AURA-EE) et le Parc Naturel Régional de Chartreuse. J'ai donc pu régulièrement assister à des réunions avec des membres de chacune de ces organisations.

Le titre complet de ce projet est : « *Métabolisme des biomasses d'origine agricole : représentations multi-échelles, analyse des vulnérabilités et évaluation par les acteurs des territoires.* ». Par métabolisme, il faut entendre les flux de matière et d'énergie en jeu dans les processus de production, transformation et valorisation des Biomasses d'Origine Agricole (BOA). Ainsi l'objectif de ce projet est d'analyser les flux de matière de toutes les filières concernant les BOA afin d'évaluer les vulnérabilités des territoires en terme de dépendance à d'autres territoires, d'empreinte écologique ou encore de concurrence d'usage. Autrement dit, de répondre à des questions telles que : dans quelles mesures les filières de valorisation des BOA d'un territoire parviennent-elles à satisfaire les besoins de sa population ? sans transfert de vulnérabilité vers d'autres territoires ? Quels compromis s'observent et quels leviers de changement peut-on envisager ? L'originalité de ce projet tiens aussi dans de régulières délibérations collectives entre chercheurs et parties prenantes de la production, transformation, et valorisation des BOA, afin de confronter les résultats théoriques de ce projet avec la réalité du terrain. Ces analyses se feront à l'échelle nationale mais aussi à l'échelle du Plan Alimentaire interTerritorial « région alpine » (PAIT) qui est une zone élargie autour de Grenoble.

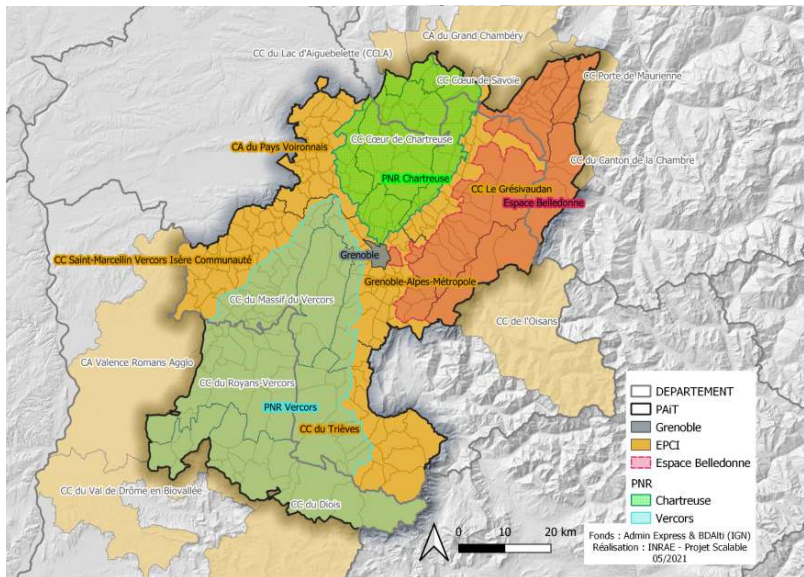


Figure 2: Plan du PAIT

Ainsi, Jean-Yves Courtonne et Julien Alapetite sont en charge de la représentation multi-échelle des filières de biomasse d'origine agricole. J'ai donc poursuivi ce travail avec eux. Un certain nombre de filières ayant déjà été modélisées (céréales, bois, ...) , j'ai travaillé durant mon stage sur les filières viande, lait et paille qui feront chacune l'objet d'un chapitre.

J'ai voulu travailler sur ce projet car les problématiques associées à l'alimentation comme la proximité de sa production, sa qualité ou encore son faible impact carbone [4] sont à mon sens des questions primordiales.

J'ai particulièrement apprécié la réunion de lancement de ce projet durant laquelle différentes personnes impliquées dans l'activité agricole de la région ont pu prendre la parole. Ils comprenaient bien les enjeux d'un tel projet et semblaient enthousiastes à sa réalisation. En effet, une composante importante de ce projet est d'inclure les différents acteurs locaux (paysans, coopératives, distributeurs, industriels, ...) dans le déroulement du projet Scalable pour qu'ils puissent juger les différentes analyses et résultats produits.

1.e) Le télétravail

Point non négligeable du stage, il s'est déroulé dans sa plus grande partie en télétravail suite aux directives gouvernementales pour lutter contre la pandémie du Covid-19. J'ai été présent sur Grenoble seulement 4 semaines, et présent à l'INRIA une dizaine de jours. Je pense que cela a considérablement modifié l'expérience du stage. Cela a été positif en terme de temps personnel : en évitant par exemple les temps de transports mais aussi en travaillant depuis différents endroits en France. Cela a été négatif par rapport à l'isolement et à la motivation : il est moins réjouissant d'aller travailler avec peu de contacts humains. Cela dit Jean-Yves et Julien ont été très disponibles pour moi. Nous avons convenu d'une réunion fixe hebdomadaire le lundi matin pour faire un point

sur mon travail. De plus, je pouvais également les joindre à tout moment suivant mes questionnements et mon avancée dans mon travail.

Voici ci dessous une collection de capture d'images pour mieux illustrer mon quotidien² !

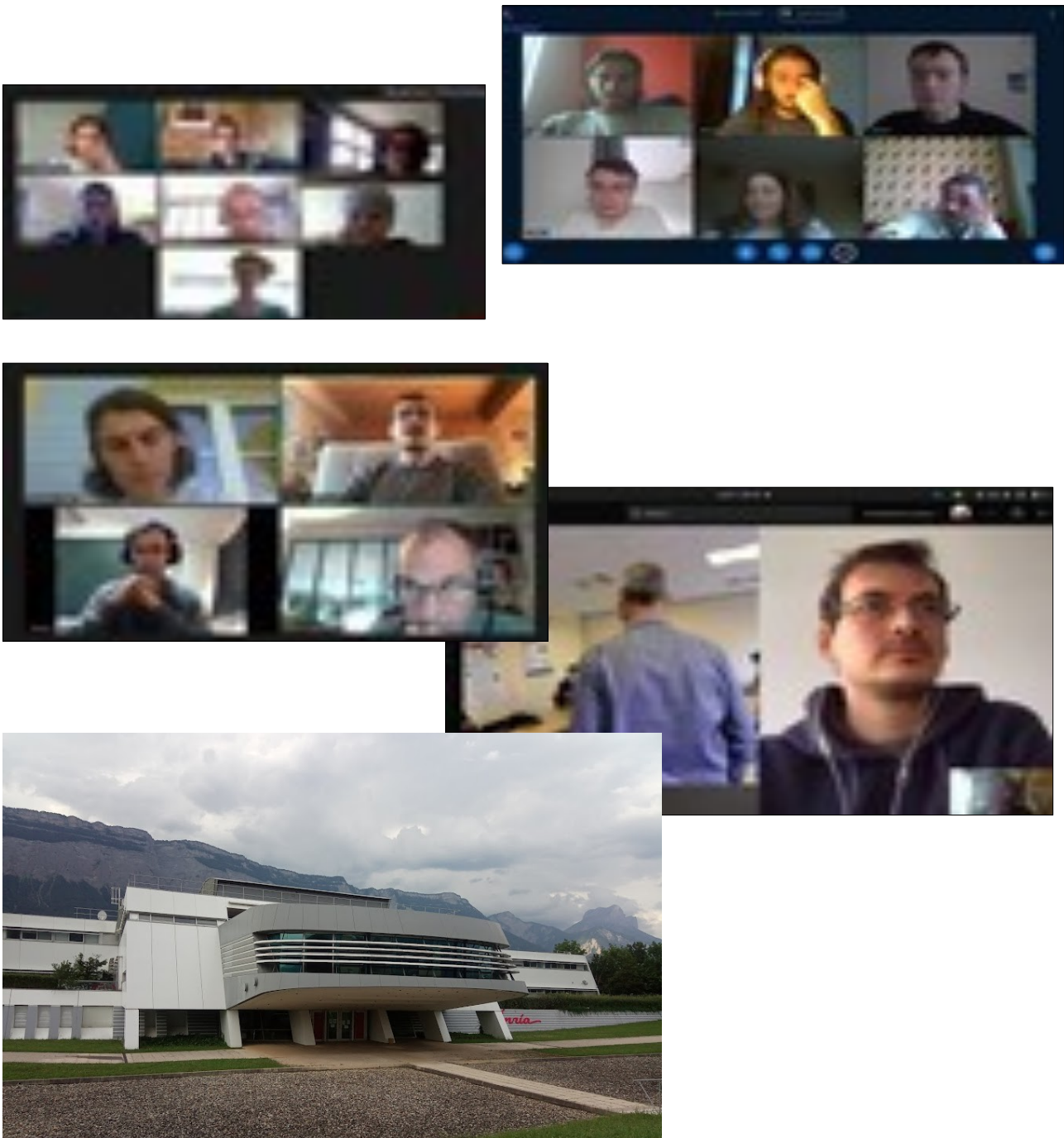


Figure 3: Quelques captures d'écran de mes réunions + une photo de l'INRIA

2 Les photos sont floutées pour des questions de droits à l'image.

2) Méthodologie

Après avoir présenté le contexte global du stage, nous allons maintenant voir comment concrètement réaliser une Analyse de Flux de Matière (AFM) avec la méthodologie AF-Filière³.

2.a) Cycle de travail pour une analyse de flux de matière

Une AFM se réalise en plusieurs étapes que l'on peut réitérer plusieurs fois afin d'affiner l'analyse. Matériellement, l'AFM se traduit par un fichier Excel organisé en différents onglets sur lesquels je reviendrai plus tard.

Afin d'illustrer plus clairement mon propos, prenons l'exemple fictif de la filière pomme.

- La définition de la structure de la filière

Il s'agit de définir l'ensemble des produits de la filière, l'ensemble des secteurs de la filière ainsi que l'ensemble des flux pouvant exister entre produits et secteurs. On décrit les flux pouvant exister à l'aide de Table Emploi Ressource (TER) : la table emploi définit quel secteur consomme quel produit, la table ressource définit quel secteur génère quel produit.

Pour la pomme, nos produits seront : Pommes, Compote, Jus, Cidre. Nos secteurs seront : Pommiers, Industries, Consommation

Table 1 : Table Emploi Ressource pour l'exemple de la filière pomme.

	Emploi				Ressource		
	Pommiers	Industries	Consommation		Pommiers	Industries	Consommation
Pommes	■	■	■	Pommes	■	■	■
Compote	■	■	■	Compote	■	■	■
Jus	■	■	■	Jus	■	■	■
Cidre	■	■	■	Cidre	■	■	■

Emploi : Les industries consomment des pommes. Les pommes, compotes, jus et cidre sont consommés par les consommateurs finaux. Voici donc la liste des flux pouvant exister : Pommes → Industries, Pommes → Conso, Compote → Conso, Jus → Conso, Cidre → Conso.

Ressource : Les pommiers produisent des pommes, les industries produisent des compotes, du jus ou du cidre. Voici donc les flux pouvant exister : Pommiers → Pommes, Industries → Compote, Industries → Jus, Industries → Cidre.

³ Il existe tout un ensemble de méthodologies pour réaliser des AFM, de même qu'il existe différents types d'AFM. À titre d'exemple, on pourra consulter cette AFM de la filière viande en Italie [5]

- La collecte des données

Afin de poursuivre l'AFM, un travail de recherche et d'extraction des données est nécessaire. Il existe de nombreuses sources de données. Au niveau européen on trouve la base de donnée Procom gérée par Eurostat, qui recense la production/exportation/importation de nombreux produits. Au niveau national, les mêmes données de production sont gérées par l'Institut national de la statistique et des études économiques (l'INSEE). Concernant les statistiques agricoles, elles sont produites par l'Agreste (service statistique du ministère de l'agriculture et de la pêche) . On peut aussi trouver de nombreuses données par l'intermédiaire d'interprofessions tel que l'Idèle (institut de l'élevage). Concernant les imports/exports, ils sont comptabilisés dans la base de données SitraM (Système d'Information sur le TRANsport de Marchandises) sur laquelle je reviendrai plus tard. Une autre manière de récupérer des informations est de questionner directement les acteurs d'une filière.

Il est important pour chaque donnée de bien connaître la zone qu'elle recouvre (France métropolitaine, DOM, régions, ...) et sa période (année, trimestre, mois, ...). De plus, toute donnée doit être associée à une incertitude que l'on estime en fonction de la fiabilité de sa source. Ces données peuvent être ensuite entrées dans notre modèle sous différentes formes :

- une valeur de flux :

Par exemple, si l'on sait que la production de pomme nationale est de 1 700 000 tonnes de pommes. Alors on pourra écrire : (Pommiers → Pommes) = 1 700 kt avec une incertitude de 10 %.

