

Actualisation des méthodes d'ajustement des caractères du contrôle de performances à la ferme dans le Programme canadien pour l'amélioration génétique des porcs

Patrick GAGNON (1), Laurence MAIGNEL (2), Pius MWANSA (2), Brian SULLIVAN (2)

(1) Centre de développement du porc du Québec Inc, Place de la Cité, tour Belle Cour, 2590, boul. Laurier, bureau 450, Québec, Québec, G1V 4M6, Canada

(2) Centre canadien pour l'amélioration des porcs Inc, Ferme expérimentale centrale, Édifice 75, 960 avenue Carling, Ottawa, Ontario, K1A 0C6, Canada

pgagnon@cdpq.ca

Actualisation des méthodes d'ajustement des caractères du contrôle de performances à la ferme dans le Programme canadien pour l'amélioration génétique des porcs

Depuis plusieurs années, le poids des carcasses de porcs est en constante augmentation en Amérique du Nord. Dans ce contexte, les programmes d'évaluation génétique doivent s'adapter en s'éloignant de l'approche traditionnelle d'ajustement des performances à un poids unique pour adopter des approches plus flexibles et prenant en compte les profils de croissance et de dépôt des différents tissus. L'objectif de l'étude était de comparer différentes approches permettant de modéliser plusieurs variables de performances en phase de croissance à l'échelle individuelle. Un total de 1 998 animaux de race pure (Duroc, Yorkshire et Landrace) ont été pesés à au moins 7 reprises entre 30 et 160 kg de poids vif au cours de 13 essais en ferme. Des mesures par ultrasons pour l'épaisseur de lard et de muscle dorsal et le gras intramusculaire ont également été récoltées. Des modèles linéaires, ou linéaires avec point de rupture ou encore établis selon l'équation de Gompertz ont été comparés : la fonction de Gompertz a été retenue pour simuler le comportement sigmoïdal de certaines variables (poids et épaisseur de muscle) alors que pour d'autres, une fonction linéaire a été retenue (épaisseur de lard et gras intramusculaire). Pour l'ajustement individuel, un algorithme de classification a été appliqué et comparé aux approches actuellement utilisées dans le Programme canadien pour l'amélioration des porcs. La mise en application de cette nouvelle approche dans le programme d'évaluation fera appel à l'apprentissage automatique (machine learning) ce qui permettra à l'algorithme de s'améliorer lorsque la taille de la base de données de référence augmentera. Cette approche permet également de fournir des indicateurs de la précision des prédictions fournies.

Revision of adjustment methods for on-farm performance traits in the Canadian Swine Improvement Program

For several years, the weight of hog carcasses has been steadily increasing in North America. In this context, genetic evaluation programs must adapt by moving away from the traditional approach of adjusting performance to a single target weight and instead adopt more flexible approaches that consider growth and tissue deposition patterns. The aim of the study was to compare different approaches to model several time-dependent performance variables during the growth phase at the individual level. A total of 1,998 purebred animals (Duroc, Yorkshire and Landrace), were weighed at least 7 times from 30-160 kg live weight during 13 on-farm trials. Ultrasound measurements for backfat thickness and muscle depth and loin intramuscular fat were also collected. Linear models, linear models with breakpoints and the Gompertz equation were compared. The Gompertz function was selected to simulate the sigmoidal behaviour of some variables (weight and muscle depth), while a linear function was used for others (backfat thickness and intramuscular fat). For individual adjustments, a classification algorithm was applied and compared to approaches currently used in the Canadian Swine Improvement Program. The application of this new approach in the evaluation program will use machine learning, which will allow the algorithm to improve as the size of the reference database increases. This approach also makes it possible to provide precision criteria for the predictions provided.

INTRODUCTION

Depuis plusieurs années, le poids des carcasses de porcs est en constante augmentation en Amérique du Nord. Au Canada, le poids moyen de carcasse est passé de 86,4 kg en 1999 à 103,3 kg en 2019 (Agriculture et Agro-Alimentaire Canada, 2020). Les programmes de sélection ont dû s'adapter rapidement à cette évolution du poids de marché cible. Par exemple, les critères de sélection pour les animaux de race pure (âge, épaisseur de lard, épaisseur de muscle), auparavant ajustés à 100 kg de poids vif, sont désormais ajustés à 120 kg et, avec l'augmentation constante du poids de marché, cet ajustement doit être continuellement actualisé à la hausse. Une autre conséquence de cette évolution est l'augmentation de la gamme de poids des animaux testés en ferme car, si le poids moyen au sondage augmente, il existe toujours un besoin de tester des animaux légers, en particulier les mâles entiers. Le Programme canadien pour l'amélioration des porcs (PCAP) centralise actuellement des données de contrôle de performances en ferme pour des animaux pesant de 75 à 160 kg de poids vif. Les facteurs d'ajustement à poids constant utilisés sont régulièrement vérifiés et éventuellement actualisés afin de prendre en compte les effets de la sélection (par exemple l'amélioration de la croissance et la réduction de l'adiposité ont un impact sur les facteurs d'ajustement pour chacun de ces caractères). Historiquement, les animaux de race pure étaient contrôlés en ferme dans un intervalle de poids très restreint et leurs performances étaient standardisées par des ajustements linéaires. Dans le contexte actuel, il apparaît nécessaire d'actualiser non seulement les paramètres mais aussi les méthodes d'ajustement étant donné la large gamme de poids et le fait que la cinétique de dépôt de poids, gras et muscle est loin d'être linéaire, en particulier en fin de croissance.

