Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

Institut National Polytechnique de Toulouse (INP Toulouse)

Présentée et soutenue par

Van Hung (Huynh) TRAN

Le 11 Oct 2018

NEW GENETIC LONGITUDINAL MODELS FOR FEED EFFICIENCY

École doctorale et discipline ou spécialité

ED SEVAB : Pathologie, Toxicologie, Génétique et Nutrition

Unité de recherche

UMR1388 - Génétique, Physiologie et Systèmes d'Elevage (GenPhySE) - INRA

Directrices de thèse : Ingrid DAVID, Hélène GILBERT

Jury

M. FARNIR Frédéric, professeur à l'université vétérinaire de Liège : Rapporteur
Mme DUPONT-NIVET Mathilde, directrice de recherche, INRA, Jouy-en-Josas : Rapporteuse
Mme PHOCAS Florence, directrice de recherche, INRA, Jouy-en-Josas : Examinatrice
M. BOUQUET Alban, ingénieur, IFIP, Le Rheu : Examinateur
Mme DAVID Ingrid, chargée de recherche, INRA, Toulouse : Directrice de thèse

RESUME

Bien que des approches non génétiques et génétiques aient été mises en œuvre pour améliorer l'efficacité alimentaire des animaux au cours des dernières années, le coût des aliments constitue encore la plus grande partie des coûts de production de nombreux systèmes d'élevage. De plus, une efficacité alimentaire non optimale augmente l'impact environnemental de l'élevage par gaspillage d'aliments. Au cours des dernières décennies, les progrès dans les technologies à haut débit pour la gestion des animaux, notamment le développement de distributeurs automatiques d'aliment, permettent d'obtenir de nombreuses mesures répétées au cours du temps (données longitudinales) de la consommation individuelle des animaux. L'objectif de cette thèse était de développer de nouveaux modèles génétiques pour mieux quantifier le potentiel génétique des animaux pour l'efficacité alimentaire en utilisant ce type de données. Les données de 2435 porcs Large White en croissance provenant d'une expérience de sélection divergente pour la consommation résiduelle ont été utilisées au cours de cette thèse. Dans cette population, les mâles étaient pesés chaque semaine alors que les femelles et les mâles castrés étaient pesés tous les mois entre 10 et 23 semaines d'âge. Dans une première étape, nous avons comparé différentes approches pour prédire les poids corporels hebdomadaires manquants de ces animaux afin d'améliorer l'évaluation de leur efficacité alimentaire. Pour la période testée, une interpolation quasi linéaire basée sur les semaines adjacentes est la meilleure approche pour traiter les poids corporels manquants dans notre ensemble de données. Dans un second temps, différents modèles longitudinaux tels que les modèles de régression aléatoire (RR), les modèles antédependants structuraux (SAD) et les modèles de type « character process » ont été comparés pour l'analyse de l'efficacité alimentaire. La comparaison a été réalisée en se basant sur des critères d'ajustement aux données (Log vraisemblance, Critère d'Information Bayésien), sur les estimations des composantes de variances (estimations de l'héritabilité, des variances génétiques et des corrélations génétiques entre semaines) et sur la capacité prédictive (coefficients de concordance de Vonesh) de chaque modèle. Les résultats ont montré que le modèle SAD est le plus parcimonieux pour l'indice de consommation (IC) et la consommation résiduelle, deux mesures d'efficacité alimentaire. Ce modèle fournit également des capacités prédictives similaires à celles des autres modèles. Un critère de sélection combinant les prédictions des valeurs génétiques hebdomadaires de chaque animal a été proposé pour des applications pratiques de ces modèles dans un objectif de sélection. En outre, nous avons évalué comment l'information génomique pouvait améliorer la précision des prédictions des valeurs génétiques des animaux pour le gain moyen quotidien et la consommation résiduelle, en appliquant des approches génomiques « single-step » aux modèles RR et SAD. Les résultats obtenus ont montré que les précisions étaient faibles et les biais de prédiction importants pour les deux caractères, et qu'ils n'étaient pas améliorés par l'apport de l'information génomique. Enfin, nous avons montré que la sélection divergente pour la consommation résiduelle avait un impact majeur sur les trajectoires de l'indice de consommation et de la consommation résiduelle pendant la croissance dans chaque lignée. En conclusion, cette thèse a montré que la sélection

pour des trajectoires d'efficacité alimentaire est faisable avec les informations disponibles actuellement. Des études supplémentaires sont nécessaires pour mieux évaluer le potentiel de l'information génomique avec ces derniers modèles et pour valider les stratégies de sélection sur ces trajectoires d'efficacité alimentaire au cours du temps dans la pratique.

ABSTRACT

Although non-genetic and genetic approaches heavily improved feed efficiency in the last decades, feed cost still contributes to a large proportion of pork production costs. In addition, the limited efficiency of feed use not only increases the environmental impact due to the waste of feed. Over the last decades, advances in high-throughput technologies for animal management, including automatic self-feeders, created a proliferation of repeated data or longitudinal data. The objective of this thesis was to develop new genetic models to better quantify the genetic potential of animals for feed efficiency using longitudinal data on body weight (BW), feed intake and body composition of the animals. Data from 2435 growing Large White pigs from a divergent selection experiment for residual feed intake (RFI) were used. In this population, males were weighted every week whereas females and castrated males were weighted every month at the beginning of the test (10 weeks of age) and more often towards the end of the test (23 weeks of age). In a first step, different approaches investigated how to predict missing weekly BW for intermediate stages. For the tested period, a quasi linear interpolation based on the adjacent weeks is the best approach to deal with missing BW in our dataset. In a second step, different longitudinal models, such as random regression (RR) models, structured antedependence models (SAD) and character process models, in which the covariances between weeks are accounted for, were compared. The comparison focused on best-fit to the data criteria (Loglikelihood, Bayesian Information Criterion), on variance components estimations (heritability estimates, genetic variances and genetic correlations between weeks) and on predictive ability (Vonesh concordance coefficients). The results showed that SAD is the most parsimonious model for feed conversion ratio (FCR) and for RFI, two measures of feed efficiency. The SAD model also provided similar predictive abilities as the other models.A selection criterion combining the weekly breeding values was proposed for practical applications to selection. In addition, we evaluated the potential of genomic information to improve the accuracy of breeding value predictions for average daily gain and residual feed intake, applying single step genomic approaches to the RR and SAD models. In our dataset, prediction accuracies was low for both traits, and was not much improved by genomic information. Finally, we showed that divergent selection for RFI had a major impact on the FCR and RFI profile trajectories in each line. In conclusion, this thesis showed that selection for trajectories of feed efficiency is feasible with the current available information. Further work is needed to better evaluate the potential of genomic information with these models, and to validate strategies to select for these trajectories in practice.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This work is part of the H2020 Feed-a-Gene project, financially supported by the European Union's H2020 Program under grant agreement no. 633531 and Animal genetics divisions, INRA, France. This thesis is also for the purpose to demand the label 'EIR-A" in the frame of the program of "Ecole Internationale de Recherche d'Agreenium (EIR-A)."

I would like to express my gratitude to my supervisors, "my scientific mothers," Ingrid and Hélène for their tremendous support, and their availability during the past three years.

I would like to send my thanks to all members of the jury: Pr. Frédéric Farnir, Dr. Florence Phocas, Dr. Mathilde Dupont-Nivet, Dr. M. Alban Bouquet for agreeing to evaluate this work.

I am especially greatful to all member of my thesis commitee : Mme Christèle Robert-Granié, M. Ludovic Brossard, M. Rémi Servien, Mme Hélène Croiseau-Leclerc pour m'avoir donné beaucoup de conseils pendant les dernières années.

Thank Dr. Jack Dekkers and his group at ISU for their great help during my internship .

Dr. Zulma is acknowledged for being accepted as my professional mentor during my thesis.

Thank my statistics professor, Mr. Réne Ecochard for the first lessons about linear mixed models; Catherine Larzul, for accepting as my "directrice de thèse." for the first year of my Ph.D; Andres Legara, my co-office, for always being friendly with me (Muchas gracias !).

Thank to Modgen teams and INRA staffs for always supporting me.

Je suis reconnaisant envers tous mes amis et mes proches qui m'ont soutenu au cours de cette thèse : équipe Modgen, Céline, Julien et petite Arianne, Emilie (maman), papi (Gilbert), Xavier, Claire, Marc, Silvia, Llibertat et autres thésards, stagiaires à Genphyse.

Merci aux tous mes professeurs en France et au Vietnam.

Merci aux mes amis vietnamiens et français, surtout, anh Huân & chị Lê đã bên cạnh e suốt 3 năm qua !

Last but by no means least, I would like to thank my family for encouraging me over the 28 years.

Castanet Tolosan, August 16th, 2018

Van Hung (Huynh) TRAN

LIST OF TABLES (excluding publications)

LIST OF FIGURES (excluding publications)

LIST OF ACRONYMS

ADG	Average Daily Gain
ADFI	Average Daily Feed Intake
AIC	Akaike Information Criterion
AMBW	Average Metabolic Body Weight
AWFI	Average Weekly Feed Intake
BFT	Backfat Thickness
BIC	Bayesian Information Criterion
BLUP	Best Linear Unbiased Prediction
BW	Body Weight
EBV	Estimated Breeding Value
FI	Feed Intake
GEBV	Genomic Estimated Breeding Value
LMA	Loin Muscle Area
MSEP	Mean Square Error of Predictions
RR	Random Regression
RR-OP	Random Regression Model using Legendre Orthogonal Polynomials
RR-SL	Random Regression Model using Splines
SAD	Structured Antedependence
SD	Standard Deviation
SE	Standard Error
SVM	Support Vector Machines
VCC	Vonesh Concordance Coefficient
WDFI	Weekly Averages of Daily Feed Intake

Chapter 1 - LITERATURE REVIEW

GENETIC MODELS FOR FEED EFFICIENCY

1.1. Feed efficiency - opportunities and challenges

According to the United Nations Population Fund, the global population reached 7.6 billion in 2017, and is expected to reach 9.1 billion by 2050 (Godfray et al., 2010). To comply the food security requirements of these people, we need to increase by 70% our food production in a context of increases in energy price, water depletion, loss of farmlands due to urbanization, and flooding and droughts caused by climate change (Nikos Alexandratos and Bruinsma, 2012). Feed efficiency is defined as the ability to transform input (feed) into output (such as body weight). Improving it decreases the requirements for resources related to input production (water, farmland), and reduces nutrient losses in manure. Thus, feed efficiency is a key factor to increase the profitability of livestock production, and to decrease its environmental impact. Although both genetic and non-genetic approaches (including nutritional management) have been conducted over the last decades to increase feed efficiency, feed costs still account for a significant part of animal production (for pork production: 75% in the Occidental countries, 70% in France) (Patience et al., 2015; Gilbert et al., 2017).

Different criteria have been proposed in the literature to quantify feed efficiency, such as feed conversion ratio (FCR) and residual feed intake (RFI) (Patience et al., 2015). In general, feed efficiency traits are heritable. They are affected by many factors, depending on the animal itself, such as animal genetics, sex, immunological status, growth rate, and on external factors, such as feed ingredients, nutrient composition, thermal environment, feed processing and delivery, etc (Patience et al., 2015), and also gut microbiome (Quan et al., 2018). In few studies, it has been reported that feed intake and feed efficiency traits change over time (David et al., 2015; Shirali et al., 2015). A selection strategy accounting for these changes could provide better genetic gains than the current selection. Due to lack of repeated measurements for this phenotype and its components (for instance, feed intake and body weight gains over successive periods), selection is usually applied to one single record of feed efficiency over a given test period. Thanks to automatic devices combined with electronic identification (i.e., automatic self-feeders), repeated records related to growth (body weight) and feed consumption during a test period are now available. It thus makes it possible to calculate a repeated measurement of feed efficiency at the individual level.

This type of data has specific properties: repeated data for each animal are correlated. Therefore, there is a need for special statistical methods that can not only account for changes of the mean curve over time, but can also handle change of covariances between measurements. To meet

these constrainsts, the analysis has to be “efficient” in treating a large number of observations for each subject.

1.2. Feed efficiency, related traits and genetic evaluation

In growing animals, there are several criteria to evaluate feed efficiency based on input (feed intake) and output (e.g. weight gain and its composition). The two main criteria in pigs are feed conversion ratio and residual feed intake. Some other measures are reported in the literature to quantify feed efficiency: partial efficiency of growth (Kellner, 1909), relative growth rate (Fitzhugh and Taylor, 1971), Kleiber ratio (Kleiber, 1947), and residual daily weight gain (Koch et al., 1963; Coyne et al., 2017), but their use is scarce.

1.2.1. Feed conversion ratio

Feed conversion ratio is the feed consumed per unit of body weight gain over a given period of growth (Brody, 1945), i.e. the inverse of feed efficiency. A low FCR is thus desirable in animal production. Since FCR is the ratio between feed intake (average daily feed intake ADFI) and body weight gain (average daily gain ADG), there might be two animals with the same FCR but with different ADFI and ADG. The consequences of selection for FCR on correlated traits such as ADG, ADFI and body composition are usually difficult to predict: the statistical properties of ratios are not optimum for breeding, which essentially assumes linear dependencies between traits. Only very recently, Shirali et al., (2018) proposed an exact Bayesian approach to predict genetic gains for FCR without assumptions or approximations. To overcome this limit, several criteria with different statistical properties have been proposed.

1.2.2. Residual feed intake

Residual feed intake is the difference between the observed feed intake and the expected feed intake based on animal requirements for maintenance and production. The notion of “residual feed intake” was mentioned for the first time in 1941 by Byerly (1941). Then, RFI was firstly proposed for evaluating feed efficiency in beef cattle by Koch et al. (1963), and later introduced in laying hens (Bordas and Merat, 1981), in pigs (de Haer et al., 1993), and other species as well as rabbits (Larzul and de Rochambeau, 2005); rainbow trout (Silverstein et al., 2005); broilers (Aggrey and Rekaya, 2013); ducks (Drouilhet et al., 2014). Depending on how the RFI definition is applied and to which species and production type, there are different approaches to compute RFI. Generally, RFI is calculated as the residual of a multiple linear regression that

includes as covariates indicators of production requirements (ADG and an estimator of body weight gain composition in species where its variability is sufficient, such as backfat thickness in pigs (BFT)), and of maintenance requirements (average metabolic body weight (AMBW)). An animal with reduced RFI is then an animal that eats less than predicted based on its performance, so it is more efficient.

Depending if the multiple linear regression accounts for phenotypic or genetic correlations, RFI is independent of the traits used in the regression at the genetic or at the phenotypic level. Thus, selection for reduced RFI should not have detrimental effects on animal growth or size, which is one of the main advantages of using RFI as a measure of feed efficiency trait instead of FCR. The other main advantage of this trait is that it has linear relationships with other traits of the breeding objectives, so it is easier to compute weights for selection indexes and predict genetic gains than for FCR. In selection experiments in pigs, selection for low RFI has been associated with lower feed intakes and leaner carcasses, with moderate effects on growth rate (Young and Dekkers, 2012; Gilbert et al., 2017).

Due to practical constraints, the use of RFI in animal selection could suffer from some limitations. First, the errors in measures of ADG, BFT, and AMBW potentially result into "noise" contributing to RFI. Secondly, if some covariates are missing, RFI cannot be calculated for a given individual. Finally, because body composition measurements are needed for RFI, that are not used to compute FCR, so it is more expensive to collect. However, body composition is usually measured *in vivo* on candidates or on carcass in most pig selected populations, so the actual additional costs are null compared to a classical pig selection scheme.

1.2.3. Heritabilities and genetic correlations between traits related to feed efficiency

Heritabilities reported in the literature for feed efficiency traits in pigs are presented in Table 1.1. In general, FCR has moderate to high heritability estimates (ranging from 0.15 to 0.45). The heritability estimates for RFI range from 0.10 to 0.44 (Nguyen et al., 2005; Gilbert et al., 2007; Cai et al., 2008; Saintilan et al., 2013).

Phenotypic and genetic correlations between feed efficiency and other traits are shown in Tables 1.2. As mentioned in many studies (Gilbert et al., 2007; Hoque et al., 2007; Cai et al., 2008), FCR is highly correlated with ADFI, ADG at the phenotypic level whereas RFI is generally independent or has low correlations with growth and body composition traits (ADG and BFT) at both the phenotypic and the genotypic levels. In addition, a high correlation between RFI and FCR (0.84, Hoque et al., 2007, 0.71, Gilbert et al., 2017) is estimated, confirming that selection for decreasing RFI improves FCR (Hoque et al., 2009).

Table 1.1. Heritability estimates of feed efficiency traits in pigs

Trait	Population	Heritability	Reference
FCR	Large White	0.45±0.07	Gilbert et al., 2007
	Large White (10694 dams/2342 sires)	0.30±0.03/0.30±0.06	Saintilan et al., 2013
	1047 Duroc boars	0.32±0.09	Jiao et al., 2014
	Large White, 22984 purebreds/8657 crossbred (commercial animals)	0.17±0.01 (purebred)/ 0.15±0.02 (crossbred)	Godinho et al., 2018
RFI	Large White pigs selected for growth rate on restricted feeding	0.22 to 0.24 ±0.08	Nguyen et al., 2005
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line)	0.27 to 0.36	Cai et al., 2008
	Large White	0.24±0.03	Gilbert et al., 2007
	Large White (10694 dams/2342 sires)	0.21±0.03/0.26±0.06	Saintilan et al., 2013
BW	1047 Duroc boars	0.34±0.28	Jiao et al., 2014
BFT	1047 Duroc boars	0.58±0.09	Jiao et al., 2014
	1622 Duroc pigs	0.53±0.15	Ito et al., 2018
	Large White, 22984 purebreds/8657 crossbred (commercial animals)	0.47±0.01 (purebred)/ 0.43±0.02 (crossbred)	Godinho et al., 2018
ADFI	1047 Duroc boars	0.66±0.11	Jiao et al., 2014
	3096 Large White pigs	0.11-0.55	David et al., 2015
	1622 Duroc pigs	0.42±0.17	Ito et al., 2018
	Large White, 22984 purebreds/8657 crossbred (commercial animals)	0.23±0.02 (purebred)/ 0.28±0.03 (crossbred)	Godinho et al., 2018
ADG	1642 Duroc (380 boars, 868 gilts, and 394 barrows)	0.48 ± 0.03	Hoque et al., 2009
	1047 Duroc boars	0.44±0.11	Jiao et al., 2014
	1622 Duroc pigs	0.45±0.13	Ito et al., 2018
	Large White, 22984 purebreds/8657 crossbred (commercial animals)	0.23±0.01 (purebred) 0.26±0.01 (crossbred)	Godinho et al., 2018

ADFI = Average daily feed intake; ADG = Average daily gain; AMBW = average metabolic body weight; RFI = residual feed intake; BFT = Backfat thickness

Table 1.2. Genetic correlations ± standard errors between feed efficiency and other traits in pigs

trait1/trait2	Population	Genetic correlation	Phenotypic correlation
FCR/RFI	Large White (Gilbert et al., 2007)	0.71±0.12	0.63
	1047 Duroc boars (Jiao et al, 2014)	0.53±0.31	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.82±0.02 (purebred)/ 0.55±0.06 (crossbred)	
ADFI/RFI	Large White (Gilbert et al., 2007)	0.77±0.10	0.68
	1047 Duroc boars (Jiao et al, 2014)	0.07±0.09	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.73±0.05 (purebred)/ 0.65±0.05 (crossbred)	
	Purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	0.52±0.12	0.63±0.03
ADG/RFI	Large White (Gilbert et al., 2007)	0.00±0.12	0.05
	1047 Duroc boars (Jiao et al, 2014)	-0.05±0.07	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.34±0.05 (purebred)/ 0.27±0.07 (crossbred)	
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	0.17±0.18	0.06±0.05
BFT/ADG	1622 Duroc pigs (Ito et al., 2018)	0.35±0.18	
	1047 Duroc boars (Jiao et al, 2014)	0.22±0.04	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.24±0.02 (purebred) -0.05±0.04 (crossbred)	
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	0.45±0.13	
BFT/FCR	1047 Duroc boars (Jiao et al, 2014)	-0.12±0.23	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.37±0.04 (purebred)/ 0.49±0.06 (crossbred)	
BFT/RFI	1047 Duroc boars (Jiao et al, 2014)	-0.11±0.19	
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	-0.14±0.16	-0.01±0.04
ADG/ADFI	1642 Duroc pigs (Hoque et al., 2009)	0.84±0.03	0.73
	1622 Duroc pigs (Ito et al., 2018)	0.79±0.08	
	1047 Duroc boars (Jiao et al, 2014)	0.32±0.05	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.71±0.03 (purebred)/ 0.66±0.04 (crossbred)	
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	0.88±0.05	0.73±0.02
BFT/ADFI	1622 Duroc pigs (Ito et al., 2018)	0.41±0.18	
	1047 Duroc boars (Jiao et al., 2014)	0.36±0.04	
	purebred Yorkshire pigs, a selection line for reduced RFI (low RFI line) (Cai et al., 2008)	0.57±0.10	0.49±0.04
FCR/ADG	1047 Duroc boars (Jiao et al., 2014)	-0.21±0.05	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.08±0.05 (purebred)/ -0.29±0.07 (crossbred)	
FCR/ADFI	1047 Duroc boars (Jiao et al, 2014)	0.13±0.11	
	Large White, 22984 purebreds/8657 crossbred (commercial animals) (Godinho et al., 2018)	0.71±0.03 (purebred)/ 0.49±0.06 (crossbred)	

ADFI = Average daily feed intake; ADG = Average daily gain; AMBW = Average Metabolic body weight; RFI = residual feed intake; BFT = Backfat thickness

1.3. Genetics models for the analysis of longitudinal data

Having longitudinal records for a trait means that repeated measures have been recorded on the individuals, at various times during their life. Meyer (2005) called "function-values traits" this type of data in quantitative genetics. The major advantage of such data is the possibility to describe the trait change across time. In addition to genetic aspects for selection of more accurate traits, the analysis of longitudinal data can help to provide real-time monitoring of the animal. Applying longitudinal models is expected to increase accuracy of selection (Begli et al., 2017). Nowadays, to handle longitudinal data, there are three options: parametric (best linear unbiased prediction, LASSO method, Bayesian, Bayesian LASSO, linear least-square regression models, ridge regression), semi-parametric (Niu et al., 2017) and non-parametric (Rice and Wu, 2001; Hulin and Jin-Ting, 2006) methods. In the framework of this literature review chapter, we will only focus on the parametric methods based on linear mixed models, such as random regression models, character process and antedependence models, which have been used in quantitative genetics, and we will compare them to multi-trait models.

All linear mixed models applied to repeated measurements can be decomposed into genetic and environmental components as follows:

$$y_{ij} = \mu_{ij} + u_i(t_j) + p_i(t_j) + \varepsilon_{ij} \qquad \text{(Eq.1)}$$

Where y_{ij} is the phenotype at time t_j for the animal i; μ_{ij} corresponds to fixed effects. $u_i(t_j)$ and $p_i(t_j)$ are the random genetic and permanent animal effects functions, with $\boldsymbol{u} \sim N(0, \boldsymbol{G} \otimes \boldsymbol{A})$ and $\boldsymbol{p} \sim N(0, \boldsymbol{P} \otimes \boldsymbol{I})$, where **A** is the known relationship matrix, **I** the identity matrix, and **G** and **P** the covariance matrices between measurements of traits (dimension $l \times l$, with l is the number of time points) for genetic and permanent environmental effects, respectively. Finally, ε_{ij} is the random residual term $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. The random effects are independent from one another.

1.3.1. Multiple trait models

A first model, making no assumption on the structure of **G** and **P,** is the multiple trait model (MT); this model considers **G** and **P** as unstructured matrices. In that case, the residual is formally excluded from E.q.1 to avoid identifiability problems, and is absorbed in ***p.***

Because all variance components have to be estimated, the MT approach is only used when there are few measurements per animal performed at the same time (or same age) for all

animals. In other cases, the dimensionality of the problem becomes too large. Indeed, the main issue of MT is the large number of parameters to estimate when the number of time points increases ($l(l+1)/2$ per random terms). Generally, the MT model shows convergence problems when more than six time points are considered (Rafat et al., 2011). In addition, measurements have to be performed at the same time (or age). To address this issue, the measurements can be classified into different categories before applying a MT mixed model, as proposed by Cai (2010). It has been shown that MT is much better than repeatability models (see below), that consider all time points as the repetition of the same trait, in terms of predictive ability, for instance for shell quality and monthly egg production in poultry (Kranis et al., 2007; Wolc et al., 2017). When the number of repeated measurements is too large, application of the MT model is unfeasible, and one has to reduce the number of parameters to estimate to obtain variance components of Eq.1.

There are then two options to reduce the number of parameters to estimate:

i) model the form of the covariance

- Repeatability model
- Character process (CP) model

ii) model the form of the random effect functions ($\boldsymbol{u}$(t) and $\boldsymbol{p}$(t))

- Random regression (RR) model
- Structure antedependence (SAD) model

1.3.2. Repeatability models

This model regards the repeated measurements as the same trait over time. In other words, the repeatability model assumes that the genetic and permanent environmental effects are constant over the whole test period. This results in equal variances and correlations of 1 between the random effects over time. The intra-class correlation between the records (intra-individual), the so-called repeatability, is defined as follows:

$$repeatability = \frac{\sigma_u^2 + \sigma_p^2}{\sigma_u^2 + \sigma_p^2 + \sigma_e^2}$$

in which σ_u^2, σ_p^2, σ_e^2 are the variances with respect to genetic, permanent environmental and error terms, respectively. The main advantage of this model is its simplicity in terms of computational requirements. This model has been used in the genetic evaluation of milk production up to 1999 (Interbull 2000), or very recently to evaluate genetic parameters for nutrient digestibility in pigs (Ouweltjes et al., 2018). However, the main assumption of the

repeatability model (same trait over time) is strong and questionable for many traits, so alternative models have been proposed.

1.3.3. Character process models

This category of models constructs directly the variances covariances matrix under the asumption that the genetic correlation gradually decreases with increasing distance between measurements (Pletcher and Geyer, 1999). In order to reduce the number of parameters to estimate, CP aims to model the covariance as functions of time. For these models, the genetic covariances can be decomposed into: $\mathrm{cov}\left(\boldsymbol{u}(t_1),\boldsymbol{u}(t_2)\right) = \mathrm{var}(\boldsymbol{u}(t_1)) \times \mathrm{var}\,\boldsymbol{u}(t_2) \times \rho_u\left(|t_2 - t_1|\right)$, and similarly for the permanent effects. The correlation between measurements $\rho_u\left(|t_2 - t_1|\right)$ under CP models may be modelled as a uniform, a first-order autoregressive, a Gaussian, a Cauchy, a standard normal or a bilateral exponential function (Pletcher and Geyer, 1999), and they can be modeled with homogenous or heterogenous variances (David et al., 2015).

For genetic evaluation, CP models were firstly developed by Pletcher and Geyer (1999). Jaffrézic (2000) made an interesting contribution with regard to testing different covariance functions and applied CP models to quantitative genetics. David et al. (2015) examined this type of model on the genetic components of feed intake of Large White pigs and rabbits. Compared to random regression model, CP is much better at estimating the genetic effects. Precisely, the CP models can adequately capture a correlation that declines rapidly to zero as values become further separated in time (Jaffrézic, 2000). However, CP model is worse at approximating the correlation structure (Jaffrézic and Pletcher, 2000; David et al., 2015). The use of this model is limited because it needs many parameters for taking into account the nonstationary variances of the longitudinal data over time. In addition, its extension to the multiple trait cases is not straightforward.

1.3.4. Random regression models

In the field of animal breeding, the term “random regression” was mentioned for the first time by Henderson in 1982, then Dekkers proposed to use random regressions for animal additive genetic model in 1992. Dekkers presented his work at the 5th world congress of genetics applied to livestock production (WCGALP) in Guelph (Schaeffer and Dekkers, 1994; Schaeffer and Jamrozik, 2008). Then, Diggle et al., (1994) used this model for analyzing longitudinal data and estimating genetic parameters.

The RR models consider the genetic and permanent environmental terms as functions of time t_j as: $u_i(t_j) = \sum_{r=0}^{m} a_{ir} f_r(t_j)$ for the genetic effects of an individual i (same model for the permanent effect). The distribution of the random coefficients a is given as follows: $\begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \otimes \mathbf{A} \right)$ where $\mathbf{K}$ is $(m+1)\times(m+1)$ covariance matrix of the random coefficient.

The covariance matrix between time points is then $\mathbf{G = FKF}'$, where $\mathbf{F}$ is the $l \times (m+1)$ (l= number of time points) matrix of the f function for all time points. The estimated breeding value for time t_j is given by:

$$EBV_i(t_j) = \sum_{r=0}^{m} \hat{a}_{ir} f_r(t_j)$$

Several functions f have been proposed in the literature. The simplest one is $f_r(t) = t^r$, which consists in a polynomial function of time. However, this model suffers from a high correlation between the random parameters. To reduce the correlation between random coefficients, orthogonal polynomials functions are used.

Random regression models using orthogonal polynomials (RR-OP)

The most often used orthogonal function is the orthogonal Legendre polynomials:

$$f_r(t_j) = \varphi_r(q_{t_j}) = \frac{1}{2^r} \sum_{k=0}^{r} \binom{r}{k} (q_{t_i} - 1)^{r-k} (q_{t_i} + 1)^k$$

where q is the time on the [-1,1] interval: $q_{t_j} = -1 + 2\left(\frac{t_j - t_{\min}}{t_{\max} - t_{\min}} \right)$ (Schaeffer, 2004).

Meyer and Hill (1997) demonstrated the equivalence between RR models and the genetic orthogonal covariance functions which were earlier proposed by Kirkpatrick et al. (1990).

Random regression using orthogonal polynomials models have been widely used, for instance for growth traits (Zumbach et al., 2010), BW, feed intake and residual feed intakes for pigs (Coyne et al, 2017), for milk production in cattle and goats (Silva et al., 2013), for egg production in poultry (Wolc and Szwaczkowski, 2009), for volume of ejaculate in Holstein bulls (Carabaño et al., 2007), for efficiency in Holstein cattle (Spurlock et al., 2012), and for carcass traits in beef cattle (Englishby et al., 2016). Kranis et al., (2007) compared genetic parameter estimates using RR and MT models for egg production in turkeys and concluded that both models were equally effective to describe the dynamics of genetic variance over time.

Jamrozik et al. (2002) illustrated that RR models could deal with test day records from different countries in which the animal management varies from one to another.

The main limit of RR with orthogonal polynomial functions is the possibility of border effects (Jaffrézic et al., 2004; David et al., 2015), that result in inflated variances at the extreme time points. This can be improved using spline (Speidel et al., 2010) functions (see below) or considering heterogeneous residual variances. In addition, if the degree of the polynomial function is high RR can require a large number of parameters to estimate. According to Misztal (2008), in several RR models, poor starting values may make convergence impractical or generate inaccurate estimates even from within the parameter space.

Random regression models using splines

Splines consist in curves constructed from pieces of low degree polynomials, which are joined smoothly at selected points (knots) (Meyer and Kirkpatrick, 2005). They can be modeled in a RR model. Random regression using splines (RR-SL) model is used in longitudinal genetic studies, thanks to its ease of use and speed of convergence (Speidel et al., 2010). Misztal (2006) argued that splines also have the advantages of quicker convergence than Legendre polynomials, maybe because spline coefficients are sparser than those of polynomials. Thus, they are quite flexible for modeling and simple for biological interpretation (Laureano et al., 2014). However, the selection of number and position of the knots is challenging (Speidel et al., 2010). To ease these choices, some authors proposed specific types of splines such as cubic smoothing splines (White et al., 1999) for modeling the lactation curves. Alternatively, B-spline is a particular type of splines that was used in RR as covariables in many studies (Rice and Wu, 2001).

1.3.5. Structured antedependence model

Antedependence models were firstly proposed by Gabriel (1962): it considers that the observation at time *t* is a regression on that/those at previous times. Then, Zimmerman and Núñez-Antón (1997) proposed to fit a parametric function on antedependence terms in order to decrease the number of parameters to estimate, leading to structured antedependence model (SAD). Jaffrézic and Pletcher (2000) and Jaffrézic et al (2004) applied the SAD model to genetic studies and found it promising for genetic evaluation.

The SAD model includes specific terms called antedependence and innovation variance parameters. Each random function is defined by three parameters: the order of the

antedependence (α), the degree of the polynomial function for each antedependence parameter (β_1 to β_α), and the degree of the polynomial for the innovation variance (γ). The function for the random effect $\boldsymbol{u}$ is $u_i(t_j) = \sum_{s=1}^{\alpha} \theta_{sj} u_i(t_{j-s}) + e_i(t_j)$, where $\boldsymbol{\theta}_{sj} = f_s(t_j)$ is the s^{th} antedependence parameter for time t_j, and $e_i(t_j)$ is the error term for animal *i* at time t_j. $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent, and $\boldsymbol{e}(t_j) \sim N(0, \boldsymbol{A}\sigma_e^2(t_j))$, where $\sigma^2_{e(t_j)}$ is called the innovation variance. To reduce the number of parameters, continuous functions of time are assumed for antedependence parameters $\theta_{sj} = \sum_{q=0}^{\beta_s} d_{sq} t_j^q$, and for the innovation variance $\sigma_e^2(t_j) = \exp\left(\sum_{q=0}^{\gamma} b_q t_j^q\right)$, in which *d* and *b* are the coefficients for antedependence parameters and innovation variance, respectively. In this thesis, after David et al. (2015), we noted a SAD model with a given set of parameters as follows: $\text{SAD}\alpha$ - $\beta_1,\ldots,\beta_\alpha$ γ. When performing a SAD model, a Cholesky decomposition of the inverse of the variance-covariance matrix

$$\mathbf{G^{-1}=L'D^{-1}L} \text{ with: } \mathbf{L} = \begin{bmatrix} 1 & & & 0 \\ -\theta_{sj} & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} \sigma_{e(1)} & & & 0 \\ & \sigma^2_{e(2)} & & \\ & & \sigma^2_{e(3)} & \\ 0 & & & \sigma^2_{e(4)} \end{bmatrix}$$

in which **L** is a lower triangular matrix with 1s on the diagonal and the antedependence parameters $-\theta_{kj}$ as below-diagonal entries, and **D** is a diagonal matrix with innovation variances $\sigma^2_{e(t)}$.

Parameters of the SAD model can be estimated using a OWN function in ASREML(Gilmour et al., 2009), which requires the specification of a $\mathbf{G^{-1}}$ and its derivatives with respect to each of the parameters (Jaffrézic et al., 2003; David et al., 2015). To facilitate convergence and avoid identifiability problems between the structured permanent and residual covariance matrices, the residual variance is included in the variance-covariance matrix of the permanent environmental effects, as proposed in previous works (David et al., 2015).

The SAD models provide a parsimonious and flexible way for modeling longitudinal data, as only one additional parameter is needed to increase one order of antedependence (Jaffrézic et

al., 2004). They also better fit the variance-covariance structure of the data than polynomial-based random regression models. They also have better prediction of missing values than RR (Jaffrézic et al., 2004; David et al, 2015). For the SAD model, Hurvich and Tsai (1989) and Núñez-Antón and Zimmerman (2000) proposed to use a Akaike information criterion (AIC) to select the best-fitted model. Jaffrézic et al. (2003) proposed to increase the antedependence order until the additional antedependence coefficient tends to zero.

From 2017, an OWN function program for ASReml supporting the performance of single or multiple traits SAD is freely available at https://zenodo.org/record/192036. This contributed as a new tool for extending the use of the SAD model for genetic evaluation and to other fields.

1.3.6. Best-fitted data model - goodness of fit

The purpose of comparison of models is to find the best-fitted model for a given dataset with the fewest parameters. For nested models, the best model within a category of model can be selected using a likelihood ratio test. To compare non-nested models, many criteria have been proposed, combining information such as the restricted maximum likelihood values (logL), the number of covariance parameters in the model (c), the number of fixed effects (p), and the number of observations (N) of every analysis. Among others, the Akaike information criterion (AIC= -2logL +2p; Akaike, 1973), and the Bayesian information criteria (BIC = -2logL + c×log(N-p)); Schwarz, 1978) are popular. The model with the lowest AIC or BIC is considered as best fitting the data, given the model parsimony. The major difference between BIC and AIC is that BIC searches a compromise between the number of parameters and the amount of information available. Thus, when considering model parsimony, BIC may be a better indicator than AIC (Flores and van der Werf, 2015).

As a genetic model should adequately model the genetic effects as well as the permanent environmental effects, the variance components, heritabilities and genetic correlations across times should be compared for the different tested models.

1.4. Predictive ability

Cross-validation can be used to assess the predictive ability of models. For this, the data are divided into a training set and a validation set. There are many criteria to compare the predictive ability of different models. Among them, the Vonesh Concordance coefficients (Vonesh et al., 1996) is given by the following equation:

$$VCC = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{\bar{y}} - \hat{y}_i)^2 + n \times (\bar{y} - \hat{\bar{y}})^2}$$

where n is the number of animals with predicted phenotypes; y_i is the observed values; $\bar{y}$ is the average of observed values; $\hat{y}_i$ is the predicted values of y_i; $\hat{\bar{y}}$ is the average of predicted values.

Other criteria are: mean square error of predictions (MSEP) (Andonov et al., 2013), correlation between predicted and observed phenotypes, correlation between predicted EBV and phenotypes or corrected phenotypes, and accuracy (correlation between EBV and corrected phenotypes divided by the square root of the heritability, often used to evaluate genomic prediction) (Begli et al., 2017).

1.5. Including genomic information in genetic evaluations

To predict animal breeding values, the Best Linear Unbiased Prediction (BLUP) approach, which uses a pedigree relationship matrix **A** to quantify the additive genetic effects, is commonly used since the 1980s, because it is simple and has low computational requirements. Over the last decades, high-throughput technologies allowed the identification of thousands of anonymous variants on the genome, such as single nucleotide polymorphisms (SNP). Recently, the GBLUP approach has been proposed by VanRaden et al. (2009), that includes such marker genotypes in a relationship matrix. Then, Legarra et al. (2009), Misztal et al. (2009) and Christensen and Lund (2010) simultaneously proposed single-step approaches in which the accuracy of genomic predictions can be improved by combining genomic information and information from traditional pedigree. This approach accounts for the fact that not all animals are genotyped, but their studies have been essentially limited to datasets with single records per animal.

Until now, there are few published studies (e.g., Wolc et al., 2013; Koivula et al., 2015; Kang et al., 2017) that included genomic information for genetic evaluation of longitudinal traits. For instance, Jiang et al., (2015) used a multiple trait Bayesian multivariate antepependence model for genomic prediction, Begli et al. (2016) showed that a GBLUP with β-spline random regression models for ADFI and RFI traits in a chicken F_2 population provided a higher prediction accuracy than that of BLUP. The use of genomic information is an increasing trend sustained by new computational techniques and power.