- une borne maximum et/ou minimum pour un flux :

Par exemple, si l'on sait que la production de compote de pomme seule est de 151 kt mais qu'il existe aussi une production de compote de pomme en mélange avec d'autres fruits.

Alors on pourra écrire : $\min(\text{Industries} \rightarrow \text{Compote}) = 151 \text{ kt}$

- une relation linéaire entre 2 flux :

Par exemple, si l'on sait que sur toutes les pommes partant vers les différentes industries, 10 % d'entre elles partent pour la production de cidre.

Alors on pourra écrire : $10 \% * (\text{Pommes} \rightarrow \text{Industries}) = (\text{Industries} \rightarrow \text{Cidre})$

- La réconciliation des données et l'estimation des incertitudes

L'étape suivante est la réconciliation des données. En entrée, nous avons une structure de filière représentée par les tables emploi-ressources (équivalentes à des matrices) partiellement remplies. En sortie nous voulons retrouver ces tables emploi-ressources remplies et respectant un certain nombre de contraintes.

La première contrainte est la conservation de la masse. Nous voulons pour chaque produit que la somme de ses productions et imports soit égale à la somme de ses consommations et exportations.

De même pour les secteurs industriels : ce qui sort dans un secteur doit être égal à ce qui entre⁴. Il est donc nécessaire de travailler avec des unités assurant la conservation de la masse.

Par exemple si on utilise une unité avec un taux d'humidité constant, on ne prend pas en considération les pertes d'eau liées au séchage et on assure donc la conservation de la masse (sinon il faudra modéliser les flux d'eau).

Ensuite nous pouvons rajouter des contraintes spécifiques à une filière, ce sont les relations linéaires entre 2 flux présentées précédemment.

Enfin il y a une contrainte de positivité et de nullité : un flux négatif ne peut exister et tous les flux ne pouvant pas exister doivent rester nuls.

L'objectif est donc de minimiser l'écart entre les valeurs d'entrée et de sortie du modèle en respectant cet ensemble de contraintes. Pour mesurer cet écart de valeur, la méthode des moindres carrés pondérés a été choisie. La pondération prenant en compte l'incertitude sur ces flux.

Si l'on considère le vecteur x représentant les valeurs d'entrée mesurées, le vecteur \tilde{x} représentant les valeurs de sortie correspondantes et Q une matrice carrée diagonale contenant les écart-type de chaque valeur alors voici la fonction objectif F :

$$F(x) = (\tilde{x} - x)^T Q^{-1} (\tilde{x} - x)$$

Figure 4: Fonction objectif

En supposant que notre problème contienne N valeurs d'entrées, avec σ comme écart type, alors on peut écrire la fonction objectif de cette manière :

$$F(x) = \sum_{i=0}^N \frac{(\tilde{x}_i - x_i)^2}{\sigma_i}$$

Figure 5: Fonction objectif

Notons que normalement la méthode des moindres carrés pondérés utilise la variance mais empiriquement l'écart type nous donne de meilleurs résultats.

Nous savons que la fonction objectif (l'écart entre la matrice d'entrée et de sortie) est convexe, ce qui implique l'existence d'un minimum, et quadratique (de degré 2). De plus le problème est continu. Les valeurs de sortie peuvent prendre toutes les valeurs possibles dans l'ensemble des réels positifs.

Toutes les contraintes sont linéaires. Cependant il existe des contraintes d'inégalité, ce qui implique que notre problème n'a pas de solutions analytiques simples. Dans le cas contraire, la méthode des multiplicateurs de Lagrange nous aurait donné une solution analytique. Ainsi, la minimisation sous contrainte s'opère grâce à des solveurs numériques tel que OSQP ou MOSEK qui trouvent par des méthodes numériques le minimum de notre fonction objectif et donc notre table emploi-ressources de sortie.

4 Exception faite pour les secteurs produisant des matières premières ou les secteurs de consommation car ils n'ont que des sorties ou que des entrées de produits.

Comme nous n'avons pas de solution analytique, nous utilisons la méthode de Monte Carlo pour calculer l'incertitude des valeurs de sortie. En simulant un nombre suffisamment grand de réconciliations avec un tirage aléatoire des valeurs d'entrées dans leur intervalle de confiance, on arrive au bout d'une centaine de tirages à obtenir une estimation des incertitudes de sortie suffisamment fiable.

- La représentation graphique en diagramme de Sankey

Une fois la réconciliation faite, l'étape suivante est de représenter les AFM sous la forme d'un diagramme de Sankey. Ces diagrammes de flux se lisent de la gauche (production) vers la droite (consommation). Plus un flux est épais, plus la valeur du flux qu'il représente est grande.

Voici le diagramme de Sankey pour notre exemple :

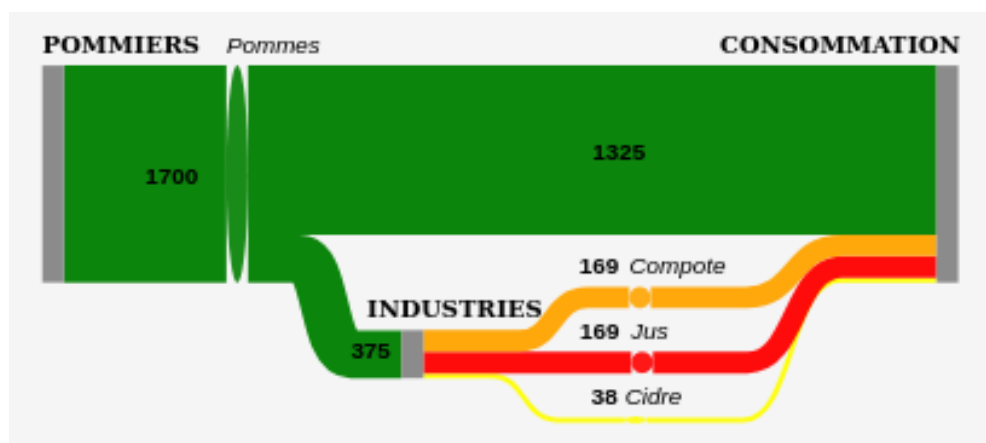


Figure 6: Diagramme de Sankey pour la filière pomme

- Analyse et confrontation des résultats aux experts de la filière

Une fois notre diagramme de Sankey prêt, nous avons finalisé une première itération de notre AFM. Il est important en premier lieu d'avoir une analyse critique des résultats de la réconciliation, puis de confronter la modélisation avec des avis d'expert ou des documents issus d'organismes spécialisés dans les filières étudiées.

Concernant l'analyse des résultats de la réconciliation, un indicateur précieux pour un flux est l'écart entre la valeur d'entrée et la valeur de sortie de la réconciliation, pondérée par l'incertitude (méthode des moindres carrés pondérés). Par exemple si un flux très certain s'écarte grandement de sa valeur initiale, on peut supposer qu'il y a un problème dans la modélisation.

Dans notre exemple, il y a trop peu de données d'entrées pour pouvoir utiliser l'indicateur indiqué ci dessus. Cela dit, il resterait de nombreuses améliorations à apporter au modèle. Il faudrait procéder à un sourçage rigoureux des données, trouver des données en plus grand nombre et plus précises, mettre un maximum à la production de compote. On pourrait aussi revoir la structure de la filière en détaillant les différentes variétés de pomme, en séparant les cidres artisanaux des cidres industriels, en détaillant les différentes consommations, en rajoutant les concentrés de pomme, etc... Il serait également nécessaire de rajouter les imports exports. On pourrait aussi envisager une modélisation à l'échelle régionale.

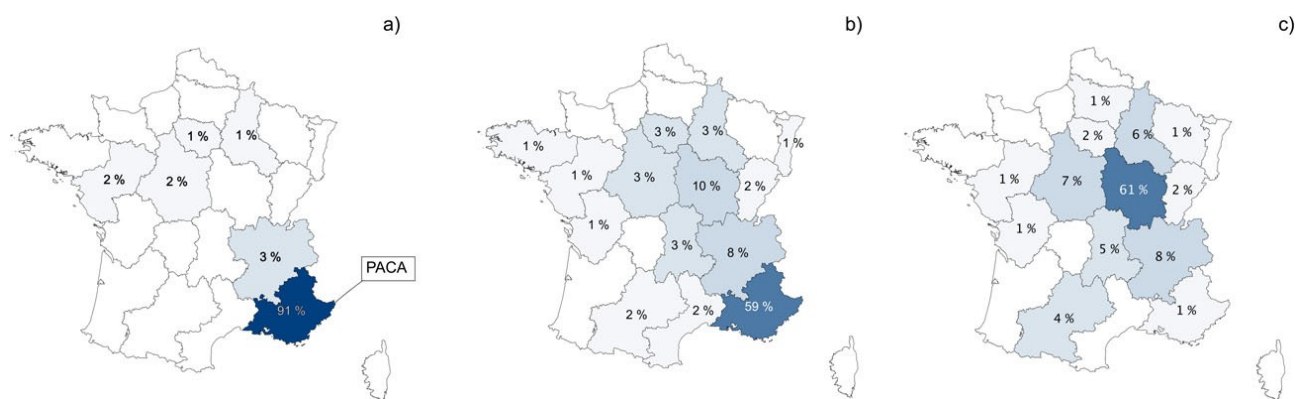
2.b) Architecture logicielle

Le processus de réconciliation décrit ci-dessus est aujourd'hui opérationnel grâce à la start-up Terriflux. La définition de la structure de la filière ainsi que la collecte des données se réalise via un fichier Excel normé. (Un onglet pour respectivement : la définition des produits, la définition des secteurs, les tables emploi-ressources, les données de flux, les minimums et maximums, les contraintes). L'algorithme de réconciliation ainsi que les codes pour l'interface web et la création des diagrammes de Sankey sont sous contrôle de source sur le gestionnaire de version Gitlab. Grâce aux fonctionnalités de Gitlab, une chaîne d'intégration et de déploiement continu est mise en place. Ainsi, à chaque mise à jour du code, ce dernier est testé puis directement mis en ligne sur le site internet terriflux.fr. Il existe 3 versions du site web correspondant chacune à 3 branches du dépôt Git : une version de développement, une version de test et une version stable. On peut donc travailler soit en local en téléchargeant directement le dépôt Git et en utilisant par exemple l'éditeur Visual Studio Code, soit travailler en ligne via l'interface du site web terriflux. L'algorithme de réconciliation prend en entrée le fichier Excel normé et correctement rempli. Il restitue en sortie le même fichier avec des onglets supplémentaires de résultats. Pour la minimisation sous contrainte, l'algorithme utilise le résolveur d'équations OSQP ou MOSEK. J'ai pour ma part essentiellement utilisé l'interface en ligne pour travailler sur mes analyses de flux de matière, même si j'ai appris à installer tout le nécessaire pour pouvoir travailler en local et éditer le code sous contrôle de source.

2.c) Exemple d'analyse

Une fois l'AFM finalisée, à quoi peut-elle bien servir ? Il est possible d'aller plus loin en calculant par exemple l'empreinte environnementale de chaque flux, en cherchant à identifier des conflits d'usage ou encore en mettant en lumière des dépendances à d'autres territoires. Prenons l'exemple de l'industrie céréalière française [6].

Après avoir fait une analyse de flux de matière de la filière céréalière française à l'échelle régionale puis couplé ces flux avec des pressions environnementales, voici différents exemples d'analyse possible :



Sur les images précédentes on peut voir : a) la provenance du pain consommé en PACA, b) la provenance de la farine utilisée pour le pain consommé en PACA, c) la provenance du blé utilisé pour le pain consommé en PACA. Il est par exemple intéressant de voir que la région PACA est quasiment indépendante pour sa production de pain, mais qu'elle ne l'est pas pour sa production de blé qui provient majoritairement de la Bourgogne. Ceci met en lumière la dépendance en blé de la région PACA à (l'ancienne) région Bourgogne.