La capacité de prédire le poids des animaux à un âge donné présente de nombreux avantages pour réaliser des ajustements de performances pour des besoins de comparaisons économiques standardisées, mais aussi, dans un contexte de sélection, pour comparer de façon fiable des animaux testés à des âges et poids parfois très différents. De nombreux modèles empiriques ont été proposés pour décrire la relation entre le poids et l'âge (Schull, 2013 ; Lonergan *et al.*, 2019). L'évolution du poids en fonction de l'âge, de la naissance à l'âge adulte, suit une courbe de type sigmoïdal, la vitesse de croissance augmentant jusqu'à un point d'inflexion puis diminuant jusqu'à ce que le poids adulte soit atteint. Wellock *et al.* (2004) ont comparé plusieurs types d'équations dans des environnements non limitants et conclu que les équations les plus adaptées sont celles utilisant un nombre limité de paramètres ayant une signification biologique, décrivant une croissance continue avec un seul point d'inflexion et une asymptote (croissance nulle) au moment de la maturité. Sur la base de ces critères, l'équation de Gompertz est couramment utilisée. Beaucoup d'autres équations ont été comparées dans différentes études (logistique, Bridges, Michealis-Menten généralisé, Richards, von Bertalanffy, etc) et les résultats varient en fonction des populations, des conditions expérimentales et du nombre de mesures disponibles. Une fois la meilleure équation trouvée et adaptée aux données, le gain de poids à un âge donné peut être calculé en utilisant la dérivée de rang 1 de cette équation.

Plusieurs études ont été consacrées à la décomposition du gain de poids en ses principales composantes (gras et muscle) pour modéliser la composition corporelle en fonction du poids vif (Schinckel & de Lange, 1996; Schinckel *et al.*, 2008). D'autres

exploitent également des données de prise alimentaire comme variables expliquant la cinétique du gain de poids, du gras et du muscle. Peu d'études sont consacrées à la cinétique du dépôt de gras intramusculaire (GIM) chez le porc. Concernant les ajustements individuels à poids constant, les courbes de croissance apportent réellement un avantage par rapport aux approches linéaires ou quadratiques. Cependant il est nécessaire de réviser régulièrement les paramètres de ces courbes, en particulier dans un contexte de sélection continue. Pour remédier à cela, des méthodes d'apprentissage automatique ont été proposées chez les bovins et le porc, afin de bâtir des algorithmes évolutifs (Alonso *et al.*, 2015 ; Alshahaf *et al.*, 2018).

Cette étude vise à définir les modèles les plus adaptés pour l'évolution du poids, de l'épaisseur de lard dorsal, de l'épaisseur de muscle et du gras intramusculaire et à proposer une approche flexible pour les utiliser dans le PCAP afin de standardiser les mesures collectées sur les porcs canadiens. Pour cela, une approche prometteuse inspirée de l'apprentissage automatique sera proposée.

1. MATERIEL ET METHODES

1.1. Données disponibles

Le projet de recherche a nécessité la collecte de mesures répétées sur 1 998 animaux de race pure (Duroc, Landrace et Yorkshire) dans 13 essais et élevés dans sept troupeaux de sélection participant au PCAP situés dans les provinces du Québec et de l'Ontario (Tableau 1). Les animaux ont été pesés autour de 30 kg, autour de 50 kg, puis pesés et mesurés aux ultrasons à 80 kg puis toutes les deux semaines jusqu'à un poids final allant jusqu'à 160 kg (poids final moyen de 140 kg). Les mesures aux ultrasons pour l'épaisseur de lard dorsal (ELD), l'épaisseur de muscle dorsal (EMD) et le pourcentage de gras intramusculaire de la longe (GIM) ont été réalisées par un technicien accrédité sur le PCAP, entre la 3^{ème} et la 4^{ème} dernières côtes.

Tableau 1 – Nombre d'animaux mesurés par race et sexe (au moins cinq mesures).