1.6. Objectives of the thesis and data

The objective of this thesis is to test genetic models dealing with longitudinal data for feed efficiency, and evaluate their prediction properties in different contexts. To provide a comprehensive view of the proposed methodologies, all the work was applied to a dataset from experimental pig lines that will be described in the next paragraphs, before introducing the chapters of the thesis.

1.6.1. Dataset

Pigs and selection experiment

Data used in this thesis come from a divergent pig selection experiment on residual feed intake. The initial population (generation G0) comprised 30 Large White litters from 30 sires and 30 dams (F0). It was the basis to select a line of pigs for low RFI (LRFI) and another line for high RFI (HRFI). The selection was described in details in Gilbert et al. (2007). Across generations, an average selection pressure of 7% was applied on males, whereas no selection pressure was applied to the dams. Data were recorded on 1286 males, 689 females and 599 castrated males over 8 generations (G0 to G7). The average inbreeding level in the last generation was 0.19 in the LRFI line and 0.18 in the HRFI line. The animals from both lines were born in two herds and all tested in the INRA experimental farm located in Rouillé (GenESI, Vienne, France).

Animal management and data collection

Four groups of 12 pigs were housed in pens equipped with single electronic feeders (ACEMA 64, Pontivy, France) in each batch of animals during growth (a batch being a group of animals born during a given week). They were acclimated to the feeder for one week before measurements could be used. Pigs were offered feed *ad libitum*, with a pelleted diet of cereals and soybean meal with 10 MJ NE/kg and 160 g CP/kg, and a minimum of 0.80 g digestible Lys/MJ NE. They had free access to water. The quantity of feed consumed was recorded after each visit to the electronic feeder. Pigs were on average 67 ± 1 days (27 ± 4 kg) at the beginning of the test, and were tested during the growing-finishing period up to 164 ± 11 days (112 ± 11 kg). As a result, up to 17 weekly records of ADFI and BW were available per animal. During the test period, body weights were recorded weekly for the males, whereas females and castrated males were weighted at 11, 15, 19 and 23 weeks of age, and more often if the test

lasted for more weeks. A total of 35 745 BW records were available from 2 583 animals over the 8 first generations (G0 to G7).

The ultrasonic backfat thickness (BFT) was measured on males at around 35 kg, 65 kg, 90 kg and 95 kg body weight, as the average of six measurements, from three points on both sides of the spine at the level of the neck, the back and the kidneys. For the females and castrated males, BFT was measured similarly but at week 11, 15, 19 and 23. Weekly averages of daily feed intake (WDFI) were computed for each animal. Outlier values for WDFI and WDFI for which more than two days of records were missing in a given week were removed from the analysis, as reported by David et al. (2015). Pigs with less than three records of WDFI and BWs were removed from the dataset. The final set consisted in 34 789 WDFI records and 22 850 BW records for 2 435 animals. A total of 3 986 animals were tracked back in the pedigree file. The steps for excluding the outlier values are presented in table 1.3.

Table 1.3. Data description

	males	females	castrated males	Total
Number of animals	1286	599	698	2583
WDFI (number of records)	17398	8786	9561	35745
BWs (number of records)	16301	3430	3766	23497
The animals with at least three records both BWs and FIs from above were kept				
	males	females	castrated males	Total
Number of animals	1186	580	669	2435
Age at beginning	67±1 days	67±1 days	67±1 days	67±1 days
BW at beginning	27±4 kg			
WDFI (number of records)	16866	8564	9329	34789
BW (number of records)	15808	3363	3679	22850
Pen (levels)	16	16	16	16
BW at the end	112±10 kg	111±7 kg	111±8 kg	111±9 kg
Age at the end	164±10 days	166±11 days	162±11 days	164±11 days
Max number of weeks	16			
Herd of birth (levels)	2			
Batch (levels)	33	29	34	66

1.6.2. Thesis organization

Based on the dataset previously described, we first investigated methodologies to predict missing BW records to improve the prediction of breeding values for feed efficiency on a weekly basis (**Chapter 2**).

In **Chapter 3**, we compared different approaches for estimating the covariance structure of longitudinal data of feed conversion ratio, and proposed criteria for animal selection using longitudinal data.

In **Chapter 4,** we used genomic and pedigree information to evaluate models for the prediction of longitudinal RFI.

In **Chapter 5**, we evaluated the impacts of divergent selection for RFI on the dynamics of FCR and RFI over time.

Finally, in **Chapter 6** a general discussion and conclusion will end this dissertation.

REFERENCES

Aggrey, S. E., and R. Rekaya. 2013. Dissection of Koch's residual feed intake: implications for selection. Poult. Sci. 92:2600–2605. doi:10.3382/ps.2013-03302.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Andonov, S., J. Ødegård, M. Svendsen, T. Ådnøy, M. Vegara, and G. Klemetsdal. 2013. Comparison of random regression and repeatability models to predict breeding values from test-day records of Norwegian goats. J. Dairy Sci. 96:1834–1843. doi:10.3168/jds.2012-5910.

Begli, H. E., R. V. Torshizi, A. A. Masoudi, A. Ehsani, and J. Jensen. 2017. Genomic dissection and prediction of feed intake and residual feed intake traits using a longitudinal model in F2 chickens. animal. 1–7. doi:10.1017/S1751731117003354.

Bordas, A., and P. Merat. 1981. Genetic variation and phenotypic correlations of food consumption of laying hens corrected for body weight and production. Br. Poult. Sci. 22:25–33. doi:10.1080/00071688108447860.

Brody, S. 1945. Bioenergetics and growth, Reinhold, New York.

Byerly, T. C. 1941. Feed and other costs of producing market eggs. Tech. Bull. n.1. The University of Maryland Agric. Exp., College Park, Maryland.

Cai, W. 2010. Quantitative genetic and statistical aspects of feed efficiency by analysis of the selection experiment for residual feed intake in Yorkshire pigs. Graduate Theses and Dissertations. 1333.

Cai, W., D. S. Casey, and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J. Anim. Sci. 86:287–298. doi:10.2527/jas.2007-0396.

Carabaño, M. J., C. Díaz, C. Ugarte, and M. Serrano. 2007. Exploring the Use of Random Regression Models with Legendre Polynomials to Analyze Measures of Volume of Ejaculate in Holstein Bulls. J. Dairy Sci. 90:1044–1057. doi:10.3168/jds.S0022-0302(07)71591-6.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2.

Coyne, J. M., D. P. Berry, K. Matilainen, M.-L. Sevon-Aimonen, E. A. Mantysaari, J. Juga, T. Serenius, and N. McHugh. 2017. Genetic co-variance functions for live weight, feed intake, and efficiency measures in growing pigs. J. Anim. Sci. 95:3822. doi:10.2527/jas2017.1408.

David, I., J. Ruesche, L. Drouilhet, H. Garreau, and H. Gilbert. 2015. Genetic modeling of feed intake. J. Anim. Sci. 93:965. doi:10.2527/jas.2014-8507.

Diggle, P. J., K.-Y. Liang, and S. L. Zeger. 1994. Analysis of longitudinal data. Clarendon Press, Oxford.

Drouilhet, L., B. Basso, M.-D. Bernadet, A. Cornuez, L. Bodin, I. David, H. Gilbert, and C. Marie-Etancelin. 2014. Improving residual feed intake of mule progeny of Muscovy ducks: Genetic parameters and responses to selection with emphasis on carcass composition and fatty liver quality1. J. Anim. Sci. 92:4287–4296. doi:10.2527/jas.2014-8064.

Englishby, T. M., G. Banos, K. L. Moore, M. P. Coffey, R. D. Evans, and D. P. Berry. 2016. Genetic analysis of carcass traits in beef cattle using random regression models. J. Anim. Sci. 94:1354–1364. doi:10.2527/jas.2015-0246.

Fitzhugh, H. A., and S. C. S. Taylor. 1971. Genetic Analysis of Degree of Maturity. J. Anim. Sci. 33:717–725. doi:10.2527/jas1971.334717x.

Flores, E. B., and J. van der Werf. 2015. Random regression test day models to estimate genetic parameters for milk yield and milk components in Philippine dairy buffaloes. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie. 132:289–300. doi:10.1111/jbg.12133.

Gabriel, K. R. 1962. Ante-dependence Analysis of an Ordered Set of Variables. Ann. Math. Stat. 33:201–212. doi:10.1214/aoms/1177704724.

Gilbert, H., J.-P. Bidanel, J. Gruand, J.-C. Caritez, Y. Billon, P. Guillouet, H. Lagant, J. Noblet, and P. Sellier. 2007. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J. Anim. Sci. 85:3182. doi:10.2527/jas.2006-590.

Gilbert, H., Y. Billon, L. Brossard, J. Faure, P. Gatellier, F. Gondret, E. Labussière, B. Lebret, L. Lefaucheur, N. Le Floch, I. Louveau, E. Merlot, M.-C. Meunier-Salaün, L. Montagne, P. Mormede, D. Renaudeau, J. Riquet, C. Rogel-Gaillard, J. van Milgen, A. Vincent, and J. Noblet. 2017. Review: divergent selection for residual feed intake in the growing pig. Anim. Int. J. Anim. Biosci. 1–13. doi:10.1017/S175173111600286X.

Gilmour, A. R., B. . Gogel, B. . Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. Available from: www.vsni.co.uk

Godfray, H. C. J., J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin. 2010. Food security: the challenge of feeding 9 billion people. Science. 327:812–818. doi:10.1126/science.1185383.

Godinho, R. M., R. Bergsma, F. F. Silva, C. A. Sevillano, E. F. Knol, M. S. Lopes, P. S. Lopes, J. W. M. Bastiaansen, and S. E. F. Guimarães. 2018. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. J. Anim. Sci. 96:817–829. doi:10.1093/jas/skx011.

de Haer, L. C. M., P. Luiting, and H. L. M. Aarts. 1993. Relations among individual (residual) feed intake, growth performance and feed intake pattern of growing pigs in group housing. Livest. Prod. Sci. 36:233–253. doi:10.1016/0301-6226(93)90056-N.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45–49. doi:10.1016/j.livsci.2008.05.016.

Hoque, M. A., K. Suzuki, H. Kadowaki, T. Shibata, and T. Oikawa. 2007. Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs. J. Anim. Breed. Genet. 124:108–116. doi:10.1111/j.1439-0388.2007.00650.x.

Hulin, W., and Z. Jin-Ting. 2006. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. John Wiley & Sons, Inc., Hoboken, NJ, USA. Available from: http://doi.wiley.com/10.1002/0470009675

Hurvich, C. ., and C. L. Tsai. 1989. HURVICH, C. M. & TSAI, C. L. (1989). Regression and time series model selection in small samples. Biometrika 76, 297-307.

Interbull 2000. National Genetic Evaluation Programmes for Dairy Production Traits Practised in Interbull Member Countries 1999–2000. Department of Animal Breeding and Genetics, Uppsala, Sweden, Bulletin 24.

Ito, T., K. Fukawa, M. Kamikawa, S. Nikaidou, M. Taniguchi, A. Arakawa, G. Tanaka, S. Mikawa, T. Furukawa, and K. Hirose. 2018. Effects of correcting missing daily feed intake values on the genetic parameters and estimated breeding values for feeding traits in pigs. Anim. Sci. J. Nihon Chikusan Gakkaiho. 89:12–20. doi:10.1111/asj.12891.

Jaffrézic, F., and S. D. Pletcher. 2000. Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. Genetics. 156:913–922.

Jaffrézic, F., R. Thompson, and W. G. Hill. 2003. Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits. Genet. Res. 82:55–65.

Jaffrézic, F., E. Venot, D. Laloë, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves. J. Anim. Sci. 82:3465–3473.

Jamrozik, J., L. R. Schaeffer, and K. A. Weigel. 2002. Genetic Evaluation of Bulls and Cows with Single- and Multiple-Country Test-Day Models. J. Dairy Sci. 85:1617–1622. doi:10.3168/jds.S0022-0302(02)74232-X.

Jiang, J., Q. Zhang, L. Ma, J. Li, Z. Wang, and J.-F. Liu. 2015. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity. 115:29–36. doi:10.1038/hdy.2015.9.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92:2377–2386. doi:10.2527/jas.2013-7338.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J.-F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity. 119:459–467. doi:10.1038/hdy.2016.91.

Kellner, O. 1909. The scientifi c feeding of animals. McMillan, New York.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979–993.

Kleiber, M. 1947. Body size and metabolic rate. Physiol. Rev. 27. stt.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of Feed Use in Beef Cattle. J. Anim. Sci. 22:486–494. doi:10.2527/jas1963.222486x.

Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784. doi:10.3168/jds.2014-8975.

Kranis, A., G. Su, D. Sorensen, and J. A. Woolliams. 2007. The Application of Random Regression Models in the Genetic Analysis of Monthly Egg Production in Turkeys and a Comparison with Alternative Longitudinal Models. Poult. Sci. 86:470–475. doi:10.1093/ps/86.3.470.

Larzul, C., and H. de Rochambeau. 2005. Selection for residual feed consumption in the rabbit. Livest. Prod. Sci. 95:67–72. doi:10.1016/j.livprodsci.2004.12.007.

Laureano, M. M. M., A. B. Bignardi, L. El Faro, V. L. Cardoso, H. Tonhati, and L. G. Albuquerque. 2014. Random regression models using different functions to model milk flow in dairy cows. Genet. Mol. Res. GMR. 13:7528–7541. doi:10.4238/2014.September.12.20.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. doi:10.3168/jds.2009-2061.

Meyer, K., and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. Livest. Prod. Sci. 47:185–200. doi:10.1016/S0301-6226(96)01414-5.

Meyer, K., and M. Kirkpatrick. 2005. Up hill, down dale: quantitative genetics of curvaceous traits. Philos. Trans. R. Soc. B Biol. Sci. 360:1443–1455. doi:10.1098/rstb.2005.1681.

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie. 123:74–80. doi:10.1111/j.1439-0388.2006.00582.x.

Misztal, I. 2008. Reliable computing in estimation of variance components. J. Anim. Breed. Genet. 125:363–370. doi:10.1111/j.1439-0388.2008.00774.x.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Nguyen, N. H., C. P. McPhee, and C. M. Wade. 2005. Responses in residual feed intake in lines of Large White pigs selected for growth rate on restricted feeding (measured on ad libitum individual feeding). J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie. 122:264–270.

Nikos Alexandratos, and J. Bruinsma. 2012. World agriculture towards 2030/2050: the 2012 revision, FAO, Food and Agriculture Organization of the United Nations. Available from: http://www.fao.org/economic/esa

Niu, F., J. Zhou, T. H. Le, and J. Z. Ma. 2017. Testing the trajectory difference in a semi-parametric longitudinal model. Stat. Methods Med. Res. 26:1519–1531. doi:10.1177/0962280215584109.

Núñez-Antón, V., and D. L. Zimmerman. 2000. Modeling nonstationary longitudinal data. Biometrics. 56:699–705.

Ouweltjes, W., L. M. G. Verschuren, J. Pijlman, R. Bergsma, D. Schokker, E. F. Knol, P. J. van der Aar, F. Molist, and M. P. L. Calus. 2018. The repeatability of individual nutrient digestibility in pigs. Livest. Sci. 207:63–67. doi:10.1016/j.livsci.2017.11.013.

Patience, J. F., M. C. Rossoni-Serão, and N. A. Gutiérrez. 2015. A review of feed efficiency in swine: biology and application. J. Anim. Sci. Biotechnol. 6. doi:10.1186/s40104-015-0031-2. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4527244/

Pletcher, S. D., and C. J. Geyer. 1999. The genetic analysis of age-dependent traits: modeling the character process. Genetics. 153:825–835.

Quan, J., G. Cai, J. Ye, M. Yang, R. Ding, X. Wang, E. Zheng, D. Fu, S. Li, S. Zhou, D. Liu, J. Yang, and Z. Wu. 2018. A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. Sci. Rep. 8:4536. doi:10.1038/s41598-018-22692-0.

Rafat, S. A., P. Namavar, D. J. Shodja, H. Janmohammadi, H. Z. Khosroshahi, and I. David. 2011. Estimates of the genetic parameters of turkey body weight using random regression analysis. Anim. Int. J. Anim. Biosci. 5:1699–1704. doi:10.1017/S1751731111000929.

Rice, J. A., and C. O. Wu. 2001. Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves. Biometrics. 57:253–259. doi:10.1111/j.0006-341X.2001.00253.x.

Saintilan, R., I. Mérour, L. Brossard, T. Tribout, J. Y. Dourmad, P. Sellier, J. Bidanel, J. van Milgen, and H. Gilbert. 2013. Genetics of residual feed intake in growing pigs: Relationships with production traits, and nitrogen and phosphorus excretion traits. J. Anim. Sci. 91:2542–2554. doi:10.2527/jas.2012-5687.

Schaeffer, L. r., and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Congr. Genet. Appl. Livest. Prod. 18:443-446.

Schaeffer, L. r., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145–146. doi:10.1111/j.1439-0388.2008.00748.x.

Schwarz, G. 1978. Estimating the Dimension of a Model. Ann. Stat. 6:461–464. doi:10.1214/aos/1176344136.

Shirali, M., C.-A. Duthie, A. Doeschl-Wilson, P. W. Knap, E. Kanis, J. A. van Arendonk, and R. Roehe. 2013. Novel insight into the genomic architecture of feed and nitrogen efficiency measured by residual energy intake and nitrogen excretion in growing pigs. BMC Genet. 14:121. doi:10.1186/1471-2156-14-121.

Shirali, M., V. H. Nielsen, S. H. Møller, and J. Jensen. 2015. Longitudinal analysis of residual feed intake and BW in mink using random regression with heterogeneous residual variance. Anim. Int. J. Anim. Biosci. 9:1597–1604. doi:10.1017/S1751731115000956.

Shirali, M., P. F. Varley, and J. Jensen. 2018. Bayesian estimation of direct and correlated responses to selection on linear or ratio expressions of feed efficiency in pigs. Genet. Sel. Evol. 50:33. doi:10.1186/s12711-018-0403-0.

Silverstein, J. T., M. Hostuttler, and K. P. Blemings. 2005. Strain differences in feed efficiency measured as residual feed intake in individually reared rainbow trout, Oncorhynchus mykiss (Walbaum). Aquac. Res. 36:704–711. doi:10.1111/j.1365-2109.2005.01278.x.

Speidel, S. E., R. M. Enns, and D. H. Crews. 2010. Genetic analysis of longitudinal data in beef cattle: a review. Genet. Mol. Res. GMR. 9:19–33. doi:10.4238/vol9-1gmr675.

Spurlock, D. M., J. C. M. Dekkers, R. Fernando, D. A. Koltes, and A. Wolc. 2012. Genetic parameters for energy balance, feed efficiency, and related traits in Holstein cattle. J. Dairy Sci. 95:5393–5402. doi:10.3168/jds.2012-5407.

VanRaden, P. M., C. P. V. Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

Vonesh, E. F., V. M. Chinchilli, and K. Pu. 1996. Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. Biometrics. 52:572–587. doi:10.2307/2532896.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and Environmental Smoothing of Lactation Curves with Cubic Splines. J. Dairy Sci. 82:632–638. doi:10.3168/jds.S0022-0302(99)75277-X.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. 2013. Analysis of egg production in layer chickens using a random regression model with genomic relationships. Poult. Sci. 92:1486–1491. doi:10.3382/ps.2012-02882.

Wolc, A., J. Arango, P. Settar, N. P. O'Sullivan, and J. C. M. Dekkers. 2017. Repeatability vs. multiple-trait models to evaluate shell dynamic stiffness for layer chickens. J. Anim. Sci. 95:9–15. doi:10.2527/jas.2016.0618.

Wolc, A., and T. Szwaczkowski. 2009. Estimation of genetic parameters for monthly egg production in laying hens based on random regression models. J. Appl. Genet. 50:41–46. doi:10.1007/BF03195650.

Young, J. M., and J. C. M. Dekkers. 2012. The genetic and biological basis of residual feed intake as a measure of feed efficiency. In: Feed efficiency in swine. Wageningen Academic Publishers, Wageningen. p. 153–166. Available from: https://link.springer.com/chapter/10.3920/978-90-8686-756-1_7

Zimmerman, D. L., and V. Núñez-Antón. 1997. Structured Antedependence Models for Longitudinal Data. In: T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. D. Wolfinger, editors. Modelling Longitudinal and Spatially Correlated Data. Vol. 122. Springer New York, New York, NY. p. 63–76. Available from: http://link.springer.com/10.1007/978-1-4612-0699-6_6

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie. 127:93–99. doi:10.1111/j.1439-0388.2009.00827.x.

Chapter 2 - DEALING WITH MISSING BODY WEIGHTS IN GROWING PIGS

2.1. Predicting missing body weights to improve the prediction of estimated breeding values

Today, with advances in self-feeder and electronic identification, the measurement of ADFI is relatively easily performed in routine. Unfortunately, the situation is slightly different for BW. Depending on the animal management strategy (measurement schedule), animals may not be measured at all time points, resulting in incomplete data. In addition, there may be errors in the recording, due to machine troubles and identification defaults, rodent activity, moisture, dust in the environment, and behaviors of the animals (Zumbach et al., 2010; Jiao et al., 2014), resulting in discarding some (automatic) records from the analysis. Altogether, that leads to non-monotonous missing data patterns. The issues related to missing data are still challenging because they can affect the genetic estimations.

In our dataset, the missing proportion of weekly BWs of the females and castrated males was high (60%) (Table 1.3). As BW are essential inputs to calculate feed efficiency in growing pigs, we wanted to quantify how missing BW impact the estimation of genetic parameters for FCR. To address this question, we performed a simulation of missing BW patterns and proposed an approach for handling missing BW in our dataset. This work was published in Journal of Animal Science (2017).

This chapter is organized as follows: firstly, the paper I presents results of a simulation study using a random regression model for genetic evaluation of longitudinal FCR. Secondly, to complement these results, we provide results from the same simulation study but using a simple repeatability model.

2.2. Article I: How to improve breeding values prediction for feed conversion ratio in case of incomplete longitudinal body weights

(Article I: Huynh-Tran, V. H., H. Gilbert, and I. David. 2017. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J. Anim. Sci. 95:39–48. doi:10.2527/jas.2016.0980)

How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights[1]

V. H. Huynh Tran,[2] H. Gilbert, and I. David

GenPhySE, INRA, Université de Toulouse, INPT, ENVT, Castanet-Tolosan, France

ABSTRACT: With the development of automatic self-feeders, repeated measurements of feed intake are becoming easier in an increasing number of species. However, the corresponding BW are not always recorded, and these missing values complicate the longitudinal analysis of the feed conversion ratio (FCR). Our aim was to evaluate the impact of missing BW data on estimations of the genetic parameters of FCR and ways to improve the estimations. On the basis of the missing BW profile in French Large White pigs (male pigs weighed weekly, females and castrated males weighed monthly), we compared 2 different ways of predicting missing BW, 1 using a Gompertz model and 1 using a linear interpolation. For the first part of the study, we used 17,398 weekly records of BW and feed intake recorded over 16 consecutive weeks in 1,222 growing male pigs. We performed a simulation study on this data set to mimic missing BW values according to the pattern of weekly proportions of incomplete BW data in females and castrated males. The FCR was then computed for each week using observed data (obser_FCR), data with missing BW (miss_FCR), data with BW predicted using a Gompertz model (Gomp_FCR), and data with BW predicted by linear interpolation (interp_FCR). Heritability (h^2) was estimated, and the EBV was predicted for each repeated FCR using a random regression model. In the second part of the study, the full data set (males with their complete BW records, castrated males and females with missing BW) was analyzed using the same methods (miss_FCR, Gomp_FCR, and interp_FCR). Results of the simulation study showed that h^2 were overestimated in the case of missing BW and that predicting BW using a linear interpolation provided a more accurate estimation of h^2 and of EBV than a Gompertz model. Over 100 simulations, the correlation between obser_EBV and interp_EBV, Gomp_EBV, and miss_EBV was 0.93 ± 0.02, 0.91 ± 0.01, and 0.79 ± 0.04, respectively. The heritabilities obtained with the full data set were quite similar for miss_FCR, Gomp_FCR, and interp_FCR. In conclusion, when the proportion of missing BW is high, genetic parameters of FCR are not well estimated. In French Large White pigs, in the growing period extending from d 65 to 168, prediction of missing BW using a Gompertz growth model slightly improved the estimations, but the linear interpolation improved the estimation to a greater extent. This result is due to the linear rather than sigmoidal increase in BW over the study period.

Key words: feed efficiency, longitudinal data, pig, quantitative genetic

J. Anim. Sci. 2017.95:39–48
doi:10.2527/jas2016.0980

INTRODUCTION

Although genetics and management of pigs have been improved in recent decades, feed still accounts for around two-thirds of the production costs in western countries (Agriculture and Horticulture Development Board, 2016; Patience et al., 2015). In addition, feed efficiency is a trait of importance in several species (Nardone et al., 2010). In practice, feed efficiency is generally expressed as its inverse trait, the feed conversion ratio (**FCR**), which corresponds to the ratio of feed intake (**FI**) to BW gain (Losinger, 1998). Today, with

[1]Our work is part of project H2020 Feed-a-Gene. The Feed-a-Gene Project is funded by the European Union's H2020 Program under grant agreement 633531. The authors would like to thank the animal genetics division at INRA, France, and the European project H2020 Feed-a-Gene for financial support for funding Huynh Tran's PhD thesis.

[2]Corresponding author: van-hung.tran@inra.fr

Received September 5, 2016.
Accepted November 1, 2016.

the development of automatic self-feeders and electronic identification, repeated measurements of FI and BW are available in many species, making it possible to analyze longitudinal FCR. Analysis of the individual profile of FCR over time can improve the genetic evaluation of this trait (Shirali et al., 2012). However, when FI is not recorded at the same time as BW, BW may be missing for substantial parts of the period to be analyzed. For instance, in an experimental French Large White pig population, male pigs are weighed every week during the growing period, whereas females and castrated males are weighed every month, meaning 60% of weekly BW are missing for the females and castrated males. Missing BW records can complicate the analysis of longitudinal FCR. Although mixed-effect regression models are supposed to be quite robust to missing data (Gibbons et al., 2010), estimation of variance components has proved to be erratic when some records are missing (Nobre et al., 2003). Little is known about the impact of missing BW on the estimation of the genetic parameters for FCR. Therefore, on the basis of the missing BW pattern observed in the French Large White pig experiment, the objectives of this study were to evaluate whether missing BW records have an impact on the estimation of genetic parameters for FCR and if the use of a Gompertz model or linear interpolation to predict the missing BW can improve estimation of the genetic parameters for FCR.

MATERIAL AND METHODS

Data

For the current study, data were collected in accordance with the national regulations of animal care in agriculture in France. Body weight records and FI of 2,503 growing French Large White pigs (1,222 males, 594 females, and 687 castrated males) were used in this study. This population is described in detail in Gilbert et al. (2007). Animal management was the following: animals born in a given farrowing batch were gathered at weaning (28 d of age) in the same postweaning unit. At 10 wk of age, 48 pigs were moved to a growing-finishing room with 4 pens per batch equipped with single-place electronic feeders (ACEMA 64, Pontivy, France; Labroue et al., 1997)). Twelve animals of the same sex were allotted to each pen. Animals were provided with an ad libitum pelleted diet based on cereals and soybean meal containing 10 MJ NE/kg and 160 g CP/kg, with a minimum of 0.80 g digestible Lys/MJ NE. The BW and age at the beginning of the test averaged 24.9 ± 3.8 kg and 67 ± 1 d, respectively. The average BW and age at the end of the test were 115.3 ± 10.9 kg and 168 ± 13 d. The pigs were allowed to acclimate to the feeders for about a week, so the records of the first week of the test period were removed from the data set.

During the 16 consecutive weeks (from wk 2 to 17) of the test period, males were weighed weekly, and the majority of females and castrated males were weighed monthly. This resulted in a weekly proportion of missing BW of up to 60% in the females and castrated males in comparison with males, whose missing BW records were low (6%). The details of the available weekly BW of the females and castrated males are presented in Fig. 1. The individual FI of each animal was recorded automatically each time it used the feeder. Weekly averages of daily feed intake (**WDFI**) were then computed for each animal. The outlier values of WDFI and WDFI for which more than 2 d of records were missing in a given week were removed from the analysis, as reported by David et al. (2015). The final data set comprised 16,301 weekly BW records and 17,398 WDFI for the male pigs, 3,430 weekly BW records and 8,786 WDFI for the females, and 3,766 weekly BW records and 9,561 WDFI for the castrated males.

Analyses

Our aim was to measure the impact of missing BW records on estimations of the genetic parameters of FCR and to explore how to improve these estimations. To this end, we compared the estimations of genetic parameters and breeding values of FCR under 4 scenarios: FCR computed using complete observed BW and WDFI data (**obser_FCR**), FCR computed with missing BW data (**miss_FCR**), and FCR computed using missing BW replaced by predicted values obtained using a "by nearest" linear interpolation (**interp_FCR**) or using a Gompertz model (**Gomp_FCR**). For this purpose, we used the male data as the reference (very low proportion of missing BW records) and simulated randomly a pattern of missing values of BW on this data set by mimicking the same pattern of proportions of BW missing values per week as those observed in castrated males and females. For example, the full male data set contained 100% BW available for wk 5. After simulating the pattern of missing values, only 6.8% remained available for further analyses.

The FCR was then calculated for each animal i and week j as follows:

$$\mathrm{FCR}_{ij} = \frac{\mathrm{WDFI}_{ij}}{\mathrm{ADG}_{ij}},$$

where WDFI_{ij} is the WDFI of animal i in week j and ADG_{ij} is the ADG of animal i at week j ($j \in [4,13]$) calculated over a 4-wk period as follows:

$$\mathrm{ADG}_{ij} = \frac{\mathrm{BW}_{ij+2} - \mathrm{BW}_{ij-2}}{\mathrm{age}_{ij+2} - \mathrm{age}_{ij-2}}.$$

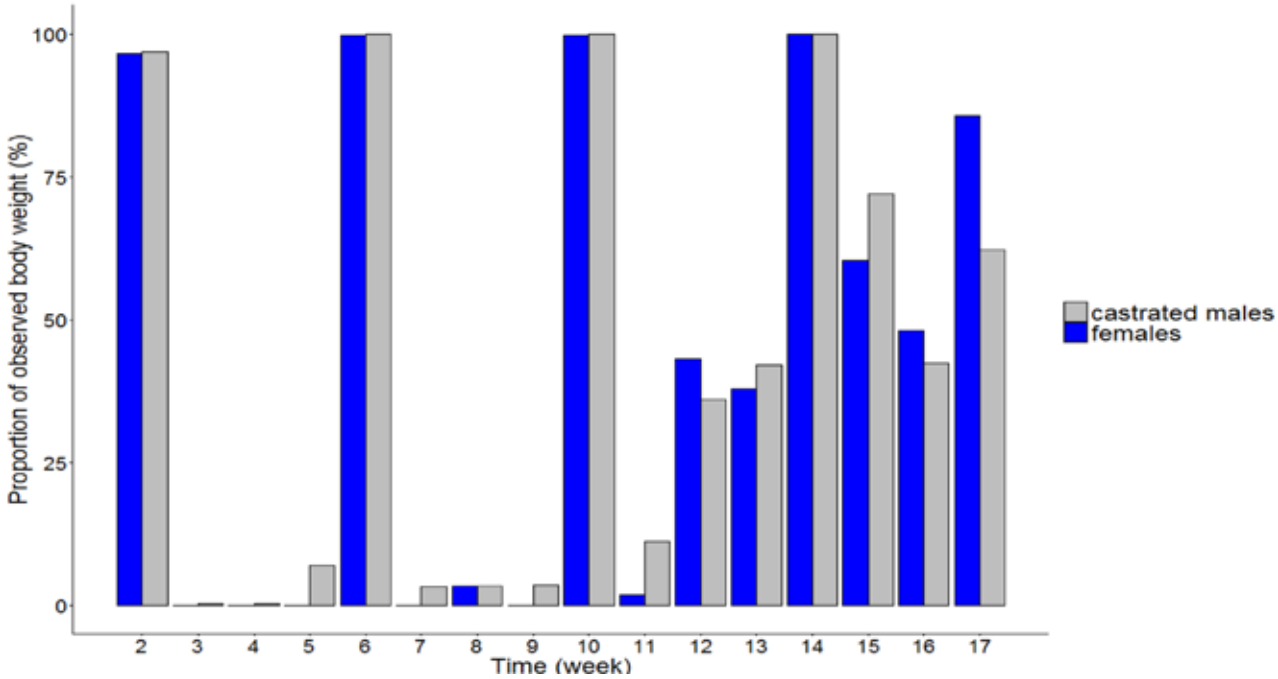


Figure 1. Proportion of BW available for females and castrated males per week from wk 2 (11 wk of age) to wk 17.

The BW used in the last formula differed depending on the scenario. In the first scenario, (obser_FCR), BW_{ij+2} and BW_{ij-2} corresponded to the measured BW. In the second scenario, (miss_FCR), if BW_{ij-2} or BW_{ij+2} was considered to be missing during the simulation of the missing data, then ADG_{ij} and hence $miss\_FCR_{ij}$ were not calculated but were considered missing. In the third scenario (interp_FCR), if BW_{ij-2} was missing, then BW_{ij-3} was used to compute ADG_{ij}. If BW_{ij-3} was also missing, then BW_{ij-1} was used to compute ADG_{ij}. If none were available, ADG_{ij} was considered missing. Similarly, missing BW_{ij+2} were replaced primarily by BW_{ij+3} and then by BW_{ij+1} if BW_{ij+3} was missing. If both were missing, then ADG_{ij} was considered missing. This method is equivalent to a linear interpolation except if BW_{ij-1} or $BW_{i,j+1}$ was used. In the following, the method is referred to as the by nearest linear interpolation. In the fourth scenario (Gomp_FCR), missing BW was predicted by a Gompertz model (Porter et al., 2010; Cai et al., 2011). For this purpose, the Gompertz model was fitted to the data with missing BW records using the NLIN procedure of SAS (SAS 9.4; SAS Inst. Inc., Cary, NC). The Gompertz model is given by

$$BW_{ij} = A_i * e^{-C_i * e^{-B_i j}}$$

where A_i is the asymptotic or maximum growth response (mature weight), B_i is the growth rate constant, and C_i is log(mature weight/birth weight). This model was fitted to each individual animal separately to estimate individual parameters A_i, B_i, and C_i. Subsequently, $\hat{A}_i, \hat{B}_i$, and $\hat{C}_i$ were used in the Gompertz formula to predict the missing BW. To reduce the effect of outliers and leverage points, BW was predicted from Gompertz model only when $\hat{A}_i, \hat{B}_i$, and $\hat{C}_i$ were between the 1st and 99th percentiles of distribution of each parameter.

In the genetic analysis, FCR_{ij} less than 0 and greater than 6 were considered outliers and were discarded from the analysis. The goal was to compare miss_EBV, interp_EBV, and Gomp_EBV obtained from miss_FCR, interp_FCR, and Gomp_FCR and see how they were correlated with obser_EBV of obser_FCR. To estimate the genetic parameters, a random regression model using Legendre polynomials (**RR-PL**) was fitted to a repeated FCR.

The RR-PL is given by

$$FCR_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \sum_{k=1}^{m} a_{ik}\varphi_{kj} + \sum_{k=1}^{n} p_{ik}\varphi_{kj} + \varepsilon_{ij}$$

where FCR_{ij} is obser_FCR, miss_FCR, interp_FCR, or Gomp_FCR for individual i at week j; $\boldsymbol{\beta}$ is the vector of fixed effect; a_{ik} and p_{ik} are the kth random regression coefficients for genetic and permanent environmental effects for animal i, respectively, with $\mathbf{a} \sim N(0, \mathbf{G} \otimes \mathbf{A})$ and $\mathbf{p} \sim N(0, \mathbf{P} \otimes \mathbf{I})$, where $\mathbf{A}$ is the known relationship matrix, $\mathbf{I}$ is an identity matrix whose order is equal to the total number of individuals, $\mathbf{G}$ is the (co)variance matrix of the additive random regression coefficients, and $\mathbf{P}$ is the (co) variance matrix of the random permanent environmental regression coefficients; φ_{kj} is the $(k-1)$th Legendre polynomial in week j; and m and n are the orders of regression for the genetic and permanent environmental effects, respectively. The permanent effect reflects the nongenetic individual effects that are correlated across repetitions.

Table 1. Percentage of missing BW and missing feed conversion ratios (FCR) in the simulation study (median across 100 simulation replicates) depending on the scenario

	Scenario[1]			
Item	Observed	Missing	Interpolation	Gompertz
Missing BW,[2] %	5.8	61.5 [60.8–62.1]		7.3 [6.6–7.7]
Missing FCR,[3] %	12.5	77.2 [76.1–78.2]	24.8 [24.1–25.9]	8.8 [7.5– 9.6]

[1]Observed = available data; missing = data with a simulated pattern of missing BW; interpolation = ADG calculated using the by nearest interpolation; Gompertz = missing BW predicted using a Gompertz model. The numbers in brackets show the minium to the maximum.

[2]Percentage of missing BW over a period of 14 wk (from wk 2 to 15).

[3]Percentage of missing FCR over a period of 10 wk (from wk 4 to 13).

The covariance components and genetic parameters were estimated using the REML approach with ASReml software (Gilmour et al., 2009). All the fixed effects and 1-way interaction of biological relevance included in the model were selected beforehand in a stepwise manner using nested models that were compared with a likelihood ratio test. The fixed effects retained in the models were the week of observation (10 levels), the pen (96 levels), the batch (32 levels), the age, and BW of the animal at the beginning of the test. Likelihood ratio tests were used to choose the best polynomial orders for genetic and permanent environmental effects on the animal. Legendre polynomials of orders 3 and 2 were retained to model the genetic and permanent environmental effects, respectively. Heritability was computed for each week j as the ratio of the genetic to the total variance:

$$h_j^2 = \frac{\sum_{k=1}^{m} \varphi_{kj}^2 \boldsymbol{G}_{kk}}{\sum_{k=1}^{m} \varphi_{kj}^2 \boldsymbol{G}_{kk} + \sum_{k=1}^{n} \varphi_{kj}^2 \boldsymbol{P}_{kk} + \sigma_\varepsilon^2},$$

where σ_ε^2 is the residual variance. Standard errors of estimates of genetic parameters were computed in ASReml using the method proposed by Fischer et al. (2004). Pearson correlations were used to compare the breeding values (EBV) for the different FCR: obser_EBV, miss_EBV, Gomp_EBV, and interp_EBV. The simulation of missing BW and the genetic analysis of each FCR were repeated 100 times. The mean and SD of the correlation coefficients and heritability of the 100 simulations were computed.

