COMMUNAUTE FRANCAISE DE BELGIQUE
UNIVERSITE DE LIEGE – GEMBLOUX AGRO-BIO TECH

Innovative algorithms to combine phenotypic, genealogical and genomic information originating from diverse sources

Jérémie VANDENPLAS

Essai présenté en vue de l'obtention du grade
de docteur en sciences agronomiques et ingénierie biologique

Promoteur: Nicolas Gengler

2014

COMMUNAUTE FRANCAISE DE BELGIQUE
UNIVERSITE DE LIEGE – GEMBLOUX AGRO-BIO TECH

Innovative algorithms to combine phenotypic, genealogical and genomic information originating from diverse sources

Jérémie VANDENPLAS

Essai présenté en vue de l'obtention du grade
de docteur en sciences agronomiques et ingénierie biologique

Promoteur: Nicolas Gengler

2014

Vandenplas Jérémie. (2014). Innovative algorithms to combine phenotypic, genealogical and genomic information originating from diverse sources. (PhD Dissertation in English). Gembloux, Belgium, Gembloux Agro-Bio Tech, University of Liege, 190p., 27 tabl., 3 fig.

Abstract

Along with technical developments, international exchanges of genetic material (e.g., frozen semen, embryos) have increased since the 1970s. However, genetic evaluations are traditionally based on phenotypic and genealogical data which are internally recorded, i.e., within well defined borders. Because imported (i.e., external) genetic material is usually strongly selected in their respective populations, internal genetic evaluations for external animals could be biased and less accurate if external data used for their selection is ignored. Moreover, comparison of internal and external animals based on their internal and external estimates of genetic merit is needed to select and potentially import the most suitable ones according to the internal breeding goal. However, such comparison is usually not possible among internal and external genetic evaluations due, e.g., to differences among units of measurement. Thereby, several approaches and algorithms have been developed to render internal and external genetic evaluations comparable, and to combine or blend phenotypic and genealogical data and external information, i.e., estimates of genetic merit and associated reliabilities. Furthermore, the recent development of genomic selection also increased needs for combining phenotypic, genealogical and genomic data and information. Therefore, the aim of this thesis was first to develop innovative algorithms to combine diverse sources of phenotypic, genealogical and genomic data and information, and second to test them on simulated and real data in order to check their correctness. Based on a Bayesian view of the linear mixed models and addressing several issues highlighted by previous studies, systems of equations combining simultaneously diverse sources of data and external information were developed for (multivariate) genetic and single-step genomic evaluations. Double counting of contributions due to relationships and due to records were considered as well as computational burden. The performances of the developed systems of equations were evaluated using simulated datasets and real datasets originating from genetic (genomic) evaluations for Holstein cattle and for show jumping horses. The different results showed that the developed equations integrated and blended several sources of information in a proper way into a genetic or a single-step genomic evaluation.

It was also observed that double counting of contributions due to relationships and due to records was (almost) avoided. Furthermore, more reliable estimates of genetic merit were also obtained for external animals and for their relatives after integration of external information. Also, the developed equations can be easily adapted to complex models, such as multivariate mixed models. Indeed, it was shown that external information correlated to the internal phenotypic traits was properly integrated using the developed equations. Finally, research of this thesis led to the development of a genomic evaluation system for Holstein cattle in the Walloon Region of Belgium for production traits, as well as for other traits, like somatic cell score. Based on the research of this thesis, future research topics, e.g., concerning integration of correlated external information and of genomic information, were finally presented.

Vandenplas Jérémie. (2014). Algorithmes innovants pour combiner des informations phénotypiques, généalogiques et génomiques provenant de différentes sources. (Thèse de doctorat en anglais). Gembloux, Belgique, Gembloux Agro-Bio Tech, Université de Liège, 190p., 27 tabl., 3 fig.

Résumé

Suite aux progrès techniques, les échanges internationaux de matériel génétique (par exemple, la semence congelée ou les embryons), ont augmenté depuis les années 1970. Toutefois, les évaluations génétiques sont traditionnellement basées sur des données phénotypiques et généalogiques qui sont enregistrées à un niveau interne, c'est-à-dire dans des frontières bien définies. Parce que le matériel génétique importé (appelé ci-après externe) est habituellement fortement sélectionné dans leurs populations respectives, les évaluations génétiques internes pour les animaux externes pourraient être biaisées et moins précises si les données externes utilisées pour leur sélection sont ignorées. En outre, la comparaison des animaux internes et externes en fonction des estimations internes et externes de leurs valeurs génétiques est nécessaire pour sélectionner et, potentiellement, importer les plus appropriés en fonction de l'objectif de reproduction interne. Cependant, une telle comparaison n'est généralement pas possible entre les évaluations génétiques internes et externes en raison, par exemple, des différences entre les unités de mesure utilisées pour mesurer les phénotypes. Ainsi, plusieurs approches et algorithmes ont été développés pour rendre comparables des évaluations génétiques internes et externes, ou pour combiner des données phénotypiques et généalogiques ainsi que de l'information externe, c'est-à-dire les estimations de valeurs génétiques et les fiabilités associées. De plus, l'évolution récente de la sélection génomique augmente également les besoins de combinaisons de données phénotypiques et généalogiques et d'informations génomiques. Par conséquent, l'objectif de cette thèse a été, premièrement, de développer des algorithmes innovants pour combiner diverses sources de données et d'informations phénotypiques, généalogiques et génomiques et, deuxièmement, de tester ces algorithmes sur des données réelles et simulées afin de vérifier leur exactitude. Fondée sur une vision bayésienne des modèles mixtes linéaires et reposant sur plusieurs questions soulevées par des études précédentes, des systèmes d'équations combinant simultanément diverses sources de données et d'informations externes ont été élaborés pour des évaluations génétiques et génomiques de type « single-step », potentiellement multi-caractères. Les doubles comptages de contributions dus aux liens de parenté entre

les animaux externes et dus aux données ainsi que la charge de calcul ont été examinés. Les performances des systèmes d'équations développés ont été évaluées en utilisant des jeux de données simulées et des données réelles provenant des évaluations génétiques (génomiques) pour les bovins Holstein et pour les chevaux de saut d'obstacle. Les différents résultats ont montré que les équations développées intègrent et combinent plusieurs sources d'information d'une manière appropriée pour les évaluations génétiques et génomiques de type « single-step ». Il a également été observé que les doubles comptages des contributions dus aux liens de parenté entre animaux externes et dus aux données étaient (presque) évités. En outre, des estimations plus fiables de valeurs génétiques ont également été obtenues pour les animaux externes et pour les animaux qui leur sont apparentés après l'intégration de l'information externe. De plus, les équations développées peuvent être facilement adaptées à des modèles complexes, tels que les modèles mixtes multi-caractères. En effet, il a été montré que l'information externe corrélée avec les caractères phénotypiques internes est bien intégrée en utilisant les équations développées. Enfin, la recherche de cette thèse a conduit à la mise en place d'un système d'évaluation génomique pour bovins Holstein en Région Wallonne (Belgique) pour, notamment, les caractères de production. Suite aux recherches menées lors de cette thèse, des sujets de recherche futurs, par exemple, concernant l'intégration d’informations externes corrélées et d'informations génomiques, ont finalement été présentés.

ACKNOWLEDGMENTS

While I was finishing my Master thesis in 2010, Prof. Nicolas Gengler (University of Liege, Gembloux Agro-Bio Tech) proposed me to continue a research on the same topic through a PhD. Accepting this opportunity was the right thing to do for me because, during these last years, I travelled a lot, I met a lot of interesting people and, of course, I learned a lot about animal breeding and genetics, statistics, maths and programming. Therefore, I would like to acknowledge first my supervisor, Prof. Nicolas Gengler, for having proposed me this opportunity.

All travels, knowledge and achievement of this PhD would not have been possible without the help of many other people. Thereby, I also acknowledge all the thesis committee members, Dr Steven Janssens (Katholieke Universiteit Leuven), Prof. Yves Beckers, Prof. Yves Brostaux, Prof. Rodolphe Palm, Prof. André Théwis and Prof. Nicolas Gengler (all from University of Liege, Gembloux Agro-Bio Tech) for their advises and encouragement. I am also grateful to each member of the jury who read and reviewed this PhD manuscript. The thesis jury included Prof. Frédéric Francis (Chair of the jury, University of Liege, Gembloux Agro-Bio Tech), Dr Ole F. Christensen (Aarhus University), Dr Steven Janssens, Prof. Ignacy Misztal (University of Georgia), Prof. Yves Beckers, Prof. Yves Brostaux, and Prof. Nicolas Gengler. Special acknowledgments are addressed to Prof. Yves Brostaux and Dr Ole F. Christensen for reviewing the first draft of this manuscript and for their helpful comments. I would also like to address my deepest gratitude to Dr Catherine Bastin (University of Liege, Gembloux Agro-Bio Tech), Dr Steven Janssens and Marie-Laure Vanrobays for their revision of this manuscript. I also address my acknowledgments to the following people who took time for revising some of my manuscripts: Dr Catherine Bastin, Dr Bernd Buske, Dr Frédéric Colinet, Dr Pierre Faux, Marie-Laure Vanrobays (all from University of Liege) and Ryan Ann Davis (University of Georgia).

I would like to address special thanks to all people who minutely answered to my questions regarding animal breeding and genetics. There are numerous scientists, and among them, I would like to personally acknowledge Dr Ole F. Christensen, Dr Steven Janssens, Dr Andres Legarra (French National Institute for Agricultural Research), Dr Han Mulder (Wageningen UR) and Dr Paul VanRaden (United States Department of Agriculture). Thank you for all the time you took to answer to my questions and for

sharing your knowledge. However, as said before, I also learned a lot about statistics, maths and programming. All the accomplishments would not have been possible without the help of numerous people. In addition to people previously mentioned, I would extend my thanks to Dr Ignacio Aguilar (Instituto Nacional de Investigación Agropecuaria), Dr Mark Bull (University of Edinburgh), Prof. Catherine Charles (University of Liege, Gembloux Agro-Bio Tech), Dr David Colignon (University of Liege), Dr François Damien (University catholique de Louvain), Dr Pierre Faux, Dr Klaus Gaertner (Weierstrass Institute for Applied Analysis and Stochastics), Prof. George Karypis (University of Minnesota), Prof. Ignacy Misztal and Dr Shogo Tsuruta (University of Georgia). Also, most of research would not have been possible without the different people who provided me datasets. I especially acknowledge Danielle Arts (Royal Dutch Sporthorse Studbook), Dr Steven Janssens, and Dr Anne Ricard (French National Institute for Agricultural Research), Dr Gregor Gorjanc and Marija Špehar (University of Ljubljana), as well as Dr Catherine Bastin, Dr Frédéric Colinet, Marie Dufrasne, Dr Hedi Hammami, Sylvie Vanderick and Marie-Laure Vanrobays (all from University of Liege, Gembloux Agro-Bio Tech).

My gratitude also goes to Dr Gregor Gorjanc, Prof. Ignacy Misztal and Dr Han Mulder for organizing my scientific stays at the University of Ljubljana, at the University of Georgia and at Wageningen UR, respectively. These stays were very fruitful on several levels. The success of these scientific stays was also made possible thanks to all the members of the different research teams, and I am especially grateful to El Hamidi Abdel Hay, Daniela Lourenco, Marija Špehar and Joy (H.) Wang.

For the achievement of this PhD, several financial supports were required. I would like to acknowledge the financial support:

- of the National Fund for Scientific Research (FRS-FNRS, Brussels, Belgium) for my position as a Research Fellow since October 2011,
- of the National Fund for Scientific Research, of the University of Liege and of Wallonie-Bruxelles International through grants provided for scientific stays, and
- from the European Commission, Directorate-General for Agriculture and Rural Development, under Grand Agreement 211708 and from the Commission of the European Communities (FP7, KBBE-2007-1) through the RobustMilk project, and of the Ministry of Agriculture of Walloon Region of Belgium (Service Public de Wallonie, Direction générale opérationnelle “Agriculture, Ressources naturelles et Environnement” – DGARNE) through different projects for research conducted

during this PhD.

I am grateful to all the members of the Animal Science Unit of Gembloux Agro-Bio Tech (University of Liege) who helped, advised and encouraged me during these last years. Thank you for your friendship and the friendly atmosphere.

The achievement of this PhD was also made possible thanks to all the after-work activities. They were numerous and were held in many places, like Athens, Bratislava, Copenhagen, Gembloux and Phoenix (to name a few). Because I am afraid to forget someone, I will no longer mention names. However, thank you all of you for all these great memories. The concerned people will recognize themselves.

Final and special thanks are addressed to all the people who were not concerned by this PhD but who accompanied me during these last years. Especially, I would like to acknowledge Eric and his family, Didier, François and Marc for their friendship. Working and discussing with you about other things than matrices, maths or genetics is very interesting. I would also like to acknowledge my parents, brothers and sister, and all the other members of my family for their encouragements and advices. Finally, my greatest thanks are addressed to Hilde who supports me all these years and to André who pointed his finger at my computer when I was distracted during the writing of this manuscript.

To conclude, my first idea for the acknowledgments was to write something short and simple and what I would really like to say to everybody. However, it is not the traditional way to act like this. Therefore, I will finish the acknowledgments as I thought first to write them: thank you everybody for your help, advices, encouragements, teaching and friendship during all these years, related or not to this PhD. And I am sure this will not take an end with the achievement of my PhD thesis.

CONTENTS

Chapter I. GENERAL INTRODUCTION

CONTEXT

One of the major objectives in animal breeding is to estimate the genetic merit of animals for traits of interest in order to rank them. Based on these rankings, the best ones can be selected and mated to finally generate an improved generation with optimal performances regarding these traits of interest. Genetic merits can be estimated through different statistical methods, such as selection index (SI; Hazel, 1943), mixed models (Henderson, 1984) or Bayesian methods (Gianola and Fernando, 1986). Under some assumptions, these statistical methods allow to animal breeders to obtain estimates of genetic merit (EGM) either for animals (e.g., estimated breeding values; EBV) or for what they transmit to their progeny (e.g., expected progeny differences, predicted transmitting abilities). All these methods aim to achieve the highest reliability (REL) for the EGM with regards to available data. Nevertheless, whatever method used, the aim is the same, i.e., ranking animals to choose the best ones in order to generate improved progeny.

Whereas SI, mixed models and Bayesian methods could give equivalent results under some assumptions, the linear mixed models developed by Henderson (e.g., Henderson, 1984) are commonly used for genetic evaluations since the 1970s, replacing the SI theory popularized by Hazel (Powell and Norman, 2006). Several properties of the linear mixed models can explain their widespread use. A first property of linear mixed models is to differentiate the effects between “fixed” and “random” effects, following the frequentist school. These effects have “Best Linear Unbiased Estimate” (BLUE) and “Best Linear Unbiased Prediction” (BLUP) properties, respectively. Other properties of linear mixed models are their easy adaptation to non-normally distributed data, their potential application in a Bayesian context, or their easy understanding (Robinson, 1991; VanRaden, 2001; Powell and Norman, 2006).

Traditionally, phenotypic and genealogical data are recorded following official recording schemes performed on populations within well defined borders (e.g., country borders). Based on these data, genetic evaluations are performed for these populations. Such genetic evaluations are hereafter called “internal” genetic evaluations because they are only based on internal data, i.e., collected within established borders. Internal genetic evaluations are characterized by their own scale, units of measurement and/or genetic bases. However, since the 1970s, technical developments, like frozen semen or embryos, increase exchanges of genetic material worldwide, leading to internationalization of

breeding schemes and breeds. Within well defined borders, the internal population might have a large proportion of genes from external populations. Such situations are especially observed in sport horse (Arnason, 2013) and dairy cattle breeding (Philipsson, 1987). Hence, different issues may arise if genetic material is widespread outside borders while genetic evaluations are performed within borders. Firstly, because imported (i.e., external) genetic material is usually strongly selected, internal EGM for external animals could be biased with an internal genetic evaluation if external data used for their selection in external populations is ignored (VanRaden, 2012). Nevertheless, although one of the major objectives in animal breeding is to internally predict genetic merits of animals with the highest REL by using all available data, internal genetic evaluations are usually performed using only data collected internally. Several reasons leading to internal genetic evaluations that ignore external data are mentioned below. Secondly, selection and importation of the most suitable external genetic material according to the internal breeding needs and goals require the comparison of animals through their own EGM and REL. However, such a comparison is usually not possible between internal and external populations due to differences among scales, units of measurement and genetic bases of genetic evaluations (Weigel and Rekaya, 2000).

One way to solve both issues is to use simultaneously all available phenotypic and genealogic data, i.e., from all concerned populations, to get unbiased EGM through a joint genetic evaluation. Some studies, for example in sport horse breeding (e.g., Furre et al., 2013) and in dairy cattle breeding (e.g., Banos et al., 1992; Weigel and Rekaya, 2000), showed results of joint genetic evaluations. However, usually, joint genetic evaluations cannot be performed because data from the different populations are not available in the same dataset for several reasons, like political roadblocks, or because data cannot be merged due to inconsistencies. Moreover, even if data can be combined in the same dataset, joint genetic evaluations could not be performed due to computing or logistical problems (Powell and Sieber, 1992). Nevertheless, comparison of genetic material is still needed. To make it feasible, instead of performing joint evaluations based on the combination of raw data, that are mostly unavailable, genetic merits can be approximated by converting or combining the available information, i.e., EGM and associated REL obtained for each population. Therefore, different approaches and algorithms converting or combining EGM and their associated REL across populations were developed over the years to improve accuracies of internal genetic evaluations and to render genetic merits of animals comparable across populations in order to select the most appropriate genetic

material in a widespread pool of genes. These developed approaches and algorithms were mainly derived from SI theory (e.g., VanRaden, 2001), mixed models methodology (e.g., Schaeffer, 1985) and Bayesian statistics (e.g., Gianola and Fernando, 1986).

For a few decades, molecular data at the deoxyribonucleic acid (DNA) level have been considered in genetic evaluations as an additional source of data that permits to improve genetic progress through both an increase of accuracy of selection and a decrease of generation intervals. Selection based on molecular information was first based on molecular genetic marker information (e.g., microsatellites) and was called marker-assisted selection (Fernando and Grossman, 1989). In 2001, Meuwissen et al. (2001) proposed to use genome-wide dense marker maps including several thousands of single nucleotide polymorphisms (SNP) to estimate genetic merits of animals. This led to the recent massive development of the so-called genomic selection in many species. The increasing availability of bi-allelic SNP data and the subsequent increasing amount of information derived from this data source (i.e., genomic EGM and associated REL) highlighted the necessity to develop approaches and algorithms for combining sources of phenotypic, genealogical and genomic data and information.

AIM OF THE THESIS

The aim of this thesis was to develop innovative algorithms to combine phenotypic, genealogical as well as genomic data and information originating from diverse sources and to test them on simulated and real data in order to check their correctness.

THESIS OUTLINE

This thesis is a compilation of published scientific papers proposing algorithms that combine different sources of data and information and investigating their use in simulated and real contexts. Firstly, a literature review of the different approaches and algorithms that render EGM and associated REL comparable or to combine them is provided in Chapter II. Then, a detailed comparison of different Bayesian approaches integrating external information into genetic evaluations is provided (Chapter III). Based on this comparison, some improvements are proposed (Chapter III), mainly to limit computational burden and to avoid double counting of contributions due to relationships. In Chapter IV, the resulting improved Bayesian approach is implemented in the context of the Belgian genetic evaluation for jumping horses. Bayesian approaches require

alterations of expectations and of (co)variances for random effects of linear mixed models. However, most available software packages based on linear mixed models used in animal breeding do not allow for such alterations. Therefore, a method is proposed to allow for those alterations while using available software packages (Chapter V). This method is based on the use of an extended data file and a user supplied (co)variance matrix (Chapter V). In Chapter VI, a unified method integrating and blending several sources of information into a genetic evaluation is developed and tested on simulated and real data. In addition to integrate and blend several sources of information, the developed method allows to take into account double counting of contributions due to records. An implementation of this latter method is the Walloon single-step genomic evaluation integrating Walloon and multiple across country evaluation (MACE) information, and is presented in Chapter VII. Chapter III to Chapter VII show improvements and implementations of Bayesian approaches that integrate several sources of external information into an internal genetic or single-step genomic evaluation. These investigations are performed in a context where internal and external information were provided for the same trait, although these approaches were developed to integrate correlated external information, i.e., to integrate external information from a certain trait correlated to the internal phenotype traits. Therefore, the first part of Chapter VIII presented the results of a study which integrates correlated external information into an internal multivariate evaluation for a simulated case. A comparative study among the different approaches that combine simultaneously external information and internal data is then detailed in the second part of Chapter VIII. Finally, implications, future research topics and a general conclusion are presented in Chapter IX.

THESIS FRAMEWORK

The research of this thesis was initiated in October 2010. The first academic year 2010-2011 was mainly dedicated to the NovaUdderHealth project financed by the Ministry of Agriculture of Walloon Region of Belgium (Service Public de Wallonie, Direction générale opérationnelle “Agriculture, Ressources naturelles et Environnement” – DGARNE) and to the FP7 European project RobustMilk. Since October 2011, this research has been supported by a fellowship (“Research Fellow”) funded by the National Fund for Scientific Research (FRS-FNRS, Belgium).

REFERENCES

Arnason, T. 2013. Bright future for research in horse breeding! *J. Anim. Breed. Genet.* 130:167–169.

Banos, G., G.R. Wiggans, and J.A.B. Robinson. 1992. Comparison of methods to rank bulls across countries. *J. Dairy Sci.* 75:2560–2568.

Fernando, R.L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467–477.

Furre, S., A. Viklund, J. Philipsson, B. Heringstad, and O. Robison. 2013. Joint genetic evaluation of Nordic sporthorses. Page 358 in Book of abstracts of the 64th annual meeting of the European Association for Animal Production, Nantes, France. Wageningen Academic Publishers, Wageningen, The Netherlands.

Gianola, D., and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63:217–244.

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. *Genetics* 28:476–490.

Henderson, C.R. 1984. Applications of linear models in animal breeding. 2nd ed. University of Guelph, Guelph, ON, Canada.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

Philipsson, J. 1987. Standards and procedures for international genetic evaluations of dairy cattle. *J. Dairy Sci.* 70:418–424.

Powell, R.L., and H.D. Norman. 2006. Major advances in genetic evaluation techniques. *J. Dairy Sci.* 89:1337–1348.

Powell, R.L., and M. Sieber. 1992. Direct and indirect conversion of bull evaluations for yield traits between countries. *J. Dairy Sci.* 75:1138–1146.

Robinson, G.K. 1991. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* 6:15–32.

Schaeffer, L.R. 1985. Model for international evaluation of dairy sires. *Livest. Prod. Sci.* 12:105–115.

VanRaden, P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. *J. Dairy Sci.* 84 (E. Suppl.):E47–E55.

VanRaden, P.M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. *Interbull Bull.* 45:29–33.

Weigel, K.A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J. Dairy Sci.* 83:815–821.

Chapter II. STRATEGIES TO RENDER COMPARABLE AND TO COMBINE RESULTS FROM DIFFERENT GENETIC AND GENOMIC EVALUATIONS: A REVIEW

Exchange of genetic materials among populations at an (inter)national level rapidly increased with the development of artificial insemination and frozen embryos, leading to an increasing necessity to render comparable or to combine estimates of genetic merit (e.g., estimated breeding values) and their associated reliabilities provided for the various populations. Combination of different sources of information became even more crucial with the development of genomic evaluations. Therefore, the objective of this Chapter was to review different approaches and algorithms developed in order to solve issues regarding comparison and combination of several genetic and genomic evaluations.

INTRODUCTION

The issue of comparing or combining estimates of genetic merit (EGM) and associated reliabilities (REL) arose from the first exchanges of genetic material among populations. Populations may be considered at a farm level (e.g., Henderson, 1975) or at a country level (e.g., Schaeffer, 1985). Different approaches and algorithms were developed to solve this issue and the objective of this Chapter is to review them. All reviewed approaches and algorithms were sorted following two strategies. Firstly, EGM and associated REL provided for external populations, hereafter called external information, can be rendered comparable or combined with internal EGM and associated REL after performing internal and external evaluations. These post evaluation approaches were described hereafter in the subsection "Post evaluation approaches". Secondly, external information provided for external populations can be combined simultaneously with internal phenotypic and genealogic data in internal genetic evaluations performed for internal populations. These approaches were described hereafter in the subsection "Simultaneous combinations". Also, it was noted that most of the reviewed approaches and algorithms were developed for (multi-breed) genetic evaluations in dairy and beef cattle. With the advent of genomic selection, needs to combine genomic information with phenotypic and genealogical data and information have appeared, and several previous approaches were adapted. Therefore, a subsection is assigned to approaches adapted and developed in the context of genomic selection.

POST EVALUATION APPROACHES

CONVERSION EQUATIONS

The oldest and simplest approach to render EGM and associated REL comparable across two populations (e.g., from two countries) is the use of a regression-based conversion equation which converts EGM from one exporting (i.e. external) population to the scale, units of measurement and genetic base of an importing (i.e. internal) population. In the context of dairy cattle, the first formula of conversion equations recommended by the International Dairy Federation in 1981 (Gravert, 1983) was of the form:

$$\mathbf{y} = a\mathbf{1} + b\mathbf{x}$$

where $\mathbf{y}$ is the vector of internal EGM, $\mathbf{x}$ is the vector of external EGM, a is the intercept and b is the slope of the conversion equation.

The intercept a can be considered as the difference in genetic base between the two populations. It is valid as long as the genetic bases of the two populations are fixed or are moving at the same rate (Philipsson, 1987). If equal REL of EGM are observed in both populations (i.e., the same number of observations and the same heritability) and if no genotype x environment interaction exists (i.e., genetic correlations between the two populations is equal to 1), the slope b is equal to the ratio of the standard deviations in genetic merit as expressed in the two populations (Wilmink et al., 1986; Philipsson, 1987; Powell and Sieber, 1992). The slope b can be considered as the relationships between scales and definitions of EGM (e.g., estimated breeding values (EBV) or predicted transmitting abilities (PTA)) of the two populations, i.e., as a scaling factor (Powell, 1988; Powell and Sieber, 1992).

However, the previous conditions are usually not fulfilled in practice and several approaches (e.g., Goddard, 1985; Philipsson et al., 1986; Wilmink et al., 1986; Powell, 1988) were proposed to estimate the intercept a and the slope b. These approaches also take into account the REL associated with the EGM from each population. In this context, the slope b also considers genetic correlations among populations that could be lower than 1 due to different heritabilities and definitions of traits (Philipsson, 1987). Approaches were also proposed to approximate REL associated with converted EGM depending on REL associated with external EGM, on genetic correlations among populations and on accuracy of conversion equations (Goddard, 1985; Powell et al., 1994). It is noted that genetic correlations lower than 1 as well as preferential treatments are mainly responsible for the non-reciprocity of the conversion equations (Powell et al., 1994).

To compare different conversion equations, Philipsson et al. (1986) defined desirable properties. Thus, methods should 1) give unbiased estimates of both the intercept a and the slope b, 2) consider the difference in REL from each population, 3) allow for the possibility for a genetic correlation less than 1 between the true genetic merits in each population, and 4) minimize the variance of differences between converted EGM and true values in the external population. Because accuracy of the conversion equations is influenced by preferential matings of external animals, by preferential treatments for some animals, and by suitability of animals selected for the estimation of a and b (Powell et al., 1994), Wickham and Philipsson (1990a) proposed recommendations for the estimation of both the intercept a and the slope b in the context of dairy cattle.

Following these recommendations, bulls with data selected for the estimation should 1) be born within a period of 10 year before the birth year of the youngest selected bull, 2) have daughters in at least 20 herds in each population, and 3) be associated with EGM having REL equal or higher than 75% in both populations. Furthermore, most recent data should be used and, if sufficient number of bulls is available, bulls initially sampled in the exporting populations (i.e., according to the gene flow) should be selected for the estimation of a and b. Recommendations for cases with a non-random use of bulls, or a correlation between EGM lower than 0.75, or a number of selected bulls lower than 20 or a number of common used bulls very low, were also proposed (Wickham and Philipsson, 1990a). All these recommendations lead to the fact that some internal populations did not have enough animals from the external population proven in their own internal population. Therefore, conversions of EGM were performed through a third population having enough common proven animals with the external and internal populations (Wickham and Philipsson, 1990b).

To summarize, conversion equations are simple, easy to apply and provide results for use in internal populations. However, estimations of a and b are mostly based on a small number of animals being evaluated in the two populations. Also, conversion equations can only be applied to render genetic evaluations of two populations comparable at a time, mostly for only one direction, and may not be accurate for animals with extremely high merit (Banos and Sigurdsson, 1996). Furthermore, relationships among animals are not taken into account by the conversion equations and external information is not propagated to relatives. Finally, conversion equations do not remove the issue of animals associated with more than one EGM within a population.

WEIGHTED AVERAGES

Conversion equations do not allow for the consideration of external information associated with an animal into its internal evaluation nor for the propagation of this external information to its progeny, leading overall to a loss of REL. Hence, other approaches and algorithms were developed to combine external and internal information. Thereby, in the context of Holstein dairy cattle, Wiggans et al. (1992) proposed an approach to combine US and Canadian bull evaluations based on the decomposition of a bull's evaluation between parent averages (PA) and progeny contributions (PC). After conversion of Canadian evaluations and PC from the Canadian trait to the US trait by using conversion equations, US and Canadian evaluations were combined as a weighted

average of either a combined PA or a PA from the US or the Canadian evaluations, depending on the availability, and US and Canadian PC adjusted for the bull's contribution. The adjustment of PC for the bull's contribution was due to the fact that the bull's contribution to PC through PA in the progeny EGM could be important. Because combination of PC across countries could change the bull's evaluation, it should be removed before combination. Also, the process was done from the oldest bull to the youngest one in order to propagate additional information from the oldest bulls to the youngest ones. Estimated REL associated with combined evaluations were a function of the sum of daughter equivalents (DE) from parents' contributions from the combined evaluations, and DE from PC for both US and Canadian evaluations. However, because the approach considered only bulls, cows' evaluations provided from the national evaluations could disagree with the combined evaluation of their sire. Moreover, the approach did not adjust progeny's evaluations for changes in bulls' EGM and foreign cows were also not considered.

Derived from the equations of the random genetic effects, Mrode et al. (1996) proposed a similar procedure, solving some disadvantages of the Wiggans' method, to combine United Kingdom and converted foreign evaluations. Combinations of evaluations for bulls and also for cows were performed as a weighted average of PA, yield deviations and PC. Evaluations of progeny were adjusted for changes in evaluations of their parents. Because changes in parents' evaluations affect only PA, only a weighted difference between PA from combined evaluations and PA from internal evaluations was added to the progeny's evaluations. Specific rules were defined for progeny with unknown parents. Combined REL for bulls, cows and progeny were estimated from the decomposition in different contributions of the national and foreign information, expressed in DE, similarly to Wiggans et al. (1992).

To summarize, both methods approximate animal model estimates. In order words, both methods took all relationships among animals into account, and this led to an increase of REL and eliminated the problem of animals associated with more than one EGM within a country. However, both methods did not convert external information and conversion equations were still needed.

LINEAR MIXED MODELS

MULTIPLE ACROSS COUNTRY EVALUATION

In the context of dairy cattle breeding, the first method based on linear mixed models to analyze jointly national evaluations from several countries was called multiple-country evaluation (MCE) and proposed by Schaeffer (1985). This method provided international estimates for all bulls in all participating countries. The MCE was based on a single-trait model assuming equal heritability across countries, several interactions as unimportant (e.g., genotype by environment interactions), a diagonal (co)variance matrix for the residual effect and unbiased internal (national) evaluations. The MCE had the advantage to use a pedigree relationship matrix across countries increasing connectedness among countries resulting in better estimates of international EGM. Furthermore, MCE allowed for the simultaneous comparison among a large number of countries (unlike conversion equations), based on a large number of daughters per bull in multiple countries. MCE also allowed for the prediction of genetic merits on the scale of each country (Schaeffer, 1985). Furthermore, compared to conversion equations, all information can be used instead of only information related to animals evaluated at least in two countries.

Since some assumptions of MCE were unrealistic (Schaeffer, 1994), Schaeffer (1994) proposed the multiple across country evaluation (MACE), which is a multiple-trait model for which similar traits in different countries are considered as different traits. In addition to the advantages of MCE, MACE overcomes the disadvantages. Indeed, MACE allows for different scales, for different units of measurement, for different heritabilities and genetic parameters for each country, and for genetic correlations between countries lower than one. Genetic correlations less than one account for 1) differences between statistical models used for genetic evaluations in several countries and 2) genotype by environment interactions. Different rankings of animals in the participating countries can be therefore observed (especially due to the consideration of genotype by environment interactions) and the degree of difference among the rankings is dependent on the genetic correlations among countries (Banos and Sigurdsson, 1996).

The model proposed by Schaeffer (1994) is a sire-maternal grandsire model and is described for a country *i* as follows:

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{c}_i + \mathbf{Z}_i\mathbf{Q}\mathbf{g}_i + \mathbf{Z}_i\mathbf{s}_i + \mathbf{e}_i \qquad \text{(II.1)}$$

where $\mathbf{y_i}$ is the vector of observations, $\mathbf{c_i}$ is the vector of country of evaluation effect, $\mathbf{g_i}$ is the vector of genetic groups of bull effect, $\mathbf{s_i}$ is the vector of genetic merits of bull effect, $\mathbf{e_i}$ is the vector of residuals, and $\mathbf{X_i}$, $\mathbf{Z_i}$, $\mathbf{Q}$ are incidence matrices.

The (co)variance matrices of $\mathbf{s}$ and $\mathbf{e}$ for *n* countries are, respectively,

$$Var(\mathbf{e}) = Var\begin{bmatrix}\mathbf{e_1}\\ \mathbf{...}\\ \mathbf{e_n}\end{bmatrix} = \begin{bmatrix}\mathbf{D_1}\sigma^2_{e_1} & \mathbf{...} & \mathbf{0}\\ \mathbf{...} & \mathbf{...} & \mathbf{...}\\ \mathbf{0} & \mathbf{...} & \mathbf{D_n}\sigma^2_{e_n}\end{bmatrix} \text{ and}$$

$$Var(\mathbf{s}) = Var\begin{bmatrix}\mathbf{s_1}\\ \mathbf{...}\\ \mathbf{s_n}\end{bmatrix} = \begin{bmatrix}\mathbf{A}s_{11} & \mathbf{...} & \mathbf{A}s_{1n}\\ \mathbf{...} & \mathbf{...} & \mathbf{...}\\ \mathbf{A}s_{n1} & \mathbf{...} & \mathbf{A}s_{nn}\end{bmatrix}$$

where $\sigma^2_{e_i}$ is the residual variance for country *i*, $\mathbf{D_i}$ is a diagonal matrix with elements equal to 1 divided by the number of daughters of a bull, s_{ij} is the sire (co)variance between country *i* and *j*, and $\mathbf{A}$ is a sire-maternal grandsire additive relationship matrix.

The residual (co)variance matrix was assumed to be diagonal although covariances among observations within a country are not zero. Covariances among observations and among countries equal to zero assume that national evaluations are performed from independent data sets. Also, observations used for the model (II.1) should represent unregressed measures of progeny performances corrected for several effects (e.g., herd effects, genetic merit of mates) in each country. Suggested observations were national EGM (Schaeffer, 1994), deregressed proofs (DRP; Rozzi et al., 1990) and daughter yield deviations (DYD; Schaeffer, 1994). Comparison of these three estimates as observations for the model (II.1) were performed by Sigurdsson and Banos (1995) and these authors recommended the use of DRP as observations. For DRP, effective daughter contributions (EDC) of bulls were suggested as weighting factor (Fikse and Banos, 2001).

A major limitation of MACE is that it can combine only one trait for a bull within a country because one of the assumptions is that residuals are not correlated among countries. However, because more and more evaluations were changed from single-trait to multiple-trait, an extension of MACE to considerer multiple traits within a country was needed (e.g., Schaeffer et al., 2000). Therefore, Schaeffer (2001) proposed a multiple-trait MACE (MT-MACE) allowing multiple traits within a country. Nevertheless, de-regression steps for multiple-trait evaluations could be difficult, especially because each country could have a different number of traits. Furthermore, de-regression steps could

not be harmonized among countries if it is performed by each country (Schaeffer, 2001). To avoid these difficulties, some studies (e.g., Sullivan and Wilton, 2001; Liu et al., 2004) proposed modifications to MT-MACE to simplify its use.

To summarize, (MT-)MACE has several advantages, like the simultaneous comparison among a large number of countries, the use of a pedigree relationship matrix, the use of all available information, and the prediction of genetic merits on the scale of each country. However, some limitations of (MT-)MACE exist, like de-regression steps (Schaeffer, 2001) and the definition of traits following the country borders instead of environment differences (e.g., climate, management; Weigel and Rekaya, 2000). Furthermore, in the context of dairy cattle, MACE provides international EBV only for bulls, leading to potential issues. Indeed, their publications as "official" by a country can lead to conflicts if national EBV for the same bulls are much different, and, therefore, disagree with EBV associated with close relatives (e.g., progeny, cows, bulls without international EBV; Täubert et al., 1999). The consideration of only bulls by MACE is also a problem to evaluate without bias females out of a foreign dam as well as females with a highly selected foreign sire without national progeny's data in comparison to females having a local origin (Pedersen et al., 1999). Thereby, although MACE solves issues concerning combinations of national EBV for most bulls, a need to propagate and to integrate MACE results into national genetic evaluations appeared.

BLENDING ALGORITHM

Publications of MACE EBV for some bulls together with national EBV for other bulls, cows and young animals could lead to conflicts. Therefore, Täubert et al. (2000) proposed an iterative algorithm to combine national and MACE EBV simultaneously for all animals. Their algorithm was based on the equations of the part of random effects of the mixed model equations (MME):

$$\left(\mathbf{D} + \mathbf{A}^{-1}\lambda\right)\hat{\mathbf{u}} = \mathbf{r}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements equal to performances equivalents for bulls with MACE EBV, $\mathbf{A}^{-1}$ is the inverse of the relationships matrix, λ is the ratio of error to genetic variances, $\hat{\mathbf{u}}$ is the vector of blended EBV, and $\mathbf{r}$ is the right hand side (RHS) of the equation.

To summarize, knowing the RHS for bulls with MACE EBV and for other bulls, cows and young animals, the blending algorithm combines iteratively national and MACE EBV simultaneously for all animals by weighting MACE EBV through

performances equivalents. Täubert et al. (2000) observed that blended EBV of progeny of bulls were influenced in the same way as those of their sires and that conflicts were solved.

SELECTION INDEX

Based on selection index (SI) theory, methods were developed to combine different genetic evaluations related to the same trait or to correlated traits. A first SI approach was developed by Weigel et al. (1998) to provide an evaluation of the productive life (PL) trait in dairy cattle. PL is related to culling data of progeny which are extensively available only relatively late in the life of a dairy sire. Also, it has a low heritability. Therefore, REL associated with PTA for young bulls are low. Therefore, Weigel et al. (1998) developed a SI approach to combine early indirect PL information obtained from correlated traits and direct PL information. An indirect PTA for PL ($\hat{u}_{ind}$) was obtained as follows:

$$\hat{u}_{ind} = Cov(u_{PL}, \mathbf{u_{MT}})' Var(\mathbf{u_{MT}})^{-1} \hat{\mathbf{u}}_{\mathbf{MT}}$$

where $Cov(u_{PL}, \mathbf{u_{MT}})$ is the covariance between PTA for PL (u_{PL}) and the true transmitting abilities for correlated traits ($\mathbf{u_{MT}}$), $Var(\mathbf{u_{MT}})$ is the variance of $\mathbf{u_{MT}}$, and $\hat{\mathbf{u}}_{\mathbf{MT}}$ is the vector of multiple-trait Best Linear Unbiased Prediction (BLUP) predictions of correlated traits.

The reliability of $\hat{u}_{ind}$ was calculated as follows:

$$REL_{ind} = Cov(u_{PL}, \mathbf{u_{MT}})' Var(\mathbf{u_{MT}})^{-1} Var(\hat{\mathbf{u}}_{\mathbf{MT}}) Var(\mathbf{u_{MT}})^{-1} Cov(u_{PL}, \mathbf{u_{MT}}) / Var(u_{PL}).$$

Direct and indirect PL predictions, $\hat{u}_{dir}$ and $\hat{u}_{ind}$ respectively, were then combined in a weighted mean where weights were function of REL associated with $\hat{u}_{dir}$ and $\hat{u}_{ind}$. Weights also accounted for the lack of independence between direct and indirect evaluations because some progeny had both direct and indirect observations. Also, many traits correlated with PL may be highly correlated among them. Because high correlations could lead instability of indirect predictions, a principal component procedure on a correlation matrix was applied to discard redundant traits.

Later, VanRaden et al. (2000) and VanRaden (2001) proposed another SI approach to combine genetic evaluations of the same or correlated traits. It consists of combining estimates of Mendelian samplings (MS) from each different genetic evaluation through SI and then to add the combined predictions to PA expressed on the desired scale. Therefore,

for an animal associated with different predictions, its combined EGM, $\hat{u}_{comb}$, is equal to:

$$\hat{u}_{comb} = \hat{u}_{PA} + Cov(u_{MS}, \hat{\mathbf{u}}_{\mathbf{MS}})' Var(\hat{\mathbf{u}}_{\mathbf{MS}})^{-1} \hat{\mathbf{u}}_{\mathbf{MS}} \qquad \text{(II.2)}$$

where $\hat{u}_{PA}$ is the estimated PA on the scale of interest, $Cov(u_{MS}, \hat{\mathbf{u}}_{\mathbf{MS}})$ is the covariance between the true MS (u_{MS}) and the vector of predicted MS ($\hat{\mathbf{u}}_{\mathbf{MS}}$) and $Var(\hat{\mathbf{u}}_{\mathbf{MS}})$ is the variance of $\hat{\mathbf{u}}_{\mathbf{MS}}$.

A combination of genetic evaluations can be done for each animal from the oldest animal to the youngest animal by updating first $\hat{u}_{PA}$ and then by applying the equation (II.2) (VanRaden, 2001). Some rules were defined to estimate $\hat{u}_{PA}$ if parents were unknown, similarly to Mrode et al. (1996). With such an approach, information from foreign parents and progeny is propagated to domestic progeny. However, some information is still lost because information from foreign progeny does not contribute back to the parents of the considered animal. For the reliabilities of $\hat{u}_{comb}$, they can be approximated from the decomposition of additional information between different contributions, expressed in DE.