Race	Sexe ¹	Poids	ELD/EMD	GIM
Duroc	C	75	54	0
	F	277	245	93
	M	316	268	86
Landrace	C	115	84	0
	F	276	252	56
	M	171	163	16
Yorkshire	C	115	89	0
	F	332	313	89
	M	163	159	17

¹ C : castrats ; F : femelles ; M : mâles entiers.

1.2. Méthodologie

L'approche générale proposée est de supposer qu'un caractère donné est défini par une équation de référence dont les paramètres varient d'un individu à l'autre. Mathématiquement, la valeur d'un caractère Y à un poids w s'exprime comme suit :

$$Y(w) = f_{\vec{x}, \vec{\beta}}(w) \quad [\text{Équation 1}]$$

où f est l'équation de référence, \vec{x} est le vecteur d'identification de l'animal et de son groupe (race, sexe, sélectionneur, ferme, année de naissance, etc.) et $\vec{\beta}$ est le vecteur des paramètres de l'équation. Il est supposé que la forme générale de f est

commune à tous les animaux et que les différences entre les animaux se situent au niveau du vecteur de paramètres $\vec{\beta}$.

1.2.1. Équation de référence

Pour un caractère donné, l'équation de référence était sélectionnée parmi différentes fonctions candidates qui possédaient au moins une des deux caractéristiques suivantes : (i) Représenter conceptuellement l'évolution du caractère et/ou (ii) Être basé sur des fonctions déjà utilisées par le PCAP (Tableau 2). Le nombre d'observations par animal ne permettait pas la sélection de fonctions ayant trop de paramètres. C'est pourquoi par exemple l'équation de Richards, une courbe de croissance à 4 paramètres (France *et al.*, 1996), n'a pas été considérée. C'est plutôt la fonction de Gompertz, courbe de croissance sigmoïdale à 3 paramètres, et sa version simplifiée à 2 paramètres qui ont été testées (Teleken *et al.*, 2017). Pour tous les caractères, à l'exception de l'âge, une équation linéaire, déjà utilisée comme ajustement, était testée de même qu'une équation linéaire avec point de rupture à 120 kg. L'utilisation de cette dernière équation se justifiait par le fait que le programme de sélection ajuste actuellement les caractères à un poids vif de 120 kg et qu'elle permettait de modéliser indirectement la non-linéarité tout en conservant une approche basée sur le mode d'ajustement actuel.

1.2.2. Ajustement individuel

Le choix de l'ajustement individuel revient à la stratégie d'estimation du vecteur de paramètres $\vec{\beta}$. Les approches déterministe et bayésienne ont été utilisées. Notez que dans les exemples d'applications de la présente étude, il est supposé que pour l'animal testé, la prédiction d'un caractère à un poids cible w_1 s'effectue à partir d'une unique observation au poids w_0 . Les deux approches pourraient tout de même être appliquées dans des cas où plusieurs observations sont disponibles.

L'approche déterministe implique que $\vec{\beta}|\vec{X}, Y(w_0)$, soit le vecteur de paramètres pour un individu du groupe défini dans \vec{X} et avec une observation $Y(w_0)$, prend un résultat unique. L'approche est appliquée en deux étapes. Premièrement, un vecteur de paramètres, noté $\vec{\beta}|\vec{X}$, est défini en fonction du groupe (sexe ou race-sexe dans le présent projet). Ensuite, un facteur de correction est appliqué aux paramètres pour que la fonction f passe par le point de mesure $Y(w_0)$. La fonction ainsi modifiée peut prédire le caractère Y à n'importe quel poids w . Cette approche a été appliquée uniquement sur les fonctions linéaires et/ou celles actuellement utilisées dans le programme canadien.

Avec l'approche bayésienne, $\vec{\beta}|\vec{X}, Y(w_0)$ suit une distribution de probabilité qui dépend du groupe de l'animal testé et s'ajuste selon l'observation $Y(w_0)$. Dans la présente étude, la distribution empirique de $\vec{\beta}|\vec{X}, Y(w_0)$ est estimée par une méthode inspirée des k plus proches voisins (k NN pour *k-nearest neighbours*) :

1. Pour chaque animal faisant partie de la population de référence (avec mesures répétées), estimer $\vec{\beta}$. Dans cette étude, la population de référence est la base de données récoltées sur les 13 essais, en excluant les animaux de l'essai de l'animal testé ;

2. Calculer une fonction objectif L qui exprimera la distance entre l'observation $Y(w_0)$ pour l'animal testé et la valeur donnée par $f_{\vec{x}, \vec{\beta}}(w_0)$ provenant de chaque animal de la population de référence. Cette fonction L peut aussi pénaliser ou exclure les animaux de la population de référence qui ne sont pas issus du même groupe que l'animal testé ;
3. Fixer un seuil s et un nombre d'animaux minimal k_0 . Retenir toutes les valeurs de $L < s$. Si le nombre de valeurs retenues est plus petit que k_0 , alors retenir les k_0 plus petites valeurs de L . Le nombre d'animaux retenu est noté par k . Dans la présente étude, k_0 a été fixé à 10 et la valeur de s dépendait du caractère ;
4. Ces k valeurs retenues mènent à k vecteurs de paramètres $\vec{\beta}$ qui sont considérés comme équiprobables pour l'animal testé ;
5. Les k fonctions $f_{\vec{x}, \vec{\beta}}(w)$ résultantes peuvent être utilisées pour estimer la variable d'intérêt pour l'animal testé à n'importe quel poids w .