In the second step of the analysis, we estimated the genetic parameters and EBV for the full data set (males, females, and castrated males, 32,552 BW records) that "naturally" contained missing BW. For this data set, 3 FCR were computed for each animal and week, miss_FCR, Gomp_FCR, and interp_FCR, using the methods described above. As the full data set naturally contained missing values, the previous miss_FCR also corresponded to observed data. As described for the simulations, the EBV and heritability were estimated using the RR-PL model. Gender was added to the models as a fixed effect. The correlation between Gomp_EBV, miss_EBV, or interp_EBV and the heritability of Gomp_FCR, miss_FCR, and interp_FCR were estimated as described above for the simulation study.

RESULTS

Simulation

A detailed description of missing BW data and resulting FCR in the 4 scenarios are given in Table 1. In the initial data set, 11,790 observations of WDFI over the 10-wk period (wk 4 to 13) were available to calculate FCR. The proportion of missing BW over the 14-wk period (wk 2 to 15) used to calculate FCR varied depending on the scenario; the proportion was low in the observed scenario (5.8%), slightly higher in the Gompertz scenario (7.3%), and very high in the missing scenario (61.5%). The percentage of missing BW in the scenario corrected using the by nearest linear interpolation was the same as in the missing scenario since BW is not replaced with this approach. A huge proportion (77.2%) of FCR was missing in the missing scenario. Correction using the by nearest linear interpolation scenario greatly reduced the proportion of missing FCR (median of 24.8%), but it remained higher than for the observed data (12.5%). Finally, the Gompertz scenario had the lowest percentage of missing FCR (median of 8.8%). It should be noted that missing Gomp_FCR values were due to extreme values of $\hat{A}, \hat{B}$, and $\hat{C}$ or to outlier values of Gomp_FCR.

The genetic variances and genetic correlations across the 10 consecutive weeks obtained with the RR-PL model in the different scenarios are listed in Table 2. The genetic variances obtained with obser_FCR, interp_FCR, and Gomp_FCR were quite similar, ranging from 0.01 to 0.09 for obser_FCR, from 0.02 to 0.09 for Gomp_FCR, and from 0.01 to 0.08 for interp_FCR, depending on the week. Higher genetic variances were obtained in the first week (0.04 for obser_FCR and Gomp_FCR, 0.05 for interp_FCR) and

Table 2. Mean and SD over 100 simulations of the additive genetic variance (on the diagonal) and genetic correlations (above the diagonal) of FCR over the 10-wk period in the different scenarios

Week	4	5	6	7	8	9	10	11	12	13
obser_FCR[1]										
4	0.04	0.89	0.70	0.48	0.14	−0.24	−0.48	−0.58	−0.56	−0.40
5		0.03	0.94	0.79	0.44	−0.06	−0.39	−0.51	−0.43	−0.19
6			0.03	0.94	0.67	0.18	−0.18	−0.33	−0.27	−0.06
7				0.03	0.88	0.13	0.13	−0.05	−0.07	0.04
8					0.02	0.84	0.57	0.37	0.23	0.13
9						0.01	0.92	0.76	0.54	0.25
10							0.01	0.95	0.75	0.41
11								0.03	0.91	0.64
12									0.04	0.89
13										0.09
miss_FCR[1]										
4	0.03 ± 0.01	0.26 ± 0.16	0.04 ± 0.15	−0.12 ± 0.12	−0.35 ± 0.08	−0.23 ± 0.14	−0.18 ± 0.14	−0.21 ± 0.14	−0.37 ± 0.11	−0.44 ± 0.17
5		0.10 ± 0.03	0.97 ± 0.02	0.85 ± 0.06	0.03 ± 0.15	−0.75 ± 0.09	−0.87 ± 0.05	−0.90 ± 0.04	−0.77 ± 0.08	0.31 ± 0.24
6			0.13 ± 0.04	0.95 ± 0.02	0.21 ± 0.14	−0.65 ± 0.12	−0.81 ± 0.08	−0.84 ± 0.06	−0.69 ± 0.09	0.39 ± 0.21
7				0.06 ± 0.02	0.49 ± 0.11	−0.40 ± 0.17	−0.60 ± 0.14	−0.66 ± 0.11	−0.54 ± 0.11	0.37 ± 0.20
8					0.02 ± 0.00[2]	0.59 ± 0.13	0.37 ± 0.15	0.29 ± 0.15	0.23 ± 0.12	−0.01 ± 0.16
9						0.09 ± 0.02	0.97 ± 0.01	0.93 ± 0.02	0.76 ± 0.06	−0.32 ± 0.20
10							0.20 ± 0.06	0.99 ± 0.00[2]	0.83 ± 0.05	−0.32 ± 0.21
11								0.23 ± 0.06	0.90 ± 0.03	−0.21 ± 0.22
12									0.13 ± 0.02	0.21 ± 0.19
13										0.20 ± 0.07
Gomp_FCR[1]										
4	0.04 ± 0.01	0.62 ± 0.06	0.26 ± 0.01	0.10 ± 0.10	0.00 ± 0.10	−0.06 ± 0.07	−0.10 ± 0.07	−0.18 ± 0.07	−0.35 ± 0.08	−0.57 ± 0.09
5		0.03 ± 0.00[2]	0.91 ± 0.02	0.78 ± 0.04	0.44 ± 0.08	−0.19 ± 0.07	−0.50 ± 0.05	−0.61 ± 0.04	−0.63 ± 0.05	−0.45 ± 0.08
6			0.03 ± 0.00[2]	0.96 ± 0.01	0.64 ± 0.06	−0.09 ± 0.07	−0.47 ± 0.05	−0.58 ± 0.04	−0.54 ± 0.05	−0.24 ± 0.08
7				0.02 ± 0.00[2]	0.83 ± 0.03	0.16 ± 0.06	−0.24 ± 0.06	−0.37 ± 0.06	−0.34 ± 0.07	−0.08 ± 0.09
8					0.01 ± 0.00[2]	0.69 ± 0.04	0.34 ± 0.08	0.19 ± 0.09	0.15 ± 0.10	−0.18 ± 0.11
9						0.02 ± 0.00[2]	0.91 ± 0.02	0.82 ± 0.04	0.71 ± 0.06	0.43 ± 0.10
10							0.03 ± 0.00[2]	0.98 ± 0.01	0.87 ± 0.03	0.52 ± 0.18
11								0.05 ± 0.01	0.95 ± 0.01	0.64 ± 0.06
12									0.07 ± 0.01	0.84 ± 0.03
13										0.09 ± 0.01
interp_FCR[1]										
4	0.05 ± 0.01	0.93 ± 0.02	0.80 ± 0.05	0.63 ± 0.07	0.36 ± 0.09	0.05 ± 0.10	−0.19 ± 0.10	−0.39 ± 0.09	−0.57 ± 0.09	−0.68 ± 0.11
5		0.03 ± 0.00[2]	0.95 ± 0.01	0.81 ± 0.04	0.49 ± 0.09	0.08 ± 0.10	−0.23 ± 0.09	−0.44 ± 0.07	−0.60 ± 0.06	−0.65 ± 0.12
6			0.02 ± 0.00[2]	0.93 ± 0.02	0.66 ± 0.07	0.24 ± 0.11	−0.09 ± 0.11	−0.33 ± 0.09	−0.49 ± 0.08	−0.54 ± 0.13
7				0.02 ± 0.00[2]	0.87 ± 0.03	0.55 ± 0.09	−0.23 ± 0.12	−0.02 ± 0.12	−0.23 ± 0.10	−0.38 ± 0.13
8					0.01 ± 0.00[2]	0.88 ± 0.03	0.66 ± 0.08	0.44 ± 0.10	0.18 ± 0.10	−0.10 ± 0.11
9						0.02 ± 0.00[2]	0.94 ± 0.02	0.79 ± 0.05	0.56 ± 0.09	0.21 ± 0.12
10							0.03 ± 0.01	0.95 ± 0.01	0.78 ± 0.06	0.45 ± 0.12
11								0.04 ± 0.01	0.93 ± 0.02	0.68 ± 0.08
12									0.06 ± 0.01	0.89 ± 0.03
13										0.08 ± 0.02

[1]Here obser_FCR = feed conversion ratio (FCR) computed using available data; miss_FCR = FCR computed using data with a simulated pattern of missing BW; interp_FCR = FCR computed using ADG calculated with the by nearest interpolation; Gomp_FCR = FCR computed using missing BW predicted by a Gompertz model.

[2]SD < 0.00.

in the last 2 wk (0.04 to 0.09 for obser_FCR, 0.07 to 0.09 for Gomp_FCR, and 0.06 to 0.08 for interp_FCR) of the test period in comparison with the middle of the growing period (wk 5 to 10, ranging from 0.01 to 0.03). Except for wk 4 and 8, the genetic variances obtained for miss_FCR were generally higher (ranging from 0.06 to 0.23) than those obtained in the other scenarios. The same comparison between scenarios and the same pattern of changes in variance over time were observed for the permanent environmental effect (result not shown).

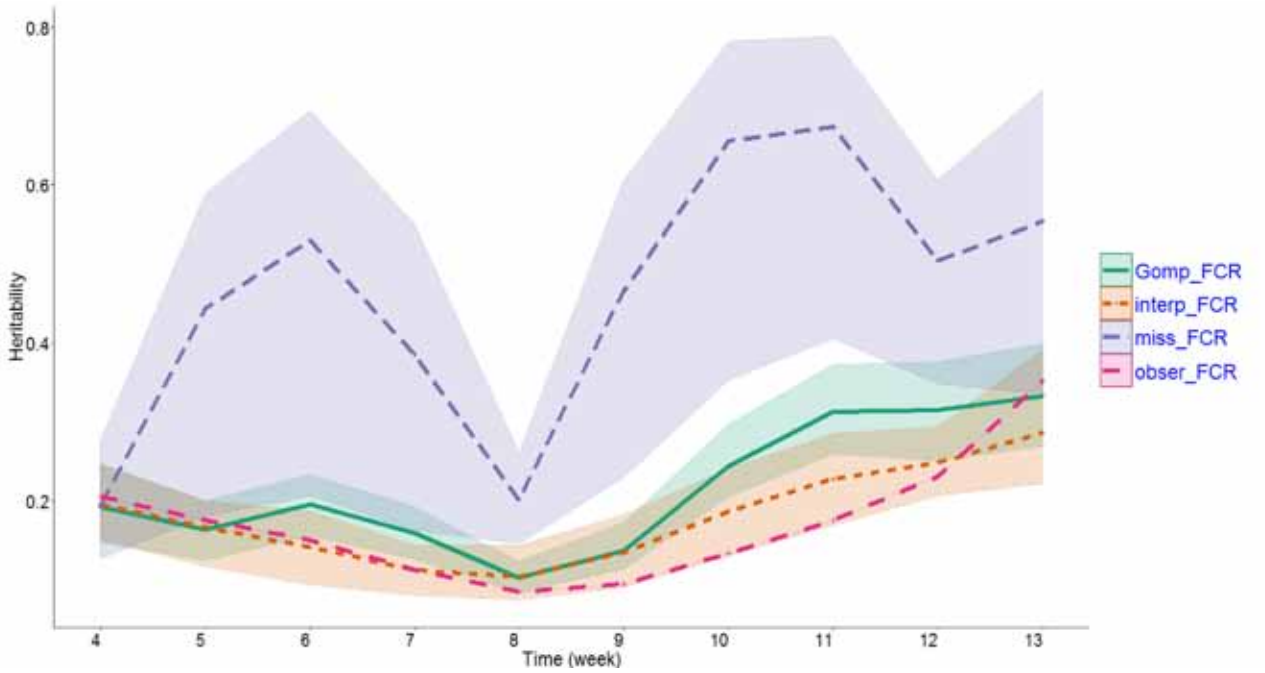


Figure 2. Changes in heritability of feed conversion ratio (FCR) over time in the different missing data scenarios. The shaded area delimits the 2.5 and 97.5 percentiles (100 iterations). Here obser_FCR = FCR computed using available data; miss_FCR = FCR computed using data with a simulated pattern of missing BW; interp_FCR = FCR computed using ADG calculated with the by nearest interpolation; Gomp_FCR = FCR computed using missing BW predicted by a Gompertz model.

The genetic correlations varied depending on the length of the interval between measurements. They ranged from −0.58 to 0.95 for obser_FCR, from −0.90 to 0.99 for miss_FCR, from −0.63 to 0.98 for Gomp_FCR, and from −0.68 to 0.95 for interp_FCR (Table 2). The 1-wk interval correlations were generally high. They were higher than 0.84 for interp_FCR and obser_FCR and ranged from 0.62 to 0.98 for Gomp_FCR. Nonetheless, low genetic correlations between FCR measured in 2 successive weeks were sometimes obtained for miss_FCR (0.26 between wk 4 and 5, 0.21 between wk 12 and 13) but with high SD. In general, the pattern of changes in the genetic correlations with the length of the interval between measurements was similar for obser_FCR, Gomp_FCR, and interp_FCR. The genetic correlation was highly positive for short time intervals, tended to decrease with the length of the interval, and became negative, resulting in an opposite correlation when the interval between the 2 measurement weeks was more than 4 to 5 wk. The genetic correlations for miss_FCR followed a similar pattern with lower genetic correlations for some of the short time intervals and strong negative correlation for others (−0.90 between wk 5 and 11). For 2 given weeks, the SD of the genetic correlation coefficients of miss_FCR was higher than those of Gomp_FCR and interp_FCR. Among interp_FCR and Gomp_FCR, the SD of correlation coefficients of interp_FCR was higher than those of Gomp_FCR from wk 9 to 13. The same patterns of correlation were observed for the permanent environmental effect (result not shown). We also observed that the means of residual variance of interp_FCR and obser_FCR were comparable (0.12, 0.12) and higher than the residual variance of miss_FCR (0.10) and double that of Gomp_FCR (0.06).

In the simulation study, changes in heritability over time for obser_FCR, miss_FCR, interp_FCR, and Gomp_FCR are illustrated in Fig. 2. The pattern of heritability over time was quite similar for interp_FCR and obser_FCR. Heritabilities obtained for Gomp_FCR followed a similar pattern, but estimates were generally higher, except for wk 4, 5, 8, and 13. The 3 curves (heritabilities of obser_FCR, interp_FCR, and Gomp_FCR) tended to decrease up to wk 8 and then to increase and reach a maximum value in wk 13. Compared to heritability of obser_FCR, the heritability of interp_FCR was slightly higher in wk 9, 10, and 11. On the other hand, the heritability estimates obtained for miss_FCR were higher than those obtained with the other scenarios at all time points (except week 4) and reached very high values. The SD of the heritabilities across simulations was much larger for miss_FCR than for Gomp_FCR or interp_FCR.

The mean and SD of correlations among the EBV in the different scenarios with 100 simulations are presented in Table 3. The correlations between obser_EBV and the other EBV were significantly different from 1. We observed a higher average correlation between obser_EBV and interp_EBV (0.93) than between obser_EBV and Gomp_EBV (0.91) and between

Table 3. Pearson correlation ± SD between the EBV obtained for obser_FCR and those obtained for miss_FCR, Gomp_FCR, and interp_FCR in the simulation study of missing BW (100 replicates)

Week	obser_FCR, miss_FCR[1]	obser_FCR, Gomp_FCR[1]	obser_FCR, interp_FCR[1]
4	0.86 ± 0.02	0.86 ± 0.01	0.86 ± 0.01
5	0.49 ± 0.02	0.78 ± 0.02	0.81 ± 0.02
6	0.39 ± 0.02	0.71 ± 0.03	0.79 ± 0.02
7	0.57 ± 0.03	0.80 ± 0.02	0.84 ± 0.02
8	0.85 ± 0.04	0.93 ± 0.01	0.90 ± 0.02
9	0.76 ± 0.04	0.91 ± 0.01	0.90 ± 0.02
10	0.68 ± 0.04	0.83 ± 0.02	0.87 ± 0.02
11	0.69 ± 0.03	0.81 ± 0.02	0.86 ± 0.01
12	0.77 ± 0.02	0.85 ± 0.01	0.88 ± 0.01
13	0.63 ± 0.03	0.84 ± 0.02	0.83 ± 0.02
All weeks	0.79 ± 0.04	0.91 ± 0.01	0.93 ± 0.01

[1]Here obser_FCR = feed conversion ratio (FCR) computed using available data; miss_FCR = FCR computed using data with a simulated pattern of missing BW; interp_FCR = FCR computed using ADG calculated with the by nearest interpolation; Gomp_FCR = FCR computed using missing BW predicted by a Gompertz model.

obser_EBV and miss_EBV (0.79). The correlation of obser_EBV with other predictions varied depending on the week from 0.39 to 0.86 for miss_EBV, from 0.71 to 0.93 for Gomp_EBV, and from 0.79 to 0.90 for interp_EBV. The lowest correlation between EBV was observed in wk 6 in all cases. The correlations between obser_EBV and EBV obtained in the other scenarios were generally the highest in wk 4, 8, and 12.

Full Data Set

The percentage of missing BW in the full data set was 34%, leading to 41.6% missing FCR. Correction using the by nearest interpolation reduced this percentage to 9% and to 0.3% using the Gompertz model to predict missing BW. The genetic variances and correlations obtained for the full data set (results not shown) were in the same range as those obtained for only males (obser_FCR in the simulation study). The heritabilities of miss_FCR, Gomp_FCR, and interp_FCR using the full data set are listed in Table 4. The patterns of heritability over time were similar in the 3 scenarios. Heritability decreased until wk 8 or 9 and increased again until the end of the test. These values ranged from 0.14 to 0.34 (miss_FCR), 0.15 to 0.35 (interp_FCR), and 0.17 to 0.35 (Gomp_FCR), which are higher than those obtained for only males (observed FCR in the simulation study). The heritabilities of Gomp_FCR were slightly higher than those of obser_FCR and miss_FCR from wk 5 to 7. In general, the SE were similar for miss_FCR, interp_FCR, and Gomp_FCR. The correlations between EBV obtained for the different FCR with the full data set are listed in Table 5. The correlations for all the weeks between EBV were high (>0.96). Within weeks, the correlations between EBV were lower but always higher than 0.87.

Table 4. Heritability over time (10 wk) for the full data set using miss_FCR, Gomp_FCR, and interp_FCR

Week	miss_FCR[1] (h^2 ± SE)	Gomp_FCR[1] (h^2 ± SE)	interp_FCR[1] (h^2 ± SE)
4	0.33 ± 0.04	0.33 ± 0.04	0.34 ± 0.04
5	0.30 ± 0.03	0.35 ± 0.03	0.31 ± 0.03
6	0.26 ± 0.03	0.33 ± 0.03	0.25 ± 0.03
7	0.21 ± 0.02	0.25 ± 0.02	0.19 ± 0.02
8	0.16 ± 0.02	0.17 ± 0.02	0.15 ± 0.02
9	0.14 ± 0.02	0.16 ± 0.02	0.17 ± 0.02
10	0.15 ± 0.03	0.20 ± 0.03	0.20 ± 0.03
11	0.18 ± 0.03	0.23 ± 0.03	0.21 ± 0.03
12	0.24 ± 0.04	0.23 ± 0.03	0.24 ± 0.04
13	0.34 ± 0.05	0.31 ± 0.04	0.35 ± 0.05

[1]Here miss_FCR = feed conversion ratio (FCR) computed using observed data with missing BW; interp_FCR = FCR computed using ADG calculated with the by nearest interpolation; Gomp_FCR = FCR computed using missing BW predicted by a Gompertz model.

DISCUSSION

The purpose of this study was to evaluate the impact of missing BW values on the estimation of genetic parameters of longitudinal FCR and to evaluate techniques to improve these estimations. In our experimental pig population, the BW of females and castrated males was recorded only monthly, which complicated the estimation of genetic parameters for weekly FCR. In the current study, we predicted the missing BW to improve the genetic evaluation for FCR. Two methods were tested. The first uses a growth curve model. Several growth models (von Bertalanffy, Richards, logistic, etc.) are proposed in the literature (e.g., Strathe et al., 2010; Coyne et al., 2015), among which we chose the Gompertz model. This growth model has been widely used to study growth curves in pigs (Koivula et al., 2008; Strathe et al., 2010; Cai et al., 2011; Coyne et al., 2015) and in other species (Narinc et al., 2010; Podisi et al., 2013; Goldberg and Ravagnolo, 2015). It is reported to be a suitable approach for data extrapolation (Koivula et al., 2008; Coyne et al., 2015) and requires fewer parameters to obtain the equivalent data fit than corresponding linear models (Archontoulis and Miguez, 2015). Using the Gompertz model to fit the BW over time, we assumed that the pigs developed normally and that their growth followed the classical sigmoidal curve. The second method we tested is a less elaborate approach to predict missing BW. If ADG for week j could not be computed due to missing BW values at week $j + 2$ or $j - 2$, then we used the nearest available BW in the adjacent weeks and modified the time

Table 5. Pearson correlations between the EBV obtained for the full data set using miss_FCR, Gomp_FCR, and interp_FCR

Week	interp_FCR, Gomp_FCR[1]	miss_FCR, Gomp_FCR[1]	miss_FCR, interp_FCR[1]
4	0.92	0.93	0.96
5	0.89	0.89	0.94
6	0.90	0.90	0.93
7	0.92	0.92	0.93
8	0.96	0.95	0.93
9	0.96	0.93	0.92
10	0.93	0.90	0.92
11	0.93	0.91	0.93
12	0.95	0.95	0.93
13	0.92	0.91	0.87
All weeks	0.96	0.94	0.97

[1]Here miss_FCR = feed conversion ratio (FCR) computed using observed data with missing BW; interp_FCR = FCR computed using ADG calculated with the by nearest interpolation; Gomp_FCR = FCR computed using missing BW predicted by a Gompertz model.

interval accordingly to calculate ADG. This is equivalent to a linear interpolation of BW using the 2 nearest BW records, which leads to changes in ladder steps of ADG over time. A slightly different option would have been to perform a "real" linear interpolation of BW before computing ADG (Zumbach et al., 2010).

We used a RR-PL model to study the repeated measurements of FCR. Different approaches are available to account for the correlation between successive measurements for the estimation of genetic parameters (character process model, spline model, structured antedependence model; Jaffrézic et al., 2003; Jaffrézic, 2004; Borquis et al., 2013; Xie and Zimmerman, 2013). However, the RR-PL model is one of the most frequently used approaches in longitudinal genetic studies thanks to its ease of use and speed of convergence (Speidel et al., 2010). Such models have, for instance, been widely used for growth traits in pigs (Zumbach et al., 2010), for milk production in cattle and goats (Silva et al., 2013), for egg production in poultry (Wolc and Szwaczkowski, 2009), for volume of ejaculate in Holstein bulls (Carabaño et al., 2007), and for carcass traits in beef cattle (Englishby et al., 2016).

The moderate heritability obtained for the observed records (obser_FCR) in male pigs is in line with estimates obtained in a previous study in the same population (0.24; Saintilan et al., 2012). The higher heritability values obtained for the full data set (males + females + castrated males) are in line with the higher value of heritability in castrated males than in males reported by Saintilan et al. (2012) for FCR (0.41 vs. 0.24). In accordance with the conclusions of these authors, we assumed that FCR corresponded to the same trait in the 3 genders. For obser_FCR, Gomp_FCR, and interp_FCR, we observed the same pattern of heritability over time, which tended to be higher at the beginning and at the end of the test period. This trend could be due to the "border effect" problem previously reported for RR models (i.e., an increase in variance at the borders of the test space; Sesana et al., 2010; Wolc et al., 2011; David et al., 2015).

Generally, our results show that the genetic correlations estimated between adjacent weeks were high. They decreased with an increase in the interval between weeks and reached high negative values for weeks separated by a longer period, which is unlikely to reflect the true correlations between these long periods. This pattern may reflect a compensatory growth phenomenon (Fabian et al., 2004; Kamalakar et al., 2009). Nonetheless, the phenomenon is generally observed when the animals' feed is restricted at the beginning of the measurement period, which was not the case in our study. Another possible explanation is that RR-PL models provided biased estimates of the correlations, and in fact, RR-PL models cannot handle a correlation pattern that decreases asymptotically to zero (Jaffrézic et al., 2004). In that case, the correlations become negative and subsequently increase again (David et al., 2015), as observed in our case.

In the simulation study, the heritabilities of FCR obtained with a missing-values pattern (miss_FCR) were very different from those obtained with the other FCR (for most weeks the 95% quantile for the heritabilities of miss_FCR and obser_FCR did not overlap), indicating that a high proportion of missing values leads to an overestimation of the genetic parameters. This was related to a combination of decreased residual variance and increased genetic variance compared to interp_FCR and obser_FCR. In addition, the SD of the heritability of miss_FCR was much higher than that of obser_FCR, Gomp_FCR, and interp_FCR, certainly because of the less accurate available measurements. Finally, the EBV estimated for miss_FCR were quite different from those of obser_FCR.

Accurate prediction of the missing BW is thus necessary to obtain accurate estimations and predictions for selection. Results of the simulation study showed that the by nearest linear interpolation provided better estimates of heritabilities and EBV (obser_FCR corresponding to the "true" heritabilities and EBV) than the prediction of missing BW using a Gompertz model, which tended to overestimate them. This result is probably explained by the fact that in the growth period in our particular data set, the increase in BW over time was quasi-linear and not sigmoidal, as assumed in the Gompertz model (Porter et al., 2010). The Gompertz model was therefore not the most appropriate model to fit the missing BW during the period of measurements. Compared to the nearest approach, the Gompertz model

has been cited as a reference to provide a good evaluation of individual growth dynamics (Koivula et al., 2008), suggesting that selection of growth curve parameters can be envisaged. The purpose of our study was different: The growth curve was used as a means to predict missing BW and to use the missing data together with the original available BW in the genetic analysis. We thus conclude that because it smooths individual variability, the Gompertz model is not appropriate in this situation. However, it should be noted that for a longer test period (i.e., including the sigmoidal portion of the growth curve) or when larger proportions of BW are missing in a data set, leading to high proportions of missing FCR with the by nearest linear interpolation, the use of Gompertz models could be appropriate.

The highest correlations between EBV were obtained for wk 4, 8, and 12 because the proportion of missing BW was low in wk 2, 6, 10, and 14, which were used to compute FCR for wk 4, 8, and 12 (the proportions of missing FCR in these weeks were lower: 29%, 26%, and 22%, respectively). It should be noted that the heritabilities obtained for the different FCR were also the closest for these weeks for the same reason. This explains the dips observed in the heritability of miss_FCR in wk 4, 8, and 12.

The proportion of missing FCR in the full data set was 41.6%, which was reduced to 9% with the by nearest linear interpolation and to 0.3% with the Gompertz approach. In the latter, the proportion of missing FCR was negligible because prior to our study, animals for which the Gompertz model showed convergence problems were removed from the analysis. Thus, the proportion of missing Gomp_FCR in the full data set corresponded to only the proportion of FCR values <0 or >6. In contrast to the results with the simulated data, the heritabilities obtained with the full data set with miss_FCR were not much larger than those obtained for Gomp_FCR or interp_FCR. This might be due to the overall lower proportion of missing FCR and, particularly, the absence of weeks with more than 95% of missing values. In fact, the highest proportion of missing FCR values per week was 59% for the full data set, meaning there was sufficient information per week to estimate the parameters of the RR-PL model. Nonetheless, for the sake of simplicity, we assumed the same changes in FCR and heritability over time for the 3 genders in our RR-PL model. These assumptions are questionable. Changes in FCR and heritabilities over time that differ between genders may be more realistic: A genetic correlation close to 1 was estimated by Saintilan et al. (2012) between genders at the test level, but growth dynamics could differ between genders. In that case, the effect on heritability and EBV estimates of missing BW records in a given gender would probably be larger and in the range of those obtained in the simulated study. In such a situation, the prediction of missing BW to calculate weekly FCR is very useful to obtain accurate longitudinal estimates of FCR.

Conclusion

This study showed that 61.5% of missing BW led to a major overestimation in heritability and EBV for longitudinal FCR. Using the Gompertz model to predict the BW reduced this phenomenon. However, in growth periods with a quasi-linear increase in BW over time, the by nearest approach provided better estimations of genetic parameters.

LITERATURE CITED

Agriculture and Horticulture Development Board. 2016. 2014 pig cost of production in selected countries: AHDB market intelligence. http://www.pork.ahdb.org.uk/media/74797/cost-of-production-web-2014.pdf. (Accessed 26 August 2016.)

Archontoulis, S. V., and F. E. Miguez. 2015. Nonlinear regression models and applications in agricultural research. Agron. J. 107:786–798. doi:10.2134/agronj2012.0506

Borquis, R. R. A., F. R. de A. Neto, F. Baldi, N. Hurtado-Lugo, G. M. F. de Camargo, M. Muñoz-Berrocal, and H. Tonhati. 2013. Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. J. Dairy Sci. 96:5923–5932. doi:10.3168/jds.2012-6023

Cai, W., H. Wu, and J. C. M. Dekkers. 2011. Longitudinal analysis of body weight and feed intake in selection lines for residual feed intake in pigs. Asian-Australas. J. Anim. Sci. 24:17–27. doi:10.5713/ajas.2011.10142

Carabaño, M. J., C. Díaz, C. Ugarte, and M. Serrano. 2007. Exploring the use of random regression models with Legendre polynomials to analyze measures of volume of ejaculate in Holstein bulls. J. Dairy Sci. 90:1044–1057. doi:10.3168/jds.S0022-0302(07)71591-6

Coyne, J. M., D. P. Berry, E. A. Mäntysaari, J. Juga, and N. McHugh. 2015. Comparison of fixed effects and mixed model growth functions in modelling and predicting live weight in pigs. Livest. Sci. 177:8–14. doi:10.1016/j.livsci.2015.03.031

David, I., J. Ruesche, L. Drouilhet, H. Garreau, and H. Gilbert. 2015. Genetic modeling of feed intake. J. Anim. Sci. 93:965–977. doi:10.2527/jas.2014-8507

Englishby, T. M., G. Banos, K. L. Moore, M. P. Coffey, R. D. Evans, and D. P. Berry. 2016. Genetic analysis of carcass traits in beef cattle using random regression models. J. Anim. Sci. 94:1354–1364. doi:10.2527/jas.2015-0246

Fabian, J., L. I. Chiba, L. T. Frobish, W. H. McElhenney, D. L. Kuhlers, and K. Nadarajah. 2004. Compensatory growth and nitrogen balance in grower-finisher pigs. J. Anim. Sci. 82:2579–2587. doi:10.2527/2004.8292579x

Fischer, T. M., A. R. Gilmour, and J. H. J. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Genet. Sel. Evol. 36:363–369. doi:10.1186/1297-9686-36-3-363

Gibbons, R. D., D. Hedeker, and S. DuToit. 2010. Advances in analysis of longitudinal data. Annu. Rev. Clin. Psychol. 6:79–107. doi:10.1146/annurev.clinpsy.032408.153550

Gilbert, H., J. P. Bidanel, J. Gruand, J. C. Caritez, Y. Billon, P. Guillouet, H. Lagant, J. Noblet, and P. Sellier. 2007. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J. Anim. Sci. 85:3182–3188. doi:10.2527/jas.2006-590

Gilmour, A. R., B. J. Gogel, and B. R. Cullis. 2009. ASReml user guide release 3.0. VSN Int., Hemel Hempstead, UK.

Goldberg, V., and O. Ravagnolo. 2015. Description of the growth curve for Angus pasture-fed cows under extensive systems. J. Anim. Sci. 93:4285–4290. doi:10.2527/jas.2015-9208

Jaffrézic, F. 2004. Multivariate character process models for the analysis of two or more correlated function-valued traits. Genetics 168:477–487. doi:10.1534/genetics.103.019554

Jaffrézic, F., R. Thompson, and W. G. Hill. 2003. Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits. Genet. Res. 82:55–65. doi:10.1017/S0016672303006281

Jaffrézic, F., E. Venot, D. Laloe, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves. J. Anim. Sci. 82:3465–3473. doi:10.2527/2004.82123465x

Kamalakar, R. B., L. I. Chiba, K. C. Divakala, S. P. Rodning, E. G. Welles, W. G. Bergen, C. R. Kerth, D. L. Kuhlers, and N. K. Nadarajah. 2009. Effect of the degree and duration of early dietary amino acid restrictions on subsequent and overall pig performance and physical and sensory characteristics of pork. J. Anim. Sci. 87:3596–3606. doi:10.2527/jas.2008-1609

Koivula, M., M.-L. Sevón-Aimonen, I. Strandén, K. Matilainen, T. Serenius, K. J. Stalder, and E. A. Mäntysaari. 2008. Genetic (co) variances and breeding value estimation of Gompertz growth curve parameters in Finnish Yorkshire boars, gilts and barrows. J. Anim. Breed. Genet. 125:168–175. doi:10.1111/j.1439-0388.2008.00726.x

Labroue, F., R. Guéblez, and P. Sellier. 1997. Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace growing pigs. Genet. Sel. Evol. 29:451. doi:10.1186/1297-9686-29-4-451

Losinger, W. C. 1998. Feed-conversion ratio of finisher pigs in the USA. Prev. Vet. Med. 36:287–305. doi:10.1016/S0167-5877(98)00094-4

Nardone, A., B. Ronchi, N. Lacetera, M. S. Ranieri, and U. Bernabucci. 2010. Effects of climate changes on animal production and sustainability of livestock systems. Livest. Sci. 130:57–69. doi:10.1016/j.livsci.2010.02.011

Narinc, D., E. Karaman, M. Z. Firat, and T. Aksoy. 2010. Comparison of non-linear growth models to describe the growth in Japanese quail. J. Anim. Vet. Adv. 9:1961–1966. doi:10.3923/javaa.2010.1961.1966

Nobre, P. R. C., I. Misztal, S. Tsuruta, J. K. Bertrand, L. O. C. Silva, and P. S. Lopes. 2003. Analyses of growth curves of Nellore cattle by multiple-trait and random regression models. J. Anim. Sci. 81:918–926. doi:10.2527/2003.814918x

Patience, J. F., M. C. Rossoni-Serão, and N. A. Gutiérrez. 2015. A review of feed efficiency in swine: Biology and application. J. Anim. Sci. Biotechnol. 6:33. doi:10.1186/s40104-015-0031-2

Podisi, B. K., S. A. Knott, D. W. Burt, and P. M. Hocking. 2013. Comparative analysis of quantitative trait loci for body weight, growth rate and growth curve parameters from 3 to 72 weeks of age in female chickens of a broiler-layer cross. BMC Genet. 14:22. doi:10.1186/1471-2156-14-22

Porter, T., E. Kebreab, H. Darmani Kuhi, S. Lopez, A. B. Strathe, and J. France. 2010. Flexible alternatives to the Gompertz equation for describing growth with age in turkey hens. Poult. Sci. 89:371–378. doi:10.3382/ps.2009-00141

Saintilan, R., P. Sellier, Y. Billon, and H. Gilbert. 2012. Genetic correlations between males, females and castrates for residual feed intake, feed conversion ratio, growth rate and carcass composition traits in Large White growing pigs. J. Anim. Breed. Genet. 129:103–106. doi:10.1111/j.1439-0388.2011.00972.x

Sesana, R. C., A. B. Bignardi, R. R. A. Borquis, L. El Faro, F. Baldi, L. G. Albuquerque, and H. Tonhati. 2010. Random regression models to estimate genetic parameters for test-day milk yield in Brazilian Murrah buffaloes: Milk yield in Brazilian Murrah buffaloes. J. Anim. Breed. Genet. 127:369–376. doi:10.1111/j.1439-0388.2010.00857.x

Shirali, M., A. Doeschl-Wilson, P. W. Knap, C. Duthie, E. Kanis, J. M. van Arendonk, and R. Roehe. 2012. Nitrogen excretion at different stages of growth and its association with production traits in growing pigs. J. Anim. Sci. 90:1756–1765. doi:10.2527/jas.2011-4547

Silva, F. G., R. A. Torres, L. F. Brito, R. F. Euclydes, A. L. P. Melo, N. O. Souza, J. I. Ribeiro Jr., and M. T. Rodrigues. 2013. Random regression models using Legendre orthogonal polynomials to evaluate the milk production of Alpine goats. Genet. Mol. Res. 12:6502–6511. doi:10.4238/2013.December.11.1

Speidel, S. E., R. M. Enns, and D. H. Crews. 2010. Genetic analysis of longitudinal data in beef cattle: A review. Genet. Mol. Res. 9:19–33. doi:10.4238/vol9-1gmr675

Strathe, A. B., A. Danfaer, H. Sorensen, and E. Kebreab. 2010. A multilevel nonlinear mixed-effects approach to model growth in pigs. J. Anim. Sci. 88:638–649. doi:10.2527/jas.2009-1822

Wolc, A., E. Barczak, J. Wójtowski, P. Ślósarz, and T. Szwaczkowski. 2011. Genetic parameters of body weight in sheep estimated via random regression and multi-trait animal models. Small Rumin. Res. 100:15–18. doi:10.1016/j.smallrumres.2011.05.009

Wolc, A., and T. Szwaczkowski. 2009. Estimation of genetic parameters for monthly egg production in laying hens based on random regression models. J. Appl. Genet. 50:41–46. doi:10.1007/BF03195650

Xie, Y., and D. L. Zimmerman. 2013. Antedependence models for nonstationary categorical longitudinal data with ignorable missingness: Likelihood-based inference. Stat. Med. 32:3274–3289. doi:10.1002/sim.5763

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127:93–99. doi:10.1111/j.1439-0388.2009.00827.x

2.3. Evaluating the impact of the missing BW with a simple repeatability model

In addition to the results presented in article I, we also investigated if missing BW records have an impact on the EBV estimation for FCR when a simple repeatability model is used. The heritability for FCR (mean ± SE) obtained with the simple repeatability model applied to the full longitudinal dataset was 0.05 ± 0.01, which is lower than those reported in the literature for FCR recorded on the test period (Saintilan et al., 2013; Gilbert et al., 2017), and lower than those of the RR models reported in above paper (Huynh-Tran et al., 2017).

Using the same dataset, approach and notations as in article I, the results obtained for the simple repeatability model are the following: On simulations, the mean of heritability estimates (± SE) over 100 simulation were: 0.05 ± 0.01, 0.05 ± 0.01 and 0.05 ± 0.02 for Gomp_FCR, inter_FCR and miss_FCR, respectively. The correlations between EBV using Gomp_FCR and interp_FCR with obser_EBV were the same (0.96 ± 0.01), whereas the correlation between EBV from miss_FCR and obser_FCR was lower (0.89 ± 0.02). On the complete data, the correlation between EBV obtained for Gomp_FCR and miss_FCR was also equal to that of EBV with inter_FCR and miss_FCR (0.97 for both of them). This suggested that the missing BW also had an impact on EBV predictions when a simple repeatability model is used. However, in the case of the simple repeatability model the quasi-linear interpolation was equivalent to the Gompertz model for improving the prediction, which was not the case with random regression results.

