



**HAL**  
open science

## Algorithme de prédiction en temps réel de la consommation alimentaire journalière chez la truie en lactation

Raphaël Gauthier, Christine Largouët, Laurence Rozé, Jean-Yves Dourmad

► **To cite this version:**

Raphaël Gauthier, Christine Largouët, Laurence Rozé, Jean-Yves Dourmad. Algorithme de prédiction en temps réel de la consommation alimentaire journalière chez la truie en lactation. 53. Journées de la Recherche Porcine, Ifip; Inrae, Feb 2021, En ligne, France. pp.127-132. hal-03134418

**HAL Id: hal-03134418**

**<https://inria.hal.science/hal-03134418>**

Submitted on 8 Feb 2021

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

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

# Algorithme de prédiction en temps réel de la consommation alimentaire journalière chez la truie en lactation

Raphaël GAUTHIER (1), Christine LARGOUËT (2), Laurence ROZE (3), Jean-Yves DOURMAD (1)

(1) PEGASE, INRAE, Institut Agro, 35590, Saint Gilles, France

(2) Institut Agro / INRIA, Université de Rennes, CNRS, IRISA, F-35000 Rennes

(3) Université de Rennes, INSA, INRIA, CNRS, IRISA, Rennes, France

[raphael.gauthier@inrae.fr](mailto:raphael.gauthier@inrae.fr)

## Algorithme de prédiction en temps réel de la consommation alimentaire journalière chez la truie en lactation

Le développement d'algorithmes de prédiction de la consommation alimentaire journalière est essentiel à la mise en œuvre des stratégies d'alimentation de précision en temps réel. Compte tenu de l'absence de modèle mécaniste de prédiction de la consommation chez la truie en lactation, une nouvelle approche est proposée combinant une prédiction en temps réel avec un apprentissage « hors ligne » des comportements alimentaires de la truie. Une base de données de 39 090 lactations, provenant de six exploitations et contenant les consommations des 20 jours après mise-bas, a été utilisée pour (i) identifier des groupes (clusters) de truies présentant un comportement alimentaire similaire et (ii) tester trois fonctions de prédiction de la consommation. L'homogénéité des clusters obtenus par apprentissage « hors ligne » a été évaluée selon les indices Silhouette et Calinski-Harabasz. Les méthodes de prédiction ont été évaluées avec l'erreur moyenne et l'erreur quadratique moyenne (RMSE) déterminées par jour et par truie. Les clusters les plus homogènes sont obtenus lorsque la base est divisée en deux groupes. La trajectoire du premier groupe est caractérisée par une augmentation continue de la consommation au cours de la lactation, et la seconde, par un plateau atteint à partir du 10<sup>ème</sup> jour. L'erreur moyenne par truie, obtenue en utilisant deux trajectoires et la meilleure fonction de prédiction, est de -0,08 kg/j, avec une RMSE de 1,06 kg/j. Bien que la variabilité individuelle soit élevée, l'utilisation des trajectoires améliore la prédiction de la consommation. En pratique, l'apprentissage des trajectoires peut être renouvelé régulièrement, tandis que la procédure de prédiction, peu gourmande en puissance de calcul, peut être intégrée dans le système d'alimentation de précision.

## Algorithm for real-time prediction of daily feed intake in lactating sows

Developing algorithms able to predict daily feed intake is essential for implementing precision-feeding strategies in real time. Given the lack of a mechanistic model to predict feed intake in lactating sows, a new approach that combined real-time prediction with off-line learning of sow feeding behaviours was developed. A database of 39,090 lactations from 6 farms that contained the first 20 post-farrowing feed intake values was used to (i) identify groups of sows with similar feeding behaviour and (ii) test three functions to predict feed intake. The homogeneity of clusters obtained by off-line learning was assessed according to the Silhouette and Calinski-Harabasz scores. The prediction functions were evaluated by calculating mean error (ME) and root mean square error (RMSE) per day and per sow. The clusters with the best homogeneity were obtained by dividing the database into two groups. The trajectory of feed intake of the first group increased continuously during lactation, while that of the second plateaued from day 10 onwards. The ME per sow obtained for these two trajectories using the best function was -0.08 kg/d, and the corresponding RMSE was 1.06 kg/d. Although individual variability was high, the use of trajectories improved the prediction of feed intake. In practice, learning of trajectories may be recalculated regularly, while the real-time prediction function, which requires fewer computing resources, could be embedded into the smart feeder.

## INTRODUCTION

L'efficacité des systèmes d'alimentation porcins dépend en grande partie de l'équilibre entre les apports alimentaires et les besoins nutritionnels des animaux. A un stade donné, les porcs sont généralement nourris avec un régime alimentaire unique formulé pour couvrir les besoins de l'essentiel de la population. Selon le concept d'élevage de précision (Vranken et Berckmans, 2017), la prise en compte de la variabilité entre animaux apparaît comme un levier majeur permettant de développer des systèmes d'alimentation plus durables (Pomar *et al.*, 2019 ; Gaillard *et al.*, 2020) et d'améliorer la durabilité et la compétitivité des élevages. La capacité de ces nouveaux systèmes de précision à répondre de manière plus adéquate aux besoins individuels des animaux a déjà été démontrée chez le porc en croissance (Cloutier *et al.*, 2015).

