

Nom : VIALANEIX	Prénom : Nathalie	Âge : 44
-----------------	-------------------	----------

## 1. INFORMATIONS COMPLÉMENTAIRES CONCERNANT LE PARCOURS SUR LES 5 DERNIÈRES ANNÉES

### Séjours à l'étranger

Juillet 2018 : Monash University, Melbourne, Australie (2 semaines, visite dans le laboratoire de Dianne Cook)

## 2. PROFIL D'ACTIVITÉ : Répartition de votre temps selon les activités

Taux d'activité	<b>100 % ETP</b>
Répartition des activités selon leur type	% du temps (rapporté à 100)
A. Production de connaissances	60
B. Expertise et mobilisation des connaissances	15
C. Formation par la recherche, formation initiale et continue	10
D. Animation ou direction de collectifs, de grands instruments, de ressources, de programmes ou de réseaux	15

*Note : Tableau basé sur une mesure de mon temps de travail par type d'activités réalisée entre janvier 2020 et juillet 2020.*

### Intitulé de la thématique de recherche et/ou des missions

Analyse et intégration de données omiques volumineuses et multi-échelles en appui à l'amélioration de l'élevage et de la santé humaine

## 3. ANALYSE DE L'ACTIVITÉ SUR LA PÉRIODE

*Remarque préliminaire :* Les mentions [#Topic] font référence aux domaines thématiques prioritaires et aux orientations de politique générale du document d'orientation [#inra2025] et montrent comment mes activités s'inscrivent dans les objectifs prioritaires d'INRAE.

### Positionnement général

J'ai été recrutée à INRAE en février 2014 sur un concours blanc CR1 blanc dont le projet portait sur les divers aspects de l'analyse et de l'inférence de réseaux pour des applications en biologie, notamment moléculaire. Depuis mon recrutement, je suis rattachée à l'unité MIAT (Mathématiques et Informatique Appliquées de Toulouse), dans l'équipe SaAB (Statistique et Algorithmique pour la Biologie).

---

Ce recrutement s'est accompagné d'une mobilité thématique caractérisée par une part croissante de travaux à l'interface avec la biologie moléculaire et des applications en élevage et santé humaine. D'un point de vue scientifique, ces cinq dernières années ont également été caractérisées par un élargissement du champ de mes recherches de l'utilisation des réseaux comme outil pour analyser les interactions multi-échelles au sein du vivant au développement de méthodes variées pour l'analyse et l'intégration de données omiques avec une attention particulier porté à leur volume (et en particulier, à la grande dimension). C'est ce projet que j'ai défendu lors du concours INRAE DR2 2018 et que je poursuis actuellement. En particulier, comme je le décrirai plus précisément dans ce rapport, j'ai surtout mis l'accent, durant ces dernières années, sur les approches d'analyses exploratoires. Les recrutements récents au sein de l'équipe SaAB (en particulier, Céline Brouard, recrutée en 2019 sur un poste de chargée de recherche en machine learning et Élise Maigné, recrutée en 2020 sur un poste d'ingénieure d'études en statistique) créent une dynamique favorable au développement et à la réalisation de ces objectifs. La thématique « machine learning pour la biologie prédictive », dans laquelle mon projet personnel s'inscrit naturellement, constitue d'ailleurs un des trois axes méthodologiques principaux du projet de l'équipe SaAB présenté à l'HCERES en 2019. Enfin, mes activités scientifiques sont fortement ancrées dans l'axe méthodologique AM2 (Développer la modélisation intégrative et prédictive de systèmes complexes, sur des échelles allant de la molécule au paysage agricole) et au au champ thématique CT1 (Bioinformatique et modélisation pour la biologie des systèmes et de synthèse) du département MathNum mais ont aussi des connexions avec les deux autres axes méthodologiques du département (« Maîtriser les enjeux liés aux données, de leur acquisition à l'extraction de connaissances » et « Concevoir à partir de modèles, pour mieux comprendre, prédire, contrôler ou piloter des systèmes »).

Je suis aussi engagée dans plusieurs missions complémentaires d'appui à la recherche, à la fois dans le domaine de l'encadrement et de la formation, dans le domaine de l'animation scientifique et également dans le domaine de la prise en charge de responsabilités administratives. Mes activités de formation, d'appui aux biologistes et de connexion avec la communauté statistique nationale et internationale s'inscrivent dans les orientations [\[#OpenScience\]](#) (connecter les infrastructures de recherche, participer au développement d'un modèle éditorial ouvert, produire des ressources pour l'enseignement numérique, anticiper les emplois et compétences pour la modélisation et l'analyse de données massives, ...) et [\[#OpenInra\]](#) (développer les partenariats à l'échelle locale, nationale et internationale, s'ouvrir vers l'enseignement supérieur, ...) du document [\[#INRA2025\]](#).

## Bilan scientifique

### Positionnement

Mon dossier de recherche présente la particularité d'être à la frontière entre plusieurs disciplines : mes travaux comportent à la fois des aspects :

- théoriques et méthodologiques dans le domaine de l'apprentissage statistique (méthodes neurones, méthodes à noyau et plus récemment, inférence de réseaux), y compris des actes de conférences internationales avec comité de lecture et publication ;
- applicatifs importants qui se traduisent par des publications dans des journaux situés en dehors de mon domaine principal de compétences (revues d'histoire, revues de biologie) ;
- de développement logiciel : j'ai conscience de la nécessité de développer des outils informatiques qui rendent accessibles facilement les méthodes statistiques développées et je participe donc activement au développement de packages R, logiciel incontournable en statistique et analyse de données. Je travaille dans un esprit de diffusion de mes productions logicielles, en utilisant les outils de versionnement (git), des entrepôts de code publics. Je suis également attachée à la généricité et à la reproductibilité du code produit (rapports RMarkdown, docker...).

---

L'originalité de mon positionnement scientifique en statistique réside dans une triple compétence en mathématiques (formation initiale via l'agrégation et doctorat en mathématiques), puis en *machine learning* et statistique computationnelle (à l'interface statistique / informatique), avec une expertise acquise sur les données omiques et les problèmes biologiques. Cette singularité est une compétence reconnue qui se matérialise par des invitations diverses (par exemple, récemment, conférence européenne sur R, eRum 2018, sur le thème « *Learning from (dis)similarity data* »), par des sollicitations à donner des cours de master sur le sujet, des participations à des jurys de recrutement (CR Inria, MCF, PR) ou de thèse ou ma participation à un groupe prospective d'experts auprès de l'INSMI (CNRS) sur l'intelligence artificielle. J'ai également été sollicitée pour participer à deux chapitres d'un livre de synthèse sur l'apprentissage (méthodes connexionnistes Nathalie VILLA-VIALANEIX et CANU 2018 et données massives Nathalie VILLA-VIALANEIX et ROSSI 2018) coordonné par la Société Française de Statistique.

### Thématiques scientifiques

Mes thématiques de recherche s'articulent autour de 4 axes principaux, pour lesquels je fournirai un bilan détaillé sur la période dans la suite de ce rapport.

- *machine learning* (apprentissage) et fouille de données avec un intérêt particulier pour les données non numériques et en particulier les graphes (réseaux)
- biostatistique et intégration de données omiques
- statistique fonctionnelle (Functional Data analysis : FDA)

Le point commun de mes travaux de recherche est l'utilisation et le développement de méthodes issues de l'apprentissage statistique (méthodes à noyaux, réseaux de neurones ...) dans le but d'analyser, de modéliser et d'intégrer des données complexes et hétérogènes (données fonctionnelles, données relationnelles, données de comptage...).

Une part non négligeable de mes publications est à l'interface entre deux ou trois de ces thématiques et l'analyse de l'évolution de mes activités sur la période de l'évaluation montre une évolution dans leurs importances respectives : si la thématique *machine learning* est présente de manière constante, l'analyse des données fonctionnelles tend à devenir une part secondaire de mes activités de recherche. En revanche, comme expliqué dans l'introduction, la thématique plus générale de l'analyse et l'intégration de données omiques (que j'aborde parfois sous l'angle de l'utilisation de réseaux pour modéliser des interactions entre entités biologiques mais qui est plus générale) prend une part croissante dans mes activités.

### *Machine learning* et fouille de données



#### Éléments clés et bilan sur la thématique « *Machine learning* et fouille de données »

**Principaux collaborateurs INRAE** : Jérôme Mariette et Céline Brouard (MIAT, Math-Num) ; **National** : Université Dauphine (notamment Fabrice Rossi et Madalina Olteanu), Pierre Neuvial (Institut de Mathématiques de Toulouse), Marie Chavent (Université de Bordeaux)

**Doctorants Jérôme Mariette** (soutenue en 2017), **Alyssa Imbert** (soutenue en 2018) et **Nathanaël Randriamihamison** (débutée en 2018)

**Publications** Publications dans des journaux à comité de lecture [BAR-HEN et al. 2015 ; MONTASTIER et al. 2015 ; OLTEANU et Nathalie VILLA-VIALANEIX 2015b ; OLTEANU et Nathalie VILLA-VIALANEIX 2015c ; Nathalie VILLA-VIALANEIX et RUIZ-GAZEN 2015 ;

GENUER, POGGI, TULEAU-MALOT et al. 2017; MARIETTE, OLTEANU et al. 2017; COTTRELL et al. 2018; IMBERT, VALSESIA, LE GALL et al. 2018; MARIETTE et Nathalie VILLA-VIALANEIX 2018; MARTI-MARIMON, VIALANEIX et al. 2018; AMBROISE, DEHMAN, NEUVIAL et al. 2019], cinq chapitres de livre [Nathalie VILLA-VIALANEIX, Liaubet LIAUBET et al. 2016; Nathalie VILLA-VIALANEIX et CANU 2018; Nathalie VILLA-VIALANEIX et ROSSI 2018; MARIETTE, OLTEANU et al. 2020; MARIETTE et VIALANEIX 2020], deux articles soumis [IMBERT, VIALANEIX et al. 2020; RANDRIAMIHAMISON, VIALANEIX et al. 2020] et quatre packages R (**SOMbrero** [VIALANEIX, MAIGNÉ et al. 2020], **mixKernel** [MARIETTE et Nathalie VILLA-VIALANEIX 2020], **RNAseqNet** [VIALANEIX et IMBERT 2020] et **adjclust** [NEUVIAL et al. 2020])

L'aspect thématique le plus transversal de mes activités de recherche est le développement et l'utilisation d'approches issues du *machine learning* (apprentissage automatique, intelligence artificielle) et de la fouille de données. Cette thématique est restée centrale dans mes activités, s'orientant progressivement vers l'étude de méthodes d'analyse exploratoire et de visualisation pour des données complexes non numériques (réseaux en particulier) et des applications dans le domaine de la biologie. Elle s'inscrit dans le cadre de l'orientation générale [#OpenScience-3] pour le développement des approches prédictives en biologie. Mes activités dans ce domaine se sont beaucoup organisées autour de l'analyse de données décrites par des noyaux ou par des dissimilarités. Ce type d'approches permet de prendre en compte des données hétérogènes et qui se présentent sous des formes variées (données numériques ou textuelles, réseaux, ...) et a montré son efficacité et sa pertinence pour l'analyse et l'intégration de données en biologie<sup>1</sup>.

De manière plus spécifique, mes travaux les plus notables dans cette direction sont l'extension d'une méthode neuronale d'analyse exploratoire, les cartes auto-organisatrices, à des données décrites par des noyaux ou des dissimilarités. Ce travail initial a débouché sur l'étude de diverses extensions de cette approche dans [OLTEANU et Nathalie VILLA-VIALANEIX 2015b; OLTEANU et Nathalie VILLA-VIALANEIX 2015c; Nathalie VILLA-VIALANEIX et RUIZ-GAZEN 2015; MARIETTE, OLTEANU et al. 2017], notamment pour leur utilisation dans le domaine de la visualisation mais aussi pour améliorer le traitement massif des données, deux thématiques d'intérêt pour leur utilisation pour l'analyse exploratoire de données non numériques. Ces travaux sont implémentés dans le package R de cartes auto-organisatrices **SOMbrero** ([VIALANEIX, MAIGNÉ et al. 2020], plus de 7000 téléchargements et pour lequel Jérôme Mariette donne régulièrement des sessions de formation et dont l'interface web embarquée a été entièrement revue au printemps 2020 avec l'appui d'Élise Maigné).

Une part importante des applications possibles des méthodes à noyaux concerne les données qui se présentent sous la forme de graphes (ou réseaux). Dans ce cadre, j'ai proposé des approches pour inférer les structures de co-expression entre gènes en présence de données multi-niveaux ([MONTASTIER et al. 2015], en présence d'individus manquants ([IMBERT, VALSESIA, LE GALL et al. 2018], thèse d'Alyssa Imbert et package R **RNAseqNet** [VIALANEIX et IMBERT 2020]) ou lorsque des informations de co-localisation spatiale des gènes sont disponibles [MARTI-MARIMON, VIALANEIX et al. 2018]. Les deux premiers articles ont été réalisés dans le cadre d'une collaboration avec avec Nathalie Viguerie (I2MC, INSERM, Toulouse) et Armand Valsesia (Nestlé, Lausanne, Suisse) pour l'étude de la réponse à un régime chez les obèses ([#Food-2] comportements alimentaires et nutrition humaine) et ont donné lieu à la rédaction d'un article présentant la découverte d'un gène dont l'expression est lié à l'amaigrissement ([IMBERT, VIALANEIX et al. 2020], soumis pour publication et actuellement en révision. Le dernier article a été réalisé dans le cadre d'un projet sur la maturité périnatale chez le porc ([#3Perf-2], couplage génotypage/phénotypage pour l'amélioration de la performance et bien être animal).

Mes travaux récents dans le domaine développent des approches issues du machine learning pour l'analyse et l'intégration de données omiques. En particulier, dans le cadre de la thèse de Jérôme Ma-

<sup>1</sup>Schölkopf B., Tsuda K. and Vert J.P. (2004) *Kernel Methods in Computational Biology*, MIT Press, London, UK.