2.4. Conclusion

This study showed that:

- When the proportion of missing BW is high, the genetic parameters of FCR are not well estimated.
- With the random regression model, prediction of missing BW using a Gompertz growth model slightly improved the EBV prediction but the quasi-linear interpolation was much better in our dataset where the trend is linear rather than sigmoidal.
- With a simple repeatability model, predicting missing BW using a quasi-linear interpolation or a Gompertz model led to the same improvement in EBV prediction for FCR.

REFERENCES

Gilbert, H., Y. Billon, L. Brossard, J. Faure, P. Gatellier, F. Gondret, E. Labussière, B. Lebret, L. Lefaucheur, N. Le Floch, I. Louveau, E. Merlot, M.-C. Meunier-Salaün, L. Montagne, P. Mormede, D. Renaudeau, J. Riquet, C. Rogel-Gaillard, J. van Milgen, A. Vincent, and J. Noblet. 2017. Review: divergent selection for residual feed intake in the growing pig. Anim. Int. J. Anim. Biosci. 1–13. doi:10.1017/S175173111600286X.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92:2377–2386. doi:10.2527/jas.2013-7338.

Saintilan, R., I. Mérour, L. Brossard, T. Tribout, J. Y. Dourmad, P. Sellier, J. Bidanel, J. van Milgen, and H. Gilbert. 2013. Genetics of residual feed intake in growing pigs: Relationships with production traits, and nitrogen and phosphorus excretion traits. J. Anim. Sci. 91:2542– 2554. doi:10.2527/jas.2012-5687.

Tran, V. H. H., H. Gilbert, and I. David. 2017. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J. Anim. Sci. 95:39–48. doi:10.2527/jas.2016.0980.

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127:93–99. doi:10.1111/j.1439-0388.2009.00827.x.

Chapter 3 - LONGITUDINAL FEED CONVERSION RATIO ANALYSES

3.1. About feed conversion ratio in pigs

Although feed conversion ratio is widely used in pig breeding programs and repeated measurements of FCR are sometimes available, few study undertook the genetic evaluation of longitudinal FCR. Our goal was to compare different approaches for longitudinal analyses, with application to FCR.

To reach this goal, we used the data of the 1186 male pigs whose proportion of missing weekly BW (thus FCR) was low (see previous chapter). Means and standard deviations of FCR per week are decribed in Fig 3.1. Different models that can handle changes of the covariance structures for FCR over time were compared: the structured antedependence model (SAD) and the random regression model using Legendre orthogonal polynomial (RR-OP) in the published paper, and the random regression models using spline functions (RR-SL) as a complementary model to evaluate its potential to better handle border effects in the complementary results presented after the paper. The fit-to-the-data, the predictive ability and the genetic parameter estimates of the different models were compared. Finally, because multiple EBV are difficult to handle to take selection decisions, we proposed in the paper a criterion for animal selection based on the longitudinal EBV .

This chapter is organized in the following order :

- Comparison of the RR-OP and SAD models on longitudinal FCR (article II)
- Evaluation of a criterion to select from longitudinal EBV (article II)
- Additional results related to the RR-SL model
- Comparison of the predictive ability of the different models

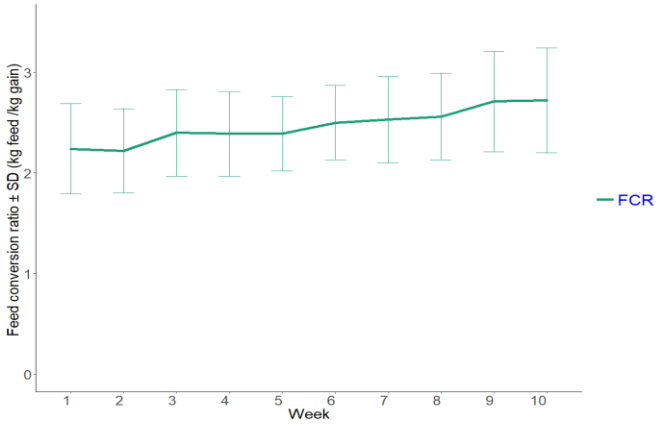


Figure 3.1. Mean and standard deviations of feed conversion ratio over 10 weeks for male pigs.

3.2. Article II: Genetic structured antedependence model and random regression model using orthogonal polynomials for feed conversion ratio in growing Large White pigs

(**Article II:** Huynh-Tran, V. H., H. Gilbert, and I. David. 2017. Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs. J. Anim. Sci. 95:4752–4763. doi:10.2527/jas2017.1864.)

Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs[1]

V. H. Huynh-Tran,[2] H. Gilbert, and I. David

GenPhySE, INRA, Université de Toulouse, INPT, ENVT, 31326 Castanet-Tolosan cedex, France

ABSTRACT: The objective of the present study was to compare a random regression model, usually used in genetic analyses of longitudinal data, with the structured antedependence (SAD) model to study the longitudinal feed conversion ratio (FCR) in growing Large White pigs and to propose criteria for animal selection when used for genetic evaluation. The study was based on data from 11,790 weekly FCR measures collected on 1,186 Large White male growing pigs. Random regression (RR) using orthogonal polynomial Legendre and SAD models was used to estimate genetic parameters and predict FCR-based EBV for each of the 10 wk of the test. The results demonstrated that the best SAD model (1 order of antedependence of degree 2 and a polynomial of degree 2 for the innovation variance for the genetic and permanent environmental effects, i.e., 12 parameters) provided a better fit for the data than RR with a quadratic function for the genetic and permanent environmental effects (13 parameters), with Bayesian information criteria values of −10,060 and −9,838, respectively. Heritabilities with the SAD model were higher than those of RR over the first 7 wk of the test. Genetic correlations between weeks were higher than 0.68 for short intervals between weeks and decreased to 0.08 for the SAD model and −0.39 for RR for the longest intervals. These differences in genetic parameters showed that, contrary to the RR approach, the SAD model does not suffer from border effect problems and can handle genetic correlations that tend to 0. Summarized breeding values were proposed for each approach as linear combinations of the individual weekly EBV weighted by the coefficients of the first or second eigenvector computed from the genetic covariance matrix of the additive genetic effects. These summarized breeding values isolated EBV trajectories over time, capturing either the average general value or the slope of the trajectory. Finally, applying the SAD model over a reduced period of time suggested that similar selection choices would result from the use of the records from the first 8 wk of the test. To conclude, the SAD model performed well for the genetic evaluation of longitudinal phenotypes.

Key words: feed efficiency, longitudinal data, pigs, random regression, selection criterion, structured antedependence model

J. Anim. Sci. 2017.95:4752–4763
doi:10.2527/jas2017.1864

INTRODUCTION

Feed efficiency is a benchmark for profitability in pig farming because the cost of feed represents about two-thirds of total production costs. It also reduces the negative effects of livestock farming on the environment (Patience et al., 2015; Gilbert et al., 2017). With the development of automatic equipment, individual feed intake (**FI**) and BW values can be repeatedly measured during the production period in some species. The analysis of repeated records can provide more accurate estimations in a genetic selection context than simple trait analyses (Boligon et al., 2011). To analyze such longitudinal data, genetic models

[1]Our work is part of the H2020 Feed-a-Gene project funded by the European Union's H2020 Programme under grant agreement number 633531. The authors would like to thank the Animal Genetics Division at INRA, France, and the European project H2020 Feed-a-Gene for financial support for Huynh-Tran's PhD thesis. The staff of the experimental facilities GenESI (UE1372, Surgères, France) is acknowledged for breeding the animals and for data collection.

[2]Corresponding author: van-hung.tran@inra.fr

Received June 27, 2017.
Accepted August 28, 2017.

should account for the covariance structures of the repeated records with few parameters to estimate. The random regression (**RR**) model is widely used, even if it presents various drawbacks such as higher variances at the beginning and the end of the studied period, so-called border effect problems (Jaffrézic et al., 2004; Meyer, 2005). The structured antedependence (**SAD**) model also deals with the correlation structure of data and has been shown to better fit covariance structures than RR models (Jaffrézic and Pletcher, 2000; Jaffrézic et al., 2004; David et al., 2015). Up to now, it has been less widely used than the RR model due to the lack of tools; however, user-friendly software (David et al., 2017) is now freely available (https://zenodo.org/record/896377; accessed 20 Sep. 2017). For selection purposes, an interpretable eigenvalue decomposition of the additive genetic matrix **K** of the RR coefficients of the RR model has been proposed to summarize the individual genetic potential over time as 1 or 2 values (Van Der Werf et al., 1998), capturing features such as persistency or area under the curve when applied to lactation curves (Togashi and Lin, 2006). To our knowledge, no methods for summarizing breeding values from the SAD model have yet been proposed. The objective of our study was to compare RR and SAD models for the genetic analysis of repeated measures of the feed conversion ratio (**FCR**) in growing pigs and to propose criteria for animal selection for the SAD model.

MATERIAL AND METHODS

Data were collected in accordance with the applicable national regulations on livestock welfare in France.

Pigs and Data Collection

The present study includes data from 1,186 Large White boars over 8 generations of divergent selection for residual FI raised after weaning in the Rouillé INRA experimental farm (GenESI, Vienne, France). The selection process was described in detail by Gilbert et al. (2007). The data used were collected from candidate boars tested in groups of 12 in pens equipped with single electronic feeders (ACEMA 64; Skiold Acemo, Pontivy, France). Pigs were age 67 ± 1 d (25 ± 4 kg) at the beginning of the test and were tested during the growing–finishing period up to 168 ± 13 d (115 ± 11 kg). The records collected during the first week of the test, when pigs acclimated to the feeders, were discarded from the analysis. Animals were fed ad libitum with a pelleted diet of cereals and soybean meal with 10 MJ NE/kg and 160 g CP/kg, and a minimum of 0.80 g digestible Lys/MJ NE.

During the 14 consecutive weeks (from wk 2 to 15) of the test period, animals were weighed weekly. The individual FI of each animal was automatically recorded each time it used the feeder. Weekly averages of the daily FI (**WDFI**) were then computed for each animal. The WDFI outlier values and WDFI for which more than 2 d of records were missing in a given week were removed from the analysis, as reported by David et al. (2015). The FCR was calculated for each animal i and week j ($j \in \{4, \ldots, 13\}$) as follows (Huynh-Tran et al., 2017):

$$FCR_{ij} = WDFI_{ij}/ADG_{ij},$$

in which $WDFI_{ij}$ is the WDFI of animal i for week j and ADG_{ij} is the ADG of animal i for week j ($j \in \{4, \ldots, 13\}$) estimated over a 4-wk period as follows:

$$ADG_{ij}=(BW_{ij+2}-BW_{ij-2})/(age_{ij+2}-age_{ij-2}),$$

in which BW_{ij} and age_{ij} are the BW and the age of animal i at week j, respectively. Only animals with at least 3 measures of FCR over the 10-wk period (wk 4 to 13) were retained for analysis. Extreme values of FCR (<0 and >4.5) were considered outliers and set as missing. The final data set comprised 11,790 weekly FCR values for 1,186 male growing pigs available from wk 4 to 13 of the test. For the sake of simplicity, we will denote $t_j \in \{1, \ldots, 10\}$ instead of $j \in \{4, \ldots, 13\}$ hereafter. A total of 3,986 animals was included in the pedigree.

Data Analysis

Estimations of Genetic Parameters. Repeated longitudinal FCR measurements were analyzed using the RR and SAD models. Both models can be written, for animal i at time t_j, as

$$FCR_{ij} = \mu_i(t_j) + u_i(t_j) + p_i(t_j) + \varepsilon_{ij}, \qquad [1]$$

in which $\mu_i(t_j)$ is the fixed effect at time t_j; $u_i(t_j)$ and $p_i(t_j)$ are the random genetic and permanent environmental animal effects functions with $\mathbf{u} \sim N(0, \mathbf{G} \otimes \mathbf{A})$ and $\mathbf{p} \sim N(0, \mathbf{P} \otimes \mathbf{I})$, in which **A** is the known relationship matrix; **I** the identity matrix; and **G** and **P** the covariance matrices between weekly measurements of FCR (of dimension 10 × 10) for genetic and permanent environmental effects, respectively. Finally, ε_{ij} is the random residual effect $\varepsilon \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$. The random functions were independent from one another.

In the RR model, for a given random effect $u_i(t_j)$, the general form of the random function of order m is $u_i(t_j) = \sum_{k=0}^{m} a_{ik}\varphi_k(t_j)$, in which a_{ik} is the $(k + 1)^{th}$ RR coefficient for the genetic effects for animal i, with **a** ~

$N(0,\mathbf{K}\otimes\mathbf{A})$, in which **K** is the covariance matrix of the additive RR coefficients, and $\varphi_k(t_j)$ is the $(k+1)^{th}$ Legendre polynomial at time t_j. In the RR model, the relationship between **K** and **G** is given by $\mathbf{G_{RR}} = \varphi\mathbf{K}\varphi'$, in which φ is the $n \times (m+1)$ (in which n is the number of time points) matrix of the Legendre polynomials for all time points.

In the SAD model, each random function is defined by 3 parameters: the order of the antedependance (α), the degree of the polynomial for each antedependence parameter (β_1 to β_α), and the degree of the polynomial for the innovation variance (γ). The function for the random effect **u** is $u_i(t_j) = \sum_{s=1}^{\alpha} \theta_{sj} u_i(t_{j-s}) + e_i(t_j)$, in which θ_{sj} is the s^{th} antedependence parameter for time t_j and $e_i(t_j)$ is the error term for animal i at time t_j; **u** and **e** are independent and $e(t_j) \sim N(0, \mathbf{A}\sigma_e^2(t_j))$. To reduce the number of parameters in the SAD model, continuous functions of time were assumed for antedependence parameters $\theta_{sj} = \sum_{q=0}^{\beta_s} d_{sq} t_j^q$ and for the innovation variance $\sigma_e^2(t_j) = \exp\left(\sum_{q=0}^{\gamma} b_q t_j^q\right)$, in which d_{sq} and b_q are the coefficients for antedependence parameters and innovation variance.

We noted a SAD model with a given set of parameters as follows: SADα − β_1, …, β_αγ. To facilitate convergence and avoid identifiability problems between the structured random permanent environmental effect and the residual covariance matrices (Wang, 2013), the residual term ε_{ij} was removed from Eq. [1] for the SAD model. The residual variance is therefore, for this approach, included in the covariance matrix of the permanent environmental effect.

Covariance components were estimated for both models using the REML method using ASReml software (Gilmour et al., 2009). Estimations for SAD models were computed using the OWN function that allows users of ASReml to model their own variance structure, as proposed by David et al. (2017). The fixed effects included were the same for both models, as previously described by Huynh-Tran et al. (2017).

Both the degree of the polynomial functions for the RR approach and the order and degrees of the antedependence functions in the SAD approach were selected by comparing nested models using likelihood ratio tests. Once the best model for each approach was identified, the data-fitting capacity of the selected RR and SAD models was compared using the Bayesian information criteria (**BIC**; Schwarz, 1978): BIC = $-2\ln(L) + c \times \ln(N-p)$, in which L is the REML of the model, N is the number of observations, and p and c are the number of fixed effects and covariance parameters, respectively. The approach (RR or SAD) with the lowest BIC was considered the best fit for the data.

We compared the heritability and the estimated the genetic covariance matrices obtained using the 2 approaches. The heritability estimates were computed for each week j as $h_j^2 = \hat{\mathbf{G}}_{jj} / \left(\hat{\mathbf{G}}_{jj} + \hat{\mathbf{P}}_{jj} + \sigma_e^2\right)$, in which $\hat{\mathbf{P}}$ and $\hat{\mathbf{G}}$ are the estimates of matrices **P** and **G**, respectively; σ_e^2 is included for only the RR model. Standard errors of heritability estimates were calculated for the RR model in ASReml using the method proposed by Fischer et al. (2004). For the SAD model, analytical expressions of the SE are more difficult to obtain. Therefore, we used a bootstrap procedure to obtain SE for this model. The bootstrap steps were as follows:

1. In iteration l, sample a vector $\mathbf{v}_l$ of antedependence parameters and innovation covariance parameters (for instance, d_{s0} to $d_{s\beta_s}$, b_0 to b_γ are the parameters related to the genetic variance) using multivariate sampling $\mathbf{v}_l \sim \text{MVN}(\hat{\mathrm{v}}, \mathbf{V})$, in which $\hat{\mathrm{v}}$ is the vector of estimates, **V** is their covariance matrix estimated using ASReml, and MVN is multivariate normal distribution;
2. Using $\mathbf{v}_l$, calculate the genetic and permanent environmental variances for each time point (David et al., 2015) and then their heritabilities;
3. Repeat steps 1 and 2 10,000 times to obtain a vector of estimated heritabilities; and
4. Based on the vector of estimated heritabilities, calculate the mean and SE for the heritability (Efron and Hastie, 2016).

The EBV for each time point for the RR and SAD models were obtained as follows: EBV obtained with the SAD model (**EBV_SAD**) were provided in the ASReml outputs and the EBV obtained with the RR model (**EBV_RR**) were computed using the estimations of the individual regression coefficients provided by ASReml as $\text{EBV\_RR}_i(t_j) = \sum_{k=0}^{m} \hat{a}_{ik}\varphi_k(t_j)$. We denoted $\mathbf{sEBV\_RR}_i$ and $\mathbf{sEBV\_SAD}_i$ as the sum of the EBV_RR and EBV_SAD, respectively, for an animal i over the test period.

We compared these EBV with each other and with the overall breeding values (the EBV from the animal model using the FCR computed over the 10-wk period; **cEBV**) obtained by analyzing the FCR for the entire test period computed as the ratio of the ADFI during the 10 wk of test over the ADG for the same period. This overall FCR was analyzed using an animal mixed model: $\text{FCR}_i = \mu_i + u_i + \varepsilon_i$, in which FCR_i is the overall FCR for the entire test period for animal i, μ_i is the fixed effect, u_i is the animal additive genetic effect of animal i, and ε_i is the residual term.

Selection Criterion. Next, computations addressed the issue of defining for each model a criterion to select the best animals based on their 10 weekly

EBV values. First, the patterns of EBV variation over time were described using a trajectory classification approach that classified animals into different trajectory groups using a *k*-means approach with the Euclidean distance. This method used a hill-climbing algorithm jointly with expectation–maximization. The optimal number of clusters was chosen according to the Calinski–Harabatz criterion (Genolini et al., 2015).

Next, the information contained in the 10 EBV was summarized in a reduced number of n_{ind} independent variables using an eigendecomposition of the **G** matrices estimated using the RR and the SAD approaches. This method decomposes the covariance matrix into a set of independent eigenvectors and associated eigenvalues. Each eigenvalue represents the amount of variance explained by the associated eigenvector (Kirkpatrick et al., 1990). Summarized breeding values (**SBV**) associated to the *p*th eigenvector (**SBV*p*$_i$**) were calculated for each animal *i* by multiplying the coresponding eigenvector with the vector of EBV_i. The SBV were denoted **SBV_RR*p*** and SBV_SAD*p* when obtained from the *p*th eigenvector of the **G** matrices of the RR and SAD models, respectively.

For the RR model, we also calculated SBV obtained from the eigendecomposition of the **K** matrix as recommended for this model (Meyer and Hill, 1997; Van Der Werf et al., 1998). These summarized EBV were denoted **SBV_RRK*p***. The SBV_RRK*p* for animal *i* was given by $\text{SBV\_RRK}p_i = \sum_{l=0}^{m} \hat{a}_{il} k_{pl}$, in which k_{pl} is the l^{th} element of the p$^{\text{th}}$ eigenvector of **K** ($p = 0, \ldots, m$). The eigendecomposition of the **K** matrix instead of the **G** matrix has the advantage of producing SBV that can be interpreted in regard to their variation over time. Actually, eigenfunctions of time can be obtained by multiplying the eigenvectors of **K** with the Legendre polynomials (Schnyder et al., 2001; Englishby et al., 2016). The **K** matrix is also usually of reduced dimension compared with **G**.

We then characterized, within each approach, the connection between the EBV group trajectory and the different SBV. In addition, to validate the interpretation of the SBV obtained with the 2 **G** matrices, they were compared with the SBV_RRK*p*. Finally, we compared these SBV with cEBV (EBV for the full test period FCR), sEBV_RR, and sEBV_SAD.

Appropriate Period for Estimating Longitudinal Feed Conversion Ratio. Lastly, we investigated whether FI, and therefore FCR, could be measured over a shorter period without compromising the description of the dynamic of FCR over time, to maintain the possibility to select for features of this dynamic. Reducing the time period for FI recording would allow collecting of records for more animals for this trait, and therefore FCR, and potentially increase the genetic gain. We first defined 3 different 5-wk periods with FCR records (initial period, wk 1 to 5; intermediate period, wk 3 to 7; and late period, wk 6 to 10) to be analyzed using the SAD model. The corresponding SBV were then computed as previously described for this model. The period providing the SBV with the highest correlation with the first and second SBV obtained from the genetic covariance matrix **G** with the SAD model (**SBV_SAD1** and **SBV_SAD2**) obtained over the entire 10-wk period was considered the best period for recording FCR.

Next, starting from the previous best 5-wk period, the number of weeks used in the analysis was increased by 1 wk at a time from 5 to 8 wk, and the same comparison was applied to determine the minimum number of weeks needed to provide a "satisfactory" SBV for FCR.

RESULTS

Estimation of Genetic Parameters

After selection, the RR model of degree 2 for genetic and permanent environmental effects was retained as the best model within the RR category and required 13 parameters to be estimated. Meanwhile, for the SAD approach, SAD1–22 was selected as the best SAD model for genetic and permanent environmental effects and required 12 parameters to be estimated. The BIC values for the best RR and SAD models were −9,838 and −10,060, respectively, indicating that the SAD approach provided the best fit for the data.

The changes in heritabilities over time are shown in Fig. 1. The heritability estimates were generally higher with the SAD model than with the RR model. They ranged from 0.22 to 0.46 (SE 0.03–0.06) for the SAD model and from 0.08 to 0.33 (SE 0.02–0.07) for the RR model. The heritabilities obtained with the RR model decreased up to wk 5 and then increased again toward the end of the test. For the SAD model, the heritability estimates were quite high at the beginning, decreased to a minimum at wk 8 (0.21 ± 0.03), and then increased before the end of the test period to reach values similar to the RR estimations. The ranges of SE were similar for the 2 approaches, from 0.02 to 0.07 for the RR model and from 0.03 to 0.06 for the SAD model.

The genetic correlations estimated for FCR over the 10 wk using the RR and SAD models are presented in Fig. 2. The genetic correlations between 2 given weeks depended on the time interval between the weeks. The shorter the interval, the higher the correlation. Correlations ranged from −0.39 to 0.98 for the RR model and from 0.08 to 0.83 for the SAD model. Consecutive week correlations were high and positive for both models, ranging from 0.91 to 0.98 for the RR

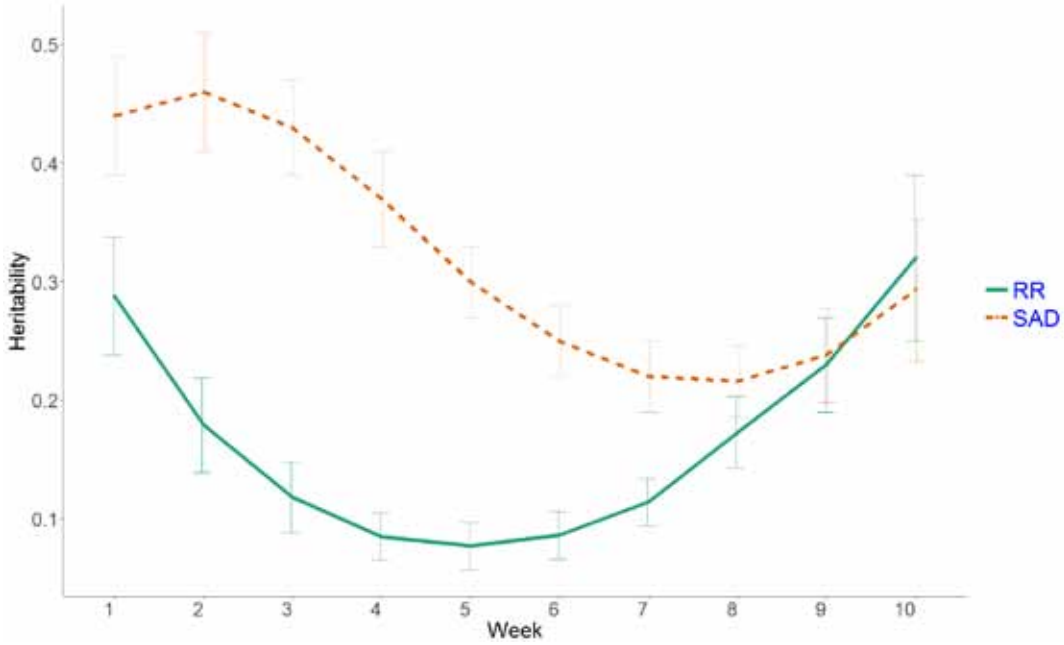


Figure 1. Changes of heritability estimates for feed conversion ratio over time under the random regression (RR) model using Legendre orthogonal polynomials and the structured antedependence (SAD) model. Standard errors are indicated as bars for each point estimate.

model and from 0.68 to 0.83 for the SAD model. For the RR model, the genetic correlations decreased as the interval between the weeks increased and became negative, resulting in negative correlations when the time interval between weeks was more than 5 to 6 wk. For the SAD model, the correlations decreased with the time interval but remained positive.

Selection Criterion

Estimated Breeding Value Trajectory Classification. The Spearman correlation between the weekly EBV_SAD and weekly EBV_RR over 10 wk for all animals was 0.95. The individual EBV trajectories under the 2 models (RR and SAD) were classified into 3 groups as shown in Fig. 3. The 3 patterns of EBV trajectories were similar for both models. Cohen's kappa agreement between the models was 0.80. The first EBV trajectory pattern was a continuous EBV increase over time with a weak slope and a low initial value (35.4 and 40.3% of the animals for the RR and SAD models, respectively; "A" group). The second pattern also reflected an increase of the EBV over time but had a steeper initial slope and higher initial value (34.6 and 32.6% of the animals for the RR and SAD models, respectively; "B" group). The last EBV trajectory pattern simply reflected a constant EBV over time (30.0 and 27.1% of the animals for the RR and SAD models, respectively; "C" group).

Selection Criterion Using Summarized Breeding Values. The approach based on eigendecomposition showed that the 2 first eigenvalues of $\mathbf{G}_{\mathbf{RR}}$ and genetic covariance matrix for SAD model ($\mathbf{G}_{\mathbf{SAD}}$) explained 90 and 73% of the genetic variation, respectively. The correlations between the SBV obtained with the different approaches are presented in Fig. 4. It should be noted that, depending on the program used to compute the eigendecomposition, matrices of eigenvectors of opposite signs can be obtained for the same initial correlation matrix. Therefore, we chose the signs of eigenvectors matrices to maximize the number of positive correlations with cEBV. The first summarized breeding values obtained from the matrix K with the RR model was highly correlated with SBV_RR1 (0.99) and SBV_SAD2(0.99), whereas the second SBV obtained from the coefficients covariance matrix **K** with the RR model (**SBV_RRK2**) was highly correlated with the second summarized breeding value obtained from the genetic covariance matrix **G** with the RR model (**SBV_RR2**; 0.99) and SBV_SAD1 (0.88) and also with sEBV_RR (0.96), sEBV_SAD (0.92), and cEBV (0.93).

In addition, the plots of the first 2 SBV depending on the trajectory clusters previously identified for the RR and SAD models (see Fig. 3) are presented in Fig. 5. For both approaches, the first 2 SBV were sufficient to describe the EBV trajectory types: for instance, for the SAD approach, animals in group A had low SBV_SAD1 values, animals in group B had high SBV_SAD1 and high SBV_SAD2 values, and animals in group C had high SBV_SAD1 and low SBV_SAD2 values. This suggested that SBV_SAD1 captured the average values of EBV over time, whereas SBV_SAD2 captured the slope of the EBV curve. The correspondences between SBV_RR*p* and EBV trajectories obtained with the RR

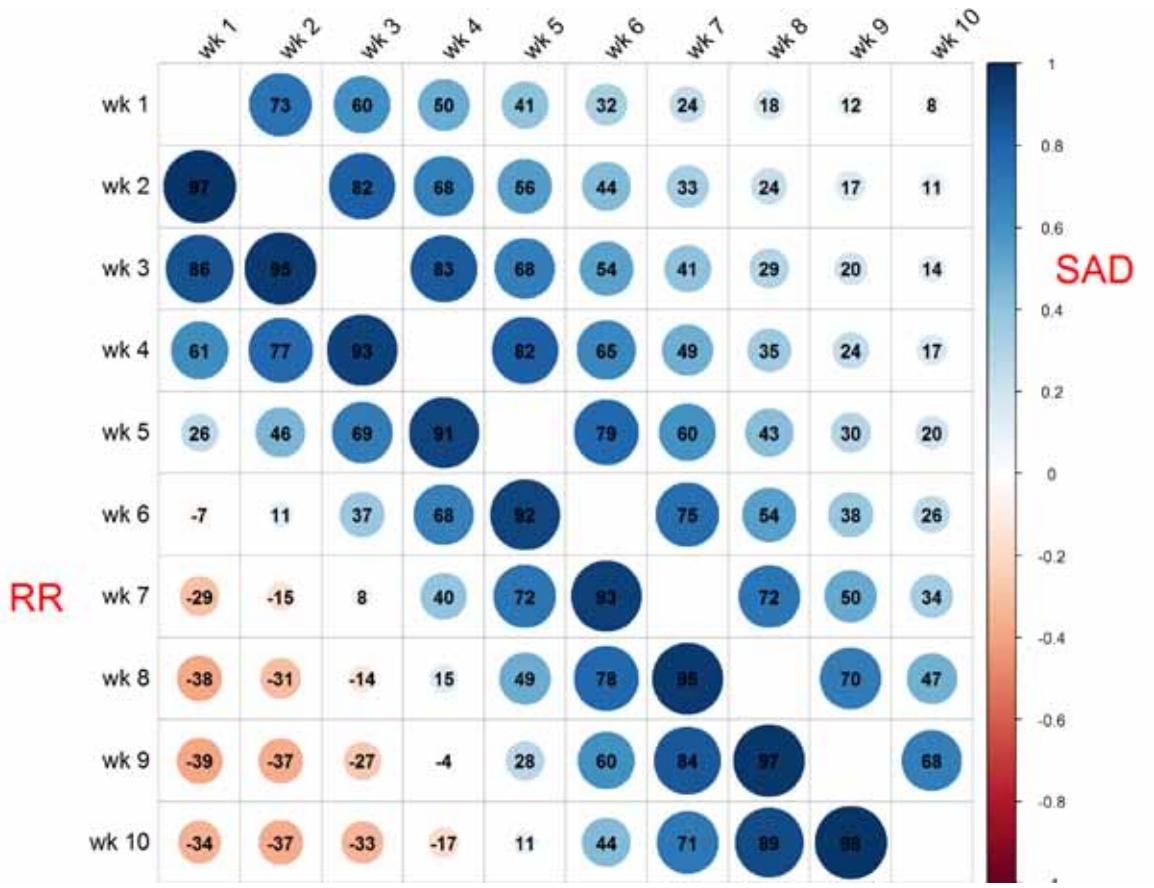


Figure 2. Genetic correlation estimates (x100) between times estimated with the random regression (RR) model (below the diagonal) and the structured antedependence (SAD) model (above the diagonal). The magnitude and sign of the correlations are indicated with darker and larger circles and blue (positive) or red (negative) colors, respectively.

approach also showed a clear distribution of the individuals from each group trajectory according to combinations of SBV_RR*p*. Finally, the plot of the eigenfunctions (Fig. 6) showed that the first eigenfunction was negative during the first 2 wk, then was positive from wk 2 until wk 9, and became negative again and decreased until the end of the test. The second eigenfunction was always positive and stable from wk 1 to 5, then increased, and reached a maximum at the end of the test.

Appropriate Period for Estimating Longitudinal Feed Conversion Ratio. The correlations between SBV_SAD1 and SBV_SAD2 obtained for reduced test periods and SBV_SAD1 and SBV_SAD2 obtained for the whole test period were estimated. The SBV_SAD1 related to the middle period had a higher correlation to SBV_SAD1 for the whole test period than that related to the first 5-wk period (0.93 vs. 0.89, respectively), whereas its correlation with the SBV_SAD2 was lower (0.67 vs. 0.69 for wk 1 to 5 and wk 3 to 7, respectively). When the evaluation period was extended by 1 wk toward the beginning or toward the end of the test period, these correlations did not increase for the middle period, contrary to those of the first 5-wk period extended for wk 6 (results not shown). Therefore, only results for the extended periods starting at the beginning of the test are reported. In this situation, the correlation between SBV_SAD1 (SBV_SAD2) for the reduced period and SBV_SAD1 (SBV_SAD2) for the whole test period increased with the number of weeks included, from 0.89 (0.69; wk 1 to 5) to 0.98 (0.87; wk 1 to 8; Fig. 7).

DISCUSSION

Estimation of Genetic Parameters

Using the BIC, the SAD model showed a slightly better fit to the data than the RR model. Furthermore, the predictive ability 1 wk ahead, computed as proposed by David et al. (2015), was similar for the 2 models (average Vonesh concordance coefficient = 0.39 for both). The SAD model provided higher heritability estimates than the RR model. Similar results have been found in the literature for other traits (Jaffrézic et al., 2004; David et al., 2015). The lower values of heritability obtained with the RR model might be a consequence of the border effect problem associated with this model, which is eliminated in

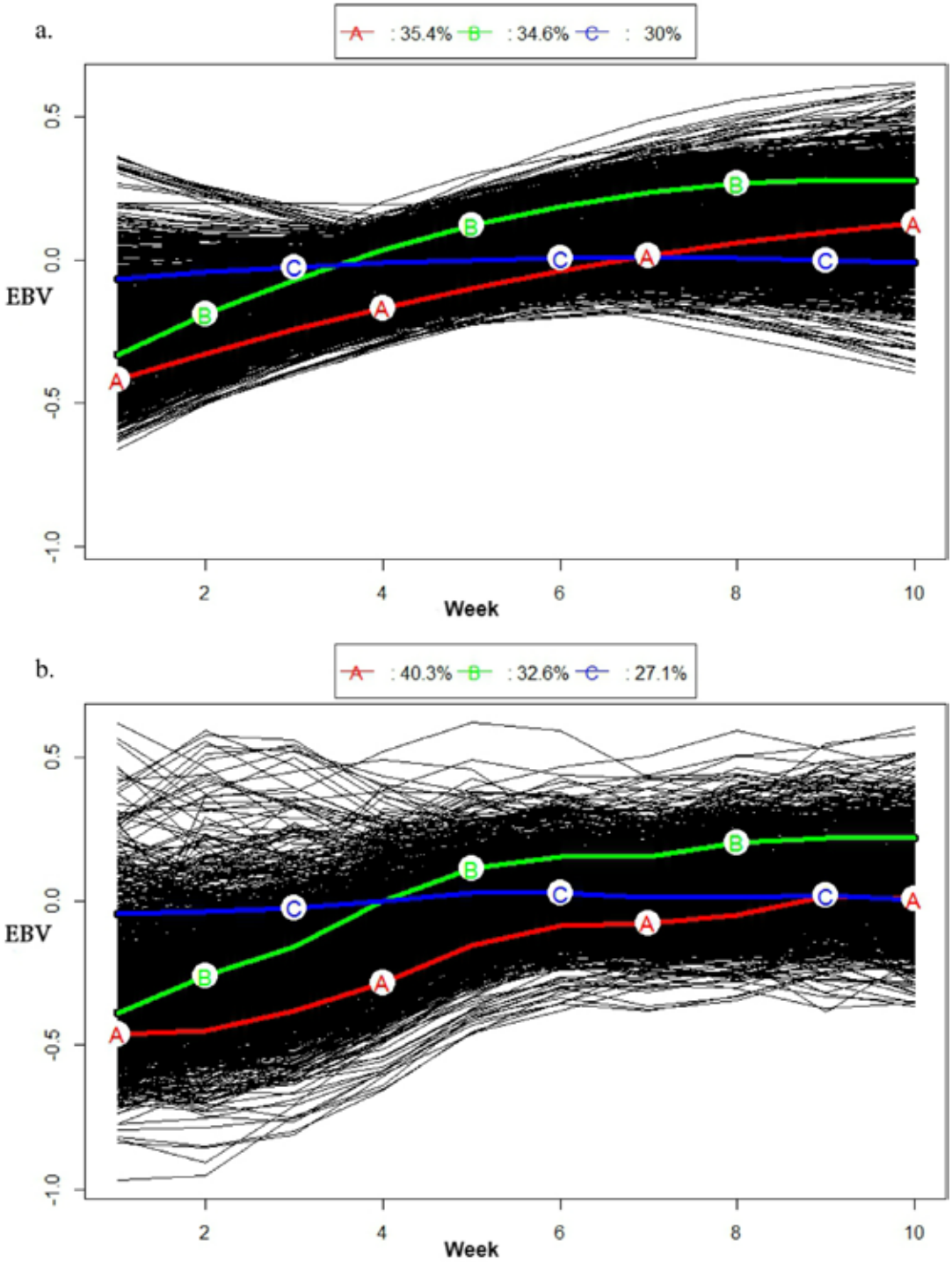


Figure 3. Individual EBV trajectories (in black) and group trajectories resulting from nonhierarchical *k*-means clustering analyses with 3 groups obtained with the random regression (a) and structured antedependence models (b). The proportion of individuals gathered in each group is indicated above each graph.

the SAD model that combines the antedependence parameters and innovation variances (Jaffrézic et al., 2004), suggesting a greater confidence in the genetic parameters obtained with the SAD model. This was reinforced by results from a multiple trait model with a diagonal covariance matrix applied to weekly FCR. Heritabilities obtained with the SAD model were closer to those of the multiple trait model than heritabilities of the RR model with this multiple trait model (average absolute difference = 0.09 vs. 0.15, respectively), the heritabilities being systematically lower with the RR model. Nonetheless, it should be noted that the computing time of the SAD model for each iteration is longer than the one of the RR model (2.7 times longer, on average) but SAD models generally converge with fewer iterations. Consequently, on our data set, the total computing time of the SAD model was 1.2 times longer than for the RR model.