Chez la truie en lactation, les besoins nutritionnels sont généralement très élevés (Noblet *et al.*, 1990), mais aussi fortement variables (Gauthier *et al.*, 2019). L'alimentation de précision semble donc être une stratégie intéressante pour améliorer l'équilibre entre les apports et les besoins individuels, mais le développement de cette approche nécessite notamment de pouvoir prédire la consommation individuelle journalière (CIJ) des truies (Gauthier *et al.*, 2019). Compte tenu des nombreux facteurs régulant l'appétit et agissant sur la prise alimentaire, il n'existe toutefois à ce jour aucun modèle mécaniste de prédiction de la consommation alimentaire en espèce porcine (Black *et al.*, 2009).

Dans cette étude, nous présentons une nouvelle approche de prédiction de la CIJ des truies en lactation, tirant profit de la disponibilité croissante de cette information en élevage du fait de l'utilisation de distributeurs d'aliment connectés (Piñeiro *et al.*, 2019). L'approche s'appuie sur des techniques d'apprentissage automatique permettant d'extraire des trajectoires de consommation à partir de grandes quantités de données. Ces trajectoires sont ensuite combinées avec une procédure de prédiction de la CIJ utilisable en temps réel. Cette publication a pour objectifs de présenter le fonctionnement de cette approche et d'en évaluer la qualité.

## 1. MATERIEL ET METHODES

### 1.1. Base de données de séries chronologiques

Une série chronologique (SC) univariée, notée  $F_t = f_{1,t}, \dots, f_{j,t}, \dots, f_{n,t}$ , est une séquence de valeurs numériques indexées selon le temps. Dans cette étude, chaque valeur  $f_{j,t}$  représente la CIJ d'une truie  $t$  pour une journée  $j$ ,  $j \in [1, 20]$ . Les 20 premières CIJ après mise-bas ont été enregistrées entre avril 2013 et juin 2019 dans six élevages commerciaux, dont les places de maternités étaient équipées du système Gestal® (JYGA Technologies Inc., Québec, Canada). La base de données ainsi constituée totalise 39 090 lactations. Le nombre de SC, la moyenne et l'écart-type des CIJ ont été calculés pour les six élevages. L'évolution de la CIJ au cours du temps a également été décrite.

### 1.2. Présentation de l'approche

L'approche proposée vise (i) à définir des groupes de truies ayant la même trajectoire de consommation pendant la lactation pour ensuite (ii) permettre une prédiction de la CIJ (Figure 1). La définition des groupes de truies fait appel à une procédure « hors ligne » d'apprentissage non supervisé, ayant pour objectifs de définir des clusters homogènes de SC et de

calculer un « prototype » représentant la trajectoire de consommation moyenne de chaque cluster. Pour prédire la CIJ à un jour donné, une trajectoire est d'abord affectée à la truie, puis la CIJ est prédite en temps réel en fonction de sa trajectoire sur la base des consommations enregistrées depuis sa mise-bas. La CIJ est ensuite mesurée par l'automate de distribution et enregistrée en fin de journée afin d'être utilisée les jours suivants.

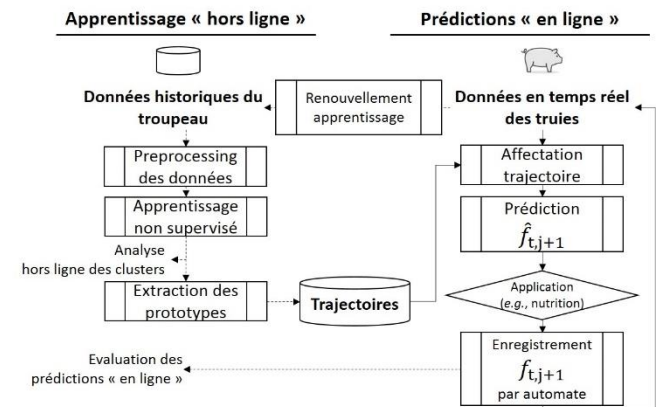


Figure 1 - Approche méthodologique pour la prédiction au jour de la consommation individuelle des truies en lactation