In a dairy cattle context, the proposed SI approach was used to approximate an international evaluation (VanRaden et al., 2000; VanRaden, 2001) as well as to combine predictions related to correlated traits (VanRaden, 2001). Regarding the context of international evaluations, the SI approach can provide international evaluations for cows, which was not the case for MACE. Another advantage compared to MACE is that an international relationship matrix is not needed. Results showed that small differences were observed between the SI approach and MACE for bulls. Correlations were about 0.99. However, REL associated with the SI approach was higher because sire and dam information was integrated instead of only sire and maternal grandsire information (VanRaden, 2001). Regarding the context of multi-trait evaluations, the comparison of the approaches proposed by Weigel et al. (1998) and VanRaden (2001) showed that gains in REL were higher with the SI approach developed by VanRaden (2001) because the VanRaden's approach included parents, animal and progeny information (VanRaden, 2001).

To summarize, the SI approaches are approximate methods combining accurately different sources of information related to the same or correlated traits and need lower computational needs than other methods based on mixed models, e.g., MACE or a joint BLUP evaluation. Also, unlike MACE, estimates are provided for all animals and

information provided by cows is considered. However, the SI approaches may not take the different effects of selection into account as correctly as a joint evaluation (VanRaden, 2001).

SIMULTANEOUS COMBINATIONS

ABSORPTION OF EQUATIONS

An approach to combine simultaneously external information and internal phenotypic and genealogic data is to integrate external information into a genetic evaluation by considering a genetic evaluation using internal and external data and by absorbing the equations related to the external data. Based on an algorithm writing directly the inverse of a relationship matrix and therefore allowing the use of an animal model, the approach based on the absorption of equations (hereafter called absorption based approach) was first implicitly proposed by Henderson (1975) in order to incorporate artificial insemination (AI) sire evaluations based on records of artificially sired daughters in other herds into intraherd predictions, as an alternative to an interherd genetic evaluation. The proposed method allowed the comparison of cows across herds and accounts for non-random usage of sires (Henderson, 1975; Bolgiano et al., 1983). While it does not seem to be a problem nowadays, an interherd genetic evaluation based on an animal model for a particular breed in a specific subpopulation was not computationally feasible at that time (Bolgiano et al., 1983). Based on external information associated with a sire, the approach consisted of adding the value $n(1-r)/(4-h^2)$ to the diagonal element of the sire's equation in the left hand side (LHS) of the internal MME where n is the number of effective daughters calculated from the sire's external REL, r is the repeatability, and h^2 is the heritability of the considered trait. The value $\frac{(1-r)}{(4-h^2)h^2}(4+(n-1)h^2)\hat{u}$, where $\hat{u}$ is the sire's external EBV, was added to the element in the RHS corresponding to the sire's equation in the internal evaluation. In the context of milk yields evaluations for dairy cattle, this element is equal to 0 for sires with no external information. Later, Quaas (1979) and Van Vleck (1982) proposed two different derivations of the method suggested by Henderson (1975). An application was also proposed to estimate genetic values for cows within a herd and to compare them among different herds for the Dairy Herd Improvement Association herds in the United States (Bolgiano et al., 1983).

To summarize, the absorption based approach is straightforward to incorporate external information of sires into an intraherd genetic evaluation and to propagate this information to sires' progeny in the considered herd. However, although approximations were proposed for more complicated cases (Henderson, 1975; Van Vleck, 1982), this method is difficult to generalize (Quaas and Zhang, 2006). Furthermore, the approach assumes that external information is expressed as deviation from the genetic merit of the base population, which could not be a trivial problem (Henderson, 1975).

PSEUDO-RECORDS

Bonaiti and Boichard (1995) proposed a method that includes external information into an internal genetic evaluation for dairy cattle using a single-trait animal model. This method consists of adding a number of virtual daughters for each bull associated with external information to the internal phenotypic and pedigree datasets. The second parents of the virtual daughters are assumed to be unknown and belonged to a genetic group which is related to the a-factor of the conversion equation. Each additional virtual daughter is associated with a pseudo-record representing the genetic merit of the bull. Pseudo-records could be DYD or DRP derived from external EBV of the bulls (Bonaiti and Boichard, 1995). Because external information of related bulls could be included in the internal evaluation, double counting of contributions due to relationships could appear. Therefore, the number of additional virtual daughters for the *i*th external animal is equivalent to the number of DE computed from external REL of the *i*th animal subtracted by the number of DE associated with its external pedigree index. The subtraction is needed to avoid double counting of contributions due relationships. One limitation of this method is that pseudo-records are assumed to belong to the same trait as the internal phenotypes. Thereby, DYD or EBV must be first converted to the internal unit and scale, e.g., with conversion equations, before their inclusion in the internal genetic evaluation. It is noted that Bonaiti and Boichard (1995) showed the similarities of their proposed pseudo-records based approach with the absorption based approach (Henderson, 1975). It is also noted that the effect of external information will never be zero, although the weight for external information of an animal decreases when more and more internal data is collected (Pedersen et al., 1999).

Pedersen et al. (1999) applied the pseudo-records based approach proposed by Bonaiti and Boichard (1995) in the context of the Danish genetic evaluation for Holsteins, Jerseys and Red Danish cattle. External information included MACE EBV and REL,

provided by International Bull Service (Interbull, Uppsala, Sweden), for sires and, for cows, external EBV and REL provided by the country of origin. Before the inclusion, external EBV for cows were converted to the Danish scale. Also, REL were restricted to a maximum of 80% for MACE EBV for sires and were assumed to be equal to 30% for converted external EBV for external cows. Such restrictions could limit double counting of Danish information. The authors concluded that including external information in the Danish evaluation led to avoid differences between national and MACE EBV and to the possible comparison of females with local or external origins.

Based on a pseudo-records based approach, VanRaden (2012), VanRaden and Tooker (2012) and VanRaden et al. (2014) proposed to include MACE information in an internal genetic evaluation by using one pseudo-record for each bull weighted by DE, instead of one pseudo-record for each virtual daughter. The pseudo-records were expressed as DRP and were estimated using a one-animal-at-a-time deregression method, similarly to Bonaiti and Boichard (1995). Because MACE information associated with bulls that have internal and external daughters include contributions from both internal and external information, VanRaden et al. (2014) proposed to use internal EBV instead of PA to compute DRP for those bulls and to subtract internal DE from the total amount of DE associated with MACE EBV in order to estimate the amount of external DE. Double counting of internal information was therefore avoided. For multiple-trait models, VanRaden et al. (2014) proposed to adapt their approach by adding, for each *i*th external animal, the product $\mathbf{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}}\mathbf{G}_{\mathbf{0}}^{-\mathbf{1}}\mathbf{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}}$ to the *i*th external animal's elements of the LHS where $\mathbf{\Delta}_{\mathbf{VRi}}$ is a diagonal matrix associated to the *i*th external animal with a diagonal element for the *j*th trait equal to $REL_{ij}/(1-REL_{ij})$, REL_{ij} is the reliability of *i*th external animal for the *j*th trait, and $\mathbf{G}_{\mathbf{0}}$ is the genetic (co)variance matrix among traits. The product $(\mathbf{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}}\mathbf{G}_{\mathbf{0}}^{-\mathbf{1}}\mathbf{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}})\mathbf{y}_{\mathbf{di}}$ is also added to the *i*th external animal's elements of the RHS where $\mathbf{y}_{\mathbf{di}}$ is the vector that includes DRP for each trait for the *i*th external animal. VanRaden and Tooker (2012) tested their approach on US Holstein data and concluded that their approach was simple and enough accurate to include external information into national evaluations. While this approach showed similarities with the absorption based method proposed by Henderson (1975) for univariate analyses, results from these two approaches could differ in practice because the computation of DRP and their consideration into the MME are different between the two approaches.

Another pseudo-records based approach was developed by Tier et al. (1999) to integrate external information into BREEDPLAN, a genetic evaluation system for beef cattle (www.breedplan.une.edu.au). Similarly to Bonaiti and Boichard (1995), the authors added virtual progeny with pseudo-records for animals associated with external information. The number of virtual daughters was limited to a maximum in order to allow that 1) internal records were reflected in internal EGM and 2) pseudo-records were computed from the multiplication between an approximated multi-trait external LHS and a vector of external EGM. This approach to compute pseudo-records is similar to deregression approaches, such as the one proposed by Calus et al. (2014). Similarly to the other pseudo-records based approaches, Tier et al. (1999) observed a correct integration of external information into their genetic evaluation system.

To summarize, different authors proposed to include accurately external information in internal genetic evaluations by adding to internal datasets weighted pseudo-records associated with animals or with their virtual daughters. The similarities between the absorption based approaches and pseudo-records based approaches were detailed (Bonaiti and Boichard, 1995). However, because DYD are not usually available, a deregression step is needed to estimate DRP. While the deregression approaches used by some authors were based on a one-animal-at-a-time deregression method, better approaches could be used (VanRaden et al., 2014), which is mostly not a trivial problem, as already discussed before.

BAYESIAN APPROACHES

Henderson (1984) proposed MME modified to incorporate previous estimates of fixed effects, like breed or sex differences, associated with a non-singular (co)variance matrix. Later, Gianola and Fernando (1986) derived general MME from the Bayesian methodology. Unlike the previous proposed system of equations (Henderson, 1984), the derivation of Gianola and Fernando (1986) concerned “fixed” and “random” effects because those effects are not distinguished in a Bayesian context. Assuming the following prior multivariate normal (MVN) distributions:

- for the “fixed” effects, $[\boldsymbol{\beta}|\mathbf{B}] \sim MVN(\mathbf{b}, \mathbf{B})$,

where $\mathbf{b}$ is a mean vector and $\mathbf{B}$ is a (co)variance matrix,

- for the “random” effects, $[\mathbf{u}|\mathbf{G}] \sim MVN(\mathbf{g}, \mathbf{G})$,

where $\mathbf{g}$ is a mean vector and $\mathbf{G}$ is a (co)variance matrix, and

- for the residual effect, $[\mathbf{e}|\mathbf{R}] \sim MVN(\mathbf{0}, \mathbf{R})$,

where $\mathbf{R}$ is a (co)variance matrix, the developed equations were written as follows (Gianola and Fernando, 1986):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} + \mathbf{B}^{-1} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{G}^{-1}\mathbf{g} \end{bmatrix} \qquad \text{(II.3)}$$

where $\mathbf{y}$ is the vector of records, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are vectors of estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$ related to the records through the incidence matrices $\mathbf{X}$ and $\mathbf{Z}$, respectively.

If a non-informative prior is considered for $\boldsymbol{\beta}$ (i.e. $\mathbf{B}^{-1} \to \mathbf{0}$) and $\mathbf{g} = \mathbf{0}$, the system of equations (II.3) simplifies to traditional Henderson‘s MME (e.g., Dempfle, 1977; Henderson, 1984; Gianola and Fernando, 1986; Robinson, 1991).

A proposed application using the system of equations (II.3) was the updating of $\boldsymbol{\beta}$ and $\mathbf{u}$ estimated from data increasing sequentially over time. A joint analysis of all data or an analysis of actual data with prior distributions based on a previous analysis of previous data would give the same results if the prior and posterior distributions are in the same family (Gianola and Fernando, 1986). A second application was proposed in the context described by Henderson (1975), which is the incorporation of AI sire evaluations based on records of artificially sired daughters in other herds into intraherd predictions. In this case, the vector $\mathbf{g}$ and the matrix $\mathbf{G}$ are partitioned between herd animals, which are called internal animals (described by the subscript I) since they are not associated with external information, and AI sires, so-called external animals (described by the subscript E), leading to $\mathbf{g} = \begin{bmatrix} \mathbf{g_I} \\ \mathbf{g_E} \end{bmatrix}$ and $\mathbf{G} = \begin{bmatrix} \mathbf{G_{II}} & \mathbf{G_{IE}} \\ \mathbf{G_{EI}} & \mathbf{G_{EE}} \end{bmatrix}$ where the vector $\mathbf{g_I}$, equal to $\mathbf{0}$, is the vector of prior means related to internal animals, the vector $\mathbf{g_E}$ is the vector of the external animals’ EGM associated with the prediction error (co)variance matrix $\mathbf{G_{EE}}$ and the matrices $\mathbf{G_{II}}$, $\mathbf{G_{EI}}$ and $\mathbf{G_{IE}}$ are functions of the additive relationships between internal and external animals. The matrix $\mathbf{B}^{-1}$ is considered as equal to $\mathbf{0}$. Because $\mathbf{G_{EE}}$ is unknown or could lead to difficulties for the computation of $\mathbf{G}^{-1}$, $\mathbf{G_{EE}}$ could be taken as a diagonal matrix with elements equal to approximates or real values of diagonal elements of the inverse needed to solve the system of equations (II.3) (Gianola and Fernando, 1986).

Later, Bayesian approaches were proposed in the context of multi-breed genetic

evaluations for beef cattle to combine data with prior literature estimates to estimate across-breed genetic values (Klei et al., 1996). In the same context, the Bayesian approach was extended to integrate external information associated with animals originating from another breed than the main breed considered in the internal multi-breed evaluation into this multi-breed genetic evaluation (e.g., Quaas and Zhang, 2001, 2006; Legarra et al., 2007). Animals originating from another breed are mostly bulls accurately evaluated in their own external system and having few progeny in the internal multi-breed genetic evaluation. While Quaas and Zhang (2006) and Legarra et al. (2007) proposed 2 different Bayesian derivations to integrate external information into internal genetic evaluations by considering external information as priors of $\mathbf{u}$, the proposed system of equations (II.4) had a similar compact notation to the system of equations (II.3):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} & \mathbf{G}^{-1}\mathbf{Q} \\ \mathbf{0} & \mathbf{Q}'\mathbf{G}^{-1} & \mathbf{Q}'\mathbf{G}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}+\mathbf{G}^{-1}\mathbf{g} \\ \mathbf{Q}'\mathbf{G}^{-1}\mathbf{g} \end{bmatrix} \qquad \text{(II.4)}$$

where $\hat{\mathbf{b}}$ is a vector of genetic base differences among genetic evaluations and $\mathbf{Q}$ is the incidence matrix relating the elements of $\hat{\mathbf{b}}$ to the animals.

Similarly to Gianola and Fernando (1986), the vector $\mathbf{g}$ and the matrix $\mathbf{G}^{-1}$ were partitioned between internal animals and external animals as $\mathbf{g}=\begin{bmatrix} \mathbf{g}_\mathbf{I} \\ \mathbf{g}_\mathbf{E} \end{bmatrix}$ and $\mathbf{G}^{-1}=\begin{bmatrix} \mathbf{G}^{\mathbf{II}} & \mathbf{G}^{\mathbf{IE}} \\ \mathbf{G}^{\mathbf{EI}} & \mathbf{G}^{\mathbf{EE}} \end{bmatrix}$. Concerning $\mathbf{G}^{-1}$, instead to approximate $\mathbf{G}_{\mathbf{EE}}$ as a diagonal matrix (Gianola and Fernando, 1986) for the computation of $\mathbf{G}^{-1}$, Quaas and Zhang (2006) and Legarra et al. (2007) proposed to directly approximate $\mathbf{G}^{-1}$ as follows:

$$\mathbf{G}^{-1}=\begin{bmatrix} \mathbf{G}^{*\mathbf{II}} & \mathbf{G}^{*\mathbf{IE}} \\ \mathbf{G}^{*\mathbf{EI}} & \mathbf{G}^{*\mathbf{EE}}+\boldsymbol{\Lambda} \end{bmatrix}=\mathbf{G}^{*-1}+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda} \end{bmatrix}$$

where $\mathbf{G}^{*-1}=\begin{bmatrix} \mathbf{G}^{*\mathbf{II}} & \mathbf{G}^{*\mathbf{IE}} \\ \mathbf{G}^{*\mathbf{EI}} & \mathbf{G}^{*\mathbf{EE}} \end{bmatrix}$ is the inverse of the additive (co)variance matrix and $\boldsymbol{\Lambda}$ is a matrix that must be approximated.

Concerning the product $\mathbf{G}^{-1}\mathbf{g}$, Quaas and Zhang (2006) and Legarra et al. (2007) computed the product as $\mathbf{G}^{-1}\mathbf{g}=\begin{bmatrix} \mathbf{0} \\ \left(\mathbf{G}_{*\mathbf{EE}}^{-1}+\boldsymbol{\Lambda}\right)\mathbf{g}_\mathbf{E} \end{bmatrix}$ where $\mathbf{G}_{*\mathbf{EE}}^{-1}$ is the inverse of the additive genetic (co)variance matrix that only accounts for the relationships among external

animals. The differences between the two Bayesian approaches (Quaas and Zhang, 2006; Legarra et al., 2007) concern mainly the calculations of the matrix $\mathbf{\Lambda}$. Legarra et al. (2007) considered the matrix $\mathbf{\Lambda}$ as a diagonal variance matrix and computed the diagonal elements of $\mathbf{\Lambda}$ as the difference between the diagonal elements of the prediction error (co)variance matrix and the diagonal elements of $\mathbf{G}^{-1}_{*\mathbf{EE}}$. Quaas and Zhang (2006) considered $\mathbf{\Lambda}$ as a block diagonal variance matrix with one block per external animal. Block diagonals for the *i*th animal were equal to $\mathbf{\Delta_{Qi}G_0^{-1}\Delta_{Qi}}$ where the matrix $\mathbf{G_0}$ is a matrix of genetic (co)variances among traits. The matrix $\mathbf{\Delta_{Qi}}$ is a diagonal matrix with elements δ_{ij} equal to $\sqrt{REL_{ij}/(1-REL_{ij})}$, where REL_{ij} is the reliability for the *j*th trait of *i*th external animal. It is noted that the block elements of the matrix $\mathbf{\Lambda}$ computed by Quaas and Zhang (2006) are equivalent to the elements computed by VanRaden et al. (2014) and added to the external animal's elements of the LHS (i.e., $\mathbf{\Delta^{0.5}_{VRi}G_0^{-1}\Delta^{0.5}_{VRi}}$).

Based on Legarra et al. (2007), an application to integrate MACE information for Holstein bulls into an internal dairy cattle genetic evaluation was proposed (Gengler and Vanderick, 2008). However, some limitations concern the proposed Bayesian approaches. Firstly, the proposed approaches did not take into account several double counting of contributions, e.g., due to relationships among external animals and due to records. Indeed, an EBV of an animal combines information from its own records (i.e., contributions due to own records) and from records of all relatives through its parents and its progeny (i.e., contributions due to relationships; VanRaden, 2001; Misztal and Wiggans, 1988). Therefore, integration of EBV for relatives can lead to counting several times the same contributions due to relationships and biases the internal genetic evaluation. Double counting of contributions due to records could appear if external information resulted from an external evaluation that combined both external and internal records. In this case, some contributions due to records would be considered several times if external information was combined with internal records. To our knowledge, in the context of Bayesian approaches, double counting of contributions due to relationships received very little consideration (e.g., Jones and Goddard, 1990). Only Gengler and Vanderick (2008) proposed an additional pre-processing step to avoid double counting of contributions due to records, despite the fact that this issue could be a major issue for common sources of external information (e.g., MACE information). Secondly, the proposed Bayesian approaches were developed to take differences among genetic bases

into account. However, Legarra et al. (2007) assumed that differences among genetic bases were equal to zero to improve software's performances. Furthermore, in addition to be expressed on a different genetic base, external information may be also expressed in other scales or units of measurement than the internal ones. For example, in the context of dairy cattle, it was noted that EBV are mostly reported as the average of lactation yields for three lactations while genetic evaluations are mostly based on test-day models (Gengler and Vanderick, 2008). Thereby, approaches must be developed to avoid these issues, similarly to other approaches detailed previously. Thirdly, in many situations, integration of several sources of external information into an internal genetic evaluation may be needed, although, to our knowledge, it has not been studied yet. Fourthly, the derivations of Legarra et al. (2007) and Quaas and Zhang (2006) assumed that external information came from a similar theoretical genetic evaluation, which is not necessarily the case.

To summarize, Bayesian approaches based on MME were one of the first approaches proposed to combine simultaneously external information and internal data. Different advantages may be observed as the avoidance of an explicit deregression step or the possibility to avoid high computational needs to estimate breeding values from data accumulating over time. However, Bayesian approaches integrating external information are not commonly used in animal breeding and several issues must be studied. Furthermore, most of current software packages available in animal breeding do not permit the application of, for example, the system of equations (II.3).

COMBINATIONS IN GENOMIC SELECTION

Currently, genomic information provided by panels of several thousands of single nucleotide polymorphisms (SNP) can be used following multi-step or single-step approaches. The multi-step approaches (e.g., VanRaden et al., 2009) consist of 1) estimating SNP effects based on (pseudo-)phenotypic data related to non-candidate genotyped animals, 2) calculating estimates of genetic merit (mostly called direct genomic values (DGV)) for candidate genotyped animals based on estimates of SNP effects and 3) combining genomic information, expressed as DGV, with traditional EBV (i.e., EBV estimated only from phenotypic and genealogic data) or with phenotypic and genealogical data. Following the traits and their availability, (pseudo-)phenotypic data may include traditional phenotypic records, DYD or DRP. The third step is necessary to avoid the publication of several EGM per animal and shows the need of approaches and

algorithms to combine EGM originating from diverse sources. The single-step approaches consist of replacing the traditional pedigree-based relationship matrix used in MME by a relationship matrix combining pedigree-based relationships and genomic relationships (Aguilar et al., 2010; Christensen and Lund, 2010). However, unlike the multi-step approaches allowing for the combination of genomic information with external information, single-step approaches do not allow for the integration of external information, like traditional MME. Therefore, for both approaches, methods and algorithms were developed to combine internal, external and genomic information. Most approaches and algorithms were derived from approaches already detailed previously and, therefore, will not be detailed extensively.

For the multi-step approaches, proposed methods concerned the combination of DGV with traditional EBV or phenotypic and genealogical data. Regarding the combination of DGV and EBV, a first class of approaches is based on SI theory (e.g., VanRaden et al., 2009; Harris and Johnson, 2010). The SI approaches are mostly used in a dairy cattle context. A second class of approaches combining DGV and EBV is based on bivariate (random) models (Mäntysaari and Strandén, 2010). Phenotypic data included DYD for conventional phenotypic and genealogic information and DGV as pseudo-records for genomic information. Heritability associated with DGV was assumed to be (close to) 1 and genetic correlation between DYD and DGV was assumed to be equal to the square root of the predictive ability of DGV. Regarding the combination of DGV and conventional phenotypic and genealogic data, different approaches were proposed. Similarly to Bonaiti and Boichard (1995), the first ones consisted of calculating pseudo-records and associated weights from DGV and associated REL and to include them in the conventional dataset considering pseudo-records as own records and from the same trait (Ducrocq and Liu, 2009; Liu et al., 2009; Ducrocq and Patry, 2010). Therefore, it was assumed that genomic and conventional information had the same genetic variance. A second class of approaches consisted of calculating pseudo-records and associated weights from DGV and including them in conventional genetic evaluation as a correlated trait. Similarly to Mäntysaari and Strandén (2010), DGV were considered as pseudo-records with a heritability equal or close to 1 by some authors (e.g., Kachman, 2008; Johnston et al., 2009; Špehar et al., 2013; Stoop et al., 2013). Contrarily to these authors, Stoop et al. (2011) proposed to include pseudo-records derived from DGV with an heritability equal to the predictive reliability of the prediction equation and with a genetic correlation with the trait of interest equal to 1. A mass-selection model was considered for

the pseudo-records. Finally, Bayesian approaches were also proposed to integrate DGV into conventional evaluations. In the context of dairy cattle, Gengler and Verkenne (2007) proposed to consider DGV as prior information for polynomials of order 0 associated with random polygenic additive effects ($\mathbf{u}_0$) of a multi-trait multi-lactation test-day model. It was assumed that $E(\mathbf{u}_0) = \mathbf{u}_{DGV}$ and $Var(\mathbf{u}_0) = \mathbf{G}_{00}$ where $E(\mathbf{u}_0) = \mathbf{u}_{DGV}$ is the vector of DGV and $\mathbf{G}_{00}$ is the additive genetic (co)variance matrix associated with $\mathbf{u}_0$. Reliabilities associated with DGV were not considered. In the context of beef cattle, DGV and associated REL were integrated into a multibreed genetic evaluation following the approach proposed by Quaas and Zhang (2006; Hyde et al., 2013). Their results showed that this approach can lead to inappropriate scaling of DGV or double counting of contributions between DGV and conventional data.

For the single-step evaluations, approaches concern integration of external polygenic information into single-step genomic evaluations, instead of combination or integration of genomic information into conventional genetic evaluations. To our knowledge, proposed approaches integrating external information into single-step genomic evaluations were only pseudo-records based approaches (VanRaden, 2012; Přibyl et al., 2013). These approaches were developed to integrate MACE EBV in a dairy cattle context by taking into account possible double counting of contributions among different sources. Only one pseudo-record and associated weight per animal was derived from external information and added to the conventional data.

At an international level, in the dairy cattle context, possible exchange of genotypes and predicted maker/SNP effects could lead to internal genomic evaluations dependent from each other while an assumption for MACE is that internal information originates from independent datasets. Therefore, a modified MACE, so called genomic MACE (GMACE), was developed to avoid this assumption, i.e., to account for non-zero residual correlations among genomic predictions across countries. GMACE was also extended to an animal model because genomic information associated with cows is increasing. GMACE is still under development (Sullivan and VanRaden, 2009; Sullivan and Jakobsen, 2014).

DISCUSSION AND CONCLUSION

This review highlights different approaches and algorithms that render genetic evaluations originating from different sources comparable or that combine these

evaluations to a single one. Each approach or algorithm was developed in a particular context (e.g., in a dairy cattle context) to solve the same common issue, that is the impossibility to run a joint genetic evaluation due to political roadblocks, inconsistencies among datasets, computing or logistic problems. It was also worth noting that some approaches, like MACE, SI and Bayesian approaches or pseudo-records based approaches, were extended to other contexts (e.g., the genomic context). All approaches and algorithms were sorted following the way external information was combined.

Firstly, external information can be combined with internal EGM and REL after performing internal evaluations. Advantages of post evaluation combinations are that 1) internal evaluations can be performed without being dependent on the publication of external information and 2) internal EGM and REL may be used to develop conversion equations without running additional evaluations free of external evaluations. In an international context, internal information free of external information is also a condition to perform unbiased MACE. However, these advantages could also be a disadvantage because external information does not contribute to the estimation of fixed and other random effects in the internal evaluation, which could create potential biases.

Secondly, external information can be simultaneously combined with internal genealogical and phenotypic data into an internal genetic evaluation. Under some assumptions, it was previously shown that the absorption based approaches and the pseudo-records based approaches are equivalent. A similar other type of approaches that simultaneously combine external information and internal data are the Bayesian approaches, also previously described. Differences observed among results are due to approximations applied according to the considered approach. Differences are also observed regarding their implementations. Bayesian and pseudo-records based approaches could be generalized more easily than absorption based approaches. Also, absorption based approaches and Bayesian approaches are not dependent of an explicit deregression step. However, an advantage of all the simultaneous combinations approaches is that external information contributes to the estimation of all effects included in the internal evaluation. Therefore, contributions due to external information are propagated to all animals included in the internal evaluation and related to animals associated with external information. Another advantage is that only one process is needed to combine all available information. A disadvantage is that internal evaluations must be performed after the publications of external information. Another disadvantage is that additional internal evaluations which do not integrate external evaluations must be

performed if, for example, conversion equations must be developed.

Finally, for both kinds of approaches and methods, external information must be studied carefully to avoid mainly bias of internal information and double counting of contributions due to relationships among external animals and due to same records. Furthermore, conversion equations seem to be still needed for most approaches to convert external information on the scale, units of measurement and genetic base of the internal populations. Approaches based on conversion equations are thus dependent on them. Few approaches, like MACE, proposed solutions to this issue through the use of genetic correlations between external and internal information. To our knowledge, most approaches, like Bayesian or pseudo-records based approaches, could use similar solutions to avoid the use of conversion equations and to be independent on them.

To conclude, a joint genetic evaluation would be preferred in comparison to approximated combinations of external information and internal information or datasets. However, although high performance computing facilities are more and more available, joint evaluations cannot be performed mostly due to political issues. Therefore, development of approaches and algorithms combining several sources of information are still needed and must be still studied given all assumptions and issues described previously.

REFERENCES

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Banos, G., and A. Sigurdsson. 1996. Application of contemporary methods for the use of international data in national genetic evaluations. *J. Dairy Sci.* 79:1117–1125.

Bolgiano, D.C., R.L. Quaas, and L.D. Van Vleck. 1983. Variances of prediction errors for differences of estimated genetic values of cows in the same and in different herds. *J. Dairy Sci.* 66:1370–1376.

Bonaiti, B., and D. Boichard. 1995. Accounting for foreign information in genetic evaluation. *Interbull Bull.* 11:4pp.

Calus, M., G. Schopen, and R. Veerkamp. 2014. Simultaneous deregression of cow and bull breeding values. *Interbull Bull.* 48:63–67.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not

genotyped. *Genet. Sel. Evol.* 42:2.

Dempfle, L. 1977. Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayésiens. *Genet. Sel. Evol.* 9:27–32.

Ducrocq, V., and Z. Liu. 2009. Combining genomic and classical information in national BLUP evaluations. *Interbull Bull.* 40:172–177.

Ducrocq, V., and C. Patry. 2010. Combining genomic and classical information in national BLUP evaluation to reduce bias due to genomic pre-selection. *Interbull Bull.* 41:33–36.

Fikse, W.F., and G. Banos. 2001. Weighting factors of sire daughter information in international genetic evaluations. *J. Dairy Sci.* 84:1759–1767.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Gengler, N., and C. Verkenne. 2007. Bayesian approach to integrate molecular data into genetic evaluations. *Interbull Bull.* 37:37–41.

Gianola, D., and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63:217–244.

Goddard, M. 1985. A method of comparing sires evaluated in different countries. *Livest. Prod. Sci.* 13:321–331.

Gravert, H.O. 1983. IDF recommended procedure for international comparison of the merit of dairy cattle. *Int. Dairy Fed. Bull.* 165:3–6.

Harris, B.L., and D.L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *J. Dairy Sci.* 93:1243–1252.

Henderson, C.R. 1975. Use of all relatives in intraherd prediction of breeding values and producing abilities. *J. Dairy Sci.* 58:1910–1916.

Henderson, C.R. 1984. Applications of linear models in animal breeding. 2nd ed. University of Guelph, Guelph, ON, Canada.

Hyde, L.R., W.R. Shafer, S.C. McGuire, M. Saatchi, and D.J. Garrick. 2013. ASA experience with incorporating genomics into genetic evaluation. Pages 10–14 in Proceedings of the 10th Genetic Prediction Workshop. Beef Improvement Federation, Kansas City, MO.

Johnston, D.J., B. Tier, and H.-U. Graser. 2009. Integration of DNA markers into BREEDPLAN EBVS. Pages 30–33 in Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics, Barossa

Valley, Australia. Association for the Advancement of Animal Breeding and Genetics.

Jones, L.P., and M.E. Goddard. 1990. Incorporating overseas information in proofs for bulls. Pages 345–348 in Proceedings of the 8th Conference of the Australian Association of Animal Breeding and Genetics. Association for the Advancement of Animal Breeding and Genetics.

Kachman, S.D. 2008. Incorporation of marker scores into national genetic evaluations. Pages 92–98 in Proceedings of the 9th Genetic Prediction Workshop. Beef Improvement Federation, Kansas City, MO.

Klei, L., R.L. Quaas, E.J. Pollak, and B.E. Cunningham. 1996. Multiple-breed evaluation. Pages 106–113 in Proceedings of the 28th Research Symposium and Annual Meeting. Beef Improvement Federation.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

Liu, Z., F. Reinhardt, and R. Reents. 2004. A multi-trait MACE model for international bull comparison based on daughter yield deviations. *Interbull Bull.* 32:46–52.

Liu, Z., F. Seefried, F. Reinhardt, and R. Reents. 2009. A simple method for correcting the bias caused by genomic pre-selection in conventional genetic evaluation. *Interbull Bull.* 40:184–188.

Mäntysaari, E.A., and I. Strandén. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluations. CD–ROM Comm. 353 in Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Misztal, I., and G.R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71(Suppl. 2):27–32.

Mrode, R.A., G.J.T. Swanson, and M.S. Winters. 1996. Method and effects of incorporating foreign information into United Kingdom production evaluations. *Anim. Sci.* 63:373–379.

Pedersen, G.A., J. Pedersen, U.S. Nielsen, and P. Madsen. 1999. Experiences of blending foreign information in the national genetic evaluation. *Interbull Bull.* 22:61–66.

Philipsson, J. 1987. Standards and procedures for international genetic evaluations of dairy cattle. *J. Dairy Sci.* 70:418–424.

Philipsson, J., B. Danell, L.R. Schaeffer, M. Schneeberger, H. Schulte-Coerne, and J.B.M.

Wilmink. 1986. Methods of transforming proofs. *Interbull Bull.* 1:12–22.

Powell, R.L. 1988. Comparison of procedures to convert sire evaluations between countries. *J. Dairy Sci.* 71:1609–1615.

Powell, R.L., and M. Sieber. 1992. Direct and indirect conversion of bull evaluations for yield traits between countries. *J. Dairy Sci.* 75:1138–1146.

Powell, R.L., G.R. Wiggans, and P.M. VanRaden. 1994. Factors affecting calculation and use of conversion equations for genetic merit of dairy bulls. *J. Dairy Sci.* 77:2679–2686.

Přibyl, J., P. Madsen, J. Bauer, J. Přibylová, M. Šimečková, L. Vostrý, and L. Zavadilová. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *J. Dairy Sci.* 96:1865–1873.

Quaas, R.L. 1979. Incorporating AI sire proofs into intraherd BLUP of breeding values. Mimeo. Dept. Anim. Sci., Cornell Univ., Ithaca, NY.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Quaas, R.L., and Z.W. Zhang. 2001. Incorporating external information in multibreed genetic evaluation. *J. Anim. Sci.* 79(Suppl. 1):342. (Abstr.).

Robinson, G.K. 1991. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* 6:15–32.

Rozzi, P., L.R. Schaeffer, E.B. Burnside, and W. Schlote. 1990. International evaluation of Holstein-Friesian dairy sires from three countries. *Livest. Prod. Sci.* 24:15–28.

Schaeffer, L.R. 1985. Model for international evaluation of dairy sires. *Livest. Prod. Sci.* 12:105–115.

Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.

Schaeffer, L.R. 2001. Multiple trait international bull comparisons. *Livest. Prod. Sci.* 69:145–153.

Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker, and J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144.

Sigurdsson, A., and G. Banos. 1995. Dependent variables in international sire evaluations. *Acta Agric. Scand. Sect. - Anim. Sci.* 45:209–217.

Špehar, M., K. Potočnik, and G. Gorjanc. 2013. Accuracy of genomic prediction for milk

production traits with different approaches in a small population of Slovenian Brown bulls. *Livest. Sci.* 157:421–426.

Stoop, M., H. Eding, M.L. van Pelt, L.C. de Haer, and G. de Jong. 2013. Using Pseudo-observations to combine Genomic and Conventional Data in the Dutch-Flemish National Evaluation. *Interbull Bull.* 47:106–110.

Stoop, W.M., H. Eding, M.L. van Pelt, and G. de Jong. 2011. Combining genomic and conventional data in the Dutch national evaluation. *Interbull Bull.* 44:169–172.

Sullivan, P.G., and J.H. Jakobsen. 2014. GMACE without variance estimation. *Interbull Bull.* 48:46–49.

Sullivan, P.G., and P.M. VanRaden. 2009. Development of genomic GMACE. *Interbull Bull.* 40:157–161.

Sullivan, P.G., and J.W. Wilton. 2001. Multiple-trait MACE with a number of traits per country. *Interbull Bull.* 27:68–72.

Täubert, H., H. Simianer, and K. Karras. 1999. Problems including Interbull breeding values in national evaluations. *Interbull Bull.* 22:57–60.

Täubert, H., H. Simianer, and K. Karras. 2000. Blending Interbull breeding values into national evaluations A new approach. *Interbull Bull.* 25:53–56.

Tier, B., D.J. Johnston, H.U. Graser, and M.E. Goddard. 1999. Incorporating foreign estimated breeding values into BREEDPLAN. Pages 199–202 in Proceedings of the 13th Conference of the Association for the Advancement of Animal Breeding and Genetics, Mandurah, Australia. Association for the Advancement of Animal Breeding and Genetics.

VanRaden, P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. *J. Dairy Sci.* 84 (E. Suppl.):E47–E55.

VanRaden, P.M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. *Interbull Bull.* 45:29–33.

VanRaden, P.M., R.L. Powell, and U. Emanuelson. 2000. Investigation of an alternative approach to international evaluations. *Interbull Bull.* 25:18–21.

VanRaden, P.M., C.P. Van 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.

VanRaden, P.M., and M.E. Tooker. 2012. Methods to include foreign information in national evaluations. *J. Dairy Sci.* 95:446.

VanRaden, P.M., M.E. Tooker, J.R. Wright, C. Sun, and J.L. Hutchison. 2014.

Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952-7962.

Van Vleck, L.D. 1982. Derivation of Henderson's method of incorporating artificial insemination sire evaluations into intraherd prediction of breeding values. *J. Dairy Sci.* 65:284–286.

Weigel, K.A., T.J. Lawlor Jr, P.M. Vanraden, and G.R. Wiggans. 1998. Use of linear type and production data to supplement early predicted transmitting abilities for productive life. *J. Dairy Sci.* 81:2040–2044.

Weigel, K.A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J. Dairy Sci.* 83:815–821.

Wickham, B.W., and J. Philipsson. 1990a. Background *Interbull Bull.* 4:3–4.

Wickham, B.W., and J. Philipsson. 1990b. Interbull recommendations. *Interbull Bull.* 4:4–11.

Wiggans, G.R., P.M. VanRaden, and R.L. Powell. 1992. A method for combining United States and Canadian bull evaluations. *J. Dairy Sci.* 75:2834–2839.

Wilmink, J.B.M., A. Meijering, and B. Engel. 1986. Conversion of breeding values for milk from foreign populations. *Livest. Prod. Sci.* 14:223–229.

Chapter III. COMPARISON AND IMPROVEMENT OF DIFFERENT BAYESIAN PROCEDURES TO INTEGRATE EXTERNAL INFORMATION INTO GENETIC EVALUATIONS

The previous Chapter reviewed various approaches and algorithms comparing or combining different genetic evaluations. One promising approach is the Bayesian approach that combines simultaneously external information (i.e., estimated breeding values and associated reliabilities provided by an external genetic evaluation) with internal phenotypic and genealogic data. Advantages of Bayesian approaches that were highlighted were the possible generalization to complex models, the avoidance of explicit deregression steps or the propagation of external information to all animals. However, while Bayesian approaches seem to be promising, Chapter II also highlighted some limitations, such as, e.g., double counting of contributions due to relationships. Therefore, the aim of this Chapter was first to review two Bayesian approaches that were recently proposed and, second, to enhance the proposed Bayesian approaches, mainly regarding computational burden and double counting of contributions due to relationships.

From: Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

ABSTRACT

The aim of this research was to compare different Bayesian procedures to integrate information from outside a given evaluation system, hereafter called external information, and in this context estimated breeding values (EBV), into this genetic evaluation, hereafter called internal evaluation, and to improve the Bayesian procedures to assess their potential to combine information from diverse sources. The 2 improvements were based on approximations of prior mean and variance. The first version of modified Bayesian evaluation considers all animals as animals associated with external information. For animals that have no external information (i.e., internal animals), external information is predicted from available external information. Thereby, propagation of this external information through the whole pedigree is allowed. Furthermore, the prediction of external information for internal animals allows large simplifications of the computational burden during setup and solving of mixed model equations. However, double counting among external animals (i.e., animals associated with available external information) is not avoided. Double counting concerns multiple considerations of contributions due to relationships by integration of external EBV for related external animals and is taken into account by the second version of modified Bayesian evaluation. This version includes the estimation of double counting before integration of external information. To test the improvements, 2 dairy cattle populations were simulated across 5 generations. Milk production for the first lactation for each female was simulated in both populations. Internal females were randomly mated with internal males and 50 external males. Results for 100 replicates showed that rank correlations among Bayesian EBV and EBV based on the joint use of external and internal data were very close to 1 for both external and internal animals if all internal and external animals were associated with external information. The respective correlations for the internal evaluation were equal to 0.54 and 0.95 if no external information was integrated. If double counting was avoided, mean squared error, expressed as a percentage of the internal mean squared error, was close to zero for both external and internal animals. However, computational demands increased when double counting was avoided. Finally, the improved Bayesian procedures have the potential to be applied for integrating external EBV, or even genomic breeding values following some additional assumptions, into routine genetic evaluations to evaluate animals more reliably.

Key words: Bayesian approach, dairy cow, integration, external information

INTRODUCTION

Theoretical properties of currently used methods to assess the genetic value of domestic animals depend on certain conditions. One of the most important is that all available information has to be used simultaneously to obtain unbiased estimates (e.g., Henderson, 1984). However, this is often not the case in practice, for many potential reasons. The most important issue is the unavailability of raw data (e.g., recorded and evaluated in another country) or the complexity of computations that require the use of multi-step, sequential, or distributed computing. Both issues are frequent in modern breeding, especially in dairy cattle breeding, because international exchange of genetic material (e.g., frozen semen and embryos) is extremely widespread. Until now, basic genetic evaluations are mostly based on local data, potentially followed by an international second step, as performed by the International Bull Service (Interbull, Uppsala, Sweden) for dairy breed sires. However, the accuracy of local evaluations may be limited for animals with few local data. Furthermore, the current massive development of genomic selection exacerbates this issue, because potentially more different genetic evaluations may exist, and the need to combine those sources of information increases. Current methods used in the context of dairy cattle are mostly selection index based on VanRaden (2001) to combine different sources of information (e.g., Gengler and VanRaden, 2008).

Another promising class of methods is based on Bayesian methods originating from the work from Klei et al. (1996) in the context of multibreed genetic evaluations for beef cattle. In this context, Bayesian means that the prior distribution of breeding values is changed according to what is known from an external source. Later, Quaas and Zhang (2001) and Legarra et al. (2007) proposed 2 different Bayesian derivations to incorporate external information, including external genetic breeding evaluations and their associated accuracies, into the internal evaluation. The integration of external information leads to an improved ranking of animals with external information (so-called external animals) in the internal evaluation, which is more similar to the ranking of a hypothetical joint evaluation of internal and external animals. Another advantage of this integration is that accuracies of estimated breeding values (EBV) for external animals are more reliable compared with those of the internal evaluation. Furthermore, this improvement of accuracies and rank correlations of external animals between the internal and joint evaluations depends on the external accuracy of prior information (Quaas and Zhang, 2001, 2006; Zhang et al., 2002;

Legarra et al., 2007) but also on several hypotheses used in the implementation. For example, current implementations do not take into account the double counting among external animals. However, an EBV of an animal combines information from its own records and from records of all relatives through its parents and its offspring (Misztal and Wiggans, 1988; VanRaden, 2001). Integrated external information of this animal and a close relative into the same genetic evaluation may be counted double if this external information contains both contributions due to relationships. Furthermore, until now, only few proposals exist to put these methods in the context of dairy cattle breeding, whereas they can be used in many situations and as a way to integrate genomic prediction (e.g., Gengler and Verkenne, 2007).