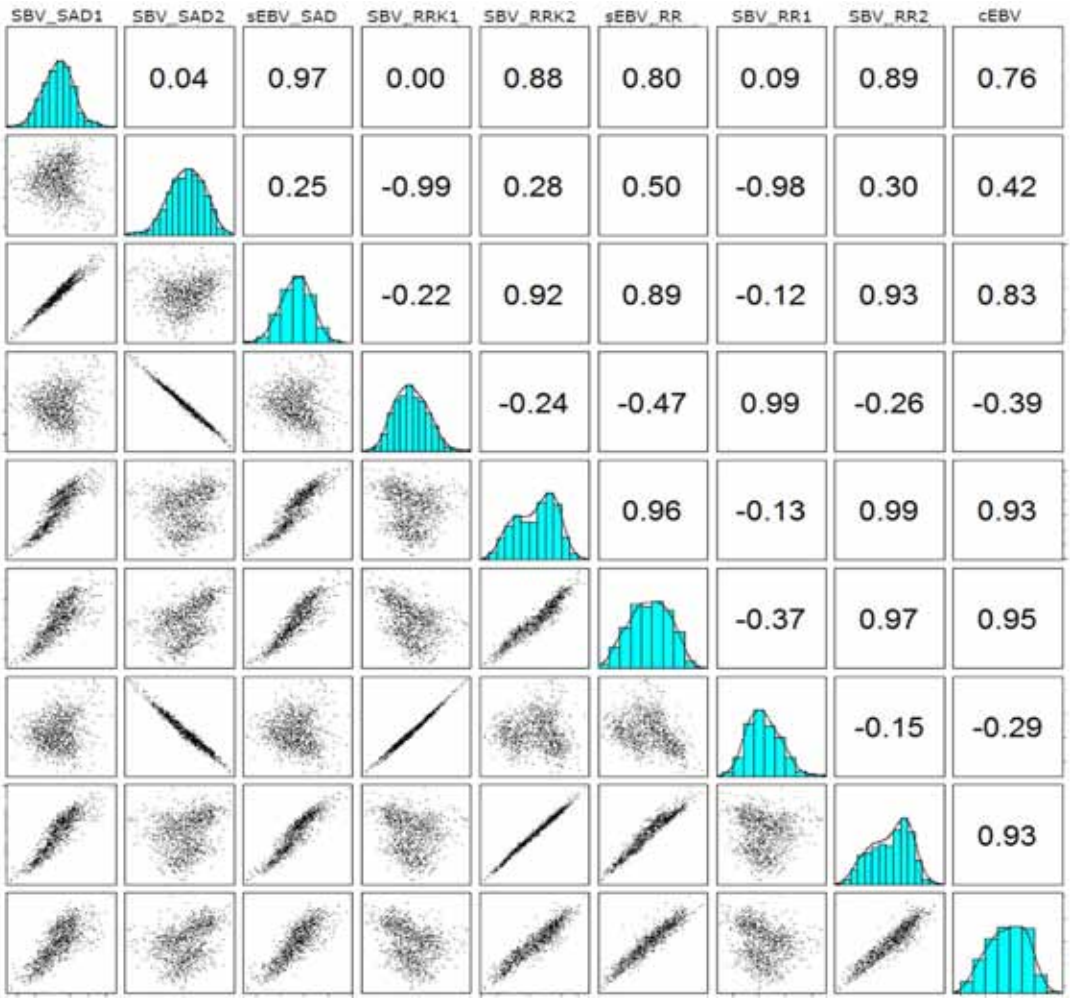


Figure 4. Distributions of the summarized breeding values (SBV_SAD1 and SBV_SAD2 = first and second summarized breeding values, respectively, obtained from the genetic covariance matrix **G** with the structured antedependence [SAD] model; SBV_RRK1 and SBV_RRK2 = first and second summarized breeding values, respectively, obtained from the coefficients covariance matrix **K** with the random regression [RR] model; SBV_RR1 and SBV_RR2 = first and second summarized breeding values, respectively, obtained from the genetic covariance matrix **G** with the RR model), the sums of EBV over the 10 wk (sEBV_RR and sEBV_SAD = sum of EBV obtained with the RR and SAD models, respectively), and the EBV from the animal model using the feed conversion ratio computed over the 10-wk period (cEBV; on the diagonal), joint distributions of these estimates (below the diagonal), and Spearman correlations between the estimates (above the diagonal).

We used a bootstrap procedure to compute the SE of heritability obtained with the SAD model. It is also feasible to use Taylor expansion to obtain an approximate SE. Nonetheless, the formula becomes complex when the order of the antedependence increases.

The heritabilities obtained with the SAD model (from 0.22 to 0.46) at different weeks were in line with those reported in the literature for FCR values recorded over the full growing period on earlier generations of the same population (0.24 ± 0.06; Saintilan et al., 2012) and in other Large White/Yorkshire populations: 0.26 ± 0.07 (Bunter et al., 2010), 0.30 ± 0.03 (Saintilan et al., 2013), and 0.32 ± 0.05 (Do et al., 2013). The changes of heritabilities with time are consistent with the assumption that different genes can be associated with FCR at different stages of growth, as suggested by Shirali et al. (2013) for residual FI and FCR.

For the SAD and RR models, the genetic correlations decreased as the time interval between measurements increased. They became negative in the case of the RR model, although this is unlikely to reflect the true correlations between these distant periods. It has been previously reported that because the RR model cannot handle correlations that asymptotically tend to 0, it provides biased estimates of the correlations for distant time intervals (Jaffrézic et al., 2004).

In such cases, the correlations become negative, as observed in previous studies (David et al., 2015). It should be noted that considering heterogeneous residual variance with time in the RR model did not modify these negative value estimates and did not reduce the border effects problem (results not shown). The positive genetic correlations over time estimated with the SAD model suggest that efficient animals with low FCR values at the beginning of the test period tend

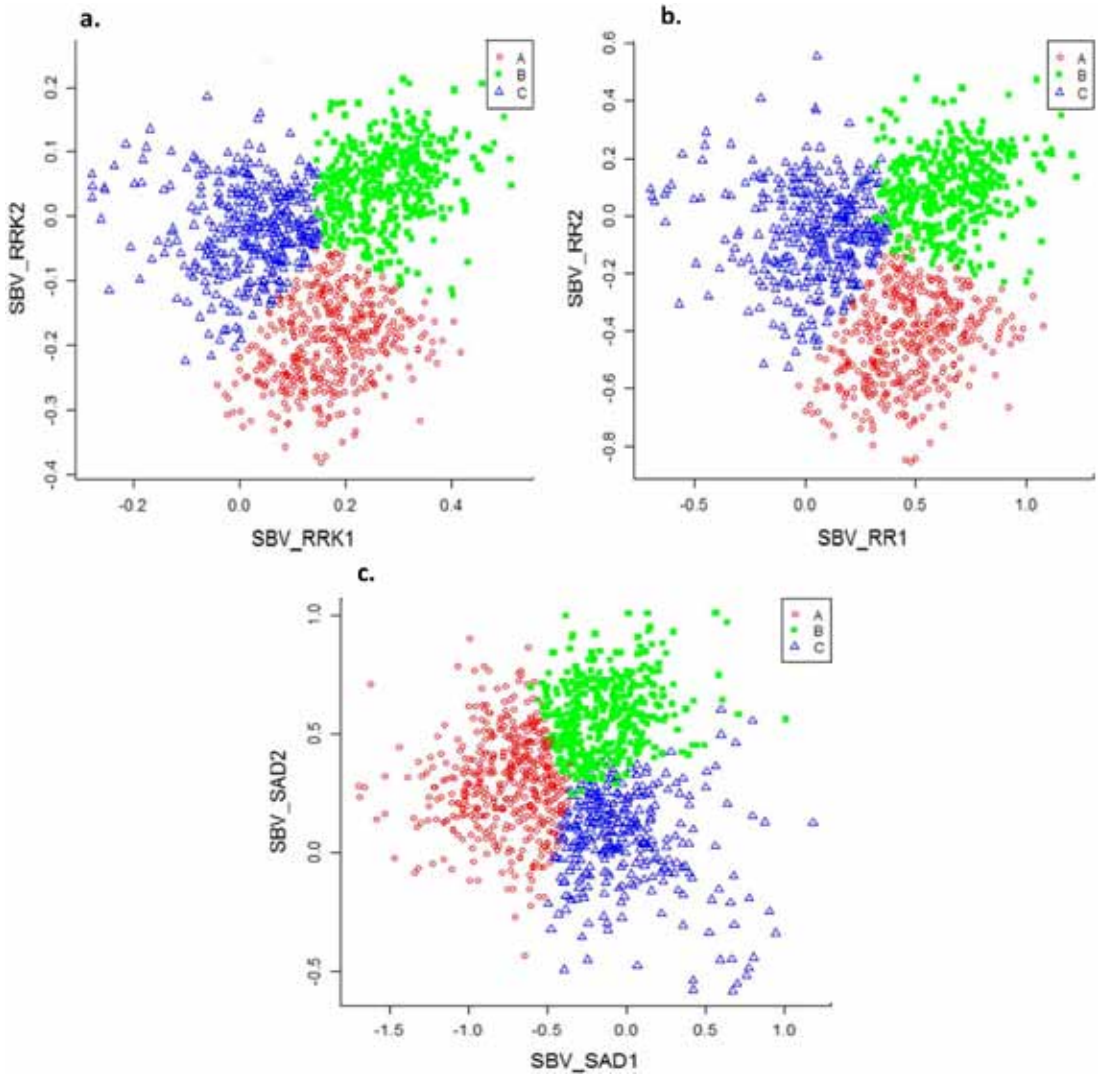


Figure 5. Scatterplots of the individual first and second summarized breeding values obtained from the estimated covariance matrices **K** with the random regression (RR) model (a; SBV_RRK1 and SBV_RRK2, respectively), from the genetic covariance matrix **G** with the RR model (b; SBV_RR1 and SBV_RR2, respectively), and from the genetic covariance matrix **G** with the structured antedependence (SAD) model (c; SBV_SAD1 and SBV_SAD2, respectively). The groups of trajectories to which each individual belongs as determined using the nonhierarchical *k*-means approach (see Fig. 3) applied to the longitudinal EBV from the RR model (a and b) and from the SAD model (c) are indicated as red circles (A group), green squares (B group), and blue triangles (C group).

to also have a lower FCR toward the middle of the test, but more independent results seem to be expected toward the end of the test. Henryon et al. (2002) estimated the genetic correlations between FCR values in growing rainbow trout at different time points over a 215-d period. Most of the correlations were positive and high. No reports about such genetic correlations for FCR could be found in the literature for pigs.

Selection Criterion

Modeling longitudinal data yields more accurate EBV due to the inclusion of repeated records over time and consideration of the covariance structure of the data (Boligon et al., 2011). However, the main difficulty of selection based on repeated measurement analysis is the obtention of as many EBV as time points used for the evaluation. The general idea is, therefore, to summarize these multiple EBV into a smaller set of new composite dimensions with a minimum loss of information (Van Der Werf et al., 1998), as successfully applied by Buzanskas et al. (2013). Ideally, 1 or 2 indexes can capture the individual EBV trajectory profiles to ease animal selection.

In the current study, a classification approach was used to identify different typical EBV trajectories from the SAD and RR approaches, as earlier proposed to cluster egg production curves at the phenotypic level by Savegnago et al. (2011) and milk yield profiles at the genetic level by Savegnago et al. (2016). This trajectory classification is proposed in our study as a complementary analysis to describe the group trajectories and better comprehend the animal profiles as compared with the selection objectives of a breeding

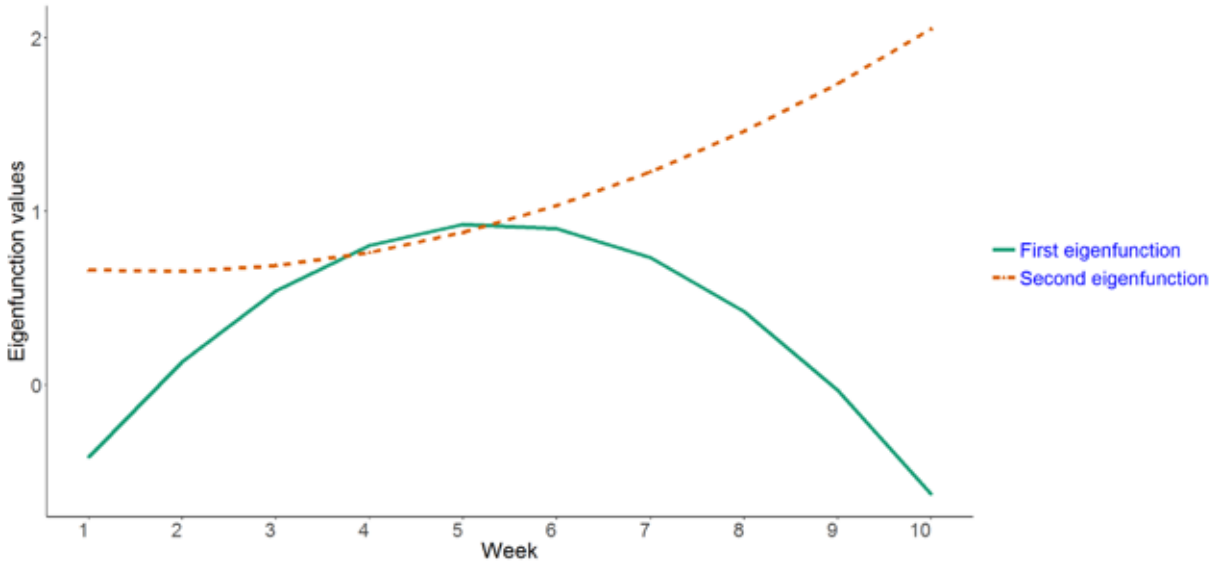


Figure 6. Two first eigenfunctions (unitless) associated with the covariance matrix **K** of the random regression model represented over the 10 wk of the test.

program. To summarize the EBV, we applied an eigendecomposition of the **G** genetic covariance matrices from the RR and SAD models, as originally applied to the **K** matrix of the RR models (Van Der Werf et al., 1998). The eigendecomposition has the advantage of accounting for the genetic covariances between weeks, which is not the case when using the average of the weekly EBV. By extracting the main axes of covariability among the EBV along time, 1 or 2 eigencomponents usually capture almost all the additive genetic variation in level and shape of the genetic curve, at least when applied to lactation curves (Druet et al., 2005; Togashi and Lin, 2006). Our results show that the first 2 SBV obtained from the eigendecomposition of the **G** matrix of the SAD model provided information similar to that of the eigendecomposition of the **K** matrix from the RR model. These SBV could, therefore, be similarly interpreted based on the eigenfunctions from the **K** matrix or the trajectory classification applied to the weekly EBV. As a result, combinations of the 2 first SBV are sufficient to describe the 3 groups of trajectories. It suggests that animals within a trajectory share genetic features that drive the dynamics of their feed efficiency during growth.

Despite differences in the estimation of the genetic parameters between the 2 approaches, the selection results were very similar for the RR and SAD models and could be confirmed by computing correlations between different SBV and with cEBV. In practice, one of the SBV was related to the average level of FCR during the test period (SBV_SAD1 and SBV_RR2) and the other one was related to the slope of the curve over time (SBV_RR1 and SBV_SAD2). In spite of this high concordance between the 2 approaches, the first 2 eigenvectors according to $\mathbf{G}_{\mathbf{SAD}}$ explained only about 73% of the genetic variation, which is rather lower than for the RR model (90%).

As expected from earlier studies (Kirkpatrick et al., 1990; Meyer and Kirkpatrick, 2005), the use of the **K** matrix and the **G** matrix of the RR model to calculate SBV led to very similar results. In the present study, the first eigenfunction changed sign with time. This suggests that selection for this first component would have opposite effects for the intermediate period compared with the extreme periods (2 wk at the beginning and 2 wk at the end of the trajectory). The second eigenfunction increased with time and was always positive. This means that selection for SBV_RRK2 would lead to selection in the same direction for all the time points, with a higher weight at the end of the testing period in comparison with the beginning of the period. Due to the high correlation between the first SBV obtained from the coefficients covariance matrix **K** (**SBV_RRK1**) and SBV_RR1 or SBV_SAD2, it confirmed our interpretation of SBV_RR1 and SBV_SAD2 as indicators of the slope of the feed efficiency curve.

To summarize, SBV can be used for selection purposes. To fully evaluate their potential, the estimation of genetic correlations with other production traits would provide a better insight on the use of the trajectories for selection. Indeed, it can be assumed that animals from the A group (low average FCR but a regular increase over time) would show a different fat content at slaughter than animals from the C group of similar average FCR, so selection for different FCR trajectories would consolidate breeding objectives on carcass composition. Further comparison of responses to selection for the traits of the breeding objective using different indexes options (cEBV and two first SBVs associated to the two first eigenvector of the matrix G)

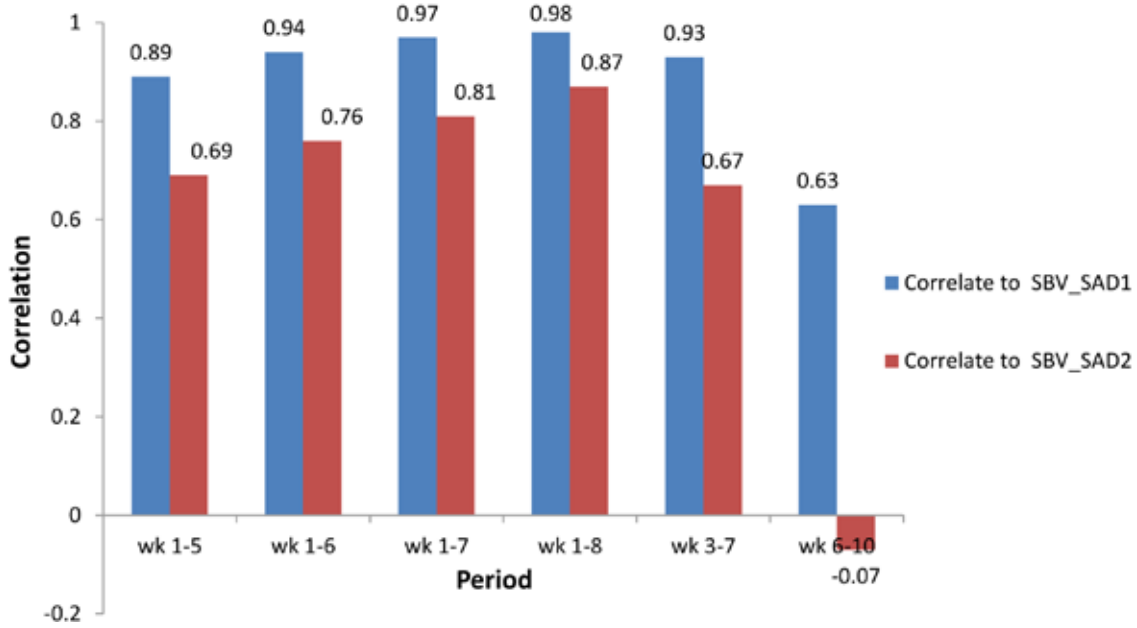


Figure 7. Correlations between the first (blue bars) and second (red bars) summarized breeding values obtained from the genetic covariance matrix **G** with the structured antedependence model (SBV_SAD1 and SBV_SAD2, respectively) for the full test period and those obtained from records from wk 1 to 5, wk 1 to 6, wk 1 to 7, wk 1 to 8, wk 3 to 7, and wk 6 to 10.

or a combination of two among them, would clarify the possible selection strategies.

Appropriate Period for Estimating Longitudinal Feed Conversion Ratio

The accuracy of the estimation of genetic parameters heavily relies on the quantity of data available. On the other hand, the cost of individual FI measures is high. Therefore, there is a trade-off between parsimony, complexity of the analysis, and potential bias, so choices need to be made. The goal is, therefore, to reduce the duration of the test period for FI with a minimum loss of accuracy for animal selection for FCR dynamic features, to test more pigs and increase the genetic gain (Begli et al., 2016). Wetten et al. (2012) proposed to use information on early periods of FI combined with information on growth to reduce the test period. In the current study, a similar conclusion was reached: the first weeks of test showed better correlations to selection criteria obtained with the whole test period than the middle and the last periods. The selection accuracy could be increased stepwise by extending the evaluation from 5 to 8 wk of duration. Further studies are required to better understand the link between the genetic gain, the costs associated with different strategies, and the changes in prediction accuracy due to a combined reduction of the duration of the test period and a greater number of pigs tested.

Conclusion

The current study indicates that the SAD model is promising for genetic selection: 1) it requires fewer parameters to fit the covariance matrices than the RR model and 2) it is not associated with the border effect problems and negative correlation estimates observed with the RR model. The use of SBV is a solution for animal selection applicable with the SAD model. The results of this study also suggest that a reduction of the duration of the FI test period to reduce measurement costs is probably feasible to select for feed efficiency.

LITERATURE CITED

Begli, H. E., R. V. Torshizi, A. A. Masoudi, A. Ehsani, and J. Jensen. 2016. Longitudinal analysis of body weight, feed intake and residual feed intake in F_2 chickens. Livest. Sci. 184:28–34. doi:10.1016/j.livsci.2015.11.018

Boligon, A. A., F. Baldi, M. E. Z. Mercadante, R. B. Lobo, R. J. Pereira, and L. G. Albuquerque. 2011. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. Genet. Mol. Res. 10:1227–1236. doi:10.4238/vol10-2gmr1087

Bunter, K. L., W. Cai, D. J. Johnston, and J. C. M. Dekkers. 2010. Selection to reduce residual feed intake in pigs produces a correlated response in juvenile insulin-like growth factor-I concentration. J. Anim. Sci. 88:1973–1981. doi:10.2527/jas.2009-2445

Buzanskas, M. E., R. P. Savegnago, D. A. Grossi, G. C. Venturini, S. A. Queiroz, L. O. C. Silva, R. A. A. Torres Júnior, D. P. Munari, and M. M. Alencar. 2013. Genetic parameter estimates and principal component analysis of breeding values of reproduction and growth traits in female Canchim cattle. Reprod. Fertil. Dev. 25:775–781. doi:10.1071/RD12132

David, I., H. Garreau, E. Balmisse, Y. Billon, and L. Canario. 2017. Multiple-trait structured antedependence model to study the relationship between litter size and birth weight in pigs and rabbits. Genet. Sel. Evol. 49:11. doi:10.1186/s12711-017-0288-3

David, I., J. Ruesche, L. Drouilhet, H. Garreau, and H. Gilbert. 2015. Genetic modeling of feed intake. J. Anim. Sci. 93:965–977. doi:10.2527/jas.2014-8507

Do, D. N., A. B. Strathe, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. J. Anim. Sci. 91:4069–4079. doi:10.2527/jas.2012-6197

Druet, T., F. Jaffrézic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37:257–271. doi:10.1186/1297-9686-37-4-257

Efron, B., and T. Hastie. 2016. The jackknife and bootstrap. In: Computer age statistical inference: Algorithms, evidence, and data science. Cambridge Univ. Press, Cambridge, UK. p. 155–180. doi:10.1017/CBO9781316576533.011

Englishby, T. M., G. Banos, K. L. Moore, M. P. Coffey, R. D. Evans, and D. P. Berry. 2016. Genetic analysis of carcass traits in beef cattle using random regression models. J. Anim. Sci. 94:1354–1364. doi:10.2527/jas.2015-0246

Fischer, T. M., A. R. Gilmour, and J. H. J. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Genet. Sel. Evol. 36:363–369. doi:10.1186/1297-9686-36-3-363

Genolini, C., X. Alacoque, M. Sentenac, and C. Arnaud. 2015. kml and kml3d: R packages to cluster longitudinal data. J. Stat. Softw. 65:1–34. doi:10.18637/jss.v065.i04

Gilbert, H., J.-P. Bidanel, J. Gruand, J.-C. Caritez, Y. Billon, P. Guillouet, H. Lagant, J. Noblet, and P. Sellier. 2007. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J. Anim. Sci. 85:3182–3188. doi:10.2527/jas.2006-590

Gilbert, H., Y. Billon, L. Brossard, J. Faure, P. Gatellier, F. Gondret, E. Labussière, B. Lebret, L. Lefaucheur, N. L. Floch, I. Louveau, E. Merlot, M.-C. Meunier-Salaün, L. Montagne, P. Mormede, D. Renaudeau, J. Riquet, C. Rogel-Gaillard, J. van Milgen, A. Vincent, and J. Noblet. 2017. Review: Divergent selection for residual feed intake in the growing pig. Animal 11(9):1427–1439. doi:10.1017/S175173111600286X

Gilmour, A. R., B. J. Gogel, and B. R. Cullis. 2009. ASReml user guide release 3.0. VSN Int. Ltd., Hemel Hempstead, UK.

Henryon, M., A. Jokumsen, P. Berg, I. Lund, P. B. Pedersen, N. J. Olesen, and W. J. Slierendrecht. 2002. Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within a farmed population of rainbow trout. Aquaculture 209:59–76. doi:10.1016/S0044-8486(01)00729-3

Huynh-Tran, V. H., H. Gilbert, and I. David. 2017. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J. Anim. Sci. 95:39–48. doi:10.2527/jas.2016.0980

Jaffrézic, F., and S. D. Pletcher. 2000. Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. Genetics 156:913–922.

Jaffrézic, F., E. Venot, D. Laloë, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves. J. Anim. Sci. 82:3465–3473. doi:10.2527/2004.82123465x

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993.

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37:473. doi:10.1186/1297-9686-37-6-473

Meyer, K., and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. Livest. Prod. Sci. 47:185–200. doi:10.1016/S0301-6226(96)01414-5

Meyer, K., and M. Kirkpatrick. 2005. Up hill, down dale: Quantitative genetics of curvaceous traits. Philos. Trans. R. Soc., B 360:1443–1455. doi:10.1098/rstb.2005.1681

Patience, J. F., M. C. Rossoni-Serão, and N. A. Gutiérrez. 2015. A review of feed efficiency in swine: Biology and application. J. Anim. Sci. Biotechnol. 6:33. doi:10.1186/s40104-015-0031-2

Saintilan, R., I. Mérour, L. Brossard, T. Tribout, J. Y. Dourmad, P. Sellier, J. Bidanel, J. van Milgen, and H. Gilbert. 2013. Genetics of residual feed intake in growing pigs: Relationships with production traits, and nitrogen and phosphorus excretion traits. J. Anim. Sci. 91:2542–2554. doi:10.2527/jas.2012-5687

Saintilan, R., P. Sellier, Y. Billon, and H. Gilbert. 2012. Genetic correlations between males, females and castrates for residual feed intake, feed conversion ratio, growth rate and carcass composition traits in Large White growing pigs. J. Anim. Breed. Genet. 129:103–106. doi:10.1111/j.1439-0388.2011.00972.x

Savegnago, R. P., S. L. Caetano, S. B. Ramos, G. B. Nascimento, G. S. Schmidt, M. C. Ledur, and D. P. Munari. 2011. Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. Poult. Sci. 90:2174–2188. doi:10.3382/ps.2011-01474

Savegnago, R. P., G. B. do Nascimento, G. J. de Magalhães Rosa, R. L. R. de Carneiro, R. C. Sesana, L. El Faro, and D. P. Munari. 2016. Cluster analyses to explore the genetic curve pattern for milk yield of Holstein. Livest. Sci. 183:28–32. doi:10.1016/j.livsci.2015.11.010

Schnyder, U., A. Hofer, F. Labroue, and N. Künzi. 2001. Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs. Genet. Sel. Evol. 33:635–658. doi:10.1186/1297-9686-33-6-635

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464. doi:10.1214/aos/1176344136

Shirali, M., C.-A. Duthie, A. Doeschl-Wilson, P. W. Knap, E. Kanis, J. A. van Arendonk, and R. Roehe. 2013. Novel insight into the genomic architecture of feed and nitrogen efficiency measured by residual energy intake and nitrogen excretion in growing pigs. BMC Genet. 14:121. doi:10.1186/1471-2156-14-121

Togashi, K., and C. Y. Lin. 2006. Selection for milk production and persistency using eigenvectors of the random regression coefficient matrix. J. Dairy Sci. 89:4866–4873. doi:10.3168/jds.S0022-0302(06)72535-8

Van Der Werf, J. H., M. E. Goddard, and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81:3300–3308. doi:10.3168/jds.S0022-0302(98)75895-3

Wang, W. 2013. Identifiability of linear mixed effects models. Electron. J. Stat. 7:244–263. doi:10.1214/13-EJS770

Wetten, M., J. Ødegård, O. Vangen, and T. H. E. Meuwissen. 2012. Simultaneous estimation of daily weight and feed intake curves for growing pigs by random regression. Animal 6:433–439. doi:10.1017/S1751731111001832

3.3. Application of the RR-SL model to longitudinal FCR

As stated in the discussion of the previous article, we tested different options to reduce the border effect problem observed with RR-OP. The first easy one was to consider heterogeneous residual variances with time. Such modification of the residual variance did not solve the border effect problem (fig 3.2) of the variance component estimations.

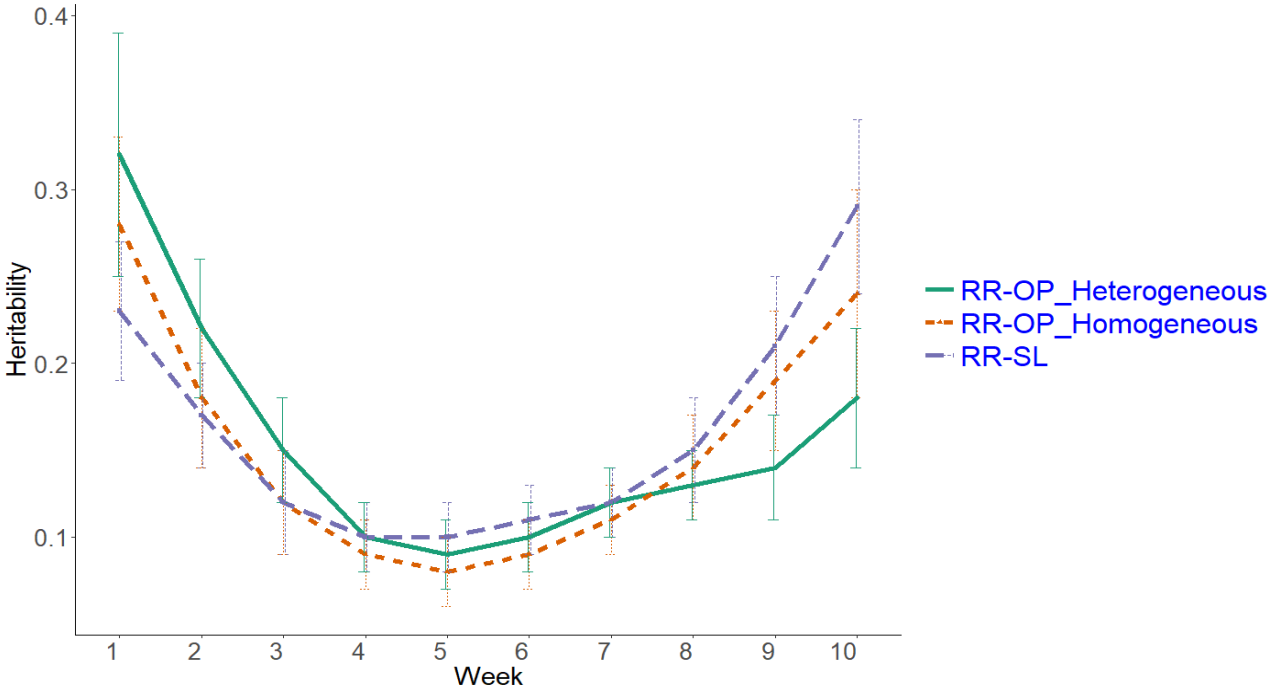


Figure 3.2. Heritability estimates of random regression using Legendre orthogonal polynomials with homogeneous residual variances (RR-OP_Homogenenous) and heterogeneous residual variances (RR-OP_Hetegenenous), and with random regression using splines (RR-SL). The green, blue and orange vertical bars are the standard error (SE) of the heritability estimates.

Another solution tested to reduce the border effect problem was to use splines instead of orthogonal polynomials in the random regression models.

3.3.1. Material and methods

The model described in Eq.1 in the previous article was applied to longitudinal FCR of male pigs with a random regression model using natural cubic splines (RR-SL) to account for the covariance structure of the data.

To select the number and positions of the knot points, we first increased the number of knots evenly distributed, starting with 2, until the loglikelihood of the model did not increase

anymore. Then the position of the knots was changed iteratively to the neighbouring weeks, and the model with the maximal likelihood was retained. It was a RR-SL with 3 knots at weeks 2, 5 and 8 (9 parameters).

3.3.2. Results and discussion

The BIC of the RR-SL model was -9860, lower than the one of the RR-OP model, but higher than the SAD model. The pattern of heritability over time obtained with the RR-SL model was similar to the one obtained with the RR-OP model (Fig. 3.2). Once again, we observed an increase of the genetic and permanent effect variances at the beginning and at the end of the test period. The genetic correlations obtained with the RR-SL model (Fig. 3.3) were in line with those of the RR-OP model.

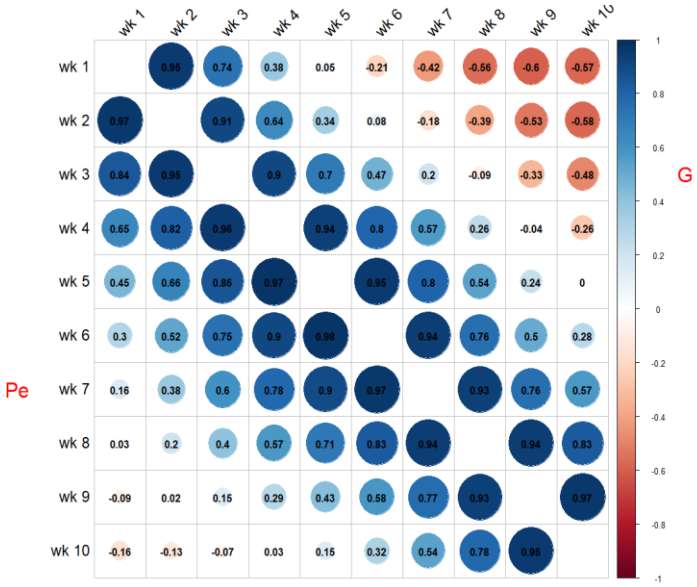


Figure 3.3. Genetic correlations (above the diagonal) and permanent environmental correlations (below the diagonal) between times, estimated with the random regression model using natural cubic splines. The magnitude and sign of the correlations are indicated with darker and larger circles and blue (positive) or red (negative) colors, respectively.

These results confirmed that the SAD model provided the better fit-to-the-data for FCR. Conversely to the expectation, applying a RR-SL model did not reduce the border effect problem observed with the RR-OP model. The discrepancy between the heritability and genetic correlations obtained with the SAD and the RR models is not satisfactory. To get a clearer idea about the true heritability and the true genetic correlations, we wanted to apply a multiple trait model to the data. The advantage of such model is to make no assumption about

the form of the covariance structure of the data. However, since the number of time points was large, the MT model never converged.

3.4. Predictive ability one week ahead

In order to help real time management of animals and to reduce waste of feed, it would be interesting to anticipate the feed conversion ratio of the animals. To do so, we compared the phenotypic predictive ability of the different models one week ahead. To perform this comparison, we predicted for the animals of the last generation (164 animals) the observations of the next week given their observations of previous weeks and all the observations for the other animals from previous generations. The predictive ability for each week (5 to 9) was estimated using the Vonesh concordance coefficient (VCC) for each model as defined in chapter 1, using data from week 1 to *w* in the last generation to predict weeks *w*+1, with *w* = 4,…, 8.

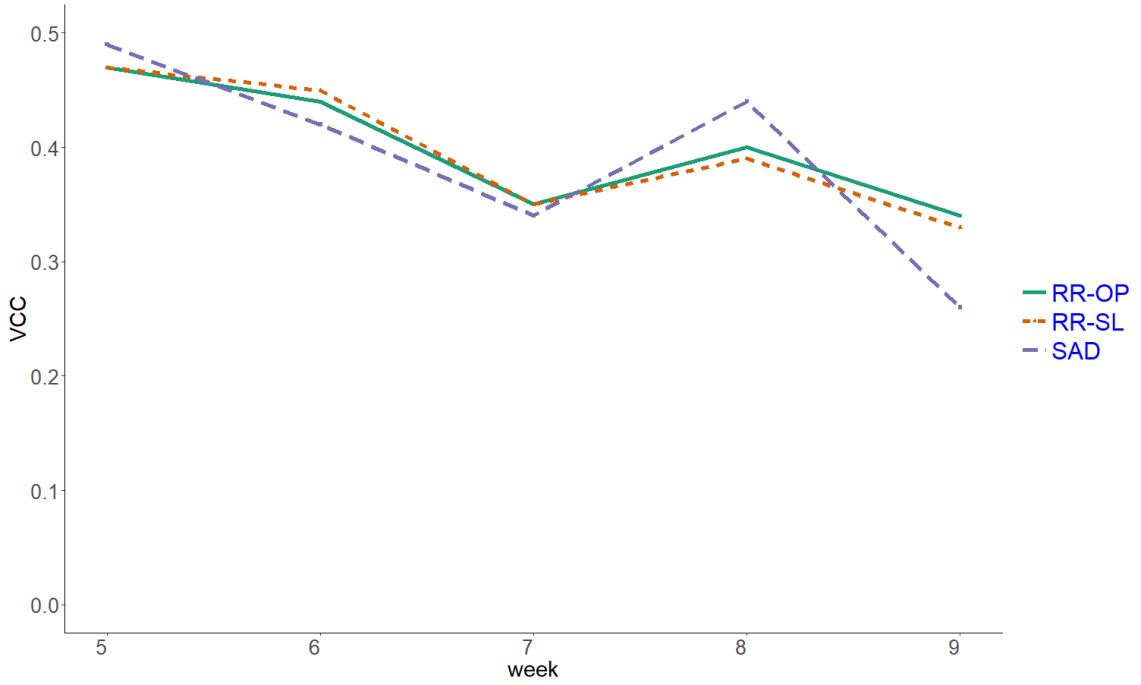


Figure 3.4. Average Vonesh concordance coefficients for different models computed for periods from week 5 to week 9. RR-OP = random regression using orthogonal polynomial; SAD = structured antedependence model, RR-SL = random regression using splines with 3 knots at weeks 2, 5 and 8.

The number of weekly FCR records of the last generation ranged from 115 to 164 depending on the week. Similar patterns of VCC were obtained for all models. The VCC decreased from

week 5 to week 7, slightly increased in week 8, and decreased again in week 9 where the number of records available was the lowest.