### 1.3. Apprentissage hors ligne et extraction de prototypes

En apprentissage non supervisé, l'identification de groupes homogènes au sein de grandes quantités de données, sans connaissance *a priori* sur la nature de ces groupes, relève des problèmes traités par les algorithmes de *clustering*. Dans cette étude, l'algorithme de clustering k-Shape (Paparrizos et Gravano, 2016) a été utilisé pour identifier des clusters homogènes de SC et en extraire des trajectoires, dans un sous-ensemble représentant 80 % des 39 090 SC de la base de données. Les 20% restant sont utilisés pour évaluer la précision de la prédiction en ligne. L'algorithme k-Shape permet de repérer efficacement la similarité de deux SC même lorsque celles-ci présentent une distorsion ou une translation de leurs valeurs selon l'axe des ordonnées. Chaque SC a donc été normalisée (centrée-réduite) de telle manière que la moyenne et l'écart-type soient égaux à 0 et 1, afin d'identifier la similarité dans le comportement alimentaire des truies malgré de possibles différences dans leur niveau moyen de consommation. L'algorithme k-Shape permet également de repérer efficacement les similarités de deux SC, dont les valeurs présentent un décalage selon l'axe des abscisses. Les index des SC pour lesquels leur similarité est maximisée sont évalués par calcul de la corrélation croisée normalisée, une mesure utilisée notamment en traitement de signal pour mesurer le décalage de deux signaux dans le temps (Paparrizos et Gravano, 2016).

L'algorithme k-Shape considère en entrée un paramètre  $k$  qui correspond au nombre de clusters à identifier. Dans cette étude,  $k$  représente donc le nombre de « comportements alimentaires » détectés dans la base de données. Comme cette valeur n'est pas connue *a priori*, l'algorithme k-Shape a été entraîné pour plusieurs valeurs de  $k$ , comprises entre 2 et 8. Pour chaque valeur de  $k$ , l'homogénéité des clusters obtenus a été évaluée avec les indices Silhouette (Rousseeuw, 1987) et Calinski-Harabasz (Calinski et Harabasz, 1974). L'indice Silhouette varie entre -1 et +1, où la valeur +1 indique que les clusters sont bien séparés entre eux et que les SC au sein d'un même cluster sont très similaires, tandis que la valeur -1 qualifie

la situation opposée. Pour le score Calinski-Harabasz, plus le score est élevé, plus la séparation des clusters est nette.

L'apprentissage hors ligne a été réalisé par élevage, et globalement sur l'ensemble du jeu d'apprentissage. La stratégie d'apprentissage par élevage a permis de calculer des prototypes spécifiques à chaque élevage (FS), tandis que la stratégie globale (G) a permis d'établir des prototypes sur l'ensemble de la base de données d'apprentissage. Chaque prototype a ensuite été exprimé en kilogramme par jour pour obtenir des trajectoires de consommation alimentaire pour chaque élevage, et sur l'ensemble des élevages.

#### 1.4. Assignment d'une truie à une trajectoire, prédiction en temps réel des CIJ, et évaluation

Pour une truie donnée, la procédure de prédiction en temps réel de la consommation, sur la base de données journalières, repose (i) sur l'assignation de la truie à l'une des trajectoires identifiées hors ligne, et (ii) sur la prédiction de la CIJ à partir de la trajectoire assignée et de ses CIJ mesurées depuis la mise-bas. L'assignation d'une truie à une trajectoire fait intervenir la mesure de similarité (Shape-Based Distance, SBD, Paparrizos et Gravano, 2016) entre les consommations passées de la truie et les trajectoires identifiées hors ligne. La CIJ est ensuite prédite à partir des fonctions 1f, utilisant la trajectoire assignée et la dernière consommation connue de la truie, et 2f, utilisant la trajectoire assignée et les deux dernières consommations connues de la truie. La CIJ est également prédite à partir d'une 3<sup>ème</sup> fonction naïve, appelée Persistance, pour laquelle la prédiction de la CIJ correspond à la dernière consommation enregistrée de la truie. Les différentes méthodes de prédiction, reposant (i) sur la combinaison des différentes stratégies d'apprentissage des trajectoires (FS et G, pour différentes valeurs de k) avec les fonctions de prédiction (1f, 2f), et (ii) sur la méthode Persistance sans apprentissage préalable des trajectoires, ont été évaluées avec l'erreur moyenne par truie ( $EM_t$ ) et par jour ( $EM_j$ ), et avec l'erreur quadratique moyenne par truie ( $RMSE_t$ ) et par jour en pourcentage ( $RMSEP_j$ ). Les méthodes de prédiction utilisant les trajectoires déterminées hors ligne sont dénommées ci-après FS.1f, FS.2f, G.1f et G.2f, selon les différents apprentissages réalisés.

## 2. RESULTATS

### 2.1. Description des données de consommation par élevage

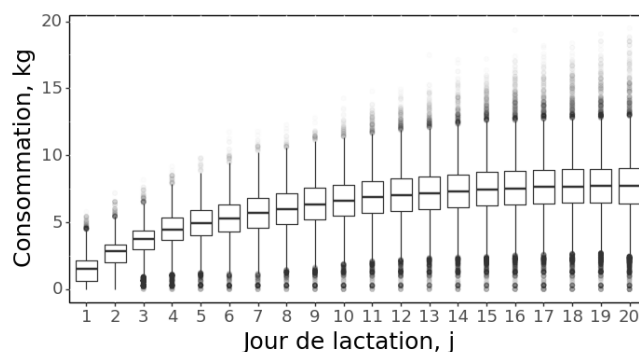
La consommation moyenne varie d'un élevage à l'autre, et est la plus élevée dans l'élevage 2 et la plus faible dans l'élevage 1 (Tableau 1). Elle est en moyenne de  $6,05 \pm 1,29$  kg/j sur une période de 20 jours *post-partum*.