---

riette (IE PF bioinformatique de Toulouse), nous avons proposé une méthode d'intégration de données omiques volumineuses et de natures variées qui a été appliquée à l'analyse des données métagénomiques issues du projet TARA Oceans (MARIETTE et Nathalie VILLA-VIALANEIX 2018 et package R, **mix-Kernel** [MARIETTE et Nathalie VILLA-VIALANEIX 2020]).

La question de l'efficacité des approches lorsque les données sont massives ou volumineuses a également été abordée dans d'autres travaux [GENUER, POGGI, TULEAU-MALOT et al. 2017; AMBROISE, DEHMAN, NEUVIAL et al. 2019], pour les forêt aléatoires et pour une méthode à noyaux de classification hiérarchique (CAH) sous contrainte. Ce dernier travail se poursuit dans le cadre de la thèse de Nathanaël Randriamihamison, qui a pour but l'utilisation de cette approche de CAH contrainte à la comparaison de données haut-débit de conformation spatiale de la chromatine (donnée HiC). Cette thèse a d'ores et déjà donné lieu à l'écriture d'un article [RANDRIAMIHAMISON, VIALANEIX et al. 2020] qui aborde les aspects théoriques de l'utilisation de ce type d'approche pour des données décrites par des similarités arbitraires. Il a été accepté pour publication dans *Journal of Classification* sous réserve de modifications mineures (re-soumis en juillet 2020).

## Intégration et analyse de données omiques

### Éléments clés et bilan sur la thématique « Intégration et analyse de données omiques »

**Principaux collaborateurs INRAE** : GenPhySE (en particulier Sylvain Foissac, Laurence Liaubet, Thomas Faraut...), Laurence Huc (Toxalim, AH) ...; **National** : Sébastien Déjean (Institut de Mathématiques de Toulouse), Nathalie Viguerie (I2MC, INSERM Toulouse), Joost Van-Meerwijk (CPTP, INSERM Toulouse); **International** : Armand Valsesia (Nestlé, Lausanne, Suisse)

**Projets coordonnés EpiFun** (financé par le métaprogramme SelGen, INRAE, 2018/2020, co-coordonné avec Thomas Faraut, GenPhySE INRA), **ASTERICS** (financé par la Région Occitanie, 2020/2022)

**Doctorante Valérie Sautron** (soutenue en 2016)

**Publications** publications dans des journaux à comité de lecture [SAUTRON, TERENINA, GRESS, LIPPI, BILLON, LARZUL et al. 2015; BOLTON et al. 2017; TERENINA et al. 2017; FOISSAC, DJEBALI, MUNYARD, VIALANEIX, RAU, MURET et al. 2019]

Comme expliqué plus haut, une majorité des développements méthodologiques auxquels je participe en *machine learning* sont destinées à l'analyse et l'intégration de données omiques et principalement ou même sont motivés par un problème finalisé posé par des partenaires biologistes. Je ne reprendrai donc pas le descriptif de ces travaux, activités et collaborations dans cette partie, pour me focaliser sur les travaux finalisés complémentaires (qui ne requièrent pas des développements méthodologiques importants, par exemple) et sur les projets dont l'objectif principal est le développement et la diffusion de méthodes d'intégration de données omiques pour l'exploration de données et la biologie prédictive.

D'un point de vue finalisé, dans le cadre du projet SUSoSTRESS j'ai co-encadré, avec Elena Terenina (GenPhySE, INRA, Toulouse) la thèse de Valérie Sautron (soutenue en octobre 2016), lors de laquelle nous avons étudié l'impact de différents types de stress chez le cochon en focalisant sur les aspects longitudinaux [SAUTRON, TERENINA, GRESS, LIPPI, BILLON, LARZUL et al. 2015; TERENINA et al. 2017]. Cette question est d'intérêt pour l'amélioration des conditions d'élevage et la baisse de la surmortalité ([#3Perf-2], santé et bien-être animal, exploration des nouvelles méthodes d'amélioration et de sélection). J'ai également pris en charge la majeure partie de l'analyse statistique (différentielle) et du développement d'outils R dans le cadre du projet Fr-Agencode<sup>2</sup>, projet pilote français de l'initiative

---

<sup>2</sup><http://www.fragencode.org>

---

internationale FAANG<sup>3</sup> sur l'annotation fonctionnelle des animaux d'élevage ([#OpenScience-2], améliorer la mutualisation des données et développer les partenariats à l'échelle européenne). Actuellement, je coordonne la tâche d'intégration de données (WP leader) du projet METAhCOL, qui s'intéresse à l'étude de l'impact de polluants alimentaires sur le développement de cancer du colon ([#Food-2], exposition environnementale et sûreté des chaînes alimentaires).

Mon activité dans le domaine de la biostatistique va au-delà de l'appui aux biologistes puisque j'ai une activité constante de veille sur les données produites qui pourraient nécessiter le développement de méthodes d'analyse spécifique. En particulier, ces dernières années, je me suis beaucoup intéressée à l'analyse de *données Hi-C* (données de co-localisation nucléaire qui se présentent sous la forme de comptages entre paires de positions dans le génome, comptages qui sont utilisés comme des proxy de la mesure d'une proximité spatiale) et plus récemment, à l'analyse de *données Single Cell* (données d'expression de gènes obtenues par cellule unique, collaboration dans le cadre d'une étude sur les relations entre diabète et cellules du thymus avec l'INSERM).

## Statistique fonctionnelle (FDA)



### Éléments clés et bilan sur la thématique « Statistique fonctionnelle »

**Principaux collaborateurs INRAE (ex)MIAT** : Victor Picheny ; **INRAE autre** : Rémi Servien (InTheRes, SA), Laurence Liaubet (GenPhySE, GA) ; **International** : Noslen Hernández (CENATAV, Havana, Cuba)

**Doctorante Gaëlle Lefort** (débutée en 2018)

**Production** Publications dans des journaux à comité de lecture [HERNÁNDEZ et al. 2015 ; Nathalie VILLA-VIALANEIX, HERNANDEZ et al. 2016 ; DOU et al. 2017 ; LEFORT, Laurence LIAUBET, CANLET, TARDIVEL et al. 2019 ; PICHENY et al. 2019], article soumis [LEFORT, SERVIEN, QUESNEL et al. 2020] et deux packages R (**SISIR** ; [VIALANEIX, PICHENY et al. 2020] et **ASICS** [LEFORT, SERVIEN, TARDIVEL et al. 2020a])

L'objet d'étude de la statistique fonctionnelle sont des données décrites par des variables qui sont des courbes. Ce type de données se retrouve de manière très naturelle dans de nombreux problèmes par le biais de l'enregistrement de phénomènes continus et en particulier en chimie où de nombreuses données sont acquises sous forme de spectres (par exemple, spectres haut-débit RMN en métabolomique). La difficulté d'analyser ce type de données a plusieurs sources : la dimension infinie de l'espace sous-jacent dans lequel les variables explicatives sont définies (qui conduit à des problèmes mal posés en statistique), la forte corrélation des mesures obtenues entre divers points de mesure de la variable fonctionnelle, le problème des déformations (décalage, amplification du signal, etc) entre les diverses courbes mesurées pour divers individus.

Cette thématique de recherche est moins présente dans mes travaux récents à l'exception de travaux réalisés dans le cadre de deux problématiques importantes pour INRAE. Dans [Nathalie VILLA-VIALANEIX, HERNANDEZ et al. 2016 ; DOU et al. 2017 ; LEFORT, Laurence LIAUBET, CANLET, TARDIVEL et al. 2019], je me suis intéressée à l'utilisation de données métabolomiques haut débit acquises par technique <sup>1</sup>H RMN (Raisonnance Magnétique Nucléaire <sup>1</sup>H). En particulier, nous y avons montré l'intérêt de l'utilisation de ce type de données en biologie prédictive, notamment dans le contexte de l'impact du changement climatique sur l'élevage porcin ([#3Perf-2] santé et bien-être animal, [#3Perf-3] élaborer des références et indicateurs permettant d'évaluer les choix en terme d'agriculture et [#Climat-1] identification d'indicateurs de la capacité d'adaptation des plantes au changement climatique) et avons proposé une approche (implémentée dans un package R) de quantification automatique des métabolites à partir d'un spectre [LEFORT, SERVIEN, TARDIVEL et al. 2020a].

---

<sup>3</sup><https://www.faang.org/>

---

Dans le cadre de la thèse de Gaëlle Lefort, ces travaux sont une voie prometteuse pour l'évaluation des performances en élevage [#3Perf] et nous les avons mises en œuvre pour répondre à la problématique de la surmortalité dans les élevages porcins (bien être animal [#3Perf-2]), dans un article soumis pour publication (et actuellement en relecture après révision) [LEFORT, SERVIEN, QUESNEL et al. 2020]. Ces travaux se poursuivent avec des perspectives pour améliorer la méthode de quantification en utilisant l'information provenant de plusieurs spectres ou des données de spectres à deux dimensions acquis en « pool ».

Par ailleurs, dans [PICHENY et al. 2019], je me suis intéressée à identifier, dans une courbe climatique (par exemple, évapotranspiration journalière pendant une année), les périodes critiques pour la plante, en terme de contrôle des paramètres liés au climat ([#3Perf-3] élaborer des références et indicateurs permettant d'évaluer les choix en terme d'agriculture et [#Climat-1] identification d'indicateurs de la capacité d'adaptation des plantes au changement climatique). Nous avons proposé une approche automatique basée sur une regression semi-paramétrique pénalisée qui permet la détection automatique des zones du domaine de définition du prédicteur fonctionnel qui expliquent le mieux une quantité numérique d'intérêt (package R SISIR [VIALANEIX, PICHENY et al. 2020]). Nous l'avons aussi mise en œuvre sur des données issues du modèle de simulation SUNFLO<sup>4</sup> dans lequel la sortie d'intérêt était le rendement du tournesol. Nous poursuivons actuellement ces travaux par l'étude d'une méthode à base de forêts aléatoires.

## Activités d'animation au service de la recherche

Cette section présente de manière plus détaillée mes activités d'appui à la recherche, principalement tournées vers l'animation et la formation. J'ai choisi la non exhaustivité pour décrire mes actions dans les activités qui me semblent les plus représentatives et les plus structurantes pour mon projet professionnel futur.

### useR! 2019 (conférence internationale, 1200 participants)

<http://user2019.r-project.org>

J'ai présidé le comité d'organisation de la conférence internationale useR! qui a eu lieu du 9 au 12 juillet 2019 à Toulouse. Cette conférence est la conférence annuelle de la fondation R, langage de programmation devenu incontournable en statistique, sciences des données et bioinformatique.

Le comité d'organisation était composé pour moitié de membres d'INRAE (MIAT, InTheReSE, ODR, TSE) et pour moitié de membres de l'Institut de Mathématiques de Toulouse et de TSE (non INRAE). La conférence a rassemblé plus de 1200 personnes (dont environ 20% de Français), dont la moitié issue du milieu académique et l'autre moitié issue du secteur privé, d'origines thématiques très variées (statistique, développement, machine learning & IA, bioinformatique, . . .). Au niveau national, elle a été soutenue par la Société Française de Statistique (organisation d'un Datathon notamment) et par quatre GDR (BIM, MaDICS, MASCOT NUM et Statistique & Santé qui ont organisé des prix posters). Le budget total de la conférence s'est élevé à environ 700k€.

L'organisation de useR! a demandé 2 ans et demi de travail (la proposition d'organisation a été déposée auprès de la fondation R en janvier 2017). Pour celle-ci, je me suis occupée de la coordination de l'équipe d'organisation, en maintenant une activité de suivi permanente sur tous les sujets liés à l'organisation de la conférence (web, recherche de sponsors, gestion des soumissions et des inscriptions, budget, lieu, ...). J'ai également assuré moi-même un certain nombre de tâches comme la mise en place des outils de collaboration, l'interface avec la fondation R et les différentes entités gravitant autour (comme le groupe de promotion de la diversité, « R ForwaRds », l'interface avec la communauté scientifique française pour leur participation à la conférence (SFdS et 4 GDR), la prise en charge

---

<sup>4</sup>Casadebaig P., Guilioni L., Lecœur J., Christophe A., Champolivier L. and Debaeke P. (2011) SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments. *Agricultural and Forest Meteorology*, **151**(2), 163-178.

---

d'une partie des réponses sur l'email de contact de la conférence, l'interface avec les invités et les sponsors...

Le bilan de cette organisation, qui a demandé une énergie considérable à tous les membres du comité d'organisation durant les 6 derniers mois avant la conférence, est globalement très positif : la conférence a été un succès et elle a permis de renforcer l'image de la communauté française en statistique computationnelle et bioinformatique à l'international. Elle a également renforcé l'image d'INRAE comme acteur important sur ces thématiques.