The first aim of this research was to compare different Bayesian approaches for their potential to combine information from diverse sources and the second aim was to improve existing Bayesian approaches to integrate external information into genetic evaluations. Focus was thereby given to the simplification of the computational burden and the avoidance of double counting among external animals.

MATERIALS AND METHODS

THEORETICAL BACKGROUND

Different concepts that will be used in this study are defined as follows:

(1) Internal data was defined as data used only for internal evaluations (e.g., milk records in a given country A).

(2) External data was defined as additional data not directly used in internal evaluations (e.g., milk records in another given country B).

(3) Internal information was related to information obtained from an evaluation based only on internal data (e.g., local EBV in country A).

(4) External information was related to information obtained from an evaluation based only on external data and free of internal information (e.g., foreign EBV or genomic EBV obtained in country B). Finally, all animals were distinguished between internal and external animals.

(5) An internal animal was an animal associated with only internal data and internal information (e.g., locally used sires in country A).

(6) An external animal was an animal associated with external data and information and also having internal data and information or being relative to the evaluation of

internal animals (e.g., foreign sires also used in country A in addition to country B or genotyped animals from country B relevant to country A).

The main reason for the application of Bayesian procedures is to obtain solutions as close as possible to those of a hypothetical joint evaluation of all external and internal animals including their data. This is performed by integrating external information into the internal genetic evaluation instead of using only internal data. The considered external information in this context was available external EBV and their associated accuracies obtained from only external data ($\mathbf{y_E}$). Both will be used to define the prior distribution of the internal EBV of the external animals ($\mathbf{u_E}$). This prior distribution can be defined in a generic way as $p(\mathbf{u_E}|\mathbf{y_E}) = MVN(\boldsymbol{\mu}_0 - \mathbf{Ub}, \mathbf{G}^*)$ where MVN means multivariate normal, $\boldsymbol{\mu}_0$ is the vector of external EBV of a joint genetic evaluation of all internal and external animals based only on external data $\mathbf{y_E}$, $\mathbf{G}^*$ is the matrix of prediction error (co)variances of these EBV, $\mathbf{b}$ is a vector of base differences between external and internal EBV, and $\mathbf{U}$ is an incidence matrix relating base differences to animals.

If E and I refer to external and internal evaluations, respectively, and based on Legarra et al. (2007), a generic model can be written leading to these mixed model equations (III.1), representing this multi-trait modified mixed model:

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_I} & \mathbf{0} \\ \mathbf{Z'_I R_I^{-1} X_I} & \mathbf{Z'_I R_I^{-1} Z_I + G^{*-1}} & \mathbf{G^{*-1} U} \\ \mathbf{0} & \mathbf{U' G^{*-1}} & \mathbf{U' G^{*-1} U} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_\mathbf{I} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'_I R_I^{-1} y_I} \\ \mathbf{Z'_I R_I^{-1} y_I + G^{*-1} \boldsymbol{\mu}_0} \\ \mathbf{U' G^{*-1} \boldsymbol{\mu}_0} \end{bmatrix} \qquad \text{(III.1)}$$

where $\mathbf{y_I}$ is the vector of internal observations, $\boldsymbol{\beta}_\mathbf{I}$ is the vector of fixed effects, $\mathbf{u}$ is the vector of random genetic effects of the external and internal random genetic effects, $\mathbf{X_I}$ and $\mathbf{Z_I}$ are the incidence matrices for internal fixed effects and animals, respectively, and $\mathbf{R_I}$ is the (co)variances matrix for the internal residual effects.

Legarra et al. (2007) showed that $\mathbf{G}^{*-1} = \begin{bmatrix} \mathbf{G}^{\mathbf{EE}} + \boldsymbol{\Lambda} & \mathbf{G}^{\mathbf{EI}} \\ \mathbf{G}^{\mathbf{IE}} & \mathbf{G}^{\mathbf{II}} \end{bmatrix}$ where $\boldsymbol{\Lambda}$ is equal to

$$\boldsymbol{\Lambda} = \mathbf{D}^{-1} - \mathbf{G}_{\mathbf{EE}}^{-1} \qquad \text{(III.2)}$$

where the matrix $\mathbf{D}$ is the matrix of prediction error (co)variances of the external information estimated from a genetic evaluation of all external animals based only on external data which did not include relationships between the internal animals, and $\mathbf{G}_{\mathbf{EE}}^{-1}$ is the inverse of the additive genetic (co)variance matrix that only accounts for the

relationships among external animals. It is important to note that the matrix $\mathbf{G}_{EE}^{-1}$ is different from $\mathbf{G}^{EE}$, because the latter also includes contributions from internal progeny of external animals. Differences between $\mathbf{G}_{EE}^{-1}$ and $\mathbf{G}^{EE}$ can be illustrated by writing the inverse of $\mathbf{G}$ in block form:

$$\begin{aligned}\mathbf{G}^{-1} &= \begin{bmatrix}\mathbf{G}_{EE} & \mathbf{G}_{EI}\\ \mathbf{G}_{IE} & \mathbf{G}_{II}\end{bmatrix}^{-1}\\ &= \begin{bmatrix}\mathbf{G}^{EE} & \mathbf{G}^{EI}\\ \mathbf{G}^{IE} & \mathbf{G}^{II}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{G}_{EE}^{-1}+\mathbf{G}_{EE}^{-1}\mathbf{G}_{EI}\mathbf{G}^{II}\mathbf{G}_{IE}\mathbf{G}_{EE}^{-1} & -\mathbf{G}_{EE}^{-1}\mathbf{G}_{EI}\mathbf{G}^{II}\\ -\mathbf{G}^{II}\mathbf{G}_{IE}\mathbf{G}_{EE}^{-1} & (\mathbf{G}_{II}-\mathbf{G}_{IE}\mathbf{G}_{EE}^{-1}\mathbf{G}_{EI})^{-1}\end{bmatrix}\end{aligned}$$

where $\mathbf{G}^{-1}$ is the inverse of the additive genetic (co)variance matrix $\mathbf{G}$ that accounts for all the relationships among all external and internal animals.

It has also been shown that

$$\mathbf{D}^{-1} = \mathbf{Z'}_{E}\,\mathbf{R}_{E}^{-1}\mathbf{Z}_{E} + \mathbf{G}_{EE}^{-1} \qquad \text{(III.3)}$$

and that

$$\mathbf{G}^{*-1}\boldsymbol{\mu}_{0} = \begin{bmatrix}\mathbf{D}^{-1}\boldsymbol{\mu}_{E}\\ \mathbf{0}\end{bmatrix} \qquad \text{(III.4)}$$

where $\boldsymbol{\mu}_{E}$ is a vector of external EBV from a genetic evaluation of all external animals based only on external data which did not include relationships between the internal animals (Legarra et al., 2007).

FOUR DIFFERENT IMPLEMENTATIONS

To be used, the generic system of equations (III.1) often needs to be simplified. In fact, usually only functions of external prediction error variances (PEV; e.g., reliabilities), are available for approximating $\mathbf{D}$ and $\mathbf{G}^{*}$. Furthermore, $\boldsymbol{\mu}_{0}$ is an unknown vector that needs to be estimated for some implementations. In this study, 4 different Bayesian implementations using gradually better approximations of prior mean and prior variance are compared (Table III-1). The differences were related to the animals providing external information and to the way that the prior mean and variance were defined. The first implementation, hereafter called Legarra-type Bayesian evaluation (LBE), was the simplest one from a computational standpoint, as only external PEV were considered to approximate $\mathbf{D}$. The second implementation, hereafter called Quaas-type Bayesian

evaluation (QBE), included covariances among traits. Both implementations defined prior means based on external EBV obtained from the genetic evaluation of all external animals based only on external data, which did not include relationships among internal animals. The QBE can be computationally simplified as shown in the third implementation, hereafter called first version of modified Bayesian evaluation (FBE). Finally, the last implementation, hereafter called second version of modified Bayesian evaluation (SBE), approximated and used the across-animal covariances that are not reported in practice while existing in $\mathbf{D}$.

Table III-1. Main differences concerning the prior mean and variance among Legarra-type Bayesian evaluation, Quaas-type Bayesian evaluation, first version of modified Bayesian evaluation, and second version of modified Bayesian evaluation

Item		Implementations[1]			
		LBE	QBE	FBE	SBE
Animals providing external EBV[1]		External	External	External, Internal	External, Internal
Prior mean					
External animals	Type	External EBV[2]	External EBV	External EBV	External EBV
	Origin	EEE[3]	EEE	EEE	EEE
Internal animals	Type	-	-	External EBV	External EBV
	Origin	-	-	SI[4]	SI
Prior variance					
Type		PEV	PEC[5]	PEC[5]	PEC[5]
Origin		EEE	EEE	JEE[6]	JEE
Relationships		-	Among external animals only	Among all external and internal animals	Among all external and internal animals
Double counting among external animals		-	-	-	Accounted

[1]LBE = Bayesian evaluation following Legarra et al. (2007) and using external EBV and prediction error variances (PEV) associated with external sires obtained from the external evaluation. QBE = Bayesian evaluation following Quaas and Zhang (2006) and using external EBV and PEV associated with external sires obtained from the external evaluation. FBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used. SBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used and the double counting among external animals was avoided.
[2]External EBV = EBV adjusted for base differences among external and internal information.
[3]EEE = genetic evaluation of all external animals based only on external data that did not include relationships among the internal animals.
[4]SI = selection index.
[5]PEC = prediction error covariances among traits (for FBE) and among traits and animals (for SBE).
[6]JEE = a posteriori joint genetic evaluation of all internal and external animals based only on external data.

LBE

Legarra et al. (2007) proposed a Bayesian implementation to integrate prior information into an internal genetic evaluation. The prior mean of the implementation was

defined as $\boldsymbol{\mu}_{\mathbf{E}} - \mathbf{U}_{\mathbf{E}}\mathbf{b}_{\mathbf{E}}$ where $\boldsymbol{\mu}_{\mathbf{E}} = \mathrm{E}(\mathbf{u}_{\mathbf{E}}|\mathbf{y}_{\mathbf{E}})$, $\mathbf{U}_{\mathbf{E}}$ is an incidence matrix relating base differences to external animals, and $\mathbf{b}_{\mathbf{E}}$ is a vector of base differences among external and internal EBV for all the external animals. The prior variance $\mathbf{D}$ was approximated by a diagonal matrix in which diagonal elements were equal to PEV associated to every external evaluation. Furthermore, this approximation of $\mathbf{D}$ implied another approximation to estimate $\boldsymbol{\Lambda}$. Because $\boldsymbol{\Lambda} = \mathbf{D}^{-1} - \mathbf{G}_{\mathbf{EE}}^{-1}$, all relationships needed to be ignored, and only diagonal elements of the matrix $\mathbf{G}_{\mathbf{EE}}$ were used. If nondiagonal elements in $\mathbf{G}_{\mathbf{EE}}$ were taken into account, the matrix $\mathbf{G}^{*}$ could be non-semi-positive definite (Legarra et al., 2007; Gengler and Vanderick, 2008).

From a computational standpoint, the LBE method is rather simple to set up as the matrix $\mathbf{D}$ is considered diagonal. Gengler and Vanderick (2008) reported that LBE could be easily integrated into a test-day model for dairy cattle genetic evaluations with few modifications of the code of the used programs and with reasonable convergence. However, the method needs to compute the base differences among external and internal information, which was estimated by Gengler and Vanderick (2008) before using the external information. This strategy avoids the computationally expensive integrated estimation of base differences.

QBE

Quaas and Zhang (2006) developed another Bayesian procedure to incorporate external information into a multibreed evaluation. They used a prior mean defined as $\boldsymbol{\mu}_{\mathbf{E}} - \mathbf{U}_{\mathbf{E}}\mathbf{b}_{\mathbf{E}}$ and a prior variance $\mathbf{D}$ approximated by $\mathbf{D} \approx \mathrm{Var}(\mathbf{u}_{\mathbf{E}}|\mathbf{y}_{\mathbf{E}}) = \mathrm{PEV}(\mathbf{u}_{\mathbf{E}}|\mathbf{y}_{\mathbf{E}})$. Hence, following Quaas and Zhang (2006) and equation (III.2), the matrix $\mathbf{D}^{-1}$ was equal to $\mathbf{D}^{-1} = \mathbf{G}_{\mathbf{EE}}^{-1} + \boldsymbol{\Lambda} = \left(\mathbf{A}_{\mathbf{EE}}^{-1} \otimes \mathbf{G}_{\mathbf{0}}^{-1}\right) + \boldsymbol{\Lambda}$, where $\mathbf{A}_{\mathbf{EE}}^{-1}$ was the inverse of the matrix that only accounts for the relationships among external animals and $\boldsymbol{\Lambda}$ was taken as a block diagonal variance matrix with one block for each external animal. The different block diagonals are equal to $\boldsymbol{\Delta}_{\mathbf{i}}\mathbf{G}_{\mathbf{0}}^{-1}\boldsymbol{\Delta}_{\mathbf{i}}$ for i = 1, 2, …, N with N external animals. The matrix $\mathbf{G}_{\mathbf{0}}$ is a matrix of genetic (co)variances among traits, and $\boldsymbol{\Delta}_{\mathbf{i}}$ is a diagonal matrix with elements $\sqrt{\delta_{ij}}$ with j = 1, 2, …, n traits. The element δ_{ij} is equal to the ratio of $REL_{ij}/(1-REL_{ij})$, where REL_{ij} is the reliability associated to the external proof $\boldsymbol{\mu}_{\mathbf{E}}$ for the jth trait of ith external animal.

The QBE implementation estimates the base differences between external and internal EBV in a different way to equation (III.1). Base differences in QBE are estimated as $\hat{\mathbf{b}}_{\mathbf{E}} = -(\mathbf{U'}_{\mathbf{E}}\,\mathbf{D}^{-1}\mathbf{U}_{\mathbf{E}})^{-1}\mathbf{U'}_{\mathbf{E}}\,\mathbf{D}^{-1}(\hat{\mathbf{u}}_{\mathbf{E}} - \boldsymbol{\mu}_{\mathbf{E}})$ (Zhang et al., 2002; Quaas and Zhang, 2006). If $\mathbf{U}$ is partitioned between external animals ($\mathbf{U}_{\mathbf{E}}$) and internal animals ($\mathbf{U}_{\mathbf{I}}$) and by replacing $\mathbf{G}^{*-1}$ by $\begin{bmatrix} \mathbf{G}^{\mathbf{EE}} + \boldsymbol{\Lambda} & \mathbf{G}^{\mathbf{EI}} \\ \mathbf{G}^{\mathbf{IE}} & \mathbf{G}^{\mathbf{II}} \end{bmatrix}$ in the mixed model equations (III.1), it can be shown (Appendix III-1) that estimation of $\hat{\mathbf{b}}_{\mathbf{E}}$ by QBE is equivalent to the computation of $\hat{\mathbf{b}}$ using mixed model equations (III.1). Except for this difference, differences between approximations of LBE and QBE mainly concern the matrix $\mathbf{D}$ and the consideration of the whole (co)variances matrix $\mathbf{G}_{\mathbf{EE}}$.

It is important to note that, from equations (III.2) and (III.3), $\boldsymbol{\Lambda} = \mathbf{Z'}_{\mathbf{E}}\,\mathbf{R}_{\mathbf{E}}^{-1}\mathbf{Z}_{\mathbf{E}}$. For the *j*th trait of *i*th external animal, the diagonal element of the matrix $\mathbf{Z'}_{\mathbf{E}}\,\mathbf{R}_{\mathbf{E}}^{-1}\mathbf{Z}_{\mathbf{E}}$ is equal to the number of records the animal *i* has for this trait multiplied by the inverse of the error variance of this *j*th trait $\sigma^2_{e_j}$ (Mrode, 2005). However, this number of records can be estimated by the effective number of records, so-called records equivalent (RE), as $RE_{ij} = \frac{\sigma^2_{e_j}}{\sigma^2_{u_j}} * \delta_{ij}$ where $\sigma^2_{u_j}$ is the genetic variance for the *j*th trait. Thereby, the diagonal elements $\sqrt{\delta_{ij}}$ are equal to $\sqrt{RE_{ij} * \frac{\sigma^2_{u_j}}{\sigma^2_{e_j}}}$.

Furthermore, QBE (Table III-1) also needs the computation of the inverse of the relationship matrix $\mathbf{A}_{\mathbf{EE}}$ that only accounts for the relationships among external animals, $\mathbf{A}_{\mathbf{EE}}^{-1}$. The matrix $\mathbf{A}_{\mathbf{EE}}^{-1}$ could be computed efficiently by first establishing directly $\mathbf{A}_{\mathbf{EE}}$ through an algorithm based on Colleau (2002) followed by its inversion with optimized subroutines (Misztal et al., 2009; Aguilar et al., 2011). However, $\mathbf{A}_{\mathbf{EE}}$ might be dense, and its storage, as well as its inversion, might not be possible or could take too much computational burden because the number of external animals could be very high. Furthermore, the direct computation of $\mathbf{A}_{\mathbf{EE}}^{-1}$ might not be possible using simplified rules, as relationships among all ancestors without external breeding will be absorbed in this matrix. Given these differences, QBE is slightly more complicated to implement than LBE.

FBE

The definition of prior mean and variance in QBE has the shortcoming that, as noted above, the computation of $\mathbf{A}_{\mathbf{EE}}^{-1}$ is more difficult than the establishment of the inverse of the relationship matrix among all external and internal animals. The consideration of all animals has also the effect that the definition of prior mean needs to include external EBV for all animals. To consider these issues, FBE was developed. The approximation concerns the terms of the left hand side of the equation (III.4) instead of the terms of the right hand side, as is done in LBE and QBE. The unknown vector $\boldsymbol{\mu}_{\mathbf{0}}$ can be approximated as follows. Let $\boldsymbol{\mu}_{\mathbf{I}}$ be an unknown vector of external EBV of all the internal animals of a joint genetic evaluation of all internal and external animals based only on external data. Because this evaluation is only based on external data, and because $\boldsymbol{\mu}_{\mathbf{0}} = \left[\boldsymbol{\mu}_{\mathbf{E}}' \quad \boldsymbol{\mu}_{\mathbf{I}}'\right]'$ and $p(\boldsymbol{\mu}_{\mathbf{I}} | \boldsymbol{\mu}_{\mathbf{E}}) = MVN\left(\mathbf{G}_{\mathbf{IE}}\mathbf{G}_{\mathbf{EE}}^{-1}\boldsymbol{\mu}_{\mathbf{E}}, (\mathbf{G}^{\mathbf{II}})^{-1}\right)$, $\boldsymbol{\mu}_{\mathbf{I}}$ can be approximated as well as $\boldsymbol{\mu}_{\mathbf{0}}$. Therefore, all internal and external animals are considered as having external information. As detailed in Table III-1, this feature distinguishes the 2 implementations found in the literature (LBE and QBE) and the new implementations FBE and SBE. In these 2 implementations, FBE and SBE, the RE associated with the predicted breeding values are set to 0, because the predicted breeding values are only based on relationships and do not bring any additional information. Hence, because all internal and external animals are considered as external animals, $\mathbf{G}^{*-1} = \begin{bmatrix} \mathbf{G}^{\mathbf{EE}} + \boldsymbol{\Lambda} & \mathbf{G}^{\mathbf{EI}} \\ \mathbf{G}^{\mathbf{IE}} & \mathbf{G}^{\mathbf{II}} \end{bmatrix}$ and $\boldsymbol{\Lambda} = \mathbf{Z}'_{\mathbf{E}}\mathbf{R}_{\mathbf{E}}^{-1}\mathbf{Z}_{\mathbf{E}}$, $\mathbf{G}^{*-1}$ can be simplified as $\mathbf{G}^{*-1} = \mathbf{G}^{-1} + \boldsymbol{\Lambda}$, where the diagonal blocks for the internal animals of the matrix $\boldsymbol{\Lambda}$ of FBE are equal to zero and the diagonal blocks for the external animals of the matrix $\boldsymbol{\Lambda}$ of FBE are equal to the corresponding diagonal blocks of $\boldsymbol{\Lambda}$ used by QBE.

From a computational standpoint, FBE has different advantages. First, because all the animals are considered to have external information, the inverse of the whole relationship matrix $\mathbf{A}$ is also used in the right hand side of FBE instead of the relationship matrix $\mathbf{A}_{\mathbf{EE}}^{-1}$ used by QBE to estimate $\mathbf{D}^{-1}$. This extension of the relationship matrix requires an estimation of external breeding values for the internal animals (e.g., using selection index theory). However, this estimation is computationally feasible. The extension of the external breeding values leads to the setup of only a single inverted relationships matrix. Second, the integration of prior information for all animals in FBE

leads to a hidden advantage for this implementation. Because all internal animals are associated with prior information, the prediction of their breeding values through FBE is influenced by the same constant difference that may exist between external information and breeding values that have to be estimated. Therefore, the equation to estimate genetic base differences among the different evaluations can be eliminated from the system of equations, because this effect becomes confounded with the general mean. A proof of this is given in the Appendix III-1. Both advantages and differences of FBE in comparison to LBE and QBE allow large simplifications of the computational burden during setup and solving of mixed model equations, making their use again easier with complicated models and large data sets.

SBE

The knowledge of only the external PEV, or functions of these, means that the across-animal covariances are not correctly considered by LBE, QBE, and FBE, which leads to double counting among external animals. In this context, double counting means multiple considerations of parts of integrated external EBV for related external animals. In SBE, double counting is taken into account through an additional two-step algorithm (TSA), for which the aim is to estimate corrected RE for the external animals independent from contributions due to relationships. Hence, only the approximation of $\mathbf{\Lambda}$ for the external animals changes in SBE compared with FBE. So, the block diagonal of $\mathbf{\Lambda}$ for each internal animal *i* is equal to zero as those of the third implementation FBE, while the block diagonal of $\mathbf{\Lambda}$ for each external animal *i* is equal to $\mathbf{\Delta_i G_0^{-1} \Delta_i}$, where $\mathbf{\Delta_i}$ is a diagonal matrix with elements $\sqrt{\mathbf{RE}^*_{\mathbf{j}ii} * \frac{\sigma^2_{u_j}}{\sigma^2_{e_j}}}$, where *j* = 1, 2, …, *n* traits, and $\mathbf{RE^*_j}$ is a diagonal matrix with diagonal elements equal to RE only due to own records for the *j*th trait.

Double counting among external animals can appear if external information - in this context, external EBV - of an animal and a close relative are integrated into the same genetic evaluation. This double counting is due to the fact that external information of those animals combines contributions due to own records and due to relationships (Misztal and Wiggans, 1988; VanRaden, 2001). To avoid this double counting, it is necessary to separate the contributions due to records from the contributions due to relationships in the external information for each external animal using TSA, in which the

2 steps are based on the algorithm A1 of Misztal and Wiggans (1988). However, the aim of the current study was different from that of Misztal and Wiggans (1988), and some modifications were necessary. The first modification concerns the fact that the 2 steps of the TSA include all relationships between external animals and their ancestors instead of only the relationships between an animal and its parents. Second, the estimated RE due to records by the algorithm A1 is obtained from a model in which all effects are absorbed into animals' effects. This leads to lower RE than the corresponding diagonal elements of the matrix $\mathbf{Z'}_{\mathbf{E}}\,\mathbf{R}_{\mathbf{E}}^{-1}\mathbf{Z}_{\mathbf{E}}$. To resolve this problem, an absorption matrix $\mathbf{M}$ is created from the RE due to records estimated by the first step of the TSA. Therefore, for the *j*th trait, the first step of the TSA separates contributions due to records and contributions due to relationships following the algorithm A1 of Misztal and Wiggans (1988). Based on the contributions due to records, an absorption matrix $\mathbf{M}$ has to be developed that is taken into account by the second step of the TSA to estimate RE for the external animals independently from contributions due to relationships or correlated traits. The TSA must be repeated for each trait and is detailed in the Appendix III-1.

The SBE shares the advantages with FBE explained above. However, as already explained, an additional advantage is that theoretically all double counting among external animals is avoided. The disadvantage is that the TSA needs to be implemented, which may be computationally challenging.

SIMULATED DATA

The 4 different Bayesian methodologies described above were tested using simulated data. For this purpose, an external and an internal population were each simulated from 30 male founders and 120 female founders. Each population included about 1000 animals distributed over 5 generations. For each population, the sires were randomly selected from available males for each generation. The maximum number of males mated in each generation was 25. All females existing in the pedigree were randomly mated with the selected males to simulate each new generation. However, these matings could not be realized if the coefficient of relationship between 2 animals was 0.5 or higher, as well as if the female had already 3 descendants. Furthermore, a male could be mated during at most 2 years.

In regard to the external population, external females were randomly mated only with external males. In each generation, 60% of external male offspring were randomly culled. In regard to the internal population, internal females were randomly mated with

internal males and a subset of external males. This subset included the first 50 sires that had the most offspring in the external population. In each generation, 99% of internal male offspring were randomly culled.

As the phenotypic trait, milk production for the first lactation was simulated for each female in both populations following Van Vleck (1994). A nested herd effect within population was randomly assigned to each record under the condition that each herd included about 40 females. Phenotypic variance and heritability were assumed to be 3.24 * 10^6 kg^2 and 0.25, respectively.

Using the simulated data, the following 7 genetic evaluations were performed:

(1) The joint evaluation was a regular BLUP evaluation based on external and internal pedigree and data. This evaluation was assumed the reference.

(2) The external evaluation was a regular BLUP evaluation based on external pedigree and data.

(3) The internal evaluation was a regular BLUP evaluation based on internal pedigree and data.

Concerning the 4 Bayesian evaluations,

(4) the Legarra-type Bayesian evaluation was a LBE using external EBV and PEV associated with external sires, obtained from external evaluation (2) inside the internal evaluation, and

(5) the Quaas-type Bayesian evaluation was a QBE using external EBV and PEV associated with external sires, obtained from external evaluation (2) inside the internal evaluation.

(6) The first version of modified Bayesian evaluation was an FBE using external EBV and PEV associated with external sires inside the internal evaluation where external EBV for all animals (internal and external) were predicted and used, and

(7) the second version of modified Bayesian evaluation was an SBE using external EBV and PEV associated with external sires inside the internal evaluation, where external EBV and PEV for all animals (internal and external) were predicted and used but applying the TSA algorithm to avoid double counting.

The simulation was replicated 100 times. For external and internal animals, comparisons between the joint evaluation and the 6 others were based on (1) Spearman rank correlation coefficients (r), (2) mean squared errors (MSE) expressed as a percentage of internal MSE, (3) regression coefficients (a), and on (4) coefficients of determination (R^2). All parameters were the average of 100 replicates.

RESULTS AND DISCUSSION

The 100 simulated external and internal populations included 1052 animals each on average. The external information integrated into the internal genetic evaluation for one external animal corresponded to 10 effective daughters on average. This number of effective daughters may seem low, but it is the lower bound of the effective number of daughters one might expect when a sire is evaluated from genomic prediction. Results for r, MSE, a, and R^2 illustrating the prediction of joint breeding values are shown in Table III-2 for the external animals (i.e., the 50 external sires associated with external information integrated through a Bayesian evaluation), and in Table III-3 for all the internal animals (i.e., animals associated with only internal information). To visualize effects of the integration of external information, EBV of the 50 external animals for one randomly chosen simulation are plotted in Figure III-1.

Table III-2. Rank correlations (r) and mean squared errors (MSE) expressed as a percentage of the internal MSE between joint evaluation and an external evaluation, an internal evaluation, and 4 different Bayesian procedures, regression coefficients (a), and coefficients of determination (R^2) of the regression of the joint evaluation on the 6 other evaluations[1]

Genetic evaluation	r ± SD	MSE ± SD	a ± SD	R^2 ± SD
Without external information[2]				
Internal	0.54 ± 0.12	100.00 ± 28.20	0.72 ± 0.16	0.35 ± 0.13
External	0.78 ± 0.08	55.06 ± 19.49	0.82 ± 0.10	0.64 ± 0.12
With external information[3]				
LBE	0.96 ± 0.02	11.76 ± 6.32	0.88 ± 0.05	0.94 ± 0.02
QBE	0.99 ± 0.005	1.33 ± 0.62	0.99 ± 0.02	0.99 ± 0.004
FBE	>0.99 ± 0.002	0.49 ± 0.25	1.01 ± 0.01	>0.99 ± 0.002
SBE	>0.99 ± 0.001	0.19 ± 0.09	1.00 ± 0.01	>0.99 ± 0.001

[1]All data are presented for external animals associated to external information integrated through a Bayesian evaluation. Reported results are averages and standard deviations over 100 replicates.
[2]Internal = internal genetic evaluation; external = external genetic evaluation.
[3]LBE = Bayesian evaluation following Legarra et al. (2007) and using external EBV and prediction error variances (PEV) associated with external sires obtained from the external evaluation. QBE = Bayesian evaluation following Quaas and Zhang (2006) and using external EBV and PEV associated with external sires obtained from the external evaluation. FBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used. SBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used and the double counting among external animals was avoided.

First, concerning the 50 external animals, rank correlations between joint evaluation and the 4 Bayesian implementations increased at least by 43% to be >0.96. Therefore, the integration of external information led to an improved ranking of external

animals in the internal evaluation (i.e., more similar ranking compared with the ranking of the joint evaluation), which was expected, especially by Legarra et al. (2007) and Quaas and Zhang (2006). Concerning all the internal animals, even if rank correlations increased only by 4%, integrationof external information for external animals related to the internal population led to rank internal animals almost identically to their ranking obtained with the joint evaluation.

Table III-3. Rank correlations (r) and mean squared errors (MSE) expressed as a percentage of the internal MSE between joint evaluation, and an internal evaluation and 4 different Bayesian procedures, and regression coefficients (a), and coefficients of determination (R^2) of the regression of the joint evaluation on the 6 other evaluations[1]

Genetic evaluation	r ± SD	MSE ± SD	a ± SD	R^2 ± SD
Without external information				
Internal[2]	0.95 ± 0.02	100.00 ± 33.52	0.95 ± 0.03	0.91 ± 0.03
With external information[3]				
LBE	0.99 ± 0.003	12.48 ± 6.27	0.98 ± 0.01	0.99 ± 0.01
QBE	>0.99 ± 0.000	1.36 ± 0.71	1.00 ± 0.004	>0.99 ± 0.001
FBE	>0.99 ± 0.000	0.79 ± 0.52	1.00 ± 0.003	>0.99 ± 0.000
SBE	>0.99 ± 0.000	0.26 ± 0.23	1.00 ± 0.002	>0.99 ± 0.000

[1]All data are presented for internal animals associated to only internal information. Reported results are averages and standard deviations over 100 replicates.
[2]Internal = internal genetic evaluation.
[3]LBE = Bayesian evaluation following Legarra et al. (2007) and using external EBV and prediction error variances (PEV) associated with external sires obtained from the external evaluation. QBE = Bayesian evaluation following Quaas and Zhang (2006) and using external EBV and PEV associated with external sires obtained from the external evaluation. FBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used. SBE = Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used and the double counting among external animals was avoided.

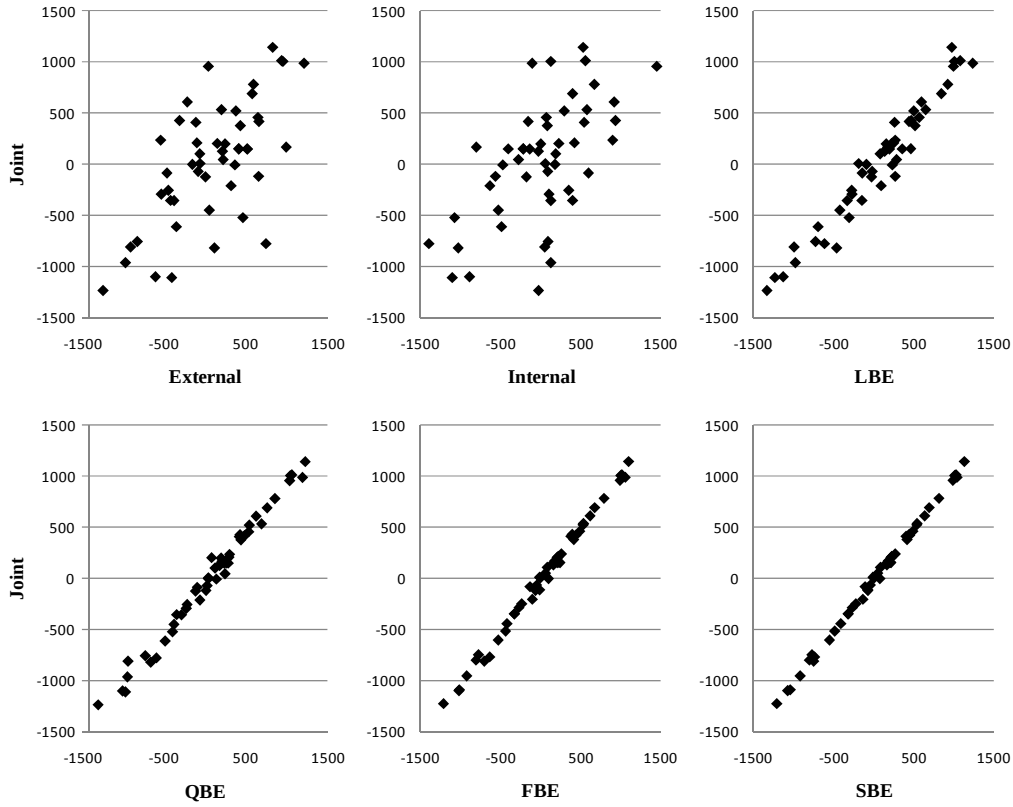

Figure III-1. Examples from one randomly chosen simulation showing EBV of the 50 external animals between joint evaluation and external and internal evaluations, Legarra-type Bayesian evaluation (LBE; i.e., a Bayesian evaluation following Legarra et al. (2007) and using external EBV and prediction error variances (PEV) associated with external sires obtained from the external evaluation), Quaas-type Bayesian evaluation (QBE; i.e., a Bayesian evaluation following Quaas and Zhang (2006) and using external EBV and PEV associated with external sires obtained from the external evaluation), first version of modified Bayesian evaluation (FBE; i.e., a Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used), and second version of modified Bayesian evaluation (SBE; i.e., a Bayesian evaluation using external EBV and PEV associated with external sires obtained from the external evaluation where external EBV for all internal and external animals were predicted and used and the double counting among external animals was avoided).

Second, according to the 4 estimated parameters r, MSE, a, and R^2, the integration of external information for the 50 external animals led to better predictions of the joint evaluation through all Bayesian implementations for all 50 external animals as well as for all internal animals. However, whereas integrated external information was identical for the 4 Bayesian implementations, significant differences were found among the 4 Bayesian procedures concerning prediction accuracy for breeding values. Breeding value prediction compared with the reference method (i.e., the joint evaluation) was poorest for the LBE

method for the 50 external animals as well as for the internal animals. This can be explained by the approximation of the matrix $\mathbf{D}$ by LBE. It approximates the latter matrix by a diagonal matrix in which diagonal elements are equal to PEV, ignoring prediction error covariances associated with every external evaluation. In contrast, the 4 parameters associated with FBE and SBE showed that integration of all relationships between the 50 external animals and all the internal animals for the approximation of $\mathbf{D}$ allowed the propagation of external information through the whole pedigree. Consequently, internal animals related to their external relatives were predicted better, too. This propagation is not possible in current methods based on selection index theory (VanRaden, 2001; Gengler and VanRaden, 2008), where information is combined on an animal-by-animal basis. Therefore, FBE and SBE have here a clear advantage compared with current methods. Furthermore, integration of all relationships allowed us to compute only one relationship matrix $\mathbf{A}^{-1}$ that takes into account all relationships among internal and external animals, whereas QBE needs the computation of the matrix $\mathbf{A}_{\mathbf{EE}}^{-1}$ as well as the matrix $\mathbf{A}^{-1}$. One can assume that numerically the setting up of $\mathbf{A}^{-1}$ using the usual rules is easier and numerically more stable than computing $\mathbf{A}_{\mathbf{EE}}^{-1}$ for potentially several thousands of animals.

Third, the 4 parameters showed that the SBE led to breeding values most similar to those estimated by the joint evaluation for the external animals. Values of R^2, a, and r were close to 1 with only few variation among replicates (SD <0.001). Mean squared error was the lowest of the 4 Bayesian implementations and showed that the application of TSA avoided double counting among external animals. Outliers of breeding values for external animals were limited. For the internal animals, QBE, FBE, and SBE were similar following the 4 parameters. Nevertheless, MSE was the lowest for SBE and showed the importance of the double counting among external animals on internal animals. However, with regards to r, a, and R^2 for the external and the internal animals estimated by FBE, double counting could be ignored if contributions due to relationships are low compared with contributions due to own records. If this is not the case, as for genomic information, TSA should be applied.

Fourth, no assumption was made about the difference of the amount of information between external and internal information. An external animal could get more information from the internal than from the external data. Integration of external information led to better predictions for breeding values obtained by the joint evaluation.

Therefore, integration of external information seems to be important even if the amount is low.

Finally, the developed methods could be used in different settings. Many situations exist where local (internal) evaluations would benefit from the integration of external information (e.g., Gengler and Vanderick, 2008). Because the developed methods can be used for multi-trait and other complex models, they allow the use of external information to improve the accuracy of evaluations for correlated, but only locally available traits as fine milk composition traits, such as free fatty acids, milk proteins, and other minor constituants (e.g., Gengler et al., 2010). In the context of genomic selection, integration of external genomic information into routine genetic evaluations could be done using the proposed methods after some adaptations (e.g., Gengler and Verkenne, 2007). Furthermore, as an anonymous reviewer reported, the matrix $\mathbf{G}^{*-1}$ is very similar to the inverse of the matrix $\mathbf{H}$ used in the single-step genomic evaluations and included both pedigree-based relationships and differences between pedigree-based and genomic-based relationships (Aguilar et al., 2010; Christensen and Lund, 2010). In a different setting and after taking precautions to avoid double counting because of the use of the same data, regular genetic evaluation results from a larger population could also be used as external priors in gene effect discovery studies (e.g., Buske et al, 2010) or any other studies requiring accurate estimation of a polygenic effect jointly with marker, single nucleotide polymorphisms, or gene effects.

CONCLUSIONS

According to these results, rankings of animals were most similar to those of a joint evaluation after the integration of all relationships and the application of the TSA to avoid double counting among external animals through SBE. It proved that the TSA worked well, although the creation of the absorption matrix $\mathbf{M}$ did not take into account the fixed effects considered in the external evaluation, which were unknown, but only one hypothetical unobserved fixed effect. The results based on our simulation showed that the Bayesian procedures FBE and QBE also worked well, with FBE having some computational advantages. Finally, with some adaptations and adjustments, FBE and SBE could be applied to integrate external information into routine genetic evaluations, SBE having additional advantages but being computationally more demanding.

ACKNOWLEDGMENTS

This research received financial support from the European Commission, Directorate-General for Agriculture and Rural Development, under Grand Agreement 211708 and form the Commission of the European Communities, FP7, KBBE-2007-1. This paper does not necessarily reflect the view of these institutions and in no way anticipates the Commission's future policy in this area. Jeremie Vandenplas, research fellow, and Nicolas Gengler, senior research associate, acknowledge the support of the National Fund for Scientific Research (Brussels, Belgium) for these positions and for the additional grants 2.4507.02F, F4552.05, and 2.4.623.08.F. Additional financial support was provided by the Ministry of Agriculture of Walloon Region of Belgium (Service Public de Wallonie, Direction générale opérationnelle "Agriculture, Ressources naturelles et Environnement" - DGARNE) through research projects D31-1207, D31-1224/S1, and D31-1233/S1. The authors are grateful to the University of Liege (SEGI facility) for the use of the NIC3 supercomputer facilities. The authors thank the members of the Animal Breeding and Genetics Group (Animal Science Unit, Gembloux Agro-Bio Tech, University of Liege), and in particular Bernd Buske and Ryan Ann Davis (University of Georgia, Athens) for reviewing the manuscript. Useful comments from three anonymous reviewers are also acknowledged.

REFERENCES

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128:422–428.

Buske, B., M. Szydlowski, and N. Gengler. 2010. A robust method for simultaneous estimation of single gene and polygenic effects in dairy cows using externally estimated breeding values as prior information. *J. Anim. Breed. Genet.* 127:272–279.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.

Colleau, J.-J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol.* 34:409–421.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Gengler, N., S. Vanderick, V. Arnould, and H. Soyeurt. 2010. Genetic evaluation for milk fat composition in the Walloon Region of Belgium. *Interbull Bull.* 42:81–84.

Gengler, N., and P.M. VanRaden. 2008. Strategies to incorporate genomic prediction into populationwide genetic evaluations. *J. Dairy Sci.* 91(E-Suppl. 1):506. (Abstr.).

Gengler, N., and C. Verkenne. 2007. Bayesian approach to integrate molecular data into genetic evaluations. *Interbull Bull.* 37:37–41.

Henderson, C.R. 1984. Applications of linear models in animal breeding. 2nd ed. University of Guelph, Guelph, ON, Canada.

Klei, L., R.L. Quaas, E.J. Pollak, and B.E. Cunningham. 1996. Multiple-breed evaluation. Pages 106–113 in Proceedings of the 28th Research Symposium and Annual Meeting. Beef Improvement Federation.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

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.

Misztal, I., and G.R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71(Suppl. 2):27–32.

Mrode, R.A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publishing, Wallingford, UK.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Quaas, R.L., and Z.W. Zhang. 2001. Incorporating external information in multibreed genetic evaluation. *J. Anim. Sci.* 79(Suppl. 1):342. (Abstr.).

VanRaden, P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. *J. Dairy Sci.* 84 (E. Suppl.):E47–E55.

Van Vleck, L.D. 1994. Algorithms for simulation of animal models with multiple traits

and with maternal and non-additive genetic effects. *Braz. J. Genet.* 17:53–57.