**Tableau 1** - Nombre de séries chronologiques, consommation moyenne (kg/j) et écart-type par élevage

Élevages	Nombre	Moyenne, kg	Ecart type
1	7 872	5,14	0,87
2	3 467	6,98	1,42
3	9 111	6,50	1,12
4	651	6,22	1,13
5	10 692	6,08	1,26
6	7 297	5,99	1,28
Tous	39 090	6,05	1,30

La consommation augmente progressivement (Figure 2). Pendant les 5 premiers jours de lactation, la consommation moyenne augmente rapidement de 2 à 5 kg, puis elle continue à augmenter plus lentement avec un plateau atteint en 3<sup>ème</sup>

semaine de lactation. La variabilité de la CIJ est importante et augmente au cours de la lactation.



**Figure 2** - Consommations journalières pour 39 090 lactations

### 2.2. Identification « hors ligne » des trajectoires

L'indice Silhouette était le plus élevé pour une valeur de  $k=2$ , aussi bien lorsque l'apprentissage était réalisé séparément dans chacun des six élevages (FS), que globalement (G) sur tous les élevages. Pour  $k=2$ , les scores Silhouette variaient de 0,16 à 0,22. Pour  $k \in [3, 5]$ , les indices Silhouette diminuaient régulièrement, et pour  $k \in [6, 8]$ , les valeurs étaient inférieures à 0,10. Dans la stratégie FS, l'indice Calinski-Harabasz était maximal pour  $k=2$  dans les élevages 1, 2, 4, 5, 6. Pour l'élevage 3, cet indice était maximal pour  $k=3$ . Dans la stratégie G, cet indice était maximal pour  $k=2$ . Sur la base de ces observations,  $k=2$  correspond au meilleur paramètre permettant d'obtenir les clusters les plus homogènes possibles, à la fois dans les approches FS et G. La figure 3 représente pour chaque élevage et l'ensemble des élevages, les deux clusters obtenus, avec les prototypes normalisés correspondant. Dans chaque cluster, le prototype correspondant est décrit par une SC synthétisant la tendance suivie par les SC du cluster.

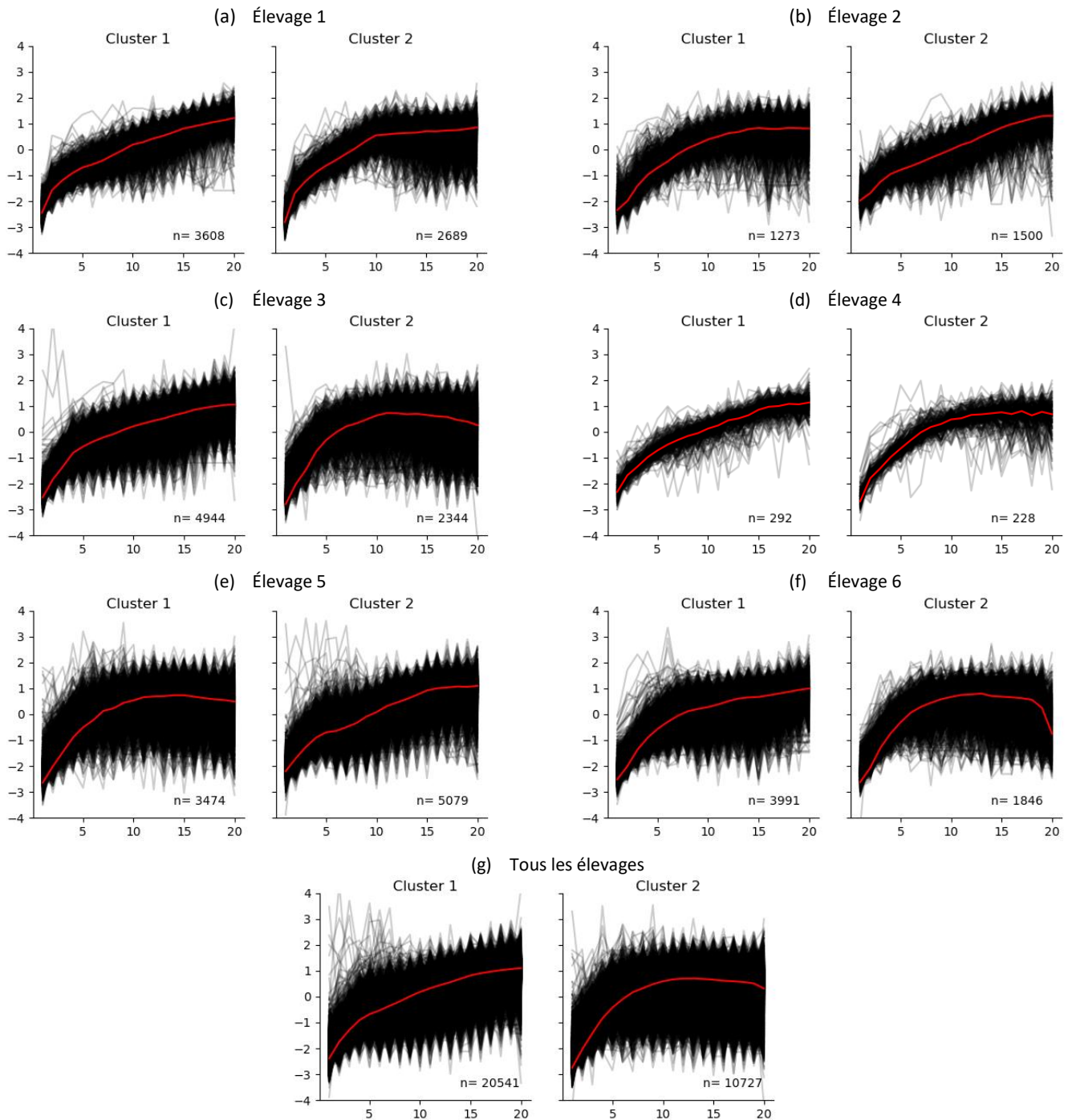
### 2.3. Evaluation des méthodes de prédiction « en ligne »