Dans les exemples d'application présentés, la valeur prédite à un poids w_1 est la médiane des k valeurs $f_{\vec{x}, \vec{\beta}}(w_1)$. L'approche bayésienne est d'autant plus intéressante qu'elle donne une mesure de la variabilité de la prédiction. Appliquée à une population de référence qui évolue, cette approche est considérée comme de l'apprentissage automatique supervisé.

1.3. Comparaison des modèles de prédiction

Le Tableau 2 présente les modèles de prédiction analysés. D'autres modèles ont été testés durant l'étude, mais seuls les plus pertinents sont présentés dans cet article. Les modèles A0, G0 et M0 sont les modèles utilisés dans le PCAP avant le début de l'étude. Seul le modèle pour l'âge (A1) a fait l'objet d'une mise à jour jusqu'à maintenant, grâce aux données du projet. Les autres modèles sont en cours d'évaluation et candidats pour intégrer le PCAP. Il est à noter que le programme canadien n'utilise actuellement aucun modèle d'ajustement pour le GIM. Pour chaque caractère, les modèles ont été évalués pour leur capacité à prédire une valeur à un poids w_1 à partir d'une observation à un poids w_0 , tous deux situés dans l'intervalle 75-160 kg.

L'évaluation des équations de référence était basée sur la distribution des résidus, que l'on souhaite centrée à 0 et peu dispersée. La moyenne, l'écart type et la racine carrée de l'erreur quadratique moyenne (REQM) des résidus ont été calculés pour chaque classe de poids (75-90, 90-105, 105-120, 120-140 et 140-160 kg). Le nombre de paramètres des équations était aussi considéré.

La performance de la méthode d'ajustement peut aussi être évaluée à partir de la distribution des résidus, en comparant les modèles provenant de la même équation de référence. De plus, pour les ajustements par l'approche bayésienne, des diagrammes de Talagrand ont été produits afin d'évaluer si la distribution des k prédictions couvre adéquatement les valeurs à prédire. La variabilité de la distribution a été mise en relation avec le résidu afin d'évaluer s'il est possible de prédire la qualité de la prédiction.

Tableau 2 – Modèles de prédiction analysés pour tous les caractères à l'étude

Variable	Modèle	f^1	Groupe	Ajustement	Source des données ²
Âge, jours	A0	$Y(w_1) = Y(w_0) - \frac{w_0 - w_1}{\beta_0 w_0 / Y(w_0)}$	Sexe (F, C/M)	Déterministe	Historique
	A1	Comme A0, avec point de rupture à 120 kg	Sexe (F, C/M)	Déterministe	Projet
	A2	$w = \beta_1 \exp \left\{ \log \left(\frac{1.5}{\beta_1} \right) \exp(-\beta_2 Y(w)) \right\}$	Sexe (F, C/M)	Bayésien	Projet
	A3	Comme A2	Sexe (F, C, M)	Bayésien	Projet
	A4	Comme A2	Race et sexe	Bayésien	Projet
ELD, mm	G0	$Y(w_1) = Y(w_0) \frac{\beta_1 + \beta_2 w_1}{\beta_1 + \beta_2 w_0}$	Race et sexe	Déterministe	Historique
	G1	Comme G0	Race et sexe	Déterministe	Projet
	G2	Comme G0, avec point de rupture à 120 kg	Race et sexe	Déterministe	Projet
	G3	$Y(w) = \beta_1 + \beta_2 w$	Race et sexe	Bayésien	Projet
EMD, mm	M0	$Y(w_1) = Y(w_0) \frac{\beta_1 + \beta_2 w_1}{\beta_1 + \beta_2 w_0}$	Race et sexe	Déterministe	Historique
	M1	Comme M0	Race et sexe	Déterministe	Projet
	M2	Comme M0, avec point de rupture à 120 kg	Race et sexe	Déterministe	Projet
	M3	$Y(w) = \beta_1 \exp \left\{ \log \left(\frac{25}{\beta_1} \right) \exp(-\beta_2 w) \right\}$	Race et sexe	Bayésien	Projet
GIM, %	P1	$Y(w_1) = Y(w_0) \frac{\beta_1 + \beta_2 w_1}{\beta_1 + \beta_2 w_0}$	Sexe (F, M)	Déterministe	Projet
	P2	$Y(g_1) = Y(g_0) \frac{\beta_1 + \beta_2 g_1}{\beta_1 + \beta_2 g_0}$	Sexe (F, M)	Déterministe	Projet
	P3	$Y(w) = \beta_1 + \beta_2 w$	Sexe (F, M)	Bayésien	Projet
	P4	$Y(g) = \beta_1 + \beta_2 g$	Sexe (F, M)	Bayésien	Projet