Zhang, Z.W., R.L. Quaas, and E.J. Pollak. 2002. Simulation study on the effects of incorporating external genetic evaluation results. CD–ROM Comm. 20–14 in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

APPENDIX III-1

EQUIVALENCE OF MIXED MODEL EQUATIONS CONSIDERING THE ESTIMATION OF BASE DIFFERENCES

Assume that external information is available for both internal and external animals from a joint genetic evaluation of all internal and external animals based only on external data and that the vectors of the base differences between the internal genetic evaluation and the joint genetic evaluation are $\hat{\mathbf{b}}_{\mathbf{E}}$ for the external animals and $\hat{\mathbf{b}}_{\mathbf{I}}$ for the internal animals. Therefore, the Bayesian mixed model equations (III.1) can be written as

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_{IE}} & \mathbf{X'_I R_I^{-1} Z_{II}} & \mathbf{0} & \mathbf{0} \\ \mathbf{Z'_{IE} R_I^{-1} X_I} & \mathbf{Z'_{IE} R_I^{-1} Z_{IE} + G^{EE} + \Lambda} & \mathbf{G^{EI}} & \mathbf{(G^{EE} + \Lambda)U_E} & \mathbf{G^{EI}U_I} \\ \mathbf{Z'_{II} R_I^{-1} X_I} & \mathbf{G^{IE}} & \mathbf{Z'_{II} R_I^{-1} Z_{II} + G^{II}} & \mathbf{G^{IE}U_E} & \mathbf{G^{II}U_I} \\ \mathbf{0} & \mathbf{U'_E (G^{EE} + \Lambda)} & \mathbf{U'_E G^{EI}} & \mathbf{U'_E (G^{EE} + \Lambda)U_E} & \mathbf{U'_E G^{EI}U_I} \\ \mathbf{0} & \mathbf{U'_I G^{IE}} & \mathbf{U'_I G^{II}} & \mathbf{U'_I G^{IE}U_E} & \mathbf{U'_I G^{II}U_I} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{I}} \\ \hat{\mathbf{u}}_{\mathbf{E}} \\ \hat{\mathbf{u}}_{\mathbf{I}} \\ \hat{\mathbf{b}}_{\mathbf{E}} \\ \hat{\mathbf{b}}_{\mathbf{I}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'_I R_I^{-1} y_I} \\ \mathbf{Z'_{IE} R_I^{-1} y_I + (G^{EE} + \Lambda)\mu_E + G^{EI}\mu_I} \\ \mathbf{Z'_{II} R_I^{-1} y_I + G^{IE}\mu_E + G^{II}\mu_I} \\ \mathbf{U'_E (G^{EE} + \Lambda)\mu_E + U'_E G^{EI}\mu_I} \\ \mathbf{U'_I G^{IE}\mu_E + U'_I G^{II}\mu_I} \end{bmatrix}$$

where $\mathbf{Z_{IE}}$ and $\mathbf{Z_{II}}$ are the incidence matrices for the external and internal animals, respectively.

Because $\boldsymbol{\Lambda} = \mathbf{D^{-1}} - \mathbf{G_{EE}^{-1}}$, $\mathbf{G^{IE}} - \mathbf{G^{II}(G^{II})^{-1}G^{IE}} = \mathbf{0}$, $\mathbf{G^{EE}} - \mathbf{G^{EI}(G^{II})^{-1}G^{IE}} = \mathbf{G_{EE}^{-1}}$ and $\hat{\mathbf{u}}_{\mathbf{I}} = -\mathbf{(G^{II})^{-1}G^{IE}}\hat{\mathbf{u}}_{\mathbf{E}}$, the development of the fourth equation leads to $\mathbf{U'_E D^{-1}}(\hat{\mathbf{u}}_{\mathbf{E}} - \boldsymbol{\mu}_{\mathbf{E}}) + \mathbf{U'_E (G^{EE} + \Lambda)U_E}\hat{\mathbf{b}}_{\mathbf{E}} + \mathbf{U_E G^{EI} U_I}\hat{\mathbf{b}}_{\mathbf{I}} = \mathbf{0}$, and the development of the fifth equation leads to $\mathbf{U_I}\hat{\mathbf{b}}_{\mathbf{I}} = -\mathbf{(G^{II})^{-1}G^{IE}U_E}\hat{\mathbf{b}}_{\mathbf{E}}$.

After absorption of the fifth equation, the vector $\hat{\mathbf{b}}_{\mathbf{E}}$ is estimated as $\hat{\mathbf{b}}_{\mathbf{E}} = (\mathbf{U'_E D^{-1} U_E})^{-1}\mathbf{U'_E D^{-1}}(\boldsymbol{\mu}_{\mathbf{E}} - \hat{\mathbf{u}}_{\mathbf{E}})$. Furthermore, it can be shown that $\mathbf{G^{IE}}\boldsymbol{\mu}_{\mathbf{E}} + \mathbf{G^{II}}\boldsymbol{\mu}_{\mathbf{I}} = \mathbf{0}$ and $\mathbf{(G^{EE} + \Lambda)}\boldsymbol{\mu}_{\mathbf{E}} + \mathbf{G^{EI}}\boldsymbol{\mu}_{\mathbf{I}} = \mathbf{D^{-1}}\boldsymbol{\mu}_{\mathbf{E}}$. The equivalent mixed model equations can be written

as:

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_{IE}} & \mathbf{X'_I R_I^{-1} Z_{II}} & \mathbf{0} \\ \mathbf{Z'_{IE} R_I^{-1} X_I} & \mathbf{Z'_{IE} R_I^{-1} Z_{IE} + G^{EE} + \Lambda} & \mathbf{G^{EI}} & \mathbf{D^{-1} U_E} \\ \mathbf{Z'_{II} R_I^{-1} X_I} & \mathbf{G^{IE}} & \mathbf{Z'_{II} R_I^{-1} Z_{II} + G^{II}} & \mathbf{0} \\ \mathbf{0} & \mathbf{U'_E D^{-1}} & \mathbf{0} & \mathbf{U'_E D^{-1} U_E} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_I \\ \hat{\mathbf{u}}_E \\ \hat{\mathbf{u}}_I \\ \hat{\mathbf{b}}_E \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} y_I} \\ \mathbf{Z'_{IE} R_I^{-1} y_I + D^{-1} \boldsymbol{\mu}_E} \\ \mathbf{Z'_{II} R_I^{-1} y_I} \\ \mathbf{U'_E D^{-1} \boldsymbol{\mu}_E} \end{bmatrix} .$$

ELIMINATION OF BASE DIFFERENCE EQUATIONS IN FBE AND SBE

The derivation is based on the estimation of $\hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}$ instead of $\hat{\mathbf{u}}$ and $\hat{\mathbf{b}}$ separately. The associated mixed model equations can be obtained through a few steps. First, using some rearrangements, the development according to the first, second and third lines of the Bayesian mixed model equations (III.1) leads to, respectively:

$$\mathbf{X'_I R_I^{-1} X_I \hat{\boldsymbol{\beta}}_I + X'_I R_I^{-1} Z_I}\left(\hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}\right) = \mathbf{X'_I R_I^{-1}}\left(\mathbf{y_I} + \mathbf{Z_I U}\hat{\mathbf{b}}\right) \quad \text{(A3-1)}$$

$$\mathbf{Z'_I R_I^{-1} X_I \hat{\boldsymbol{\beta}}_I} + \left(\mathbf{Z'_I R_I^{-1} Z_I + G^{*-1}}\right)\left(\hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}\right) - \mathbf{G^{*-1}\boldsymbol{\mu}_0} = \mathbf{Z'_I R_I^{-1}}\left(\mathbf{y_I} + \mathbf{Z_I U}\hat{\mathbf{b}}\right) \quad \text{(A3-2)}$$

$$\mathbf{U'G^{*-1}}\left(\hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}\right) - \mathbf{U'G^{*-1}\boldsymbol{\mu}_0} = \mathbf{0} \quad \text{(A3-3)}$$

Therefore, solutions for $\hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}$ and $\hat{\mathbf{b}}$ can be obtained by solving jointly (A3-4) and (A3-5):

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_I} \\ \mathbf{Z'_I R_I^{-1} X_I} & \mathbf{Z'_I R_I^{-1} Z_I + G^{*-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_I \\ \hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'_I R_I^{-1}}\left(\mathbf{y_I} + \mathbf{Z_I U}\hat{\mathbf{b}}\right) \\ \mathbf{Z'_I R_I^{-1}}\left(\mathbf{y_I} + \mathbf{Z_I U}\hat{\mathbf{b}}\right) + \mathbf{G^{*-1}\boldsymbol{\mu}_0} \end{bmatrix} \quad \text{(A3-4)}$$

$$\mathbf{U'G^{*-1}U}\hat{\mathbf{b}} = \mathbf{U'G^{*-1}}\left(\boldsymbol{\mu}_0 - \hat{\mathbf{u}}\right) \quad \text{(A3-5)}$$

If all the animals contribute to the base differences, then $\mathbf{U}$ represents a summing matrix and $\hat{\mathbf{b}}$ the vector of weighted average base differences between $\boldsymbol{\mu}_0$ and $\hat{\mathbf{u}}$. Given this, $\mathbf{U}\hat{\mathbf{b}}$ represents a vector of constants added to each EBV and $\mathbf{Z_I U}\hat{\mathbf{b}}$ represents a vector of constants added to each record. Therefore, the following reparameterization can be used: $\hat{\mathbf{u}}^* = \hat{\mathbf{u}} + \mathbf{U}\hat{\mathbf{b}}$.

Furthermore, adding the same constants to each EBV will not change the rankings, and rankings will be thereby invariant to the used constants. The constants added to the records will also only change estimates of fixed effects. Those different estimates of fixed effects will have no effect on animal rankings because all animals are affected by the

same constant. For these reasons, (A3-4) can be rewritten as follows:

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_I} \\ \mathbf{Z'_I R_I^{-1} X_I} & \mathbf{Z'_I R_I^{-1} Z_I + G^{*-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{I}}^{*} \\ \hat{\mathbf{u}}^{*} \end{bmatrix} = \begin{bmatrix} \mathbf{X'_I R_I^{-1} y_I} \\ \mathbf{Z'_I R_I^{-1} y_I + G^{*-1} \mu_0} \end{bmatrix}$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{I}}^{*}$ represents the new fixed effects computed by ignoring the constant in the records.

TSA

The estimation of RE independent from contributions due to relationships or correlated traits is performed by the following TSA; the TSA must be repeated for each trait.

The first step of the TSA is solved iteratively as follows:

1) For each animal i, $\mathbf{H}_{\mathbf{1}_{ii}}^{[\mathbf{0}]} = RE_{ij}$,

   where $\mathbf{H_1}$ is a diagonal matrix with RE of each external animal *i* based on the external PEV for the *j*th trait.

2) $\mathbf{Q}_{\mathbf{1}}^{[\mathbf{1}]} = \mathbf{H}_{\mathbf{1}}^{[\mathbf{0}]}$.

3) k=1.

4) $\mathbf{P}^{[\mathbf{k}]} = \left(\mathbf{Q}_{\mathbf{1}}^{[\mathbf{k}]} + \mathbf{A}^{*-1}\lambda_j\right)^{-1}$,

   where $\mathbf{A}^{*-1}$ is the inverse of the relationship matrix that accounts for the relationships between external animals and their ancestors and $\lambda_j = \frac{\sigma_{e_j}^2}{\sigma_{u_j}^2}$ for with $\sigma_{e_j}^2$ and $\sigma_{u_j}^2$ are the error variance and the genetic variance for the *j*th trait, respectively.

5) $\mathbf{H}_{\mathbf{2}}^{[\mathbf{k}]} = \left(diag\left(diag\left(\mathbf{P}^{[\mathbf{k}]}\right)\right)\right)^{-1} - \mathbf{I}\lambda$.

6) $\mathbf{S}^{[\mathbf{k}]} = \mathbf{H}_{\mathbf{1}}^{[\mathbf{0}]} - \mathbf{H}_{\mathbf{2}}^{[\mathbf{k}]}$.

7) If $\mathbf{S}^{[\mathbf{k}]}$ is not sufficient small, perform for each animal *i*:

   a) $\mathbf{Q}_{\mathbf{1}ii}^{[\mathbf{k+1}]} = \mathbf{Q}_{\mathbf{1}ii}^{[\mathbf{k}]} + \mathbf{S}^{[\mathbf{k}]}$.

b) If any diagonal element in $\mathbf{Q}_1^{[k+1]}$ is negative, set it to 0.

c) k=k+1.

d) Repeat from 4).

8) For each animal i, perform:

a) $\mathbf{X}_i = 1$ if $\mathbf{Q}_{1ii}^{[k]} \neq 0$.

b) $\mathbf{X}_i = 0$ if $\mathbf{Q}_{1ii}^{[k]} = 0$.

9) $\mathbf{M} = \mathbf{X_d} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$,

where the matrix **M** is the absorption matrix based on the contributions due to own records.

The second step of the TSA is solved iteratively as follows:

1) For each animal i, $\mathbf{H}_{1_{ii}}^{[0]} = RE_{ij}$.

2) $\mathbf{Q}_2^{[1]} = \sqrt{\mathbf{H}_1^{[0]}} * \mathbf{M} * \sqrt{\mathbf{H}_1^{[0]}}$.

3) k=1.

4) $\mathbf{P}^{[k]} = \left(\mathbf{Q}_2^{[k]} + \mathbf{A}^{*-1}\lambda_j\right)^{-1}$.

5) $\mathbf{H}_2^{[k]} = \left(diag\left(diag\left(\mathbf{P}^{[k]}\right)\right)\right)^{-1} - \mathbf{I}\lambda$.

6) $\mathbf{S}^{[k]} = \mathbf{H}_1^{[0]} - \mathbf{H}_2^{[k]}$.

7) If $\mathbf{S}^{[k]}$ is not sufficient small, perform for each animal i:

a) $\mathbf{Q}_{2ii}^{[k+1]} = \mathbf{Q}_{2ii}^{[k]} + \mathbf{S}^{[k]}$.

b) If any diagonal element in $\mathbf{Q}^{[k+1]}$ is negative, set it to 0.

c) $\mathbf{Q}_{2ij}^{[k+1]} = \sqrt{\mathbf{Q}_{2ii}^{[k+1]} * \mathbf{M}_{ii}^{-1}} * \mathbf{M}_{ij} * \sqrt{\mathbf{M}_{jj}^{-1} * \mathbf{Q}_{2jj}^{[k+1]}}$.

d) k=k+1.

e) Repeat from 4).

8) If $\mathbf{Q}_2^{[k+1]}$and $\mathbf{Q}_2^{[k]}$ are close enough, perform for each animal i,

$$\mathbf{RE}^{*}_{\mathbf{j}_{ii}} = \mathbf{Q}^{[\mathbf{k}+\mathbf{1}]}_{\mathbf{2}ii} * \mathbf{M}^{-\mathbf{1}}_{ii},$$

where $\mathbf{RE}^{*}_{\mathbf{j}}$ is a diagonal matrix with diagonal elements equal to RE only due to own records for the *j*th trait.

Chapter IV. An integration of external information for foreign stallions into the Belgian genetic evaluation for jumping horses

The previous Chapter proposed a Bayesian procedure to integrate external information into a genetic evaluation considering computational burden and double counting of contributions due to relationships among external animals. However, this procedure was tested only on simulated data. Therefore, the objective of this Chapter was to apply the proposed Bayesian procedure to integrate external information provided by France and the Netherlands for foreign stallions into the Belgian genetic evaluation for show jumping horses.

From: Vandenplas, J., S. Janssens, N. Buys, and N. Gengler. 2013. An integration of external information for foreign stallions into the Belgian genetic evaluation for jumping horses. *J. Anim. Breed. Genet.* 130:209–217. (Used by permission of the Journal of Animal Breeding and Genetics)

SUMMARY

The aim of this study was to test the integration of external information, i.e. foreign estimated breeding values (EBV) and the associated reliabilities (REL), for stallions into the Belgian genetic evaluation for jumping horses. The Belgian model is a bivariate repeatability Best Linear Unbiased Prediction animal model only based on Belgian performances, while Belgian breeders import horses from neighbouring countries. Hence, use of external information is needed as prior to achieve more accurate EBV. Pedigree and performance data contained 101 382 horses and 712 212 performances, respectively. After conversion to the Belgian trait, external information of 98 French and 67 Dutch stallions was integrated into the Belgian evaluation. Resulting Belgian rankings of the foreign stallions were more similar to foreign rankings according to the increase of the rank correlations of at least 12%. REL of their EBV were improved of at least 2% on average. External information was partially to totally equivalent to 4 years of contemporary horses' performances or to all the stallions' own performances. All these results showed the interest to integrate external information into the Belgian evaluation.

Key words: Bayesian approach, external information, integration, jumping horses

INTRODUCTION

The Belgian sport horse population is situated at the crossroads of different sport horse populations, which leads to a mix of the European genes. Artificial insemination facilitates the use of foreign stallions since the 1980s. For these reasons, Belgium seems to be one of the centres for European sport horse breeding (Ruhlmann et al., 2009a). Three Belgian studbooks of warmblood horses are involved, that is the Belgian Warmblood Horse Studbook (BWP), the Royal Belgian Sports Horse Society (sBs) and the Studbook Zangersheide. For all three, the improvement in the performances in show jumping is an important breeding objective (Koenen et al., 2004).

Since 1998, a genetic evaluation for show jumping horses is implemented in Belgium, and as in most other European countries, the estimated breeding values (EBV) are based on national information only (Koenen and Aldridge, 2002; Janssens et al., 2007), whereas the Belgian sport horse population is clearly linked with other foreign studbooks. This may lead to a limited reliability (REL) of EBV for horses with few Belgian records and to inappropriate breeders' choices of a stallion on the international

scene. Similar issues exist in other countries as sport horse breeding is very international. Based on experiences in dairy cattle breeding, for which 'Interbull' (Uppsala, Sweden) provides sire EBV for dairy cattle from different countries, an international group of scientists and breeding organizations, called 'Interstallion', was created in 1998 to achieve reliable breeding values across countries for sport horse stallions. Within the framework of this group, Ruhlmann et al. (2009b) concluded that an international evaluation of jumping horses is feasible. However, such an international genetic evaluation combining all information sources is not yet available, and one option is to integrate external information into the local genetic evaluation. Different theoretical approaches exist to do this. In the case of multibreed genetic evaluations for beef cattle, Klei et al. (1996) proposed a Bayesian approach where external information is considered as prior information for the local evaluation. Two different Bayesian derivations were proposed by Quaas and Zhang (2006) and Legarra et al. (2007). Recently, Vandenplas and Gengler (2012) proposed some improvements to these methods, especially to take into account the double counting among related external animals. However, some issues arise before the implementation of a Bayesian procedure, like the independence of the external evaluations from the internal one or the similarities between the external and internal evaluated traits (Gengler and Vanderick, 2008).

The first aim of this study was to apply a Bayesian approach to integrate external information, i.e., foreign EBV and their associated REL, for stallions into the Belgian genetic evaluation of show jumping horses, and the second aim was to test the model adequacy and the predictive ability of the applied method.

MATERIALS AND METHODS

Performance data on show jumping were provided by the horse riding organization for national level competitions, the Royal Belgian Federation for Equestrian Sports (KBRSF), and by the horse riding organization for recreational level competitions, the Rural Riding Association (LRV). The available performance data (data I) included 710 212 performances from 44 755 competitive horses during the period 1991–2009. Performances in show jumping consisted of ranking of horses participating in show jumping competitions converted into normalized score by a Blom's approximation (Janssens et al., 2007). These performances were also considered as two traits in terms of competition levels, that is the KBRSF level and the LRV level. The KBRSF level was considered as the Belgian breeding goal trait (hereafter called Belgian trait). The pedigree

file, a combination of pedigree records provided by sBs and BWP, included 101 382 registered horses. The following bivariate repeatability Best Linear Unbiased Prediction (BLUP) animal model was applied to perform the Belgian genetic evaluation (evaluation A; Janssens et al. 1997, 2007):

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Zp} + \mathbf{e} \qquad \text{(IV.1)}$$

where $\mathbf{y}$ was the vector of performances, $\boldsymbol{\beta}$ was the vector of fixed effects, $\mathbf{u}$ was the vector of random additive genetic effects, $\mathbf{p}$ was the vector of random permanent environmental effects and $\mathbf{e}$ was the vector of residuals. $\mathbf{X}$ and $\mathbf{Z}$ were incidence matrices relating performances to fixed effects and to random effects, respectively.

Fixed effects were the age of the participating horse, its sex and the show jumping event organized by the KBRSF or by the LRV in which it participated. Variance components for the random permanent environmental and genetic effects used for this study were those estimated by Janssens et al. (1997). Heritability was equal to 0.10 for performances at the KBRSF level (i.e. the Belgian trait) and to 0.11 for performances at the LRV level. Genetic correlation between these two traits was equal to 0.63.

Despite the fact that two traits are evaluated, only breeding values for performances in KBRSF level estimated using both performances in KBRSF level and LRV level (i.e. EBV estimated by the evaluation A for the Belgian trait; EBV_A) of stallions approved by BWP and/or sBs are published on a standardized scale following the recommendations of 'Interstallion' (Interstallion, 2005; Janssens and Buys, 2008).

Reliabilities of EBV_A based on data I (EBV_{AI}; REL_{AI}; Table IV-1) were computed using the equation:

$$REL = 1 - PEV/\sigma_g^2 \qquad \text{(IV.2)}$$

where σ_g^2 is the genetic variance for the Belgian trait and PEV is the prediction error variance obtained from the diagonal element of the inverted left hand side of the mixed model equations (IV.1).

Available external information consisted of external EBV (EBV_E) and their associated external REL (REL_E) for stallions approved by BWP having a published Belgian index, born after 1978 and originally registered in a Dutch or a French studbook. External information on 98 French stallions and 67 Dutch stallions was provided by the Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique (France) and the Royal Dutch Sport Horse (the Netherlands), respectively. However, because EBV_E were not the same trait and not expressed on the same scale as

the Belgian trait, precorrections were needed before its integration into the Belgian genetic evaluation.

Table IV-1. Performed genetic evaluations, estimated breeding values (EBV) and reliabilities (REL)

Genetic evaluation[1]	Datasets[2]		
	I	II	III
A	Evaluation AI, EBV_{AI}, REL_{AI}	Evaluation AII, EBV_{AII}, REL_{AII}	Evaluation AIII, EBV_{AIII}, REL_{AIII}
B	Evaluation BI, EBV_{BI}, REL_{BI}	Evaluation BII, EBV_{BII}, REL_{BII}	Evaluation BIII, EBV_{BIII}, REL_{BIII}

[1]A = Belgian genetic evaluation; B = Bayesian evaluation.
[2]I = complete data; II = data for which all performances after 31 December 2005 were assumed to be missing; III = data for which all the French and Dutch stallions' own performances were assumed to be missing.

First, EBV_E were converted to the Belgian trait and scale of the year 2009 for both countries. This conversion was performed separately for Dutch and French stallions following the method detailed by Goddard (1985) that regressed previously deregressed internal EBV on external EBV. The 2 samples to estimate conversion equations for the Dutch and French stallions included all Dutch stallions (i.e., 47) and French stallions (i.e., 93) having both an EBV_{AI} and an EBV_E, respectively. REL of the converted EBV (REL_c) were estimated from all the REL_E provided by France and by the Netherlands following the method detailed by Goddard (1985) that took into account the error in estimating the true regression equation and the variance of the converted EBV (EBV_c) about the true regression equation. External information with a REL_c lower than 0.01 was set to missing. It is noted that genetic correlation coefficients for traits between Belgium and the exporting countries were needed for the conversion following Goddard (1985). Because no genetic correlation coefficient was available for the pair Belgium/the Netherlands (Ruhlmann et al., 2009b), the genetic correlation coefficients for traits were approximated by the Pearson correlation coefficient between Dutch EBV_E and EBV_{AI} estimated for the Dutch stallions for the pair Belgium/the Netherlands and by the Pearson correlation coefficient between French EBV_E and EBV_{AI} estimated for the French stallions for the pair Belgium/France. Second, EBV_E had to be free from internal information to avoid double counting between external and internal information (Gengler and Vanderick, 2008). The literature review of Koenen (2002) and van Veldhuizen (1997) for the Dutch genetic evaluation and of Tavernier (1991) and Ricard (1997) for the French genetic evaluation showed that France and the Netherlands never use the same phenotypic

information (i.e., same show jumping competitions) as Belgium for their respective genetic evaluations for show jumping. Following the literature, this second condition seemed to be respected. Third, as external information was associated with related stallions, double counting of information among related external stallions could exist. Therefore, the integration of external information was performed following the second version of modified Bayesian evaluation detailed by Vandenplas and Gengler (2012). This approach allows simplifications of the computational burden and takes into account double counting among related animals thanks to the estimation of the contributions due to relationships. These contributions were estimated by a two-step algorithm taken into account all relationships between the foreign stallions and their ancestors.

The equations system of the Belgian model (IV.1) integrating external information (evaluation B) can be written as:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{*-1}} & \mathbf{Z'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} & \mathbf{Z'R^{-1}Z+P^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y+G^{*-1}\mu_0} \\ \mathbf{Z'R^{-1}y} \end{bmatrix} \quad \text{(IV.3)}$$

where $\mathbf{R}$ was the residual (co)variance matrix, $\mathbf{P}$ was the (co)variance matrix for the permanent environment, $\boldsymbol{\mu}_0$ was the vector of EBVc and $\mathbf{G}^*$ was the matrix of prediction error (co)variances of EBV_c.

Because the second version of modified Bayesian evaluation (Vandenplas and Gengler, 2012) was applied, the inverse of $\mathbf{G}^*$ was equal to $\mathbf{G}^{*-1} = \mathbf{G}^{-1} + \boldsymbol{\Lambda}$ where the matrix $\mathbf{G}^{-1}$ is the inverse of the additive genetic (co)variances matrix and the matrix $\boldsymbol{\Lambda}$ was a block diagonal variance matrix with one block diagonals per horse. For the *N* stallions associatied to external information, the different block diagonals were equal to $\boldsymbol{\Delta}_\mathbf{i}\mathbf{G}_\mathbf{0}^{-1}\boldsymbol{\Delta}_\mathbf{i}$ for $i = 1, \ldots, N$. The matrix $\mathbf{G}_\mathbf{0}$ was the matrix of genetic (co)variances among traits and $\boldsymbol{\Delta}_\mathbf{i}$ was a diagonal matrix with elements equal to $\sqrt{RE_{ij} * \frac{\sigma^2_{u_j}}{\sigma^2_{e_j}}}$ for $j = 1, 2$ traits where $\sigma^2_{e_j}$ was the error variance of this *j*th trait, $\sigma^2_{u_j}$ was the genetic variance for the *j*th trait, and RE_{ij} was equal to the value of records equivalents (RE) only due to own records for the *j*th trait. RE were estimated thanks to the algorithm taking into account double counting among related animals (Vandenplas and Gengler, 2012).

To approximate REL of breeding values estimated by the evaluation B based on data I for the Belgian trait (EBV_{BI}; REL_{BI}), the following procedure was applied. First, for

each stallion that had external information, one virtual performance was added to the performance data and weighted by the value of RE independent from contributions due to relationships. The weight for the real performances in the performance data was equal to 1. An additional level for each fixed effect of the Belgian genetic evaluation was created and assigned to the virtual performances to ensure that they had no influence on the genetic evaluation. PEV were estimated by the inversion of the left hand side of the mixed model equations, and REL_{BI} were calculated using the equation (IV.2). All the genetic evaluations and computations of PEV were performed using the BLUPF90 program family (Misztal, 2012) modified to integrate external information by taking into account double counting among related animals.

Descriptive statistics were computed to characterize integrated external information and its influence on the ranking of the horses. With regard to REL_E, REL_c and RE, mean and standard deviations (SD) as well as the number of foreign stallions associated with a non-zero REL_E, a non-zero REL_c and non-zero RE were described. Pearson correlation coefficients between EBV_{AI} and EBV_E and the coefficients of determination of the conversion equations for the Dutch and French stallions were also computed, as well as Spearman rank correlation coefficients among EBV_{AI}, EBV_{BI} and EBV_E, for all the horses, for the French stallions, for the Dutch stallions and for the 100 best stallions. This latter group included the 100 best-ranked stallions following the evaluation AI, born after 1979 and associated with a REL_{AI} equal or higher than 0.75. The model adequacy was tested by the comparison of accuracy and precision of the evaluations AI and BI through comparisons of mean bias (MB), mean squared error of prediction (MSEP) and Pearson correlation coefficients between observed and estimated performances associated with the Belgian trait ($r_{y:\hat{y}}$; Tedeschi, 2006). Considering all the horses, MB and MSEP were expressed as a percentage of the average performance of all the performances. Considering the 100 best stallions, the Dutch and French stallions, MB and MSEP were expressed as a percentage of the average performance of their performances, respectively.

To test the predictive ability of the applied method, subsets II and III were created and evaluations A and B were performed based on these two subsets (i.e., evaluations AII, AIII, BII and BIII). Resulting EBV for the Belgian trait are called EBV_{AI}, EBV_{AII}, EBV_{AIII}, EBV_{BI}, EBV_{BII} and EBV_{BIII} (Table IV-1). Subset II consisted of all performances before 31 December 2005 included. All other performances were assumed to be missing

(i.e., 34.5% of all the performances). It simulated the predictive ability of the method routinely applied. Subset III consisted of the complete data except for all French and Dutch stallions' own performances assumed to be missing (i.e., 0.27% of all the performances). It simulated the predictive ability of the method applied for stallions with no own Belgian performances. REL of EBV_{AII}, EBV_{AIII}, EBV_{BII} and EBV_{BIII} (REL_{AII}, REL_{AIII}, REL_{BII} and REL_{BIII}, respectively) were approximated as described previously for evaluations A and B. Pearson correlation coefficients (r_{II}) between EBV_{AI} and EBV_{AII}, and EBV_{BI} and EBV_{BII}, as well as Pearson correlation coefficients (r_{III}) between EBV_{AI} and EBV_{AIII}, and EBV_{BI} and EBV_{BIII}, were estimated. Furthermore, variances of the differences (VAR_{II}) between EBV_{AI} and EBV_{AII}, and EBV_{BI} and EBV_{BII}, as well as variances of the differences (VAR_{III}) between EBV_{AI} and EBV_{AIII}, and EBV_{BI} and EBV_{BIII}, were also estimated. r_{II}, r_{III}, VAR_{II} and VAR_{III} were estimated for all the horses included in the genetic evaluations, for the 100 best stallions and for the French and Dutch stallions.

Means and SD of REL_{AI}, REL_{AII}, REL_{AIII}, REL_{BI}, REL_{BII} and REL_{BIII} were computed for all horses, for the French stallions and for the Dutch stallions.

RESULTS AND DISCUSSION

DESCRIPTIVE STATISTICS

Among the 98 French and 67 Dutch stallions associated with external information, only 97 French and 54 Dutch stallions had a non-zero REL_c. Furthermore, REL_c decreased at least by 57% compared with REL_E (Table IV-2). This decrease was expected because, first, the coefficients of determination of the conversion equations and, second, the Pearson correlation coefficients between EBV_{AI} and EBV_E in the conversion equations were low to moderate (Table IV-2). Powell et al. (1994) surveyed different countries to determine expressions of REL_c associated with EBV_c in the context of dairy cattle. Several countries considered that REL_c were equal to REL_E. However, because REL_c must be integrated into a genetic evaluation, the variances of the sample regression equations must be taken into account as it was detailed by Goddard (1985). Therefore, lower variances of the sample regression equations and higher coefficients of determination, that is more accurate conversion equations, would be desirable to estimate reliable EBV_c (i.e., to obtain higher REL_c). Restrictions on data used for the comparisons of genetic evaluations between countries were formulated to improve the accuracy of

conversions methods (e.g., Powell and Sieber, 1992; Powell et al., 1994). However, owing to the low number of foreign stallions having both an EBV_{AI} and an EBV_E (i.e., 93 French stallions and 47 Dutch stallions) and the low average reliabilities associated with EBV_{AI} and EBV_E, most of the recommended restrictions could not be respected. Furthermore, the use of the Pearson correlation coefficients between EBV_{AI} and EBV_E for the French stallions and for the Dutch stallions led to an underestimation of REL_c because Calo et al. (1973) showed that genetic correlation coefficients for traits were higher than the corresponding Pearson correlation coefficients. This can be confirmed for the pair Belgium/France for which the genetic correlation coefficient was previously estimated between 0.76 and 0.88 (Ruhlmann et al., 2009b), while the Pearson correlation coefficient was equal to only 0.71 (Table IV-2). Therefore, it would be interesting to have estimates of the genetic correlation coefficients for the missing pair Belgium/the Netherlands because imprecise estimates of genetic correlation coefficients could lead to inexact REL_c (Calo et al., 1973; Powell et al., 1994).

Table IV-2. Coefficients of determination (R^2) of the conversion equations and Pearson correlation coefficients (r) between internal and external estimated breeding values. Means and standard deviations (SD) of nonzero external reliabilities (REL_E), non-zero reliabilities of a converted estimated breeding value (REL_c) and non-zero record equivalents free of contributions due to relationships (RE) as well as the number of foreign stallions (Nb) associated with non-zero REL_E, REL_c and RE

	French stallions	Dutch stallions
R^2	0.46	0.28
r	0.71	0.59
REL_E		
Nb	98	67
Mean	0.59	0.42
SD	0.19	0.18
REL_c		
Nb	97	54
Mean	0.27	0.07
SD	0.10	0.06
RE		
Nb	97	50
Mean	2.75	0.51
SD	1.44	0.52

As shown before, REL_E and the accuracy of the conversion equations have an effect on REL_c, but also on the improvement of the genetic evaluation. Indeed, a simplified system of mixed model equations ($\mathbf{Cs} = \mathbf{r}$) integrating external information (i.e., prior information) can be written as:

$$\left(\mathbf{C}+\mathbf{V}^{-1}\right)\mathbf{s}=\mathbf{r}+\mathbf{V}^{-1}\boldsymbol{\mu} \qquad \text{(IV.4)}$$

where $\mathbf{C}$, $\mathbf{r}$ and $\mathbf{s}$ are the left hand side, the right hand side and the vector of solutions of the mixed model equations, respectively, $\boldsymbol{\mu}$ is the vector of mean prior information and $\mathbf{V}$ is the prior variance matrix.

On the one hand, in the case of highly accurate prior information (i.e., in this case, high REL_c needed high REL_E and accurate conversion equations), the value of $\mathbf{V}$ will be close to zero, $\mathbf{V}^{-1}$ will be large and, therefore, the equations system (IV.4) will tend to $\mathbf{V}^{-1}\mathbf{s}=\mathbf{V}^{-1}\boldsymbol{\mu}$ and the estimate of $\mathbf{s}$ to $\hat{\mathbf{s}}\approx\boldsymbol{\mu}$. On the other hand, in the case of low accurate prior information, (i.e., in this case, low REL_c estimated from low REL_E and/or poorly accurate conversion equations), the prior information is non informative, the value of $\mathbf{V}$ will be large, $\mathbf{V}^{-1}$ will tend to zero and, therefore, the equations system (IV.4) will tend to $\mathbf{Cs}=\mathbf{r}$ and the estimate of $\mathbf{s}$ to $\hat{\mathbf{s}}\approx\mathbf{C}^{-1}\mathbf{r}$, similarly to a system which do not integrate prior information. Between these two extreme cases, $\hat{\mathbf{s}}$ can be considered as a weighted average of the combination of data and prior information (Klei et al., 1996).

With regard to RE (Table IV-2), the number for each stallion was a function of REL_c and of the relationships with other stallions. The low number of RE can be explained by the low-to-moderate REL_c but also by the low-to-moderate coefficients of determination of the conversion equations and Pearson correlation coefficients between EBV_{AI} and EBV_E. Furthermore, 4 Dutch stallions that were highly related to other ones were associated with RE equal to 0 after estimation of contributions due to relationships, while this was not the case for the French stallions. Hence, considering contributions due to relationships seems necessary to avoid double counting of external information.

Concerning the rankings of the horses, the rank correlation between EBV_{AI} and EBV_{BI} for all horses (Table IV-3) showed that the integration of external information for foreign stallions into the Belgian genetic evaluation influenced very slightly the ranking of the whole population. These modifications can be explained by the animal model: all the relationships between the foreign stallions and other horses were taken into account. These relationships caused an effect of foreign information on related horses through the foreign stallions. However, these modifications were small because external information was integrated only for about 0.2% of the horses. Concerning the foreign stallions, the integration of external information led to a change of their rankings according to the rank correlations between EBV_E and EBV_{AI} or EBV_{BI} (Table IV-3). The Belgian ranking of

the foreign stallions was more similar to the ranking in their country of birth when the external information was integrated, as expected and shown by Quaas and Zhang (2006). However, because average quantity of French information was higher in terms of RE than the Dutch information (Table IV-2), the increase in the rank correlation coefficients between EBV_E and EBV_{AI} and between EBV_E and EBV_{BI} for the French stallions was higher than the corresponding increase for the Dutch ones, as according to the theory (Zhang et al., 2002; Legarra et al., 2007). Finally, because the Belgian ranking of foreign horses was more similar to the ranking in their country of birth and because 24 foreign stallions were considered in the group of the 100 best stallions, the rank correlation between EBV_{AI} and EBV_{BI} for the best stallions (Table IV-3) showed a change in their ranking of 2%. The reranking was mainly due to the reranking of the foreign stallions, but also due to the reranking of foreign stallions' relatives. The best stallions associated with external information gained four ranks on average in the ranking of the 100 best stallions, and a gain of 23 ranks was the largest reranking for a foreign best stallion. Integration of external information also led to a gain of 18 ranks for a stallion not associated with external information but related to several foreign stallions.

Table IV-3. Spearman rank correlation coefficients between breeding values (EBV) estimated by the Belgian genetic evaluation based on the complete data I (EBV_{AI}) and EBV estimated by a Bayesian[1] evaluation based on data I (EBV_{BI}), Spearman rank correlation coefficients between EBV_{AI} and external EBV (EBV_E), and Spearman rank correlation coefficients between EBV_{BI} and EBV_E for all horses, for the 100 best stallions, for the French stallions and for the Dutch stallions

Group of horses	Nb	Spearman rank correlations		
		EBV_{AI} /EBV_{BI}	EBV_{AI} /EBV_E	EBV_{BI} /EBV_E
All horses	101 382	>0.99	-	-
Best stallions	100	0.98	-	-
French stallions	98	0.87	0.69	0.90
Dutch stallions	67	0.95	0.61	0.73

[1]Bayesian: Belgian genetic evaluation integrating external information by a Bayesian approach.

MODEL ADEQUACY

Considering all horses, as well as the best stallions, the comparison of MB, MSEP and $r_{y:\hat{y}}$ did not show a reduction in the level of precision and accuracy of the Belgian model when external information was included because MB were close to 0% and MSEP were equal for the two genetic evaluations. Similar values were also estimated for $r_{y:\hat{y}}$

associated with the evaluations AI and BI (Table IV-4).

With regard to the French stallions, the bias of 1.32% for the evaluation AI was reduced when external information was integrated (Table IV-4). However, it was not confirmed by the associated MSEP, and thereby, the adequacy of the model for the French stallions was not improved, but also not diminished, by the integration of external information.

With regard to the Dutch stallions, it seems that the integration of Dutch information improved the adequacy of the model for their genetic evaluation because MB and MSEP of the evaluation BI were slightly lower than MB and MSEP of the evaluation AI (Table IV-4). Furthermore, $r_{y:\hat{y}}$ of the evaluation BI was slightly higher than the one of the evaluation AI.

Finally, the integration of external information did not diminish the adequacy of the Belgian model for all horses, for the best stallions and for foreign stallions. However, owing to the low amount of external information (only for about 0.2% of the horses) and to the low average REL_E, the model adequacy was not improved or only weakly for the Dutch stallions.

Table IV-4. Number of performances (Nb) associated with the Belgian trait for all horses, the 100 best stallions, the French stallions and the Dutch stallions. Mean bias (MB), mean squared errors of prediction (MSEP) and Pearson correlation coefficients between observed and estimated performances ($r_{y:\hat{y}}$) for the evaluation AI[1] and the evaluation BI[2] applying all the performances, performances associated with the 100 best stallions, French performances and Dutch performances for the Belgian trait. MB and MSEP are expressed as a percentage of the average performance of all the performances and performances associated with the 100 best stallions, the French and Dutch stallions, respectively

	All horses		Best stallions		French stallions		Dutch stallions	
	AI	BI	AI	BI	AI	BI	AI	BI
Nb	350 907		2749		1322		414	
MB	2.00e-3	-4.00e-8	0.06	0.06	1.32	1.06	1.49	1.36
MSEP	2.74	2.74	0.67	0.67	2.26	2.26	2.38	2.37
$r_{y:\hat{y}}$	0.50	0.50	0.26	0.26	0.36	0.36	0.49	0.50

[1]Belgian genetic evaluation based on the complete data I.
[2]Bayesian evaluation based on the complete data I.

PREDICTIVE ABILITY

Considering all horses, the evaluation BII had a similar or slightly worse predictive ability than the evaluation AII according to r_{II} and VAR_{II} (Table IV-5). This

low difference between predictive abilities was expected because external information was integrated into the evaluation for only 165 foreign stallions, whereas 44 755 horses have performances among the 101 382 horses in the pedigree. However, if the evaluations were based on subset III, i.e., if only performances of the foreign stallions were assumed to be missing, there was a slight advantage for the evaluation B, especially in terms of VAR_{II}. Furthermore, r_{III} close to 1 can be explained by the fact that only 0.27% of all the performances were assigned to missing values in subset III (Table IV-5), hence their limited overall influence.

Table IV-5. Pearson correlation coefficients (r_{II}) and variances of differences (VAR_{II}) between EBV_{AI}[1] and EBV_{AII}, and EBV_{BI} and EBV_{BII}, and Pearson correlation coefficients (r_{III}) and variances of differences (VAR_{III}) between EBV_{AI} and EBV_{AIII}, and EBV_{BI} and EBV_{BIII} for all horses, for the 100 best stallions, for the French stallions and the Dutch stallions

Group of horses	Genetic evaluation	Datasets			
		I-II		I-III	
		r_{II}	VAR_{II} (x 10^{-3})	r_{III}	VAR_{III} (x 10^{-3})
All horses	A	0.89	4.49	>0.99	0.10
	B	0.89	4.58	>0.99	0.06
Best stallions	A	0.80	9.34	0.98	1.18
	B	0.82	10.24	0.99	1.02
French stallions	A	0.96	3.37	0.89	8.05
	B	0.99	2.23	0.98	3.48
Dutch stallions	A	0.93	7.12	0.95	5.73
	B	0.95	6.58	0.97	4.31

[1]EBV_{ij}: Estimated breeding values where *i* refers to the type of the genetic evaluation (i.e. A = Belgian genetic evaluation and B = Bayesian evaluation) and *j* refers to the used data (i.e. I = complete data, II = data for which all performances after 31 December 2005 were assumed to be missing, and III = data for which all French and Dutch stallions' own performances were assumed to be missing).

Considering the French stallions, r_{II}, VAR_{II}, r_{III} and VAR_{III} showed that the predictive ability of breeding values was improved when French information was integrated (Table IV-5). The high r_{II} (0.96) between EBV_{AI} and EBV_{AII} also showed that performances after 2005 influenced less the genetic evaluation of the French stallions compared with the genetic evaluation for all horses. It can be explained by the facts that the French stallions had few performances after 2005 and few relationships with Belgian horses competing after 2005. Regarding the evaluations based on the subset III, there were an increase of r_{III} around 10% and a reduction of VAR_{III} of 57% for the evaluations B in comparison with r_{III} and VAR_{III} of the evaluations A, respectively. Following these

results, integrated external information was almost equivalent to the French stallions' own performances. These results suggest that the integration of external information could be very interesting in the case of imported stallions having no or few own Belgian performances (e.g., young imported stallions or confirmed foreign stallions imported through their semen).