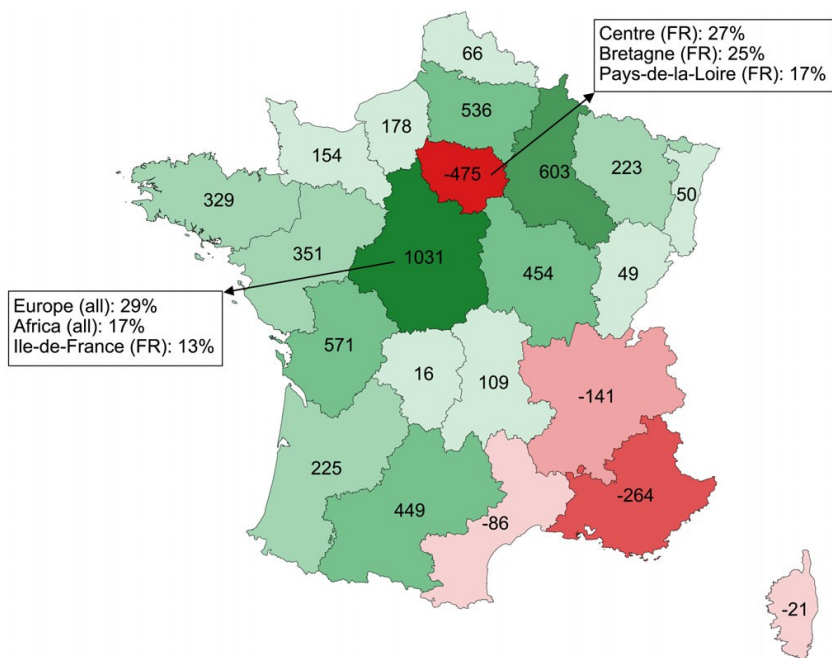


Figure 8: Import/Export net de surface agricole par région

Sur cet autre exemple. On peut voir l'export net de surface agricole par région française en kilohectares (kha). Les nuances de rouge correspondent aux régions importatrices nettes tandis que les nuances de vert correspondent aux régions exportatrices nettes. On peut donc voir pour chaque région importatrice, la surface de terre externe à la région nécessaire à son approvisionnement en céréales, et pour chaque régions exportatrices, la surface de terre dédiée à l'exportation de céréale. Il est intéressant de constater sans surprise que la région Île-de-France, très urbaine, est importatrice nette, tandis que la région centre, très agricole, est exportatrice nette. On peut également voir le détail des plus grandes provenances d'importations ou destinations d'exportations pour ces régions.

Nous pourrions voir ultérieurement dans l'étude sur la filière paille un exemple de conflit d'usage.

3) Mes Études

Maintenant que nous avons vu les détails sur la réalisation d'une analyse de flux de matière, voici les études successives que j'ai menées durant mon stage.

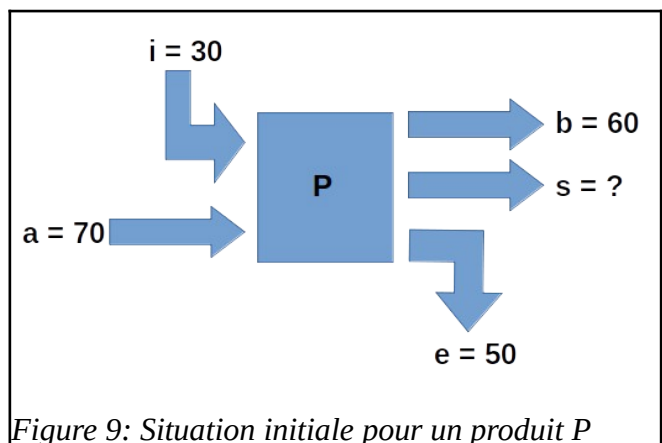
3.a) Filière Viande

Ma première analyse de flux de matière s'est portée sur la filière viande. Une AFM avait déjà été réalisée par le Bureau d'Analyse Sociétale pour une Information Citoyenne (le BASIC). Cela m'a aidé pour modéliser les différents produits et secteurs de cette filière.

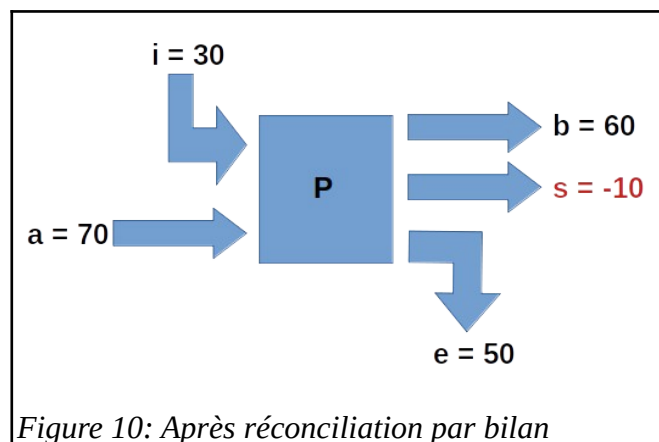
Le BASIC ne disposant pas de la technologie de réconciliation au moment de l'AFM, les flux ont été réconciliés manuellement par bilan. Le problème lorsque l'on procède de cette manière est que l'on fait reposer l'erreur sur un seul flux. On peut donc se retrouver avec par exemple un flux négatif.

Exemple :

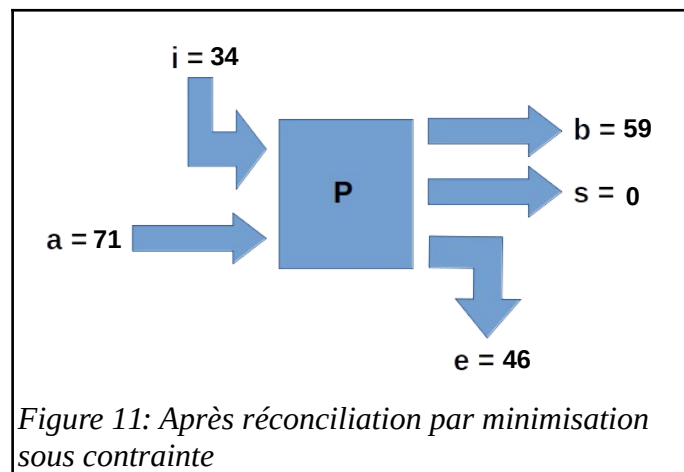
Nous avons un produit P avec les flux suivants :



En réconciliant par bilan,
on calcul : $70+30 = 100$ en entrée.
Donc, par bilan matière,
on doit avoir 100 en sortie.
Ainsi le flux de sortie manquant :
 $s = 100 - 60 - 50 = -10$.
Cela n'est pas possible !



En utilisant la réconciliation de données par minimisation sous contraintes et en supposant par exemple que les flux i et e seraient plus incertains que les autres flux, alors on obtiendrait sûrement un résultat de ce type plus réaliste :



Pour la définition de la structure de la filière, je me suis inspiré de ce que le BASIC avait déjà produit. Cela dit j'ai dû la faire évoluer tout au long de la construction de l'AFM. Pour simplifier mon travail, j'ai traité d'une part la volaille et d'autre part le bétail avant de fusionner ces 2 filières.

Pour la collecte des données, l'essentiel des données de production, importations et exportations viennent de la base de donnée européenne ProdCom. C'est aussi à partir de cette nomenclatures (NACE Rév.2) que j'ai défini mes produits. Voici des exemple : *Carcasses de coqs et poules, non découpées, fraîches ou réfrigérées / Morceaux de porcins, jambons et épaules, congelés / saucisses, saucissons et produits similaires, de viande, d'abats ou de sang / ...* Pour la filière viande complète, 173 produits et 33 secteurs sont définis au total. J'ai également extrait des données des Statistiques Agricoles Annuelles de l'Agreste. On y retrouve les effectifs d'animaux finis. Les animaux finis sont les animaux destinés à l'abattage⁵. On trouve également des données de l'Agreste pour les imports/exports, taux de graisse et taux d'abats. Concernant les données de consommation, j'ai repris celles du BASIC calculées à partir du modèle PARCEL et des données de l'étude INCA3 menée par l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail).

Un second point important est l'unité choisie. Afin de pouvoir déterminer, de façon cohérente, les flux de matière au sein des filières animales, des coefficients ont été introduits pour tenir compte des actions menées sur les viandes (éviscération, transformation...) et ramener les volumes par rapport à une base commune : la Tonne Équivalent Carcasse ou TEC. Par équivalent carcasse, il faut comprendre animal saigné, décapité, écorché, éviscéré [7]. Par exemple, le coefficient TEC pour un gros bovins vivant est de 0,5. Ce qui signifie que si le bovin vivant pèse une tonne. À la pesée en sortie d'abattoir il pèsera 500kg. Inversement, le coefficient TEC pour un saucisson sera de 1,8. Donc pour produire 100kg de saucisson, 180 kg de carcasse seront nécessaire. Cela s'explique par des pertes de poids liées au désossage ou au séchage. Utiliser cette unité permet de ne pas prendre en considération dans le modèle tous les co-produits tels que le sang, les abats, la graisse, la peau, les cornes, ...

5 Il a été intéressant pour chaque filière de se familiariser avec un nouveau vocabulaire

Les données ProdCom étant données en kg, il a été nécessaire d'appliquer à chaque produit un coefficient de conversion TEC. Des tables de coefficients TEC données par l'Agreste existent. Le problème étant que l'Agreste utilise la Nomenclature Combinée 8 (NC8) alors que Eurostat utilise la Classification statistique de Produits associés par Activité (CPA ou NACE Rév.2). Il a donc fallu utiliser une table de conversion NC8/CPA. J'ai pour cela écrit un code python déterminant le bon coefficient TEC pour chaque produit. La Nomenclature NC8 étant plus précise que la norme CPA, certains produits avaient plusieurs coefficients TEC. J'ai pour ces produits, choisi d'utiliser une moyenne arithmétique des coefficients. Voici le logigramme de mon algorithme :

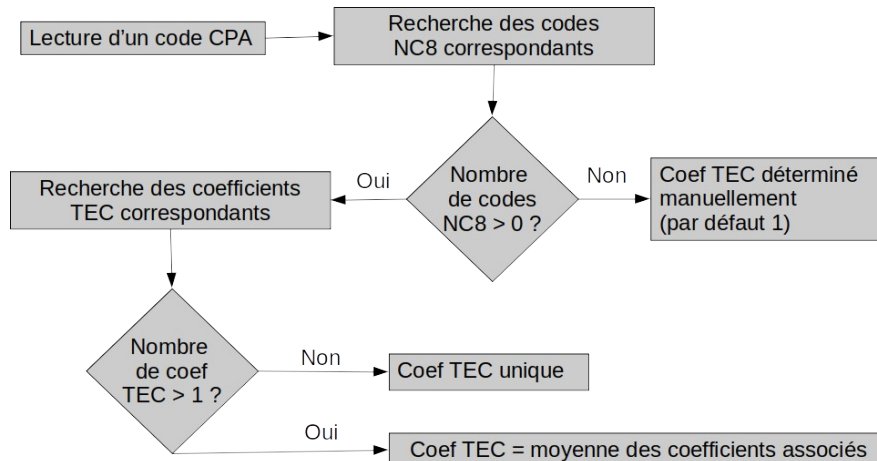


Figure 12: logigramme de l'algorithme de détermination des coefficients TEC

Après cette phase de collecte des données a suivi l'étape de réconciliation et d'analyse des résultats. J'ai dû itérer un grand nombre de fois pour résoudre des bugs, remanier l'organisation de la filière et enfin aboutir à une modélisation cohérente.

Voici quelques chiffres pour appréhender la taille du problème. Pour la version la plus aboutie (avec les co-produits), ce sont 179 produits, 33 secteurs et 53 contraintes qui ont été définis. La version nationale contient 905 variables dont 141 avec des données initiales. La version régionalisée contient 35055 variables dont 538 avec des données initiales.

Voici un exemple de remaniement de l'organisation d'une filière : les catégories Prodcom pour la viande peuvent être dans l'état « non découpé » : *Canards, oies et pintades non découpés*, ou dans l'état « découpé » : *Morceaux de canards, d'oies et de pintades*. Ma première analyse de la filière était la suivante :



Figure 13: schéma initial de la filière volaille

Cela posait problème au niveau de la réconciliation car les 3 flux ci dessus étaient très différents. Ils devaient s'écarter, compte tenu de leur incertitude, bien au delà de leur valeur initiale pour pouvoir être cohérents entre eux. Après quelques recherches [8] et réflexions, j'ai réorganisé la filière de cette manière :

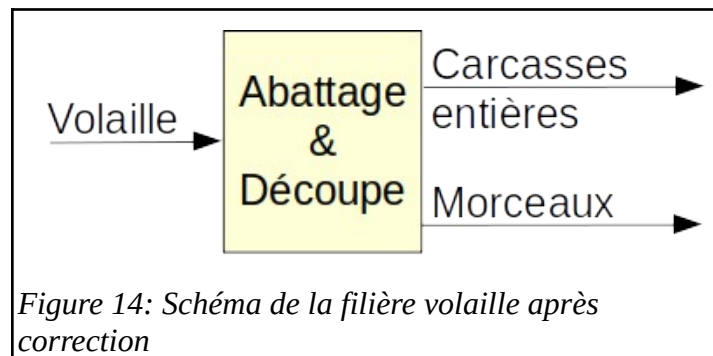


Figure 14: Schéma de la filière volaille après correction

J'ai compris que d'une part les activités d'abattage et de découpe sont regroupées mais aussi, qu'à l'issue des ces activités, on peut obtenir des morceaux de viande ou de la viande entière. Cela m'a permis de correctement réorganiser la filière et d'obtenir une bonne réconciliation. C'est à dire un écart suffisamment faible entre valeurs d'entrée et valeurs de sortie.