Pour chacune des méthodes de prédiction reposant sur les trajectoires déterminées hors ligne, les  $EM_j$  les plus faibles sont obtenues pour  $k=2$ . La valeur la plus faible est obtenue avec la méthode G.1f.2. Avec cette méthode, 75 % des valeurs  $EM_t$  étaient comprises entre -0,10 et +0,05 kg/j.

L'effet du stade de lactation sur  $EM_j$  est présenté à la figure 4. Pour FS.1f, FS.2f, G.1f et G.2f, avec  $k=2$ , la valeur  $EM_j$  était généralement négative en début de lactation, légèrement positive ensuite, et négative à nouveau jusqu'à la fin de la lactation. A titre de comparaison, les erreurs  $EM_j$  obtenues avec la méthode Persistance étaient presque toujours négatives avec d'importantes erreurs les premiers jours de lactation. Les erreurs  $EM_j$  variaient entre -0,30 et 0,20 kg/j pour les méthodes FS.1f, FS.2f, G.1f et G.2f, avec  $k=2$ . Elles étaient comprises entre -1,20 kg/j et 0,10 kg/j pour la méthode Persistance.

Pour chacune des méthodes FS.1f, FS.2f, G.1f et G.2f, les plus faibles  $RMSE_t$  étaient également obtenues pour  $k=2$  avec, respectivement 1,11, 1,06, 1,13 et 1,07 kg/j. En comparaison, la  $RMSE_t$  obtenue avec la méthode Persistance était de 1,21 kg/j. La plus faible  $RMSE_t$  était obtenue avec la méthode FS.2f avec  $k=2$ . Avec cette méthode, 75 % des  $RMSE_t$  étaient comprises entre 0,75 et 1,30 kg/j.

Pour toutes les méthodes, la valeur de  $RMSEP_j$  diminue au cours des 5 premiers jours de lactation, puis elle se stabilise à un niveau faible par la suite. Le premier jour, elle était la plus faible pour FS.1f.2 et FS.2f.2 et représentait 45,6% de la CIJ moyenne,



**Figure 3** - Comparaison des clusters et des prototypes normalisés pour chaque élevage et tous les élevages ensemble, pour  $k=2$ .

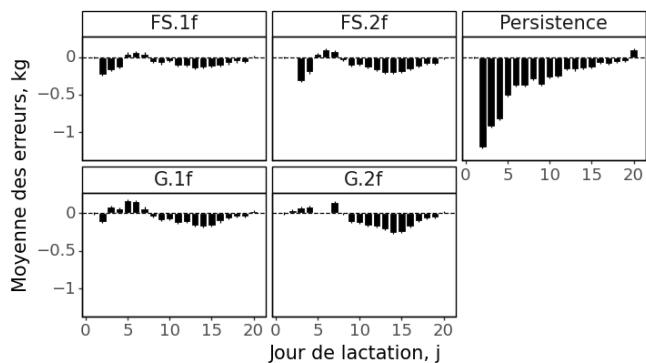
alors que pour G.1f.2, G.2f.2 et Persistance, elle atteignait 69,3% de la moyenne. Entre le 1<sup>er</sup> et le 3<sup>ème</sup> jour, la  $RMSEP_d$  diminue rapidement et atteint environ 20 % pour l'ensemble des méthodes basées sur les trajectoires identifiées hors ligne, alors qu'elle reste supérieure à 32% avec la méthode Persistance. Entre le 5<sup>ème</sup> et le 20<sup>ème</sup> jour, la  $RMSEP_d$  reste presque constante. Sur cette période les moyennes des  $RMSEP_d$ , avec  $d \in [5,20]$ , pour les méthodes Persistance, FS.1f.2, G.1f.2, FS.2f.2 et G.2f.2 étaient voisines avec respectivement 18,8%, 18,5%, 18,6%, 17,2%, and 17,2%.