Considering the Dutch stallions, the lower r_{II} and higher VAR_{II} compared with the r_{III} and VAR_{III}, respectively, for the evaluations A led to higher influence of performances recorded after 2005 for their genetic evaluation compared with their own records (Table IV-5). The low number of Dutch stallions with performances, that is <60% of the Dutch stallions, can explain this observation. This is also explained by the high r_{III} between EBV_{AI} and EBV_{AIII} (0.95). However, following r_{III} and VAR_{III} (Table IV-5), the predictive ability was improved when external information was integrated into the evaluation based on the subset III. Regarding the evaluations based on the subset II, the increase in r_{II} and the slight improvement in VAR_{II} also confirmed the improvement in the predictive ability when Dutch information was integrated into the Belgian genetic evaluation.

RELIABILITIES

Means and SD of REL were calculated for all the genetic evaluations (Table IV-6). All the genetic evaluations had a minimum and a maximum REL equal to 0.00 and 0.99, respectively. As expected since external information was only integrated for 165 stallions, the integration of external information did not influence on average the genetic evaluation for all the horses.

It is noted that the accuracy of the procedure applied to estimate REL_{BI}, REL_{BII} and REL_{BIII} depended on the accuracy of REL_c and therefore on accuracies of the conversion equations and of the genetic correlation coefficients for traits between Belgium and the exporting countries. Because the genetic correlation coefficient was unknown for the pair Belgium/the Netherlands, they were estimated from EBV_{AI} and EBV_E of the foreign stallions and errors linked with the estimation of REL_c were also introduced into the estimation of REL associated with EBV_B. Again, these errors show the need to estimate genetic correlation coefficients for traits between countries to perform an unbiased genetic evaluation.

Table IV-6. Means and standard deviations (SD) of reliabilities $(REL)_{AI}$[1], REL_{AII}, REL_{AIII}, REL_{BI}, REL_{BII} and REL_{BIII} for all horses, for the French stallions and for the Dutch stallions

Group of horses	Genetic evaluation	Datasets					
		I		II		III	
		Mean	SD	Mean	SD	Mean	SD
All horses	A	0.21	0.17	0.17	0.16	0.21	0.17
	B	0.21	0.17	0.17	0.16	0.21	0.17
French stallions	A	0.58	0.23	0.54	0.24	0.52	0.26
	B	0.61	0.20	0.58	0.20	0.57	0.21
Dutch stallions	A	0.51	0.26	0.46	0.26	0.47	0.27
	B	0.52	0.25	0.47	0.25	0.49	0.26

[1]REL_{ij}: Reliabilities of estimated breeding values where *i* refers to the type of genetic evaluation (i.e. A = Belgian genetic evaluation and B = Bayesian evaluation) and *j* refers to the used data (i.e. I = complete data, II = data for which all performances after 31 December 2005 were assumed to be missing, and III = data for which all French and Dutch stallions' own performances were assumed to be missing).

Regarding to the Dutch and French stallions, the average REL_{BI} were improved compared with the average REL_{AI}. Additionally to the influence of the imprecision because of accuracy of regression equations and of the unknown genetic correlations as explained previously, the improvement in average REL also depended on the range of REL_c and thereby on the range of REL_E. Zhang et al. (2002) also concluded that the amount of improvement depends on REL_E for a simulation for beef cattle.

For the particular case of the French stallions, the integration of external information led to an increase (5%) of the average REL for the genetic evaluations based on data I (Table IV-6). Then, the average REL_A decreased when performances were assumed to be missing, as it was expected. This reduction was higher for REL_{AIII}. The own performances were more informative for the French stallions than the contemporary horses' performances recorded after 2005, as already observed. Nevertheless, the integration of French information led to an average REL_{BII} and REL_{BIII} equal or close to the average REL_{AI}. So, external information was on average at least equivalent to the Belgian performances related to these stallions.

Regarding to the Dutch stallions, average REL_{AII} and REL_{AIII} confirmed the higher influence of performances after 2005 for their genetic evaluation compared with their own performances (Table IV-6). Furthermore, average REL_{BII} and REL_{BIII} compared with average REL_{AI} showed that the integration of Dutch information was not totally equivalent to the missing information, despite it was not insignificant in each case. Indeed, there was an increase of the average REL of at least 2% when Dutch information was integrated into the Belgian evaluation.

CONCLUSION

According to these results, external information, that is foreign EBV and their associated REL, for French and Dutch stallions was partially to totally equivalent to 4 years of contemporary horses' performances or to their own performances in show jumping. Its integration did not diminish the adequacy of the Belgian model for all horses, as well as for foreign stallions. It also improved the predictive ability and the accuracy of EBV for the foreign stallions. The resulting Belgian ranking of the foreign stallions was more similar to their foreign ranking according to their country of birth, according to the Spearman rank correlations. All these results showed the interest to integrate external information into the Belgian genetic evaluation for show jumping, especially for imported stallions having no or few Belgian performances (e.g., young imported stallions or confirmed foreign stallions imported through their semen). However, estimates of genetic correlations for traits among countries, as well as accurate conversion equations, are needed for a more accurate Belgian Bayesian evaluation.

ACKNOWLEDGEMENTS

Jérémie Vandenplas, as a Research Fellow, acknowledges the support of the National Fund for Scientific Research (Brussels, Belgium) for this position and for the additional grants 2.4507.02F(2) and F4552.05 (2.4.623.08.F). The authors acknowledge Anne Ricard (Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, France) and Danielle Arts (Royal Dutch Sport Horse, the Netherlands) to provide the external information to realize this research. The authors are also grateful to University of Liege (SEGI facility) for the use of NIC3 supercomputer facilities. The authors thank the anonymous reviewers for the useful comments.

REFERENCES

Calo, L.L., R.E. McDowell, L.D. VanVleck, and P.D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J. Anim. Sci.* 37:676–682.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Goddard, M. 1985. A method of comparing sires evaluated in different countries. *Livest. Prod. Sci.* 13:321–331.

Interstallion. 2005. Recommendations on choice of scale and reference population for publication of breeding values in sport horse breeding. Accessed Mar. 19, 2012. http://www.biw.kuleuven.be/genlog/livgen/docs/publicationscale.pdf.

Janssens, S., and N. Buys. 2008. Genetische index voor springen - GSIn2008. Accessed Mar. 19, 2012. http://www.biw.kuleuven.be/genlog/livgen/chgs_ebv/GenetischeSpringindex_2008.pdf.

Janssens, S., N. Buys, and W. Vandepitte. 2007. Sport status and the genetic evaluation for show jumping in Belgian sport horses. Page 151 in Book of Abstracts of the 58th Annual Meeting of the European Association for Animal Production, Dublin, Ireland. Wageningen Academic Publishers, Wageningen, The Netherlands.

Janssens, S., D. Geysen, and W. Vandepitte. 1997. Genetic parameters for show jumping in Belgian sporthorses. 5pp. in Book of Abstracts of the 48th Annual Meeting of the European Association for animal production, Vienna, Austria. Wageningen Academic Publishers, Wageningen, The Netherlands.

Klei, L., R.L. Quaas, E.J. Pollak, and B.E. Cunningham. 1996. Multiple-breed evaluation. Pages 106–113 in Proceedings of the 28th Research Symposium and Annual Meeting. Beef Improvement Federation.

Koenen, E.P.C. 2002. Genetic evaluations for competition traits of warmblood sport horses. Note presented at World Breeding Federation for Sport Horses seminar, Budapest, Hungary. Accessed Mar. 21, 2012. http://www.biw.kuleuven.be/genlog/livgen/docs/genetic-evaluation.pdf.

Koenen, E.P.C., and L.I. Aldridge. 2002. Testing and genetic evaluation of sport horses in an international perspective. Page 23 in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Koenen, E.P.C., L.I. Aldridge, and J. Philipsson. 2004. An overview of breeding objectives for warmblood sport horses. *Livest. Prod. Sci.* 88:77–84.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

Misztal, I. 2012. BLUPF90 family of programs. Accessed Mar. 21, 2012. http://nce.ads.uga.edu/~ignacy/newprograms.html.

Powell, R.L., and M. Sieber. 1992. Direct and indirect conversion of bull evaluations for yield traits between countries. *J. Dairy Sci.* 75:1138–1146.

Powell, R.L., G.R. Wiggans, and P.M. VanRaden. 1994. Factors affecting calculation and

use of conversion equations for genetic merit of dairy bulls. *J. Dairy Sci.* 77:2679–2686.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Ricard, A. 1997. Breeding evaluations and breeding programs in France. 9 pp. in Book of Abstracts of the 48th Annual Meeting of the European Association for animal production, Vienna, Austria. Wageningen Academic Publishers, Wageningen, The Netherlands.

Ruhlmann, C., E. Bruns, E. Fraehr, J. Philipsson, S. Janssens, K. Quinn, E.T. Hellsten, and A. Ricard. 2009a. Genetic connectedness between seven European countries for performance in jumping competitions of warmblood riding horses. *Livest. Sci.* 120:75–86.

Ruhlmann, C., S. Janssens, J. Philipsson, E. Thorén-Hellsten, H. Crolly, K. Quinn, E. Manfredi, and A. Ricard. 2009b. Genetic correlations between horse show jumping competition traits in five European countries. *Livest. Sci.* 122:234–240.

Tavernier, A. 1991. Genetic evaluation of horses based on ranks in competitions. *Genet Sel. Evol* 23:159–173.

Tedeschi, L.O. 2006. Assessment of the adequacy of mathematical models. *Agric. Syst.* 89:225–247.

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

van Veldhuizen, A.E. 1997. Breeding value estimation for riding horses in Netherlands. 8pp. in Book of Abstracts of the 48th Annual Meeting of the European Association for Animal Production, Vienna, Austria. Wageningen Academic Publishers, Wageningen, The Netherlands.

Zhang, Z.W., R.L. Quaas, and E.J. Pollak. 2002. Simulation study on the effects of incorporating external genetic evaluation results. CD–ROM Comm. 20–14 in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Chapter V. SHORT COMMUNICATION: ALTERATION OF PRIORS FOR RANDOM EFFECTS IN GAUSSIAN LINEAR MIXED MODELS

Bayesian approaches integrating external information (i.e., estimated breeding values and associated reliabilities provided by an external genetic evaluation) into an internal genetic evaluation were proposed in the previous Chapters and were based on the alteration of both the mean and the (co)variance of the prior multivariate normal distributions of random effects of linear mixed models. However, most software packages available in animal breeding community do not permit such alterations, and, thereby, they do not permit the implementation of the proposed Bayesian approaches that integrate external information. Furthermore, source codes of most software are usually unavailable, making the implementation of the Bayesian approaches proposed in the previous Chapters impossible. Therefore, the aim of this Chapter was to propose a method to alter both the mean and the (co)variance of the prior distributions of random effects of linear mixed models in the framework of currently available software packages. Based on two datasets, the method was tested with three software packages.

From: Vandenplas, J., O.F. Christensen, and N. Gengler. 2014. Short communication: Alteration of priors for random effects in Gaussian linear mixed models. *J. Dairy Sci.* 97:5880-5884.

ABSTRACT

Linear mixed models, for which the prior multivariate normal distributions of random effects are assumed to have a mean equal to 0, are commonly used in animal breeding. However, some statistical analyses (e.g., the consideration of a population under selection in a genomic breeding scheme, multiple-trait predictions of lactation yields, and Bayesian approaches integrating external information into genetic evaluations) need to alter both the mean and (co)variance of the prior distributions and, to our knowledge, most software packages available in the animal breeding community do not permit such alterations. Therefore, the aim of this study was to propose a method to alter both the mean and (co)variance of the prior multivariate normal distributions of random effects of linear mixed models while using currently available software packages. The proposed method was tested on simulated examples with 3 different software packages available in animal breeding. The examples showed the possibility of the proposed method to alter both the mean and (co)variance of the prior distributions with currently available software packages through the use of an extended data file and a user-supplied (co)variance matrix.

Key words: Prior distribution, Bayesian, linear mixed model

SHORT COMMUNICATION

Currently, Henderson's mixed models methods and Best Linear Unbiased Prediction (BLUP; Henderson, 1975) are commonly used in animal breeding. The typical linear mixed model is written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad \text{(V.1)}$$

where $\mathbf{y}$ is the vector of records, $\boldsymbol{\beta}$ and $\mathbf{u}$ are vectors of fixed and random effects related to the records through the incidence matrices $\mathbf{X}$ and $\mathbf{Z}$, respectively, and $\mathbf{e}$ is the vector of residuals.

It is assumed that the expectations are $E\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and the (co)variance matrices are $Var\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$ where $\mathbf{G}$ is the (co)variance matrix associated with $\mathbf{u}$ and $\mathbf{R}$ is the (co)variance matrix associated with $\mathbf{e}$. The estimates of $\boldsymbol{\beta}$; that is, $\hat{\boldsymbol{\beta}}$, and the predictions of $\mathbf{u}$; that is, $\hat{\mathbf{u}}$, can be obtained solving the mixed-model equations written

as follows (Henderson, 1950):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \qquad \text{(V.2)}$$

In the case of BLUP, the (co)variance matrices $\mathbf{G}$ and $\mathbf{R}$ are assumed to be known.

Although $\mathbf{u}$ is assumed to have an expectation equal to **0**, a need exists to alter this expectation in some statistical analyses. For example, Bayesian approaches integrating external information (i.e., estimated breeding values (EBV) and associated reliabilities obtained from a foreign genetic evaluation) into a local genetic evaluation alter both $E[\mathbf{u}]$ and $Var[\mathbf{u}]$ (e.g., Gianola and Fernando, 1986; Quaas and Zhang, 2006; Legarra et al., 2007). For such approaches, $E[\mathbf{u}]$ is equal to the foreign EBV and $Var[\mathbf{u}]$ represents the associated matrix of prediction error (co)variances. Another example is the consideration of a population under selection in a genomic breeding scheme by assuming $E[\mathbf{u}] \neq \mathbf{0}$ for the genotyped animals. Indeed, they may have an expectation different from 0 if selection occurred (Vitezica et al., 2011). Also, Schaeffer and Jamrozik (1996) proposed a multiple-trait procedure for predicting lactation yields for dairy cows based on an alteration of $E[\mathbf{u}]$ with information from groups of cows sharing the same production characteristics. However, to our knowledge, most software packages currently available in animal breeding do not permit alterations of expectations of random effects, whereas they may allow the use of a user supplied covariance matrix. Therefore, the aim of this study was to propose a method to alter both the expectations and (co)variances of random effects while using software packages currently available in animal breeding. The development of the proposed method was based on a Bayesian view of linear mixed models.

BAYESIAN VIEW OF LINEAR MIXED MODELS

Bayes estimators for linear (mixed) models and their relations with BLUP were discussed by several authors (e.g., Lindley and Smith, 1972; Dempfle, 1977; Gianola and Fernando, 1986; Sorensen and Gianola, 2002). From a Bayesian view, all fixed and random effects are considered as random. However, the terms “fixed” and “random” will still be used below to differentiate $\boldsymbol{\beta}$ from $\mathbf{u}$, respectively.

For the linear model (V.1), the following prior multivariate normal (MVN) distributions are assumed:

$$[\boldsymbol{\beta}|\mathbf{B}] \sim MVN(\mathbf{b}, \mathbf{B}),$$

where $\mathbf{b}$ is a mean vector and $\mathbf{B}$ is a (co)variance matrix,

$$[\mathbf{u}|\mathbf{G}] \sim MVN(\mathbf{g}, \mathbf{G}),$$

where $\mathbf{g}$ is a mean vector, and

$$[\mathbf{e}|\mathbf{R}] \sim MVN(\mathbf{0}, \mathbf{R}).$$

Assuming that all the (co)variance matrices are known, the joint posterior density of $\boldsymbol{\beta}$ and $\mathbf{u}$ can be written as follows:

$$f(\boldsymbol{\beta}, \mathbf{u}|\mathbf{y}, \mathbf{B}, \mathbf{G}, \mathbf{R}) \propto \exp\left\{-\frac{1}{2}\begin{pmatrix}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) + (\boldsymbol{\beta}-\mathbf{b})'\mathbf{B}^{-1}(\boldsymbol{\beta}-\mathbf{b}) \\ + (\mathbf{u}-\mathbf{g})'\mathbf{G}^{-1}(\mathbf{u}-\mathbf{g})\end{pmatrix}\right\}.$$

Because this joint posterior distribution is multivariate normal, its mean equals its mode, and $\boldsymbol{\beta}$ and $\mathbf{u}$ can be estimated by differentiating the joint posterior distribution with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$ and setting the derivatives equal to zero. From this, as shown by Gianola and Fernando (1986), the following equation is obtained:

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}+\mathbf{B}^{-1} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}+\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}+\mathbf{G}^{-1}\mathbf{g}\end{bmatrix}. \qquad \text{(V.3)}$$

If a noninformative prior is considered for $\boldsymbol{\beta}$ (i.e., $\mathbf{B}^{-1} \to \mathbf{0}$) and $\mathbf{g} = \mathbf{0}$, the system of equations (V.3) simplifies to traditional mixed-model equations (V.2).

ALTERATION OF PRIORS FOR RANDOM EFFECTS

In the following development, it is assumed that $\mathbf{g}$ and $\mathbf{G}^{-1}$ are known, $\mathbf{g} \neq \mathbf{0}$ and a non-informative prior for $\boldsymbol{\beta}$ is considered. Therefore, the system of equations (V.3) is written as follows:

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}+\mathbf{G}^{-1}\mathbf{g}\end{bmatrix}. \qquad \text{(V.4)}$$

Although most available software packages allow the use of a user-supplied (co)variance matrix as $\mathbf{G}$, or its inverse $\mathbf{G}^{-1}$, most of them do not allow that $\mathbf{g} \neq \mathbf{0}$. Thereby, the system of equations (V.4), allowing an alteration of the default mean, cannot be solved with current software packages. A way to solve this issue is to develop a system of equations equivalent to the system of equations (V.4) that can be solved by current software packages. Therefore, below we define $\mathbf{y_P}$, a vector of pseudo-records (i.e., records corrected for all other effects than $\mathbf{u}$); $\mathbf{X_P}$ and $\mathbf{Z_P}$, 2 incidence matrices relating pseudo-records to $\boldsymbol{\beta}$ and $\mathbf{u}$, respectively; $\mathbf{R_P}$, a residual (co)variance matrix associated to

the pseudo-records; and $\mathbf{G}^*$, a (co)variance matrix associated with $\mathbf{u}$ conditional on pseudo-records. Assuming that $\mathbf{X_p} = \mathbf{0}$, $\mathbf{Z_P'R_P^{-1}y_P} = \mathbf{G^{-1}g}$ and $\mathbf{G^{-1}} = \mathbf{G^{*-1}} + \mathbf{Z_P'R_P^{-1}Z_P}$, the system of equations (V.4) is equivalent to

$$\begin{bmatrix} \mathbf{X'R^{-1}X} + \mathbf{X_P'R_P^{-1}X_P} & \mathbf{X'R^{-1}Z} + \mathbf{X_P'R_P^{-1}Z_P} \\ \mathbf{Z'R^{-1}X} + \mathbf{Z_P'R_P^{-1}X_P} & \mathbf{Z'R^{-1}Z} + \mathbf{G^{*-1}} + \mathbf{Z_P'R_P^{-1}Z_P} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} + \mathbf{X_P'R_P^{-1}y_P} \\ \mathbf{Z'R^{-1}y} + \mathbf{Z_P'R_P^{-1}y_P} \end{bmatrix}. \quad \text{(V.5)}$$

The system of equations (V.5) can be written using compact notation as

$$\begin{bmatrix} \mathbf{X^{*'}R^{*-1}X^*} & \mathbf{X^{*'}R^{*-1}Z^*} \\ \mathbf{Z^{*'}R^{*-1}X^*} & \mathbf{Z^{*'}R^{*-1}Z^*} + \mathbf{G^{*-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X^{*'}R^{*-1}y^*} \\ \mathbf{Z^{*'}R^{*-1}y^*} \end{bmatrix}$$

where $\mathbf{X^*} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X_P} \end{bmatrix}$, $\mathbf{Z^*} = \begin{bmatrix} \mathbf{Z} \\ \mathbf{Z_P} \end{bmatrix}$, $\mathbf{y^*} = \begin{bmatrix} \mathbf{y} \\ \mathbf{y_P} \end{bmatrix}$ and $\mathbf{R^{*-1}} = \begin{bmatrix} \mathbf{R^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R_P^{-1}} \end{bmatrix}$.

Because the equivalent system of equations (V.5) has the same form as the system of equations (V.2), it can be solved using current software packages allowing the use of a user-supplied (co)variance matrix as $\mathbf{G}^*$, or its inverse $\mathbf{G}^{*-1}$, and a data file containing actual records ($\mathbf{y}$) and extended with pseudo-records ($\mathbf{y_P}$). Therefore, $\mathbf{y_P}$ and $\mathbf{G}^{*-1}$ must be computed before solving the system of equations (V.5) as follows (after some computational simplifications):

(1) Set up the matrix $\mathbf{G}^{-1}$ and the vector $\mathbf{g}$;

(2) Compute $\boldsymbol{\theta} = \mathbf{G}^{-1}\mathbf{g}$;

(3) Set up the matrix $\mathbf{Z_P}$ which is a diagonal matrix with (a) $\mathbf{Z}_{\mathbf{P}_{ii}} = 1$ if $\boldsymbol{\theta}_i \neq 0$ or (b) $\mathbf{Z}_{\mathbf{P}_{ii}} = 0$ if $\boldsymbol{\theta}_i = 0$, where i = 1, …, n and n is the number of levels associated with $\mathbf{u}$;

(4) Compute the (co)variance matrix $\mathbf{R_P}$ as $\mathbf{R_P} = \mathbf{Z_P}(\mathbf{I} \otimes \mathbf{R_0})\mathbf{Z_P}$ where $\mathbf{I}$ is an identity matrix of size k equal to the number of records for each trait (for simplicity assumed to be the same across traits) and $\mathbf{R_0}$ is the residual (co)variance matrix between traits for 1 record;

(5) Compute the vector of pseudo-records $\mathbf{y_P}$ as $\mathbf{y_P} = \mathbf{R_P}\boldsymbol{\theta}$;

(6) Compute the (co)variance matrix $\mathbf{G}^{*-1}$ as $\mathbf{G}^{*-1} = \mathbf{G}^{-1} - \mathbf{R_P^{-1}}$ where $\mathbf{R_P^{-1}}$ is the generalized inverse of $\mathbf{R_P}$.

Each pseudo-record in $\mathbf{y_P}$ must be added to the data file. Because each pseudo-record can be considered as 1 record corrected for all effects other than $\mathbf{u}$, all fixed

effects are set to 0, leading to $\mathbf{X_p} = \mathbf{0}$, except for the effect $\mathbf{u}$, for which it is equal to the level associated with the pseudo-record. With this approach, the system of equations (V.5), equivalent to the system of equations (V.4), can be solved using current software packages.

For a univariate analysis, step (4) is not needed and the computations of steps (5) and (6) can be simplified to $\mathbf{y_P} = \boldsymbol{\theta}\sigma_e^2$ and $\mathbf{G}^{*-1} = \mathbf{G}^{-1} - \mathbf{Z_P}\left(\sigma_e^2\right)^{-1}$, respectively, where σ_e^2 is the residual variance. For a multivariate analysis, software packages may only allow the use of a user-supplied matrix $\mathbf{A}$ (e.g., a relationships matrix), or its inverse $\mathbf{A}^{-1}$, instead of $\mathbf{G}$ or $\mathbf{G}^{-1}$, such that $\mathbf{G}^{-1}$ is equal to $\mathbf{G}^{-1} = \mathbf{A}^{-1} \otimes \mathbf{G_0^{-1}}$, where $\mathbf{G_0^{-1}}$ is the inverse of a known (co)variance matrix between traits. Because no matrix $\mathbf{A}^*$, or $\mathbf{A}^{*-1}$, can be found such that the computed $\mathbf{G}^{*-1}$ is equal to $\mathbf{G}^{*-1} = \mathbf{A}^{*-1} \otimes \mathbf{G_0^{-1}}$, a canonical transformation of the multivariate model must be performed (e.g., Quaas, 1984). Therefore, a vector of transformed observations $\mathbf{y_T}$ is defined such that $\mathbf{y}_{\mathbf{T}i} = \mathbf{T}\mathbf{y}_i$ for the *i*th animal (*i* = 1, 2, ..., *k*), with a matrix $\mathbf{T}$ satisfying $\mathbf{T}\mathbf{R_0}\mathbf{T}' = \mathbf{I}$ and $\mathbf{T}\mathbf{G_0}\mathbf{T}' = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix, and in addition, a vector of transformed prior means $\mathbf{g_T}$ must be computed as $\mathbf{g}_{\mathbf{T}i} = \mathbf{T}\mathbf{g}_i$ for *i* = 1, 2, ..., *k*. For the *j*th transformed variable, the system of equations (V.4) can be then written as

$$\begin{bmatrix} \mathbf{X}_j'\mathbf{X}_j & \mathbf{X}_j'\mathbf{Z}_j \\ \mathbf{Z}_j'\mathbf{X}_j & \mathbf{Z}_j'\mathbf{Z}_j + \mathbf{G}_{\mathbf{T}j}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{T}j} \\ \hat{\mathbf{u}}_{\mathbf{T}j} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_j'\mathbf{y}_{\mathbf{T}j} \\ \mathbf{Z}_j'\mathbf{y}_{\mathbf{T}j} + \mathbf{G}_{\mathbf{T}j}^{-1}\mathbf{g}_{\mathbf{T}j} \end{bmatrix}, \qquad \text{(V.6)}$$

where $\mathbf{G}_{\mathbf{T}j}^{-1} = \mathbf{A}^{-1}\mathbf{D}_{jj}^{-1}$ is the inverse of the (co)variance matrix associated with $\mathbf{u}_{\mathbf{T}j}$ and $\mathbf{D}_{jj}^{-1}$ is the inverse of the *j*th diagonal element of $\mathbf{D}$.

Because each analysis for the *j*th transformed variable is a univariate analysis, the limitation of the use of a user supplied matrix $\mathbf{A}$ or $\mathbf{A}^{-1}$ for a multivariate analysis is avoided. Thereby, the system of equations (V.6) can be solved for each *j*th transformed variable with the system of equations (V.5). The vector $\mathbf{y_P}$ and the matrix $\mathbf{G}^{*-1}$ specific to the *j*th transformed variable (i.e., $\mathbf{y}_{\mathbf{P}j}$ and $\mathbf{G}_j^{*-1}$) can be computed as described previously, and a matrix $\mathbf{A}_{\mathbf{j}}^{*-1}$ can be found such that $\mathbf{G}_j^{*-1} = \mathbf{A}_j^{*-1}\mathbf{D}_{jj}^{-1}$. Solutions for the system of equations (V.4) expressed on the original scale are equal to $\hat{\boldsymbol{\beta}}_i = \mathbf{T}^{-1}\hat{\boldsymbol{\beta}}_{\mathbf{T}i}$ and $\hat{\mathbf{u}}_i = \mathbf{T}^{-1}\hat{\mathbf{u}}_{\mathbf{T}i}$ for *i* = 1, 2, ..., *k*.

If a software package does not accept levels of effects defined as 0 for a (pseudo-) record (e.g., DMU; Madsen and Jensen, 2012), solutions of (V.5) can be obtained by adding a dummy level for those effects to replace the level defined as 0 and then by writing a regression model where effects including a dummy level are each considered as a regression without intercept nested within the effects. The associated covariables are equal to 0 for the dummy levels; otherwise, they are equal to 1.

EXAMPLE 1: NUMERICAL EXAMPLE

Consider a sample of 6 animals as designed in Table V-1. Milk yields (kg) for animals 2 to 6 are reported (Table V-1). The assumed model was a univariate model with a fixed herd effect and a random additive genetic effect. The residual and additive genetic variances were assumed to be 750 000 and 250 000 kg^2, respectively. The prior mean vector ($\mathbf{g}$) and the vector of estimated pseudo-records ($\mathbf{y_P}$) are reported in Table V-1. The variance $\mathbf{G^{-1}}$ was equal to $\mathbf{G^{-1}} = \mathbf{A^{-1}}\left(\sigma_u^2\right)^{-1}$ where $\mathbf{A^{-1}}$ is the inverse of the relationship matrix and σ_u^2 is the additive genetic variance. The estimates of the fixed herd effect and the predictions of the random effect were obtained solving the system of equations (V.4) and (V.5). Solved with the free software package GNU Octave (Eaton et al., 2011), both systems of equations provided the same solutions. These results were expected because the 2 systems of equations are equivalent.

Table V-1. Design and solutions for example 1

Animal	Sire	Dam	Herd	Milk yields	Prior mean	Pseudo-records	Estimated breeding values
1	-	-	-	-	0	-700	-208.150
2	-	-	1	8000	200	500	434.95
3	1	-	1	5500	200	1000	-80.29
4	1	2	2	6000	200	600	122.73
5	-	2	2	6500	0	-400	109.72
6	-	3	2	7000	0	-400	3.62

EXAMPLE 2: INTEGRATION OF EXTERNAL INFORMATION BASED ON A BAYESIAN APPROACH

The example 2 tested the proposed method to integrate external information following the second version of modified Bayesian evaluation (SBE; Vandenplas and Gengler, 2012). The software packages used for this example were ASReml with the !BLUP option (release 3.0; Gilmour et al., 2009), BLUPF90 (version 1.45; Misztal,

2013), and DMU4 from the DMU package (version 6, release 5.1; Madsen and Jensen, 2012).

A local sample of animals and a foreign one, both including about 2400 animals distributed over 6 generations, were simulated each from 50 male founders and 200 female founders. For both samples, all females were randomly mated with males randomly selected from available males to simulate the next generation. The maximum number of males mated in each generation was 40 and a male could be mated during a maximum of 2 generations. Furthermore, these matings could not be realized if the coefficient of relationship between 2 animals was higher or equal to 0.5, and if the female had already 3 progeny. Concerning the foreign sample, foreign females were only mated with foreign males. In each generation, 60% of foreign male offspring were randomly culled. Concerning the local sample of animals, local females were mated with local males and a subset of foreign males, including the first 150 sires that had most offspring in the foreign population. In each generation, 99% of local male offspring were randomly culled. Records for milk yield (kg) for the first lactation were simulated for each female in both samples (Van Vleck, 1994). A herd effect nested within sample was randomly assigned to each female under the condition that each herd included about 40 females. Residual and additive genetic variances were assumed to be $6 * 10^4$ and $2 * 10^4$ kg^2, respectively.

Simulation of data and a foreign genetic evaluation (i.e., a BLUP evaluation (V.2) based on foreign pedigree and data) were performed with GNU Octave software. Integration of external information into the local genetic evaluation (i.e., an evaluation based on local pedigree and data) was performed following the SBE. In the context of SBE, external information included foreign EBV for the 150 sires and EBV predicted from the foreign ones for local animals (summarized in $\mathbf{g}$) and associated record equivalents only due to own records (RE; i.e., effective numbers of records free of contributions due to relationships estimated by a 2-step algorithm; Vandenplas and Gengler, 2012). Because RE are only due to own records, RE for local animals were equal to zero. The matrix $\mathbf{G}^{-1}$ associated with $\mathbf{g}$ is the matrix of prediction error (co)variances and was approximated as $\mathbf{G}^{-1} = \mathbf{A}^{-1}\left(\sigma_u^2\right)^{-1} + \mathbf{\Lambda}$ where $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements equal to $\mathbf{\Lambda}_{ii} = \mathbf{RE}_i\left(\sigma_e^2\right)^{-1}$ for the *i*th animal (Vandenplas and Gengler, 2012). The SBE was also performed with GNU Octave software and obtained solutions were considered as reference.

To test the proposed approach with ASReml, BLUPF90, and DMU4, local pedigree and data files as well as foreign EBV for the 150 sires and predicted EBV for local animals and associated RE were created. Setting up of $\mathbf{G}^{-1}$, $\mathbf{g}$, $\mathbf{Z_P}$ and $\mathbf{R_P}$, and computations of $\boldsymbol{\theta}$, $\mathbf{y_P}$ and $\mathbf{G}^{*-1}$ were performed with GNU Octave software. Files containing the user-supplied (co)variance matrix were written following specific rules for each software package and the pseudo-records in $\mathbf{y_P}$ were added to the data file with appropriate levels for the different effects.

The system of equations (V.5), based on the extended data and user-supplied (co)variance matrix, was solved with ASReml, BLUPF90, and DMU4. For DMU4, solutions were obtained with the equivalent nested regression model. The system of equations was solved by direct computation proposed by the 3 software packages used. Solutions were compared with those estimated with GNU Octave software, measuring the relative errors $\Delta = \left| \frac{sol_{ij} - sol_{Oj}}{sol_{Oj}} \right|$ where sol_{Oj} was the estimate for the *j*th level of the effects estimated with GNU Octave and sol_{ij} was the estimate for the *j*th level of the effects estimated with 1 of the 3 software packages used (i.e., *i* = ASReml, BLUPF90, or DMU4). The Δ values ranged from 0.00 to $4.96 * 10^{-6}$ for ASReml, from 0.00 to $9.78 * 10^{-8}$ for BLUPF90, and from 0.00 to $3.83 * 10^{-4}$ for DMU4. Average Δ were $9.11 * 10^{-7}$ for ASReml, $2.54 * 10^{-10}$ for BLUPF90, and $2.52 * 10^{-6}$ for DMU4. Differences between Δ can be explained by the precision considered by the software packages for integer and real variables and by the writing format specifications of the software packages for the solutions.

In conclusion, these 2 examples showed the possibility to alter both mean and (co)variance of the prior distribution associated with random effects for linear mixed models equations with current software packages commonly available in the animal breeding community through the use of an extended data file and a user-supplied (co)variance matrix.

ACKNOWLEDGEMENTS

J. Vandenplas, as a research fellow, and N. Gengler, as a former senior research associate, acknowledge the support of the National Fund for Scientific Research (Brussels, Belgium) for these positions. Additional financial support was provided by the

Ministry of Agriculture of Walloon Region of Belgium (Service Public de Wallonie, Direction générale opérationnelle "Agriculture, Ressources naturelles et Environnement" – DGARNE) through research projects D31-1207, D31-1224, and D31-1274. Useful comments from the 2 reviewers are also acknowledged.

REFERENCES

Dempfle, L. 1977. Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayésiens. *Genet. Sel. Evol.* 9:27–32.

Eaton, J.W., D. Bateman, S. Hauberg, and R. Wehbring. 2011. GNU Octave. A high-level interactive language for numerical computations. 3rd ed. Free Software Foundation, Inc., Boston, MA.

Gianola, D., and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63:217–244.

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

Henderson, C.R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* 21:309–310. (Abstr.).

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

Lindley, D.V., and A.F.M. Smith. 1972. Bayes estimates for the linear model. *J. R. Stat. Soc. Ser. B Methodol.* 34:1–41.

Madsen, P., and J. Jensen. 2012. A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, release 5.1. University of Aarhus, Faculty Agricultural Sciences (DJF), Dept of Genetics and Biotechnology, Research Centre Foulum, Tjele, Denmark.

Misztal, I. 2013. BLUPF90 family of programs. Accessed Jul. 3, 2012. http://nce.ads.uga.edu/wiki/doku.php.

Quaas, R.L. 1984. Linear prediction. Pages 1–76 in BLUP school handbook. Use of mixed models for prediction and for estimation of (co)variance components. Animal Genetics and Breeding Unit, University of New England, New South Wales, Australia.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Schaeffer, L.R., and J. Jamrozik. 1996. Multiple-trait prediction of lactation yields for dairy cows. *J. Dairy Sci.* 79:2044–2055.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer-Verlag, New York, NY.

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

Van Vleck, L.D. 1994. Algorithms for simulation of animal models with multiple traits and with maternal and non-additive genetic effects. *Braz. J. Genet.* 17:53–57.

Vitezica, Z.G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. *Genet. Res.* 93:357–366.

Chapter VI. UNIFIED METHOD TO INTEGRATE AND BLEND SEVERAL, POTENTIALLY RELATED, SOURCES OF INFORMATION FOR GENETIC EVALUATION

Chapter III proposed an algorithm that considers double counting of contributions due to relationships among external animals when their associated external information was integrated into an internal genetic evaluation. However, as highlighted in Chapter II, external information may be based on data shared by the external and internal evaluations and double counting of contributions due to records may thus appear. Furthermore, Chapter II also highlighted that all proposed Bayesian approaches, including those proposed in Chapter III, do not allow the integration of more than one source of external information. Therefore, based on a Bayesian approach, a method integrating and blending simultaneously several sources of information into an internal genetic evaluation while avoiding double counting of contributions due to relationships and due to records was proposed and tested on both simulated and real data in this Chapter.

From: Vandenplas, J., F.G. Colinet and N. Gengler. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. *Genet. Sel. Evol.* 46:59.

ABSTRACT

BACKGROUND

A condition to predict unbiased estimated breeding values by best linear unbiased prediction is to use simultaneously all available data. However, this condition is not often fully met. For example, in dairy cattle, internal (i.e. local) populations lead to evaluations based only on internal records while widely used foreign sires have been selected using internally unavailable external records. In such cases, internal genetic evaluations may be less accurate and biased. Because external records are unavailable, methods were developed to combine external information that summarizes these records, i.e. external estimated breeding values and associated reliabilities, with internal records to improve accuracy of internal genetic evaluations. Two issues of these methods concern double counting of contributions due to relationships and due to records. These issues could be worse if external information came from several evaluations, at least partially based on the same records, and combined into a single internal evaluation. Based on a Bayesian approach, the aim of this research was to develop a unified method to integrate and blend simultaneously several sources of information into an internal genetic evaluation by avoiding double counting of contributions due to relationships and due to records.

RESULTS

This research resulted in equations that integrate and blend simultaneously several sources of information and avoid double counting of contributions due to relationships and due to records. The performance of the developed equations was evaluated using simulated and real datasets. The results showed that the developed equations integrated and blended several sources of information well into a genetic evaluation. The developed equations also avoided double counting of contributions due to relationships and due to records. Furthermore, because all available external sources of information were correctly propagated, relatives of external animals benefited from the integrated information and, therefore, more reliable estimated breeding values were obtained.

CONCLUSIONS

The proposed unified method integrated and blended several sources of information well into a genetic evaluation by avoiding double counting of contributions due to relationships and due to records. The unified method can also be extended to other

types of situations such as single-step genomic or multi-trait evaluations, combining information across different traits.

BACKGROUND

Simultaneous use of all available data by Best Linear Unbiased Prediction (BLUP) is a condition to predict unbiased estimated breeding values (EBV; Henderson, 1984). However, this condition is not often fully met. For example, in dairy cattle, while foreign bulls are often widely used, e.g., through artificial insemination, evaluating populations based only on internal phenotypic data (i.e., internal records) will lead to potentially biased and less accurate evaluations (Bonaiti and Boichard, 1995). The reason is that external phenotypic data used to select these foreign bulls are not available at the internal level. Multiple across country evaluation (MACE), performed at an international level by International Bull Service (Interbull, Uppsala, Sweden), allows EBV, for each population scale, to be aggregated into a single ranking for international dairy sires. However, this has no influence on internal evaluations. These issues are also relevant in the setting of current developments of genomic multi-step or single-step prediction methods (e.g., Aguilar et al., 2010; Christensen and Lund, 2010; VanRaden, 2012).

Because external phenotypic data are not available at the internal level, methods were developed to combine external information, i.e. external EBV and associated reliabilities (REL), with internal data to improve accuracy of internal genetic evaluations. A first type of approaches is based on performing, a posteriori, an additional step after the genetic evaluation at the internal level. These approaches combine external and internal EBV based on selection index theory (e.g., VanRaden, 2001), based on mixed model theory (e.g., Täubert et al., 2000) or based on bivariate evaluations (e.g., Mäntysaari and Strandén, 2010). One of the problems of a posteriori approaches is that external information used for selection will not contribute to the estimation of fixed effects at the internal level, which can create potential biases. A second type of approaches combines external information simultaneously with internal phenotypic data in genetic evaluations at the internal level. Simultaneous combination of external information and internal phenotypic data can be carried out using different methods. However, to our knowledge, the following two approaches are the most used. First, external information can be directly included by converting this information into pseudo-records for fictive daughters of external animals (e.g., Bonaiti and Boichard, 1995). Similar approaches were proposed to include external information into internal single-step genomic evaluations (e.g.,

VanRaden, 2012; Přibyl et al., 2013). Second, external information can be directly included by changing both the mean and (co)variance of the prior distributions of genetic effects in a Bayesian approach, as mentioned, for example, by Gianola and Fernando (1986). Quaas and Zhang (2001, 2006) and Legarra et al. (2007) proposed two Bayesian derivations to integrate external information into internal genetic evaluations in the context of multi-breed genetic evaluations for beef cattle. These two derivations consider external information as priors of internal genetic effects. Vandenplas and Gengler (2012) compared these two derivations and proposed some improvements that concerned mainly double counting of contributions due to relationships among external animals. Indeed, an EBV of an animal combines information from its own records (i.e., contributions due to own records) and from records of all relatives through its parents and its progeny (i.e., contributions due to relationships; VanRaden, 2001; Misztal and Wiggans, 1988). Therefore, integration of EBV for relatives can cause the same contributions that are due to relationships to be counted several times, which can bias genetic evaluations at the internal level.

Both types of approaches, i.e., that combine available information a posteriori or simultaneously, raise another issue if the external information results from an evaluation that combines external and internal records, which is that some contributions due to records will be considered several times when external information is combined with internal records. Although this is a major issue for common sources of external information (e.g., MACE information), to our knowledge, only a few studies have proposed solutions to the double counting of contributions due to records (e.g., VanRaden, 2012; Gengler and Vanderick, 2008; VanRaden and Tooker, 2012). The proposed solutions were developed as an additional pre-processing step before integration of external information. Furthermore, in many situations, integration of several sources of external information into genetic evaluations at the internal level may be needed but this has not been studied to our knowledge. In such cases, double counting of contributions due to records could be worse if external information from several evaluations were, at least partially, based on the same internal records, and/or on the same external records, and integrated into the same genetic evaluation.

Thus, the aim of this research was to develop a unified method to integrate and blend simultaneously several, potentially related, external sources of information into an internal genetic evaluation based on a Bayesian approach. In order to achieve this aim, methods were developed to avoid double counting of contributions due to relationships

and due to records generated by the integration of several sources of information. This resulted in modified mixed model equations (MME) that integrate and blend simultaneously several sources of information and avoid double counting of contributions due to relationships and due to records. The performance of the developed equations was evaluated using simulated and real datasets.