## PF Biostatistique de Toulouse et Pôle IMABS

Localement, je me suis engagée de manière croissantes ces dernières années dans les activités d'animation à l'interface entre statistique (et mathématiques, informatiques et sciences des données en général) et biologie (et agronomie, écologie) et ce, principalement dans deux directions :

- la **coordination de la PF Biostatistique de Toulouse**. La plateforme de biostatistique de Toulouse, membre de Genotoul, se veut un carrefour de compétences pour appuyer la recherche en biologie. Elle est multi-sites et multi-instituts, et ses membres actifs sont localisés à INRAE et dans divers laboratoires de l'Université Paul Sabatier. J'en suis la coordinatrice principale, avec l'aide de Sébastien Déjean (IMT/Université Toulouse 3) et David Rengel (IPBS/CNRS Toulouse). Les activités principales de la plateforme sont :
  - des **formations continues en statistique et biostatistique**, à destination des non spécialistes, particulièrement des biologistes et bioinformaticiens d'INRAE. Ces formations sont réalisées dans le cadre des formations de la plateforme biostatistique de Toulouse et/ou en partenariat avec la plateforme bioinformatique de Toulouse. J'ai eu un investissement croissant sur ces aspects et j'ai repris la **responsabilité des formations dispensées par la plateforme biostatistique de Toulouse** en 2015 et entièrement revu leurs contenus et leur organisation. Ce travail a permis de renouveler l'offre de formations, en coordination avec le service de formation continue d'INRAE de Toulouse.
  - l'**appui à la recherche en biologie** par co-encadrement d'étudiants ou par la collaboration à des projets de recherche. C'est dans ce cadre que s'inscrivent les projets METAhCOL (dont j'ai parlé dans mon bilan scientifique) et ASTERICS (dont je parle dans la partie Perspectives). Dans ces deux projets, je suis, respectivement, leader de WP et coordinatrice principale.
  - l'**animation scientifique** avec notamment, l'organisation d'un groupe de travail mensuel <http://nathalievialeix.eu/biopusces/> et la co-organisation de la journée régionale annuelle « Bioinfo / Biostat ».
- Lors de mon recrutement à l'unité MIAT en 2014, j'ai souhaité m'investir sur des activités similaires au sein de l'INRA pour mieux m'insérer dans le collectif. En particulier, j'ai co-organisé avec Victor Picheny, le *séminaire hebdomadaire de l'unité*. Après diverses actions pour le relier aux divers acteurs de la communauté scientifique en mathématiques et informatique de Toulouse, nous avons réussi à obtenir une fréquentation importante, qui n'est pas limitée aux membres de l'unité mais s'élargit à des collègues d'autres unités ou des universités de Toulouse. C'est donc naturellement que je me suis engagée dans la **coordination du pôle IMABS/INRAE Toulouse** (Informatique et Mathématiques pour les AgroBioSciences – <https://www6.inrae.fr/imabs>) en 2018. IMABS a vocation à soutenir l'objectif du centre INRAE Occitanie Toulouse de se positionner comme un acteur majeur pour les approches prédictives en biologie, dans le cadre des priorités nationales de l'Institut. Dans ce cadre, entourée d'un comité de pilotage multi-unités, nous avons réalisé des actions de diffusion de l'information (newsletters, environ 3 ou 4 par an) et des actions d'animation (organisation de séminaires de centre ou de journées thématiques



---

en relation avec la communication de centre, notamment la journée « IA & Agriculture » qui a été organisée en partenariat avec l'institut de convergence #DigitAg).

## 4. PERSPECTIVES

**D'un point de vue scientifique**, mon objectif est d'intensifier encore mes activités dans le domaine du développement de méthodes statistiques et d'intelligence artificielle pour l'intégration des données omiques. Dans ce cadre, les travaux de développement de méthodes exploratoires que j'ai présentés dans la partie « *Machine learning* et fouille de données » trouveront une valorisation et une diffusion naturelle lors du projet ASTERICS qui démarre en septembre 2020 et qui a pour objectif de compléter les outils bioinformatique proposés par la PF BIOINFO de Toulouse par un portail web permettant de réaliser des analyses intégratives (dans un premier temps principalement exploratoires) de manière intuitive et simple ([#OpenScience-3], biologie des système). L'acceptation de ce projet est le fruit de 2 ans de travail de montage, au nom de la PF Biostatistique de Toulouse que je coordonne (voir la partie « Animation » de ce rapport) et en collaboration avec la PF BIOINFO et une société privée, Hyphen-Stat <https://www.hyphen-stat.com/>, spécialiste de l'analyse statistique de données biologiques.

Si je souhaite continuer à développer des approches exploratoires, je projette également d'intensifier mon travail en direction des approches prédictives d'intégration de données. D'un point de vue méthodologique, j'ai déjà commencé à m'engager dans cette direction, d'une part en collaboration avec Céline Brouard et Jérôme Mariette, pour proposer des approches de sélection de variables adaptées aux noyaux dans des cadres non supervisé et de régression à sortie multiple ou non numérique. Un article est actuellement en cours d'écriture sur ce sujet qui se base sur des approches pénalisées et des algorithmes d'optimisation par approche proximale.

En parallèle, j'ai commencé à m'intéresser au développement d'approches utilisant des informations de type « relations » ou « réseaux » pour l'amélioration de la prédiction phénotypique. C'était en particulier l'objectif initial du projet EpiFun (financé par le métaprogramme SelGen, 2018/2020) qui a permis la constitution d'un consortium européen sur la thématique plus générale de l'intégration de données omiques pour la prédiction phénotypique et a abouti au dépôt en mars 2020 d'un projet OBELICS au programme H2020 BIOTECH-07-2020. Ce projet, qui a demandé un investissement conséquent, n'a malheureusement pas été retenu mais je suis actuellement en cours de développer certaines des problématiques que nous avons prévues d'aborder dans ce projet. En particulier, j'ai organisé, depuis le printemps 2020, un groupe de lecture sur le sujet des méthodes neuronales utilisant des informations de graphes pour la prédiction au sein de l'unité (groupe de lecture inter-équipes). Dans la poursuite de ce travail, je projette, avec Céline Brouard, de proposer un sujet de master sur cette thématique à l'automne 2020 qui pourrait éventuellement déboucher sur un sujet de thèse.

**D'un point de vue des activités d'appui à la recherche**, je souhaite intensifier mes activités en direction de la structuration locale des interfaces entre statistique, mathématiques, informatique et sciences des données d'une part et biologie, agriculture, agronomie et écologie de l'autre (AgroBioSciences). En effet, le centre INRAE Occitanie Toulouse met l'accent sur le développement de l'analyse des données pour les AgroBioSciences (projet de Schéma Stratégique de Site 2018-2022). Je souhaite renforcer et aider à structurer le tissu local toulousain de compétences en statistique appliquées à la biologie. Ce besoin, qui fait partie des objectifs d'INRAE [#OpenInra-1] et des missions affichées du département MathNum, est également dicté par le déséquilibre local actuel entre besoins en appui pour l'analyse statistique de données et forces disponibles pour couvrir ces besoins.

Dans ce cadre, je poursuivrai mes activités de coordination de la plateforme biostatistique de Toulouse. Après avoir réorganisé l'offre de formation de la plateforme et avoir engagé la plateforme dans des projets financés qui visent à améliorer les services offerts en biologie, mes objectifs sont de renforcer les relations entre PF Biostatistique et PF BIOINFO à Toulouse, et de faire progressivement croître l'investissement de la PF Biostatistique dans l'appui aux projets locaux en AgroBioSciences.

---

Le projet ASTERICS sera une première étape pour une meilleure synergie entre les deux plateformes, synergie nécessaire pour pouvoir répondre à la difficulté de la gestion et de l'exploitation des très nombreuses données produites par les méthodes récentes de séquençage haut débit.

D'autre part, les activités d'IMABS ont naturellement été en retrait depuis le confinement mais le déplacement de l'unité MIAT dans un nouveau bâtiment pourvu d'une salle de séminaire en partie dédiée à cette animation et d'une salle de formation devrait permettre de poursuivre les efforts engagés. Il devrait permettre l'accueil d'invités qui voudraient s'engager sur des projets interdisciplinaires et donc favoriser les échanges et les collaborations entre méthodologistes et experts des AgroBioSciences. Les champs thématiques méthodologiques couverts par IMABS sont larges et les domaines d'application le sont tout autant. Il est donc compliqué d'assurer une animation scientifique suffisamment générale pour qu'elle s'adresse à cette large communauté et suffisamment spécifique pour qu'elle puisse générer de l'intérêt. Le succès de la journée IA & Agriculture organisée en février 2020 montre que cette adéquation est possible et permet de rassembler théoriciens, méthodologistes, spécialistes des AgroBioSciences et acteurs du secteur privé. L'objectif est de poursuivre sur cette lancée : le comité de pilotage a proposé des actions autour des thèmes de la reproductibilité ou du calcul haute performance pour lesquels nous envisageons l'organisation de journées thématiques avec des partenaires hors INRAE.

## ANNEXES (LISTE DES PRODUCTIONS)

### 1. Production de connaissances

#### 1.1 Production de connaissances originales reconnues dans la/les communauté(s) scientifique(s) de référence

##### 1.1.1 Articles à destination des communautés scientifiques validés par les pairs et publiés dans une revue à comité de lecture ou dans un entrepôt de documents en accès libre (comme par exemple les Peer Community In)

- Bar-Hen, Avner, Nathalie Villa-Vialaneix, and Hubert Javaux (2015). "Analyse statistique des profils et de l'activité des participants d'un MOOC". In: *Revue Internationale des Technologies en Pédagogie Universitaire* 12.1-2, pp. 11–22.
- Hernández, Noslen, Rolando J. Biscay, Nathalie Villa-Vialaneix, and Isneri Talavera (2015). "A non parametric approach for calibration with functional data". In: *Statistica Sinica* 25, pp. 1547–1566. DOI: [10.5705/ss.2013.242](https://doi.org/10.5705/ss.2013.242).
- Montastier, Emilie, Nathalie Villa-Vialaneix, Sylvie Caspar-Bauguil, Petr Hlavaty, Eva Tvřizicka, Ignacio Gonzalez, Wim H.M. Saris, Dominique Langin, Marie Kunesova, and Nathalie Viguerie (2015). "System model network for adipose tissue signatures related to weight changes in response to calorie restriction and subsequent weight maintenance". In: *PLoS Computational Biology* 11.1, e1004047. DOI: [doi:10.1371/journal.pcbi.1004047](https://doi.org/10.1371/journal.pcbi.1004047).
- Olteanu, Madalina and Nathalie Villa-Vialaneix (2015b). "On-line relational and multiple relational SOM". In: *Neurocomputing* 147, pp. 15–30. DOI: [10.1016/j.neucom.2013.11.047](https://doi.org/10.1016/j.neucom.2013.11.047).
- (2015c). "Using SOMbrero for clustering and visualizing graphs". In: *Journal de la Société Française de Statistique* 156.3, pp. 95–119. URL: <http://publications-sfds.math.cnrs.fr/index.php/J-SFds/article/view/473>.
- Sautron, Valérie, Elena Terenina, Laure Gress, Yannick Lippi, Yvon Billon, Catherine Larzul, Liaubet Liaubet, Nathalie Villa-Vialaneix, and Pierre Mormède (2015). "Time course of the response to ACTH in pig: biological and transcriptomic study". In: *BMC Genomics* 16.961, PMC4650497. DOI: [10.1186/s12864-015-2118-8](https://doi.org/10.1186/s12864-015-2118-8).
- Villa-Vialaneix, Nathalie and Anne Ruiz-Gazen (2015). "Beyond multi-dimensional data in model visualization: high-dimensional and complex nonnumeric data". In: *Statistical Analysis and Data Mining* 8.4. Discussion paper, pp. 232–239. DOI: [10.1002/sam.11274](https://doi.org/10.1002/sam.11274).

- 
- Villa-Vialaneix, Nathalie, Noslen Hernandez, Alain Paris, Cécile Domange, Nathalie Priymenko, and Philippe Besse (2016). “On combining wavelets expansion and sparse linear models for regression on metabolomic data and biomarker selection”. In: *Communications in Statistics - Simulation and Computation* 45.1, pp. 282–298. DOI: [10.1080/03610918.2013.862273](https://doi.org/10.1080/03610918.2013.862273).
- Bolton, Jennifer, Emilie Montastier, Jérôme Carayol, Sophie Bonnel, Lucile Mir, Marie-Adline Marques, Arne Astrup, Wim Saris, Jason Iacovoni, Nathalie Villa-Vialaneix, Armand Valsesia, Dominique Langin, and Nathalie Viguerie (2017). “Molecular biomarkers for weight control in obese individuals subjected to a multi-phase dietary intervention”. In: *The Journal of Clinical Endocrinology and Metabolism* 102.8, pp. 2751–2761. DOI: [10.1210/jc.2016-3997](https://doi.org/10.1210/jc.2016-3997).
- Dou, Samir, Nathalie Villa-Vialaneix, Laurence Liaubet, Yvon Billon, Mario Giorgi, Hélène Gilbert, Jean-Luc Gourdine, Juliette Riquet, and David Renaudeau (2017). “<sup>1</sup>H NMR-based metabolomic profiling method to develop plasma biomarkers for sensitivity to chronic heat stress in growing pigs”. In: *PLoS ONE* 12.11, e0188469. DOI: [10.1371/journal.pone.0188469](https://doi.org/10.1371/journal.pone.0188469).
- Genuer, Robin, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix (2017). “Random forests for big data”. In: *Big Data Research* 9, pp. 28–46. DOI: [10.1016/j.bdr.2017.07.003](https://doi.org/10.1016/j.bdr.2017.07.003).
- Mariette, Jérôme, Madalina Olteanu, and Nathalie Vialaneix (2017). “Efficient interpretable variants of online SOM for large dissimilarity data”. In: *Neurocomputing* 225, pp. 31–48. DOI: [10.1016/j.neucom.2016.11.014](https://doi.org/10.1016/j.neucom.2016.11.014).
- Terenina, Elena, Valérie Sautron, Caroline Ydier, Darya Bazovkina, Amélie Sevin-Pujol, Laure Gress, Yannick Lippi, Claire Naylies, Yvon Billon, Catherine Larzul, Liaubet Liaubet, Pierre Mormède, and Nathalie Villa-Vialaneix (2017). “Time course study of the response to LPS targeting the pig immune response gene networks”. In: *BMC Genomics* 18, p. 988. DOI: [10.1186/s12864-017-4363-5](https://doi.org/10.1186/s12864-017-4363-5).
- Cottrell, Marie, Madalina Olteanu, Fabrice Rossi, and Nathalie Villa-Vialaneix (2018). “Self-organizing maps, theory and applications”. In: *Revista Investigacion Operacional* 39.1, pp. 1–22. URL: <https://rev-inv-ope.univ-paris1.fr/fileadmin/rev-inv-ope/files/39118/39118-01.pdf>.
- Imbert, Alyssa, Armand Valsesia, Caroline Le Gall, Claudia Armenise, Gregory Lefebvre, Pierre-Antoine Gourraud, Nathalie Viguerie, and Nathalie Villa-Vialaneix (2018). “Multiple hot-deck imputation for network inference from RNA sequencing data”. In: *Bioinformatics* 34.10, pp. 1726–1732. DOI: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).
- Imbert, Alyssa and Nathalie Vialaneix (2018). “Décrire, prendre en compte, imputer et évaluer les valeurs manquantes dans les études statistiques : une revue des approches existantes”. In: *Journal de la Société Française de Statistique* 159.2, pp. 1–55. URL: <http://journal-sfds.fr/article/view/681>.
- Mariette, Jérôme and Nathalie Villa-Vialaneix (2018). “Unsupervised multiple kernel learning for heterogeneous data integration”. In: *Bioinformatics* 34.6, pp. 1009–1015. DOI: [10.1093/bioinformatics/btx682](https://doi.org/10.1093/bioinformatics/btx682).
- Marti-Marimon, Maria, Nathalie Vialaneix, Valentin Voillet, Martine Yerle-Bouissou, Yvette Lahbib-Mansais, and Laurence Liaubet (2018). “A new approach of gene co-expression network inference reveals significant biological processes involved in porcine muscle development in late gestation”. In: *Scientific Report* 8, p. 10150. DOI: [10.1038/s41598-018-28173-8](https://doi.org/10.1038/s41598-018-28173-8).
- Ambroise, Christophe, Alia Dehman, Pierre Neuvial, Guillem Rigau, and Nathalie Vialaneix (2019). “Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics”. In: *Algorithms for Molecular Biology* 14, p. 22. DOI: [10.1186/s13015-019-0157-4](https://doi.org/10.1186/s13015-019-0157-4).
- Foissac, Sylvain, Sarah Djebali, Kylie Munyard, Nathalie Vialaneix, Andrea Rau, Kevin Muret, Diane Esquerre, Matthias Zytnicki, Thomas Derrien, Philippe Bardou, Fany Blanc, Cédric Cabau, Elisa Crisci, Sophie Dhorne-Pollet, Françoise Drouet, Thomas Faraut, Ignacio Gonzáles, Adeline Goubil, Sonia Lacroix-Lamande, Fabrice Laurent, Sylvain Marthey, Maria Marti-Marimon, Raphaëlle Mormal-Leisenring, Florence Mompарт, Pascale Quere, David Robelin, Magali San-