### 3. DISCUSSION

#### 3.1. Identification hors ligne des trajectoires

L'algorithme k-Shape a été utilisé pour identifier des groupes homogènes de comportements alimentaires en lactation. La

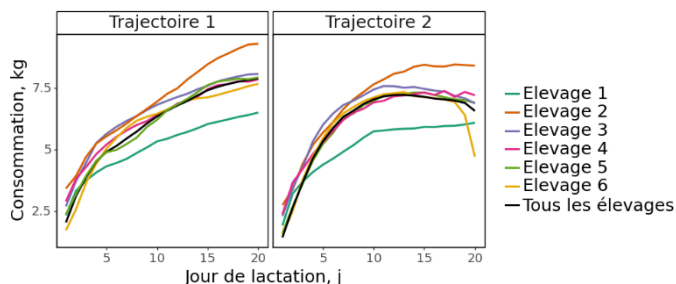
valeur de  $k=2$  maximise l'indice Silhouette dans chacun des six élevages, et également à l'échelle globale. L'indice Silhouette est strictement positif (0,20 en moyenne) ce qui indique que les SC sont plus proches de leur cluster d'appartenance que des autres. La relative faiblesse de l'indice suggère néanmoins que les clusters peuvent se chevaucher sur une certaine période de la lactation et que certaines SC peuvent présenter des similarités avec d'autres clusters. En effet, dans les différents clusters, l'évolution de la consommation alimentaire jusqu'au 5<sup>ème</sup> jour de lactation est assez similaire, ce qui reflète l'augmentation progressive de la ration allouée par l'éleveur. Cela pourrait être lié aux pratiques d'alimentation, à l'augmentation progressive de leurs besoins nutritionnels due à l'augmentation de la production laitière (NRC, 2012 ; Gauthier *et al.* 2019), à l'adaptation progressive de leur tube digestif (Theil, 2015).



**Figure 4** - Effet du stade de lactation sur l'erreur moyenne selon chaque méthode de prédiction

L'utilisation de plus de deux clusters n'a pas permis d'obtenir des ensembles de clusters plus homogènes, l'indice Silhouette diminuant de manière presque linéaire avec l'augmentation de la valeur de  $k$ . Le score Calinski-Harabasz confirme également ce nombre optimal de clusters. L'algorithme  $k$ -Shape a permis de bien contrôler les distorsions et les décalages selon l'axe des ordonnées entre SC, comme le montre la figure 3 avec des données normalisées.

Bien que la variabilité résiduelle des SC normalisées soit élevée, les résultats obtenus avec  $k$ -Shape suggèrent que les SC peuvent être classées en un nombre limité de trajectoires, indépendamment du niveau de consommation alimentaire et de la variabilité individuelle des truies. Une des trajectoires identifiées dans chaque élevage séparément ou dans tous les élevages simultanément décrit une augmentation rapide de la consommation pendant les 7 premiers jours, suivie d'une augmentation plus lente et presque linéaire par la suite. Elle représente entre 54 % et 68 % des SC de la base d'apprentissage, selon les élevages. La seconde trajectoire décrit une augmentation curvilinéaire de la consommation d'aliments pendant les 7 premiers jours, suivie d'un plateau, à partir du 10<sup>e</sup> jour environ. Cette 2<sup>nd</sup>e trajectoire représente entre 32 et 46 % des SC.



**Figure 5** - Comparaison des trajectoires obtenues pour chaque élevage et tous les élevages ensemble, pour  $k=2$ .

A partir de ces trajectoires, des courbes de consommation alimentaire ont été obtenues pour chaque élevage, et pour tous les élevages simultanément (Figure 5). Il existe très peu d'études disponibles dans la littérature sur les trajectoires de consommation alimentaire des truies en lactation, principalement parce que les CIJ sont très rarement enregistrées, sauf récemment avec l'utilisation de distributeurs connectés (Piñeiro *et al.*, 2019). À notre connaissance, seule l'étude de Koketsu *et al.* (1996) a été réalisée sur un grand nombre de truies (environ 25 000 lactations), les consommations journalières étant enregistrées manuellement par les éleveurs. La durée moyenne de la lactation (19 jours) était similaire à celle de la présente étude, et la consommation moyenne (5,2 kg/j) était inférieure de 15%. A partir de ces