METHODS

INTEGRATION OF SEVERAL SOURCES OF EXTERNAL INFORMATION

Assume an internal genetic evaluation (referred to with the subscript E_0) based on internal data (i.e., a set of phenotypic records: $\mathbf{y}_{E_0}$) that provides internal information (i.e., EBV and associated REL obtained from the evaluation E_0). Also, assume an *i*th external genetic evaluation (*i* = 1, 2, ..., *N*, referred to with the subscript E_i) that is based on the *i*th source of external data (i.e., the *i*th set of phenotypic records not used by evaluation E_0 and free of internal data: $\mathbf{y}_{E_i}$) and that provides the *i*th source of external information, i.e., all available external EBV (EBV_{Ei}) and associated REL (e.g., EBV and associated REL obtained from evaluation E_1 based only on external data E_1, and EBV and associated REL obtained from evaluation E_2 based only on external data E_2). In addition to be free of internal data, it is also assumed that each *i*th source of external data was free of the other *N*-1 sources of external data. These assumptions lead to each *i*th source of external information to be free of internal data and information, as well as of the *N*-1 other external data and information.

Two groups of animals, hereafter called external and internal animals, are defined according to the *i*th source of external information. Therefore, for each *i*th source of external information, external animals (subscript A_i with *i* = 1, 2,…, *N*) are defined as animals that are associated with this *i*th source of external information and for which internal data and/or information is available or that have relationships with animals involved in the internal evaluation E_0. All animals that are not defined as external animals for the *i*th source of external information are defined as internal animals (subscript A_i^0). Internal animals are then defined as animals associated with only internal information when considering the *i*th source of external information. It is noted that external animals may be associated with different sources of external information and that an animal may be considered as external for the *i*th source of external information and internal for the *N*-

1 other sources of external information because the definitions of external and internal animals depend only on the source of external information considered. Those definitions are summarized in Table VI-1. In addition, because pedigree information for animals can be easily integrated into a genetic evaluation, it is assumed that the same complete pedigree information could be used for all animals for each genetic evaluation. Concerning the notation of matrices in the following sections (e.g., $\mathbf{X}_{\mathbf{E_i}(\mathbf{A_l})}$), the subscript E_i refers to the *i*th source of external information and the subscript within brackets (A_l) refers to the *l*th group of animals.

Table VI-1. Concepts related to the terminology of internal and external animals and information

Data[1]	Pedigree	
	Internal animals	External animals
Internal data	Internal evaluation Internal information[2]	Internal - external evaluations Internal - external information
External data	-	External evaluation External information

[1]Data = set of phenotypic records;
[2]Information = estimated breeding values and associated reliabilities.

The *N* sources of external information must be integrated into the internal evaluation E_0. For external animals associated with the *i*th source of external information, all EBV_{Ei} are summarized by the vector of external EBV, $\hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i})}$, and by the prediction error (co)variance matrix, $\mathbf{D}_{\mathbf{E_i}(\mathbf{A_i})}$. Because $\hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i})}$ could be estimated with an equivalent external genetic evaluation that includes the internal animals in the pedigree through a genetic (co)variance matrix extended to all animals for the *i*th source of external information, $\mathbf{G}_{\mathbf{E_i}} = \begin{bmatrix} \mathbf{G}_{\mathbf{E_i}(\mathbf{A_i^0 A_i^0})} & \mathbf{G}_{\mathbf{E_i}(\mathbf{A_i^0 A_i})} \\ \mathbf{G}_{\mathbf{E_i}(\mathbf{A_i A_i^0})} & \mathbf{G}_{\mathbf{E_i}(\mathbf{A_i A_i})} \end{bmatrix}$, the vector of external EBV for all internal and external animals for the *i*th source of external information is estimated as:

$$\hat{\mathbf{u}}_{\mathbf{E_i}} = \begin{bmatrix} \hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i^0})} \\ \hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i})} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{\mathbf{E_i}(\mathbf{A_i^0 A_i})} \mathbf{G}^{-1}_{\mathbf{E_i}(\mathbf{A_i A_i})} \hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i})} \\ \hat{\mathbf{u}}_{\mathbf{E_i}(\mathbf{A_i})} \end{bmatrix}.$$

A modified set of multi-trait mixed model equations that integrate *N* sources of external information, each summarized by $\hat{\mathbf{u}}_{\mathbf{E_i}}$ and its associated prediction error (co)variance matrix $\mathbf{D}_{\mathbf{E_i}}$ for the *i*th source of external information, can be written as [See Additional file 1 for the derivation of the equations]:

$$\begin{bmatrix} \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{X}_{\mathbf{E_0}} & \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{Z}_{\mathbf{E_0}} \\ \mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{X}_{\mathbf{E_0}} & \mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{Z}_{\mathbf{E_0}} + \mathbf{G}_{\mathbf{E_0}}^{-1} + \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1} - \mathbf{G}_{\mathbf{E_i}}^{-1}\right) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{E_0}} \\ \hat{\mathbf{u}}_{\mathbf{E_0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{y}_{\mathbf{E_0}} \\ \mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{y}_{\mathbf{E_0}} + \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1}\hat{\mathbf{u}}_{\mathbf{E_i}}\right) \end{bmatrix}, \quad \text{(VI.1)}$$

where $\mathbf{X}_{\mathbf{E_0}}$ and $\mathbf{Z}_{\mathbf{E_0}}$ are incidence matrices relating records in $\mathbf{y}_{\mathbf{E_0}}$ to the vector of fixed effects $\hat{\boldsymbol{\beta}}_{\mathbf{E_0}}$ and the vector of random additive genetic effects $\hat{\mathbf{u}}_{\mathbf{E_0}}$, respectively, $\mathbf{G}_{\mathbf{E_0}}^{-1}$ is the inverse of the internal additive genetic (co)variance matrix associated with the internal genetic evaluation E_0 that includes all internal and external animals and $\mathbf{R}_{\mathbf{E_0}}^{-1}$ is the inverse of the residual (co)variance matrix.

For the approximation of $\mathbf{D}_{\mathbf{E_i}}^{-1}$, it can be shown that [See Additional file 1]: $\mathbf{D}_{\mathbf{E_i}}^{-1} = \mathbf{G}_{\mathbf{E_i}}^{-1} + \mathbf{Z'}_{\mathbf{E_i}}\mathbf{R}_{\mathbf{E_i}}^{-1}\mathbf{Z}_{\mathbf{E_i}}$, where $\mathbf{Z}_{\mathbf{E_i}}$ is the incidence matrix relating records of *i*th external data to internal and external animals and $\mathbf{R}_{\mathbf{E_i}}^{-1}$ is the residual (co)variance matrix for the *i*th source of external information. Thereby, $\mathbf{D}_{\mathbf{E_i}}^{-1}$ is approximated by $\mathbf{D}_{\mathbf{E_i}}^{-1} = \mathbf{G}_{\mathbf{E_i}}^{-1} + \boldsymbol{\Lambda}_{\mathbf{E_i}}$, where $\boldsymbol{\Lambda}_{\mathbf{E_i}}$ is a block diagonal variance matrix with one block per animal (Quaas and Zhang, 2006; Vandenplas and Gengler, 2012) and $\boldsymbol{\Lambda}_{\mathbf{E_i}} \approx \mathbf{Z}_{\mathbf{E_i}}^{'}\mathbf{R}_{\mathbf{E_i}}^{-1}\mathbf{Z}_{\mathbf{E_i}}$. Each diagonal block of $\boldsymbol{\Lambda}_{\mathbf{E_i}}$ is equal to $\boldsymbol{\Lambda}_{\mathbf{E_i(j)}}\mathbf{R}_{\mathbf{0}}^{-1}\boldsymbol{\Lambda}_{\mathbf{E_i(j)}}$ for j = 1, 2, …, J animals, where the matrix $\mathbf{R}_{\mathbf{0}}$ is a matrix of residual (co)variance among traits and the jth matrix $\boldsymbol{\Lambda}_{\mathbf{E_i(j)}}$ is a diagonal matrix with elements $\sqrt{RE_{ijk}}$ where k = 1, 2, ..., K traits. Element RE_{ijk} is the effective number of records, i.e., record equivalents, for the jth animal for the kth trait associated with the ith source (Misztal and Wiggans, 1988; Vandenplas and Gengler, 2012). Record equivalents express the quantity of contributions due to relationships and/or due to records considered for the evaluation of an animal. For internal animals, RE_{ijk} is equal to 0 because all contributions are only due to the relationships among external and internal animals. For external animals, if double counting of contributions due to relationships among them is not taken into account, $RE_{ijk} = \frac{1-h_k^2}{h_k^2} * \frac{REL_{ijk}}{1-REL_{ijk}}$ for the jth animal for the kth trait associated with the ith source, where h_k^2 is the heritability of the kth trait (Misztal and Wiggans, 1988; VanRaden and Wiggans, 1991). If double

counting of contributions due to relationships among external animals is taken into account, RE_{ijk} only expresses the amount of contributions due to records and can be estimated through a two-step algorithm (TSA; Vandenplas and Gengler, 2012). The first step of this TSA determines external animals associated with external information that includes only contributions due to relationships. The second step estimates the amount of contributions due to records (expressed as RE) for external animals associated with information that combines both contributions due to relationships and own records. Note that the proposed approximation of $\mathbf{Z}'_{\mathbf{E_i}}\mathbf{R}^{-1}_{\mathbf{E_i}}\mathbf{Z}_{\mathbf{E_i}}$ differs from the approximation proposed by Quaas and Zhang (2006). Indeed, they proposed to approximate each diagonal block of $\mathbf{\Lambda}_{\mathbf{E_i}}$ by $\mathbf{\Delta}_{\mathbf{Qi(j)}}\mathbf{G}_{\mathbf{0}}^{-1}\mathbf{\Delta}_{\mathbf{Qi(j)}}$, where the matrix $\mathbf{G_0}$ is a matrix of genetic (co)variance among traits and $\mathbf{\Delta}_{\mathbf{Qi(j)}}$ is a diagonal matrix with elements:

$\sqrt{\delta_{ijk}} = \sqrt{REL_{ijk}/(1-REL_{ijk})}$.

Also, the multi-trait MME (VI.1) that integrate *N* sources of external information differ from the usual multi-trait MME only by the terms $\sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1} - \mathbf{G}_{\mathbf{E_i}}^{-1}\right)$ and $\sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1}\hat{\mathbf{u}}_{\mathbf{E_i}}\right)$:

$$\begin{bmatrix} \mathbf{X}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{X}_{\mathbf{E_0}} & \mathbf{X}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{Z}_{\mathbf{E_0}} \\ \mathbf{Z}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{X}_{\mathbf{E_0}} & \mathbf{Z}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{Z}_{\mathbf{E_0}} + \mathbf{G}^{-1}_{\mathbf{E_0}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{E_0}} \\ \hat{\mathbf{u}}_{\mathbf{E_0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{y}_{\mathbf{E_0}} \\ \mathbf{Z}'_{\mathbf{E_0}}\mathbf{R}^{-1}_{\mathbf{E_0}}\mathbf{y}_{\mathbf{E_0}} \end{bmatrix}. \qquad \text{(VI.2)}$$

Furthermore, it was previously assumed that the whole pedigree is available for all genetic evaluations. The additive genetic (co)variance matrices that include all internal and external animals are then equal for all genetic evaluations (i.e., $\mathbf{G}_{\mathbf{E_0}} = \mathbf{G}_{\mathbf{E_1}} = \mathbf{G}_{\mathbf{E_2}} = \ldots = \mathbf{G}_{\mathbf{E_N}}$). Nevertheless, each internal or external genetic evaluation could be performed as a single-step genomic evaluation (e.g., Aguilar et al., 2010; Christensen and Lund, 2010) without modifications to the Bayesian derivation [See Additional file 1] because assumptions on the different matrices $\mathbf{G}_{\mathbf{E_i}}$ were not limiting. Such cases would lead to $\mathbf{G}_{\mathbf{E_0}} \neq \mathbf{G}_{\mathbf{E_i}}$. For example, integration of external information provided by the usual MME into a single-step genomic evaluation would lead to $\mathbf{G}_{\mathbf{E_0}} \neq \mathbf{G}_{\mathbf{E_i}}$ because $\mathbf{G}_{\mathbf{E_0}}$ would include genomic information (Aguilar et al., 2010; Christensen and Lund, 2010), unlike $\mathbf{G}_{\mathbf{E_i}}$.

INTEGRATION OF SEVERAL SOURCES OF EXTERNAL INFORMATION BY AVOIDING DOUBLE COUNTING OF CONTRIBUTIONS DUE TO RECORDS

Assumptions stated in the previous section led to each source of external information to be obtained from an external evaluation that was based only on external data and free of internal data and information, as well as of the *N*-1 other external data and information. In practice, this assumption is not necessarily valid because a source of external information may be obtained from an external evaluation based on external data and/or information and also on internal data and/or information (e.g., EBV and associated REL obtained in country E_1 based on external data E_1 and on internal data E_0). Thus, double counting of contributions due to records between internal and external information must be taken into account, as detailed below.

For the *i*th source of external information, internal information included into external information (subscript I_i) associated with the external animals can be summarized as $\hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i})}$, i.e., the vector of internal EBV associated with external animals for which external information included both external and internal information, and by $\mathbf{D}_{\mathbf{I_i}(\mathbf{A_i})}$, the prediction error (co)variance matrix associated with $\hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i})}$.

A modified set of multi-trait mixed model equations that integrate several sources of external information and take double counting of contributions due to records between external and internal information into account, can be written as follows [See Additional file 2]:

$$\begin{bmatrix} \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{X}_{\mathbf{E_0}} & \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{Z}_{\mathbf{E_0}} \\ \mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{X}_{\mathbf{E_0}} & \begin{matrix}\mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{Z}_{\mathbf{E_0}} + \mathbf{G}_{\mathbf{E_0}}^{-1} + \\ \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1} - \mathbf{G}_{\mathbf{E_i}}^{-1}\right) - \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{I_i}}^{-1} - \mathbf{G}_{\mathbf{I_i}}^{-1}\right)\end{matrix} \end{bmatrix}, \qquad \text{(VI.3)}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{E_0}} \\ \hat{\mathbf{u}}_{\mathbf{E_0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{y}_{\mathbf{E_0}} \\ \mathbf{Z'}_{\mathbf{E_0}}\mathbf{R}_{\mathbf{E_0}}^{-1}\mathbf{y}_{\mathbf{E_0}} + \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E_i}}^{-1}\hat{\mathbf{u}}_{\mathbf{E_i}}\right) - \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{I_i}}^{-1}\hat{\mathbf{u}}_{\mathbf{I_i}}\right) \end{bmatrix}$$

where $\mathbf{G}_{\mathbf{I_i}}$ is a genetic (co)variance matrix for all animals for the internal information included into the *i*th source of external information, $\hat{\mathbf{u}}_{\mathbf{I_i}} = \begin{bmatrix} \hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i^0})} \\ \hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i})} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{\mathbf{I_i}(\mathbf{A_i^0}\mathbf{A_i})}\mathbf{G}_{\mathbf{I_i}(\mathbf{A_i}\mathbf{A_i})}^{-1}\hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i})} \\ \hat{\mathbf{u}}_{\mathbf{I_i}(\mathbf{A_i})} \end{bmatrix}$ is the vector of internal EBV associated with the *i*th source of external information that includes internal information and $\mathbf{D}_{\mathbf{I_i}}^{-1}$ is the

inverse of the prediction error (co)variance matrix associated with $\hat{\mathbf{u}}_{\mathbf{I}_i}$ and approximated as detailed in the previous section.

If the *i*th source of external information does not include internal information for external animals, the vector $\hat{\mathbf{u}}_{\mathbf{I}_i}$ is undetermined and the matrix $\mathbf{D}_{\mathbf{I}_i}^{-1}$ is equal to $\mathbf{G}_{\mathbf{I}_i}^{-1}$. This leads to the system of equations (VI.1).

BLENDING SEVERAL SOURCES OF EXTERNAL INFORMATION BY AVOIDING DOUBLE COUNTING OF CONTRIBUTIONS DUE TO RECORDS

Equations to blend several sources of external information by avoiding double counting of contributions due to records among internal and external data/information can be derived from the system of equations (VI.3) by assuming that $\mathbf{y}_{\mathbf{E}_0}$ has no records (i.e. that $\mathbf{y}_{\mathbf{E}_0}$ is an empty vector). Then, the equation can be written as follows:

$$\left(\mathbf{G}_{\mathbf{E}_0}^{-1} + \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E}_i}^{-1} - \mathbf{G}_{\mathbf{E}_i}^{-1}\right) - \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{I}_i}^{-1} - \mathbf{G}_{\mathbf{I}_i}^{-1}\right)\right)\hat{\mathbf{u}}_{\mathbf{E}_0} = \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{E}_i}^{-1}\hat{\mathbf{u}}_{\mathbf{E}_i}\right) - \sum_{i=1}^{N}\left(\mathbf{D}_{\mathbf{I}_i}^{-1}\hat{\mathbf{u}}_{\mathbf{I}_i}\right). \qquad \text{(VI.4)}$$

SIMULATED EXAMPLE

The system of equations (VI.3) was tested using data simulated with the software package GNU Octave (Eaton et al., 2011). The context of the simulation was a country that imports sires from another country to generate the next generation of production animals and potential sires. Populations of the importing country (hereafter called the internal population) and of the exporting country (hereafter called the external population) were assumed to belong to the same breed. Each population included about 1000 animals distributed over five generations and was simulated from 120 female and 30 male founders. For both populations, milk yield in the first lactation was simulated for each female with progeny, following Van Vleck (1994). A herd effect nested within-population was randomly assigned to each phenotypic record. To obtain enough observations per level for the herd effect, each herd included at least 40 females. Phenotypic variance and heritability were assumed to be $3.24 * 10^6$ kg² and 0.25, respectively.

To simulate the internal and external populations, the following rules were applied to generate each new generation. First, from the second generation, both females and males older than one year old were considered as mature for breeding and a male could be mated during at most two breeding years. Second, 95% of the available females and 75% of the available males with the highest true breeding values were selected for breeding.

Third, all selected females were randomly mated with the selected males. The maximum number of males mated to produce the next generation was set to 25. Furthermore, a mating could be performed only if the additive relationship coefficient between male and female was less than 0.5 and if the female had less than three progeny.

The external population was simulated first and additional rules were applied to this population. For this population, males that were selected for mating only originated from the external population and 60% of the external male offspring with the lowest true breeding values were culled in each generation. Then, the internal population was simulated. For this population, males were selected among all available internal males and a subset of selected external sires. This subset of external sires included the first 50 sires with the highest true breeding values in the external population. Also, 99% of internal male offspring with the lowest true breeding values were culled in each generation. No female offspring was culled in either population.

Using the simulated data, three genetic evaluations were performed (Table VI-2):

(a) A joint evaluation ($EVAL_J$) was performed as a BLUP evaluation using the system of equations (VI.2) and based on external and internal pedigree and data. This evaluation was assumed to be the reference.

(b) An internal evaluation ($EVAL_I$) was performed as a BLUP evaluation using the system of equations (VI.2) and based on internal pedigree and data.

(c) An external evaluation ($EVAL_E$) was performed as a BLUP evaluation using the system of equations (VI.2) and based on external pedigree and data.

Table VI-2. Genetic evaluations performed for the simulated example

	Genetic evaluations[1]					
	J	E	I	BE	BJ	BJ-I
External pedigree	X	X				
Internal pedigree	X		X	X	X	X
External data	X	X				
Internal data	X		X	X	X	X
Integrated information (50 external sires)						
External EBV and REL				X		
Joint EBV and REL					X	X
Internal EBV and REL						X

[1]J = Joint; E = External; I = Internal; BE = Bayesian External; BJ = Bayesian Joint; BJ-I = Bayesian Joint minus Internal.

Three Bayesian evaluations that integrated information provided by $EVAL_E$ or by $EVAL_J$ for the 50 external sires into $EVAL_I$ were also performed. Because the external sires were related, double counting of contributions due to relationships existed and this

was taken into account for the three Bayesian evaluations through the TSA (Vandenplas and Gengler, 2012). Double counting of contributions due to records could also exist with the integration of information provided by $EVAL_J$ into $EVAL_I$ because $EVAL_J$ and $EVAL_I$ were partially based on the same data (i.e., internal data). The following three Bayesian evaluations were performed:

(d) A Bayesian evaluation using the system of equations (VI.1) and using EBV and prediction error variances (PEV) obtained from $EVAL_E$ associated with the 50 external sires that were used inside the internal population as external information ($EVAL_{BE}$).

(e) A Bayesian evaluation using the system of equations (VI.1) and EBV and PEV obtained from $EVAL_J$ associated with the 50 external sires as external information (hereafter called joint information) ($EVAL_{BJ}$). Although $EVAL_J$ was based on external and internal data, double counting of contributions due to records between joint and internal information was not taken into account.

(f) A Bayesian evaluation integrating joint information by using the system of equations (VI.3) and taking into account double counting of contributions due records among internal and joint information ($EVAL_{BJ-I}$). Double counting of contributions due to records among internal and joint information was taken into account by using EBV and PEV obtained from $EVAL_I$ associated with the 50 external sires.

The simulation was replicated 100 times. Comparisons between $EVAL_J$ and $EVAL_I$, $EVAL_{BE}$, $EVAL_{BJ}$, or $EVAL_{BJ-I}$ were performed separately for the 50 external sires and for the internal animals. Comparisons were based on:

(1) Spearman's rank correlation coefficients (r) of EBV obtained from $EVAL_J$ (EBV_J) with EBV obtained from $EVAL_I$ (EBV_I), $EVAL_{BE}$ (EBV_{BE}), $EVAL_{BJ}$ (EBV_{BJ}), and $EVAL_{BJ-I}$ (EBV_{BJ-I}),

(2) regression coefficients (a) of EBV_J on EBV_I, EBV_{BE}, EBV_{BJ}, and EBV_{BJ-I}, and

(3) coefficients of determination (R^2) associated with the regressions,

(4) the total amount of RE (RE_{tot}) associated with external information, joint information and joint information corrected for the included internal information, and

(5) mean squared errors (MSE) of EBV_I, EBV_{BE}, EBV_{BJ}, and EBV_{BJ-I}, expressed as a percentage of MSE obtained for EBV_I. For each replicate, the MSE obtained for EBV_I was reported to a relative value of 100 before the different computations of

MSE.

Because the TSA was applied before all three Bayesian evaluations, RE_{tot} were free of contributions due to relationships estimated by the Bayesian evaluations. For an easier understanding of the results and discussion, RE can be transformed into daughter equivalents (DE) through $DE_{ijk} = \frac{4 - h_k^2}{1 - h_k^2} * RE_{ijk}$ (VanRaden and Wiggans, 1991). All results were the average of the 100 replicates.

WALLOON EXAMPLE

Even if MACE allows the aggregation of EBV for dairy sires, internal genetic evaluations for animals not associated with MACE information (e.g., cows, calves, young sires) are not influenced by external information considered by the MACE for dairy sires and may be still biased. Therefore, integration of MACE information into internal evaluations, as well as blending of MACE and internal information, could benefit those animals. The performance of equation (VI.4) that blends MACE and internal information was evaluated in the context of the official Walloon genetic evaluation for Holstein cattle.

The Walloon example used information for milk, fat and protein yields for Holstein cattle provided by the official Walloon genetic evaluation (Auvray and Gengler, 2002; Croquet et al., 2006). The genetic variances were those used for the official Walloon genetic evaluation (Auvray and Gengler, 2002) and were equal to 280 425 kg² for milk yield, to 522.6 kg² for fat yield and to 261.5 kg² for protein yield. The respective heritabilities were equal to 0.38, 0.43 and 0.41. The pedigree file was extracted from the database used for the official Walloon genetic evaluation ($EVAL_W$) and covered up to six known ancestral generations. The extraction was performed for a randomly selected group of 1909 animals (potentially genotyped) born after 1998. The selected group included sires, cows and calves that were used or were not at the internal level. After extraction, the pedigree file contained 16 234 animals.

Internal information included EBV and associated REL estimated from data provided by the Walloon Breeding Association (EBV_W, REL_W) for the $EVAL_W$ for milk production of April 2013 (Auvray and Gengler, 2002; Croquet et al., 2006). A total of 12 046 animals were associated with an available EBV_W. External information included EBV and REL for 1981 sires provided with the official release for the April 2013 MACE performed by Interbull ($EVAL_{MACE}$, EBV_{MACE}, REL_{MACE}; Interbull, 2013). It should be noted that the Walloon Region in Belgium participated in the April 2013 MACE. Internal

and external information were harmonized between the Walloon and MACE evaluations by adjusting scales and mean differences towards the original expression of the trait in the Walloon genetic evaluations. External information was then considered to be the same trait as the internal phenotype trait.

Unlike the simulated example, no joint evaluation based on Walloon and external records was available for both external and internal animals. Because $EVAL_{MACE}$ aggregated EBV from several national genetic evaluations for sires, it was considered as the reference for the evaluated sires. Walloon and MACE information were blended by using equation (VI.4) for the following four cases: with or without consideration of double counting of contributions due to relationships and with or without consideration of double counting of contributions due to records (Table VI-3). Double counting of contributions due to relationships was possible because all animals associated with Walloon and/or MACE information were related. Double counting of contributions due to records was also possible because MACE information associated with the 1981 sires included contributions provided by $EVAL_W$. Thus, to test the importance of both double counting issues, the following four cases were evaluated:

(a) Walloon and MACE information were blended without considering double counting of contributions due to records and due to relationships ($EVAL_{BLNN}$, EBV_{BLNN}, REL_{BLNN}).

(b) Walloon and MACE information were blended by considering only double counting of contributions due to records ($EVAL_{BLRE}$, EBV_{BLRE}, REL_{BLRE}). To achieve this goal, the contribution of Walloon information into MACE information was determined based on the domestic effective daughter equivalents (EDC) associated with EBV_{MACE} and REL_{MACE} and provided with the official release for the 2013 April MACE by Interbull. MACE information free of Walloon information was reported by a domestic EDC equal to 0. A total of 601 sires were associated with an EDC greater than 0. For these 601 sires, EBV and associated REL estimated from Walloon data and contributing to the April 2013 MACE routine-run (EBV_{Wc}, REL_{Wc}) were considered by $EVAL_{BLRE}$ to take double counting of contributions due to records into account. Double counting of contributions due to relationships was not taken into account for either Walloon or MACE information.

(c) Walloon and MACE information were blended by only considering double counting of contributions due to relationships among all animals ($EVAL_{BLR}$,

EBV_{BLR}, REL_{BLR}). The TSA was therefore applied for Walloon and MACE information. Double counting of contributions due to records was not considered.

(d) Walloon and MACE information were blended by considering both double counting of contributions due to records and due to relationships ($EVAL_{BL}$, EBV_{BL}, REL_{BL}). Reliabilities for EBV_{BLNN}, EBV_{BLRE}, EBV_{BLR} and EBV_{BL} were computed using the equation $REL = 1 - PEV / \sigma_g^2$, where σ_g^2 is the genetic variance for the corresponding trait and PEV is the prediction error variance obtained from the diagonal element of the inverted left hand side (LHS) of the equation (VI.4).

Table VI-3. Bayesian evaluations performed for the Walloon example

	Bayesian evaluations			
	BLNN	BLRE	BLR	BL
Available estimated breeding values and reliabilities				
Official Walloon evaluation	X	X	X	X
Multiple Across Country Evaluation	X	X	X	X
Double counting accounted				
Records		X		X
Relationships			X	X

As explained previously, $EVAL_{MACE}$ was considered as the reference for sires evaluated through $EVAL_{MACE}$. Comparisons between $EVAL_{MACE}$ and $EVAL_{W}$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ or $EVAL_{BL}$ were performed based on:

(1) Spearman's rank correlation coefficients (r) of EBV_{MACE} with EBV_{W}, EBV_{BLNN}, EBV_{BLRE}, EBV_{BLR} and EBV_{BL},

(2) MSE of EBV_{W}, EBV_{BLNN}, EBV_{BLRE}, EBV_{BLR}, and $EVAL_{BL}$ (i.e. mean squared errors expressed as a percentage of average MSE of EBV_{W}),

(3) regression coefficients (a) and,

(4) R^2 of the regressions of $EVAL_{MACE}$ on the five other evaluations (i.e., $EVAL_{W}$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ and $EVAL_{BL}$),

(5) RE_{tot} and

(6) average REL.

Comparisons concerned two groups of sires. A first group of sires included 1212 sires that were associated with both Walloon and MACE information and had daughters with records in the Walloon region dataset (hereafter called "internally used sires"). A second group of sires included 631 sires that were associated with both Walloon and MACE information but had no daughters with records in the Walloon region dataset (i.e.,

they had only foreign, or external, daughters; hereafter called “internally unused sires”). The RE_{tot} were free of contributions due to relationships that were estimated by the Bayesian evaluations but could include contributions due to relationships that resulted from the previous genetic evaluation if the TSA was not applied.

The effect of blending MACE and Walloon information was also studied for internal animals that were not associated with MACE information and that were sired by internally used sires by considering (1) r between $EVAL_{BL}$ and $EVAL_{W}$, $EVAL_{BLNN}$, $EVAL_{BLRE}$ or $EVAL_{BLR}$, (2) RE_{tot} and (3) average REL. Three groups of internal animals were defined depending on their REL_{W}. The first group included internal animals that were associated with a REL_{W} lower than 0.50, the second group included internal animals that were associated with a REL_{W} between 0.50 and 0.75, and the third group included internal animals with a REL_{W} equal or higher than 0.75.

All blending evaluations were performed using a version of the BLUPF90 program (Misztal, 2013) modified to implement the equations (VI.1), (VI.3) and (VI.4).

RESULTS AND DISCUSSION

SIMULATED EXAMPLE

On average, each of the 100 simulated internal and external populations included 1048 animals. Results for r, MSE, a and R^2 for prediction of EBV_J are in Table VI-4 for the 50 external sires and for the internal animals.

Compared to the rankings of $EVAL_I$, integration of external or joint information for the 50 external sires led to rankings of $EVAL_{BE}$, $EVAL_{BJ}$ or $EVAL_{BJ\text{-}I}$ that were more similar to those of $EVAL_J$. Rank correlations r increased from 0.57 for $EVAL_I$ to at least 0.95 for $EVAL_{BJ}$ for the 50 external sires and from 0.93 for $EVAL_I$ to at least 0.98 for $EVAL_{BJ}$ for internal animals (Table VI-4). Furthermore, MSE, a and R^2 also showed that the integration of external or joint information for the 50 external animals with $EVAL_{BE}$, $EVAL_{BJ}$ or $EVAL_{BJ\text{-}I}$ led to better predictions of EBV_J for both external and internal animals (Table VI-4). Therefore, the observations that internals animals related to the 50 external sires were also better predicted by $EVAL_{BE}$, $EVAL_{BJ}$ and $EVAL_{BJ\text{-}I}$, compared to $EVAL_I$, revealed that the external information propagated from the 50 external sires to relatives.

Table VI-4. Average (SD in parentheses) of parameters obtained for the simulated example over 100 replicates

Concerned animals[1]	Genetic evaluation[2]	r^2	MSE[3]	a[4]	R^2 [4]	RE_{tot} [5]
Internal animals						
	$EVAL_I$	0.934 (0.021)	100.00 (28.621)	0.982 (0.042)	0.896 (0.030)	-
	$EVAL_{BE}$	>0.999 (0.000)	0.61 (0.58)	0.997 (0.005)	0.999 (0.001)	-
	$EVAL_{BJ}$	0.979 (0.005)	34.26 (7.92)	0.977 (0.024)	0.965 (0.008)	-
	$EVAL_{BJ-I}$	0.996 (0.001)	6.78 (3.02)	1.021 (0.013)	0.993 (0.002)	-
External sires						
	$EVAL_I$	0.571 (0.131)	100.00 (32.31)	0.712 (0.168)	0.391 (0.146)	-
	$EVAL_{BE}$	0.997 (0.001)	0.35 (0.22)	1.000 (0.011)	0.998 (0.002)	76.3 (5.1)
	$EVAL_{BJ}$	0.956 (0.017)	17.16 (4.18)	0.821 (0.039)	0.924 (0.030)	141.5 (7.8)
	$EVAL_{BJ-I}$	0.996 (0.002)	0.60 (0.26)	0.993 (0.012)	0.996 (0.002)	78.7 (5.1)

[1]Internal animals = animals associated with only internal information; External sires = sires associated with external information;
[2]$EVAL_I$ = BLUP evaluation based on internal pedigree and data; $EVAL_{BE}$ = Bayesian evaluation using external EBV and PEV associated with the 50 external sires used in the internal population; $EVAL_{BJ}$ = Bayesian evaluation using EBV and PEV obtained from the joint evaluation and associated with the 50 external sires; $EVAL_{BJ-I}$ = Bayesian evaluation using EBV and PEV obtained from the joint and from internal evaluations and associated with the 50 external sires to avoid double counting among internal and joint information; r = rank correlations between EBV estimated by $EVAL_J$ and by $EVAL_I$, $EVAL_{BE}$, $EVAL_{BJ}$ or $EVAL_{BJ-I;}$
[3]MSE = mean squared errors expressed as a percentage of the average internal MSE between a joint evaluation and $EVAL_I$, $EVAL_{BE}$, $EVAL_{BJ}$ or $EVAL_{BJ-I}$;
[4]a = regression coefficient and R^2 = coefficient of determination of the regression of EBV estimated by the joint evaluation on EBV estimated by $EVAL_I$, $EVAL_{BE}$, $EVAL_{BJ}$ or $EVAL_{BJ-I}$;
[5]RE_{tot} = total amount of record equivalents free of contributions due to relationships among external animals.

The RE_{tot} associated with $EVAL_{BE}$ was equal to 76.3 (which also corresponded to 381.6 DE), while the RE_{tot} associated with $EVAL_{BJ}$ was equal to 141.5 (DE = 707.7, Table VI-4). The higher RE_{tot} associated with $EVAL_{BJ}$ showed that double counting of contributions due to records was present when joint information was integrated. Indeed, joint information contained both external and internal information. The RE_{tot} associated with $EVAL_{BJ-I}$ was equal to 78.7 (DE = 393.3, Table VI-4). While this latter RE_{tot} is slightly higher (i.e., 3.1% on average) than the RE_{tot} associated with $EVAL_{BE}$, it showed that double counting was almost avoided when internal information was considered for the 50 external sires. A total of 96.4% of contributions due to records of internal information on average was removed from the joint information (Table VI-4). The remaining 3.6% of contributions due to records of internal information was double

counted by the Bayesian evaluations and may result from the estimation of contributions due to relationships and/or from the estimation of contributions due to records among joint and internal information.

Because double counting of contributions due to records between joint and internal information was almost avoided, breeding values that were estimated by $EVAL_{BJ-I}$ for all animals led to better predictions of EBV_J than $EVAL_{BJ}$, based on r, MSE, a and R^2 (Table VI-4). Rank correlations of EBV_J with EBV_{BJ} and EBV_{BJ-I} increased from 0.979 for $EVAL_{BJ}$ to 0.996 for $EVAL_{BJ-I}$ for the internal animals and from 0.956 for $EVAL_{BJ}$ to 0.996 for $EVAL_{BJ-I}$ for the 50 external animals. The MSE decreased on average from 34.3% for $EVAL_{BJ}$ to 6.8% for $EVAL_{BJ-I}$ for the internal animals and from 17.2% for $EVAL_{BJ}$ to 0.6% for $EVAL_{BJ-I}$ for the external animals. These results again showed that integration of external/joint information for the 50 external sires influenced the prediction of internal relatives through the propagation of information from the external sires to relatives. These results show that the double counting of contributions due to records also affected predictions of internal animals. Furthermore, as expected, $EVAL_{BE}$ predicted EBV_J slightly better than $EVAL_{BJ-I}$ for both external sires and internal animals, based on the corresponding r, MSE, a and R^2 (Table VI-4). The low difference in accuracy of prediction between $EVAL_{BE}$ and $EVAL_{BJ-I}$ could be attributed to the estimation of contributions due to relationships and due to records.

Based on these results, double counting of contributions due to records was almost avoided. Thus, the integration of information into a genetic evaluation by avoiding both contributions due to relationships and due to records performed well for external animals. Internal animals also benefited of the integration of information thanks to their relationships with external animals.

WALLOON EXAMPLE

Of the 12 046 animals associated with available Walloon information for the three traits, 6232 animals for milk yield, 6209 animals for fat yield, and 6212 animals for protein yield were associated with information that was based only on contributions due to relationships, as estimated by the TSA. In terms of RE, contributions due to relationships represented from 14.9% for fat yield to 16.3% for milk yield of the contributions associated with Walloon information (Figure VI-1). Among the 1981 sires associated with MACE information, two sires were associated with information that includes only contributions due to relationships for the three traits. Both these sires had

several sons among all the sires associated with an EBV_{MACE}, which explains that the contributions were considered as only due to relationships. In terms of RE, all contributions due to relationships represented on average 5.1% of the contributions associated with MACE information for the three traits. Of the 601 sires with an EBV_{Wc}, all sires were associated with information that included both contributions due to relationships and due to records. This latter observation for the 601 sires was expected because these 601 sires must have at least 10 daughters with records within 10 herds in the Walloon region to participate in the MACE evaluation.

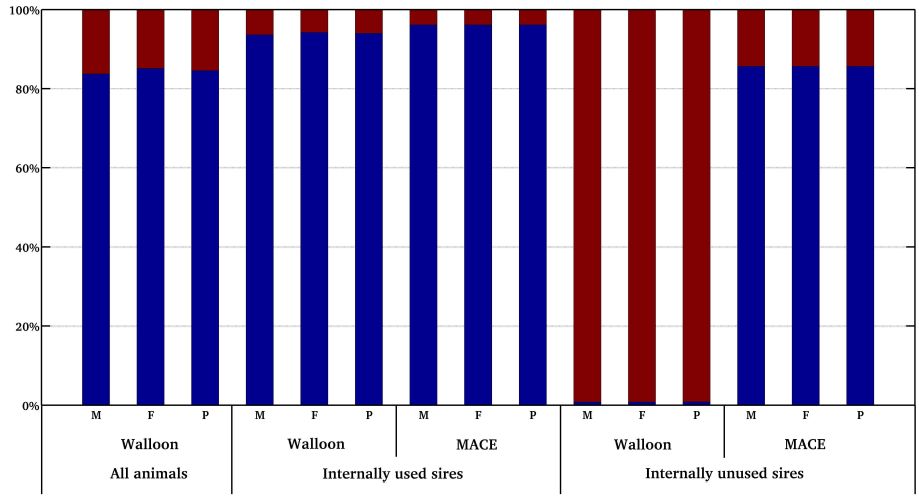

Figure VI-1. Percentage of contributions due to records and due to relationships for the Walloon example. Percentage of contributions due to records (blue squares) and due to relationships (red squares) associated with Walloon information for all animals, internally used and unused sires and associated with MACE information for internally used and unused sires for milk (M), fat yield (F) and protein (P) yields.

INTERNALLY USED SIRES

Of the internally used sires, 1212 had Walloon and MACE information and had both internal and external daughters with records. On average, each sire had 143.1 internal daughters with records. The average REL_W ranged from 0.74 to 0.76 (Table VI-5) and the average REL_{MACE} was equal to 0.88 for the three traits. Results for r, MSE, a and R^2 for prediction of EBV_{MACE} by $EVAL_{BL}$ are in Table VI-6 for the 1212 sires for milk, fat and protein yields. For the three traits, blending of Walloon and MACE information by taking double counting of contributions due to records and due to relationships into account (i.e., $EVAL_{BL}$) led to a ranking that was more similar to the MACE ranking than to the internal ranking (i.e., $EVAL_W$), although these internally used sires sired a large

number of cows with records in the Walloon region. Rank correlations increased by 0.104 points for milk yield to 0.125 points for fat yield to achieve a rank correlation between EBV_{MACE} and EBV_{BL} that ranged from 0.987 to 0.990 (Table VI-6). The MSE, a and R^2 showed that accuracy of predictions of EBV_{MACE} by EBV_W or by EBV_{BL} increased when external information was integrated. Integration of MACE information also increased the average REL by 0.14 points for fat yield to 0.16 points for milk yield (Table VI-5). This increase of average REL corresponded to an increase of 57.5, 51.4, and 50.9 DE per sire on average for milk, fat and protein yields, respectively. Also, the average REL_{BL} for the 1212 sires was 0.02 points higher than the average REL_{MACE} (Table VI-6). This difference in average REL, as well as the differences between EBV_{MACE} and EBV_{BL} based on MSE, a and R^2 (Table VI-6), can be explained by the fact that MACE did not include all information available for animals in the Walloon Region. Indeed, EBV_W of a sire was included into MACE if it had at least 10 daughters with records within 10 herds at the internal level. Therefore, EBV_W for sires that did not fulfill this requirement were not considered by MACE, but were taken into account by the four Bayesian evaluations, which provided additional information compared to MACE information. Approximations based on estimation of contributions due to relationships and theoretical assumptions of the model may also explain some of the differences between EBV_{MACE} and EBV_{BL}. For example, MACE was considered as a national genetic evaluation. These results indicate that $EVAL_{BL}$, i.e. a Bayesian evaluation that blended internal information and external information and avoided most double counting of contributions due to records and due to relationships, was successful in integrating MACE information for internally used sires.

Table VI-5. Average reliabilities (REL; SD in parentheses) associated with Walloon estimated breeding values for internally used and unused sires

Considered animals	Milk yield	Fat yield	Protein yield
Internally used sires	0.74 (0.22)	0.76 (0.21)	0.75 (0.22)
Internally unused sires	0.22 (0.10)	0.23 (0.10)	0.22 (0.10)

Double counting of contributions due to records and due to relationships were also not considered (i.e. $EVAL_{BLNN}$) or were considered separately (i.e. $EVAL_{BLRE}$ and $EVAL_{BLR}$) to study their influences on prediction of $EVAL_{MACE}$ for internally used sires. Parameters r, a and R^2 associated with $EVAL_{BLNN}$, $EVAL_{BLRE}$ and $EVAL_{BLR}$ for the 1212 sires were similar to the r, a and R^2 of $EVAL_{BL}$, although a slight advantage was observed for $EVAL_{BL}$. Therefore, the four blending evaluations led to similar rankings as MACE

for the 1212 internally used sires (i.e., rank correlations equal to 0.99 on average; Table VI-6).