¹ Le facteur de correction est inclus à la fonction f pour les modèles avec l'approche déterministe. Pour les modèles P2 et P4, la variable g est l'ELD.

² Source des données pour l'estimation des paramètres. La valeur « Historique » indique que l'équation et ses paramètres n'ont pas été modifiées. La valeur « Projet » indique que les paramètres sont estimés sur les animaux des 13 essais, à l'exception de l'essai de l'animal testé.

2. RESULTATS

2.1. Analyse des résidus

Le biais, l'écart-type et la REQM de chaque modèle sont présentés au Tableau 3.

Pour l'âge, les résultats indiquent clairement que l'ancien modèle du PCAP (A0) est plus biaisé que les autres ; il sous-estime (respectivement surestime) l'âge lorsque le poids cible est plus petit (respectivement plus grand) que le poids observé. Ce modèle a été construit à l'origine pour couvrir l'intervalle 75-115 kg (Kennedy et Uribe, 1994) et non pour des poids allant jusqu'à 160 kg. Le modèle A1, construit selon le modèle A0, mais avec un point de rupture à 120 kg pour mieux couvrir les poids plus lourds, est plus performant, mais généralement moins que les modèles basés sur l'équation de Gompertz à 2 paramètres avec méthode d'ajustement bayésienne. Pour ces modèles, le biais est généralement très faible et les modèles traitant les sexes séparément (A2 et A3) sont légèrement plus performants que le modèle par sexe et race (A4). En imposant des critères plus précis pour la définition du groupe, le nombre d'animaux par groupe issus de la population de référence diminue. C'est une cause probable de la dégradation des résultats pour le modèle A4 par rapport aux modèles A2 et A3. Pour l'ELD, les performances sont similaires entre les modèles, indiquant que le point de rupture à l'équation linéaire n'améliore pas substantiellement les résultats par rapport à l'équation linéaire. La sélection des paramètres de l'équation a un impact limité sur la distribution des résidus.

Tableau 3 – Biais absolu, écart-type et REQM moyens pour chaque modèle analysé. Entre parenthèses, la valeur la plus élevée parmi les 20 combinaisons de classes de poids cibles et observés.

Variable	Modèle ¹	Biais	Écart-type	REQM
Âge, jours	A0	5,2 (11,9)	5,3 (7,4)	7,6 (12,9)
	A1	2,1 (7,1)	5,3 (7,7)	5,9 (9,2)
	A2	0,7 (2,0)	5,5 (7,6)	5,6 (7,7)
	A3	0,7 (2,2)	5,4 (7,6)	5,5 (7,6)
	A4	0,8 (2,2)	5,7 (9,7)	5,8 (10,0)
ELD, mm	G0	0,2 (0,6)	1,9 (3,3)	1,9 (3,3)
	G1	0,2 (0,4)	1,9 (3,3)	1,9 (3,3)
	G2	0,1 (0,2)	1,9 (3,3)	1,9 (3,3)
	G3	0,2 (0,4)	1,9 (3,1)	1,9 (3,1)
EMD, mm	M0	1,1 (2,4)	4,8 (6,5)	4,9 (6,7)
	M1	0,6 (1,1)	4,8 (6,5)	4,8 (6,6)
	M2	0,2 (0,7)	4,8 (6,5)	4,8 (6,6)
	M3	0,2 (0,8)	4,7 (6,0)	4,7 (6,0)
GIM, %	P1	0,23 (0,61)	0,85 (1,5)	0,90 (1,6)
	P2	0,25 (0,62)	0,80 (1,4)	0,85 (1,5)
	P3	0,25 (0,83)	0,98 (2,5)	1,02 (2,7)
	P4	0,20 (0,48)	0,80 (1,2)	0,84 (1,3)

¹ Les modèles sont définis au **Erreur ! Source du renvoi introuvable.**

Pour l'EMD la mise à jour des paramètres (M1) de l'équation linéaire utilisée dans le PCAP (M0) a permis de diminuer le biais. La prise en compte de la non-linéarité diminue encore plus le biais, que ce soit par un point de rupture à 120 kg à l'équation linéaire (M2) ou en utilisant l'équation de Gompertz (M3).