Enfin est venu le moment de réaliser la représentation graphique en diagramme de Sankey. J'ai donc, comme pour la réconciliation, utilisé l'outil mis à disposition par Terriflux. Le travail de représentation peut être assez long, compte tenu du nombre de flux et de labels à arranger. Voici une version du diagramme de Sankey agrégé :

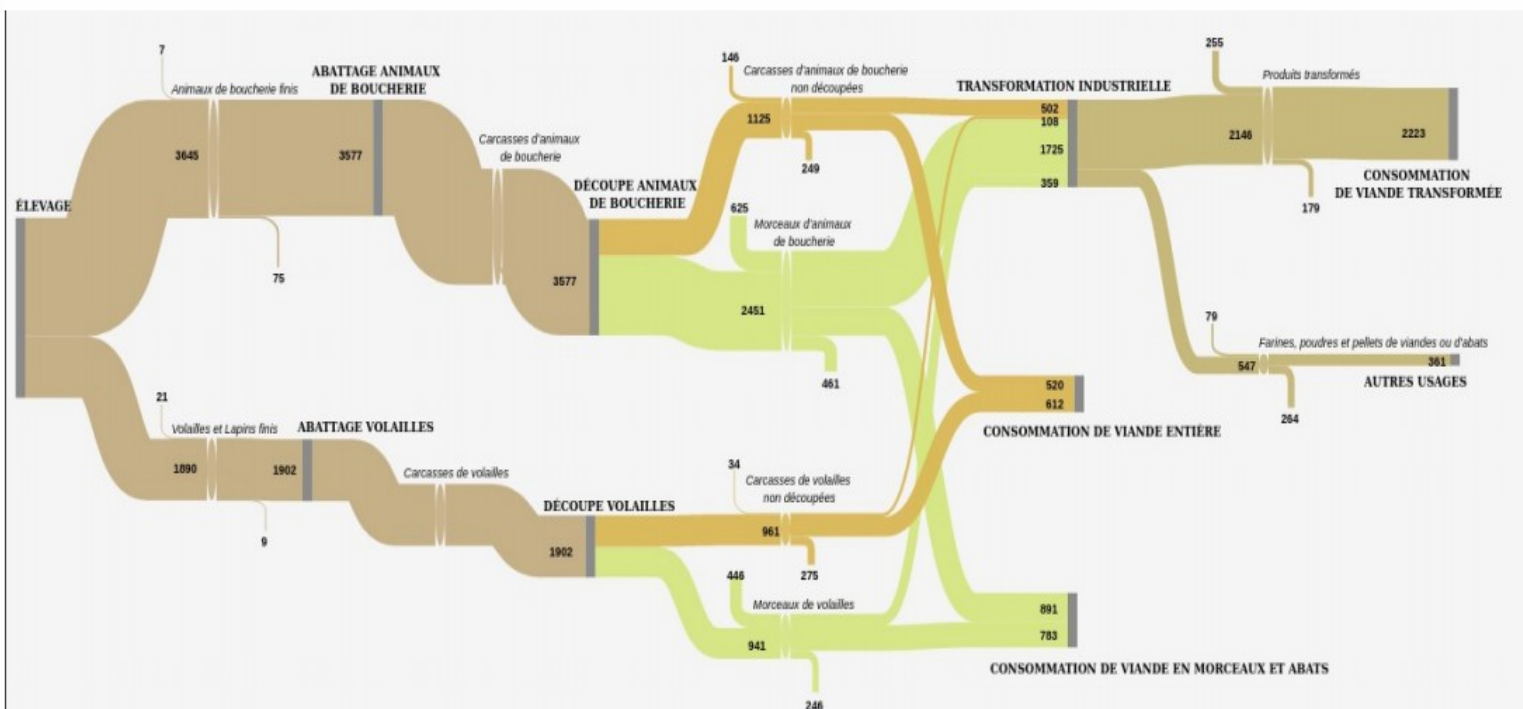


Figure 15: Diagramme de Sankey pour la filière viande simplifié

Concernant la confrontation des résultats aux experts de la filière, j'ai pu comparer la partie relative à la volaille de mon étude aux résultats d'une autre étude réalisée par l'ITAVI (organisme expert des filières avicoles). :

Table 2: Comparaison entre ma modélisation de la filière volaille et celle de l'ITAVI.

	Mes résultats (Mtec)	Résultats ITAVI (Mtec)	Écart (%)
Consommation	0,925	1,3	29
Export finaux	0,35	0,3	17
Transformation Industrielle	0,185	0,3	38
Production	1,2	1,1	9
Imports	0,43	0,6	28

Même si il y a des disparités entre nos résultats, on retrouve bien le même ordre de grandeur pour chaque secteur. Ne connaissant pas les détails de la méthodologie utilisée par l'ITAVI, il n'est pas possible de se livrer à une explication détaillée de ces différences de résultats. Elles s'expliquent probablement par des sources de données différentes, une méthode de réconciliation différente⁶, une année d'étude différente [9] (2015 pour moi contre 2018 pour l'ITAVI), un périmètre sur la volaille différent⁷. Au regard de tous ces paramètres et des résultats de l'ITAVI, je pense donc que nos résultats sont réalistes.

Plus tard dans le déroulement de mon stage, je suis revenu sur mon travail pour régionaliser mes résultats et créer une version sans coefficients TEC donc incluant les co-produits de la filière viande. Pour cette dernière version, je me suis aidé d'un article scientifique sur les co-produits [10]. De plus, hormis un plus grand nombre de produits et quelques conversions pour utiliser des tonnes et non des tonnes équivalent carcasse, il n'y pas eu de grands changements méthodologiques par rapport à ce que j'ai décrit précédemment.

Pour la régionalisation : grâce à un système d'extraction et de mise en forme des données que je décrirai dans la partie sur les codes pythons, j'ai pu récupérer les données d'imports/exports interrégionaux. Pour les données de production, j'ai utilisé ce que nous appelons des « proxys ». L'idée est de ventiler un flux national à l'échelle régionale en utilisant un indicateur pertinent. Pour ventiler les données de production, on utilise souvent les codes d'Activité Principale Exercée (APE). Nous disposons de bases de données recensant le nombre d'emplois salariés par régions et par code APE. Par exemple pour ventiler la production de *Viande non préparée de boucherie*, j'ai utilisé le nombre d'emplois salariés pour la *Transformation et conservation de la viande de boucherie* correspondant au code APE 1011Z. J'ai pu procéder de la même manière pour ventiler la production de *Viande non préparée de volaille* ou encore de *Produits transformés*. De la même manière, j'ai ventilé les données de consommation en utilisant les données de démographie régionale. Sur les diagrammes de Sankey régionaux suivant, on retrouve bien le fait que la Bretagne (figure 17) est une grande productrice de viande de boucherie ou encore que la région Île de France (figure 18) est une grande importatrice et consommatrice de viande.

6 L'ITAVI à procédé à une réconciliation par bilan donc moins robuste.

7 Par exemple prise en compte ou non de la volaille de réforme comme les anciennes poules pondeuses

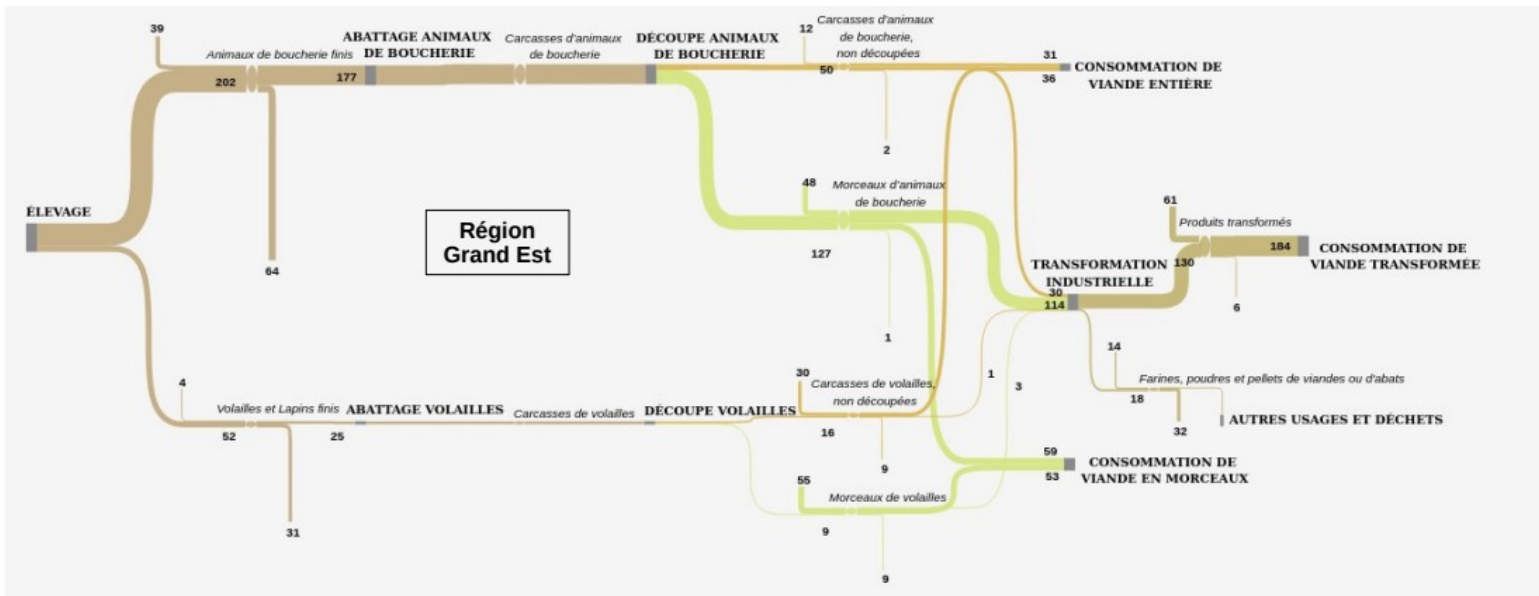


Figure 16: Diagramme de Sankey de la filière viande pour la région Grand Est (en kt)

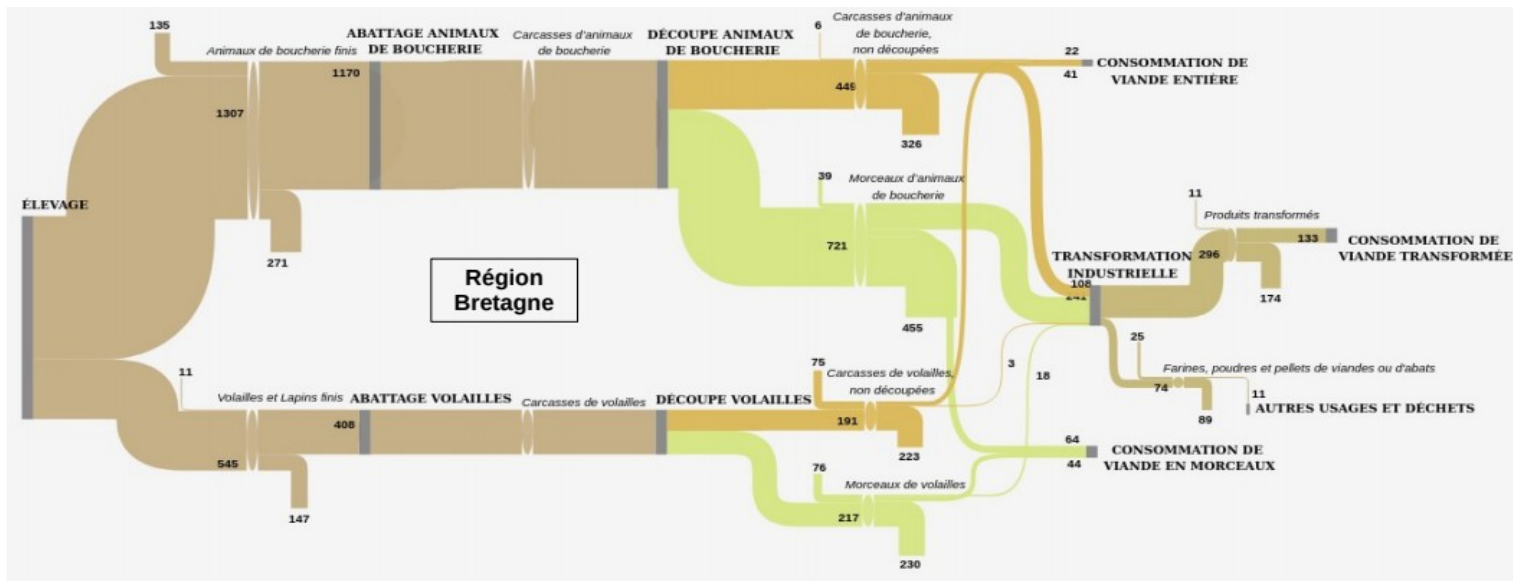


Figure 17: Diagramme de Sankey de la filière viande pour la région Bretagne (en kt)