Table VI-6. Parameters obtained for the Walloon example for 1212 internally used sires

Genetic evaluations	Milk yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.886	100.00	0.87 (0.013)	0.78	21 934.6	0.74 (0.22)
$EVAL_{BLNN}$	0.987	11.68	0.993 (0.005)	0.97	55 038.2	0.92 (0.05)
$EVAL_{BLRE}$	0.989	10.01	0.984 (0.004)	0.98	37 487.1	0.91 (0.05)
$EVAL_{BLR}$	0.988	10.57	1.004 (0.004)	0.98	52 313.0	0.91 (0.06)
$EVAL_{BL}$	0.990	8.87	0.995 (0.004)	0.98	34 141.2	0.90 (0.06)
	Fat yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.862	100.00	0.815 (0.014)	0.74	20 016.8	0.76 (0.22)
$EVAL_{BLNN}$	0.983	12.22	0.989 (0.005)	0.97	46 144.6	0.92 (0.05)
$EVAL_{BLRE}$	0.985	10.69	0.977 (0.005)	0.97	32 320.9	0.92 (0.05)
$EVAL_{BLR}$	0.985	11.12	1.004 (0.005)	0.97	43 943.6	0.91 (0.06)
$EVAL_{BL}$	0.987	9.54	0.991 (0.005)	0.97	29 631.1	0.90 (0.06)
	Protein yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.882	100.00	0.851 (0.013)	0.79	20 851.6	0.75 (0.22)
$EVAL_{BLNN}$	0.985	12.38	0.985 (0.005)	0.97	49 589.7	0.92 (0.05)
$EVAL_{BLRE}$	0.987	10.79	0.975 (0.004)	0.98	34 372.9	0.91 (0.05)
$EVAL_{BLR}$	0.986	11.26	0.996 (0.005)	0.98	47 189.5	0.91 (0.06)
$EVAL_{BL}$	0.988	9.56	0.986 (0.004)	0.98	31 434.7	0.90 (0.06)

[1]r = rank correlation between $EVAL_{MACE}$ and $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ or $EVAL_{BL}$.
[2]MSE = mean squared error expressed as a percentage of the average internal mean squared error.
[3]a = regression coefficient (SE in parentheses) and R^2 = coefficient of determination of the regression of MACE EBV on EBV estimated by $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ or $EVAL_{BL}$.
[4]RE_{tot} = total amount of record equivalents; [5]REL = average reliability (SD in parentheses).

However, double counting can be observed based on MSE, RE_{tot} and REL (Table VI-6). With regard to double counting of contributions due to relationships for the 1212 internally used sires, RE that were free of contributions due to relationships (i.e. RE that included only contributions due to records) for EBV_{MACE} were equal to 30 378 (DE = 176 578) for milk yield, 23 927 (DE = 150 772) for fat yield, and 26 338 (DE = 160 416) for protein yield. These amounts of RE free of contributions due to relationships represented 96.1% of the RE that contributed to MACE information. Considering the Walloon information for the 1212 sires, RE that included only contributions due to records represented from 93.6% of all Walloon contributions for milk yield to 94.2% for fat yield. For both Walloon and MACE information associated with the

internally used sires and for the three traits (i.e. for milk, fat and protein yields), less than 6.4% of all contributions were attributed to relationships (Figure VI-1). Such low percentages of contributions due to relationships are in agreement with selection index theory (Van Vleck, 1993). While double counting of contributions due to relationships was present for $EVAL_{BLRE}$ (i.e. the blending evaluation that considered only double counting of contributions due to records), the contributions due to relationships were small and their double counting had little effect on the prediction of EBV_{MACE} for the internally used sires, compared to $EVAL_{BL}$, based on parameters r and MSE. However, as expected, an average increase of 1% in REL_{BLRE} was observed, compared to REL_{BL}. Thus, the REL_{BLRE} were, on average, slightly overestimated.

With regard to double counting of contributions due to records, based on RE, Walloon information represented from 64.3% of the total information free of contributions due to relationships associated with $EVAL_{BL}$ for milk yield to 67.6% for fat yield (Table VI-6). Thus, integrated information free of contributions due to relationships and due to records (i.e. MACE information from which Walloon information was subtracted) represented 32.5% of the total information associated with $EVAL_{BL}$ for fat yield to 35.8% for milk yield. If double counting of contributions due to relationships was considered only, RE_{tot} associated with $EVAL_{BLR}$ ranged from 43 944 RE for fat yield to 52 313 RE for milk yield, while RE_{tot} associated with $EVAL_{BL}$ ranged from 29 631 RE for fat yield to 34 141 RE for milk yield. Thus, between 14 313 and 18 172 RE were considered twice by $EVAL_{BLR}$. However, double counting of contributions due to records affected the prediction of EBV_{MACE} for internally used sires only slightly according to all parameters evaluated (Table VI-5). The REL_{BLR} were overestimated by 1% on average for the internally used sires, compared to REL_{BL}. Furthermore, no preference was observed between $EVAL_{BLRE}$ and $EVAL_{BLR}$ based on r, MSE, a and R^2 for the three traits. Indeed, r and R^2 were similar for these two evaluations, while $EVAL_{BLRE}$ was more reliable based on MSE, but parameter a indicated that $EVAL_{BLR}$ was more reliable. However, $EVAL_{BLRE}$ had the greatest under- and overestimation of true breeding values based on parameter a. Based on these results, it can be stated that double counting of contributions due to relationships and due to records had little effect on EBV for internally used sires.

INTERNALLY UNUSED SIRES

Of the internally unused sires (i.e. that had only external daughters with records), 631 sires were associated with Walloon and MACE information. Their average REL_W

ranged from 0.22 to 0.23 for the three traits (Table VI-7) and the average REL_{MACE} was equal to 0.77. Because they had only external daughters, Walloon contributions only included contributions due to relationships and no contributions due to records. Based on RE_{tot} (Table VI-7), Walloon contributions due to records for all 631 sires were in general well estimated by the TSA, ranging from 0.79% of the Walloon total contributions for milk yield to 0.80% for protein yield (Figure VI-1). The small non-zero percentage could be attributed to approximations involved in estimating the contributions due to relationships and due to records by the TSA, such as the consideration of an unknown fixed effect (Vandenplas and Gengler, 2012). The nearly correct estimation of contributions due to relationships led to similar average REL_{MACE} and average REL_{BL} for the three traits (Table VI-7). Integration of MACE information also increased the average REL_{W} by at least 0.54 points, resulting in an average REL_{BL} equal to 0.77 for the three traits. These results for the 631 internally unused sires confirmed that MACE information already contained the main contributions due to relationships that were expressed in the Walloon information and that double counting of contributions due to relationships was mostly avoided. Not considering contributions due to relationships (i.e., $EVAL_{BLNN}$ and $EVAL_{BLRE}$) led to overestimation of average REL by at least 3% (Table VI-7).

Results for r, MSE, a and R^2 for the prediction of EBV_{MACE} by the four blending evaluations are in Table VI-7 for the 631 internally unused sires for the three traits. Blending of Walloon and MACE information led to similar rankings of the 631 sires for the four blending evaluations. Rank correlations between EBV_{MACE} and EBV for the four blending evaluations increased from 0.73 to 0.99 for milk yield, from 0.57 to 0.99 for fat yield and from 0.72 to 0.99 for protein yield. These rank correlations indicated that the blending method was also successful for sires with only external information for all three traits. These results were confirmed by a decrease of MSE by at least 96.9% and by regression coefficients close to 1.0, with an R^2 equal to 0.99 for all three traits (Table VI-7). Because double counting can be only attributed to contributions due to relationships for the 631 internally unused sires, $EVAL_{BLNN}$ and $EVAL_{BLRE}$ led to similar parameters. This was also observed for $EVAL_{BL}$ and $EVAL_{BLR}$ (Table VI-7). Differences between these two groups of evaluations were only observed based on MSE and a (Table VI-7). These two parameters showed that EBV_{MACE} for the 631 sires were slightly better predicted when contributions due to relationships were considered. However, all these results showed that contributions due to relationships had little effect on the prediction of EBV_{MACE}.

Table VI-7. Parameters obtained for the Walloon example for 631 internally unused sires

Genetic evaluations	Milk yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.725	100.00	0.667 (0.024)	0.56	2.5	0.22 (0.10)
$EVAL_{BLNN}$	0.994	3.09	0.953 (0.004)	0.99	4021.7	0.81 (0.05)
$EVAL_{BLRE}$	0.994	3.06	0.952 (0.004)	0.99	4021.7	0.81 (0.05)
$EVAL_{BLR}$	0.994	2.68	0.978 (0.004)	0.99	3172.9	0.77 (0.06)
$EVAL_{BL}$	0.994	2.68	0.977 (0.004)	0.99	3172.9	0.77 (0.06)
	Fat yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.571	100.00	0.506 (0.024)	0.40	2.0	0.23 (0.10)
$EVAL_{BLNN}$	0.992	2.28	0.95 (0.005)	0.99	3172.5	0.81 (0.05)
$EVAL_{BLRE}$	0.992	2.28	0.949 (0.005)	0.99	3172.5	0.81 (0.05)
$EVAL_{BLR}$	0.992	2.09	0.987 (0.005)	0.99	2499.1	0.77 (0.06)
$EVAL_{BL}$	0.992	2.08	0.986 (0.005)	0.99	2499.1	0.77 (0.06)
	Protein yield					
	r[1]	MSE[2]	a[3]	R^2 [3]	RE_{tot} [4]	REL[5]
$EVAL_W$	0.717	100.00	0.684 (0.025)	0.54	2.3	0.22 (0.10)
$EVAL_{BLNN}$	0.993	2.96	0.952 (0.004)	0.99	3490.3	0.81 (0.05)
$EVAL_{BLRE}$	0.993	2.95	0.951 (0.004)	0.99	3490.3	0.81 (0.05)
$EVAL_{BLR}$	0.993	2.75	0.978 (0.005)	0.99	2751.0	0.78 (0.06)
$EVAL_{BL}$	0.993	2.75	0.977 (0.005)	0.99	2751.0	0.77 (0.06)

[1]r = rank correlation between $EVAL_{MACE}$ and $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ or $EVAL_{BL}$.
[2]MSE = mean squared error expressed as a percentage of the average internal mean squared error.
[3]a = regression coefficient (SE in parentheses) and R^2 = coefficient of determination of the regression of MACE EBV on EBV estimated by $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$, $EVAL_{BLR}$ or $EVAL_{BL}$.
[4]RE_{tot} = total amount of record equivalents; [5]REL = average reliability (SD in parentheses).

VanRaden and Tooker (2012) found similar correlations between EBV_{MACE} and combined EBV for sires with only external daughters (i.e., between 0.991 and 0.994 for yield traits). Their strategy consisted of computing external deregressed proofs (DRP) from EBV_{MACE} and including one extra record based on these DRP, weighted by the associated DE for the sire. Internal contributions in MACE information for sires with internal and external daughters were considered by subtracting the number of internal DE from the total and by using internal EBV instead of parent averages from EBV_{MACE} to compute external DRP. Based on Legarra et al. (2007), Gengler and Vanderick (2008) integrated MACE information into the official Walloon genetic evaluation for milk production. External EBV were estimated by selection index theory and internal contributions were considered as in VanRaden and Tooker (2012). Thus, while these two latter approaches and the approach proposed in this study consider internal contributions

to MACE information in a similar manner [See Additional file 2], the main advantage of the proposed approach is to avoid a pre-processing deregression step or computation of external EBV.

INTERNAL ANIMALS

The effect of the integration of MACE information on predictions was also studied for internal animals that were not associated with MACE information and that were sired by internally used sires. A total of 3331 internal animals was considered. If double counting of contributions due to relationships and due to records were avoided (i.e., $EVAL_{BL}$), integration of MACE information led to an increase of the REL from 0.32 to 0.42 for milk yield and from 0.31 to 0.42 for fat and protein yields for internal animals that had a REL_W less than 0.50 (Table VI-8). These increases were equivalent to 2.4 DE for milk yield, 2.3 DE for fat yield and 2.4 DE for protein yield. On average, no increase in REL was observed for internal animals with REL_W greater than 0.50 (Table VI-9 and Table VI-10; Figure VI-2). Therefore, integration of MACE information was mostly relevant for external animals that were associated with this information and for internal animals with a low REL_W sired by external animals.

Table VI-8. Parameters for internal animals with a Walloon reliability less than 0.50 and sired by internally used sires

Traits	N[1]	Parameters[2]	Genetic evaluation				
			$EVAL_W$	$EVAL_{BLNN}$	$EVAL_{BLRE}$	$EVAL_{BLR}$	$EVAL_{BL}$
Milk yield	1948	r	0.944	0.995	0.995	0.999	1.000
		RE_{tot}	245.1	1655.2	1655.2	245.1	245.1
		REL	0.32 (0.10)	0.57 (0.06)	0.56 (0.06)	0.43 (0.07)	0.42 (0.07)
Fat yield	1694	r	0.923	0.994	0.994	0.999	1.000
		RE_{tot}	102.6	1254.9	1254.9	102.6	102.6
		REL	0.31 (0.09)	0.56 (0.06)	0.56 (0.06)	0.42 (0.08)	0.42 (0.08)
Protein yield	1786	r	0.938	0.995	0.995	0.999	1.000
		RE_{tot}	148.4	1243.5	1243.5	148.4	148.4
		REL	0.31 (0.09)	0.56 (0.06)	0.56 (0.06)	0.42 (0.08)	0.42 (0.08)

[1]N = Number of internal animals.
[2]r = rank correlation between $EVAL_{BL}$ and $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$ or $EVAL_{BLR}$; RE_{tot} = Total amount of record equivalents; REL = average reliability (SD in parentheses).

Table VI-9. Parameters for internal animals with a Walloon reliability between 0.50 and 0.74 and sired by internally used sires

Traits	N[1]	Parameters[2]	Genetic evaluation				
			$EVAL_W$	$EVAL_{BLNN}$	$EVAL_{BLRE}$	$EVAL_{BLR}$	$EVAL_{BL}$
Milk yield	1360	r	0.999	>0.999	>0.999	>0.999	1.000
		RE_{tot}	1205.7	2759.1	2759.1	1205.7	1205.7
		REL	0.55 (0.04)	0.67 (0.02)	0.67 (0.03)	0.55 (0.03)	0.55 (0.03)
Fat yield	1607	r	0.999	>0.999	>0.999	>0.999	1.000
		RE_{tot}	1322.0	3125.6	3125.6	1322.0	1322.0
		REL	0.57 (0.04)	0.68 (0.03)	0.68 (0.03)	0.57 (0.04)	0.57 (0.04)
Protein yield	1516	r	0.999	>0.999	>0.999	>0.999	1.000
		RE_{tot}	1252.0	2787.7	2787.7	1252.0	1252.0
		REL	0.56 (0.04)	0.68 (0.03)	0.68 (0.03)	0.56 (0.04)	0.56 (0.04)

[1]N = Number of internal animals.
[2]r = rank correlation between $EVAL_{BL}$ and $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$ or $EVAL_{BLR}$; RE_{tot} = Total amount of record equivalents; REL = average reliability (SD in parentheses).

Table VI-10. Parameters for internal animals with a Walloon reliability greater than 0.74 and sired by internally used sires

Traits	N[1]	Parameters[2]	Genetic evaluation				
			$EVAL_W$	$EVAL_{BLNN}$	$EVAL_{BLRE}$	$EVAL_{BLR}$	$EVAL_{BL}$
Milk yield	23	r	0.998	0.999	1.000	0.999	1.000
		RE_{tot}	132.6	156.9	156.9	132.6	132.6
		REL	0.80 (0.04)	0.82 (0.03)	0.82 (0.03)	0.80 (0.04)	0.80 (0.04)
Fat yield	30	r	0.999	1.000	>0.999	1.000	1.000
		RE_{tot}	158.8	190.6	190.6	158.8	158.8
		REL	0.81 (0.04)	0.83 (0.03)	0.83 (0.03)	0.81 (0.04)	0.81 (0.04)
Protein yield	29	r	0.999	>0.999	>0.999	1.000	1.000
		RE_{tot}	147.7	174.6	174.6	147.7	147.7
		REL	0.80 (0.04)	0.83 (0.03)	0.83 (0.03)	0.80 (0.04)	0.80 (0.04)

[1]N = Number of internal animals.
[2]r = rank correlation between $EVAL_{BL}$ and $EVAL_W$, $EVAL_{BLNN}$, $EVAL_{BLRE}$ or $EVAL_{BLR}$; RE_{tot} = Total amount of record equivalents; REL = average reliability (SD in parentheses).

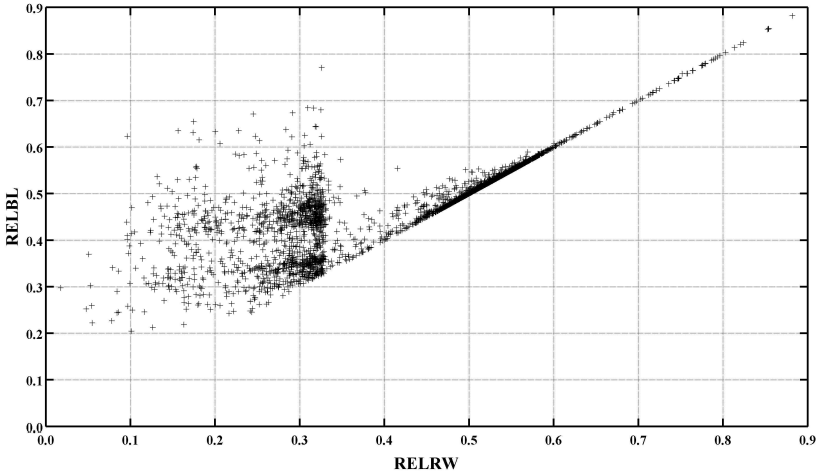

Figure VI-2. Reliabilities for internal progeny. Reliabilities associated with the Bayesian evaluation that considers double counting of contributions due to relationships and due to records (REL_{BL}) as a function of reliabilities associated with the official Walloon evaluation (REL_{RW}) for the 3331 internal animals sired by internally used sires (i.e., having daughters with records in the Walloon Region) for milk yield.

The effect of double counting was also studied in comparison to $EVAL_{BL}$ for the 3331 internal animals that were only associated with Walloon information and that were sired by internally used sires. Own contributions due to relationships for internal animals with REL_W less than 0.50 represented from 85.2% of the total contributions for milk yield to 91.8% for fat yield (Table VI-8). These percentages ranged from 55.1% for protein yield to 57.7% for fat yield for internal animals with REL_W between 0.50 and 0.75, and from 15.4% for protein yield to 16.7% for fat yield for internal animals with REL_W greater than 0.75 (Table VI-9, and Table VI-10). As stated before, these observations were as expected based on selection index theory (Van Vleck, 1993), and double counting of own contributions due to relationships was mostly present for internal animals with low REL_W. However, internal animals were also affected by double counting of contributions due to relationships and due to records that originated from their sires (and relatives) through the contributions due to relationships. Double counting that originated from their own contributions and from their sires (and relatives) could be observed based on a comparison of REL_{BLRE}, REL_{BLR} and REL_{BL} and of r between EBV_{BL} and EBV_{BLRE} or EBV_{BLR} (Table VI-8, Table VI-9, and Table VI-10). Double counting of contributions due to records that originated from sires of internal animals had minor effects on the average REL_{BLR} associated with internal animals (at most 1%) and rankings of internal animals ($r \geq 0.999$; Table VI-8, Table VI-9, and Table VI-10). However, double counting of

contributions due to relationships led to an increase of average REL by at least 0.14 points for internal animals with REL_W less than 0.50 and by at least 0.11 points for internal animals with REL_W ranging from 0.50 to 0.74. The increase of average REL was lower for internal animals with REL_W greater than 0.75 (>0.02 points; Table VI-8, Table VI-9, and Table VI-10). Although the average REL_{BLR} and REL_{BLRE} were (slightly) overestimated for both evaluations, double counting of contributions due to records and due to relationships had little effect on the ranking of internal animals compared to the ranking of $EVAL_{BL}$, regardless of the group of internal animals or trait considered. Indeed, rank correlations between $EVAL_{BL}$ and $EVAL_{BLR}$ or $EVAL_{BLRE}$ were greater than 0.99 (Table VI-8, Table VI-9, and Table VI-10). All these results show that double counting of contributions due to relationships and due to records can be ignored for the prediction of EBV for internal animals that are sired by external animals. However, all double counting must be taken into account to estimate REL accurately.

ON THE IMPLEMENTATION

Considering all groups of animals, i.e., internally used and unused sires, as well as internal animals sired by internally used sires, our results for the Walloon example suggest that contributions due to relationships can be ignored. Indeed, the different rank correlations for $EVAL_{BLRE}$ (i.e., the Bayesian evaluation that took only double counting of contributions due to records into account) were similar to the rank correlations of $EVAL_{BL}$. Furthermore, in practice, the TSA could be difficult to apply if a high number of animals is associated with external information because it requires the inversion of a, potentially, dense matrix for each iteration. However, effects of double counting of contributions due to relationships should be tested before ignoring it. For example, overestimation of REL could occur especially for traits for which contributions due to relationships would be at least as significant as contributions due to records (e.g., if the phenotypes are expensive to obtain). Furthermore, REL associated with the modified MME were estimated based on the inverted LHS. Although this was feasible for the simulated and Walloon data, this may not be feasible in most cases, and approaches that estimate REL (e.g., Misztal and Wiggans, 1988; VanRaden and Wiggans, 1991) could be modified to take into account RE (or DE) associated with external information.

The Walloon example was considered as an evaluation that blends MACE and Walloon (internal) information in the context of official Walloon genetic evaluations for Holstein cattle. However, the Walloon example can also be considered as a particular case

of an internal evaluation that has no internal data and blends only sources of external information, i.e., MACE and Walloon information, that are partially based on the same information, i.e., the Walloon information. This case can be extended to more general cases for which internal data may exist and external animals are associated with at least two sources of information (e.g., E_1 and E_2) that are partially based on the same external records or information. Double counting of external information that is shared by the sources of external information, e.g., E_1 and E_2, can be avoided by the proposed approach thanks to the knowledge and availability of EBV and associated REL that are based only on external information that is shared by the sources of external information. Nevertheless, although taking external information that is shared by different sources of external information into consideration seems to be possible with the proposed approach, this may be difficult in practice because it requires that EBV and associated REL based on shared external information are known and available.

CONCLUSIONS

The proposed unified method integrated and blended several sources of information into an internal genetic evaluation in an appropriate manner. The results also showed that the proposed method was able to avoid double counting of contributions due to records and due to relationships. Furthermore, because all available external sources of information were correctly propagated, relatives of external animals benefited from integrated information and, therefore, received more reliable EBV. The unified method could also be used in the context of single-step genomic evaluations to integrate external information to indirectly recover a large amount of external phenotypic information (Colinet et al., 2013). While the simulated and Walloon examples were univariate, the unified method was developed for multi-trait models that, e.g., allow evaluation of only internally available traits (e.g., methane emissions, fine milk composition traits, such as fatty acids, milk proteins and other minor components), using additional external information from correlated traits (e.g., traits evaluated by Interbull).

ADDITIONAL FILES

Additional file 1: Integration of two sources of external information into a genetic evaluation. This file describes a derivation that integrates two sources of external information into a genetic evaluation, based on a Bayesian view of the mixed models (Sorensen and Gianola, 2002) and similar to the Bayesian derivation of Legarra et al.

(2007) that integrates one source of external information into a genetic evaluation.

Additional file 2: Double counting between internal and external information. This file describes the development to avoid double counting of contributions due to records between internal and external information.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

JV developed the algorithms and the equations, conceived the experimental design, ran the tests and wrote the first draft. FC prepared data for the Walloon example. NG initiated and directed the research. All authors participated in writing the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

J Vandenplas, as a research fellow, and N Gengler, as a former senior research associate, acknowledge the support of the National Fund for Scientific Research (Brussels, Belgium) for these positions. Additional financial support was provided by the Ministry of Agriculture of the Walloon region of Belgium (Service Public de Wallonie, Direction générale opérationnelle "Agriculture, Ressources Naturelles et Environnement" – DGARNE) through research projects D31-1207, D31-1224/S1, and D31-1274, D31-1304 and D31-1308. Financial support for scientific visits was also provided by Wallonie Brussels International. J Vandenplas acknowledges I Misztal for hosting him at the Animal and Dairy Sciences Department of University of Georgia, G Gorjanc for hosting him at the Animal Science Department of University of Ljubljana, P Faux for helpful discussions and editing help, S Tsuruta and I Aguilar for their help concerning the BLUPF90 programs. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the National Fund for Scientific Research (Brussels, Belgium) under Grant No. 2.5020.11. The authors thank the two anonymous reviewers for the useful comments.

REFERENCES

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic

information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Auvray, B., and N. Gengler. 2002. Feasibility of a Walloon test-day model and study of its potential as tool for selection and management. *Interbull Bull.* 29:123–127.

Bonaiti, B., and D. Boichard. 1995. Accounting for foreign information in genetic evaluation. *Interbull Bull.* 11:4pp.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.

Colinet, F.G., J. Vandenplas, P. Faux, S. Vanderick, R. Renaville, C. Bertozzi, X. Hubin, and N. Gengler. 2013. Walloon single-step genomic evaluation system integrating local and MACE EBV. *Interbull Bull.* 47:203–210.

Croquet, C., P. Mayeres, A. Gillon, S. Vanderick, and N. Gengler. 2006. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89:2257–2267.

Eaton, J.W., D. Bateman, S. Hauberg, and R. Wehbring. 2011. GNU Octave. A high-level interactive language for numerical computations. 3rd ed. Free Software Foundation, Inc., Boston, MA.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Gianola, D., and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63:217–244.

Henderson, C.R. 1984. Applications of linear models in animal breeding. 2nd ed. University of Guelph, Guelph, ON, Canada.

Interbull. 2013. Interbull routine genetic evaluation for dairy production traits. April 2013. Accessed Apr. 19, 2013. http://www.interbull.org/web/static/mace_evaluations_archive/eval/prod-apr13.html.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

Mäntysaari, E.A., and I. Strandén. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluations. CD–ROM Comm. 353 in Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Misztal, I. 2013. BLUPF90 family of programs. Accessed Jul. 3, 2012. http://nce.ads.uga.edu/wiki/doku.php.

Misztal, I., and G.R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71(Suppl. 2):27–32.

Přibyl, J., P. Madsen, J. Bauer, J. Přibylová, M. Šimečková, L. Vostrý, and L. Zavadilová. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *J. Dairy Sci.* 96:1865–1873.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Quaas, R.L., and Z.W. Zhang. 2001. Incorporating external information in multibreed genetic evaluation. *J. Anim. Sci.* 79(Suppl. 1):342. (Abstr.).

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer-Verlag, New York, NY.

Täubert, H., H. Simianer, and K. Karras. 2000. Blending Interbull breeding values into national evaluations A new approach. *Interbull Bull.* 25:53–56.

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

VanRaden, P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. *J. Dairy Sci.* 84(E. Suppl.):E47–E55.

VanRaden, P.M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. *Interbull Bull.* 45:29–33.

VanRaden, P.M., and M.E. Tooker. 2012. Methods to include foreign information in national evaluations. *J. Dairy Sci.* 95:446.

VanRaden, P.M., and G.R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737–2746.

Van Vleck, L.D. 1993. Selection Index and introduction to mixed model methods. CRC Press. Boca Raton, Fl.

Van Vleck, L.D. 1994. Algorithms for simulation of animal models with multiple traits and with maternal and non-additive genetic effects. *Braz. J. Genet.* 17:53–57.

ADDITIONAL FILE 1

INTEGRATION OF TWO SOURCES OF EXTERNAL INFORMATION INTO A GENETIC EVALUATION

The following Bayesian derivation is similar to the Bayesian derivation of Legarra et al. (2007) that integrates one source of external information into an internal genetic evaluation in the context of multi-breed genetic evaluations for beef cattle.

Assume a set of animals partitioned in four groups. The first group (i.e., internal animals $A_{1,2}^0$) has only records in the internal data set ($\mathbf{y}_{E_0}$). The second group (i.e., external animals A_1) has records in the external data set, $\mathbf{y}_{E_1}$, and may have records in $\mathbf{y}_{E_0}$. The third group (i.e., external animals A_2) has records in the external data set, $\mathbf{y}_{E_2}$, and may have records in $\mathbf{y}_{E_0}$. The fourth group (i.e., external animals $A_{1,2}$) have records in both $\mathbf{y}_{E_1}$ and $\mathbf{y}_{E_2}$, and may have also records in $\mathbf{y}_{E_0}$. For the following genetic evaluations, variance components are assumed to be identical.

Concerning the notation of matrices in the following development (e.g., $\mathbf{X}_{E_i(A_l)}$), the subscript E_i refers to the *i*th source of data and the subscript within brackets (A_l) refers to the l^{th} group of animals, respectively.

Assume a hypothetical joint genetic evaluation (denoted by the subscript J) of all animals ($A_{1,2}^0$, A_1, A_2, $A_{1,2}$) including both datasets $\mathbf{y}_{E_1}$ and $\mathbf{y}_{E_2}$. Because it was assumed that $\mathbf{y}_{E_1}$ and $\mathbf{y}_{E_2}$ were pre-corrected for fixed effects, the model partitioned among the four groups of animals can be written as:

$$\begin{bmatrix} \mathbf{y}_{E_1(A_1)} \\ \mathbf{y}_{E_1(A_{1,2})} \\ \mathbf{y}_{E_2(A_2)} \\ \mathbf{y}_{E_2(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_{E_1(A_1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_{E_1(A_{1,2})} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_{E_2(A_2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_{E_2(A_{1,2})} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{J(A_{1,2}^0)} \\ \mathbf{u}_{J(A_1)} \\ \mathbf{u}_{J(A_2)} \\ \mathbf{u}_{J(A_{1,2})} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{J_1(A_1)} \\ \mathbf{e}_{J_1(A_{1,2})} \\ \mathbf{e}_{J_2(A_2)} \\ \mathbf{e}_{J_2(A_{1,2})} \end{bmatrix}$$

where $\mathbf{u}_J = \begin{bmatrix} \mathbf{u}'_{J(A_{1,2}^0)} & \mathbf{u}'_{J(A_1)} & \mathbf{u}'_{J(A_2)} & \mathbf{u}'_{J(A_{1,2})} \end{bmatrix}'$ is the vector of genetic random effects for animals $A_{1,2}^0$, A_1, A_2 and $A_{1,2}$ for the evaluation J, $\mathbf{Z}_{E_1(A_1)}$ and $\mathbf{Z}_{E_1(A_{1,2})}$ are incidence matrices relating records of $\mathbf{y}_{E_1(A_1)}$ and $\mathbf{y}_{E_1(A_{1,2})}$ to $\mathbf{u}_{J(A_1)}$ and $\mathbf{u}_{J(A_{1,2})}$, respectively, and $\mathbf{Z}_{E_2(A_2)}$ and $\mathbf{Z}_{E_2(A_{1,2})}$ are incidence matrices relating records of $\mathbf{y}_{E_2(A_2)}$ and $\mathbf{y}_{E_2(A_{1,2})}$ to

$\mathbf{u}_{J(A_2)}$ and $\mathbf{u}_{J(A_{1,2})}$, respectively, and $\mathbf{e}_{J_1}$ and $\mathbf{e}_{J_2}$ are the vectors of residuals associated with $\mathbf{y}_{E_1}$ and $\mathbf{y}_{E_2}$, respectively.

The corresponding mixed model equations (MME) can be written as:

$$\begin{bmatrix}
\mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_1)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_2)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2})} \\
\mathbf{G}_{E_0}^{(A_1 A_{1,2}^0)} & \mathbf{Z'}_{E_1(A_1)}\mathbf{R}_{E_1}^{(A_1 A_1)}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}_{E_0}^{(A_1 A_1)} & \mathbf{G}_{E_0}^{(A_1 A_2)} & \mathbf{Z'}_{E_1(A_1)}\mathbf{R}_{E_1}^{(A_1 A_{1,2})}\mathbf{Z}_{E_1(A_{1,2})} + \mathbf{G}_{E_0}^{(A_1 A_{1,2})} \\
\mathbf{G}_{E_0}^{(A_2 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_2 A_1)} & \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2 A_2)}\mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_0}^{(A_2 A_2)} & \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2 A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} + \mathbf{G}_{E_0}^{(A_2 A_{1,2})} \\
\mathbf{G}_{E_0}^{(A_{1,2} A_{1,2}^0)} & \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}_{E_1}^{(A_{1,2} A_1)}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}_{E_0}^{(A_{1,2} A_1)} & \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2} A_2)}\mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_0}^{(A_{1,2} A_2)} & \begin{matrix}\mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}_{E_1}^{(A_{1,2} A_{1,2})}\mathbf{Z}_{E_1(A_{1,2})} + \\ \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2} A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} + \\ \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2})}\end{matrix}
\end{bmatrix}$$

$$\begin{bmatrix}\hat{\mathbf{u}}_{J(A_{1,2}^0)} \\ \hat{\mathbf{u}}_{J(A_1)} \\ \hat{\mathbf{u}}_{J(A_2)} \\ \hat{\mathbf{u}}_{J(A_{1,2})}\end{bmatrix} = \begin{bmatrix}\mathbf{0} \\ \mathbf{Z'}_{E_1(A_1)}\mathbf{R}_{E_1}^{(A_1 A_1)}\mathbf{y}_{E_1(A_1)} \\ \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2 A_2)}\mathbf{y}_{E_2(A_2)} \\ \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}_{E_1}^{(A_{1,2} A_{1,2})}\mathbf{y}_{E_1(A_{1,2})} + \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2} A_{1,2})}\mathbf{y}_{E_2(A_{1,2})}\end{bmatrix}$$

(equation 1.1)

where $\mathbf{G}_{E_0}^{-1} = \begin{bmatrix} \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_1)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_2)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_1 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_1 A_1)} & \mathbf{G}_{E_0}^{(A_1 A_2)} & \mathbf{G}_{E_0}^{(A_1 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_2 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_2 A_1)} & \mathbf{G}_{E_0}^{(A_2 A_2)} & \mathbf{G}_{E_0}^{(A_2 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_{1,2} A_1)} & \mathbf{G}_{E_0}^{(A_{1,2} A_2)} & \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2})} \end{bmatrix}$ is the inverse of the additive (co)variance matrix including all animals and $\begin{bmatrix} \mathbf{R}_{E_1}^{(A_1 A_1)} & \mathbf{R}_{E_1}^{(A_1 A_{1,2})} \\ \mathbf{R}_{E_1}^{(A_{1,2} A_1)} & \mathbf{R}_{E_1}^{(A_{1,2} A_{1,2})} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{R}_{E_2}^{(A_2 A_2)} & \mathbf{R}_{E_2}^{(A_2 A_{1,2})} \\ \mathbf{R}_{E_2}^{(A_{1,2} A_2)} & \mathbf{R}_{E_2}^{(A_{1,2} A_{1,2})} \end{bmatrix}$ are the inverse of the residual (co)variance matrices associated with $\mathbf{e}_{J_1}$ and $\mathbf{e}_{J_2}$, respectively.

Assume an internal genetic evaluation (denoted by the subscript E_0) of all animals (i.e., animals $A_{1,2}^0$, A_1, A_2 and $A_{1,2}$) including only $\mathbf{y}_{E_0}$ and using the prior distribution $p\left(\hat{\mathbf{u}}_{E_0} \middle| \mathbf{y}_{E_1}, \mathbf{y}_{E_2}\right) = MVN(\boldsymbol{\mu}, \mathbf{G}^*)$ (Sorensen and Gianola, 2002) where $\mathbf{G}^*$ is the inverse of the left hand side (LHS) of the equation 1.1 and $\boldsymbol{\mu}$ is the solutions of the equation 1.1.

The model for the genetic evaluation E_0 can be written as:

$$\mathbf{y}_{E_0} = \mathbf{X}_{E_0}\boldsymbol{\beta}_{E_0} + \mathbf{Z}_{E_0}\mathbf{u}_{E_0} + \mathbf{e}_{E_0},$$

where $\mathbf{X}_{E_0}$ and $\mathbf{Z}_{E_0}$ are incidence matrices relating records in $\mathbf{y}_{E_0}$ to the vector of fixed effects $\boldsymbol{\beta}_{E_0}$ and the vector of genetic random effects $\mathbf{u}_{E_0} = \left[\mathbf{u}'_{E_0(A^0_{1,2})} \quad \mathbf{u}'_{E_0(A_1)} \quad \mathbf{u}'_{E_0(A_2)} \quad \mathbf{u}'_{E_0(A_{1,2})}\right]'$, respectively and $\mathbf{e}_{E_0}$ is the vector of residuals.

The MME can be written as:

$$\begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{X}_{E_0} & \mathbf{X}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{Z}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{X}_{E_0} & \mathbf{Z}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{Z}_{E_0} + \mathbf{G}^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{E_0} \\ \hat{\mathbf{u}}_{E_0} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{y}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}^{-1}_{E_0}\mathbf{y}_{E_0} + \mathbf{G}^{*-1}\boldsymbol{\mu} \end{bmatrix}, \quad \text{(equation 1.2)}$$

where $\mathbf{R}^{-1}_{E_0}$ is the inverse of the residual (co)variance matrix associated with $\mathbf{e}_{E_0}$.

However, the evaluation J (equation 1.1), and therefore $\mathbf{G}^*$ and $\boldsymbol{\mu}$, are unknown.

Assume that two genetic evaluations (denoted by the subscripts E_1 and E_2, respectively) for two groups of external animals (i.e., animals A_1 and $A_{1,2}$, and animals A_2 and $A_{1,2}$, respectively) which do not include in the genealogy internal animals (i.e., animals $A^0_{1,2}$ and A_2, and animals $A^0_{1,2}$ and A_1, respectively) are known. The model for the genetic evaluation E_1 of only external animals A_1 and $A_{1,2}$ including only $\mathbf{y}_{E_1}$ and which does not include in the genealogy animals $A^0_{1,2}$ and A_2 can be written as:

$$\begin{bmatrix} \mathbf{y}_{E_1(A_1)} \\ \mathbf{y}_{E_1(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{E_1(A_1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{E_1(A_{1,2})} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{E_1(A_1)} \\ \mathbf{u}_{E_1(A_{1,2})} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{E_1(A_1)} \\ \mathbf{e}_{E_1(A_{1,2})} \end{bmatrix},$$

where $\mathbf{u}_{E_1(A_1)}$ and $\mathbf{u}_{E_1(A_{1,2})}$ are the vectors of genetic random effects for animals A_1 and $A_{1,2}$ for the genetic evaluation E_1.

The MME can be written as:

$$\begin{bmatrix} \mathbf{Z}'_{E_1(A_1)}\mathbf{R}^{(A_1A_1)}_{E_1}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}^{*(A_1A_1)}_{E_1} & \mathbf{Z}'_{E_1(A_1)}\mathbf{R}^{(A_1A_{1,2})}_{E_1}\mathbf{Z}_{E_1(A_{1,2})} + \mathbf{G}^{*(A_1A_{1,2})}_{E_1} \\ \mathbf{Z}'_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_1)}_{E_1}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}^{*(A_{1,2}A_1)}_{E_1} & \mathbf{Z}'_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_{1,2})}_{E_1}\mathbf{Z}_{E_1(A_{1,2})} + \mathbf{G}^{*(A_{1,2}A_{1,2})}_{E_1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{E_1(A_1)} \\ \hat{\mathbf{u}}_{E_1(A_{1,2})} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}'_{E_1(A_1)}\mathbf{R}^{(A_1A_1)}_{E_1}\mathbf{y}_{E_1(A_1)} \\ \mathbf{Z}'_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_{1,2})}_{E_1}\mathbf{y}_{E_1(A_{1,2})} \end{bmatrix} = \mathbf{D}^{*-1}_{E_1} \begin{bmatrix} \hat{\mathbf{u}}_{E_1(A_1)} \\ \hat{\mathbf{u}}_{E_1(A_{1,2})} \end{bmatrix}$$

(equation 1.3)

where $\mathbf{G}_{E_1}^{*-1} = \begin{bmatrix} \mathbf{G}_{E_1}^{*(A_1A_1)} & \mathbf{G}_{E_1}^{*(A_1A_{1,2})} \\ \mathbf{G}_{E_1}^{*(A_{1,2}A_1)} & \mathbf{G}_{E_1}^{*(A_{1,2}A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{E_1(A_1A_1)} & \mathbf{G}_{E_1(A_1A_{1,2})} \\ \mathbf{G}_{E_1(A_{1,2}A_1)} & \mathbf{G}_{E_1(A_{1,2}A_{1,2})} \end{bmatrix}^{-1}$ is the inverse of the additive genetic (co)variance matrix for the external genetic evaluation E_1.

Similarly, the MME for the genetic evaluation E_2 of only external animals A_2 and $A_{1,2}$ including only $\mathbf{y}_{E_2}$ and which does not include in genealogy animals $A_{1,2}^0$ and A_1 can be written as:

$$\begin{bmatrix} \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2A_2)}\mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_2}^{*(A_2A_2)} & \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} + \mathbf{G}_{E_2}^{*(A_2A_{1,2})} \\ \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2}A_2)}\mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_2}^{*(A_{1,2}A_2)} & \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2}A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} + \mathbf{G}_{E_2}^{*(A_{1,2}A_{1,2})} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{E_2(A_2)} \\ \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2A_2)}\mathbf{y}_{E_2(A_2)} \\ \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2}A_{1,2})}\mathbf{y}_{E_2(A_{1,2})} \end{bmatrix} = \mathbf{D}_{E_2}^{*-1} \begin{bmatrix} \hat{\mathbf{u}}_{E_2(A_2)} \\ \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix}$$

(equation 1.4)

where $\hat{\mathbf{u}}_{E_2(A_2)}$ and $\hat{\mathbf{u}}_{E_2(A_{1,2})}$ are the vectors of genetic random effects for animals A_2 and $A_{1,2}$ for the genetic evaluation E_2, and $\mathbf{G}_{E_2}^{*-1} = \begin{bmatrix} \mathbf{G}_{E_2}^{*(A_2A_2)} & \mathbf{G}_{E_2}^{*(A_2A_{1,2})} \\ \mathbf{G}_{E_2}^{*(A_{1,2}A_2)} & \mathbf{G}_{E_2}^{*(A_{1,2}A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{E_2(A_2A_2)} & \mathbf{G}_{E_2(A_2A_{1,2})} \\ \mathbf{G}_{E_2(A_{1,2}A_2)} & \mathbf{G}_{E_2(A_{1,2}A_{1,2})} \end{bmatrix}^{-1}$ is the additive genetic (co)variance matrix for the genetic evaluation E_2.

Therefore,

$$\mathbf{D}_{E_1}^{*-1} - \mathbf{G}_{E_1}^{*-1} = \begin{bmatrix} \mathbf{D}_{E_1}^{*(A_1A_1)} & \mathbf{D}_{E_1}^{*(A_1A_{1,2})} \\ \mathbf{D}_{E_1}^{*(A_{1,2}A_1)} & \mathbf{D}_{E_1}^{*(A_{1,2}A_{1,2})} \end{bmatrix} - \begin{bmatrix} \mathbf{G}_{E_1}^{*(A_1A_1)} & \mathbf{G}_{E_1}^{*(A_1A_{1,2})} \\ \mathbf{G}_{E_1}^{*(A_{1,2}A_1)} & \mathbf{G}_{E_1}^{*(A_{1,2}A_{1,2})} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{Z'}_{E_1(A_1)}\mathbf{R}_{E_1}^{(A_1A_1)}\mathbf{Z}_{E_1(A_1)} & \mathbf{Z'}_{E_1(A_1)}\mathbf{R}_{E_1}^{(A_1A_{1,2})}\mathbf{Z}_{E_1(A_{1,2})} \\ \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}_{E_1}^{(A_{1,2}A_1)}\mathbf{Z}_{E_1(A_1)} & \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}_{E_1}^{(A_{1,2}A_{1,2})}\mathbf{Z}_{E_1(A_{1,2})} \end{bmatrix}$$

(equation 1.5)

and, similarly,

$$\mathbf{D}_{E_2}^{*-1} - \mathbf{G}_{E_2}^{*-1} = \begin{bmatrix} \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2A_2)}\mathbf{Z}_{E_2(A_2)} & \mathbf{Z'}_{E_2(A_2)}\mathbf{R}_{E_2}^{(A_2A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} \\ \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2}A_2)}\mathbf{Z}_{E_2(A_2)} & \mathbf{Z'}_{E_2(A_{1,2})}\mathbf{R}_{E_2}^{(A_{1,2}A_{1,2})}\mathbf{Z}_{E_2(A_{1,2})} \end{bmatrix}$$ (equation 1.6).