- 
- Cristobal, Gwenola Tossier-Klopp, Silvia Vincent-Naulleau, Stéphane Fabre, Marie-Hélène Pinard-Van der Laan, Christophe Klopp, Michèle Tixier-Boichard, Hervé Acloque, Sandrine Lagarrigue, and Elisabetta Giuffra (2019). “Multi-species annotation of transcriptome and chromatin structure in domesticated animals”. In: *BMC Biology* 17, p. 108. DOI: [10.1186/s12915-019-0726-5](https://doi.org/10.1186/s12915-019-0726-5).
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Patrick Tardivel, Marie-Christine Pèrè, Hélène Quesnel, Alain Paris, Nathalie Iannuccelli, Nathalie Vialaneix, and Rémi Servien (2019). “ASICS: an R package for a whole analysis workflow of 1D 1H NMR spectra”. In: *Bioinformatics* 35.21, pp. 4356–4363. DOI: [10.1093/bioinformatics/btz248](https://doi.org/10.1093/bioinformatics/btz248).
- Picheny, Victor, Rémi Servien, and Nathalie Villa-Vialaneix (2019). “Interpretable sparse sliced inverse regression for functional data”. In: *Statistics and Computing* 29.2, pp. 255–267. DOI: [10.1007/s11222-018-9806-6](https://doi.org/10.1007/s11222-018-9806-6).
- Randriamihamison, Nathanaël, Nathalie Vialaneix, and Pierre Neuvial (2020). “Applicability and interpretability of Ward’s hierarchical agglomerative clustering with or without contiguity constraints”. In: *Journal of Classification*. Forthcoming. DOI: [10.1007/s00357-020-09377-y](https://doi.org/10.1007/s00357-020-09377-y).

### 1.1.2 Articles soumis à une revue ou déposés dans un entrepôt de documents d’accès libre en attente de validation par les pairs

- Apert, Cécile, Ariel Galindo-Albarran, Sarah Castan, Claire Detraves, Jérémy Santamaria, Nicola McJannett, Manuel Lebourrier, Olivier Joffre, Simon Fillatreau, Bart Haegeman, Nathalie Vialaneix, Claire Hoede, Paola Romagnoli, and Joost P.M. van Meerwijk (2020). “IL-2 and IL-5 drive intrathymic differentiation of functionally distinct regulatory T lymphocyte-clusters”. Submitted for publication.
- Imbert, Alyssa, Nathalie Vialaneix, Julien Marquis, Julie Vion, Aline Charpagne, Sylviane Metairon, Claire Laurens, Cédric Moro, Nathalie Boulet, Ondine Walter, Gregory Lefebvre, Jörg Hager, Dominique Langin, Wim H.M. Saris, Arne Astrup, Nathalie Viguerie, and Armand Valsesia (2020). “Network analyses reveal negative link between changes in adipose tissue GDF15 and BMI during dietary induced weight loss”. Submitted for publication (in revision).
- Lefort, Gaëlle, Rémi Servien, Hélène Quesnel, Yvon Billon, Laurianne Canario, Nathalie Iannuccelli, Cécile Canlet, Alain Paris, Nathalie Vialaneix, and Laurence Liaubet (2020). “The maturity in fetal pigs using a multi-fluid metabolomic approach”. Co-last author. Submitted for publication (in revision).

### 1.1.9 Chapitres d’ouvrages

- Villa-Vialaneix, Nathalie, Liaubet Liaubet, and Magali SanCristobal (2016). “Depicting gene co-expression networks underlying eQTLs”. In: *Systems Biology in Animal Production and Health*. Ed. by H.N. Kadarmideen. Vol. 2. Switzerland: Springer International Publishing, pp. 1–31. ISBN: 978-3-540-68489-3. DOI: [10.1007/978-3-319-43332-5\\_1](https://doi.org/10.1007/978-3-319-43332-5_1).
- Laguerre, Sandrine, Ignacio Gonzáles, Sébastien Nouaille, Annick Moisan, Nathalie Villa-Vialaneix, Christine Gaspin, Marie Bouvier, Agamemnon J. Carpousis, Muriel Cocaïgn-Bousquet, and Laurence HDSAMMG Girbal (2018). “Large-scale measurement of mRNA degradation in *Escherichia coli*: to delay or not to delay”. In: *High-Density Sequencing Applications in Microbial Molecular Genetics*. Ed. by Agamemnon J. Carpousis. Vol. 612. Methods in Enzymology. Cambridge, MA, USA: Elsevier, pp. 47–66. DOI: [10.1016/bs.mie.2018.07.003](https://doi.org/10.1016/bs.mie.2018.07.003). URL: [10.1016/bs.mie.2018.07.003](https://doi.org/10.1016/bs.mie.2018.07.003).
- Villa-Vialaneix, Nathalie and Stéphane Canu (2018). “Apprentissage connexionniste”. In: *Apprentissage Statistique et Données Massives*. Ed. by Myriam Maumy-Bertrand, Gilbert Saporta, and Christine Thomas-Agnan. Société Française de Statistique. Paris, France: Éditions TECHNIP.
- Villa-Vialaneix, Nathalie and Fabrice Rossi (2018). “Méthodes pour l’apprentissage de données massives”. In: *Apprentissage Statistique et Données Massives*. Ed. by Myriam Maumy-Bertrand, Gilbert Saporta, and Christine Thomas-Agnan. Société Française de Statistique. Paris, France: Éditions TECHNIP.

---

Mariette, Jérôme, Madalina Olteanu, and Nathalie Vialaneix (2020). “Kernel and dissimilarity methods for exploratory analysis in a social context”. In: Forthcoming (book chapter).

Mariette, Jérôme and Nathalie Vialaneix (2020). “Approches à noyau pour l’analyse et l’intégration de données omiques”. In: Forthcoming (book chapter).

#### 1.1.11 Relecture d’article en open peer reviews

- 2020 : PLoS ONE

#### 1.1.10 Relecture d’articles non open peer reviews

- **Journaux**

- 2020 : Bioinformatics
- 2019 : IEEE Journal of Biomedical and Health Informatics, BMC Bioinformatics (2), NAR Genomics and Bioinformatics, Molecular Ecology Ressources
- 2018 : IEEE Transactions on Neural Networks and Learning Systems, International Journal of Pattern Recognition and Artificial Intelligence, Bioinformatics, Journal of Statistical Computation and Simulation
- 2015-2017 : liste exhaustive non disponible mais généralement environ 5 rapports de relecture par an ou plus pour les journaux : Computational Statistics and Data Analysis, Information Sciences, Information Retrieval, Journal de la Société Française de Statistique, Neurocomputing, Bernouilli, Electronic Journal of Applied Statistical Analysis, Neural Networks, Environmental Modelling and Software, IEEE Transactions on Neural Networks and Learning Systems, Pattern Recognition Letters, BMC Bioinformatics, Communication in Statistics, Computational Statistics, Statistical Applications in Genetics and Molecular Biology, Bioinformatics, International Journal of Pattern Recognition and Artificial Intelligence, Journal of Statistical Computation and Simulation, IEEE Journal of Biomedical and Health Informatics, Molecular Ecology Ressources, NAR Genomics and Bioinformatics

- **Actes de conférences**

- 2020 : European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN, 5), International Joint Conferences on Artificial Intelligence (IJCAI, 4), Journées Ouvertes de Biologie, Informatique et Mathématiques (JOBIM, 13)
- 2019 : Modèles & Analyse des Réseaux : Approches Mathématiques & Informatiques (MARAMI, 4), International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM, 2), ESANN (3), IJCAI (4)
- 2018 : ESANN (3), IJCAI (6)
- 2015-2017 : liste exhaustive non disponible mais généralement entre 10 et 20 rapports de relecture pour des conférences par an. En particulier, j’ai été membres des comités scientifiques ou de relecteurs des conférences suivants : (internationales) WSOM 2015, 2017 et 2019 ; ESANN 2015, 2016, 2017, 2018, 2019 et 2020 et IJCAI 2018, 2019 et 2020 – (nationales) MARAMI 2019, Rencontres R 2015 et 2020 (repoussées en 2021), Journées de Statistique de la SFdS 2016, Statistical Methods for PostGenomic Data (SMPGD) 2020 (pas de relectures), JOBIM 2020

#### 1.1.12 Compte-rendu d’ouvrage ou note de lecture

Villa-Vialaneix, Nathalie (2015a). “Note de consultation : “Analyse des données multidimensionnelles” (MOOC, F. Husson et al., 2015)”. In: *Statistique et Enseignement* 6.2, pp. 61–66. URL: <http://www.statistique-et-enseignement.fr/article/view/493/472>.

---

### 1.1.14 Communications à des congrès et colloques : communication orale

- Lahbib-Mansais, Yvette, Maria Marti Marimon, Valentin Voillet, Harmonie Barasc, Florence Mompart, Juliette Riquet, Sylvain Foissac, David Robelin, Hervé Acloque, Yvon Billon, Nathalie Villa-Vialaneix, Liaubet Liaubet, and Martine Yerle-Bouissou (2016a). “3D nuclear positioning of IGF2 alleles and *trans* interactions with imprinted genes in pig fetal cells”. In: *Chromosome Research (Proceedings of International Colloquium on Animal Cytogenetics and Genomics, ICACG)* (July 2–5, 2016). Ed. by M. Yerle-Bouissou and A. Pinton. Vol. 24. S1. Toulouse, France, p. 89.
- Poggi, Jean-Michel, Robin Genuer, Nathalie Villa-Vialaneix, and Christine Tuleau-Malot (2017). “Random forests for big data”. In: *10th International Conference of the ERCIM WG on Computational and Methodological Statistics* (Dec. 16–18, 2017). E1510. London, UK.
- Ambroise, Christophe, Alia Dehman, Michel Koskas, Pierre Neuvial, Guillem Rigauill, and Nathalie Vialaneix (2018). “Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics”. In: *Journées bioinfo/biostat* (Dec. 13–13, 2018). Toulouse, France.
- Genuer, Robin, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix (2018). “Random forests for big data”. In: *Data Science, Statistics & Visualization*. Wien, Austria.
- Imbert, Alyssa, Armand Valsesia, Claudia Armenise, Gregory Lefebvre, Pierre-Antoine Gourraud, Nathalie Viguerie, and Nathalie Villa-Vialaneix (2018). “Multiple hot-deck imputation for network inference from RNA sequencing data”. In: *17th European Conference on Computational Biology* (Sept. 8–12, 2018). Highlight talk. Athens, Greece.
- Lefort, Gaëlle, Laurence Liaubet, Patrick Tardivel, Cécile Canlet, Marie Tremblay-Franco, Laurent Debrauwer, Nathalie Villa-Vialaneix, and Rémi Servien (2018). “Using ASICS to quantify metabolites in 1D <sup>1</sup>H NMR spectra: an application to perinatal survival in pigs”. In: *Workshop bioinfo/biostat* (Jan. 17–18, 2018). Toulouse, France.
- Poggi, Jean-Michel, Robin Genuer, Nathalie Villa-Vialaneix, and Christine Tuleau-Malot (2018). “Random forests for big data”. In: *Joint Statistical Meeting*. Vancouver, BC, Canada.
- Lahbib-Mansais, Yvette, Maria Marti-Marimon, Nathalie Vialaneix, Sylvain Foissac, Martine Bouissou-Matet, and Laurence Liaubet (2019). “Organisation nucléaire et expression génique lors du développement chez le porc”. In: *Séminaire du réseau EpiPHASE* (June 26–27, 2019). Castanet-Tolosan, France.
- Mariette, Jérôme, Céline Brouard, Rémi Flamary, and Nathalie Vialaneix (2019). “Unsupervised variable selection for kernel methods in systems biology”. In: *Journées bioinfo/biostat* (Oct. 4–4, 2019). Toulouse, France.