The results showed that the predictive ability of SAD is in line with those of RR-OP and RR-SL. To summarize the results, we calculated the average VCC for weeks *5* to 9 as:

$$\overline{VCC} = \frac{\sum_{j=5}^{j=9} n_j * VCC_j}{\sum_{j=5}^{j=9} n_j}$$

The mean $\overline{VCC}$ was 0.39 for all models. Altogether, it indicates a low concordance between the predicted and observed phenotypes for all models. With this situation additional information would be necessary to help real time management of the animals.

3.5. Conclusion

In this chapter,

1. We performed the comparison of different models to analyze longitudinal FCR. Among them, the SAD model showed advantages in terms of number of parameters and control of the border effects which affect the RR models.
2. The structured antedependence model showed a similar predictive ability as RR-OP and RR-SL models.
3. We also proposed an approach for summarizing the longitudinal EBV into few values, that seems promising for selection.

Altogether, this chapter provides new material for animal selection based on animal's EBV profile for feed efficiency. To explicit the potential of genomic information for such longitudinal analyses, in the following chapters we exploited the genomic information in a attempt to improve the genetic prediction (chapter 4) and figure out the impact of the divergent selection for RFI at the pedigree and genomic level (chapter 5).

Chapter 4 - PEDIGREE AND GENOMIC PREDICTIONS OF LONGITUDINAL DATA FOR RESIDUAL FEED INTAKE AND AVERAGE DAILY GAIN IN GROWING PIGS

4.1. Introduction: use of genomics for longitudinal feed efficiency

In the previous chapters, the longitudinal analyses have shown their advantages for genetic evaluation based on pedigree information. In this chapter, we investigate longitudinal models (random regression, structured antedependence and multiple trait models) applied to residual feed intake. This approach using the pedigree relationship matrix will be compared with the "single step" approach, in which both pedigree and genomic information are combined into an **H** relationship matrix.

This work has been developed in collaboration between INRA, France, and Iowa State University, USA. In the frame of the "EIR-A" program (Ecole internationale de la recherche d'Agreenium", I spent three months (from Nov 2017 to Feb 2018) at the Animal Science Department, Iowa State University, USA, under the supervision of Pr. J.C.M Dekkers. This work is presented in the form of a paper in preparation to be submitted to *Genetics Selection Evolution*.

Pedigree and genomic predictions of longitudinal data for residual feed intake and average daily gain in growing pigs

V.H. Huynh-Tran*[1], J.C.M. Dekkers†, I. David*, H. Gilbert*

*GenPhySE, INRA, Université de Toulouse, INPT, ENVT, 31326 Castanet-Tolosan cedex, France

†Department of Animal Science, Iowa State University, Ames, IA 50011, USA

[1]Corresponding author: van-hung.tran@inra.fr

Abstract

The objectives of this study were i) first, to compare three models to account for longitudinal measurements: random regression (RR), structured antedependence (SAD), multi-trait (MT) models for genetic parameters and breeding value (EBV) estimations of two longitudinal traits: average daily gain (ADG) and residual feed intake (RFI), measured each week from 12 to 25 weeks of age; ii) second, to compare a usual pedigree-based animal mixed model (Best Linear Unbiased Prediction, BLUP) with a single step genomic BLUP model (HBLUP) in terms of accuracy and bias of pedigree and genomic predictions for ADG and RFI. Data from a population of 2345 growing Large White pigs from two lines divergently selected for RFI over eight generations (G0 to G7) were used for the analysis. The SAD model best fitted the data for both traits and used the lowest number of parameters (8 and 12 parameters for ADG and RFI, respectively). The heritability estimates for all models ranged from 0.07 to 0.49 for ADG, and from 0.06 to 0.35 for RFI, depending on the week of test and model. For both traits, genetic correlations between weeks were high (≥ 0.58) when weeks were separated by one week, and close to zero between extreme weeks. Negative genetic correlations between the first and last weeks of test were obtained with RR and MT models. Prediction accuracy of the three models were computed using a cross-validation approach for two scenarios in which the oldest animals (G0 to G4) were used to train the model and predict the genetic merit of the youngest animals (G5 to G7) in the high RFI and the low RFI lines separately. The prediction accuracies were low for both traits (≤0.42 for ADG and ≤0.32 for RFI) and all models and were not systematically improved by the use of genomic information. However, better accuracies were obtained for ADG than for RFI. We conclude that the greater heritability, and possibility the limited response to selection of ADG in these lines compared to RFI, could explain some of the poor prediction accuracies for RFI.

Key words: genetic, genomic prediction, residual feed intake, average daily gain, pigs

INTRODUCTION

Feed costs account for about two third of the pork production costs. To assess the productivity and feed efficiency of a pig system, average daily gain (ADG) and residual feed intake (RFI) are widely used [1]. However, genetic evaluations on ADG and RFI are mostly carried out on the average of the trait over a given period. Most studies assume that the genetic parameters such as genetic variances, and subsequently heritabilities, are constant over time, even if it is known that the genetic potential of the animal changes during the production period for these traits [2]. Over the last years, with the proliferation of monitoring systems such as self-feeders, electronic identification and molecular biotechnologies, the repetition over time of performance records (e.g. feed intake and body weight measurements) on one hand, and genome-wide single nucleotide polymorphisms (SNP) marker genotypes on the other hand, have become available for a large number of farm animals. To estimate genetic parameters, the Best Linear Unbiased Prediction (BLUP) approach using a pedigree relationship matrix to quantify the additive genetic effects is commonly used since 1980: it is simple and has low computational requirements. Recently, a GBLUP approach has been proposed [3], that includes low to medium density marker genotypes in a genomic relationship information, accounting not only for recombination events in the genotyped pedigree (i.e., linkage analysis) but also for the population linkage disequilibrium pattern in the genome, i.e., the possibility of predicting alleles at some loci on the basis of alleles at other (possibly close) loci [4]. The accuracies of genomic predictions can be improved by combining genomic information and information from traditional pedigree in a single-step BLUP approach or HBLUP approach [5–7] to combine informations from genotyped and non-genotyped animals.

Even if longitudinal models have been used to estimate genetic parameters over the last two decades, few studies evaluated genomic prediction combined with longitudinal models [8,9]. Recently, we investigated the structured antedependence model for genetic evaluation and found that it has several advantages in comparison with the random regression model [10] when applied to feed conversion ratio. In this paper, we will compare the accuracy and bias of genetic and genomic longitudinal models for the predictions of ADG and RFI.

MATERIALS AND METHODS

Data were collected in accordance with the national regulations of animal care in agriculture in France.

Animals and phenotypes

Data consisted of body weight (BW), daily feed intake (DFI) and backfat thickness (BFT) records of 2435 growing French Large White pigs (1186 males, 580 females, and 669 castrated males). These animals were from a divergent selection experiment for RFI: one line was selected for low RFI (LRFI, more efficient) and the other line was selected for high RFI (HRFI, less efficient). The development of the selected lines and performance testing procedures were described in Gilbert et al. (2007)[11].

In brief, animals born in a given farrowing batch were gathered at weaning (28 days of age) in a single post-weaning unit (UE GenESI, Rouillé, France). At 10 weeks of age, about 48 pigs were moved to a growing-finishing room with four pens per batch equipped with single-place electronic feeders (ACEMA 64, Pontivy, France). Twelve animals of same sex were allotted to each pen and had *ad libitum* access to a pelleted diet based on cereals and soybean meal containing 10 MJ NE/kg and 160 g CP/kg, with a minimum of 0.80 g digestible Lys/MJ NE. The BW and age at the beginning of the test averaged 25±4 kg and 67±1 days, respectively. The average BW and age at the end of the test were 115±11 kg and 168±13 days. All pigs were allowed to acclimate to the feeders for about a week, so the records of the first week of the test period were removed from the data set. During 16 consecutive weeks (from the 2nd week to the 17th week of the test period), males were weighed weekly, and the majority of females and castrated males were weighed monthly.

Weekly averages of daily feed intake (FI) were computed for each animal from the records of the feed intake visits to the electronic feeders. The outliers and inaccurate values (more than two days of records missing in a given week) were removed, as reported in [12]. The ultrasonic

backfat thickness (BFT) was the average of 3 measurements on both sides of the spine at the level of the neck, the back and the kidneys of the animal. For the males, BFT were measured at around 35 kg, 65 kg and 90 and 95 kg, whereas, those of females and castrated males were measured at weeks 11, 15, 19 and 23.

Weekly missing records for BW and BFT for week *j* were predicted by linear interpolations, as proposed for BW by Huynh-Tran et al. (2017) [13]. Data from the 2nd week to the 15th week of test were used to compute longitudinal ADG for week 4 to week 13, as reported in [13]. The average metabolic body weight for each week (AMBW, [14]) was similarly calculated over a 4-week period to limit the influence of inaccurate BW measurements. Thus, ten weekly records of each trait were available per animal. The values of ADG, BFT, AMBW and FI that deviated by more than 3 standard deviations (SD) from their respective means were removed from the analysis. The final dataset comprised 22250 records from 2435 animals. A total of 3986 animals was included in the pedigree file.

Genomic markers

The Illumina SNP60 Beadchip V2 (Illumina Inc., San Diego, CA) and the GeneSeek Genomic Profiler HD were used for genotyping all sires from generations G0 to G7, and dams from generations G0 to G6 (660 pigs). Only the 42780 SNP that overlapped between the two panels were used. The SNP with minor allele frequencies lower than 5%, with missing position or a call rate lower than 95% were removed from the analyses, as were excluded individuals with more than 4 % genotype inconsistencies with their parents or progeny. After this quality control, 39,557 SNPs remained for further analyses. Missing genotypes in each SNP chip panel were then imputed on the animals genotyped with the alternate panel using the pedigree information with FImpute [15]. Altogether, 64487 SNPs genotyped or imputed on all pigs were available for further analyses.

Data analyses

General model for longitudinal data analyses. Three models were used: the structured antedependence (SAD), random regression with Legendre orthogonal polynomials (RR), and multiple trait (MT) models, with either a pedigree relationship matrix or a so-called **H** matrix combining pedigree and genomic relationships. The general form of the model is the same for all cases, for animal *i* at time t_j:

$$y_i(t_j) = \mu_i(t_j) + u_i(t_j) + p_i(t_j) + \varepsilon_i(t_j) \quad \text{(Eq.1)}$$

Where $y_i(t_j)$ is the phenotype of animal *i* at time t_j, $\mu_i(t_j)$ is the vector of fixed effects at time t_j, $u_i(t_j)$ and $p_i(t_j)$ are the random genetic and permanent environmental effects for animal *i* at time t_j, with: $\boldsymbol{p} \sim N(0, \boldsymbol{P} \otimes \boldsymbol{I})$, $\boldsymbol{u} \sim N(0, \boldsymbol{J} \otimes \Sigma)$ and $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{D} \otimes \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix of appropriate size, $\boldsymbol{P}$ the 10 x 10 covariance matrix of the weekly permanent effects, and $\boldsymbol{D}$ is the 10 x 10 diagonal matrix of the residual effects across weeks. $\boldsymbol{J}$ is the 10 x 10 covariance matrix of the weekly genetic effects; Σ is **A** , the pedigree-known relationship matrix for the genetic evaluations, and **H,** the matrix combining genomic and pedigree informations, for the genomic models as proposed by Legarra et al. (2009) [5], Aguilar et al (2010) [16] and Christensen and Lund (2010) [7]: $\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}_w^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$, with $\mathbf{A}_{22}$ the pedigree-based numerator relationship matrix for the genotyped animals (660 animals), $\mathbf{G}_w = (1-\omega)\mathbf{G}^* + \omega\mathbf{A}_{22}$, in which ω is the proportion of genetic variance not captured by the markers. According to Christensen (2012) [1], the change of $(1-\omega)$ has a small effect on the accuracy of the breeding value predictions. Therefore, in this study, we used ω = 0.05, as mostly used in the literature. The matrix **G*** was obtained by scaling the genomic relationship matrix **G** [2], to equal the means of the diagonal and off-diagonal elements of the **G** and $\mathbf{A}_{22}$ matrices. The SAD, RR and MT models differed for the covariance matrices **P**, **J** and **D.**

MT covariance structures. By definition, the MT model has unstructured covariance matrices for **P** and **J**. Nonetheless, since such assumption requires a large number of parameters to estimate, we considered that ADG (FI) was the same trait for groups of two to three consecutive weeks. Four periods were then defined: weeks 1 to 3, weeks 4 to 5, weeks 6 to 7, weeks 8 to 10, leading to the following constrained **P**, **J** and **D** matrices:

$$\mathbf{P}=\begin{bmatrix}
\sigma^2_{p1} & & & & & & & & & \\
\sigma^2_{p1} & \sigma^2_{p1} & & & & & & & & \\
\sigma^2_{p1} & \sigma^2_{p1} & \sigma^2_{p1} & & & & & & & \\
\sigma_{p_1p_2} & \sigma_{p_1p_2} & \sigma_{p_1p_2} & \sigma^2_{p2} & & & & & & \\
\sigma_{p_1p_2} & \sigma_{p_1p_2} & \sigma_{p_1p_2} & \sigma^2_{p2} & \sigma^2_{p2} & & & & & \\
\sigma_{p_1p_3} & \sigma_{p_1p_3} & \sigma_{p_1p_3} & \sigma_{p_2p_3} & \sigma_{p_2p_3} & \sigma^2_{p3} & & & & \\
\sigma_{p_1p_3} & \sigma_{p_1p_3} & \sigma_{p_1p_3} & \sigma_{p_2p_3} & \sigma_{p_2p_3} & \sigma^2_{p3} & \sigma^2_{p3} & & & \\
\sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_2p_4} & \sigma_{p_2p_4} & \sigma_{p_3p_4} & \sigma_{p_3p_4} & \sigma^2_{p4} & & \\
\sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_2p_4} & \sigma_{p_2p_4} & \sigma_{p_3p_4} & \sigma_{p_3p_4} & \sigma^2_{p4} & \sigma^2_{p4} & \\
\sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_1p_4} & \sigma_{p_2p_4} & \sigma_{p_2p_4} & \sigma_{p_3p_4} & \sigma_{p_3p_4} & \sigma^2_{p4} & \sigma^2_{p4} & \sigma^2_{p4}
\end{bmatrix}$$

$$\mathbf{J}=\begin{bmatrix}
\sigma^2_{u1} & & & & & & & & & \\
\sigma^2_{u1} & \sigma^2_{u1} & & & & & & & & \\
\sigma^2_{u1} & \sigma^2_{u1} & \sigma^2_{u1} & & & & & & & \\
\sigma_{u_1u_2} & \sigma_{u_1u_2} & \sigma_{u_1u_2} & \sigma^2_{u2} & & & & & & \\
\sigma_{u_1u_2} & \sigma_{u_1u_2} & \sigma_{u_1u_2} & \sigma^2_{u2} & \sigma^2_{u2} & & & & & \\
\sigma_{u_1u_3} & \sigma_{u_1u_3} & \sigma_{u_1u_3} & \sigma_{u_2u_3} & \sigma_{u_2u_3} & \sigma^2_{u3} & & & & \\
\sigma_{u_1u_3} & \sigma_{u_1u_3} & \sigma_{u_1u_3} & \sigma_{u_2u_3} & \sigma_{u_2u_3} & \sigma^2_{u3} & \sigma^2_{u3} & & & \\
\sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_2u_4} & \sigma_{u_2u_4} & \sigma_{u_3u_4} & \sigma_{u_3u_4} & \sigma^2_{u4} & & \\
\sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_2u_4} & \sigma_{u_2u_4} & \sigma_{u_3u_4} & \sigma_{u_3u_4} & \sigma^2_{u4} & \sigma^2_{u4} & \\
\sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_1u_4} & \sigma_{u_2u_4} & \sigma_{u_2u_4} & \sigma_{u_3u_4} & \sigma_{u_3u_4} & \sigma^2_{u4} & \sigma^2_{u4} & \sigma^2_{u4}
\end{bmatrix};\quad
\mathbf{D}=\begin{bmatrix}
\sigma^2_{\varepsilon_1} & & & & & & & & & \\
0 & \sigma^2_{\varepsilon_2} & & & & & & & & \\
0 & 0 & \sigma^2_{\varepsilon_3} & & & & & & & \\
0 & 0 & 0 & \sigma^2_{\varepsilon_4} & & & & & & \\
0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_5} & & & & & \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_6} & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_7} & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_8} & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_9} & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_{10}}
\end{bmatrix}$$

In total, 21 parameters were thus necessary to estimate with the MT model.

RR and SAD covariance structures. The RR and SAD approaches reduced the number of parameters to estimate in Eq 1 by modelling the form of the random effects functions.

In the RR model, for a given random effect $u_i(t_j)$, the general form of the random function of order *s* is: $u_i(t_j)=\sum_{k=0}^{s} a_{ik}\varphi_k(t_j)$, where a_{ik} is the $(k+1)^{\text{th}}$ random regression coefficient for the genetic effects for animal *i*, with $\mathbf{a}\sim N(0,\mathbf{K}\otimes\Sigma)$ where **K** is the covariance matrix of the

additive random regression coefficients, and $\varphi_k\left(t_j\right)$ is the $(k+1)^{th}$ Legendre polynomial at time t_j. In this model, the relationships between **K** and **J** are given by $\boldsymbol{J_{RR} = \varphi K \varphi'}$, where $\boldsymbol{\varphi}$ is the (n x (s+1), n being the number of time points) matrix of the Legendre polynomials for all time points.

In the SAD model, the random effect function consists in a regression on preceding observations. The function for the random effect u is defined as: $u_i\left(t_j\right) = \sum_{w=1}^{\lambda} \psi_{wj} u_i\left(t_{j-w}\right) + e_i\left(t_j\right)$, where λ is the order of the antedependence, ψ_{wj} is the w^{th} antedependence parameter for time t_j, and $e_i\left(t_j\right)$ is the error term for animal i at time t_j $\mathbf{e}\left(t_j\right) \sim N\left(0, \boldsymbol{\Sigma}\sigma_e^2\left(t_j\right)\right)$. To reduce the number of parameters to estimate, the antedependence parameters and the variance of the error term were considered as functions of time: $\psi_{wj} = \sum_{q=0}^{\beta_w} \xi_{wq} t_j^q$, $\sigma_e^2(t_j) = exp\left(\sum_{q=0}^{\gamma} \omega_q t_j^q\right)$. The SAD model for each random term is thus defined by three parameters: the antedependence order, the degree of the polynomial for each antedependence parameter (β_1 to β_ω), and the degree of the polynomial for the innovation variance (γ).We noted a SAD model with a given set of parameters as follows: SADλ-β_1,…,β_ω γ. To ease convergence and avoid identifiability problems between the structured random permanent environmental effects and the residual covariance matrices [17], the SAD model in our study merged the residual with the permanent environmental effects.

All models were applied to the data using the REML approach in ASReml 3.0 [19]. The best models within the SAD and RR approaches were first selected by comparing nested genetic models with the pedigree relationship matrix using likelihood ratio tests (LRT). In addition, Bayesian Information Criteria (BIC) [18] were computed to compare the different models: BIC = -2ln(L) + c*ln(N-p), where L is the restricted maximum likelihood of the model, N the number of observations, and p and c the number of fixed effects and covariance parameters, respectively. The model (RR, SAD or MT) with the lowest BIC was considered as the best fit to the data. The same structures of models were used when the genomic information was further included in the analyses.

Variance components estimations. Estimates for RFI were obtained from the genetic and permanent effects of animal models applied to weekly feed intakes including time specific

ADG, AMBW, BFT as covariates, and the same fixed effects as for the model of ADG. The fixed effects included in these models were the herd, week, pen, age at the beginning of test and interaction between week and generation, and sex and batch.

Variance components were obtained with the ASReml software [19]. Standard errors of the heritabilities were computed using a standard Taylor series approximation built in ASReml with RR and MT models, and using bootstraps for the SAD estimations [10]. The genomic models required the computation of the **H** matrix. It was obtained and formated using PreGSf90 [20] and modified with R to be supplied to ASReml as a user defined relationship matrix.

Prediction accuracy and biases

To evaluate the models ability to predict breeding values, the predictive abilities of the different approaches were investigated using cross-validation. From the initial database (2435 animals), two datasets were distinguished: a training dataset (n=1169, including 469 genotyped individuals), including animals of the first five generations of selection (G0 to G4) , and two validation datasets consisting of animals of the last three generations (G5 to G7), each line being treated separately: the HRFI line (n_1=99, scenario 1) and LRFI line (n_2=93, scenario 2).

First, the models were run on the full data to obtain the corrected phenotypes (y_c) of the genotyped individuals. Because of the limited number of genotyped animals in the validation dataset having phenotypic records (only sires and not dams), corrected phenotypes of all genotyped individuals were obtained as the mean of their progeny corrected phenotypes, ignoring their own performances when available. The average number of progeny per sire or dam were 18 for LRFI pigs and 20 for HRFI pigs in the validation datasets.

Second, the same models were applied to the reduced training dataset (full dataset without phenotypes for the individuals of the validation dataset). The estimated breeding values (EBV) predicted with the pedigree matrix (EBVp) and with the genomic matrix (GEBVp) on the validation dataset were used to compute the predictive ability of each model. Two criteria were used: the accuracy and the bias of the predictions. For each week, the prediction accuracy was defined as the weighted correlation between the EBVp or GEBVp and the corrected phenotypes y_c, divided by the square root of the corresponding weekly heritability estimated from the full dataset [21]. The individual weights to compute the weighed correlations were the number of progeny. The bias corresponded to the regression coefficient of y_c on EBVp or GEBVp. Since EBVp (respectively GEBVp) was the breeding value of the parent, and y_c the average corrected phenotypes of its progeny, the expected bias was 0.5.

The EBVp and GEBVp of SAD and MT were directly obtained from the outputs of ASReml, while EBVp of RR models were computed using the estimations of the individual regression coefficients provided by ASReml as $(G)EBV\_RR_i\left(t_j\right)=\sum_{k=0}^{s}\hat{a}_{ik}\varphi_k(t_j)$.

Finally, to compare the longitudinal animal models with the classical animal models as usually applied on a single record per animal, animal models with the **A** or **H** relationship matrix were applied to FI and ADG computed over the test period. The full test RFI was given by:

$$FI_i=\mu_i+ADG_i+AMBW_i+BF_i+u_{RFI_i}+p_{RFI_i}+\varepsilon_{RFI_i}$$

In which FI_i = average daily feed intake over the test period, ADG_i and $AMBW_i$ are the ADG and AMBW of animal *i* for whole test period computed as:

$$ADG_i=\frac{BW_{i\_end}-BW_{i\_begin}}{age_{i\_end}-age_{i\_begin}}\ ,\ AMBW_i=\frac{BW_{i\_end}^{1.6}-BW_{i\_begin}^{1.6}}{1.6\times\left(BW_{i\_end}-BW_{i\_begin}\right)}$$, and BF_i is the backfat thickness of animal *i* at the end of test (week 12).

RESULTS

Goodness of fit

The number of parameters and BIC of the best model for each category are presented in Table 1. With ADG, the best SAD model was a SAD111/111 for the genetic and permanent environmental effects, respectively, with 8 parameters. The best RR model was of degree 2 for both genetic and environmental variances (13 parameters), when the multiple-trait model for the four periods had 21 parameters. The lowest BIC was obtained for the SAD model.

For RFI, the SAD122/122 model and the RR model of degree 2 for both the genetic and permanent environment effects were kept. The BIC of the three models ranged from 5984 for SAD (with 12 parameters) to 7232 for MT (with 21 parameters).

Table 4.1. Log Likelihood (LogL), number of parameters and Bayesian Information Criterion (BIC) of best models for each category: random regression (RR), structured antedependence (SAD) and multi trait (MT) models for average daily gain (ADG) and residual feed intake (RFI).

Trait	Best category model[1]	LogL	Number of parameters	BIC
ADG	RR (2/2)	-9919	13	19968
	SAD (111/111)	-8594	8	17268
	MT	-11267	21	22744
RFI	RR (2/2)	-3403	13	6936
	SAD (122/122)	-2932	12	5984
	MT	-3511	21	7232

[1]RR (2/2) = RR model with Legendre polynomials of degree 2 for genetic random effect and permanent environmental random effects; SAD (111/111) = SAD model with degree 1 for antedependence, antedependence parameters and error variances for both genetic and permanent environmental random effects

Computation of the H matrix

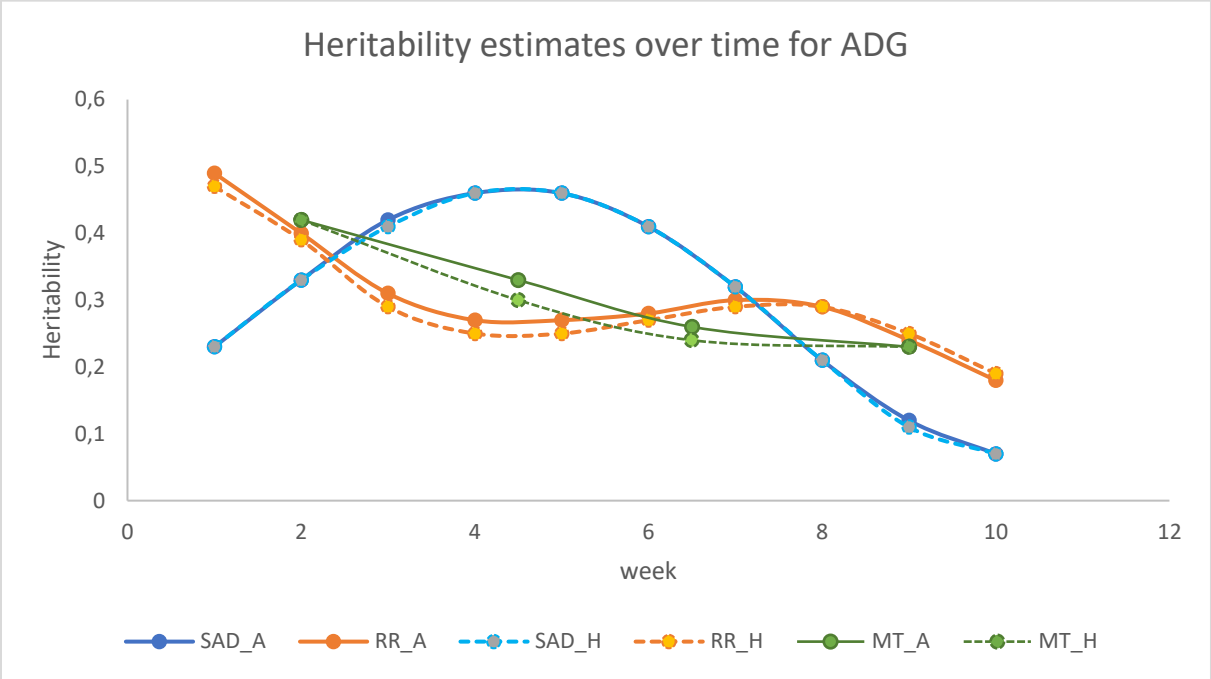
The average of the diagonal elements and off-diagonal elements of matrices $\mathbf{A}_{22}$ and **G** were 1.033 and 0.072, respectively. The correlations between elements of matrices $\mathbf{A}_{22}$ and **G** were 0.90 for all elements, and 0.91 for off-diagonal elements. Correlations between EBV from SAD and RR were 0.95 and 0.96 for RFI and ADG, respectively. Correlations between SAD and RR GEBV were 0.93 and 0.95 for RFI and ADG, respectively. Their correlations were also high with (G)EBV of MT models (>0.93 for both traits).

Heritability estimates

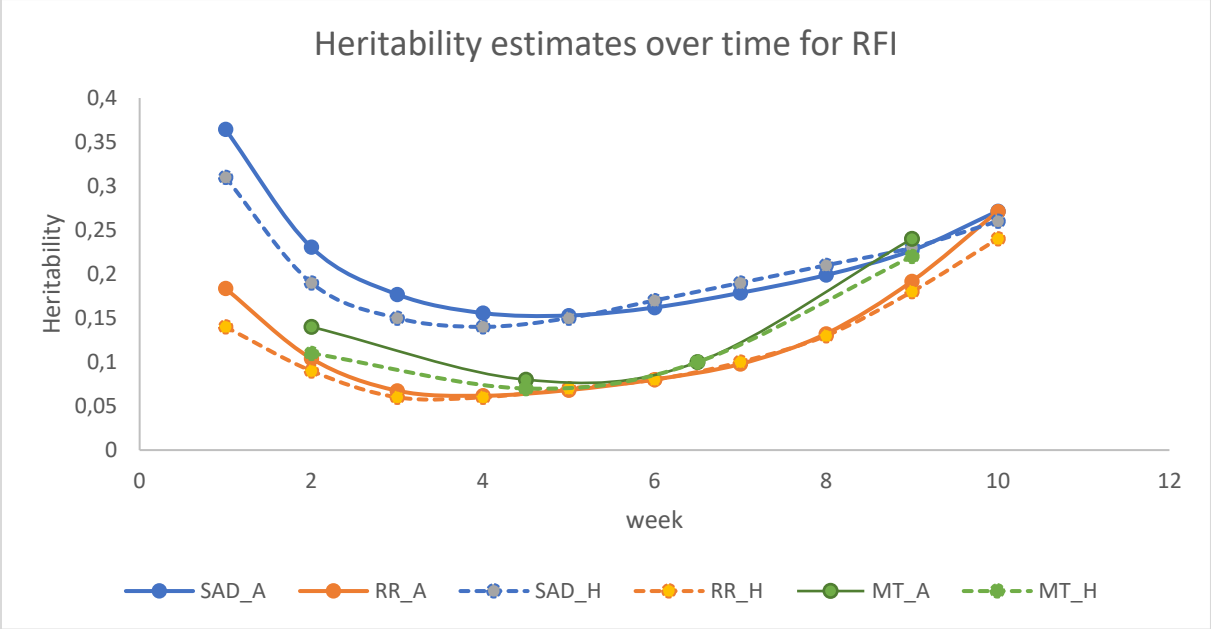
The estimates for ADG heritability obtained with the pedigree matrix varied from 0.07 to 0.42 for SAD, from 0.18 to 0.49 for RR, and from 0.42 to 0.23 for MT (Figure 1.a). The estimates followed a common trend for MT and RR models: they decreased from the beginning to the end of test, whereas those of the SAD model increased from week 1 to week 6 and then decreased until the end of test. For RFI, the heritability estimates for each model are presented in Figure 1.b. They ranged from 0.15 to 0.36, 0.18 to 0.27, and 0.08 to 0.26 for SAD, RR and MT models, respectively. The heritability estimates of the SAD model were higher than those of RR for all weeks, except week 10. The heritability estimates for MT were lower than those of the SAD model, whereas they were higher than the heritability estimates of RR.

With the **H** matrix the heritability estimates were very close to those obtained with the **A** matrix, with a maximum difference of 0.03 points for ADG and 0.05 for RFI. For the whole test period,

the heritability estimates with the **A** matrix and with the **H** matrix were 0.39±0.05 and 0.38±0.05 for ADG, and 0.21±0.03 and 0.20±0.03 for RFI.



a.



b.

Fig. 4.1. Heritability estimates with matrix **A** (solid curves) and with matrix **H** (dash curves) for average daily gain (ADG, a) and residual feed intake (RFI, b) over 10 weeks, using random regression (RR), structured antedependence (SAD) and multi trait (MT) models.

Genetic correlations

Average daily gain: The genetic correlations between weeks obtained with the genetic models were high (Fig. 2), ranging from 0.20 to 0.94 for the SAD model, and from -0.18 to 0.98 for the RR model. They generally decreased with the distance between weeks. The 1-week interval correlations ranged from 0.58 to 0.94 for SAD and from 0.93 to 0.98 for RR. The MT model showed the same patterns, with genetic correlations from -0.03 to 0.87 and high genetic correlations for 1-week intervals.

Residual feed intake: The genetic correlations between weeks obtained with the genetic models were high (Fig. 3), ranging from 0.20 to 0.90 for the SAD model, and from -0.20 to 0.97 for the RR model. They generally decreased with the distance between weeks. The 1-week interval correlations ranged from 0.61 to 0.90 for SAD and from 0.90 to 0.97 for RR. The genetic correlations of MT model ranged from -0.14 to 0.70 and genetic correlations for 1-week intervals varied from 0.36 to 0.70.

For both ADG and RFI, when the **H** matrix replaced the **A** matrix in the models, very similar genetic correlations matrices were obtained.

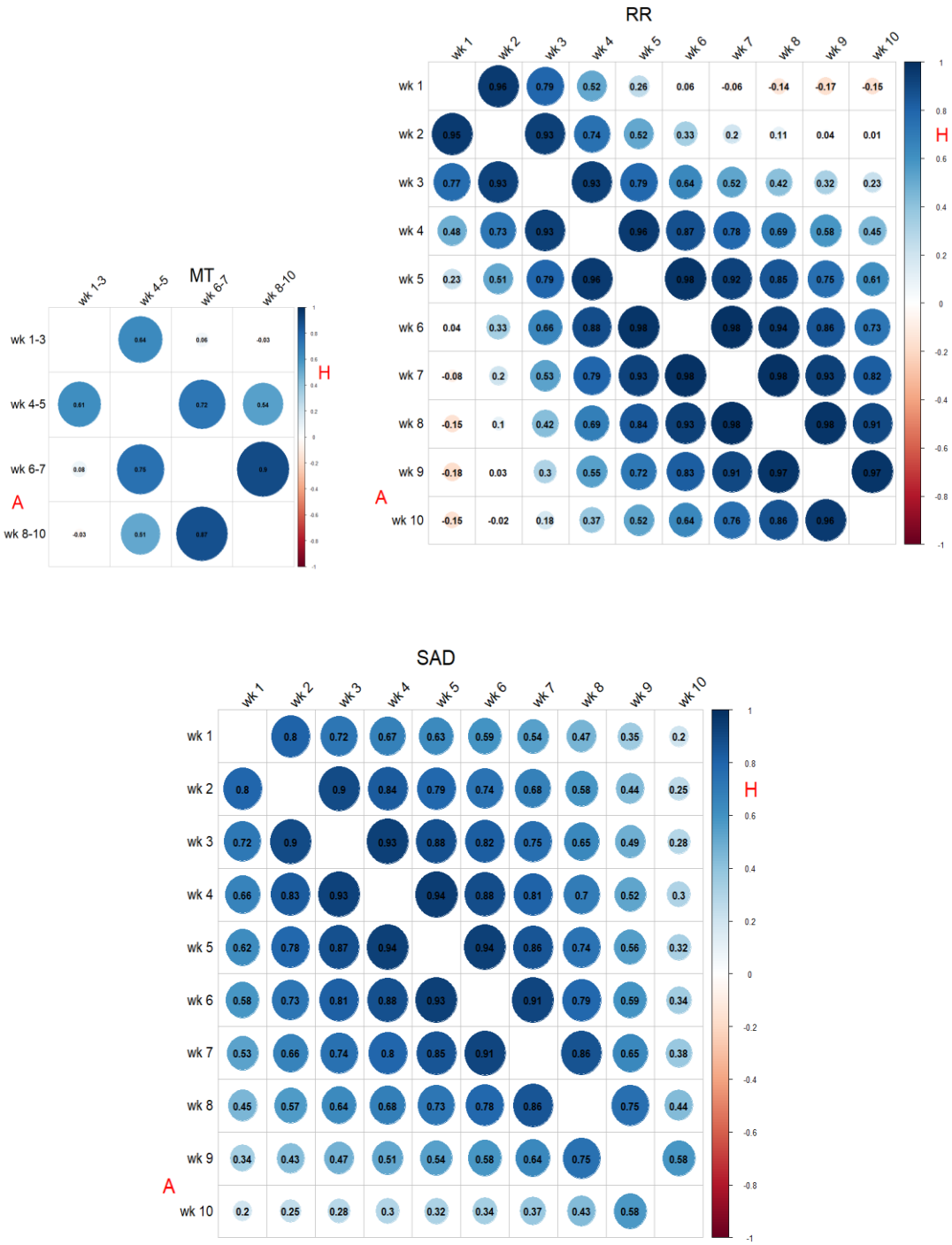


Fig. 4.2. Genetic correlation estimates between weeks (periods) with matrix **H** (above diagonal) and matrix **A** (below diagonal) for average daily gain estimated with random regression (RR), structured antedependence (SAD) and multi trait (MT) models. For MT model, the 10 weeks were combined into four periods (weeks 1 to 3, weeks 4 to 5, weeks 6 to 7, weeks 8 to 10).

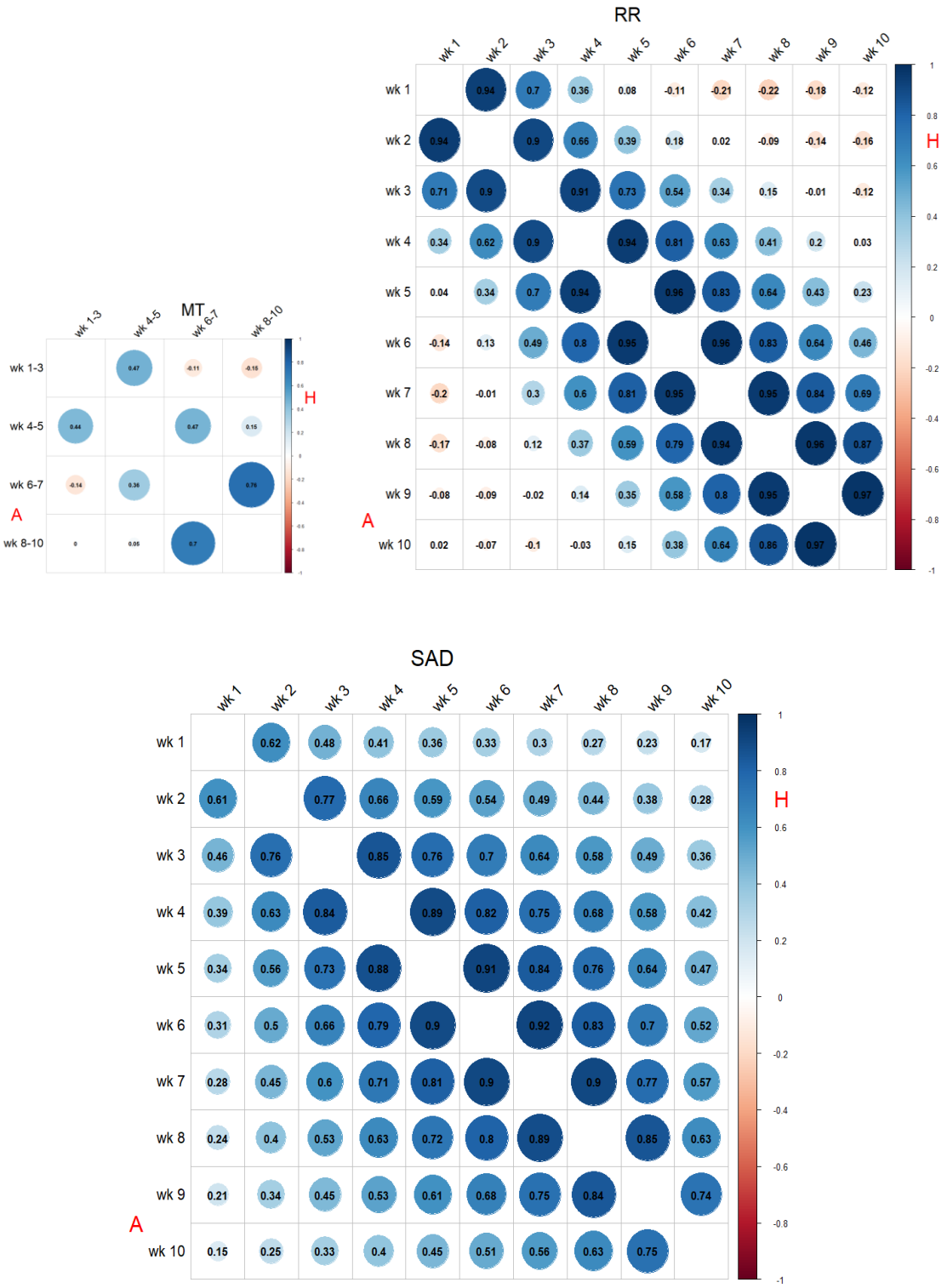


Fig. 4.3. Genetic correlation estimates between weeks (periods) with matrix **H** (above diagonal) and matrix **A** (below diagonal) for residual feed intake estimated with a random regression (RR), structured antedependence (SAD) and multi trait (MT) models. For MT model, the 10 weeks were combined into four periods (weeks 1 to 3, weeks 4 to 5, weeks 6 to 7, weeks 8 to 10).