Pour le GIM, les modèles linéaires en fonction de l'épaisseur de lard (P2 et P4) ont un REQM légèrement inférieur aux modèles linéaires en fonction du poids (P1 et P3).

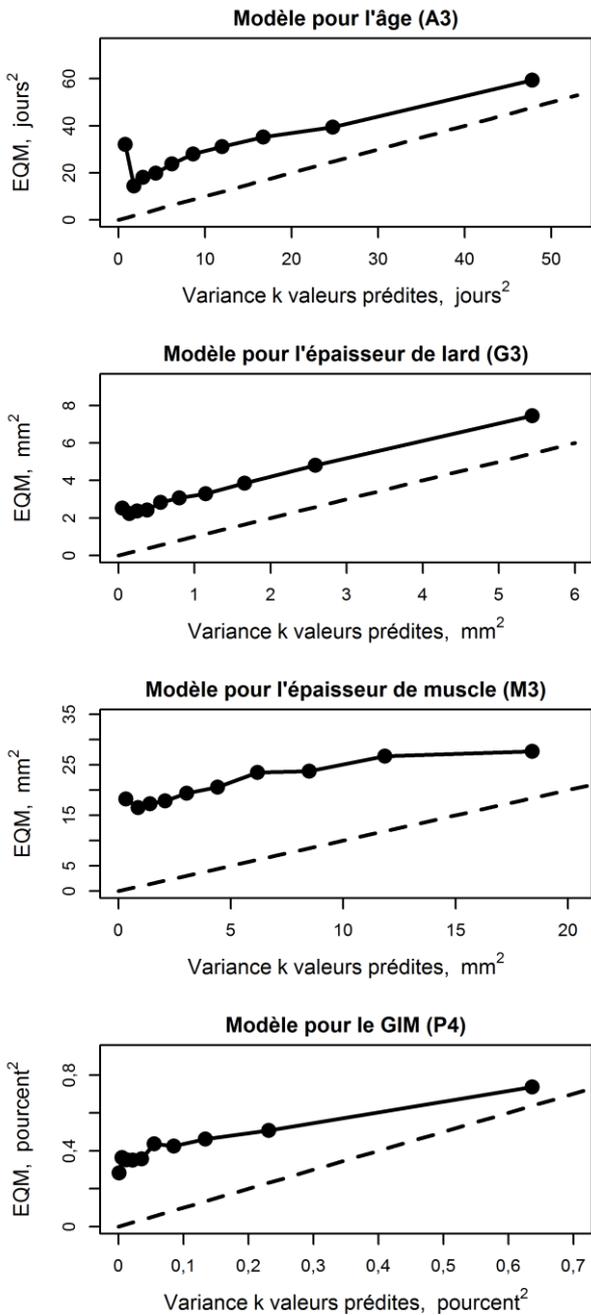


Figure 1 – EQM de la prédiction en fonction de la variance de l'ensemble des valeurs générées par l'approche bayésienne.

2.2. Analyse de la variabilité

Chaque prédiction a une incertitude que les méthodes d'ajustement déterministes, actuellement utilisées pour tous les caractères dans le PCAP, ne prennent pas en compte. Il pourrait s'avérer utile pour le programme de sélection d'accompagner chaque prédiction d'une mesure d'incertitude, que ce soit par un écart-type ou par un indice de fiabilité qui serait fournie au sélectionneur. Dans cette optique, il est pertinent de vérifier si la variabilité des *k* valeurs prédites avec la méthode d'ajustement bayésienne permet de déterminer *a priori* l'incertitude d'une prédiction.

Pour un caractère donné, toutes les prédictions effectuées à partir d'un modèle avec méthode d'ajustement bayésienne étaient réparties en 10 classes selon la variance des *k* valeurs :

variance inférieure au 10^e centile, variance entre le 10^e et le 20^e centile, et ainsi de suite jusqu'à une variance entre le 90^e et le 99^e centile (les valeurs extrêmes ont été ignorées). Donc, il y avait 10 groupes de tailles similaires pour lesquelles l'EQM de la prédiction était calculée (Figure 1).

Pour les quatre caractères, l'EQM augmente avec la variance de l'ensemble des valeurs générées, ce qui implique qu'il est possible d'avoir une indication de la fiabilité de la prédiction à partir de la méthode proposée. Cependant, même s'il y a corrélation, l'EQM est nettement supérieure à la variance de l'ensemble des valeurs générées. Cette différence pourrait être causée par (i) un biais des modèles ou (ii) une sous-dispersion générée par l'approche bayésienne (i.e. l'approche bayésienne ne reproduit pas la variabilité réelle des données). Le biais des modèles étant relativement faible (Tableau 3) la sous-dispersion est la cause la plus plausible, ce que confirme le diagramme de Talagrand (Figure 2). Pour les quatre caractères, la valeur réelle à prédire est trop souvent en dehors ou à la limite de l'ensemble généré par l'approche bayésienne.