### 1.1.15 Communications à des congrès et colloques : conférence invitée

- Villa-Vialaneix, Nathalie (2015b). “What is a MOOC?” In: *ENBIS-15 Conference* (Sept. 6–10, 2015). Ed. by J. Bischoff, B. de Ketelaere, R. Göb, K. Lurz, I. Ograjensek, A. Pievatolo, and M. Reis. ENBIS Communications and Multimedia Center, Faculty of Economics, University of Ljubljana, Slovenia, Prague, Czech Republic.
- Cottrell, Marie, Madalina Olteanu, Fabrice Rossi, and Nathalie Villa-Vialaneix (2016). “Theoretical and applied aspects of the self-organizing maps”. In: *Advances in Self-Organizing Maps and Learning Vector Quantization (Proceedings of WSOM 2016)* (Jan. 6–8, 2016). Ed. by E. Merényi, M.J. Mendenhall, and O’Driscoll P. Vol. 428. Advances in Intelligent Systems and Computing. Houston, TX, USA: Springer International Publishing Switzerland, pp. 3–26. DOI: [10.1007/978-3-319-28518-4\\_1](https://doi.org/10.1007/978-3-319-28518-4_1).
- Olteanu, Madalina and Nathalie Villa-Vialaneix (2016b). “Using SOMbrero for clustering and visualizing complex data”. In: *9th International Conference of the ERCIM WG on Computational and Methodological Statistics* (Dec. 9–11, 2016). EO083. Seville, Spain.
- Picheny, Victor, Rémi Servien, and Nathalie Villa-Vialaneix (2016a). “Interpretable sparse sliced inverse regression for functional data”. In: *Workshop “Learning with Functional Data”* (Oct. 7–7, 2016). Lille, France.

- 
- Servien, Rémi, Victor Picheny, and Nathalie Villa-Vialaneix (2016). “Interval sparsity for functional inverse regression”. In: *22nd International Conference on Computational Statistics (COMPSTAT), Satellite CRoNoS Workshop on Functional Data Analysis* (Aug. 23–26, 2016). CO0153. Oviedo, Spain.
- Villa-Vialaneix, Nathalie, Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot (2016). “Random forest for big data”. In: *jstar2016 : Journées de Statistique de Rennes* (Oct. 20–21, 2016). Rennes, France.
- Villa-Vialaneix, N. (2017). “Stochastic self-organizing map variants with the R package SOMbrero”. In: *12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (Proceedings of WSOM 2017)* (June 28–30, 2017). Ed. by J.C. Lamirel, M. Cottrell, and M. Olteanu. Keynote speaker. Nancy, France: IEEE. DOI: [10.1109/WSOM.2017.8020014](https://doi.org/10.1109/WSOM.2017.8020014).
- Villa-Vialaneix, Nathalie (2018). “Learning from (dis)similarity data”. In: *European R Users Meeting (eRum 2018)* (May 14–16, 2018). Keynote speaker. Budapest, Hungary.

### 1.1.16 Communications à des congrès et colloques : poster

- Villa-Vialaneix, Nathalie and Madalina Olteanu (2015). “Multiple dissimilarity SOM for clustering and visualizing graphs with node and edge attributes”. In: *International Conference on Machine Learning (ICML 2015), Workshop FEAST* (July 6–11, 2015). Poster.
- Lahbib Mansais, Yvette, Maria Marti-Marimon, Valentin Voillet, Harmonie Barasc, Florence Mompert, Juliette Riquet, Sylvain Foissac, David Robelin, Hervé Acloque, Yvon Billon, Nathalie Villa-Vialaneix, Liaubet Liaubet, and Martine Bouissou-Matet (2016). “3D nuclear positioning of IGF2 alleles and *trans* interactions with imprinted genes”. In: *Conference of the International Society for Animal Genetics (ISAG 2016)* (July 23–27, 2016). P8001. Poster. Salt Lake City, UT, USA.
- Lahbib-Mansais, Yvette, Maria Marti Marimon, Valentin Voillet, Harmonie Barasc, Florence Mompert, Juliette Riquet, Sylvain Foissac, David Robelin, Hervé Acloque, Yvon Billon, Nathalie Villa-Vialaneix, Liaubet Liaubet, and Martine Yerle-Bouissou (2016b). “3D nuclear positioning of IGF2 alleles and *trans*-interactions with imprinted genes in fetal pig cells”. In: *Conference on Genome Architecture in Space and Time* (June 20–24, 2016). Poster. Trieste, Italy.
- Paniaga, Lily, Adrian Leip, Nathalie Villa-Vialaneix, and Wim de Vries (2016). “Estimating nitrous oxide fluxes from agricultural soils at European scale using a crop generic meta-model”. In: *19th Nitrogen Workshop* (June 27–29, 2016). 0187. Poster. Skara, Sweden.
- Sautron, Valérie, Marie Chavent, Nathalie Viguerie, and Nathalie Villa-Vialaneix (2016a). “Multiway SIR for biological data integration”. In: *Statistical Methods for Post Genomic Data* (Feb. 11–12, 2016). Poster. Lille, France.
- Sautron, Valérie, Elena Terenina, Laure Gress, Yannick Lippi, Yvon Billon, Nathalie Villa-Vialaneix, and Pierre Mormède (2016). “Time course of the response to ACTH in pig: biological and transcriptomic study”. In: *Conference of the International Society for Animal Genetics (ISAG)* (July 23–27, 2016). P3007. Poster. Salt Lake City, UT, USA.
- Mariette, Jérôme and Nathalie Villa-Vialaneix (2017). “Unsupervised multiple kernel learning to integrate various metagenomic sources”. In: *Proceedings of 4ème Colloque de Génomique Environnementale* (Sept. 13–15, 2017). Poster. Marseille, France.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Nathalie Vialaneix, and Rémi Servien (2018). “ASICS: a new R package for identification and quantification of metabolites in complex 1D 1H NMR spectra”. In: *17th European Conference on Computational Biology* (Sept. 8–12, 2018). Poster. Athens, Greece.
- Mestre, Camille, Sarah Djebali, Thomas Faraut, Nathalie Vialaneix, Andrea Rau, Cédric Cabau, Matthias Zytnicki, Thomas Derrien, Sandrine Lagarrigue, Elisabetta Giuffra, and Sylvain Foissac (2018). “Integrative analyses of chromosome conformation, chromatin accessibility and gene

- 
- expression in human and livestock genomes”. In: *17th European Conference on Computational Biology* (Sept. 8–12, 2018). Poster. Athens, Greece.
- Ambroise, Christophe, Alia Dehman, Michel Koskas, Pierre Neuvial, Guillem Rigau, and Nathalie Vialaneix (2019). “Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics”. In: *Statistical Methods for Post Genomic Data (SMPGD 2019)* (Jan. 31–Feb. 1, 2019). Poster. Barcelona, Spain.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Hélène Quesnel, Nathalie Vialaneix, and Rémi Servien (2019). “ASICS: a new R package for identification and quantification of metabolites in complex 1H NMR spectra”. In: *useR! 2019* (July 9–12, 2019). Poster. Toulouse, France.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Nathalie Vialaneix, and Rémi Servien (2019b). “ASICS : un package R pour l’identification et la quantification de métabolites dans un spectre RMN 1H”. In: *Journées bioinfo/biostat* (Oct. 4–4, 2019). Poster. Toulouse, France.
- Lefort, Gaëlle, Laurence Liaubet, Hélène Quesnel, Cécile Canlet, Nathalie Vialaneix, and Rémi Servien (2019). “ASICS: identification and quantification of metabolites in complex 1D 1H NMR spectra”. In: *15th Annual Conference of the Metabolomics Society (Metabolomics 2019)* (June 23–27, 2019). Poster. The Hague, The Netherlands.
- Lefort, Gaëlle, Nathalie Vialaneix, Hélène Quesnel, Marie-Christine Père, Nathalie Iannuccelli, Cécile Canlet, Alain Paris, Rémi Servien, and Laurence Liaubet (2019). “Study of fetal pig maturity in relation with neonatal survival using a multi-fluids metabolomic approach”. In: *15th Annual Conference of the Metabolomics Society (Metabolomics 2019)* (June 23–27, 2019). Poster. The Hague, The Netherlands.
- Randriamihamison, Nathanaël, Pierre Neuvial, and Nathalie Vialaneix (2019b). “Classification ascendante hiérarchique, contrainte d’ordre : conditions d’applicabilité, interprétabilité des dendrogrammes”. In: *Conférence sur l’Apprentissage Automatique (CAp 2019)* (July 3–5, 2019). poster. Toulouse, France.

### 1.1.15 Communications à des congrès et colloques : abstract

- Besse, Philippe, Nathalie Villa-Vialaneix, and Anne Ruiz-Gazen (2015). “Enseigner la statistique pour l’analyse de mégadonnées”. In: *47e Journées de Statistique de la SFdS (JdS 2015)* (June 1–5, 2015). Lille, France.
- Genuer, Robin, Jean-Michel Poggi, Christine Tuleau, and Nathalie Villa-Vialaneix (2015). “Random forests and big data”. In: *47e Journées de Statistique de la SFdS (JdS 2015)* (June 1–5, 2015). Lille, France.
- Olteanu, Madalina and Nathalie Villa-Vialaneix (2015a). “Classification et visualisation de graphes avec SOMbrero”. In: *4èmes Rencontres R* (June 24–26, 2016). Grenoble, France.
- Imbert, Alyssa, Caroline Le Gall, Claudia Armenise, Gregory Lefebvre, Jörg Hager, Armand Valsesia, Pierre-Antoine Gourraud, Nathalie Viguerie, and Nathalie Villa-Vialaneix (2016). “Imputation de données manquantes pour l’inférence de réseau à partir de données RNA-seq”. In: *48e Journées de Statistique de la SFdS (JdS 2016)* (May 30–June 3, 2016). Montpellier, France.
- Imbert, Alyssa and Nathalie Villa-Vialaneix (2016). “Outils pour l’analyse et la simulation de données RNA-seq”. In: *Rencontres R 2016* (June 22–24, 2016). Lightning talk. Toulouse, France.
- Mariette, Jérôme, Hélène Chiapello, and Nathalie Villa-Vialaneix (2016). “Integrating *Tara* oceans data sets using multiple kernels”. In: *15th European Conference on Computational Biology (ECCB 2016), Workshop “Recent Computational Advances in Metagenomics (RCAM)”* (Sept. 3–7, 2016). The Hague, The Netherlands.
- Picheny, Victor, Rémi Servien, and Nathalie Villa-Vialaneix (2016b). “Parcimonie par intervalle pour la régression inverse par tranche fonctionnelle”. In: *48e Journées de Statistique de la SFdS (JdS 2016)* (May 30–June 3, 2016). Montpellier, France.



- 
- Sautron, Valérie, Marie Chavent, Nathalie Viguerie, and Nathalie Villa-Vialaneix (2016b). “Multiway-SIR for longitudinal multi-table data integration”. In: *22nd International Conference on Computational Statistics (COMPSTAT)* (Aug. 23–26, 2016). CC0200. Oviedo, Spain.
- Villa-Vialaneix, Nathalie, Christophe Bontemps, and Sébastien Dejean (2016). “Outils pour chercher de l’information sur R et se former”. In: *Rencontres R 2016*. Lightning talk. Toulouse, France.
- Djebali, Sarah, Kylie Munyard, Nathalie Villa-Vialaneix, Cédric Cabau, Andrea Rau, Elisa Crisci, Thomas Derrien, Christophe Klopp, Matthias Zytnicki, Sandrine Lagarrigue, Hervé Acloque, Sylvain Foissac, and Elisabetta Giuffra (2017). “Integrative and differential analysis of transcriptomes and chromatin accessibility regions reveals regulatory mechanisms involved in pig immune and metabolic functions”. In: *Conference of the International Society for Animal Genetics (ISAG 2017)* (July 16–21, 2017). 144. Dublin, Ireland, p. 42.
- Foissac, Sylvain, Sarah Djebali, Hervé Acloque, Philippe Bardou, Fany Blanc, Cédric Cabau, Thomas Derrien, Françoise Drouet, Diane Esquerré, Stéphane Fabre, Christine Gaspin, Ignacio González, Adeline Goubil, Christophe Klopp, Fabrice Laurent, Sylvain Marthey, Maria Marti-Marimon, Florence Mompert, Kylie Munyard, Kevin Muret, Sophie Pollet, Pascale Queré, Andrea Rau, David Robelin, Magali SanCristobal, Michèle Tixier-Boichard, Gwenola Tosser-Klopp, Nathalie Villa-Vialaneix, Silvia Vincent-Naulleau, Matthias Zytnicki, Marie-Hélène Pinard-Van der Laan, Sandrine Lagarrigue, and Elisabetta Giuffra (2017). “Profiling the landscape of transcription, chromatin accessibility and chromosome conformation of cattle, pig, chicken and goat genomes”. In: *Conference of the International Society for Animal Genetics (ISAG 2017)* (July 16–21, 2017). 20. Dublin, Ireland, p. 6.
- Marti-Marimon, Maria, Hervé Acloque, Matthias Zytnicki, David Robelin, Sarah Djebali, Nathalie Villa-Vialaneix, Ole Madsen, Yvette Lahbib-Mansais, Diane Esquerré, Florence Mompert, Martin Groenen, Martine Yerle-Bouissou, and Sylvain Foissac (2017). “Characterization of 3D genomic interactions in fetal pig muscle”. In: *Conference of the International Society for Animal Genetics (ISAG 2017)* (July 16–21, 2017). Dublin, Ireland.
- Saby-Chaban, Claire, Wenting Zhang, René Fournier, Rémi Servien, Nathalie Villa-Vialaneix, Fabien Cobière, and Sylvie Chastant-Maillard (2017). “Reprise de cyclicité post partum et performances de reproduction chez la vache laitière Prim’Holstein en France”. In: *Proceedings of Journées Nationales des Groupements Techniques Vétérinaires* (May 17, 2011–May 19, 2017). Reims, France.
- Saby-Chaban, Claire, Wenting Zhang, René Fournier, Rémi Servien, Nathalie Villa-Vialaneix, Fabien Corbière, and Sylvie Chastant-Maillard (2017). “Progesterone and betahydroxybutyrate in line measurements for a better description and understanding of Holstein cows fertility in field condition”. In: *Proceedings of 8th European Conference on Precision Livestock Farming (EC-PLF)* (Sept. 12–14, 2017). Nantes, France.
- Imbert, Alyssa and Nathalie Villa-Vialaneix (2018). “RNAseqNet : un package pour l’inférence de réseaux à partir de données RNA-seq”. In: *7èmes Rencontres R* (July 4–6, 2018). Rennes, France. URL: <https://r2018-rennes.sciencesconf.org/204748>.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Nathalie Villa-Vialaneix, and Rémi Servien (2018). “ASICS : un package R pour l’identification et la quantification de métabolites dans un spectre RMN 1H”. In: *7èmes Rencontres R* (July 4–6, 2018). Rennes, France. URL: <https://r2018-rennes.sciencesconf.org/203758>.
- Djebali, Sarah, Sylvain Foissac, Nathalie Vialaneix, Kylie Munyard, Andrea Rau, Thomas Faraut, Sandrine Lagarrigue, Hervé Acloque, and Elisabetta Giuffra (2019). “Chromatin accessibility conservation across four livestock species”. In: *Conference of the International Society for Animal Genetics (ISAG 2019)* (July 7–12, 2019). Llieda, Spain.
- Foissac, Sylvain, Sarah Djebali, Kylie Munyard, Nathalie Vialaneix, Andrea Rau, Hervé Acloque, Sandrine Lagarrigue, and Elisabetta Giuffra (2019). “Functional annotation of livestock genomes: chromatin structure and regulation of gene expression”. In: *Journal of Animal Science (Proceedings*