Prediction accuracies and bias for ADG

The accuracies of the genetic and genomic predictions are presented in Tables 2 (Scenario 1) and 3 (Scenario 2).

Predicting HRFI pigs (Table 2). Accuracies were generally low. They ranged from -0.10 to 0.13 for the pedigree predictions and varied from -0.02 to 0.15 for the genomic predictions. Some accuracies were slightly negative, as low as -0.15, especially for the first and last weeks of prediction. No clear difference was observed between **A** and **H** based models, except for the SAD model where the HBLUP seemed to improve the prediction accuracies. The accuracy was improved with HBLUP only for period 1 for the MT model (0.06 vs -0.03 for the HBLUP and BLUP models, respectively).

The biases of the RR model ranged from -0.34 to 1.20 for the BLUP, and from -0.07 to 0.68 for the HBLUP. The biases of the SAD model were quite low for BLUP (from -0.07 to 0.55), and increased with HBLUP (from -0.08 to 0.82), getting closer to the target value 0.5. The biases of the MT model ranged from -0.19 to 0.48 for BLUP, and from -0.05 to 0.40 for HBLUP. For the whole-test ADG, the accuracies (bias) were 0.28 (0.62) and 0.32 (0.75) for matrices **A** and **H**, respectively.

Predicting LRFI pigs (Table 3). The accuracy to predict EBV in the LRFI line was slightly better than to predict the HRFI pigs. The prediction accuracies were higher with RR than with SAD when the pedigree matrix was used. The HBLUP seemed to improve the prediction accuracy for the last five weeks of test with RR, while it was increased for all weeks with the SAD model. The SAD model had biases closer to 0.5 than those of RR with the **H** matrix, ranging from 0.29 to 0.89. The HBLUP with the MT model slightly improved the prediction accuracy for the last three periods, but accuracies were generally low, and the bias improvement was not clear. For ADG over the entire test, accuracies (bias) were 0.31 (0.51) and 0.42 (0.59) for BLUP and HBLUP approaches, respectively.

Table 4.2. Accuracies and biases of predictions for weekly average daily gain in the high residual feed intake line validation dataset with random regression (RR), structured antedependence (SAD) and multi trait (MT) models with the pedigree (_A) or the combination of pedigree and genomic (_H) relationship matrix.

RR						
week	Cor(y*,EBV)[1]	Accuracy_A	Bias_A	Cor(y*,GEBV)[2]	Accuracy_H	Bias_H
1	-0.05	-0.10	-0.34	0.03	0.05	0.12
2	0.00	0.00	0.00	0.04	0.06	0.19
3	0.07	0.13	1.20	0.07	0.13	0.68
4	0.06	0.11	0.65	0.05	0.10	0.38
5	0.06	0.11	0.51	0.05	0.10	0.35
6	0.07	0.13	0.52	0.06	0.12	0.31
7	0.07	0.13	0.61	0.07	0.13	0.32
8	0.06	0.11	0.59	0.06	0.11	0.35
9	0.02	0.04	0.25	0.04	0.08	0.46
10	-0.03	-0.09	-0.23	-0.01	-0.02	-0.07
SAD						
week	Cor(y*,EBV)	Accuracy_A	Bias_A	Cor(y*,GEBV)	Accuracy_H	Bias_H
1	0.06	0.13	0.55	0.02	0.05	0.23
2	0.00	0.00	0.01	0.08	0.13	0.67
3	0.03	0.05	0.16	0.10	0.15	0.82
4	0.05	0.04	0.16	0.07	0.10	0.45
5	0.02	0.07	0.24	0.07	0.10	0.42
6	0.05	0.04	0.10	0.04	0.07	0.20
7	0.04	0.07	0.16	0.05	0.09	0.29
8	0.05	0.11	0.17	0.05	0.11	0.47
9	0.03	0.09	0.09	0.05	0.15	0.60
10	-0.03	-0.10	-0.07	-0.04	-0.15	-0.08
MT						
period	Cor(y*,EBV)	Accuracy_A	Bias_A	Cor(y*,GEBV)	Accuracy_H	Bias_H
Week 1 to 3	-0.02	-0.03	-0.19	0.04	0.06	0.28
Week 4 to 5	0.07	0.12	0.46	0.06	0.11	0.39
Week 6 to 7	0.06	0.12	0.48	0.06	0.12	0.40
Week 8 to 10	0.01	0.01	0.01	-0.01	-0.01	-0.05
Animal model						
	Cor(y*,EBV)	Accuracy_A	Bias_A	Cor(y*,GEBV)	Accuracy_H	Bias_H
Whole test	0.17	0.28	0.62	0.20	0.32	0.75

[1]y* = average corrected phenotypes of genotyped individuals in the validation set using the **A** relationship matrix; EBV = estimated breeding value; GEBV = estimated breeding value from the model including the **H** relationship matrix

Table 4.3: Accuracies and biases of predictions for weekly average daily gain in the low residual feed intake line validation dataset with random regression (RR), structured antedependence (SAD) and multi trait (MT) models with the pedigree (_A) or the combination of pedigree and genomic (_H) relationship matrix.

			RR			
week	Cor(y*, EBV)[1]	Accuracy_A	Bias_A	Cor(y*, GEBV)[2]	Accuracy_H	Bias_H
1	0.08	0.11	0.34	0.05	0.08	0.17
2	0.09	0.14	0.38	0.07	0.11	0.28
3	0.06	0.13	0.27	0.07	0.12	0.29
4	0.06	0.11	0.26	0.05	0.13	0.29
5	0.05	0.10	0.29	0.04	0.09	0.23
6	0.05	0.09	0.28	0.06	0.12	0.26
7	0.04	0.07	0.20	0.04	0.08	0.18
8	0.04	0.07	0.24	0.08	0.15	0.32
9	0.04	0.09	0.34	0.08	0.16	0.35
10	0.04	0.09	0.36	0.05	0.11	0.37
			SAD			
week	Cor(y*, EBV)	Accuracy_A	Bias_A	Cor(y*, GEBV)	Accuracy_H	Bias_H
1	0.07	0.14	0.52	0.08	0.17	0.58
2	0.08	0.13	0.45	0.09	0.16	0.48
3	0.03	0.04	0.10	0.10	0.15	0.37
4	0.04	0.05	0.15	0.10	0.15	0.35
5	0.04	0.05	0.16	0.08	0.12	0.29
6	0.04	0.07	0.15	0.10	0.16	0.46
7	0.05	0.10	0.25	0.10	0.17	0.51
8	0.05	0.10	0.25	0.09	0.19	0.68
9	0.04	0.13	0.32	0.06	0.18	0.73
10	0.08	0.30	1.20	0.07	0.28	0.89
			MT			
period	Cor(y*, EBV)	Accuracy_A	Bias_A	Cor(y*, GEBV)	Accuracy_H	Bias_H
Week 1 to 3	0.08	0.12	0.33	0.07	0.11	0.33
Week 4 to 5	0.05	0.09	0.25	0.08	0.12	0.27
Week 6 to 7	0.05	0.10	0.31	0.08	0.16	0.39
Week 8 to 10	0.05	0.12	0.42	0.06	0.13	0.32
			Animal model			
	Cor(y*, EBV)	Accuracy_A	Bias_A	Cor(y*, GEBV)	Accuracy_H	Bias_H
Whole test	0.19	0.31	0.51	0.26	0.42	0.59

y* = average corrected phenotypes of genotyped individuals in the validation set using the **A** relationship matrix; EBV = estimated breeding value; GEBV = estimated breeding value from the model including the **H** relationship matrix

Prediction accuracies and bias for RFI

The accuracies and biases of the genetic and genomic predictions are shown in Tables 4 (Scenario 1) and 5 (Scenario 2). The accuracies were generally lower for RFI than for ADG.

Predicting HRFI pigs (Table 4). Based on the pedigree information, the accuracies ranged from -0.03 to 0.32, -0.06 to 0.23, and 0.08 to 0.19 for the RR, SAD and MT models, respectively. The HBLUP applied to RFI did not improve these prediction accuracies. Conversely, with the RFI over the whole test period, the HBLUP achieved a slightly higher accuracy than the BLUP (0.08 vs 0.03).

Predicting LRFI pigs (Table 5). The weekly accuracies varied from -0.16 to 0.14 for all models. With the LRFI validation dataset, the HBLUP did not improve the predictions. For the first periods, negative accuracies were obtained with MT. Biases for BLUP and HBLUP were: -0.44 to 0.57 and -0.65 to 0.13 for RR, -0.19 to 0.34 and -0.47 to 0.11 for SAD, and -0.49 to 0.12 and -0.29 to 0.22 for MT. In addition, the HBLUP had lower accuracy (0.05 vs 0.07) and lower bias (0.22 vs 0.47) when obtained than the BLUP.

Table 4.4. Accuracies and biases of predictions for weekly residual feed intake in the high residual feed intake line validation dataset with random regression (RR), structured antedependence (SAD) and multi trait (MT) models with the pedigree (_A) or the combination of pedigree and genomic (_H) relationship matrix.

			RR			
week	Cor(y* ,EBV)[1]	Accuracy_A	Bias_A	Cor(y* ,GEBV)[2]	Accuracy_H	Bias_H
1	0.05	0.11	1.25	0.05	0.14	1.10
2	0.02	0.07	0.65	0.05	0.17	1.14
3	-0.01	-0.03	-0.13	-0.02	-0.08	-0. 33
4	0.03	0.13	0.55	0.02	0.09	0.32
5	0.06	0.22	0.76	0.03	0.12	0.34
6	0.06	0.21	0.67	0.02	0.08	0.21
7	0.10	0.32	0.99	0.04	0.15	0.37
8	0.02	0.04	0.12	-0.01	-0.04	-0.07
9	0.03	0.08	0.21	0.02	0.06	0.12
10	0.05	0.10	0.28	0.04	0.09	0.18
			SAD			
week	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_H	Bias_H
1	0.03	0.04	0.23	0.02	0.03	0.15
2	0.03	0.07	0.47	0.04	0.09	0.45
3	-0.02	-0.06	-0.25	-0.02	-0.04	-0.17
4	-0.00	-0.01	-0.03	-0.01	-0.03	-0.06
5	0.04	0.10	0.37	0.03	0.08	0.24
6	0.00	0.01	0.58	0.03	0.07	0.23
7	0.09	0.23	0.82	0.05	0.11	0.34
8	0.02	0.04	0.11	0.00	0.00	0.01
9	0.03	0.06	0.16	0.04	0.07	0.16
10	0.06	0.12	0.33	0.06	0.12	0.26
			MT			
period	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_H	Bias_H
Week 1 to 3	0.03	0.08	0.88	0.04	0.12	1.19
Week 4 to 5	0.03	0.11	0.37	0.01	0.04	0.11
Week 6 to 7	0.06	0.19	0.64	0.01	0.03	0.17
Week 8 to 10	0.05	0.10	0.26	0.04	0.09	0.12
			Animal model			
	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_H	Bias_H
Whole test	0.02	0.03	0.13	0.04	0.08	0.31

y*_A = average corrected phenotypes of genotyped individuals in the validation set using the **A** relationship matrix; EBV = estimated breeding value; GEBV = estimated breeding value from the model including the **H** relationship matrix

Table 4.5. Accuracies and biases of predictions for weekly residual feed intake in the low residual feed intake line validation dataset with random regression (RR), structured antedependence (SAD) and multi trait (MT) models with the pedigree (_A) or the combination of pedigree and genomic (_H) relationship matrix.

RR						
week	Cor(y* ,EBV)[1]	Accuracy_A	Bias_A	Cor(y* ,GEBV)[2]	Accuracy_H	Bias_H
1	-0.00	-0.01	-0.05	-0.01	-0.02	-0.08
2	-0.02	-0.07	-0.44	-0.02	-0.07	-0.39
3	-0.10	-0.05	-0.26	-0.04	-0.15	-0.65
4	0.00	0.01	0.04	-0.00	-0.01	-0.07
5	0.04	0.14	0.57	-0.00	-0.01	-0.03
6	-0.01	-0.03	-0.09	-0.04	-0.16	-0.38
7	0.03	0.10	0.36	0.00	0.00	0.00
8	-0.00	-0.01	-0.02	-0.01	-0.04	-0.08
9	-0.02	-0.04	-0.12	-0.00	-0.01	-0.02
10	0.00	0.01	0.02	0.03	0.07	0.13
SAD						
week	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_H	Bias_H
1	-0.01	-0.02	-0.08	-0.02	-0.03	-0.12
2	-0.00	-0.01	-0.05	-0.03	-0.06	-0.28
3	-0.01	-0.02	-0.10	-0.05	-0.12	-0.47
4	-0.00	-0.01	-0.04	-0.02	-0.06	-0.24
5	0.03	0.08	0.34	-0.02	-0.04	-0.13
6	-0.01	-0.02	-0.06	-0.04	-0.11	-0.28
7	0.01	0.02	0.07	-0.01	-0.02	-0.06
8	-0.03	-0.06	-0.19	-0.02	-0.05	-0.11
9	-0.01	-0.03	-0.08	0.01	0.02	0.04
10	-0.01	-0.01	-0.03	0.03	0.07	0.11
MT						
period	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_A	Bias_A
Week 1 to 3	-0.02	-0.05	-0.49	-0.01	-0.03	-0.24
Week 4 to 5	0.01	0.03	0.12	-0.02	-0.08	-0.29
Week 6 to 7	0.002	0.00	0.02	-0.01	-0.02	-0.07
Week 8 to 10	0.002	0.00	0.01	0.05	0.11	0.22
Animal model						
	Cor(y* ,EBV)	Accuracy_A	Bias_A	Cor(y* ,GEBV)	Accuracy_H	Bias_H
Whole test	0.03	0.07	0.47	0.02	0.05	0.22

[1]y*_A = average corrected phenotypes of genotyped individuals in the validation set using the **A** relationship matrix; EBV = estimated breeding value; GEBV = estimated breeding value from the model including the **H** relationship matrix

DISCUSSION

Genetic parameters for longitudinal ADG and RFI

Heritabilities for longitudinal measurements were expected to be lower than for the whole test period measurement. The heritability estimates for longitudinal ADG in this study were in line with other pig studies performed on ADG for whole test period, e.g. 0.37 ± 0.05 [11], 0.48 ± 0.13 [22], 0.44 ± 0.11 [23]. Despite slightly different curves over time, the estimates were relatively similar for the three models for longitudinal ADG. The heritability estimates for longitudinal RFI ranged from 0.06 to 0.36, with low heritability for the intermediate stages with all models. The heritability estimates of the SAD were higher than those of RR models, in particular for early weeks of test, due to the fact that SAD provided higher genetic variances and lower permanent environmental variances. These values were in the range of estimations from previous studies [0.03 to 0.10 [24], 0.22-0.24 [25], 0.27-0.36 [26]].

Most studies compute RFI before estimating the variance components in a 2-step approach [23,27]. Begli et al. (2016) [28] reported a 1-step approach for estimating the genetic parameters for RFI without first calculating the phenotype. It takes a little longer for the animal models to converge but this method is more appropriate than the 2-step approach. Generally, the mean of heritability estimates for HBLUP were very close to their BLUP counterparts. Heuer et al. (2018) [29] also reported a slightly lower heritability estimate for RFI in growing Holstein heifers using a HBLUP than with a BLUP approach, but not significantly different.

In this study, genetic correlations between weeks with SAD and RR were high when weeks were close. As reported previously for FCR [10], the SAD model led to positive correlations between all time points, whereas the RR model often provided a pattern with a change of correlation sign from positive to negative when the distance between weeks increased [10,12]. It suggests that the genetic correlation patterns for RR and SAD are to some extent due to the specific properties of the models. Negative correlations between the first period and later period of test were also obtained with MT. This coincides with RR estimations, even though the MT model does not perfectly describe the genetic correlation changes, as i) it assembles multiple weeks into periods, ii) it estimates independent fixed effects for the four-period traits.

Genetic and genomic prediction

Although the single step genomic models have been recently widely used, most studies focused on traits with a single record per animal. Few applied it to longitudinal data. Koivula et al [8] performed a single step analysis using random regression in a test-day model in Nordic Red

dairy cows. Kang (2019) [9] performed a single step approach with random regression on simulated phenotypes. On pigs, only few genomic predictions of real single record phenotype are reported, such as Jiao et al (2014) and Christensen et al (2012), both on growth and feed efficiency traits.

Due to the particular design of our study, when training was run on all G0 to G4 animals of the two lines to predict the EBV or GEBV for RFI in the next generations in both lines simultaneously, very high correlations between the EBV (GEBV) and corrected phenotypes were obtained (results not shown), due to the strong selection applied to the lines [30]. In addition, as shown by Mauch et al (in prep.) the genomic content of the advanced generations heavily evolved due to combination of drift and selection, shaping very contrasted genomes in the latter generations. To circumvent this specificity, training on both lines and validating the prediction in each line separately was retained.

Slightly higher accuracies were yielded with HBLUP for the SAD model on ADG. Other models and traits did not show improvement of the accuracies due to genomic prediction. However, the accuracies were generally low, which is close to estimates in Jiao et al. (2014)[23], who reported an accuracy of 0.24 for ADG in Duroc pigs using one-record single step models. Christensen et al (2012) [31] reported higher accuracies of 0.35 for single record ADG with the GBLUP approaches than with a BLUP. However, in our study the bias was not systematically improved by the genomic approach for these models, the biases closer to 0.5 being with the **A** matrix with RR on the HRFI line and the **H** matrix with SAD on the LRFI line. However, consistent advantages of the **H** matrix on the predictions biases have been reported with the single step approach using random regression models [8]. The change of bias in MT due to the **H** matrix was not considerable, and similarly for ADG for the whole test. The advantages of the genomic approach for RFI were even lower and less consistent across validation sets. The accuracies were generally lower than for ADG, as reported by Jiao et al. (2014) [23] with an accuracy of 0.09 for RFI and 0.11 for FCR on Duroc pigs. Christensen et al. (2012) [31] reported slightly greater accuracies of 0.19 and 0.21 for RFI with genomic approaches in another Duroc population. In chicken, Begli et al. (2017) [28] reported that the accuracies (with pedigree information) for longitudinal RFI ranged from -0.11 to 0.33, and that GBLUP using high-density SNP did not improve EBV prediction for this trait. In our study, some weekly accuracies were also negative with both **A** and **H** matrices for RFI. This suggests a poor predictive ability of the longitudinal models for some time points for RFI.

Finally, for RFI over the whole test period, the prediction accuracy of HBLUP was slightly higher than those of BLUP for the HRFI validation dataset, whereas for the LRFI validation set, there was no benefit from using HBLUP, with no clear putative explanation.

The prediction accuracy of the single step approach mainly depends on the heritability of the trait, the number of phenotype records in the training population, the connectedness between of the training and validation dataset, the linkage disequilibrium between the SNP and the QTL affecting the trait, and also the genomic relationship and effective number of SNP markers [23,32]. The difference between accuracies for ADG and RFI could be explained by the fact that RFI has a lower heritability than ADG, and that the number of QTL detected for RFI is generally low (see Gilbert et al 2017 [30], for instance). Our limited accuracies, sometimes negative for RFI, could also result from the structure of the population used: a divergent selection experiment on RFI with two separated lines [11], in which the number of animals of each generation is limited, as well as the number of genotyped animals (17% of the population) and so the size of the validation set. Finally, among the 660 genotyped animals, only 90 animals had both genotyped and phenotypes, so we computed average progeny-based corrected phenotypes for genotyped individuals of the validation sets to increase the accuracies. On the other hand, the pedigree is well structured, all the animals within a line have a high level of connectedness with the others, and the number of phenotyped animals is high enough to provide good accuracies of the EBV of their genotyped parents. Ultimately, predicting longitudinal RFI seemed more difficult than predicting longitudinal ADG. From earlier studies of these populations (see Gilbert et al, 2017 for a review), it has been shown that ADG showed no correlated response to selection on RFI, whereas RFI was the main responding trait. Indeed, the genomic structure of the lines [11], which differentiated due to selection but also due to drift [33], could disadvantage prediction for the selection criterion. As a result, we can hypothesized that results obtained on ADG, which is certainly less affected by our design, are closer to real situations in which responses to selection are distributed on multiple traits thanks to multitrait indexes. Under this hypothesis, HBLUP seemed to improve the genomic accuracies of the longitudinal models. On the other hand, we can suggest that predicting EBV in one population with very divergent phenotypes compared to the training population might be difficult, or at least require extended numbers of genotypes and phenotypes to capture most of the available variability in the training population.

Longitudinal models have been implemented in animal breeding over the last 20 years, and the single step approach has been used from 2010, but reports of the combination of both are scarce. Some study showed that including the genomic information improved reliability on real data in

cattle [8] and on simulated data [9]. No genomic prediction study is available on pig repeated measurements for comparison. Altogether, the prediction accuracies of weekly (G)EBV were reduced compared to accuracies of single records models, as a result of the model complexity. However, compared to Jiao et al (2014) [23] and Christensen et al (2012) [31], the accuracies for weekly ADG remained in the range of values they obtained for the different studied traits. Finally, previous studies of longitudinal traits with genomic models used random regression models for single step evaluation. In our study, we tested the SAD model in a genomic context for the first time, as an alternative approach. The SAD is interesting because it requires fewer parameters than the RR model.

CONCLUSION

Among all models, SAD provided the best fitted data model for both ADG and RFI. The study also confirmed that low predictive accuracies are obtained in pig populations compared to cattle situations. The situation of a divergent selection experiment seemed to worsen the accuracies for the traits heavily responding to selection: the single step approach improved the genomic prediction for ADG in some cases, but this was not observed for RFI.

ACKNOWLEDGMENTS

This work is part of the H2020 Feed-a-Gene project funded by the European Union's H2020 Program under grant agreement no. 633531. V.H. Huynh-Tran acknowledges the financial support of INRA in the frame of the "Ecole Internationale de Recherche d'Agreenium" (EIR-A) program for facilitating his internship at Iowa State University. V. H. Huynh-Tran acknowledges the significant help of Dr. Zulma Vitezica for her discussing and technical help on this study.

REFERENCES

1. Patience JF, Rossoni-Serão MC, Gutiérrez NA. A review of feed efficiency in swine: biology and application. J Anim Sci Biotechnol [Internet]. 2015 [cited 2017 Mar 15];6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4527244/

2. Shirali M, Duthie C-A, Doeschl-Wilson A, Knap PW, Kanis E, van Arendonk JA, et al. Novel insight into the genomic architecture of feed and nitrogen efficiency measured by residual energy intake and nitrogen excretion in growing pigs. BMC Genet. 2013;14:121.

3. VanRaden PM, Tassell CPV, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci. 2009;92:16–24.

4. Legarra A, Robert-Granié C, Manfredi E, Elsen J-M. Performance of Genomic Selection in Mice. Genetics. 2008;180:611–8.

5. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci. 2009;92:4656–63.

6. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci. 2009;92:4648–55.

7. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. Genet Sel Evol. 2010;42:2.

8. Koivula M, Strandén I, Pösö J, Aamand GP, Mäntysaari EA. Single-step genomic evaluation using multitrait random regression model and test-day data. J Dairy Sci. 2015;98:2775–84.

9. Kang H, Zhou L, Mrode R, Zhang Q, Liu J-F. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity. 2017;119:459–67.

10. Huynh-Tran VH, Gilbert H, David I. Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs. J Anim Sci. Nov 2017b;95:4752–63.

11. Gilbert H, Bidanel J-P, Gruand J, Caritez J-C, Billon Y, Guillouet P, et al. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J Anim Sci. 2007;85:3182.

12. David I, Ruesche J, Drouilhet L, Garreau H, Gilbert H. Genetic modeling of feed intake. J Anim Sci. 2015;93:965–77.

13. Huynh-Tran VHH, Gilbert H, David I. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J Anim Sci. 2017;95:39–48.

14. Noblet J, Karege C, Dubois S, van Milgen J. Metabolic utilization of energy and maintenance requirements in growing pigs: effects of sex and genotype. J Anim Sci. 1999;77:1208–16.

15. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.

16. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. J Dairy Sci. 2010;93:743–52.

17. Wang W. Identifiability of linear mixed effects models. Electron J Stat. 2013;7:244–63.

18. Schwarz G. Estimating the Dimension of a Model. Ann Stat. 1978;6:461–4.

19. Gilmour AR, Gogel B., Cullis B., Thompson R. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. 2009; Available from: www.vsni.co.uk

20. Aguilar I, Misztal I, Tsuruta S, Legarra A, Huiyu Wang. PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. 2014;

21. Zhang C, Kemp RA, Stothard P, Wang Z, Boddicker N, Krivushin K, et al. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. Genet Sel Evol GSE. 2018;50:14.

22. Hoque MA, Kadowaki H, Shibata T, Oikawa T, Suzuki K. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest Sci. 2009;121:45–9.

23. Jiao S, Maltecca C, Gray KA, Cassady JP. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J Anim Sci. 2014;92:2377–86.

24. Shirali M, Strathe AB, Mark T, Nielsen B, Jensen J. Joint analysis of longitudinal feed intake and single recorded production traits in pigs using a novel Horizontal model. J Anim Sci. 2017;95:1050–62.

25. Nguyen NH, McPhee CP, Wade CM. Responses in residual feed intake in lines of Large White pigs selected for growth rate on restricted feeding (measured on ad libitum individual feeding). J Anim Breed Genet Z Tierzuchtung Zuchtungsbiologie. 2005;122:264–70.

26. Cai W, Casey DS, Dekkers JCM. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J Anim Sci. 2008;86:287–98.

27. Godinho RM, Bergsma R, Silva FF, Sevillano CA, Knol EF, Lopes MS, et al. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. J Anim Sci. 2018;96:817–29.

28. Begli HE, Torshizi RV, Masoudi AA, Ehsani A, Jensen J. Longitudinal analysis of body weight, feed intake and residual feed intake in F2 chickens. Livest Sci. 2016;184:28–34.

29. Heuer C, Deeb N, Sun C, Kendall D, Moreno J, Vishwanath R. Single step genomic evaluation of residual feed intake in growing Holstein heifers. Proc World Congr Genet Appl Livest Prod Vol Electron Poster Sess - Biol - Feed Intake Effic 1 665 2018. 2018;

30. Gilbert H, Billon Y, Brossard L, Faure J, Gatellier P, Gondret F, et al. Review: divergent selection for residual feed intake in the growing pig. Anim Int J Anim Biosci. 2017;1–13.

31. Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. Single-step methods for genomic evaluation in pigs. animal. 2012;6:1565–71.

32. Zhou L, Mrode R, Zhang S, Zhang Q, Li B, Liu JF. Factors affecting GEBV accuracy with single-step Bayesian models. Heredity. 2018;120:100–9.

33. Mauch E., Labrune Y, Riquet J, Gilbert H, Dekkers JC. Genetic characterization of two independent populations of pigs divergently selected for residual feed intake. Gordon Conf Quant Genet Genomics. Galveston, Texas; 2017 Mar 26;

Chapter 5 - IMPACT OF THE DIVERGENT SELECTION FOR RESIDUAL FEED INTAKE ON THE GENETIC TRAJECTORIES OF RESIDUAL FEED INTAKE AND FEED CONVERSION RATIO

During this thesis, we tested different longitudinal models on a population with a particular genetic structure, Large White pigs divergently selected for RFI. This selection could enhance the changes of trait dynamics in response to selection. We thus evaluated if the divergent selection for RFI on the entire test period affected the genetic profiles of feed efficiency, FCR after the study presented in chapter 3, and RFI after the study presented in chapter 4.

In both cases, the (G)EBV of longitudinal models were averaged per line and generation to show the changes of dynamics with selection. The work on FCR using the RR model and the pedigree information only was presented as a poster at the 11th World Congress on Genetics Applied to Livestock Production (WCGALP), Auckland, NZ (2018), and was completed with results from the SAD model. The study on changes of longitudinal RFI with selection took advantages of the outputs of RR, SAD, and MT models, using genetic and genomic EBV.

5.1. Impact of the divergent selection for RFI on longitudinal FCR

As shown in the introduction of this manuscript, FCR strongly correlates with RFI at the level of the test period. It suggests that selection for RFI would affect FCR profiles during growth.

5.1.1. Using RR-OP models

Based on the outputs of the RR-OP models presented in article II (chapter 3), this study investigated the changes of EBV patterns for FCR of the animals undergoing a divergent selection for RFI.

(**Article III**: Huynh-Tran, V. H., David, I., Billon, Y., Gilbert, H. (2018). Changes of EBV trajectories for feed conversion ratio of growing pigs due to divergent selection for residual feed intake. In: Proceedings of the 11th World Congress on Genetics Applied to Livestock Production (p. 513). Presented at 11. World Congress on Genetics Applied to Livestock Production (WCGALP), Auckland, NZL (2018-02-12 - 2018-02-16). NZL: Massey University. 1130 p.)

Changes of EBV trajectories for feed conversion ratio of growing pigs due to divergent selection for residual feed intake

V.H. Huynh-Tran[1], I. David[1], Y. Billon[2], H. Gilbert[1]

[1]GenPhySE, INRA, Université de Toulouse, INPT, ENVT, 31326 Castanet-Tolosan cedex, France
van-hung.tran@inra.fr (Corresponding author)
[2]GenESI, INRA, 17700 Surgères, France

Summary

The objective of this paper was to study the estimated breeding values (EBV) profiles for feed conversion ratio (FCR) over the growing period in eight generations of divergent selection for Residual Feed Intake (RFI) in Large White pigs. Data comprised 11790 weekly FCR collected on 1186 boars during a 10 week-period. A random regression model was used to estimate 10 week EBVs per animal. Then, the individual EBV trajectories were classified into subgroups using a *k*-means approach with a Euclidean distance. The responses to selection in the divergent lines (LRFI= low RFI; HRFI= high RFI) on the FCR dynamics were evaluated. On the one hand, the average weekly EBV over time per line and generation were considered; on the other hand the first two summarized breeding values (SBV1 & SBV2, representing the slope and the mean of the EBV curves, respectively) were computed for each individual from the eigendecomposition of the genetic covariance matrix between time points, and they were averaged for each line and generation. The results showed that individual EBV trajectories for FCR were classified into three distinct subgroups. These groups corresponded to early generations, late LRFI generations, and late HRFI generations, respectively. The more efficient pigs had smaller initial value of FCR and smaller slope of the EBV than less efficient pigs. The changes in SBV1 and SBV2 corroborated the evolution of the EBV curves for each line and generation. Further examination pointed out changes in the dynamics of growth rates associated to these responses. This study showed that selection for feed efficiency affected the dynamics of FCR during growth.

Keywords: longitudinal data, pigs, divergence selection, feed efficiency, residual feed intake

Introduction

Improving feed efficiency of pigs contributes to increased profitability of pig farming as well as reduction of its negative impacts on the environment (Patience et al. 2015, Gilbert et al. 2017). Feed conversion ratio (FCR = daily feed intake/average daily gain) and residual feed intake (RFI; difference between observed feed intake and expected feed intake for maintenance and production requirements) are two measures of feed efficiency. The RFI permits to select animals that consume less without affecting growth traits. Selection for RFI has resulted in an improvement of FCR at the phenotypic and the genetic levels (Gilbert et al., 2017). However, little is known about the effect of selection on the dynamics of feed efficiency over the growing period. The analysis of repeated records can improve estimations in a genetic selection context

for feed efficiency compared to simple trait analyses (Boligon et al., 2011). With the development of automatic self-feeders, individual feed intake and body weight values (and thus FCR) can be measured repeatedly during the growing period in groups of animals. For RFI, records require measurement of body composition to evaluate production requirements variability due to differences in protein/lipid distribution, which is usually not measured repeatedly. To evaluate the impact of selection for RFI on the dynamics of feed efficiency during growth, we studied the evolution of the estimated breeding values (EBV) profiles for FCR in divergent lines for RFI.

Material and methods

Pigs and data collection

The present study includes data from 1186 growing Large White boars over 8 generations raised after weaning in the Rouillé INRA experimental farm (GenESI, Vienne, France). From the initial generation (G0), the lines were divergently selected for RFI to produce animals with low RFI (LRFI, corresponding to more efficient animals) and high RFI (HRFI, corresponding to less efficient animals). The genetic selection process has been described in details by Gilbert et al. (2007). Pigs were tested during the growing-finishing period from 67 ± 1 day (25 ± 4 kg) to 168 ± 13 days (115 ± 11 kg). They had individual measurements for body weight (BW) every week and for feed intake (FI) every day. From the 14-week test, weekly FCRs were calculated according to Huynh-Tran et al. (2017a), resulting in 11790 weekly FCR values over 10 weeks. A total of 3986 animals were included in the pedigree.

Statistical analyses

The 10 repeated measurements of FCR were analyzed using a random regression model with polynomial of order 2 for both genetic and permanent environmental effects. The fixed effects included in the model were the week of observation (10 levels), the pen within batch (96 levels), the batch of birth (32 levels), the age and BW of the animal at the beginning of the test. Covariance components and breeding values were estimated by the restricted maximum likelihood (REML) method using the ASRemL software (Gilmour et al., 2009).

As a result of the genetic models, we obtained 10 weekly EBVs for FCR per animal. The patterns of EBV changes over time were then described using a trajectory classification approach that classified animals into different trajectory groups using a *k*-means approach with a Euclidean distance (Genolini et al., 2015). This trajectory classification was proposed to visualize the changes of animal profiles as a result of the selection on RFI. Next, the 10-EBVs vector for each animal was summarized in a reduced number of independent summarized breeding values (SBV). For animal i, SBVp_i was obtained by multiplying the p^{th} eigenvector of the eigendecomposition of the estimated genetic covariance matrix between times (**G)** with the vector of 10 week $\boldsymbol{EBV}_i$. As shown in Huynh-Tran et al. (2017b), SBV_1 and SBV_2 are related to the slope and mean of the EBV curves, respectively. SBV1 and SBV2 were averaged per line and generation to assess the evolution of the dynamics of weekly EBV for FCR in response to selection for RFI.

Results and discussion

The trajectory classification approach identified three subgroups of individual EBV trajectories, as shown in **Fig. 1.** The first group consisted in a low initial value and continuous increase over time with a weak slope (35.4 % of the pigs, group noted A). The second group also reflected an increase of the EBV over time but with a steeper initial slope and higher initial value (34.6% of the animals, group noted B). The last EBV trajectory pattern reflected a constant EBV over time (30.0% of the animals, group noted C). This classification is strongly related with the selection for RFI : the EBV trajectories of the animals from the first three generations belonged to group C, pigs from generations 4 to 7 of the LRFI line belonged to group A, and animals from generations 4 to 7 of the HRFI line to group B (Fig. 1). These results suggested that selection for RFI strongly impacted the FCR curves during the test. Altogether, the FCR curve shapes supported that pigs are more efficient to convert feed into meat at the earlier ages than at the end of the test as previously reported by Shirali et al. (2012).

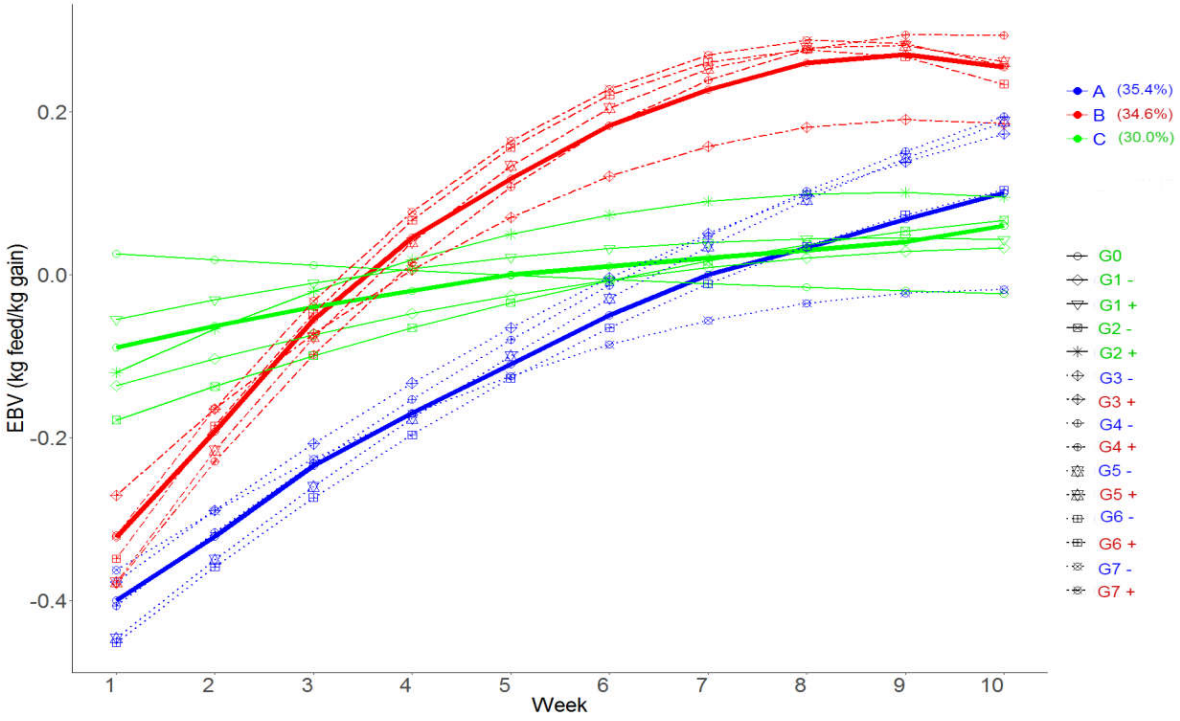


Figure 1. Mean EBV trajectories per line and generation obtained with a random regression model during the test for the RFI lines. The solid green lines (group C) are the EBV trajectories of pigs from generations G0, G1, G2, the dotted blue lines (group A) are the EBV trajectories of pigs from generations G3 to G7 of the low RFI line; the dotdashed red curves (group B) are the EBV trajectories of pigs from generations G3 to G7 of the high RFI line. The bold lines are the mean curve of group A, B and C, respectively.