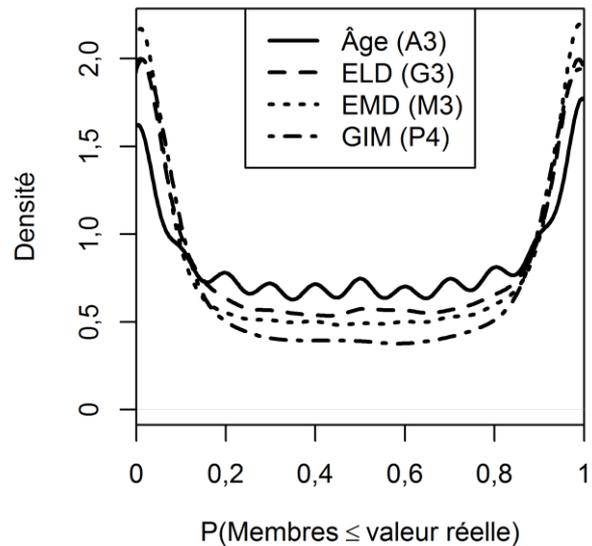


Figure 2 – Diagramme de Talagrand pour les quatre caractères. Un modèle parfait aurait une densité de 1 partout.

3. DISCUSSION

L'étude a permis de mettre à jour les équations pour le PCAP. Pour l'âge et l'EMD, des courbes de croissance (fonction de Gompertz à deux paramètres) ont été retenues. Pour l'âge, la réduction du biais est majeure : environ 4,5 jours en moyenne sur l'ensemble des strates de poids (Tableau 3) et environ 10 jours lorsque le poids cible se situe entre 75 et 90 kg et le poids d'observation entre 140 et 160 kg. La réduction du biais est plus modeste pour l'EMD, mais les résultats ont tout de même permis de détecter une non-linéarité avec seulement cinq observations par individu.

L'équation linéaire, déjà utilisée dans le PCAP, a été retenue pour l'ELD. Ce résultat n'implique pas nécessairement la linéarité, mais plutôt que la non-linéarité, si elle est présente, n'était pas assez importante pour améliorer les résultats à partir de cinq observations par animal.

Pour le GIM, la corrélation est plus forte avec l'ELD qu'avec le poids. Plus d'individus seraient nécessaires pour affiner les modèles ; pour l'instant, les castrats sont exclus et la majorité des données proviennent de la race Duroc.

L'approche d'ajustement bayésienne a montré des résultats intéressants. Pour le GIM, le biais a légèrement diminué par rapport à l'utilisation de l'approche déterministe (facteur multiplicatif) à partir d'une même équation de référence (Tableau 3). Mais la principale avancée de l'utilisation d'une telle approche est la caractérisation de la variabilité de la prédiction. À ce niveau, les résultats montrent un certain potentiel pour donner une appréciation de la fiabilité des prédictions (Figure 1), mais le plein potentiel de l'approche n'a pas encore été atteint.

Le premier élément qui permettrait l'amélioration des prédictions serait l'augmentation de la taille du jeu de données. La taille de la base de données acquise durant l'étude a permis de faire des avancées intéressantes, mais n'a pas permis de créer des groupes homogènes assez précis permettant d'avoir une prédiction ciblée pour l'animal. Ce fût le cas pour l'âge, où le modèle par sexe et par race (A4) était légèrement moins performant que les modèles par sexe (A2 et A3), et pour le GIM, où le nombre de données ne permettait pas d'envisager un modèle par race et par sexe.

L'augmentation de la taille de la base de données permettrait d'améliorer la fonction objectif L en raffinant les définitions des groupes et en intégrant plusieurs critères de performances. Par exemple, la sélection des k animaux de la population de référence pourrait se faire en considérant l'ensemble des caractères (ex. : utiliser l'ELD, l'EMD, le GIM et ultimement la prise alimentaire pour prédire l'âge à un poids donné). Pour ce qui est de la définition des groupes, elle n'a pas à se résumer au sexe et à la race, mais peut tenir compte de toute information, comme par exemple la ferme, le technicien, l'année (pour tenir compte de l'évolution génétique), les dates (pour tenir compte de l'effet de l'ambiance), etc.