données, Koketsu *et al.* (1996) ont identifié six trajectoires en fonction du niveau de consommation, de la dynamique temporelle de la consommation et de l'occurrence d'une baisse transitoire de la consommation pendant la lactation. Trois de ces trajectoires présentaient des évolutions similaires à celles des deux trajectoires identifiées dans la présente étude, avec une évolution rapide ou progressive de la consommation d'aliments dans le temps. Cependant, deux des trajectoires identifiées dans l'étude de Koketsu *et al.* (1996), celles avec une augmentation rapide et une baisse transitoire forte ou modérée de la prise d'aliments, n'ont pas été identifiées dans notre étude, même en considérant plus de deux clusters (résultats non présentés). D'après Dourmad (1991), qui a également observé une baisse transitoire de la consommation alimentaire à environ 5 jours de lactation chez des truies maigres ayant un appétit élevé et nourries *ad libitum* à partir de la mise-bas, ceci pourrait être lié à l'apparition de troubles gastro-intestinaux résultant d'une consommation alimentaire excessive non contrôlée en début de lactation. Il peut être suggéré que, dans la présente étude, l'utilisation de distributeurs connectés permettant de contrôler le risque de surconsommation aurait diminué la fréquence de ce type de comportements. Selon Koketsu *et al.* (1996), les courbes d'alimentation "rapides" ou "progressives", qui correspondent aux deux trajectoires identifiées dans la présente étude, sont celles préférables pour optimiser les performances de reproduction et de lactation, et réduire l'apparition de problèmes de retour en œstrus après le sevrage. Il est donc possible que le dispositif d'alimentation utilisé ait favorisé la fréquence de ces courbes dans notre base de données. Avec une autre base de données provenant d'élevages ayant des pratiques alimentaires différentes (par exemple avec une alimentation *ad libitum* dès la mise-bas), d'autres trajectoires auraient donc pu être identifiées. Cela souligne l'importance d'effectuer régulièrement la procédure d'apprentissage hors ligne, avec des données obtenues directement sur l'élevage.

### 3.2. Prédictions « en ligne »

La prédiction en ligne démarre par l'attribution chaque jour d'une trajectoire pour chaque truie, sur la base d'une mesure de similarité entre les trajectoires identifiées et les consommations de la truie enregistrées depuis sa mise-bas. En moyenne, pour  $k=2$ , une truie donnée est réaffectée 2,2 fois à une trajectoire différente au cours de sa lactation. Ces changements se produisent principalement vers le 5<sup>e</sup> et le 15<sup>e</sup> jour de la lactation. Cela pourrait être dû à la difficulté de comparer des trajectoires moyennes avec des SC présentant des CIJ davantage variables.

Pour chacune des méthodes de prédiction utilisant les trajectoires déterminées par apprentissage hors ligne, la qualité de la prédiction était la meilleure pour les plus faibles valeurs de  $k$ . L'augmentation du nombre de trajectoires tend donc à diminuer la qualité de la prédiction, probablement en réaffectant trop fréquemment les trajectoires des truies au cours de la lactation.

La méthode de prédiction reposant sur la fonction Persistance, sans apprentissage hors ligne, a été utilisée comme méthode de référence pour la comparaison avec les autres méthodes. La figure 4 montre clairement que cette fonction ne prédit pas efficacement la CIJ, en particulier au cours des 11 premiers jours de lactation pendant lesquels la CIJ augmente quel que soit l'élevage et le cluster. Sur cette période, les fonctions de prédiction 1f et 2f sont beaucoup plus performantes, avec des



valeurs  $EM_j$  plus faibles et centrées sur 0 et des  $RMSE_j$  plus faibles. Après 11 jours de lactation, lorsque la consommation d'aliment est stabilisée, les valeurs  $EM_j$  des différentes fonctions de prédiction sont comparables. Cependant, comme les erreurs  $EM_j$  peuvent se compenser de jour en jour, ce seul critère n'est pas suffisant pour évaluer la précision de la prédiction. L'analyse des  $RMSE$  fournit une compréhension complémentaire de la qualité prédictive des différentes méthodes. Selon les métriques  $EM_t$  et  $RMSE_t$ , les meilleures prédictions sont obtenues pour un apprentissage des trajectoires par élevage, avec  $k=2$ , et une fonction de prédiction basée sur les consommations des deux jours précédents (FS.2f.2) ( $EM_t = -0,08$  kg/j,  $RMSE_t = 1,06$  kg/j), bien que la même méthode basée sur l'ensemble des données de l'élevage (G.2f.2) soit très proche ( $EM_t = -0,04$ ,  $RMSE_t = 1,07$  kg/j). L'écart de  $RMSEP_j$  entre les fonctions 2f par rapport aux fonctions 1f est probablement due à une meilleure prédiction de la consommation en prenant en compte la moyenne entre deux prédictions.

### 3.3. Utilisation en pratique de l'approche

Comme indiqué à la figure 1, les prédictions peuvent être utilisées par toute application qui requiert de prédire la consommation journalière pendant la lactation, au niveau individuel. En alimentation de précision, la prédiction de la consommation alimentaire est nécessaire pour déterminer la teneur optimale en nutriments du régime à préparer par le distributeur connecté (Gauthier *et al.*, 2019).