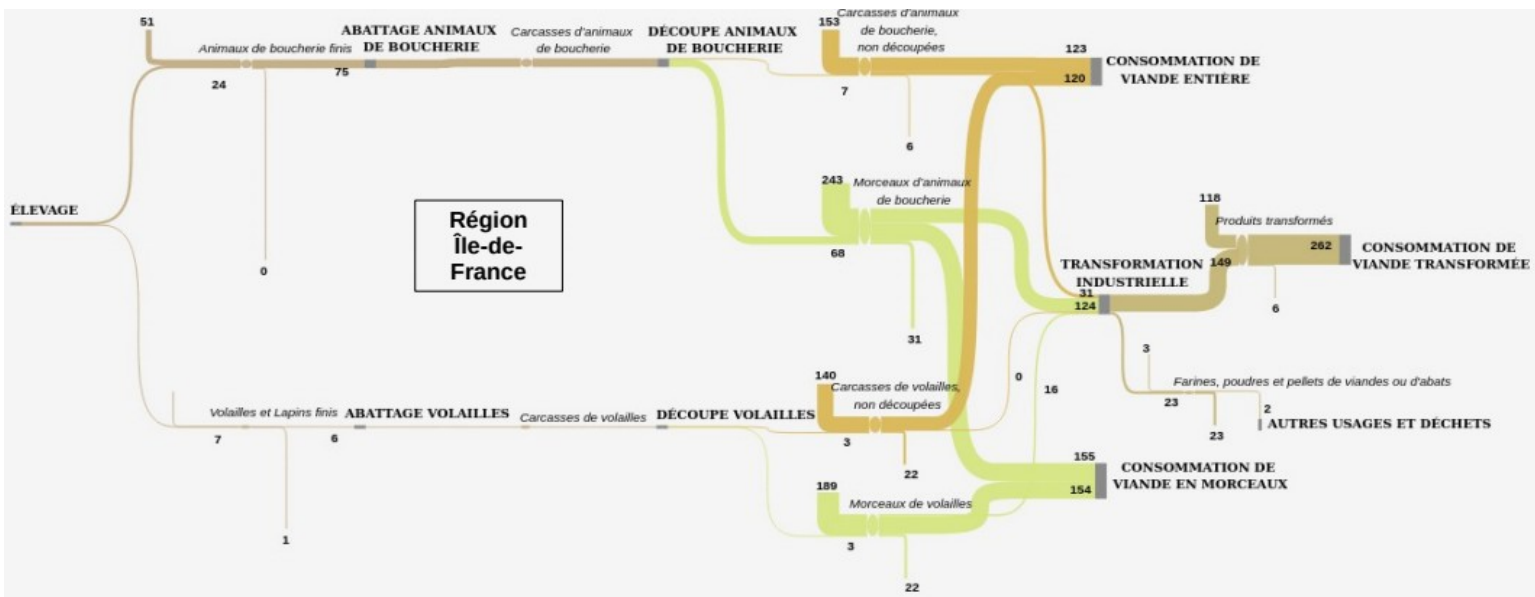


Figure 18 : Diagramme de Sankey de la filière viande pour la région Île de France (en kt)

3.b) Filière Lait

Ma deuxième étude a porté sur la filière lait. Pour cela, je suis reparti du travail d'une thésarde de l'équipe, Michela Bevione. Elle avait déjà réalisé une AFM de la filière lait au niveau national sur laquelle je me suis basé. J'ai commencé par comparer les résultats de l'AFM à ceux d'une étude produite par l'institut de l'élevage (l'Idèle) intitulée « Où va le lait ? » [11]. Seule la consommation des principaux produits de l'industrie laitière ont été comparés entre les deux études car celle de l'Idèle ne concerne pas les secteurs de production et de transformation :

Table 3 : Comparaison entre la modélisation de Michela et celle de l'idèle pour la filière lait.

	Résultats Michela (kt)	Résultats Idèle (kt)	Écart (%)
Lait conditionné	2889,34	3618,7	25,24
Beurre	529,26	494,78	6,51
Crème conditionnée	406,52	376,61	7,36
Fromages	1604,82	1573,5	1,95
Lait en poudre	265,03	302,35	14,08
Yaourt et desserts lactés	1946,37	2178,45	11,92

On remarque une similitude dans les résultats. L'écart plus grand pour le lait s'explique par la difficulté de faire concorder la nomenclature de l'Idèle avec celle de Michela. En effet l'Idèle distingue le lait vrac du lait conditionné. Plus généralement les écarts de résultat s'expliquent par une année d'étude différente (2010 pour Michela contre 2017 pour l'Idèle) ou encore une méthodologie de réconciliation différente⁸. On peut donc considérer que les 2 modèles sont plutôt concordants.

A la suite de cette comparaison avec les résultats de l'Idèle, j'ai dû procéder à des réajustements et à des améliorations du modèle. J'ai par exemple ventilé le produit fromage en différents types (pâte filée, molle, persillée, ...), ou encore j'ai distingué différents types de consommations (par les foyers, la restauration et les industries agro-alimentaires). J'ai aussi par la suite régionalisé le modèle.

Cette étude a surtout été longue pour la phase de représentation graphique en diagramme de Sankey. J'ai fait remonter beaucoup d'erreurs logiciel à Julien de Terriflux pour qu'il puisse les corriger et améliorer ses outils. Nous avons étroitement collaboré afin d'aboutir à un niveau de fonctionnalité et une représentation graphique satisfaisante.

Une des particularités de cette étude est de pouvoir convertir les résultats en matière sèche, matière grasse ou matière protéique. On peut observer que l'essentiel de la matière grasse (*figure 20*) est dans le beurre ou le fromage. L'essentiel de la matière protéique (*figure 21*) est dans le lait conditionné, la poudre de lait ou le fromage. Ou encore, le lactosérum ne contient que peu de matière sèche (*figure 22*) en comparaison avec la matière totale (*figure 19*).

⁸ Par bilan pour l'Idèle ce qui aboutit à certains flux négatifs

Matière Totale

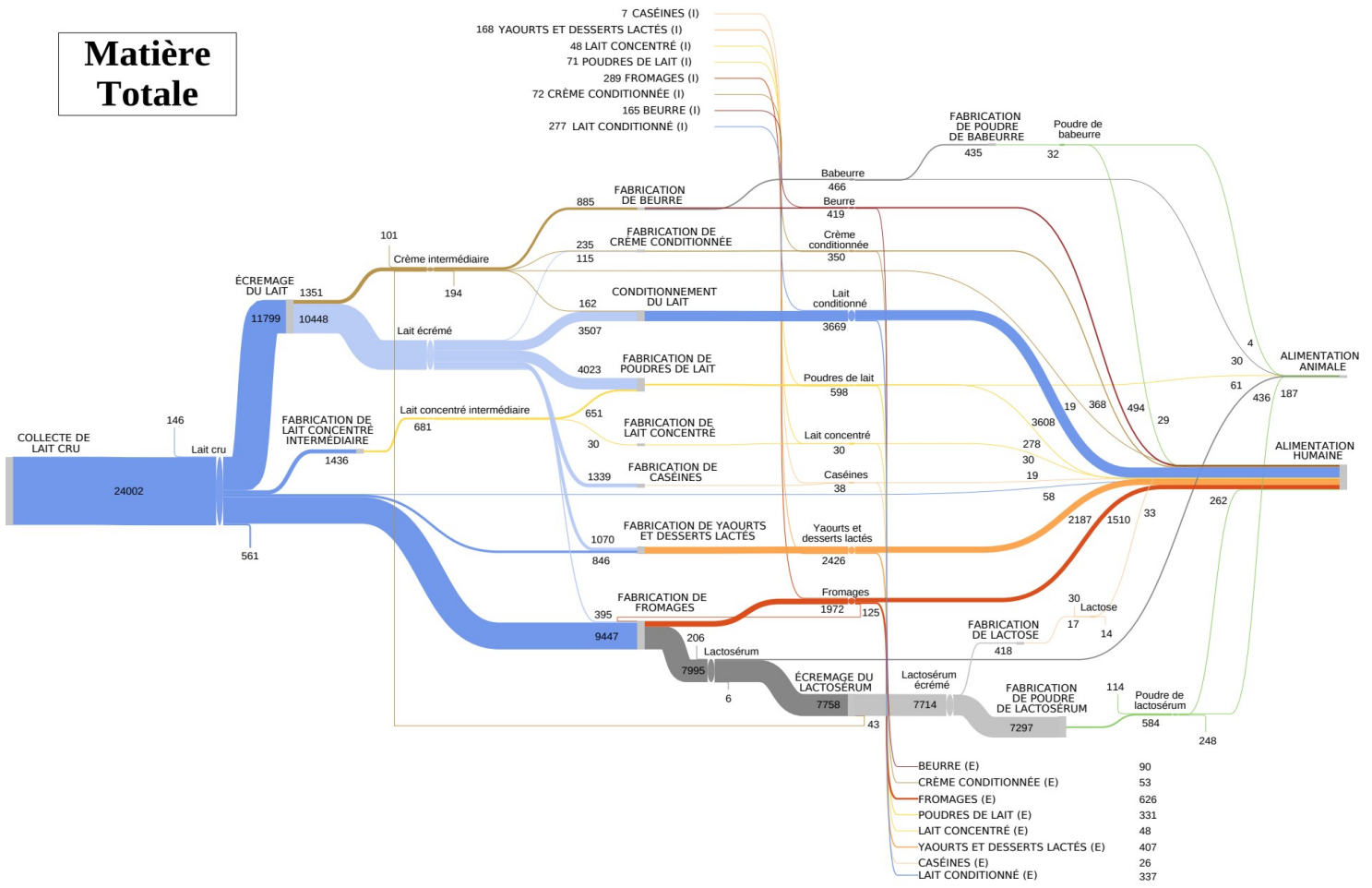


Figure 19: Diagramme de Sankey de la filière lait pour la matière totale (en kt)