- 
- of the ASAS/ASDS Midwest Joint Meeting). Vol. 97. Suppl. 2. Omaha, NE, USA: Oxford Univ. Press Inc., Cary, NC, USA, pp. 15–16. DOI: [10.1093/jas/skz122.028](https://doi.org/10.1093/jas/skz122.028).
- Foissac, Sylvain, Sarah Djebali, Nathalie Vialaneix, Matthias Zytnicki, Andrea Rau, Sandrine Lagarrigue, Hervé Acloque, and Elisabetta Giuffra (2019). “Multi-level conservation of chromosome conformation across livestock species reveals evolutionary links between genome structure and function”. In: *Conference of the International Society for Animal Genetics (ISAG 2019)* (July 7–12, 2019). Llieda, Spain.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Nathalie Vialaneix, and Rémi Servien (2019a). “ASICS : identifier et quantifier des métabolites dans un spectre RMN 1H”. In: *51èmes Journées de Statistique de la SFdS (JdS 2019)* (June 3–7, 2019). Strasbourg, France.
- Randriamihamison, Nathanaël, Pierre Neuvial, and Nathalie Vialaneix (2019a). “Classification ascendante hiérarchique, contrainte d’ordre : conditions d’applicabilité, interprétabilité des dendrogrammes”. In: *51èmes Journées de Statistique de la SFdS (JdS 2019)* (June 3–7, 2019). Strasbourg, France.
- Lefort, Gaëlle, Laurence Liaubet, Cécile Canlet, Nathalie Vialaneix, and Rémi Servien (2020). “ASICS: identification and quantification of metabolites in complex 1H NMR spectra”. In: *European RFMF Metabomeeting 2020* (Jan. 22–24, 2020). Ed. by Jules Griffin and Fabien Jourdan. Toulouse, France: RFMF and MPF, EC1.
- Lefort, Gaëlle, Nathalie Vialaneix, Hélène Quesnel, Marie-Christine Père, Yvon Billon, Laurianne Canario, Nathalie Iannuccelli, Cécile Canlet, Alain Paris, Rémi Servien, and Laurence Liaubet (2020). “Étude de la maturité des porcelets en fin de gestation en utilisant une approche métabolomique multifluide”. In: *52èmes Journées de la Recherche Porcine* (Feb. 4–5, 2020). B14. Poster. Paris, France: IFIP, INRAE. URL: <http://www.journees-recherche-porcine.com/texte/2020/bienetre/b14.pdf>.
- Randriamihamison, Nathanaël, Marie Chavent, Sylvain Foissac, Nathalie Vialaneix, and Pierre Neuvial (2020). “Classification ascendante hiérarchique sous contrainte de contiguité pour l’analyse différentielle de données Hi-C”. In: *Journées de Statistique de la SFdS (volume exceptionnel)*. Forthcoming.

### 1.1.15 Communications à des congrès et colloques : full paper

- Sibertin-Blanc, Christophe and Nathalie Villa-Vialaneix (2015). “Data analysis of social simulations outputs”. In: *Multi-Agent-Based Simulation XV (Proceedings of MABS 2014)* (May 5–6, 2014). Ed. by F. Grimaldo and E. Norling. Vol. 9002. Lecture Notes in Artificial Intelligence. Paris, France: Springer International Publishing Switzerland, pp. 133–150. DOI: [10.1007/978-3-319-14627-0](https://doi.org/10.1007/978-3-319-14627-0).
- Mariette, Jérôme and Nathalie Villa-Vialaneix (2016). “Aggregating Self-organizing maps with topology preservation”. In: *Advances in Self-Organizing Maps and Learning Vector Quantization (Proceedings of WSOM 2016)* (Jan. 6–8, 2016). Ed. by E. Merényi, M.J. Mendenhall, and O’Driscoll P. Vol. 428. Advances in Intelligent Systems and Computing. Houston, TX, USA: Springer International Publishing Switzerland, pp. 27–37. DOI: [10.1007/978-3-319-28518-4\\_2](https://doi.org/10.1007/978-3-319-28518-4_2).
- Olteanu, Madalina and Nathalie Villa-Vialaneix (2016a). “Sparse online self-organizing maps for large relational data”. In: *Advances in Self-Organizing Maps and Learning Vector Quantization (Proceedings of WSOM 2016)* (Jan. 6–8, 2016). Ed. by E. Merényi, M.J. Mendenhall, and O’Driscoll P. Vol. 428. Advances in Intelligent Systems and Computing. Houston, TX, USA: Springer International Publishing Switzerland, pp. 27–37. DOI: [10.1007/978-3-319-28518-4\\_6](https://doi.org/10.1007/978-3-319-28518-4_6).
- Mariette, Jérôme, Fabrice Rossi, Madlina Olteanu, and Nathalie Villa-Vialaneix (2017). “Accelerating stochastic kernel SOM”. In: *XXVth European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017)* (Apr. 26–28, 2017). Ed. by M. Verleysen. Bruges, Belgium: i6doc, pp. 269–274.

---

## 1.2 Produits pour la recherche mis à disposition mis à disposition de communautés scientifiques (logiciels, bases de données, matériels biologiques, etc.)

### 1.2.1 Développements informatiques à destination des communautés scientifiques (applications, bases de données, ontologies, terminologies, bibliothèques, plugins, sites web, couches graphiques...)

Josse, Julie, Imke Mayer, Nicholas Tierney, and Nathalie Vialaneix (2020). *R-miss-tastic: A resource website on missing values - Methods and references for managing missing data*. <https://rmisstastic.netlify.app>.

Lefort, Gaëlle, Rémi Servien, Patrick Tardivel, and Nathalie Villa-Vialaneix (2020a). *ASICS: Automatic Statistical Identification in Complex Spectra*. R package version 2.4.0 (2020-04-28), published on Bioconductor. URL: <https://bioconductor.org/packages/ASICS/>.

Mariette, Jérôme and Nathalie Villa-Vialaneix (2020). *mixKernel: Omics Data Integration Using Kernel Methods*. R package version 0.4 (2020-02-26), published on CRAN. URL: <http://CRAN.R-project.org/package=mixKernel>.

Neuvial, Pierre, Christophe Ambroise, Shubham Chaturvedi, Alia Dehman, Michel Koskas, Guillem Rigaiïl, and Nathalie Vialaneix (2020). *adjclust: Adjacency-Constrained Clustering of a Block-Diagonal Similarity Matrix*. R package version 0.5.99 (2020-06-18), published on CRAN. URL: <http://CRAN.R-project.org/package=adjclust>.

Vialaneix, Nathalie and Alyssa Imbert (2020). *RNAseqNet: Log-Linear Poisson Graphical Model with Hot-Deck Multiple Imputation*. R package version 0.1.4 (2020-06-08), published on CRAN. Maintainer. URL: <http://CRAN.R-project.org/package=RNAseqNet>.

Vialaneix, Nathalie, Élise Maigné, Jérôme Mariette, Madalina Olteanu, Fabrice Rossi, Laura Bendhaïba, and Julien Boelaërt (2020). *SOMbrero: SOM Bound to Realize Euclidean and Relational Outputs*. R package version 1.3 (2020-06-08), published on CRAN. Maintainer. URL: <http://CRAN.R-project.org/package=SOMbrero>.

Vialaneix, Nathalie, Victor Picheny, and Rémi Servien (2020). *SISIR: Sparse Interval Sliced Inverse Regression*. R package version 0.1-1 (2020-06-05), published on CRAN. Maintainer. URL: <http://CRAN.R-project.org/package=SISIR>.

### 1.2.2 Jeux de données

Lefort, Gaëlle, Rémi Servien, Patrick Tardivel, and Nathalie Villa-Vialaneix (2020b). *ASICSdata: Example of 1D NMR spectra data for ASICS package*. R package version 1.0.0 (2020-05-07), published on Bioconductor (data package). URL: <https://bioconductor.org/packages/ASICSdata/>.

## 1.3 Élaboration et animation de projets de recherche (académique, participatif, avec des partenaires privés ou publics)

### 1.3.1 Projets de recherches en partenariat avec un ou plusieurs partenaires académiques, socio-économiques ou pouvoirs publics

- **CO-LOcATION (Fetal maturity at the feto-maternal interface : COntribution of fetalL and maternal genOmes and tissue metAbolism perturbaTIONs)**
  - Financier : ANR (programme JCJC)
  - Date : 2021/2023
  - Responsable : Agnès Bonnet (GenPhySE, INRAE Toulouse)
  - Rôle : **participante**

- 
- **ASTERICS (A Tool for the ExploRation and Integration of omiCS data))**
    - Financier : Région Occitanie (et partenaire privé Hyphen-Stat)
    - Date : 2020/2022
    - Rôle : **coordinatrice**
  - **OBELICS (Open BiotEchnoLOGY models for the Integration of omiCS data into genotype-to-phenotype associations))**
    - Financier : Projet soumis (non accepté) au programme BIOTEC-07-2020 (H2020). 16 partenaires pour un budget de 8 millions d'euros
    - Rôle : **coordinatrice**
  - **EpiFun (Systems biology for genomic selection) (*projet d'animation*)**
    - Financier : INRA, Métaprogramme SelGen
    - Date : 2018/2020
    - Rôle : **Co-coordinatrice** (avec Thomas Faraut, GenPhySE, INRA)
  - **R-miss-tastic (A unified platform for missing values methods and workflows)**
    - Financier : R Consortium
    - Date : 2018/2020
    - Responsables : Julie Josse (École Polytechnique de Paris) et Nicholas Tierney (Monash University, Australie)
    - Rôle : *partenaire*
    - Valorisation associée : JOSSE et al. [2020](#) (site web)
  - **MILAGE (Microbiota of lamb genetically efficient)**
    - Financier : INRA, Métaprogramme MEM
    - Date : 2018/2019
    - Responsable : Christel Marie-Etancelin (GenPhySE, INRA Toulouse)
    - Rôle : *partenaire*
  - **SuBPIG (Enhancing survival at birth in piglets)**
    - Financier : INRA, Métaprogramme GISA
    - Date : 2018/2019
    - Responsable : Laurence Liaubet (GenPhySE, INRA Toulouse)
    - Rôle : *partenaire*
  - **SingleCell (*projet d'animation*)**
    - Financier : Programme Toulouse Tech Inter'Labs
    - Date : 2018/2019
    - Responsable : Cathy Maugis-Rabusseau (IMT, Université de Toulouse)
    - Rôle : *partenaire*
  - **METAhCOL (Impact of low doses mixtures of aromatic hydrocarbon on colon carcinogenesis process : focus on energy metabolism hallmark)**

- 
- Financier : ANR Environnement et Cancer
  - Date : 2017/2021
  - Responsable : Dominique Lagadic-Gossmann (IRSET, INSERM Rennes)
  - Rôle : **responsable de « work package » et d'équipe**
  - **SCALES (Inférence statistique multi-échelle et données-dépendante pour la génomique)**
    - Financier : CNRS (Appel à projets « Osez l'interdisciplinarité »)
    - Date : 2017/2019
    - Responsable : Pierre Neuvial (IMT, Université de Toulouse & CNRS)
    - Rôle : *partenaire*
    - Publications associées : [RANDRIAMIHAMISON, VIALANEIX et al. 2020] (journal), [RANDRIAMIHAMISON, NEUVIAL et al. 2019a; RANDRIAMIHAMISON, NEUVIAL et al. 2019b; RANDRIAMIHAMISON, CHAVENT et al. 2020] (conférences)
  - **FR-AgENCODE (A FAANG pilot project for the functional annotation of livestock genomes)**
    - Financier : Département INRAE/GA
    - Date : 2014/2017
    - Responsable : Elisabetta Giuffra (GABI, INRA Jouy-En-Josas) et Sylvain Foissac (GenPhySE, INRA Toulouse)
    - Rôle : *partenaire*
    - Publications associées : [FOISSAC, DJEBALI, MUNYARD, VIALANEIX, RAU, MURET et al. 2019] (journal), [DJEBALI, MUNYARD et al. 2017; FOISSAC, DJEBALI, ACLOQUE et al. 2017; MESTRE et al. 2018; DJEBALI, FOISSAC et al. 2019; FOISSAC, DJEBALI, MUNYARD, VIALANEIX, RAU, ACLOQUE et al. 2019; FOISSAC, DJEBALI, VIALANEIX et al. 2019] (conférences)
  - **memRNase (Localisation membranaire de la RNase E : rôle dans la maturation, la surveillance et la dégradation des ARNm)**
    - Financier : ANR (non thématique)
    - Date : 2014/2017
    - Responsable : Agamemnon Carpousis (Université de Toulouse)
    - Rôle : *partenaire*
    - Publication associée : LAGUERRE et al. 2018
  - **Apprentistique (Outils statistiques pour l'évaluation des performances en classification et apprentissage sur des données imprécises)**
    - Financier : CNRS (Défi interdisciplinaire « Mastodons »)
    - Date : 2017
    - Responsable : Bart Lamiroy (Université de Nancy)
    - Rôle : *partenaire*
  - **PigHeat (Adaptation des porcs à la chaleur : la voie génétique)**

- 
- Financier : ANR BIOADAPT
  - Date : 2013/2016
  - Responsable : David Renaudeau (UZR, INRA Antilles-Guyane)
  - Rôle : *partenaire*
  - Publication associée : DOU et al. 2017

- **SUSoSTRESS (Génétique moléculaire de la réponse au stress et robustesse chez le porc)**

- Financier : ANR BIOADAPT
- Date : 2013/2016
- Responsable : Pierre Mormède (GenPhySE, INRA Toulouse)
- Rôle : **responsable de « work package »**
- Publications associées : [SAUTRON, TERENINA, GRESS, LIPPI, BILLON, LARZUL et al. 2015; TERENINA et al. 2017] (journal) et [SAUTRON, CHAVENT et al. 2016a; SAUTRON, CHAVENT et al. 2016b; SAUTRON, TERENINA, GRESS, LIPPI, BILLON, Nathalie VILLA-VIALANEIX et al. 2016] (conférences)

- **Internet 3D (Réseau de gènes co-exprimés et Interactions géniques nucléaires)**

- Financier : INRA, Département GA
- Date : 2014/2015
- Responsable : Yvette Lahbib-Mansais (GenPhySE, INRA Toulouse)
- Rôle : *partenaire*
- Publications associées : [MARTI-MARIMON, VIALANEIX et al. 2018] (journal), [LAHBIB MANSAIS et al. 2016; LAHBIB-MANSAIS, MARTI MARIMON et al. 2016a; LAHBIB-MANSAIS, MARTI MARIMON et al. 2016b; MARTI-MARIMON, ACLOQUE et al. 2017; LAHBIB-MANSAIS, MARTI-MARIMON et al. 2019] (conférences)

### 1.3.3 Coordination ou participation à l’animation de consortia multi-acteurs nationaux ou internationaux (GIS, ...)

Membre du bureau (vice-secrétaire générale septembre 2014 – janvier 2017) et webmistress (juin 2011 – juillet 2016) de la **Société Française de Statistique** (SFdS, société savante).