Grâce à l'algorithme k-Shape, l'apprentissage hors ligne est relativement rapide, comparativement aux autres méthodes de clustering de SC. Le temps nécessaire dépend principalement du nombre de SC utilisées, variant de quelques minutes à quelques heures sur un ordinateur de puissance moyenne, comme ceux disponibles en exploitation. À l'inverse, les méthodes de prédiction en ligne sont très rapides et nécessitent moins d'une seconde d'exécution par truie par jour, et peu de ressources informatiques. La prédiction en ligne pourrait donc être facilement intégrée dans le système de contrôle des distributeurs connectés.

Pour démarrer le système de prédiction dans un nouvel élevage, où les données sur la consommation d'aliments ne sont pas encore disponibles, les trajectoires G obtenues à partir de

la présente étude peuvent être utilisées. Ensuite, lorsque des données de consommation alimentaire deviennent disponibles dans l'élevage en quantité suffisante (quelques centaines de SC comme dans l'élevage 4), la procédure d'apprentissage hors ligne peut alors être lancée afin de produire des trajectoires spécifiques à l'élevage. Cet apprentissage hors ligne nécessite très peu de paramètres et, d'après nos résultats, l'exploration des trajectoires avec  $k < 4$  semble être suffisante. Préalablement, un nettoyage des SC est obligatoire afin de détecter d'éventuelles anomalies dans les données, et ainsi, sécuriser l'identification hors ligne des trajectoires et la prédiction en ligne de la CIJ.

## CONCLUSION

La prédiction de SC de consommation d'aliment des animaux est une tâche difficile. Elle ouvre cependant la porte à de nombreuses applications dans la pratique pour l'alimentation de précision. Ce travail est une première contribution à l'utilisation des méthodes d'apprentissage automatique pour traiter des SC de consommations alimentaires de truies en lactation. Le « clustering » avec k-Shape permet d'extraire des trajectoires cohérentes avec celles de la bibliographie. Les résultats indiquent que les trajectoires alimentaires des truies en lactation sont peu nombreuses, bien qu'une forte variabilité individuelle subsiste au cours de la lactation. L'approche proposée est facilement interprétable et permet d'associer les consommations d'une nouvelle truie à une trajectoire connue. Enfin, ce système d'aide à la décision nécessite peu de ressources informatiques tout en étant en mesure d'apprendre de manière automatique à partir de nouvelles données.

## REMERCIEMENTS

Les auteurs remercient JYGA Technologies (Québec, Canada) et les éleveurs qui ont fourni les données utilisées dans cette étude, ainsi que Thomas Dahmen pour son aide à une phase précoce de l'étude. Celle-ci s'inscrit dans le cadre d'une thèse de doctorat du projet #DigitAg (ANR-16-CONV-0004), soutenu par l'Agence nationale française de recherche dans le cadre du programme "Investissements pour le futur"; et du programme de recherche et d'innovation Horizon 2020 Feed a Gene de l'Union européenne (n° 633531).

## REFERENCES BIBLIOGRAPHIQUES

- Black J.L., 2009. Models to predict feed intake. In: Torrallardona, D., Roura, E. (Eds.), Voluntary Feed Intake in Pigs. Wageningen Academic Publishers.
- Calinski T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods, 3, 1–27.
- Cloutier L., Pomar C., Létourneau Montminy M.P., Bernier J.F., Pomar J., 2015. Evaluation of a method estimating real-time individual lysine requirements in two lines of growing-finishing pigs. Animal, 9, 561–568.
- Gauthier R., Largouët C., Gaillard C., Cloutier L., Guay F., Dourmad J.-Y., 2019. Dynamic modeling of nutrient use and individual requirements of lactating sows. J. Anim. Sci., 97, 2886–2900.
- Koketsu Y., Dial G.D., Pettigrew J.E., King V.L., 1996. Feed intake pattern during lactation and subsequent reproductive performance of sows. J. Anim. Sci., 74, 2875–2884.
- Noblet J., Dourmad J.-Y., Etienne M., 1990. Energy utilization in pregnant and lactating sows: modeling of energy requirements. J. Anim. Sci., 68, 562–572.
- NRC, 2012. Nutrient Requirements of Swine, 11th rev. ed. Natl. Acad. Press, Washington, DC. 400 pp.
- Paparrizos J., Gravano L., 2016. k-Shape. ACM SIGMOD Record 45, 69–76.
- Piñeiro C., Morales J., Rodríguez M., Aparicio M., Manzanilla E.G., Koketsu Y., 2019. Big (pig) data and the internet of the swine things: a new paradigm in the industry. Anim. Front., 9, 6–15.
- Pomar, C., van Milgen, J., Remus, A., 2019. Precision livestock feeding, principle and practice. Poult. Pig Nutr. Challenges 21st Century 397–418.
- Rousseeuw P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math., 20, 53–65.
- Theil P.K., 2015. Transition feeding of sows, in: Farmer C. (Ed.), The Gestating and Lactating Sow. Wageningen Academic Publishers, The Netherlands, pp. 147–172.
- Vranken E., Berckmans D., 2017. Precision livestock farming for pigs. Anim. Front., 7, 32–37.