Avec l'optique d'augmenter la taille de la base de données et d'ajouter des critères à considérer, d'autres approches plus performantes d'apprentissage automatique pourraient être envisagées. Aussi, la sélection des animaux dans la population de référence pourrait se faire sur les données brutes, sans passer par une équation de référence. Il est à noter que, dans la présente étude, les animaux de référence étaient caractérisés par les paramètres de l'équation de référence et non par leurs données brutes. La prédiction des performances sans équation de référence représenterait un changement majeur dans l'approche de standardisation des données. Elle nécessiterait

sans doute des efforts de vulgarisation auprès des éleveurs et utilisateurs des performances ajustées et des valeurs génétiques. Un réel avantage cependant serait la capacité d'exploiter le plein potentiel des données brutes, notamment au niveau de la variabilité, que l'approche actuelle, basée sur les paramètres d'équation, sous-estime (Figure 2).

En outre, l'approche proposée requiert que des essais soient réalisés périodiquement chez les sélectionneurs. Le protocole de la mise en œuvre des essais (fréquence des essais par sélectionneur, nombre d'animaux par sexe et par race, nombre de mesures par animal, collectes de nouvelles données, etc.) reste à planifier avec les sélectionneurs. Elle dépendra notamment de l'approche d'ajustement qui sera retenue.

CONCLUSION

L'objectif de l'étude était de proposer pour le PCAP des modèles d'ajustement plus flexibles, en particulier qui s'adaptent à différents poids. L'équation de Gompertz à deux paramètres, qui représente conceptuellement la croissance non-linéaire de l'animal, a été retenue pour le poids et l'EMD. L'amélioration de la prédiction de l'âge à un poids donné est très importante par rapport à l'équation classique utilisée auparavant, développée sur des animaux plus légers. Pour l'ELD, l'équation linéaire peut encore s'appliquer et pour le GIM, des modèles basés sur l'ELD ont montré de bons résultats, mais plus de données seraient nécessaires pour affiner le modèle.

Au-delà de la sélection de nouveaux modèles, la présente étude visait à évaluer une nouvelle approche de standardisation évolutive pour le PCAP. La nouvelle approche bayésienne montre des résultats intéressants, notamment en ce qui a trait à la représentation de la variabilité. Or, le plein potentiel de cette approche pourra s'exprimer avec la prise de nouvelles mesures répétées et ce, de manière périodique. Elle pourrait aussi être bonifiée par l'inclusion de nouvelles variables issues de nouveaux capteurs et si nécessaire, par les nouvelles techniques d'analyse de données massives. La stratégie de mise en œuvre reste à élaborer avec les sélectionneurs.

REMERCIEMENTS

Ce projet a été réalisé avec l'aide des organisations de sélection participant au Programme canadien pour l'amélioration des porcs et avec le soutien financier du Programme Agri-Science d'Agriculture et Agroalimentaire Canada.

REFERENCES BIBLIOGRAPHIQUES

- Agriculture et Agroalimentaire Canada, 2020. Information sur le marché des viandes rouges. Rapport des poids moyens de carcasses chaudes dans les abattoirs inspectés fédéralement (009D).
- Alonso J., Villa A., Bahamonde A., 2015. Improved estimation of bovine weight trajectories using Super Vector Machine Classification. *Computers and Electronics in Agriculture*, 110, 36-41.
- Alsahaf A., Azzopardi G., Ducro B., Hanenberg E., Veerkamp R.F., Petkov N., 2018. Prediction of slaughter age in pigs and assessment of the predictive value of phenotypic and genetic information using random forest. *J. Anim. Sci.*, 96, 4935-4943.
- France, J., Dijkstra, J., Dhanoa, M. S., 1996. Growth functions and their application in animal science. *Annales de Zootechnie*, 45(Suppl1),165-174.
- Kennedy B.W. & Uribe H., 1994. Adjustments for weight at probing to 105 kg. Report to CCSI Genetics Committee, September 1994. 7 pages.
- Lonergan M., Topel D.G., & Marple D.N., 2019. Growth curves and growth patterns. *In The Science of Animal Growth and Meat Technology (Second Edition)* p71-109.
- Schinckel A. P. & de Lange C.F., 1996. Characterization of growth parameters needed as inputs for pig growth models. *J. Anim. Sci.* 74, 2021-2036.
- Schinckel A. P., Mahan D. C., Wiseman T. G., & Einstein M. E., 2008. Growth of protein, moisture, lipid, and ash of two genetic lines of barrows and gilts from twenty to one hundred twenty-five kilograms of body weight. *J. Anim. Sci.* 86, 460-471.
- Schull C.M., 2013. Modeling growth of pigs reared to heavy weights. PhD Thesis, University of Illinois at Urbana-Champaign, 143 pages.
- Teleken J.T., Cazonatto Galvão A., da Silva Robazza W., 2017. Comparing non-linear mathematical models to describe growth of different animals. *Acta Scientiarum. Animal Sciences*, 39, 73-81.
- Wellock I. J., Emmans G.C., & Kyriazakis I., 2004. Describing and predicting potential growth in the pig. *Anim. Sci.* 78, 379-388.