### 1.4 Collaborations scientifiques, formalisées ou en cours d’élaboration ; préciser les dates de mise en œuvre et si, éventuellement, elles s’inscrivent dans un projet (en référence au point 1.2) et si des publications ou valorisations y sont associées

*Pour les collaborations non formalisées, je me limite aux travaux actuellement en cours.*

#### 1.4.1 Lister les collaborations au sein de l’unité ou de INRAE

- Collaboration au sein de l’unité MIAT : Céline Brouard (sélection de variables, réseaux de neurones pour graphe), Élise Maigné (**SOMbrero**), Jérôme Mariette (sélection de variables, **SOMbrero**, *formalisé* : ASTERICS, MILAGE), PF BIOINFO (*formalisé* : ASTERICS), Matthias Zytnicki (données Hi-C)
- Collaboration INRAE hors unité :

- 
- GenPhySE, département GA : Sylvain Foissac (données Hi-C, *formalisées* : thèse Nathanaël Randriamihamison, EpiFun, SCALES, FR-AgENCODE), Laurence Liaubet (*formalisées* : thèse Gaëlle Lefort, CO-LOCATION, SuBPIG, ASTERICS, Internet3D), Agnès Bonnet (*formalisées* : CO-LOCATION), Philippe Bardou (ASTERICS), Thomas Faraut (*formalisée* : EpiFun), Christel Marie-Etancelin (*formalisée* : MILAGE), Pierre Mormède (*formalisées* : SUSoSTRESS, thèse Valérie Sautron), Yvette Lahbib-Mansais (*formalisées* : Internet3D)
  - PEGASE (Rennes), département GA : Hélène Quesnel (*formalisée* : thèse Gaëlle Lefort), David Renaudeau (*formalisée* : PigHeat)
  - InTheRes, département SA : Laurence Huc (*formalisée* : METAhCOL)
  - LBE, département Environnement et Agronomie : Rémi Servien (données fonctionnelles, *formalisée* : thèse Gaëlle Lefort)
  - TBI, département Microbiologie et chaîne alimentaire : Muriel Cocaign, Laurence Girbal, Sandrine Laguerre (*formalisée* : memRNase)
  - Projets déposés (collaborations possibles à venir) :
    - **MAGICs (exploration of the MATuration of the Gut Immune system using CITE-seq)**
      - \* Financier : Département Physiologie Animale et Systèmes d'Élevage
      - \* Date : soumis à l'appel à projet département 2020 (début 2021)
      - \* Responsable : Christelle Knudsen (GenPhySE, INRAE Toulouse)
      - \* Rôle : **participante**
    - **EFFICACY\_EPN\_LEARNING (Le machine learning pour prédire l'efficacité des nématodes entomopathogènes en fonction de leur microbiote)**
      - \* Financier : Département Santé des Plantes et Environnement
      - \* Date : soumis à l'appel à projet département 2020 (début 2021)
      - \* Responsable : Sophie Gaudriault (DGIMI, INRAE Montpellier)
      - \* Rôle : **co-responsable**
    - **METABARRIER (Targeting gut microbiota-derived metabolites to promote epithelial barrier maturation at the suckling-to-weaning transition)**
      - \* Financier : ANR (programme JCJC)
      - \* Date : soumis au programme 2021
      - \* Responsable : Martin Beaumont (GenPhySE, INRAE Toulouse)
      - \* Rôle : **participante**

#### 1.4.2 Lister les collaborations nationales non INRAE

- **Inria** : Marie Chavent (*formalisée* : thèse Nathanaël Randriamihamison)
- **CNRS** : Pierre Neuvial (données Hi-C, classification ascendante hiérarchique sous contrainte *formalisées* : projet SCALES, thèse Nathanaël Randriamihamison)
- **INSERM** : Nathalie Viguerie (*formalisée* : thèse Alyssa Imbert), Joost Van-Meerwijk (single cell), Sarah Djebali (machine learning pour la recherche de relations enhanceurs / gènes : dépôt de demande de financement de thèse en cours, *formalisée* : EpiFun)
- **Autre** : Julie Josse (École Polytechnique de Paris; *formalisée* : R-miss-tastic),

---

### 1.4.3 Lister les collaborations internationales

Nicholas Thierney (Monash University, Australie; *formalisée* : R-miss-tastic); Di Cook (Monash University, Australie; visualisation de données Hi-C)

### 1.4.4 Participation (sans responsabilité) à des réseaux thématiques, disciplinaires, sociétés savantes, organisation de congrès, ...

- Participation au réseau thématique toulousain « SingleCell » coordonné par Cathy Maugis-Rabusseau.
- Membre de la Société Française de Statistique et de la Société Francophone de Classification. Ex-membre de « Femmes & Mathématiques »

## 2. Expertise et mobilisation de connaissances

### 2.2 Valorisation des connaissances scientifiques en appui à l'innovation

#### 2.2.5 Publication et valorisation dans des journaux et plateformes destinés aux professionnels

Chastant-Maillard, Sylvie, Claire Saby-Chaban, Wenting Zhang, René Fournier, Rémi Servien, and Nathalie Villa-Vialaneix (2019). "Reprise atypique de la cyclicité ovarienne chez la vache laitière : une forte association avec la cétose". In: *Le Point Vétérinaire.fr*. URL: <https://www.lepointveterinaire.fr/publications/le-point-veterinaire/article-rural/n-393/reprise-atypique-de-la-cyclicite-ovarienne-chez-la-vache-laitiere-une-forte-association-avec-la-cetose.html>.

### 2.3 Expertise scientifique auprès de la communauté scientifique nationale et internationale

#### 2.3.1 Participation à des comités de rédactions éditoriaux

- 2019-... : membre du comité éditorial de la revue "Journal of Statistical Software"
- 2017-2019 : membre du comité éditorial de la revue "Lifestyle Genomics"

#### 2.3.3 Evaluation de projets scientifiques, de rapports de thèse, d'HDR, et d'entités de recherche à l'échelle nationale ou internationale

- **Évaluation de projets scientifiques**
  - 2020 : rapport d'expertise de thèse CIFRE
  - 2019 : 2 rapports d'expertise de sujets de postdoc de l'initiative KIM, Université de Montpellier
  - 2018 : rapport d'expertise de projet de recherche région Auvergne Rhône Alpes
  - 2016 : rapport d'expertise de sujet de thèse département INRA EFPA
- **Rapports d'HDR**
  - 2017 : Andrea Rau (Université d'Evry)
- **Rapports de thèse**
  - 2020 : Alain Virouleau (Institut Polytechnique de Paris), Audrey Hulot (Université Paris Sud) (*à venir*)



- 
- 2019 : Florian Hébert (AgroCampus Ouest), Christophe Le Priol (Université de Grenoble)
  - 2018 : Gilles Monneret (Université Pierre et Marie Curie)
  - 2017 : Virginie Stanislas (Université d’Evry), Eva Carmina Serrano Balderas (Université de Montpellier)
  - 2016 : Trung Ha (Université Paris Sud)
  - 2015 : Méлина Gallopin (Université Paris Sud), Amaury Labenne (Université de Bordeaux)

#### **2.3.4. Participation à des comités de thèses, des instances de recrutement, des instances d’évaluation individuelle ou de collectifs, jurys de concours, jurys de sélection ; participation à des conseils scientifiques**

- **Jurys HDR**

- 2019 : Sébastien Déjean (Institut de Mathématiques de Toulouse)
- 2018 : Rémi Servien (Université de Toulouse)
- 2018 : Ndèye Niang Keita (CNAM)

- **Jurys thèse**

- 2021 : Fériel Boulfani (Université Toulouse I) (*à venir*)
- 2021 : Michaël Pierrelee (Université de Marseille) (*à venir*)
- 2020 : Antoine Bichat (University d’Evry) (*à venir*)
- 2018 : May Taha (Université de Montpellier)
- 2017 : Ioana Gavra (Université Toulouse III)
- 2015 : Delphine Rieutord (Université de Grenoble)
- 2013 : Joseph El Gemayel (Université Toulouse I)
- 2008 : Marco Follador (Université de Bologne, Italie)

- **Comités de thèse**

- Florian Touitou (GenPhySE INRAE, 2020-...)
- Manon Dugué (GenPhySE INRAE, 2020-...)
- Alexandre Conanec (Université de Bordeaux, 2019-...)
- Franck Boizard (I2MC, INSERM & IRIT, Université Toulouse III, 2016/2019)
- Valentin Voillet (GenPhySE INRA, 2013/2016)
- Agnès Bonnet (GenPhySE INRA, 2011/2013)

- **Instances d’évaluation individuelle**

- 2011/2018 : Membre titulaire élue du Conseil National des Universités (CNU), 26ème section : Participation aux campagnes de qualification 2012, 2013, 2014, 2016, 2018 et aux campagnes de promotion 2012, 2013, 2015, 2017 et 2018

- **Comité de recrutement ou jury de concours**

- 2020 : Comité de sélection recrutement PR, Université Paris 1
- 2020 : Comité de sélection recrutement PR, Université Paris Dauphine
- 2019 : Comité de sélection recrutement MCF, Université d’Evry

- 
- 2019 : Comité de sélection recrutement PR, Université Paris Dauphine
  - 2018 : Jury du concours CR Inria, centre de Paris
  - 2018 : Comité de sélection recrutement MCF, Université Paris 5
  - 2015 : Comité de sélection recrutement MCF, Université Paris 1

### 2.3.5 La représentation de l'institut dans des instances ou organisations nationales, européennes ou internationales impliquées dans des réflexions stratégiques (ANR, alliances, PEER, Belmont, ...)

Depuis 2019, membre d'un groupe prospectif d'experts en Intelligence Artificielle auprès de l'INSMI (Institut Nationale des Sciences Mathématiques et de leurs Interactions, CNRS).

### 2.3.6. Organisation de congrès, participation à des comités d'organisation

- **Manifestation internationale**

- 9-12 juillet 2019 : Présidente du comité d'organisation de useR! 2019 <http://user2019.r-project.org>, 1200 participants, budget de 700k€

- **Manifestations nationales**

- décembre 2020 (à venir) : Membre du comité d'organisation des journées SeqBIM 2020 (50-100 participants attendus)
- 22-24 juin 2015 : Présidente du comité d'organisation des Rencontres R 2016 <https://r2016-toulouse.sciencesconf.org>, 250 participants

- **Manifestations locales**

- (automne 2018, 2019 et à venir 2020) : Membre du comité d'organisation de la journée Bioinfo / Biostat <https://bioinfo-biostat.sciencesconf.org/>
- 28 février 2020 : Co-organisation de la journée IA & Agriculture dans le cadre du pôle IMABS et de #DigitAg
- 4-5 décembre 2019 : Co-organisation du Chromosome conformation symposium dans le cadre du pôle IMABS

## 2.5 Contribution aux débats d'idées et actions d'information

### 2.5.1 Diffusion de connaissances (documents destinés à un public large de non spécialistes, ouvrages, films, site web, etc.)

2015 : pour la Société Française de Statistique, j'ai assuré le suivi de la réalisation d'une **brochure ONISEP « zoom sur les métiers des mathématiques et de l'informatique »** et de deux courts films associés (dont un met en scène une statisticienne de l'INRA <https://www.youtube.com/watch?v=mVRTMAtwRFA>).