Matière Grasse

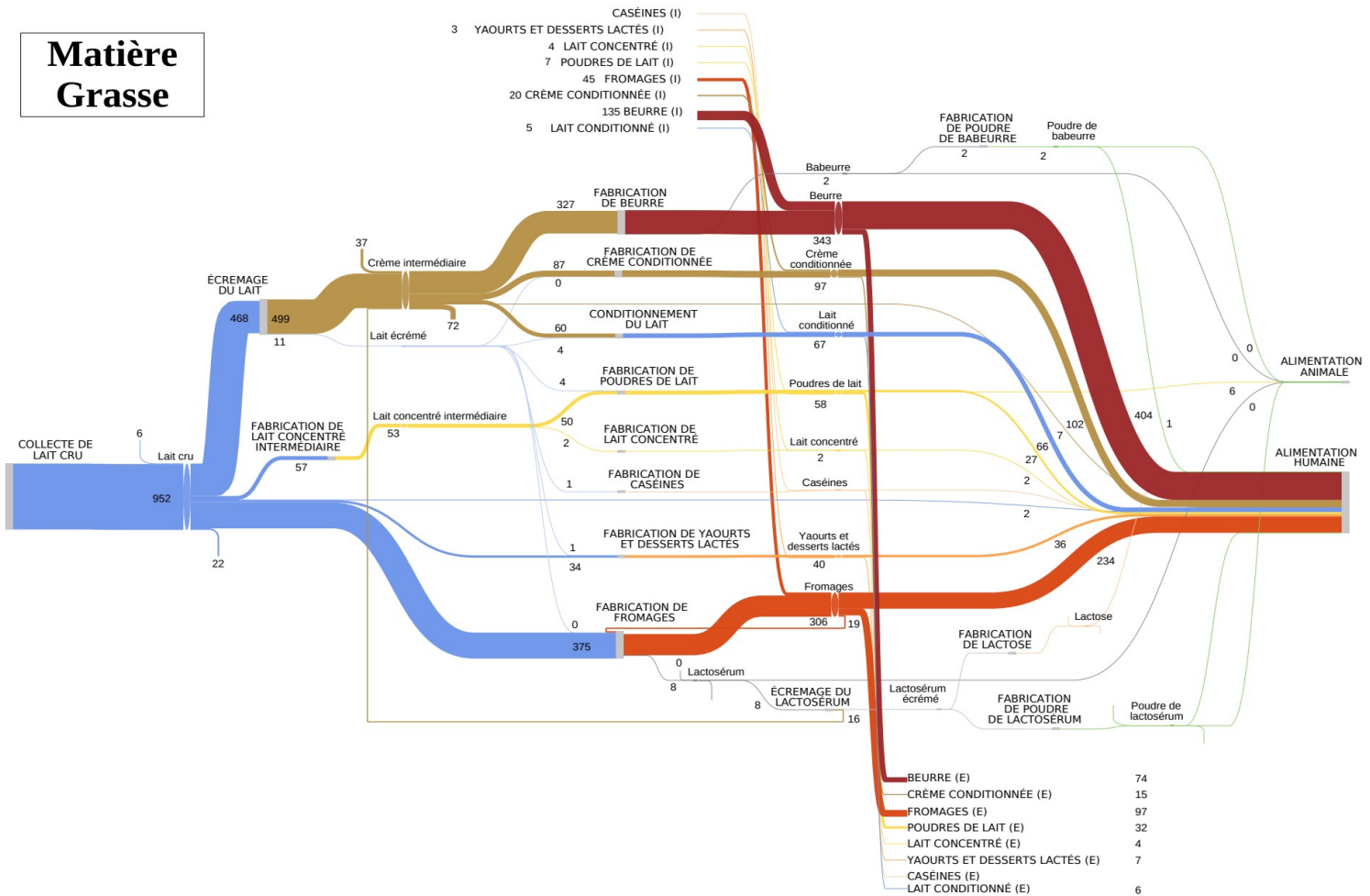


Figure 20: Diagramme de Sankey de la filière lait pour la matière grasse (en kt)

Matière Protéique

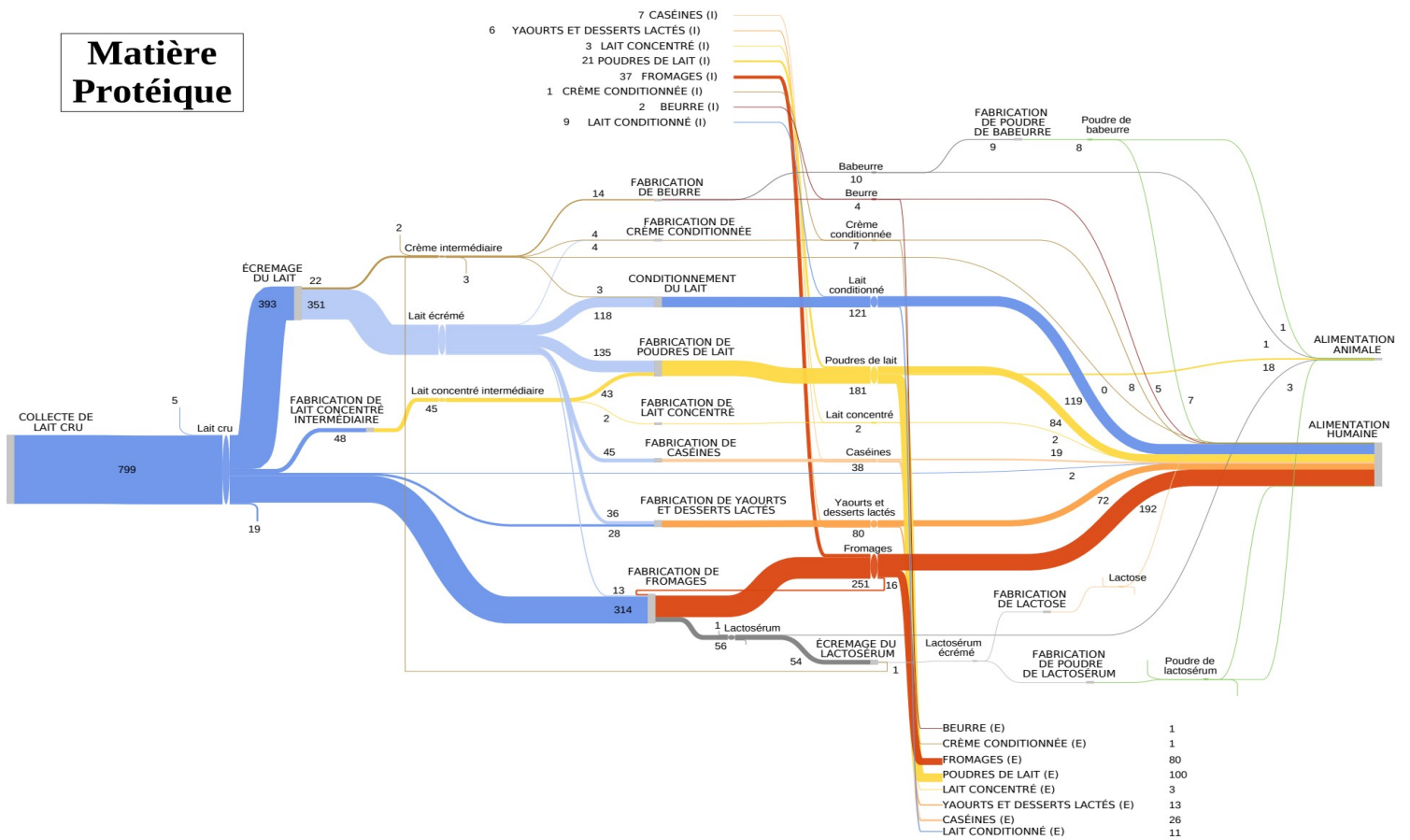


Figure 21: Diagramme de Sankey de la filière lait pour la matière protéique (en kt)

Matière Sèche

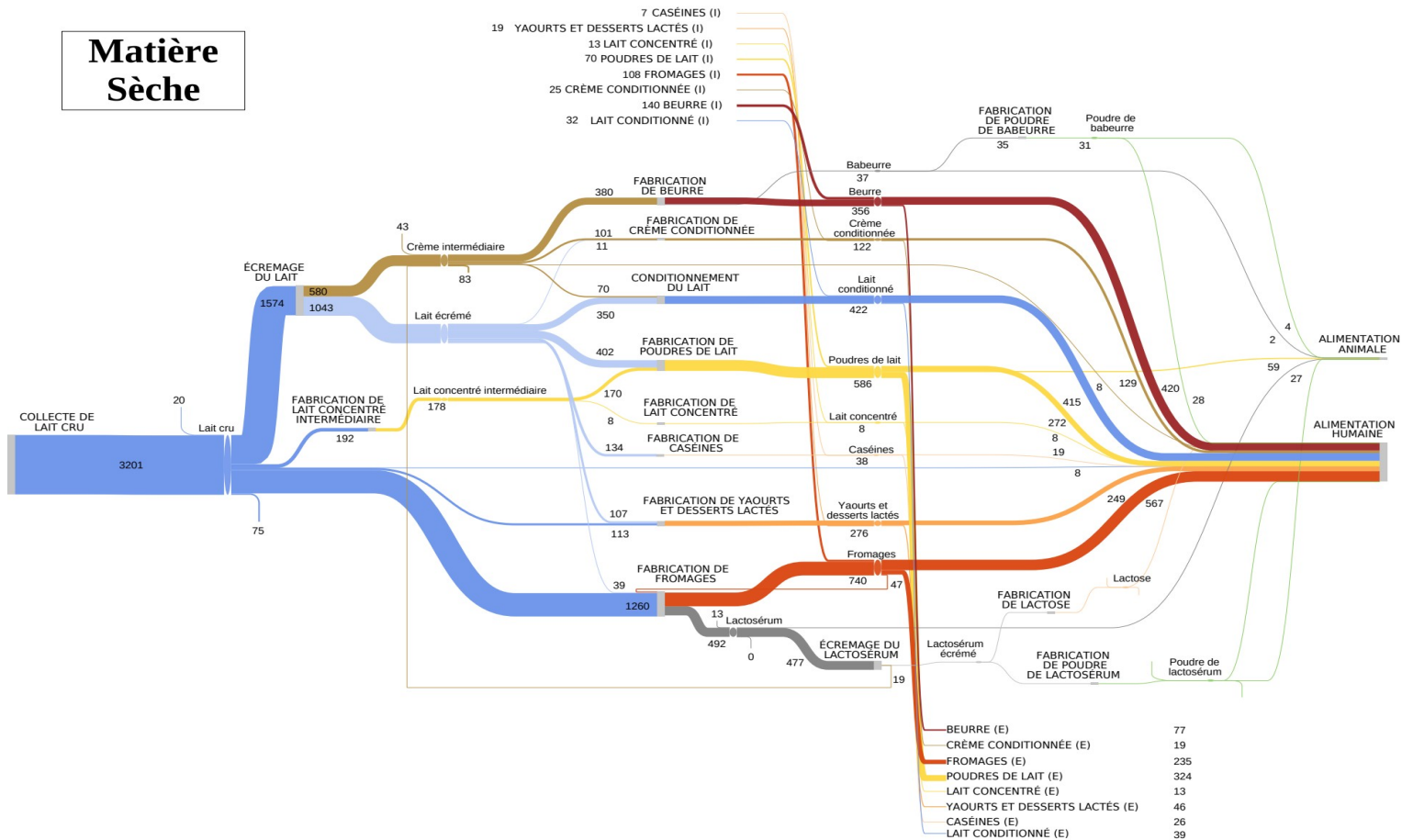


Figure 22: Diagramme de Sankey de la filière lait pour la matière sèche (en kt)

Ce travail, ainsi que celui concernant la viande, a pu servir à une autre stagiaire, Ophélie, travaillant sur les AFM au Centre International de Recherche sur l'Environnement et le Développement (CIRED). Elle travaille spécifiquement sur les questions d'import-export et de leurs empreintes environnementales associées. Elle cherche à répondre aux questions suivantes : Pourquoi la France importe 165 kt de beurre et en exporte 90 kt ? Dans quelle mesure est il possible de réduire ces flux ? Quel serait le bénéfice à en tirer en terme de réduction des émissions de gaz à effet de serre liées au transport de ces marchandises ? J'ai apprécié de pouvoir partager mon travail afin qu'il soit utile à d'autres personnes, ainsi que de découvrir le travail réalisé par d'autres laboratoires de recherche [12].

En conclusion, voici quelques chiffres concernant la filière lait. C'est 45 produits, 47 secteurs et 186 contraintes qui ont été défini. La version nationale contient 636 variables dont 96 avec des données initiales. La version régionale contient 16938 variables dont 688 avec des données initiales.

3.c) Filière Paille

L'étude de la filière paille a été très intéressante à réaliser. Bien qu'utile pour le projet Scalable, nous l'avons réalisée suite à une demande de l'ADEME pour la Bourgogne Franche Comté (BFC) sur dix céréales en couvrant dix années. Cette étude est pilotée par le Cluster Robins qui cherche à développer l'emploi de produits biosourcés dans les constructions. Et donc en particulier l'utilisation de la paille pour la construction. Cette étude est réalisée par Terriflux pour la partie AFM et par l'entreprise Accort Paille pour son expertise sur la paille de construction. En effet Accort Paille accompagne le développement économique des filières paille régionales et réalise de l'assistance à maîtrise d'ouvrage pour des projets utilisant la paille de construction. Le but de cette étude est de quantifier la ressource en paille en BFC, de structurer cette filière et de limiter les cas de conflit d'usage. Pour réaliser cette étude, Julien s'est occupé de collecter les données [13][14] tandis que je me suis occupé de la modélisation de la filière et de l'intégration des données au modèle.

- Concernant la structure de la filière paille, elle se divise en 3 parties : une première de production au niveau des champs, une deuxième de logistique comprenant le stockage et le transport routier, enfin une dernière partie de consommation, avec la paille utilisée pour la litière d'élevage, la construction, la production d'énergie, le paillage ou encore les champignonnières.

- Concernant la collecte des données, cette étude a été assez compliquée car il n'existe que peu de données précises sur la filière paille, et encore moins à l'échelle régionale. Ainsi, nous avons estimé les quantités de paille utilisées par le biais de différents modèles. Ces modèles incluent de nombreux paramètres qu'il est possible de faire varier en fonction des hypothèses prises sur les différentes étapes de la filière.