To better describe how selection for RFI affected the FCR trajectory, changes in SBV1 and SVB2 per line and generation are presented in Fig. 2. The SBV2, which is strongly correlated to the mean of EBV trajectories (0.96, Huynh-Tran et al, 2017b) quasi-linearly decreased for LRFI line (from -0.13 for G1 to -0.51 (kg feed/kg gain) for G7) and increased for HRFI line (from 0.015 for G1 to 0.17 for G7). This result was in line with the study of Gilbert et al. (2017)

that reported significant response to selection on FCR over the whole test-period in these lines. The SBV1, associated with the slope of the EBV trajectories (Huynh-Tran et al, 2017b), increased from G0 to G5, then decreased until G7 for LRFI line. Meanwhile, SBV1 increased from G0 to G4, then slightly decreased from G5 to G7 for the HRFI line. As a result, slopes of the FCR curves remained similar between lines until G3, then the slopes started diverging , the decrease of feed efficiency with time being higher in the HRFI line in comparison with the LRFI line. These results show that selection for RFI impacted the FCR curves during the test. Since FCR is a ratio, differences in changes in FCR over time between lines maybe related, as suggested by Saintilain et al. (2015), to differences between lines in changes in FI, ADG or both.

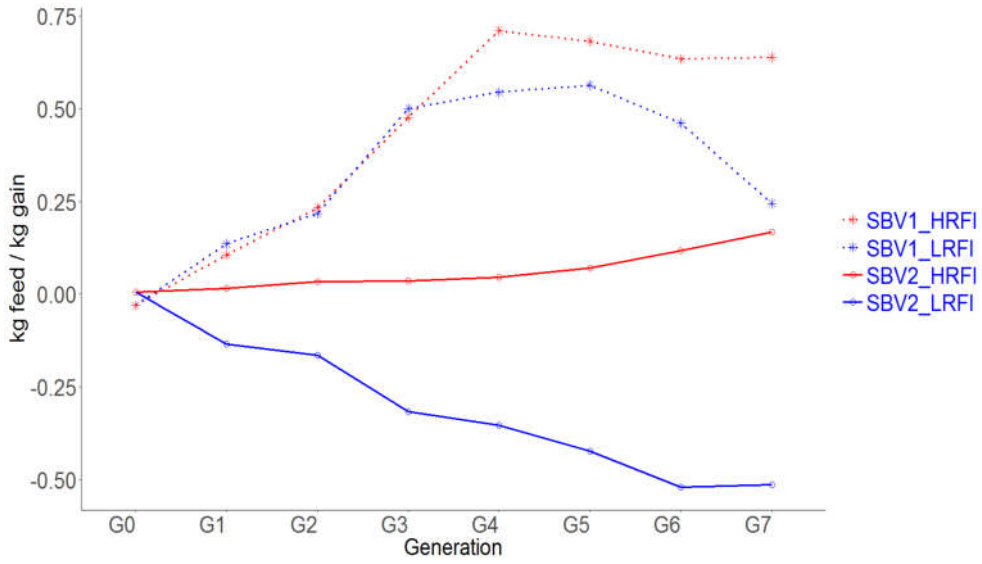


Figure 2 : Change of the summarized breeding values during 8 generations (G0 to G7). SBV1_LRFI (SBV1_HRFI), SBV2_LRFI (SBV2_HRFI): first and second summarized breeding values for feed conversion ratio (FCR) obtained from the genetic covariance matrix ***G*** *with a random regression model for the low residual feed intake line (LRFI)(high residual feed intake line (HRFI)).*

So, to further understand the differences in FCR curves between lines, the phenotypic FI and ADG curves were also examined per line and generation (results not shown). We observed the same shape of FI curve over time for both lines, with a change of magnitude due to the generations (LRFI pigs eating less feed every week than HRFI pigs). For ADG, the pattern of the ADG curves changed with the selection. The ADG curves of the first three generations (G0 to G2) were similar for both lines. From G3, the ADG patterns changed between lines. The LRFI pigs had a lower growth rate for the earlier periods and higher growth rate for the later period than the HRFI pigs. A similar difference was reported at the phenotypic level by Saintilan et al. (2015), who reported faster growing animals in the less efficient group at the beginning of the test, and faster growing animals in the more efficient group at the end of the test-period. It has been shown that this difference in growth rate is associated with difference in lipid to protein deposition ratio between the divergent RFI lines (Gilbert et al., 2017), resulting in increased leanness in the LRFI pigs.

Conclusion

This study showed that selection based on RFI had an impact on the dynamics of the FCR over time.

List of References

Boligon, A. A., F. Baldi, M. E. Z. Mercadante, R. B. et al. 2011. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. Genet. Mol. Res. GMR. 10:1227–1236.

David, I., J. Ruesche, L. Drouilhet, et al. 2015. Genetic modeling of feed intake. J. Anim. Sci. 93:965–977.

Genolini, C., X. Alacoque, M. Sentenac, and C. Arnaud. 2015. kml and kml3d: R Packages to Cluster Longitudinal Data. J. Stat. Softw. 065.

Gilbert, H., J.-P. Bidanel, J. Gruand, J.-C. et al. 2007. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J. Anim. Sci. 85:3182.

Gilbert, H., Y. Billon, L. Brossard, et al. 2017. Review: divergent selection for residual feed intake in the growing pig. Animal 11(9):1427-1439.

Gilmour, A. R., B. J. Gogel, and B. R. Cullis. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Huynh-Tran, V. H., H. Gilbert, and I. David. 2017a. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J. Anim. Sci. 95:39–48.

Huynh-Tran, V. H., H. Gilbert, and I. David. 2017b. Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs. J. Anim. Sci. (in press)

Patience, J. F., M. C. Rossoni-Serão, and N. A. Gutiérrez. 2015. A review of feed efficiency in swine: biology and application. J. Anim. Sci. Biotechnol. 6(1):33.

Saintilan, R., L. Brossard, B. Vautier, et al. 2015. Phenotypic and genetic relationships between growth and feed intake curves and feed efficiency and amino acid requirements in the growing pig. Animal. 9(1):18–27.

Shirali, M., A. Doeschl-Wilson, P. W. Knap. et al. 2012. Nitrogen excretion at different stages of growth and its association with production traits in growing pigs. J. Anim. Sci. 90:1756–1765.

5.1.2. Changes of FCR profiles with the SAD approach

The previous study with the RR model revealed that selection for RFI affects the change of FCR overtime at the genetic level. With the SAD model, similar patterns of EBV per generation as with the RR model were obtained in each line (Fig. 5.1). They were not as smooth, which reflects the properties of the models.

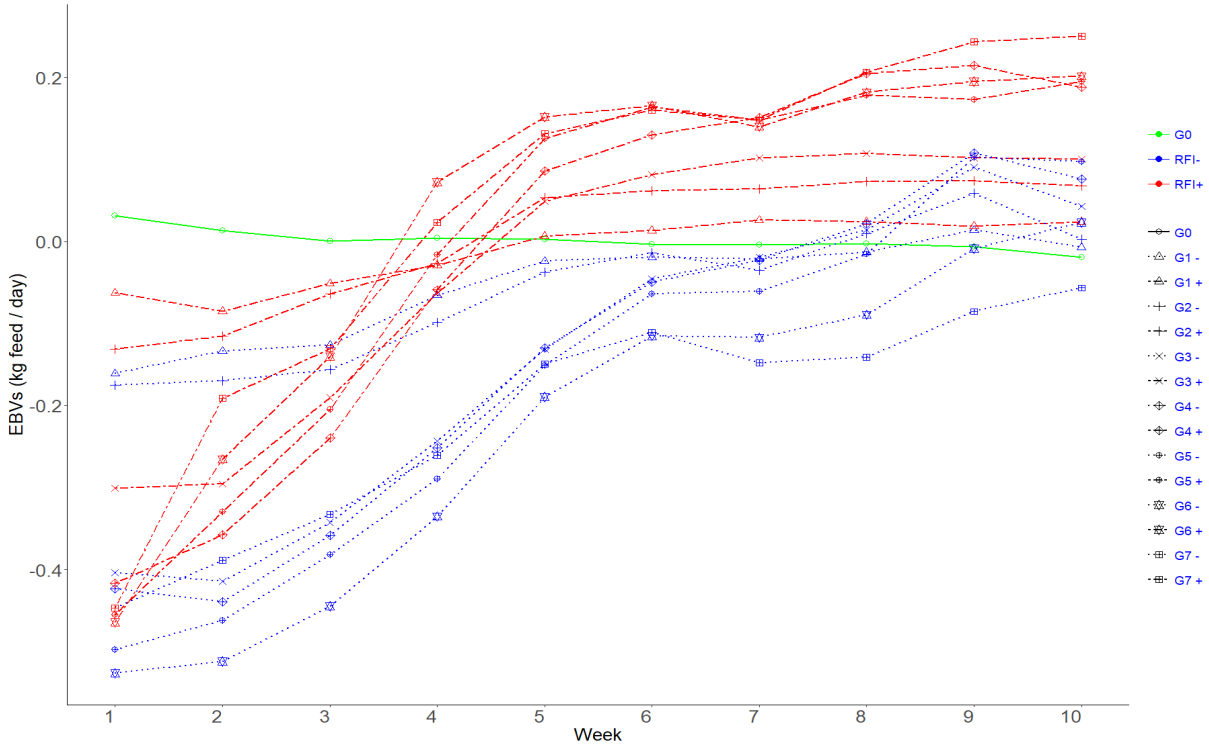


Figure 5.1. Mean EBV trajectories per line and generation obtained with a structured antedependence model (SAD 122/122 for genetic random and permanent environmental effects, respectively) during the test for the RFI lines. The solid green lines (group C) are the EBV trajectories of pigs from generations G0, G1, G2, the dotted blue lines (group A) are the EBV trajectories of pigs from generations G3 to G7 of the low RFI line; the dot-dashed red curves (group B) are the EBV trajectories of pigs from generations G3 to G7 of the high RFI line. The bold lines are the mean curve of group A, B and C, respectively.

5.2. Genetic (EBV) and genomic (GEBV) trajectories of longitudinal residual feed intake

5.2.1. Data and statistical models

Data and statistical models for the RR, SAD and MT approaches were presented in chapter 4. To examine the impact of selection for RFI during the test on the RFI trajectories, the averages of the weekly (G)EBV per combination of generation and line were computed from the results of the pedigree-based animal models and the single step models applied to the full population.

5.2.2. Results: RFI trajectories

The resulting averaged trajectories for (G)EBV for each line and generation combination are presented in Figure 5.2 for the longitudinal models RR, SAD and MT. To ease readiness, the profiles of the G0 population were set to zero, and other profiles are presented as deviations from the G0 population.

In general, the EBV and GEBV trajectories were quite similar. The trajectory patterns were smoother with RR than with SAD and MT, due to the nature of the covariance structure. The LRFI curves were below the G0 line, and the HRFI curves were above the G0 line, reflecting the changes in RFI average levels due to selection. For HRFI, on average the (G)EBV increased from the week 1, reached a peak at week 7 or 8 and then decreased until the end of the test, but not to the initial level. This pattern was amplified in the advanced generations compared to the earlier generations of selection, with differences between weeks reaching more than 500g/d in the last HRFI generation between week 1 and week 7. Conversely, the (G)EBV patterns of LRFI animals had opposite shapes, from flat to low and convex as selection continued. Additionally, the difference depending on the weeks was lower than in the HRFI line, as if the selection pressure moved the full trajectory towards lower values.

Thus, selecting for RFI at the level of the test heavily affected the dynamics of longitudinal RFI. This effect seemed more pronounced in the HRFI line.

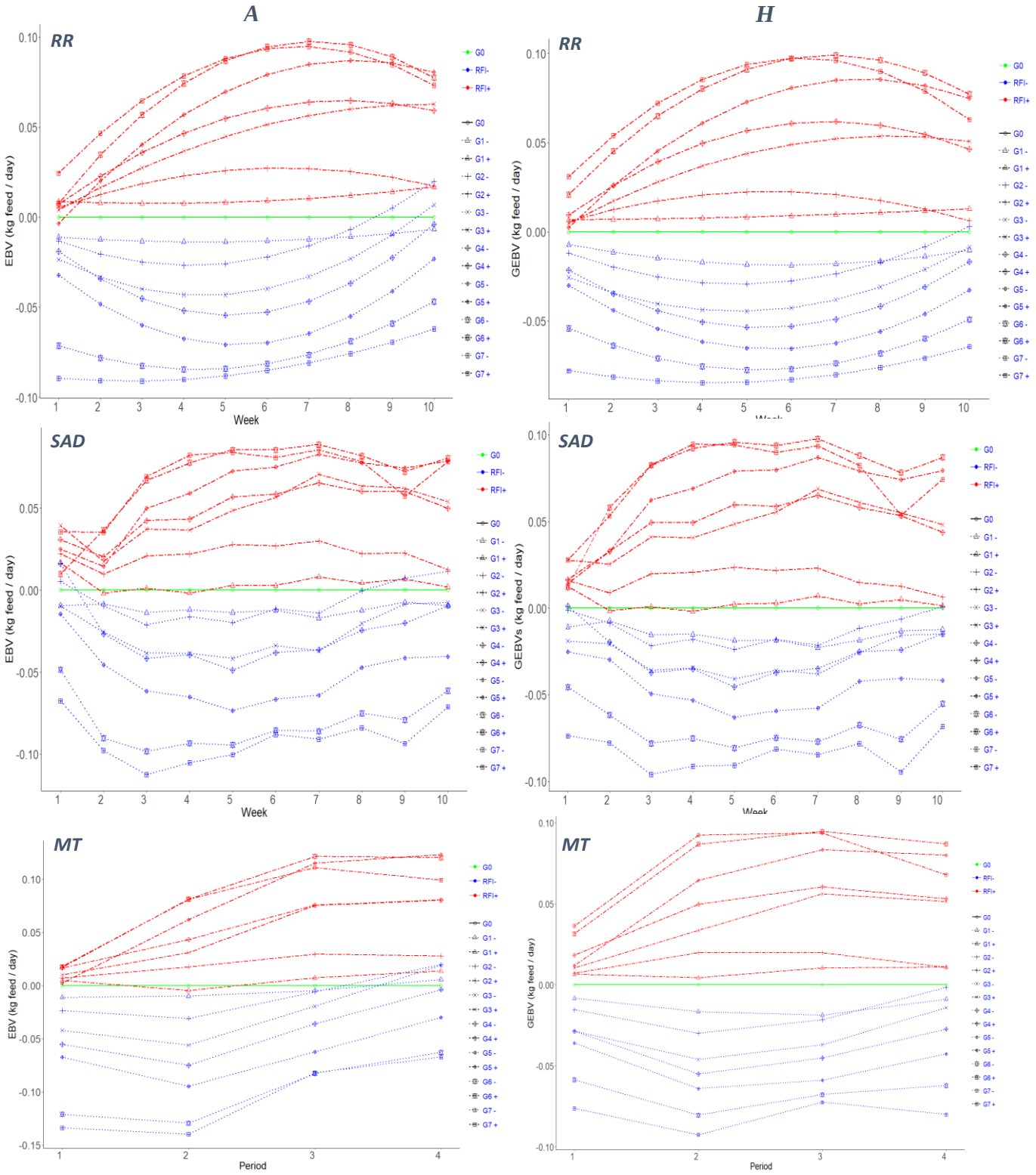


Figure 5.2. Mean RFI (G)EBV per line and generation obtained with a random regression (RR), structured antedenpence (SAD) and multi trait (MT) model for the low RFI (Gx-) and high (Gx+) RFI lines on weekly records, using pedigree-based (EBV, left) or H-based (GEBV, right) relationships matrices. The solid green lines are the trajectories of G0 animals, centered to zero; the blue dotted lines are the trajectories of pigs from generations G1 to G7 in the low RFI line; the red dotted-dashed curves are the trajectories of pigs from generations G1 to G7 in the high RFI line.

5.3. Discussion of the impacts of selection for RFI on the feed efficiency profiles during growth

These two analyses showed that selection based on RFI has an impact on the dynamics of the feed efficiency over time. Further studies in relation with the dynamics of the other major production traits might help to understand the levers explaining these changes of trajectory and how they contributed to the changes of the test RFI. Now that the EBV patterns are described, it would be interesting to consider how to select animals for feed efficiency based on the available longitudinal information (including EBV profiles, SBV, genetic correlations with other traits, etc.). In practice, the individual genetic patterns can be captured into few parameters, as we proposed the use of SBV as a criterion for animal selection. Another approach would be to use a function to summarize the phenotypic profiles into parameters, such as Gompertz performs on BW, and then work on these parameters to select animals with desired profiles. However, before using longitudinal EBV for animal selection, the costs and expected genetic gains on the different components of the objective of selection should be compared between a selection based on one EBV and one based on longitudinal EBV.

5.4. Conclusion

This study showed

i. The changes of EBV patterns for FCR over time for each line-generation combination in a divergent selection experiment for RFI. This proved that selection for RFI has an impact on the dynamics of FCR at the genetic level.

ii. Similar evolutions of EBV and GEBV for longitudinal RFI over time.

Altogether, the current study confirmed that the selection for RFI has an impact on the dynamics of longitudinal feed conversion ratio and residual feed intake.

Chapter 6 - GENERAL DISCUSSION AND PERSPECTIVES

This thesis succeeded in

1. Proposing a linear interpolation to deal with missing weekly body weights, and thus improve the genetic evaluation for feed conversion ratio.

2. Demonstrating the advantages of the SAD models for different traits of large economic importance in pigs (feed conversion ratio, residual feed intake and average daily gain).

3. Deriving a selection criterion that can be applied to any model analyzing repeated data, providing that covariance structures are estimated.

4. Showing the impact of selection for feed efficiency at the level of the test period on feed efficiency trajectories.

5. Estimating the accuracy of RFI and ADG prediction for longitudinal data using genomic information.

6.1. Divergent selection for RFI

As mentioned in the previous chapters, selection for RFI improves feed efficiency with limited impact on growth. A divergent selection for a trait creates a rapid response to selection, which allows evaluating the direct and correlated responses to a criterion for selection (Gilbert et al., 2017). As a counterpart, divergent selection presents some limits such as: i) confounding the results from genetic drift and responses to selection (Gilbert et al., 2017), ii) promoting a restructuration of the genetic variance in the selected population (Melo and Marroig, 2015), iii) changing the physiology of the individuals potentially to extreme extents, and iv) modifying the genetic patterns on other traits (Huynh-Tran et al., 2017).

All results presented in this thesis were obtained on pig lines divergently selected for RFI, a population with a particular genetic structure. Consequently, it could affect some of the results obtained. For addressing this issue, we compared when available our results for each trait, such as longitudinal variance components, to studies on other populations, most being obtained on non-divergent populations. For instance, the heritability estimates for FCR and ADG were in accordance with those obtained on Duroc in Hoque et al. (2009) and Jiao et al. (2014), and on commercial crossbred Large White (Godinho et al., 2018). The genetic parameters for RFI were in line with those reported on a Large White population selected for growth rate on restricted feeding (Nguyen et al., 2005). Thus, it is reasonable to extend our variance component estimations to different populations and to conclude that the estimations of the variance components were not affected by the divergent selection.

However, from the fourth chapter we can suspect that genomic prediction in such small lines heavily submitted to drift requires more data and/or additional effects to allow an accurate prediction of the selected traits. This especially questions the opportunity of genomic prediction across distant lines of limited sizes and different performance levels, in which drift heavily structured the genomes. Further studies would be needed to explore this limit.

6.2. Develop a SAD multiple traits model for residual feed intake

Genetic parameters of RFI were estimated from RFI phenotypes or estimated from ADFI analyses in which ADG, BFT, and AMBW were considered as covariates. Two assumptions were made during these analyses: the partial coefficients of ADG, BFT and AMBW are constant over time, and they are the same for the genetic and environmental components of each covariate. Cai et al. (2011) found that the partial coefficients of BW changed over time on pigs from a divergent selection experiment on RFI at Iowa State University. We tested this hypothesis on our data and found no significant interaction between the partial coefficient and the week of observation. However, we did not investigate if they were the same for the genetic and environmental components. Applying a multiple trait SAD model (David et al. 2017) on ADFI, ADG, AMBW, and BFT would allow testing this hypothesis. In brief, the multiple-trait SAD model would be the following for animal *i* at time t_j:

$$ADG_i(t_j) = \mu_{ADG_i}(t_j) + u_{ADGi}(t_j) + p_{ADGi}(t_j)$$
$$AMBW_i(t_j) = \mu_{AMBW_i}(t_j) + u_{AMBWi}(t_j) + p_{AMBWi}(t_j)$$
$$BFT_i(t_j) = \mu_{BFT_i}(t_j) + u_{BFTi}(t_j) + p_{BFTi}(t_j)$$
$$FI_i(t_j) = \mu_{FI_i}(t_j) + u_{FIi}(t_j) + p_{FIi}(t_j)$$

where effects included in the fixed part $\mu_i(t_j)$ are the combination of the fixed effects of each trait. The random effect functions for genetic and permanent effects for ADG, AMBW and BFT will be according to a "classical" SAD model (i.e. regression on preceding observations), while for ADFI regression on ADG, AMBW and BFT will be included in the model in order to obtain genetic and permanent effects for RFI:

$$\begin{cases} u_{FI}(t_j) = \theta_{FI(t_j)} u_{FI}(t_j - 1) + \delta_{ADG(t_j)} u_{ADG}(t_j) + \beta_{AMBW(t_j)} u_{AMBW}(t_j) + \gamma_{BFT(t_j)} u_{BFT}(t_j) + \varepsilon_{u_{FI}}(t_j) \\ p_{FI}(t_j) = \theta'_{FI(t_j)} p_{FI}(t_j - 1) + \delta'_{ADG(t_j)} p_{ADG}(t_j) + \beta'_{AMBW(t_j)} p_{AMBW}(t_j) + \gamma'_{BFT(t)} p_{BFT}(t_j) + \varepsilon'_{p_{FI}}(t_j) \end{cases}$$

in which $\theta_{FI(t_j)}$, $\varepsilon_{u_{FI}}(t_j)$ are the antedependence parameters and error term for the genetic random function of ADFI at time t_j; $\delta_{ADG(t)}$, $\beta_{AMBW(t_j)}$, $\gamma_{BFT(t_j)}$ are the cross antedependence parameters of time t_j linking genetic components of ADFI to those of ADG, AMBW and BFT, respectively. Similarly, $\theta'_{FI(t_j)}$, $\varepsilon'_{p_{FI}}(t_j)$, $\delta'_{ADG(t_j)}$, $\beta'_{AMBW(t_j)}$, $\gamma'_{BFT(t_j)}$ are the counterparts of permanent environmental effects.

Given that by definition residual feed intake is the feed intake corrected for ADG, AMBW and BFT, we can estimate the genetic and permanent effects for residual feed intake as:

$$\begin{cases} u_{RFI}\left(t_j\right) = u_{FI}\left(t_j\right) - \delta_{ADG\left(t_j\right)} u_{ADG}\left(t_j\right) - \beta_{AMBW\left(t_j\right)} u_{AMBW}\left(t_j\right) - \gamma_{BFT\left(t_j\right)} u_{BFT}\left(t_j\right) = \varepsilon_{u_{FI}}\left(t_j\right) + \theta_{FI\left(t_j\right)} u_{FI}\left(t_j - 1\right) \\ p_{RFI}\left(t_j\right) = p_{FI}\left(t_j\right) - \delta'_{ADG\left(t_j\right)} p_{ADG}\left(t_j\right) - \beta'_{AMBW\left(t_j\right)} p_{AMBW}\left(t_j\right) - \gamma'_{BFT(t)} p_{BFT}\left(t_j\right) = \varepsilon'_{p_{FI}}\left(t_j\right) + \theta'_{FI\left(t_j\right)} p_{FI}\left(t_j - 1\right) \end{cases}$$

The advantage of such model is to be much more flexible than those considering same regression coefficients for the genetic and environmental effect of ADG and BFT on ADFI. This one-step model should provide the best estimate for longitudinal analysis of RFI. Unfortunately, due to time constraints this model could not be tested in the frame of this thesis. It is one of the main model development perspectives to envisage for the longitudinal analysis of this trait.

6.3. Genomic prediction accuracy

The prediction accuracy depends on many factors such as the nature of the trait, the quality of the model, the size of the training dataset and of the validation dataset, as well as the connectedness between them. In chapter 4, we showed that the use of a single step approach with longitudinal data yielded an improved genomic accuracy for ADG but not for RFI. This could be explained by the structure of the current dataset, with two lines under divergent selection for RFI, but with no divergence for ADG. To improve the prediction accuracy in such circumstances, there are many questions related to the choice of the training population regarding to the validation population (Guo et al., 2014). Specifically, the number of individuals in each training and validations sets, and the relatedness between them are main factors that are reported as affecting the prediction accuracy. The impact of genetic drift differentiating populations when using multi-population prediction has been less explored, and could be an extension to this study. The genomic prediction accuracy is also affected by the trait heritability and marker density. Altogether, these potential impacts of the structure of the population suggest testing the single step approach for RFI on other populations not under a divergent selection for RFI, with increased size of the training and validation datasets. In addition, because the RFI lines have now more generations available with genotypes, the study could be repeated on an extended number of individuals genotyped to evaluate the relative impact of these different factors in our results. Additionally, testing the single step for RFI on other populations (not under a divergent selection for RFI) with increased size of training dataset, a validation dataset "large" enough, and with high-density markers, could be conducted in the future. For instance, repeated measurements of ADG and ADFI are available in commercial pig populations evaluated in the test station in Le Rheu, on about 2000 pigs per year.

Finally, more recent approaches proposed in the literature could improve the prediction accuracy, such as weighted single step or weighted Bayesian single – step, in which potential markers of interest are determined before applying the models and given a different weight in the models. However, these methods are ideal for traits with few associated markers. In the case of RFI, most studies reported a large number of SNPs contributing to the genetic variance of RFI (Onteru et al., 2013; Gilbert et al, 2017). Thus, the application of weighted single step might lead to limited improvement of the accuracies.

6.4. Animal selection based on the genetic trajectory

In chapter 3, with SAD model, we demonstrated that it is possible to obtain "enough" information with a reduced test period compared to the test over a 10 week-period. However, the number of records and distance between the measures should be investigated in relation with the objectives of each breeding program.

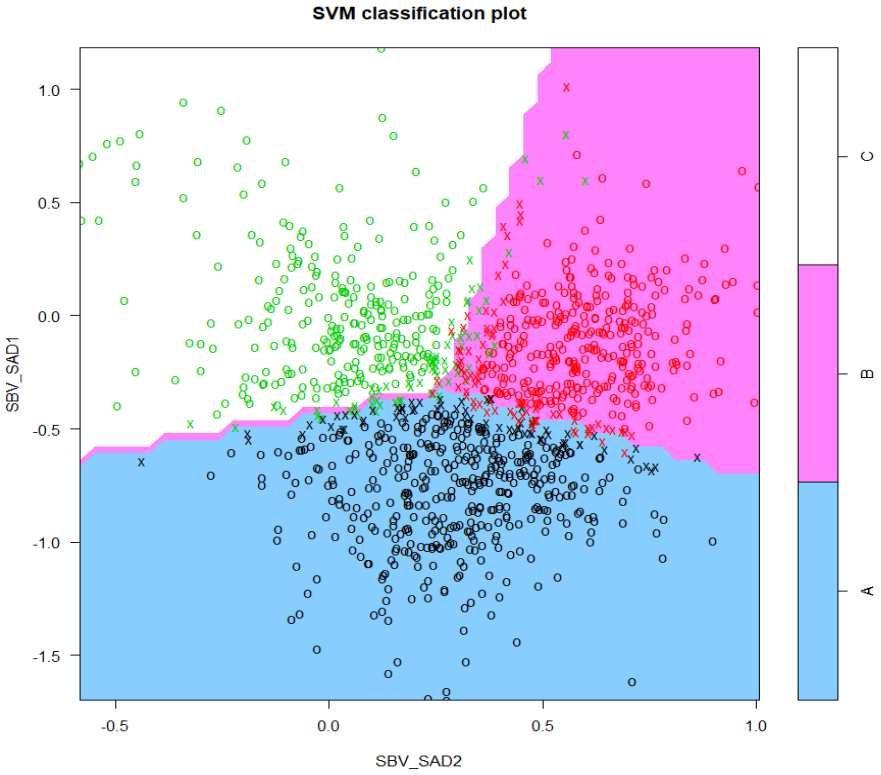


Figure 6.1. An example of using Support Vector Machines (SVM) to establish the boundaries of the trajectory profiles for all individuals, obtained on the joint distribution of their first and second summarized breeding values (SBV) for the structured antedependence model (SBV_SAD1, SBV_SAD2) In this plot, the "×" are the points directly affecting the classification line, so called support vectors. The "o" points do not affect the determination of the boundaries.

The number of EBV obtained for each animal from the longitudinal analysis is equal to the number of time points considered. Until now, there have been several options for animal selection based on longitudinal data. First, using the mean or sum of the predicted EBV over time is an option which is often used for the lactation curve in cattle. The second option is to exploit the whole EBV trajectory and to determine groups of EBV profiles (Savegnago et al., 2016). The best group of individuals have to be identified after a trajectory classification. Once

groups of animals are distinguished, boundaries between the classes can be determined, for instance using a Support Vector Machines (SVM) approach (Winters-Hilt and Merat, 2007). The principle of SVM is to maximize the total distance between the line (boundary) and the closest points in each class. Figure 6.1 presents the results of an SVM approach applied to the trajectory classification based on the EBV obtained with the SAD model for FCR (paper II). It shows that the classification without a priori on the feed efficiency level of the animals nicely separates out the pigs in different groups.

A third option to obtain a criterion to select animals is to compute a total breeding value that corresponds to a linear combination of the weekly EBV. The weights for each week could derivate from an economic study, or from the genetic covariance matrix as proposed in Huynh-Tran et al. (2017b). The objective of including different weights to blend the genetic trajectory should optimize selection and achieve breeding goals faster. However, because we showed that selection for feed efficiency can impact the trajectory patterns of the future generations, the weight of each EBV should be regularly investigated to ensure appropriate responses to selection. Finally, if SBV or other linear combination of longitudinal EBV were used as selection criteria for animal breeding, their accuracy should be evaluated. Given that they are linear combinations of EBV of known accuracy, a derivation for the SBV could be proposed with an equivalent formula for quantifying the prediction accuracy as that usually used in evaluation of the accuracy in EBV prediction.

The individual EBV trajectories obtained with RR models can be described by polynomials functions, as defined by the model. The individual trajectories of other models are by essence less smooth, but they could also be modelled with polynomial functions. If a function was available, we could summarize the EBV of each animal in a simple function with limited number of parameters, particularly in cases with large number of time points per animal. However, compared to summarized breeding values, this approach would potentially capture all the changes of the trajectories without selection of the useful ones (Huynh-Tran et al., 2017b). Thus, it would require thorough evaluation to avoid undesired responses to selection.
The combination of trajectories of different traits is finally a more ambitious possibility. It could allow joint improvement of several criteria at once, resulting in faster achievement of multiple goals.

REFERENCES

Cai, W., M. S. Kaiser, and J. C. M. Dekkers. 2011. Genetic analysis of longitudinal measurements of performance traits in selection lines for residual feed intake in Yorkshire swine. J. Anim. Sci. 89:1270–1280. doi:10.2527/jas.2010-3107.

David, I., H. Garreau, E. Balmisse, Y. Billon, and L. Canario. 2017. Multiple-trait structured antedependence model to study the relationship between litter size and birth weight in pigs and rabbits. Genet. Sel. Evol. 49:11. doi:10.1186/s12711-017-0288-3.

Gilbert, H., Y. Billon, L. Brossard, J. Faure, P. Gatellier, F. Gondret, E. Labussière, B. Lebret, L. Lefaucheur, N. Le Floch, I. Louveau, E. Merlot, M.-C. Meunier-Salaün, L. Montagne, P. Mormede, D. Renaudeau, J. Riquet, C. Rogel-Gaillard, J. van Milgen, A. Vincent, and J. Noblet. 2017. Review: divergent selection for residual feed intake in the growing pig. Anim. Int. J. Anim. Biosci. 1–13. doi:10.1017/S175173111600286X.

Godinho, R. M., R. Bergsma, F. F. Silva, C. A. Sevillano, E. F. Knol, M. S. Lopes, P. S. Lopes, J. W. M. Bastiaansen, and S. E. F. Guimarães. 2018. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. J. Anim. Sci. 96:817–829. doi:10.1093/jas/skx011.

Guo, Z., D. M. Tucker, C. J. Basten, H. Gandhi, E. Ersoz, B. Guo, Z. Xu, D. Wang, and G. Gay. 2014. The impact of population structure on genomic prediction in stratified populations. TAG Theor. Appl. Genet. Theor. Angew. Genet. 127:749–762. doi:10.1007/s00122-013-2255-x.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45–49. doi:10.1016/j.livsci.2008.05.016.

Huynh-Tran, V. H., I. David, Y. Billon, and H. Gilbert. 2018. Changes of EBV trajectories for feed conversion ratio of growing pigs due to divergent selection for residual feed intake. Proc. 11th World Congr. Genet. Appl. Livest. Prod. P 513 Present. 11 World Congr. Genet. Appl. Livest. Prod. WCGALP Auckl. NZL 2018-02-12 - 2018-02-16 NZL Massey Univ. 1130.

Huynh-Tran, V. H., H. Gilbert, and I. David. Nov 2017b. Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs. J. Anim. Sci. 95:4752–4763. doi:10.2527/jas2017.1864.

Melo, D., and G. Marroig. 2015. Directional selection can drive the evolution of modularity in complex traits. Proc. Natl. Acad. Sci. U. S. A. 112:470–475. doi:10.1073/pnas.1322632112.

Nguyen, N. H., C. P. McPhee, and C. M. Wade. 2005. Responses in residual feed intake in lines of Large White pigs selected for growth rate on restricted feeding (measured on ad libitum individual feeding). J. Anim. Breed. Genet. Z. Tierzuchtung Zuchtungsbiologie. 122:264–270.

Onteru, S. K., D. M. Gorbach, J. M. Young, D. J. Garrick, J. C. M. Dekkers, and M. F. Rothschild. 2013. Whole Genome Association Studies of Residual Feed Intake and Related Traits in the Pig. PLOS ONE. 8:e61756. doi:10.1371/journal.pone.0061756.

Savegnago, R. P., G. B. do Nascimento, G. J. de M. Rosa, R. L. R. de Carneiro, R. C. Sesana, L. El Faro, and D. P. Munari. 2016. Cluster analyses to explore the genetic curve pattern for milk yield of Holstein. Livest. Sci. 183:28–32. doi:10.1016/j.livsci.2015.11.010.

Winters-Hilt, S., and S. Merat. 2007. SVM clustering. BMC Bioinformatics. 8:S18. doi:10.1186/1471-2105-8-S7-S18.

TRAINING COURSES AND SEMINARS

Training courses

- ✓ Programming C++ (28h)
- ✓ Ecole international de recherche-Agreenium (EIR-A) (3 days, 2017, Rennes)
- ✓ Ecole international de recherche-Agreenium (EIR-A) (5 days, 2016, Montpellier)
- ✓ CSAGAD, quantitative genetics courses (sessions 1+2), AGROPARISTECH, Paris
- ✓ Inkscape (1/2 day)
- ✓ Advance quantitative genetics for animal breeding course by Pr. Toro (GenPhySE, 2017) (28h)
- ✓ English for scientific presentation (Toulouse, 2016) (28h)
- ✓ Scientific French (28h)

Seminars

- ✓ Doctoral seminar of Animal Genetic Division, INRA, Toulouse (2 days, March 2016)
- ✓ Doctoral seminar of Animal Genetic Devison, INRA, Rennes (2 days, May 2017)
- ✓ Doctoral seminar of Animal Genetic Devison, INRA, Les Mureaux (2 days, May 2018)
- ✓ Doctoral school seminar SEVAB 2016
- ✓ Doctoral school seminar SEVAB 2017

International collaboration

USA: Animal genetics department, Iowa state university (from Nov 2017 to Feb 2018)

Supervisor: Dr. Jack C.M. Dekkers

Project: “Pedigree and genomic predictions of longitudinal data”

SCIENTIFIC COMMUNICATION

PUBLICATIONS

1. **Huynh-Tran, V. H**., H. Gilbert and I. David. 2017. Genetic structured antedependence and random regression models to study longitudinal feed conversion ratio in growing French Large White pigs. J. Anim. Sci. Doi : 10.2527/jas2017.1864

2. **Huynh-Tran, V. H**., H. Gilbert and I. David. 2017. How to improve breeding value prediction for feed conversion ratio in the case of incomplete longitudinal body weights. J. Anim. Sci. 95:39–48. doi:10.2527/jas.2016.0980.

3. **Huynh-Tran, V. H**., J.C.M. Dekkers, I. David, H. Gilbert. Pedigree and genomic predictions of longitudinal data for residual feed intake and average daily gain in growing pigs (in preparation)

SCIENTIFIC CONGRESSES

1. **Huynh-Tran, V. H**., H. Gilbert and I. David, 2018. Changes of EBV trajectories for feed conversion ratio of growing pigs due to divergent selection for residual feed intake (WCGAPL 2018, Aukland, New Zealand)

2. **Huynh-Tran, V. H**., H. Gilbert and I. David, 2016. Gompertz model improves breeding value prediction for feed conversion ratio for incomplete weights (oral presentation, EAAP 2016, Belfast, UK)