### 2.5.2 Conception et animation de débats avec les citoyens (conférences ou colloques, autres manifestations)

2019 : Participation aux **Rencontres de l'innovation**, cycle de conférences organisées par le département StID de l'IUT de Perpignan, Carcassonne : « La biologie à l'ère du big data »

---

### 3. Formation par la recherche, formation initiale et continue

#### 3.1 Contribution à la formation par la recherche

##### 3.1.2 Encadrement d'étudiant.es BAC+4, BAC+5

- *Mars/Août 2020* Co-encadrement (avec Sébastien Déjean, Université Toulouse 3 et David Rengel, IPBS/CNRS Toulouse) du stage de Mallory Cals (Master 2, MAPI<sup>3</sup>, Université Toulouse 3) « Intégration de données omiques pour l'étude de l'impact des contaminants alimentaires sur le métabolisme et le développement de cancers » dans le cadre du projet ANR METAhCOL.
- *Juin/Août 2019* Co-encadrement (avec Gaëlle Lefort, MIAT, INRA Toulouse) du stage de Fanny Mathevet (4ème année INSA Toulouse, filière Génie Mathématiques et Modélisation) « Étude des effets de polluants alimentaires sur les cellules » dans le cadre du projet ANR METAhCOL.
- *Mars/Août 2019* Co-encadrement (avec Rémi Servien, InTheRes, Gaëlle Lefort, MIAT et Laurence Liaubet, GenPhySE, INRA Toulouse) du stage de Camille Guilmineau (Master 2 Ingénierie de la décision et big data, Université Catholique de l'Ouest, Angers) « Analyse de données métabolomiques pour résoudre un problème de bien-être animal » dans le cadre du projet GISA SuBPIG.
- *Mars/Août 2018* Co-encadrement (avec Victor Picheny, MIAT, INRA de Toulouse et Rémi Servien, InTheRes, INRA de Toulouse) du stage de Yousra Khattali (Master 2 Recherche Opérationnelle, Université Paul Sabatier, Toulouse) « Sélection de variables dans les données climatiques pour les modèles agro-écologiques ».
- *Mars/Août 2018* co-Encadrement avec Pierre Neuvial (Institut de Mathématiques de Toulouse & CNRS) du stage de Nathanaël Randriamihamison (Master 1 MAPI3, Université Paul Sabatier, Toulouse) « Classification hiérarchique sous contrainte de contiguïté pour l'analyse de données Hi-C » dans le cadre du projet CNRS SCALES.
- *Mars/Août 2017* Co-encadrement (avec Alyssa Imbert, MIAT, INRA de Toulouse et Nathalie Viguerie, I2MC, INSERM) du stage de Thibaut Guignard (Master 2 Data Science pour la biologie, AgroCampus Ouest) « Analyse statistique d'une protéine impliquée dans les problèmes d'obésité en relation avec l'expression des gènes ».
- *Mars/Août 2016* Co-encadrement (avec Rémi Servien, Toxalim, INRA de Toulouse et Sylvie Chastant-Maillard et Claire Saby, ENVIT) du stage de Wenting Zhang (Master 2 Economics and Mathematics, Toulouse School of Economics) « Analysis of the variation of milk progesterone concentrations in dairy cows ».
- *Mars/Août 2016* Co-encadrement (avec Laurence Liaubet et Magali San Cristobal, GenPhySE, INRA de Toulouse) du stage de Camille Champion (Master 1 MAPI3, Université Toulouse III) « Analyse de données métabolomiques ».
- *Mars/Août 2015* Co-encadrement (avec Laurence Liaubet et Valérie Sautron, GenPhySE, INRA de Toulouse) du stage de Caroline Ydier (Master 2 Ingénierie mathématiques, Université de Nantes) « Étude de la réponse au lipopolysaccharide chez le porcelet » dans le cadre du projet ANR SusOStress.
- *Mars/Août 2015* Co-encadrement (avec Victor Picheny, MIAT, INRA de Toulouse) du stage de Rolande C.B. Kpekou Tossou (M2 Statistics and Econometrics, Toulouse School of Economics) « Analyse par simulation de l'interaction climat / rendement ».

- 
- *Mai/Juin 2015* Co-encadrement (avec Nathalie Viguerie, I2MC, INSERM) du stage de Maria Paula Caldas (M1 Economics and Statistics, Toulouse School of Economics) « DiOGenes Clinical Data Analysis ».

### 3.1.3 Thèses

- 2018-... : Co-encadrement (avec Rémi Servien, InTheRes, Laurence Liaubet, GenPhySE, et Hélène Quesnel, PEGASE, INRAE Toulouse et Rennes) de la thèse de Gaëlle Lefort « Intégration de données métabolomiques par quantification de données haut débit et application à l'analyse des causes de la mortalité périnatale chez le porc », financée par les départements INRAE Math-NUM, GA et SA et par l'Institut de convergence #DigitAg.
- 2018-... : Co-encadrement (avec Pierre Neuviel, Institut de Mathématiques de Toulouse/CNRS, Marie Chavent, CQFD/Inria Bordeaux et Sylvain Foissac, GenPhySE/INRAE Toulouse) de la thèse de Nathanaël Randriamihamison « Classification hiérarchique sous contrainte de contiguïté pour l'analyse de données Hi-C », financée par le programme INRAE/Inria.
- 2015/2018 : Co-encadrement (avec Nathalie Viguerie, I2MC/INSERM Toulouse) de la thèse de Alyssa Imbert « Intégration de données hétérogènes complexes à partir de tableaux de tailles déséquilibrées », financée par la société Methodomics et la Région Occitanie. Alyssa Imbert est actuellement postdoctorante au CEA.
- 2013/2017 : Co-encadrement (avec Christine Gaspin, MIAT/INRA Toulouse) de la thèse de Jérôme Mariette « Apprentissage statistique pour l'intégration de données omiques ». Jérôme Mariette a réalisé sa thèse dans le cadre de son poste d'IE PF Bioinformatique de Toulouse, poste qu'il occupe encore actuellement.
- 2013/2017 : Co-encadrement (avec Pierre Mormède et Elena Terenina, GenPhySE/INRA Toulouse) de la thèse de Valérie Sautron « Biologie intégrative des réponses au stress et robustesse chez le porc », financée par le projet ANR SusOStress et la Région Midi-Pyrénées. Valérie Sautron est actuellement biostatisticienne dans l'entreprise IQVIA à Strasbourg.

### 3.1.5 Autres types d'encadrement

- 2016/2017 : Co-encadrement (avec Pierre Mormède, GenPhySE/INRA Toulouse) du CDD IE de Gaëlle Lefort dans le cadre du projet ANR SusOStress.
- 2015/2016 : Co-encadrement (avec Céline Noirod, MIAT/INRA Toulouse) du CDD IE de Gaëlle Lefort sur l'analyse de données de méthylation.

### 3.1.6 Mémoire d'étudiant

Co-encadrement (avec Sylvain Foissac, GenPhySE/INRA Toulouse) d'un mémoire de 4ème année d'un binôme de l'INSA Toulouse, filière Génie Mathématiques et Modélisation sur la classification non supervisée de données HiC.

## 3.2 Contribution à la formation initiale et continue

La plupart de mes enseignements et interventions sont accompagnées de matériels (supports de cours et/ou de TP) disponibles sur mon site web : <http://www.nathalievialaneix.eu>.

### 3.2.1 Contribution aux enseignements

- **M2 Statistics and Econometrics (2016/2020)** (Toulouse School of Economics, Université Toulouse Capitole) Cours et TD « Graph Mining ». Formation internationale, cours dispensés en anglais, 25 heures par an, conception et mise en œuvre.

- 
- **M1 Economics and Statistics (2014/2019)** (Toulouse School of Economics, Université Toulouse Capitole) Cours et TD « Applied multivariate data analysis – Big data analytics ». Formation internationale, cours dispensés en anglais, 10 heures par an, conception et mise en œuvre.

### 3.2.2 Contribution à la formation destinée à des professionnels

- Depuis 2016 : **coordination de l'offre de formation de la plateforme Biostatistique de Toulouse** (en collaboration avec la formation permanente INRAE Toulouse). En particulier, j'ai pris en charge les formations (conception et mise en œuvre) :
  - Initiation à la statistique (sessions en 2016, 2017, 2018, 2019 et 2020)
  - Initiation à la statistique avec R (session programmée en 2020)
  - Graphiques avancés et interactifs avec R (sessions en 2017 et 2018)
  - An introduction to network inference and mining (session en 2015)
  - Analyse statistique de données RNA-seq en coordination avec la PF Bioinformatique de Toulouse (sessions en 2015, 2017 ; une programmée en 2020)
- 11-12 février 2019 : Conception et mise en œuvre d'une formation de laboratoire INRA (LIPM/INRA Toulouse) sur « Analyse statistique de données RNA-seq »
- 26-28 avril 2017 : Participation à une formation de laboratoire INRA (LIPM/INRA Toulouse) sur « Analyse de données 'omiques – de l'expression des gènes à l'inférence de réseaux »

### 3.2.3 Contribution à la formation continue des scientifiques ou techniciens

- 27 sept.-2 oct. 2021 : 4th course on « Computational Systems Biology of Cancer : Multi-omics and Machine Learning Approaches », Institut Curie, Paris, France. Intervention programmée sur « Kernel methods and variable selection for exploratory analysis and multi-omics integration »
- 19-22 sept. 2017 : École d'automne « Mathématiques, Informatique et Biologie » organisée par le LabEx CIMI (Centre International en Mathématiques et Informatique de Toulouse). Intervention sur « An introduction to network analysis »
- 2-7 octobre 2016 : JES « Apprentissage statistique et données massives », Journées de formation thématiques organisées par la SFdS. Interventions sur « apprentissage connexionnistes » et « Méthodes pour l'apprentissage de données massives » (avec co-rédaction de deux chapitres d'un ouvrage collectif [Nathalie VILLA-VIALANEIX et CANU 2018 ; Nathalie VILLA-VIALANEIX et ROSSI 2018])
- 17-22 Juillet 2016 : École d'été organisée par le LabEx SPS (Sciences des Plantes de Saclay) « From gene expression to genomic network ». Intervention sur « Normalization and differential analysis of RNA-seq data »
- 1-6 Mai 2016 : École de printemps organisée par l'Institut des Systèmes Complexes de Toulouse, « Advanced mathematics for network analysis ». Intervention sur « Statistical approaches for graph mining »
- 21 Juin-5 Juillet 2015 : School for advanced sciences of Luchon, France « Network analysis and applications, Session I ». Intervention sur « Mining co-expression network »

---

## 4. Animation ou direction de collectifs, de grands instruments, de ressources, de programmes ou de réseaux

### 4.2 Conception ou responsabilité scientifique de ressources ou de dispositifs collectifs

#### 4.2.1 Grands instruments, observatoires, plateformes, collections de ressources biologiques, infrastructures scientifiques collectives et infrastructures de recherche (dont les e-infrastructures)

**Coordination de la Plateforme de Biostatistique de Toulouse (2014-...)** : <https://perso.math.univ-toulouse.fr/biostat/>. La plateforme de biostatistique de Toulouse, membre de Genotoul, se veut un carrefour de compétences pour appuyer la recherche en biologie. Elle est multi-sites et multi-instituts, et ses membres actifs sont localisés à INRAE et dans divers laboratoires de l'Université Paul Sabatier. J'en suis la coordinatrice principale, avec l'aide de Sébastien Déjean (IMT/Université Toulouse 3) et David Rengel (IPBS/CNRS Toulouse). Les activités principales de la plateforme sont :

- la formation ;
- l'appui à la recherche en biologie par co-encadrement d'étudiants ou par la collaboration à des projets de recherche ;
- l'animation scientifique avec notamment, l'organisation d'un groupe de travail mensuel <http://nathalieviplaneix.eu/biopuces/> et la co-organisation de la journée régionale annuelle « Bioinfo / Biostat ».

### 4.3 Responsabilité dans des réseaux thématiques ou disciplinaires, nationaux ou internationaux

Lister les réseaux en précisant leur périmètre institutionnel (acteurs INRAE, non INRAE, national, international, etc.), la nature de votre participation, et l'année du début de celle-ci.

- **Coordination du pôle IMABS/INRAE Toulouse (2018-...)** : Informatique et Mathématiques pour les AgroBioSciences – <https://www6.inrae.fr/imabs>. IMABS a vocation à soutenir l'objectif du centre INRAE Occitanie Toulouse de se positionner comme un acteur majeur pour les approches prédictives en biologie, dans le cadre des priorités nationales de l'Institut. Dans ce cadre, entourée d'un comité de pilotage multi-unités, je réalise des actions de diffusion de l'information (newsletters, environ 3 ou 4 par an) et des actions d'animation (organisation de séminaires de centre ou de journées thématiques en relation avec la communication de centre). Les activités d'IMABS devraient croître avec le déplacement de l'unité MIAT dans un nouveau bâtiment qui sera en partie dédié à cette animation (avec, en particulier, une salle de conférences prévue à cet effet) ;
- **Co-animation de NETBIO (2013-...)** (réseau thématique financé par le département Math-NUM sur l'analyse de réseaux biologiques) : la coordinatrice principale de ce réseau est Marie-Laure Martin-Magniette (INRAE). Dans ce cadre, j'aide à la recherche de financement et à la co-organisation des journées annuelles dans lesquelles j'interviens également lorsque cela est pertinent.

### 4.5 Responsabilités éditoriales et d'organisation de manifestations d'ordre scientifique

*Note : Je ne comprends pas bien la différence entre cette section et les éléments mentionnés dans les sections 2.3.1 et 2.3.6. Je recopie donc partiellement (en me limitant aux activités les plus importantes) ces deux sections ici et en précisant mon rôle.*

---

- **Organisation de manifestations scientifiques**

- 9-12 juillet 2019 : Présidente du comité d'organisation de useR! 2019 <http://user2019.r-project.org>, 1200 participants, budget de 700k€

*Rôle précis* : coordination de l'équipe d'organisation, mise en place des outils de collaboration, interface avec la fondation R et les différentes entités gravitant autour, interface avec la communauté scientifique française pour leur participation à la conférence (SFdS et 4 GDR), prise en charge d'une partie des réponses sur l'email de contact de la conférence, interface avec les invités et les sponsors... Je me suis engagée dans le suivi d'absolument toutes les activités (web, recherche de sponsors, gestion des soumissions et des inscriptions, budget, lieu, ...) liées à l'organisation de la conférence.

L'organisation de useR! a demandé 2 ans et demi de travail et mobilisé une grande partie de mon énergie pendant les 6 derniers mois avant la conférence.

- 22-24 juin 2015 : Présidente du comité d'organisation des Rencontres R 2016 <https://r2016-toulouse.sciencesconf.org>, 250 participants

*Rôle précis* : coordination de l'équipe d'organisation, interface avec les invités, recherche de sponsors, prise en charge d'une partie des réponses à l'email de contact, organisation du programme, ...

L'organisation des Rencontres R 2016 s'est étalée sur environ 1 an.

- **Responsabilités éditoriales**

- 2019-... : membre du comité éditorial de la revue "Journal of Statistical Software"
- 2017-2019 : membre du comité éditorial de la revue "Lifestyle Genomics"

## 5. Autres productions

### 5.1.2 Newsletter

5 newsletters préparées et envoyées depuis décembre 2018 pour le pôle IMABS (diffusion locale INRAE Toulouse).