- Concernant la production de paille au niveau des champs, nous avons utilisé les données de production et de collecte des grains de céréale ainsi que de production de paille à l'échelle nationale fournies par l'Agreste. Prenons l'exemple du blé. Lorsqu'il est moissonné, une partie de la tige reste au sol : c'est le **chaume**. Le reste est séparé entre **paille** et **grains** lors du battage. Enfin, des débris de paille broyés sont perdus lors des différentes opérations : c'est la **menue paille**. En connaissant la répartition en matière sèche entre chaume/menue paille/paille et grains, on peut estimer la quantité de paille produite en fonction de la quantité de grains récoltée. Cette répartition est approximative car elle dépend de la hauteur de coupe ou encore de la longueur des tiges de blé. Vient ensuite le premier potentiel conflit d'usage : le retour au sol. Pour préserver une bonne fertilité du sol, en particulier pour l'agriculture biologique en plein essor, il est important de laisser une partie de la paille moissonnée au sol pour le nourrir. Il existe certaines recommandations (environ 30 à 50% de la paille moissonnée) mais en pratique, ce serait le cours du prix de la paille qui déterminerait ce taux de retour au sol. Une fois la quantité de paille retournée au sol évaluée, on peut donc estimer la quantité de paille théoriquement disponible pour d'autres usages.

- Concernant la partie logistique. Une première inconnue est la quantité de paille stockée par les céréaliers en polyculture pour leurs élevages (estimée à 30%). Une deuxième inconnue est la quantité de paille perdue lors du stockage ou du transport. Ensuite, vient la question des échanges interrégionaux et internationaux. La base de donnée SitraM les recense sous l'intitulé : *Paille, foin, balles de céréales - Plantes fourragères*. En croisant cette information avec des experts locaux de la filière, les échanges de plantes fourragères seraient minimes. Nous avons donc considéré les données d'imports exports comme des données brutes et non comme des bornes maximales.

- Enfin, concernant la partie sur la consommation de la paille par les différents secteurs d'activité. Le premier secteur de consommation est la paille utilisée en litière d'élevage. Nous avons donc mis en place un modèle d'utilisation de la paille pour l'élevage des bovins, porcins, ovins, caprins, équins et de volaille. Pour cela nous nous sommes aidé d'un modèle pré-existant mis en place par la Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt (DRAAF) de BFC. Nous avons dû enquêter sur les fondements de ce modèle pour le comprendre et le mettre à jour. Il a fallu pour chaque type d'élevage évaluer les effectifs à partir des données de l'Agreste ou des dires d'expert, mais aussi évaluer les pratiques d'élevage avec par exemple le type de stabulation (fumier ou lisier, temps passé en étable, ...), ou encore évaluer la quantité de paille utilisée en litière par surface ou le taux de paille contenu dans le fumier. Nous avons également comparé et ajusté nos résultats avec ceux du modèle Elba de l'institut du végétal Arvalis. En effet, ce dernier fait référence pour estimer la production de fumier et de lisier. Nous avons donc choisi d'utiliser les résultats de ce modèle pour mieux calibrer le notre.

Table 4:Données du modèle Elba pour la Bourgogne Franche-Comté

	fumier bovins	lisier bovins	fumier ovins caprins	fumier porcins	lisier porcins	fumier équidés	fumier volailles
Données ELBA (t)	6 913 995	1 014 703	105 039	20 876	313 855	217 684	79 934

Les autres consommations de paille, principalement la construction et l'énergie, sont négligeables en comparaison avec la litière d'élevage. Nous avons tout de même estimé l'utilisation de la paille pour ces secteurs avec le peu de données disponibles. Par exemple nous avons estimé l'utilisation de paille pour la construction en BFC à partir d'une donnée de l'institut Nomadéis qui indique qu'environ 4600 tonnes de bottes de pailles sont valorisées dans la construction chaque année en France. Nous nous sommes également aidés d'échanges avec une stagiaire de l'école AgroParisTech, Victoria, réalisant elle aussi un modèle sur la filière paille, pour estimer les consommations de paille par secteur.

La réconciliation des données pour cette AFM est particulière car les données introduites sont paramétrables. Ainsi la réconciliation nous a permis d'affiner nos modèles et de mettre en lumière des jeux de paramètres compatibles. Voici quelques chiffres permettant d'illustrer la complexité du modèle : 96 produits, 41 secteurs et 113 contraintes ont été définis. Le fichier contient 428 variables dont 198 avec des données initiales.

D'autre part les échanges avec les experts de la filière ont été très intéressants lors de cette étude. J'ai assisté à plusieurs réunions avec le Cluster Robins et Accort Paille qui prennent part à cette étude. Il était intéressant de constater que pour eux, cette étude est stratégique. Elle permet de montrer que la filière paille construction a sa place et ne menace pas les autres secteurs. J'ai aussi participé à la réunion de présentation des résultats de notre étude avec beaucoup d'acteurs de la filière paille en BFC. Il a été intéressant de constater que nos résultats ont permis de poser une base scientifique sur laquelle discuter. Notre étude a permis d'évacuer des peurs que certains acteurs pouvaient avoir sur la concurrence pour l'utilisation de la paille et a ouvert la voie à de possibles négociations pour structurer cette filière en BFC.

Sur les diagrammes de Sankey, il est intéressant de comparer les années 2015 et 2016. L'année 2016 a eu une récolte de grains bien inférieure à l'année 2015 à cause de mauvaises conditions climatiques (2970 kt de grains en 2016 contre 4570 kt de grains en 2015). Cela a induit inévitablement une baisse de la production de paille. Du côté des élevages, le cheptel variant peu d'une année sur l'autre, la consommation de paille litière reste quasiment constante (1220 kt de paille litière en 2015 contre 1200 kt de paille litière en 2016). Ainsi la conséquence de cette mauvaise récolte a été une tension sur la disponibilité de paille pour les autres usages. En effet, elle est matérialisée dans les graphiques suivants par l'appellation GDD pour Gisement Durablement Disponible. On remarque qu'en 2016 elle est bien moindre qu'en 2015 (610kt de paille durablement disponible en 2015 contre 0kt de paille durablement disponible en 2016). Pour améliorer le modèle, il faudrait prendre en compte l'évolution du comportement des éleveurs en fonction du prix de la paille. Car la paille se faisant plus rare en 2016, il est probable que les prix ont augmenté et donc que les éleveurs ont réduit leur consommation de paille litière. Il faudrait aussi un modèle plus complet et plus robuste pour les autres usages, car la consommation de paille n'a pas pu réellement tomber à 0.

2015

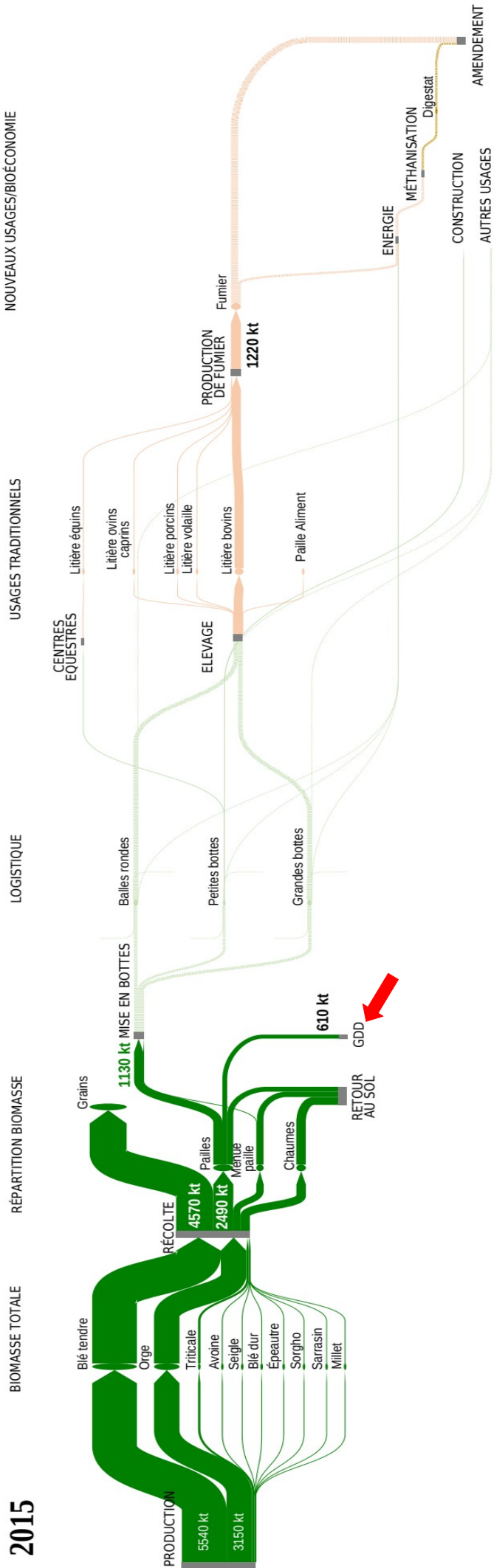


Figure 24: Diagramme de Sankey pour la filière Paille en BFC en 2015 (en kt)

2016

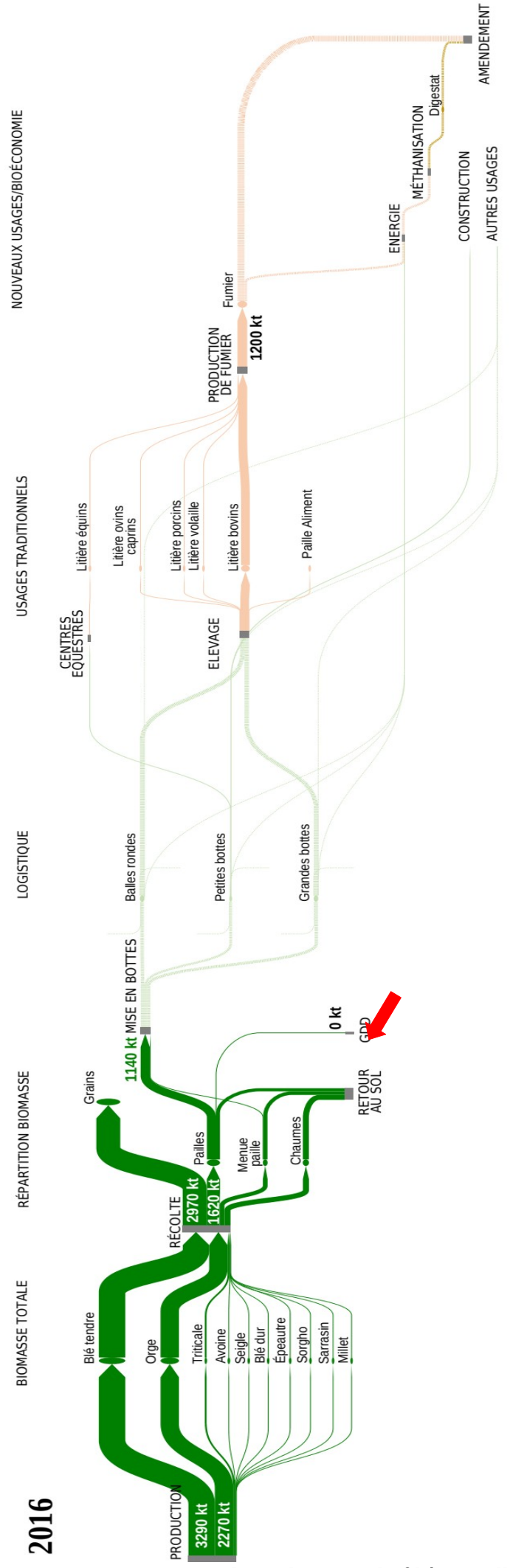


Figure 23 : Diagramme de Sankey pour la filière Paille en BFC en 2016 (en kt)

3.d) Codes Pythons

Une autre partie de mon stage aura été le développement de codes pythons. J'ai travaillé principalement sur 2 codes Jupyter Notebook permettant l'extraction et la mise en forme de données. Les données brutes étant des fichiers issus de la banque de données sur le transport SitraM, cette dernière se divise entre les données relatives au transport national et les données relatives au transport international.

Pour le transport national, les données viennent du Transport Routier de Marchandises et des Voix Navigables de France (TRM-VNF). Les données TRM sont issues d'une enquête statistique : tous les ans environ 80 000 camions sont tirés au hasard et sondés sur l'activité du véhicule pendant une semaine déterminée. Il est donc primordial de prendre en compte le nombre d'observations pour évaluer la fiabilité d'une donnée de flux. Les données VNF quant à elles sont basées sur les déclarations faites aux péages par les bateliers. Les données de fret ferroviaire ne sont plus disponibles depuis 2007 suite à un décret limitant pour l'État l'accès à ces informations.

Pour le transport international, les données sont issues des déclarations faites aux douanes. Ce sont donc tous les transports de marchandise faisant l'objet du commerce extérieur de la France qui sont recensés.

Le premier Jupyter Notebook vient lire ces fichiers d'entrée pour les mettre en forme proprement : par exemple remplacer certaines chaînes de caractère par des entiers si la colonne ne doit comporter que des entiers. Il vient aussi rajouter les codes et noms des régions associés aux départements. Enfin il stocke toutes ces données correctement formatées en local dans un fichier de type CSV et/ou dans une base de donnée PostgreSQL. Le code existant était déjà fonctionnel pour l'année 2017 mais non fonctionnel pour les autres années. Mon travail a été de reprendre ce code afin de le rendre fonctionnel pour toutes les années à partir de 2010. J'ai dû opérer un traitement différencié pour chaque année car chaque fichier a ses particularités (par exemple 2 groupes de colonnes qui sont inversés pour certaines années, ou une colonne en plus pour les années avant 2014).

Le deuxième Jupyter Notebook vient sélectionner les données nous intéressant dans la base de donnée PostgreSQL préalablement construite. Pour cela, une requête SQL contenant différents paramètres tels que les produits ou l'année d'intérêt est envoyée, puis le tableau renvoyé par la requête subit différentes opérations. Il est tout d'abord traité pour agréger les flux par régions et grouper les différents modes de transports. Ensuite, pour une région donnée, on détaille les exports par régions de destination mais on agrège toutes les régions d'import sous l'appellation « Autres régions françaises » (sinon tous les flux seront doublés). On calcule éventuellement une moyenne entre les différentes années et on calcule l'incertitude [15] en fonction du nombre d'observations pour les flux TRM. Enfin, on écrit les flux directement dans le formalisme de notre fichier Excel pour pouvoir intégrer rapidement les nouvelles données à notre modèle. Mon travail a été de reprendre ce code pour l'améliorer. J'ai tout d'abord développé l'option permettant d'extraire les données, non plus depuis la base de donnée PostgreSQL, mais depuis le fichier CSV généré par le précédent code. Le code déjà existant renvoyait les données pour toutes les régions françaises. J'ai développé une première option permettant d'extraire les données pour une ou plusieurs régions d'intérêt.

Dans la même lignée, j'ai développé une seconde option afin d'extraire les données pour un ou plusieurs départements d'intérêt. J'ai également rafraîchi l'ensemble du code, pour supprimer quelques bugs, le rendre plus lisible et plus modulaire. J'ai aussi testé ce code sur différentes régions et différentes filières afin de vérifier son bon fonctionnement et de supprimer des bugs ce qui n'avait pas été fait auparavant. Je me suis aussi assuré qu'il n'y avait pas d'erreurs au cours des différentes opérations en refaisant les calculs manuellement depuis les fichiers d'origine.

4) Conclusion

L'objectif initial de mon stage était de contribuer à la réalisation d'Analyses de Flux de Matières de filières agricoles au niveau de la France et des régions françaises. Je pense qu'au regard des études que j'ai menées, j'ai bien atteint l'objectif initialement donné. Je pense aussi avoir bien conduit mon stage car j'ai pu avancer dans mon travail de manière progressive et régulière. J'ai également pu constater une progression de mes compétences sur la durée du stage.

J'ai toutefois terminé mon stage avec des sentiments partagés. Je pense avoir surtout apprécié ce qui m'a le plus manqué : les contacts humains. Toutefois, j'ai pu apprécier l'esprit d'équipe et de collaboration, même si une grande partie du travail s'est faite à distance. A mon sens, comprendre la portée et l'intérêt de ce sur quoi l'on travail ne peut se faire qu'en interaction avec les autres. Ainsi le télétravail a pu être parfois assez difficile car cela demande beaucoup d'auto-motivation. Travailler des journées entières sur des tableaux Excel peut être assez déprimant si on en oublie la raison pour laquelle nous travaillons sur ces tableaux ! J'ai tout de même apprécié la grande liberté que confère le télétravail en libérant le travail de sa contrainte géographique bien que j'eusse préféré passer une plus grande partie de mon stage à l'INRIA.

Il reste encore beaucoup de travail à accomplir pour poursuivre le projet SCALABLE. Dans la lignée de mon stage, il faudra poursuivre les AFM pour d'autres filières de biomasses d'origine agricole comme les fruits et les légumes, le vin ou encore les pommes de terre. Il sera également nécessaire de réaliser plusieurs itérations avec les experts des filières pour améliorer les modèles. Pour cela, il est prévu de créer un référentiel partagé entre le monde académique et les acteurs des différentes filières pour s'accorder sur les différentes structures de filière et définitions de produits/secteurs. Une autre étape importante sera de coupler les AFM avec les enjeux socio-économiques (emplois et valeur ajoutée sur l'ensemble de la filière) et avec les pressions environnementales. À l'échelle du PAIT, une analyse organisationnelle précise des acteurs des filières étudiées sera menée.

Enfin, tout ces travaux amèneront des conclusions en terme de vulnérabilités matérielles et organisationnelles des territoires par rapport à leurs filières agricoles. Ces dernières permettront de répondre de façon pertinente à des questions relatives à la relocalisation telle que : Que gagne-t-on et que perd-on à relocaliser ? Quels sont les acteurs gagnant et perdant ? Quel impact sur l'environnement ? L'emploi ?

Il restera aussi de nombreuses perspectives d'améliorations des outils proposés par Terriflux sur lesquels Julien travaille activement : par exemple pouvoir visualiser sur les graphiques de Sankey les flux qui ont le plus grand écart-type entre leur valeur brute et leur valeur réconciliée ou encore visualiser les pressions environnementales.

Je pense aussi avoir eu un bon aperçu du monde de la recherche et du quotidien des chercheurs. J'ai été assez admiratif devant la détermination et la passion des membres de cette équipe pour leurs travaux. Le travail de recherche est un travail exigeant, rigoureux et de longue haleine. J'ai beaucoup apprécié de travailler avec Jean-Yves, Julien, l'équipe et nos collaborateurs. Il est certain que cette expérience m'aura enrichi personnellement !

5) Bibliographie

- [1] Jean-Yves Courtonne, 2016, Évaluation environnementale de territoires à travers l'analyse de filières.
- [2] Courtonne et al, 2015, Downscaling material flow analysis: The case of the cereal supply chain in France.
- [3] AF-Filière, 2016, Analyse des flux des filières biomasse pour des stratégies régionales de bioéconomie
- [4] Barbier et al, 2018, Programme de recherche CECAM : Contenu énergétique et carbone de l'alimentation des ménages en France.
- [5] Amicarelli et al, 2021, Material flow analysis and sustainability of the Italian meat industry, Journal of Cleaner Production.
- [6] Courtonne et al, 2016, Environmental Pressures Embodied in the French Cereals Supply Chain.
- [7] IFIP, Composition des carcasses de porcs charcutiers
- [8] Wikipédia, Article sur les Abattoirs, <https://fr.wikipedia.org/wiki/Abattoir>
- [9] volaille-francaise.fr, 2018, chiffres-clés, donne des indications sur l'évolution du marché de la volaille.
<https://www.volaille-francaise.fr/wp-content/uploads/rapport2018chiffres-cles.pdf>
- [10] Laisse, 2018, L'efficacité nette de conversion des aliments par les animaux d'élevage : une nouvelle approche pour évaluer la contribution de l'élevage à l'alimentation humaine.
- [11] Idèle et France AgriMer, 2019, L'évolution des circuits de distribution : Où va le lait de vache collecté et transformé en France ?
- [12] Barbier et al, 2018, Contenu énergétique et carbone de l'alimentation des ménages en France
- [13] CHARPENTIER, 2019, Schéma régional Biomasse Bourgogne-Franche-Comté 2019-2023
- [14] IAU île-de-France, 2018, La paille : concurrences et complémentarités des usages du gisement agricole en Île-de-France
- [15] Courtonne, J-Y., Longaretti, P-Y., Dupré, D. 2018. Uncertainties of Domestic Road Freight Statistics: Insights for Regional Material Flow Studies. Journal of Industrial Ecology

6) Annexes

6.a) Liste des abréviations

- **ADEME** : Agence De l'Environnement et de la Maîtrise de l'Énergie. www.ademe.fr
- **AFM** : Analyses de Flux de Matières.
- **AGRESTE** : Service statistique du Ministère de l'Agriculture et de la Pêche. <https://agreste.agriculture.gouv.fr/>
- **ANSES** : Agence Nationale de Sécurité sanitaire de l'alimentation, de l'Environnement et du travail.
- **APE** : Activité Principale Exercée.
- **ARVALIS** : Institut du végétal. www.arvalis-infos.fr
- **AST** : Alternatives SocioTechniques.
- **AURA-EE** : Auvergne-Rhône-Alpes Énergie Environnement. www.auvergnerhonealpes-ee.fr
- **BASIC** : Bureau d'Analyse Sociétale pour une Information Citoyenne. <https://lebasic.com>
- **BFC** : région Bourgogne Franche-Comté.
- **BOA** : Biomasses d'Origine Agricole.
- **CIRED** : Centre International de Recherche sur l'Environnement et le Développement. www.centre-cired.fr
- **CPA (ou NACE Rév.2)** : Classification statistique de Produits associés par Activité.
- **CSV** : Comma-Separated Values file.
- **DOM** : Départements d'Outre-Mer.
- **DRAAF** : Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt.
- **ELBA** : Outil de référence pour l'évaluation de ressource en Biomasse Agricole en France. <https://elba.arvalis-ext.com>
- **GDD** : Gisement Durablement Disponible.
- **IDELE** : Institut de l'élevage. <https://idele.fr>
- **IFIP** : Institut du porc. www.ifip.asso.fr
- **INRAE** : Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement. www.inrae.fr
- **INRIA** : Institut National de Recherche en Informatique et en Automatique. www.inria.fr

- **INSEE** : Institut national de la statistique et des études économiques. www.insee.fr
- **ITAVI** : Organisme expert des filières avicoles. www.itavi.asso.fr
- **NC8** : Nomenclature Combinée à 8 chiffres.
- **PACA** : région Provence-Alpes Côte d’Azur.
- **PARCEL** : outil web Pour une Alimentation Résiliente Citoyenne Et Locale. <https://parcel-app.org/>
- **PAIT** : Plan Alimentaire interTerritorial « région alpine ».
- **PRODCOM** : PRODUCTION COMMUNAUTAIRE, statistiques européennes sur la production de bien manufacturés. <https://ec.europa.eu/eurostat/web/prodcom>
- **SITRAM** : Système d’Information sur le TRANSPORT de Marchandises.
- **STEEP** : Soutenabilité, Transition, Environnement, Économie biophysique et Politiques locales. <https://files.inria.fr/steep/>
- **TEC** : Tonne Équivalent Carcasse
- **TER** : Table Emploi Ressource.
- **TRM** : Transport Routier de Marchandises.
- **VNF** : Voies Navigables de France.

6.b) Archives Numériques

- Fichiers Excels

Vous trouverez en suivant le lien ci-dessous un dossier contenant les fichiers Excels que j'ai produit. La mention « nat » indique que le fichier concerne la filière à l'échelle nationale. La mention « reg » indique le fichier concerne la filière à l'échelle régionale. Voici un bref descriptif des onglets de base. Vous trouverez plus d'informations relatives à chaque filière dans l'onglet **READ_ME**.

- **Paramètres** : Paramètres pour la réconciliation et l'affichage des données en diagramme de Sankey
- **Dim produits** : Ensemble des produits pour une filière donnée.
- **Dim secteurs** : Ensemble des secteurs pour une filière donnée.
- **flux pouvant exister** : Ensemble des flux possibles entre produits et secteurs.
- **Conversions** : Table de conversion pour passer d'une unité à l'autre (visible dans le tooltip)
- **Données** : données et leurs incertitudes concernant les flux précédemment définis.
- **min max** : borne minimale ou maximale et leurs incertitudes pour un flux donné.
- **contraintes** : relations numériques entre plusieurs flux.

<https://drive.google.com/drive/folders/1iUikHqrgbUgghiyLqiKBIBdxDm-jA7gr?usp=sharing>

- Diagramme de Sankey

Vous trouverez de nombreux diagrammes de Sankey réalisés à l'aide des outils de Terriflux en vous rendant sur <https://terriflux.fr/portfolio>. Vous trouverez en particulier les diagrammes que j'ai réalisés pour la filière viande, lait et paille.