Substituting “unknown” terms of the equation 1.1 by their corresponding terms

from the equations 1.3, 1.4, 1.5 and 1.6, the MME (equation 1.1) can be written as:

$$\begin{bmatrix} \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_1)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_2)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_1 A_{1,2}^0)} & \mathbf{D}_{E_1}^{*(A_1 A_1)} - \mathbf{G}_{E_1}^{*(A_1 A_1)} + \mathbf{G}_{E_0}^{(A_1 A_1)} & \mathbf{G}_{E_0}^{(A_1 A_2)} & \mathbf{D}_{E_1}^{*(A_1 A_{1,2})} - \mathbf{G}_{E_1}^{*(A_1 A_{1,2})} + \mathbf{G}_{E_0}^{(A_1 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_2 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_2 A_1)} & \mathbf{D}_{E_2}^{*(A_2 A_2)} - \mathbf{G}_{E_2}^{*(A_2 A_2)} + \mathbf{G}_{E_0}^{(A_2 A_2)} & \mathbf{D}_{E_2}^{*(A_2 A_{1,2})} - \mathbf{G}_{E_2}^{*(A_2 A_{1,2})} + \mathbf{G}_{E_0}^{(A_2 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2}^0)} & \mathbf{D}_{E_1}^{*(A_{1,2} A_1)} - \mathbf{G}_{E_1}^{*(A_{1,2} A_1)} + \mathbf{G}_{E_0}^{(A_{1,2} A_1)} & \mathbf{D}_{E_2}^{*(A_{1,2} A_2)} - \mathbf{G}_{E_2}^{*(A_{1,2} A_2)} + \mathbf{G}_{E_0}^{(A_{1,2} A_2)} & \mathbf{D}_{E_1}^{*(A_{1,2} A_{1,2})} - \mathbf{G}_{E_1}^{*(A_{1,2} A_{1,2})} + \\ & & & \mathbf{D}_{E_2}^{*(A_{1,2} A_{1,2})} - \mathbf{G}_{E_2}^{*(A_{1,2} A_{1,2})} + \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2})} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{u}}_{J(A_{1,2}^0)} \\ \hat{\mathbf{u}}_{J(A_1)} \\ \hat{\mathbf{u}}_{J(A_2)} \\ \hat{\mathbf{u}}_{J(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_{E_1}^{*(A_1 A_1)} \hat{\mathbf{u}}_{E_1(A_1)} + \mathbf{D}_{E_1}^{*(A_1 A_{1,2})} \hat{\mathbf{u}}_{E_1(A_{1,2})} \\ \mathbf{D}_{E_2}^{*(A_2 A_2)} \hat{\mathbf{u}}_{E_2(A_2)} + \mathbf{D}_{E_2}^{*(A_2 A_{1,2})} \hat{\mathbf{u}}_{E_2(A_{1,2})} \\ \mathbf{D}_{E_1}^{*(A_{1,2} A_1)} \hat{\mathbf{u}}_{E_1(A_1)} + \mathbf{D}_{E_1}^{*(A_{1,2} A_{1,2})} \hat{\mathbf{u}}_{E_1(A_{1,2})} + \mathbf{D}_{E_2}^{*(A_{1,2} A_2)} \hat{\mathbf{u}}_{E_2(A_2)} + \mathbf{D}_{E_2}^{*(A_{1,2} A_{1,2})} \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix}$$

(equation 1.7).

By replacing $\mathbf{G}^{*-1}$ in the equation 1.2 by the LHS of the equation 1.7 and $\mathbf{G}^{*-1}\boldsymbol{\mu}$ by the RHS of the equation 1.7, the following equations are obtained:

$$\begin{bmatrix} \mathbf{X}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{X}_{E_0} & \mathbf{X}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{Z}_{E_0} \\ \mathbf{Z}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{X}_{E_0} & \mathbf{Z}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{Z}_{E_0} + \begin{bmatrix} \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_1)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_2)} & \mathbf{G}_{E_0}^{(A_{1,2}^0 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_1 A_{1,2}^0)} & \mathbf{D}_{E_1}^{*(A_1 A_1)} - \mathbf{G}_{E_1}^{*(A_1 A_1)} + \mathbf{G}_{E_0}^{(A_1 A_1)} & \mathbf{G}_{E_0}^{(A_1 A_2)} & \mathbf{D}_{E_1}^{*(A_1 A_{1,2})} - \mathbf{G}_{E_1}^{*(A_1 A_{1,2})} + \mathbf{G}_{E_0}^{(A_1 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_2 A_{1,2}^0)} & \mathbf{G}_{E_0}^{(A_2 A_1)} & \mathbf{D}_{E_2}^{*(A_2 A_2)} - \mathbf{G}_{E_2}^{*(A_2 A_2)} + \mathbf{G}_{E_0}^{(A_2 A_2)} & \mathbf{D}_{E_2}^{*(A_2 A_{1,2})} - \mathbf{G}_{E_2}^{*(A_2 A_{1,2})} + \mathbf{G}_{E_0}^{(A_2 A_{1,2})} \\ \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2}^0)} & \mathbf{D}_{E_1}^{*(A_{1,2} A_1)} - \mathbf{G}_{E_1}^{*(A_{1,2} A_1)} + \mathbf{G}_{E_0}^{(A_{1,2} A_1)} & \mathbf{D}_{E_2}^{*(A_{1,2} A_2)} - \mathbf{G}_{E_2}^{*(A_{1,2} A_2)} + \mathbf{G}_{E_0}^{(A_{1,2} A_2)} & \mathbf{D}_{E_1}^{*(A_{1,2} A_{1,2})} - \mathbf{G}_{E_1}^{*(A_{1,2} A_{1,2})} + \\ & & & \mathbf{D}_{E_2}^{*(A_{1,2} A_{1,2})} - \mathbf{G}_{E_2}^{*(A_{1,2} A_{1,2})} + \mathbf{G}_{E_0}^{(A_{1,2} A_{1,2})} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{E_0} \\ \hat{\mathbf{u}}_{E_0} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{y}_{E_0} \\ \mathbf{Z}'_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{y}_{E_0} + \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_{E_1}^{*(A_1 A_1)} \hat{\mathbf{u}}_{E_1(A_1)} + \mathbf{D}_{E_1}^{*(A_1 A_{1,2})} \hat{\mathbf{u}}_{E_1(A_{1,2})} \\ \mathbf{D}_{E_2}^{*(A_2 A_2)} \hat{\mathbf{u}}_{E_2(A_2)} + \mathbf{D}_{E_2}^{*(A_2 A_{1,2})} \hat{\mathbf{u}}_{E_2(A_{1,2})} \\ \mathbf{D}_{E_1}^{*(A_{1,2} A_1)} \hat{\mathbf{u}}_{E_1(A_1)} + \mathbf{D}_{E_1}^{*(A_{1,2} A_{1,2})} \hat{\mathbf{u}}_{E_1(A_{1,2})} + \mathbf{D}_{E_2}^{*(A_{1,2} A_2)} \hat{\mathbf{u}}_{E_2(A_2)} + \mathbf{D}_{E_2}^{*(A_{1,2} A_{1,2})} \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix} \end{bmatrix}$$

(equation 1.8).

To simplify the system of equations 1.8, two genetic evaluations including in genealogy all animals (i.e., animals $A_{1,2}^0$, A_1, A_2 and $A_{1,2}$) equivalent to the genetic evaluations E_1 and E_2 for the external animals can be performed. Therefore, the following model of a genetic evaluation of all animals including only $\mathbf{y}_{E_1}$ can be written as:

$$\begin{bmatrix} \mathbf{y}_{E_1(A_1)} \\ \mathbf{y}_{E_1(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_{E_1(A_1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_{E_1(A_{1,2})} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{E_1(A^0_{1,2})} \\ \mathbf{u}_{E_1(A_1)} \\ \mathbf{u}_{E_1(A_2)} \\ \mathbf{u}_{E_1(A_{1,2})} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{E_1(A_1)} \\ \mathbf{e}_{E_1(A_{1,2})} \end{bmatrix},$$

where $\mathbf{u}_{E_1(A^0_{1,2})}$ and $\mathbf{u}_{E_1(A_2)}$ are the vectors of estimated random additive genetic effects for animals $A^0_{1,2}$ and A_2.

The corresponding MME can be written as follows:

$$\begin{bmatrix} \mathbf{G}^{(A^0_{1,2}A^0_{1,2})}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_1)}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_2)}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_1A^0_{1,2})}_{E_1} & \mathbf{Z'}_{E_1(A_1)}\mathbf{R}^{(A_1A_1)}_{E_1}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}^{(A_1A_1)}_{E_1} & \mathbf{G}^{(A_1A_2)}_{E_1} & \mathbf{Z'}_{E_1(A_1)}\mathbf{R}^{(A_1A_{1,2})}_{E_1}\mathbf{Z}_{E_1(A_{1,2})} + \mathbf{G}^{(A_1A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_2A^0_{1,2})}_{E_1} & \mathbf{G}^{(A_2A_1)}_{E_1} & \mathbf{G}^{(A_2A_2)}_{E_1} & \mathbf{G}^{(A_2A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_{1,2}A^0_{1,2})}_{E_1} & \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_1)}_{E_1}\mathbf{Z}_{E_1(A_1)} + \mathbf{G}^{(A_{1,2}A_1)}_{E_1} & \mathbf{G}^{(A_{1,2}A_2)}_{E_1} & \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_{1,2})}_{E_1}\mathbf{Z}_{E_1(A_{1,2})} + \mathbf{G}^{(A_{1,2}A_{1,2})}_{E_1} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{u}}_{E_1(A^0_{1,2})} \\ \hat{\mathbf{u}}_{E_1(A_1)} \\ \hat{\mathbf{u}}_{E_1(A_2)} \\ \hat{\mathbf{u}}_{E_1(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{Z'}_{E_1(A_1)}\mathbf{R}^{(A_1A_1)}_{E_1}\mathbf{y}_{E_1(A_1)} \\ \mathbf{0} \\ \mathbf{Z'}_{E_1(A_{1,2})}\mathbf{R}^{(A_{1,2}A_{1,2})}_{E_1}\mathbf{y}_{E_1(A_{1,2})} \end{bmatrix} = \mathbf{D}^{-1}_{E_1}\hat{\mathbf{u}}_{E_1}$$

(equation 1.9)

where $\mathbf{G}^{-1}_{E_1} = \begin{bmatrix} \mathbf{G}^{(A^0_{1,2}A^0_{1,2})}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_1)}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_2)}_{E_1} & \mathbf{G}^{(A^0_{1,2}A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_1A^0_{1,2})}_{E_1} & \mathbf{G}^{(A_1A_1)}_{E_1} & \mathbf{G}^{(A_1A_2)}_{E_1} & \mathbf{G}^{(A_1A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_2A^0_{1,2})}_{E_1} & \mathbf{G}^{(A_2A_1)}_{E_1} & \mathbf{G}^{(A_2A_2)}_{E_1} & \mathbf{G}^{(A_2A_{1,2})}_{E_1} \\ \mathbf{G}^{(A_{1,2}A^0_{1,2})}_{E_1} & \mathbf{G}^{(A_{1,2}A_1)}_{E_1} & \mathbf{G}^{(A_{1,2}A_2)}_{E_1} & \mathbf{G}^{(A_{1,2}A_{1,2})}_{E_1} \end{bmatrix}$ is the additive genetic (co)variance matrix taking into account all animals.

Similarly, MME for a genetic evaluation of all animals including only $\mathbf{y}_{E_2}$ can be written as:

$$\begin{bmatrix} \mathbf{G}_{E_2}^{(A_{1,2}^0 A_{1,2}^0)} & \mathbf{G}_{E_2}^{(A_{1,2}^0 A_1)} & \mathbf{G}_{E_2}^{(A_{1,2}^0 A_2)} & \mathbf{G}_{E_2}^{(A_{1,2}^0 A_{1,2})} \\ \mathbf{G}_{E_2}^{(A_1 A_{1,2}^0)} & \mathbf{G}_{E_2}^{(A_1 A_1)} & \mathbf{G}_{E_2}^{(A_1 A_2)} & \mathbf{G}_{E_2}^{(A_1 A_{1,2})} \\ \mathbf{G}_{E_2}^{(A_2 A_{1,2}^0)} & \mathbf{G}_{E_2}^{(A_2 A_1)} & \mathbf{Z'}_{E_2(A_2)} \mathbf{R}_{E_2}^{(A_2 A_2)} \mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_2}^{(A_2 A_2)} & \mathbf{Z'}_{E_2(A_2)} \mathbf{R}_{E_2}^{(A_2 A_{1,2})} \mathbf{Z}_{E_2(A_{1,2})} + \mathbf{G}_{E_2}^{(A_2 A_{1,2})} \\ \mathbf{G}_{E_2}^{(A_{1,2} A_{1,2}^0)} & \mathbf{G}_{E_2}^{(A_{1,2} A_1)} & \mathbf{Z'}_{E_2(A_{1,2})} \mathbf{R}_{E_2}^{(A_{1,2} A_2)} \mathbf{Z}_{E_2(A_2)} + \mathbf{G}_{E_2}^{(A_{1,2} A_2)} & \mathbf{Z'}_{E_2(A_{1,2})} \mathbf{R}_{E_2}^{(A_{1,2} A_{1,2})} \mathbf{Z}_{E_2(A_{1,2})} + \mathbf{G}_{E_2}^{(A_{1,2} A_{1,2})} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{u}}_{E_2(A_{1,2}^0)} \\ \hat{\mathbf{u}}_{E_2(A_1)} \\ \hat{\mathbf{u}}_{E_2(A_2)} \\ \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{Z'}_{E_2(A_1)} \mathbf{R}_{E_2}^{(A_1 A_1)} \mathbf{y}_{E_2(A_1)} \\ \mathbf{0} \\ \mathbf{Z'}_{E_2(A_{1,2})} \mathbf{R}_{E_2}^{(A_{1,2} A_{1,2})} \mathbf{y}_{E_2(A_{1,2})} \end{bmatrix} = \mathbf{D}_{E_2}^{-1} \hat{\mathbf{u}}_{E_2}$$

(equation 1.10).

However, although the vectors $\begin{bmatrix} \hat{\mathbf{u}}_{E_1(A_{1,2}^0)} \\ \hat{\mathbf{u}}_{E_1(A_2)} \end{bmatrix}$ and $\begin{bmatrix} \hat{\mathbf{u}}_{E_2(A_{1,2}^0)} \\ \hat{\mathbf{u}}_{E_2(A_1)} \end{bmatrix}$ remain unknown in practice, they could be predicted from the known vectors $\begin{bmatrix} \hat{\mathbf{u}}_{E_1(A_1)} \\ \hat{\mathbf{u}}_{E_1(A_{1,2})} \end{bmatrix}$ and $\begin{bmatrix} \hat{\mathbf{u}}_{E_2(A_2)} \\ \hat{\mathbf{u}}_{E_2(A_{1,2})} \end{bmatrix}$, e.g., through the selection index theory, respectively.

Therefore, the MME (1.1) can be written as $\left[\mathbf{G}_{E_0}^{-1} + \left(\mathbf{D}_{E_1}^{-1} - \mathbf{G}_{E_1}^{-1}\right) + \left(\mathbf{D}_{E_2}^{-1} - \mathbf{G}_{E_2}^{-1}\right)\right]\hat{\mathbf{u}}_J = \left[\mathbf{D}_{E_1}^{-1}\hat{\mathbf{u}}_{E_1} + \mathbf{D}_{E_2}^{-1}\hat{\mathbf{u}}_{E_2}\right]$ (equation 1.11).

By replacing $\mathbf{G}^{*-1}$ in the equation 1.2 by the LHS of the equation 1.11 and $\mathbf{G}^{*-1}\boldsymbol{\mu}$ by the RHS of the equation 1.11, we obtain:

$$\begin{bmatrix} \mathbf{X'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{X}_{E_0} & \mathbf{X'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{Z}_{E_0} \\ \mathbf{Z'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{X}_{E_0} & \mathbf{Z'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{Z}_{E_0} + \mathbf{G}_{E_0}^{-1} + \left(\mathbf{D}_{E_1}^{-1} - \mathbf{G}_{E_1}^{-1}\right) + \left(\mathbf{D}_{E_2}^{-1} - \mathbf{G}_{E_2}^{-1}\right) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{E_0} \\ \hat{\mathbf{u}}_{E_0} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{y}_{E_0} \\ \mathbf{Z'}_{E_0} \mathbf{R}_{E_0}^{-1} \mathbf{y}_{E_0} + \mathbf{D}_{E_1}^{-1} \hat{\mathbf{u}}_{E_1} + \mathbf{D}_{E_2}^{-1} \hat{\mathbf{u}}_{E_2} \end{bmatrix}.$$

This development could be extended to more than two sources of external information.

ADDITIONAL FILE 2

DOUBLE COUNTING BETWEEN INTERNAL AND EXTERNAL INFORMATION

Assume $\hat{\mathbf{u}}_{E_0}$ and $\mathbf{D}_{E_0}^{-1} = \mathbf{G}_{E_0}^{-1} + \boldsymbol{\Lambda}_{E_0}$ (equation 2.1), the vector of known internal EBV and the inverse of the associated prediction error (co)variance matrix obtained from the genetic evaluation E_0 based on the source E_0 including only internal information where $\mathbf{G}_{E_0}^{-1}$ is the inverse of the additive (co)variance matrix for all internal and external animals in the genetic evaluation E_0 and $\boldsymbol{\Lambda}_{E_0}$ is a block diagonal variance matrix. The vector $\hat{\mathbf{u}}_{E_1}$ and the matrix $\mathbf{D}_{E_1}^{-1} = \mathbf{G}_{E_1}^{-1} + \boldsymbol{\Lambda}_{E_1}$ (equation 2.2) are the vector of known external EBV and the inverse of the associated prediction error (co)variance matrix obtained from a genetic evaluation based on the source E_1 including external and internal information where $\mathbf{G}_{E_1}^{-1}$ is the inverse of the additive (co)variance matrix for all internal and external animals in the genetic evaluation E_1. The vector $\hat{\mathbf{u}}_{E_2}$ and the matrix $\mathbf{D}_{E_2}^{-1}$ are the vector of unknown external EBV and the inverse of the associated unknown prediction error (co)variance matrix obtained from a genetic evaluation E_2 based on the source E_2 including only external information. It is also assumed that double counting among animals due to relationships is taken into account.

Therefore, from $\boldsymbol{\Lambda}_{E_0}$ and $\boldsymbol{\Lambda}_{E_1}$, the diagonal matrix of RE expressing the amount of contributions only due to records, $\mathbf{RE}_{E_0}$ and $\mathbf{RE}_{E_1}$, can be estimated for the two sources of information E_0 and E_1, respectively. Because these RE are free of contributions due to relationships and due to correlated traits, the matrix of RE associated with the source of information E_2, $\mathbf{RE}_{E_2}$, can be estimated as follows:

$$\mathbf{RE}_{E_2} = \mathbf{RE}_{E_1} - \mathbf{RE}_{E_0} \qquad \text{(equation 2.3).}$$

It can be also written that $\boldsymbol{\Lambda}_{E_2} = \boldsymbol{\Lambda}_{E_1} - \boldsymbol{\Lambda}_{E_0}$ (equation 2.4). The unknown $\mathbf{D}_{E_2}^{-1}$ can be approximated as $\mathbf{D}_{E_2}^{-1} = \mathbf{G}_{E_2}^{-1} + \boldsymbol{\Lambda}_{E_2}$ (equation 2.5) where $\mathbf{G}_{E_2}^{-1}$ is the inverse of an unknown additive (co)variance matrix for the external source E_2. From the equations 2.1, 2.2 and 2.5, the equation 2.4 is equivalent to the equation $\mathbf{D}_{E_2}^{-1} - \mathbf{G}_{E_2}^{-1} = (\mathbf{D}_{E_1}^{-1} - \mathbf{G}_{E_1}^{-1}) - (\mathbf{D}_{E_0}^{-1} - \mathbf{G}_{E_0}^{-1})$ (equation 2.5).

Following the equations (VI.1) and assuming the lack of phenotypes in $\mathbf{y}_{E_0}$, it can

be written:

$$\left[\mathbf{G}_{E_1}^{-1} + \mathbf{D}_{E_0}^{-1} - \mathbf{G}_{E_0}^{-1} + \mathbf{D}_{E_2}^{-1} - \mathbf{G}_{E_2}^{-1}\right]\hat{\mathbf{u}}_{E_1} = \mathbf{D}_{E_0}^{-1}\hat{\mathbf{u}}_{E_0} + \mathbf{D}_{E_2}^{-1}\hat{\mathbf{u}}_{E_2} \qquad (2.6).$$

Because of the equation 2.5, the equation 2.6 can be written as follows:

$$\mathbf{D}_{E_1}^{-1}\hat{\mathbf{u}}_{E_1} = \mathbf{D}_{E_0}^{-1}\hat{\mathbf{u}}_{E_0} + \mathbf{D}_{E_2}^{-1}\hat{\mathbf{u}}_{E_2} .$$

Thereby, $\hat{\mathbf{u}}_{E_2}$ can be estimated using $\mathbf{D}_{E_2}^{-1}\hat{\mathbf{u}}_{E_2} = \mathbf{D}_{E_1}^{-1}\hat{\mathbf{u}}_{E_1} - \mathbf{D}_{E_0}^{-1}\hat{\mathbf{u}}_{E_0}$ (equation 2.7).

Because the source E_2 is free of internal information E_0, it can be integrated into the internal evaluation through the system of equations (VI.1) as follows:

$$\begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{X}_{E_0} & \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{Z}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{X}_{E_0} & \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{Z}_{E_0} + \mathbf{G}_{E_0}^{-1} + \mathbf{D}_{E_2}^{-1} - \mathbf{G}_{E_2}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{E_0} \\ \hat{\mathbf{u}}_{E_0} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{y}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{y}_{E_0} + \mathbf{D}_{E_2}^{-1}\hat{\mathbf{u}}_{E_2} \end{bmatrix}$$

(equation 2.8).

Due to the equations 2.5 and 2.7, $\mathbf{D}_{E_2}^{-1}$ and $\hat{\mathbf{u}}_{E_2}$ must not be estimated explicitly and the system of equations 2.8 can be written as follows:

$$\begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{X}_{E_0} & \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{Z}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{X}_{E_0} & \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{Z}_{E_0} + \mathbf{G}_{E_0}^{-1} + \left(\mathbf{D}_{E_1}^{-1} - \mathbf{G}_{E_1}^{-1}\right) - \left(\mathbf{D}_{E_0}^{-1} - \mathbf{G}_{E_0}^{-1}\right) \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{E_0} \\ \hat{\mathbf{u}}_{E_0} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{y}_{E_0} \\ \mathbf{Z}'_{E_0}\mathbf{R}_{E_0}^{-1}\mathbf{y}_{E_0} + \mathbf{D}_{E_1}^{-1}\hat{\mathbf{u}}_{E_1} - \mathbf{D}_{E_0}^{-1}\hat{\mathbf{u}}_{E_0} \end{bmatrix}$$

This development can be extended to integrate several sources of external information.

Chapter VII. WALLOON SINGLE-STEP GENOMIC EVALUATION SYSTEM INTEGRATING LOCAL AND MACE EBV

As reviewed in Chapter II, single-step genomic evaluations combine internal phenotypic and genealogic data with genomic data without considering external information. In Chapter VI, it has been suggested that the proposed Bayesian methods can be extended to single-step genomic evaluations. However, the inclusion of genomic data in the proposed Bayesian methods was never tested in the previous Chapters. Therefore, based on the Bayesian methods proposed in Chapter VI and based on single-step genomic evaluations, the aim of this Chapter was to test and implement a Walloon genomic evaluation for Holstein cattle that combines simultaneously genomic data and all available Walloon and external information.

Adapted from: Colinet, F.G., J. Vandenplas, P. Faux, S. Vanderick, R. Renaville, C. Bertozzi, X. Hubin, and N. Gengler. 2013. Walloon single-step genomic evaluation system integrating local and MACE EBV. *Interbull Bull.* 47:203–210.

ABSTRACT

Walloon dairy cattle could be considered as a small scale population where the majority of AI bulls are imported from several foreign countries. Single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) methods allow the simultaneous use of genomic, pedigree and phenotypic information and should reduce potential biases in the estimation of genomically enhanced breeding values (GEBV). Therefore, in the context of developing a Walloon genomic evaluation system, it was considered as the best option. However, in opposition to multi-step genomic predictions, ssGBLUP only uses local phenotypic information and is unable to use directly important other sources of information coming from abroad, e.g., multiple across country evaluation (MACE) results provided by International Bull Service (Interbull, Uppsale, Sweden). Therefore, single-step Genomic Bayesian Prediction (ssGBayes) was used as an alternative method for the Walloon genomic evaluation system. The ssGBayes approach allows combining simultaneously all available genotype, pedigree, local and foreign information in a local evaluation by considering a correct propagation of external information avoiding double counting of contributions due to relationships and due to records. In the Walloon genomic evaluation system, local information refers to Walloon estimated breeding values (EBV) and associated reliabilities (REL) and foreign information refers to MACE EBV and associated REL. Furthermore, the Bayesian approach has the advantage to directly combine EBV and REL without any deregression step. The ssGBayes method computed more accurate predictions for all types of animals. For example, for genotyped animals with low Walloon REL (<0.25) without MACE results and sired by genotyped bulls with MACE results, the average increase of REL for the studied traits was 0.39 points of which 0.14 points could be traced to the inclusion of MACE information. For other categories of genotyped animals, the contribution by MACE information was high too. The new Walloon genomic evaluation system passed the Interbull GEBV tests for several traits in July 2013. This approach has the potential to improve current genomic prediction strategies as it can be used in other settings where the combination of different sources of information is required.

Key words: Bayesian integration, MACE, genomic prediction

INTRODUCTION

Simultaneous use of all data by Best Linear Unbiased Prediction (BLUP) is a

condition to predict unbiased estimated breeding values (EBV; Henderson, 1984). However, this condition is not always fully met. For example, small scale local populations lead to evaluations based only on local data while foreign bulls are used (e.g., 87% of cows in 1st to 3rd parity in 2012 were sired by artificial insemination (AI) bulls born outside of Walloon Region of Belgium). Although these bulls were strongly preselected, foreign raw data used to select them is unavailable leading to potential biases in local evaluations. Local EBV will be also less accurate because only incomplete data (i.e., foreign raw data not included) is available. Genomic selection could increase these problems for local genomic evaluations.

Most current genomic evaluation systems are multi-step, relying heavily on the use of multiple across country evaluation (MACE) results as the primary source of foreign phenotypic information. However, these implementations of genomic prediction using MACE results mitigated these issues only for sires with high REL which are introduced during the single nucleotide polymorphisms (SNP) prediction equation estimation step.

Single-step genomic evaluations (ssGBLUP; e.g., Aguilar et al., 2010; Christensen and Lund, 2010) should reduce potential biases in the estimation of genomically enhanced breeding values (GEBV) by the simultaneous combination of genomic, pedigree and all local phenotypic information (VanRaden, 2012), also because fewer approximations are made than in multi-step methods. Therefore, in the context of developing a Walloon genomic evaluation system, ssGBLUP was considered as the best option. However, in opposition to multi-step genomic prediction, ssGBLUP uses only local phenotypic information and is unable to use directly other important sources of information provided, e.g., by MACE. Nevertheless, the recovery of such important sources of information in the Walloon genomic evaluation system was required due to the widespread use of imported AI bulls.

Therefore, in the context of the Walloon genomic evaluation system, the aim of this research was to assess the potential of a Bayesian approach, based on ssGBLUP, to simultaneously combine all available genotype, pedigree, local and foreign information in a local genomic evaluation. This approach also avoids deregression steps, allows a correct propagation of external information and avoids multiple considerations of contributions due to relationships and due to records.

MATERIALS AND METHODS

In this study, local information will refer to local EBV and associated reliabilities (REL) estimated from all available local data and foreign information will refer to MACE EBV and associated REL.

Currently, in the Walloon Region of Belgium, genomic evaluations for the Holstein breed are performed for all traits submitted to MACE. In this study, results are reported showing the strategy and the results obtained in the July 2013 run for milk, fat and protein yields, somatic cell score, longevity and two conformational traits (stature and udder support).

The genomic evaluation system implemented in the Walloon Region of Belgium consisted of several steps. First, a group of genotyped animals was defined as those animals born after the year 1998. Ancestors for these animals were extracted from the database used for the official Walloon genetic evaluation and covered up to 6 known ancestral generations. After extraction, the pedigree file contained 16 234 animals of which 1909 animals (1378 bulls and 525 cows) were genotyped. The large majority of genotyped cows were not from a selected set of local Holstein animals but were from a set of Holstein animals representing the Walloon variability in the Holstein cattle. A total of 38 604 SNP markers were selected after editing.

Local information included EBV and associated REL for cows and bulls estimated from data provided by the Walloon Breeding Association (subscript W; EBV_W, REL_W) for the official Walloon evaluation of April 2013 (Auvray and Gengler, 2002; Croquet et al., 2006). Table VII-1 shows the number of animals associated to Walloon information for which EBV_W were available for each studied trait. Foreign information included EBV and REL for sires provided by the April 2013 MACE evaluation performed by Interbull (subscript M; EBV_M, REL_M; Table VII-1).

Table VII-1. Used genetic parameters, local and foreign information available for the genomic evaluation for the seven reported traits

Trait	Heritability	Genetic variance	No. of animals			No. of genotyped animals		
			EBV_W	EBV_M	EBV_{WC}	EBV_W	EBV_M	EBV_{WC}
Milk yield	0.38	280 425	12 046	1981	601	1762	1205	278
Fat yield	0.43	523	12 046	1981	601	1762	1205	278
Protein yield	0.41	262	12 046	1981	601	1762	1205	278
SCS	0.13	0.2060	12 047	1941	575	1762	1167	261
Longevity	0.11	0.0797	11 641	1914	520	1758	1155	238
Stature	0.52	1.1984	12 671	1922	595	1706	1158	277
Udder support	0.19	0.3212	12 226	1911	573	1699	1158	277

For every trait, contributions of Walloon information into MACE were determined based on the domestic effective daughter equivalents (EDC) associated to EBV_M and REL_M as reported by Interbull. MACE information free of Walloon information had therefore a reported domestic EDC equal to 0. For all animals and traits with a domestic EDC different from 0, Walloon EBV and associated REL contributing to the April 2013 MACE routine-run (subscript Wc; EBV_{Wc}, REL_{Wc}) were considered to avoid double counting of contributions due to records (Table VII-1). Information was harmonized between the local and MACE traits by adjusting scale and mean difference towards the original expression of the trait in the Walloon genetic evaluation computations. As shown in Table VII-1, numbers of available local and foreign records were slightly different among the traits.

The Bayesian procedures that integrate multiple sources of external information into genetic evaluations were outlined by Vandenplas et al. (2014). Also, these authors outlined that their proposed systems of equations could be extended to integrate multiple sources of external information into ssGBLUP (Vandenplas et al., 2014). Thereby, their proposed equation (VI.4) that blends several sources of external information by avoiding double counting of contributions due to records and due to relationships was adapted to blend Walloon and MACE information into a ssGBLUP for each trait separately. This method, hereafter called single-step Genomic Bayesian Prediction (ssGBayes), was used as an alternative method for the Walloon genomic evaluation system. The equation associated with ssGBayes that blends genomic, Walloon and MACE information and that considers Walloon information contributing to MACE (hereafter called $ssGBayes_{W+M-Wc}$) can be written as follows:

$$\left(\mathbf{G}^{*-1} + \mathbf{\Lambda_W} + \mathbf{\Lambda_M} - \mathbf{\Lambda_{Wc}}\right)\hat{\mathbf{a}}_{\mathbf{W+M-Wc}} = \mathbf{D_W^{-1}}\hat{\mathbf{u}}_{\mathbf{W}} + \mathbf{D_M^{-1}}\hat{\mathbf{u}}_{\mathbf{M}} - \mathbf{D_{Wc}^{-1}}\,\hat{\mathbf{u}}_{\mathbf{Wc}} \qquad \text{(VII.1)}$$

where $\mathbf{G}^* = \mathbf{H}\sigma_a^2$ is the combined genomic-pedigree based (co)variance matrix, $\mathbf{H}$ is the combined genomic-pedigree based relationship matrix (e.g., Aguilar et al., 2010; Christensen and Lund, 2010), σ_a^2 is the additive genetic variance, $\hat{\mathbf{a}}_{\mathbf{W+M-Wc}}$ is the vector of Walloon GEBV based on Walloon and MACE information, $\hat{\mathbf{u}}_{\mathbf{W}}$ is the vector of EBV_W, $\hat{\mathbf{u}}_{\mathbf{M}}$ is the vector of EBV_M, $\hat{\mathbf{u}}_{\mathbf{Wc}}$ is the vector of EBV_{Wc}, $\mathbf{\Lambda_i}$ (i = W, M and Wc) is a matrix mimicking least squares part of hypothetical BLUP, and $\mathbf{D_i^{-1}}$ is the inverse of the prediction error (co)variance matrix of $\hat{\mathbf{u}}_{\mathbf{i}}$.

The inverse of the combined genomic-pedigree based relationship matrix $\mathbf{H}$ was

computed using the inverse of the additive pedigree relationship matrix and a modified genomic relationship matrix using a weight equal to 0.95 for raw genomic relationships and equal to 0.05 for pedigree relationships. For matrices compatibility, both diagonal and off-diagonal values were respectively centred on the average of diagonal and off-diagonal elements of the subpart of the additive relationship matrix among genotyped animals.

Regarding the vectors $\hat{\mathbf{u}}_\mathbf{i}$ for the 3 sources of information (i.e., W, M and Wc), it is worth noting that only some animals included in the pedigree were associated with known EBV and REL (hereafter called external animals; Table VII-1). Therefore, for each *i*th source of information, animals not associated with available EBV were called internal animals and the vector of EBV_i (*i* = W, M and Wc) for all animals included in the pedigree, $\hat{\mathbf{u}}_\mathbf{i}$, was estimated as (Vandenplas et al., 2014):

$$\hat{\mathbf{u}}_\mathbf{i} = \begin{bmatrix} \mathbf{G}_{\mathbf{i(IE)}}\,\mathbf{G}_{\mathbf{i(EE)}}^{-1}\,\hat{\mathbf{u}}_{\mathbf{i(E)}} \\ \hat{\mathbf{u}}_{\mathbf{i(E)}} \end{bmatrix}$$

where the subscript I refers to internal animals not associated with the *i*th source of information, the subscript E refers to external animals associated with the *i*th source of information, $\hat{\mathbf{u}}_{\mathbf{i(E)}}$ is the vector of EBV for the *i*th source of information associated with external animals and $\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{\mathbf{i(EE)}} & \mathbf{G}_{\mathbf{i(EI)}} \\ \mathbf{G}_{\mathbf{i(IE)}} & \mathbf{G}_{\mathbf{i(II)}} \end{bmatrix}^{-1}$ is the inverse of the pedigree-based (co)variance matrix.

For the 3 sources of information (i.e., W, M and Wc), the matrix $\mathbf{D}_\mathbf{i}^{-1}$ was approximated as $\mathbf{D}_\mathbf{i}^{-1} = \mathbf{G}^{-1} + \mathbf{\Lambda}_\mathbf{i}$ where $\mathbf{\Lambda}_\mathbf{i}$ is a diagonal variance matrix with one diagonal element per animal equal to $RE_{ij} \big/ \sigma_e^2$ for j = 1, 2, …, J animals (Vandenplas et al., 2014). The element σ_e^2 is the residual variance and the element RE_{ij} is the effective number of records, so-called record equivalents, for the *j*th animal. Record equivalents expressed the amount of contributions for an animal (Misztal and Wiggans, 1988). It is worth noting that double counting of contributions due to relationships among related animals could exist because both Walloon and MACE information were associated with related animals. Therefore, the combination of Walloon and MACE information was performed by taking into account contributions due to relationships among related animals. These contributions were estimated by a two-step algorithm (TSA; Vandenplas and Gengler, 2012). It takes into account all relationships between animals associated

with information and their ancestors. Therefore, for the internal animals, RE_{ij} is equal to 0. All contributions for these animals were only due to their relationships with external animals. For the external animals, RE_{ij} was estimated through TSA and only express the amount of contributions due to records.

Because a major feature of the Walloon genomic evaluation system is its ability to use MACE information, in comparison to ssGBLUP, the influence of the use of this information was tested. To test this influence, ssGBayes was run considering only Walloon information (ssGBayes$_W$) using the following system of equations:

$$\left(\mathbf{G}^{*-1} + \mathbf{\Lambda}_{\mathbf{W}}\right)\hat{\mathbf{a}}_{\mathbf{W}} = \mathbf{D}_{\mathbf{W}}^{-1}\hat{\mathbf{u}}_{\mathbf{W}} \qquad \text{(VII.2)}$$

where $\hat{\mathbf{a}}_{\mathbf{W}}$ is the vector of Walloon GEBV only based on Walloon information.

Approximation of genomic REL (GREL) for GEBV in genomic evaluation systems is not always straight forward (Misztal, 2013). Because the equations (VII.1) and (VII.2) associated with ssGBayes$_{W+M-Wc}$ and ssGBayes$_W$, respectively, represented hypothetical mixed model equations, the computation of REL was tested using the standard formula:

$$GREL = 1 - PEV / \sigma_g^2 \qquad \text{(VII.3)}$$

where σ_g^2 is the diagonal element of $\mathbf{G}^*$ and PEV is the prediction error variance obtained from the diagonal element of the inverted left hand side of the equations (VII.1) and (VII.2), respectively. By using diagonal elements of $\mathbf{G}^*$, the method corrected for inbreeding estimated using combined pedigree and genomic information.

The two ssGBayes were performed using BLUPF90 (Misztal, 2013) modified to implement equations (VII.1) and (VII.2).

RESULTS AND DISCUSSION

For all traits, among the approximately 12 000 animals associated with available Walloon information, around 1950 bulls were also evaluated by Interbull (Table VII-1). Walloon information for around one third of these bulls contributed to the April 2013 MACE routine-run. Table VII-1 also indicates that at least 83% of the 1378 genotyped bulls and at least 11% of the 16 234 animals in the considered pedigree file had foreign information. This large amount of additional information was incorporated in the genomic evaluation system and would allow increasing the overall accuracy of the produced GEBV.

Table VII-2 gives details on the improvement of REL when estimating (G)EBV from different sources. First, the improvement due to including only genomic information was considered. For the genotyped bulls with low reliable official Walloon EBV (REL_W < 0.50), the genomic information allowed an substantial increase of between 0.13 and 0.19 points for average REL of these bulls according to the studied traits.

The genomic information also increased average REL for the two other categories of bulls with more accurate Walloon EBV. Indeed, the average REL was increased with 0.05-0.06 points for the bulls with REL_W between 0.50 and 0.75. Even for locally well proven bulls (i.e., $REL_W \geq 0.75$), the genomic information added 0.01 to the average REL.

Considering the simultaneous combination of genomic and foreign information (i.e., $ssGBayes_{W+M-Wc}$), the increases of the averaged REL for each of the three mentioned categories of genotyped bulls (Table VII-2) were higher than those associated with $ssGBayes_W$. As expected, the highest increase of REL was observed for the bulls with the lowest REL_W. When comparing different traits, the use of $ssGBayes_{W+M-Wc}$ led to an increase of average REL between 0.20 points for longevity and 0.41 points for milk yield compared to $ssGBayes_W$. The increase was lower for genotyped bulls with REL_W included in the range [0.50-0.75[with 0.09 to 0.22 additional points of REL. Even for the already locally well proven bulls (i.e., $REL_W \geq 0.75$), $ssGBayes_{W+M-Wc}$ still provided more reliable GEBV than $ssGBayes_W$. Additional points of REL ranged from 0.02 for longevity to 0.05 for fat yield (Table VII-2).

Table VII-3 shows the improvements for genotyped animals only associated with EBV_W, i.e., without foreign information, and sired by genotyped bulls with MACE results. These genotyped animals were Walloon cows and bulls as well as foreign or Walloon bulls to be tested. Again, similarly to Table VII-2, even if $ssGBayes_W$ allowed an increase of average REL with 0.16-0.28 additional points, $ssGBayes_{W+M-Wc}$ led to higher REL. For most traits, $ssGBayes_{W+M-Wc}$ provided an average REL higher than 0.50 for these genotyped animals with a REL_W included in the range]0.00-0.25[.

Table VII-2. Average reliabilities (REL; SD in parentheses) associated to EBV_W, $GEBV_W$ and $GEBV_{W+M-Wc}$ for genotyped bulls for the seven studied traits

Trait	REL_W											
	]0.00 - 0.50[				[0.50 – 0.75[				[0.75 – 0.99]			
	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]
Milk yield	647	0.25 (0.12)	0.44 (0.09)	0.80 (0.09)	173	0.63 (0.07)	0.69 (0.06)	0.87 (0.05)	390	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
Fat yield	642	0.26 (0.12)	0.45 (0.09)	0.80 (0.09)	158	0.63 (0.07)	0.69 (0.05)	0.87 (0.04)	412	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
Protein yield	644	0.26 (0.12)	0.44 (0.09)	0.80 (0.09)	162	0.63 (0.07)	0.69 (0.06)	0.87 (0.04)	404	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
SCS	682	0.25 (0.12)	0.43 (0.09)	0.84 (0.12)	186	0.63 (0.07)	0.68 (0.06)	0.90 (0.08)	337	0.90 (0.07)	0.91 (0.06)	0.96 (0.04)
Longevity	889	0.23 (0.12)	0.36 (0.09)	0.51 (0.09)	146	0.61 (0.08)	0.66 (0.07)	0.75 (0.07)	149	0.86 (0.06)	0.87 (0.06)	0.89 (0.04)
Stature	632	0.28 (0.10)	0.46 (0.08)	0.82 (0.13)	141	0.63 (0.07)	0.69 (0.05)	0.91 (0.06)	408	0.91 (0.07)	0.92 (0.06)	0.96 (0.04)
Udder support	699	0.28 (0.10)	0.43 (0.08)	0.71 (0.14)	189	0.63 (0.07)	0.68 (0.06)	0.84 (0.08)	286	0.91 (0.08)	0.92 (0.07)	0.95 (0.04)

[1]REL obtained from Walloon polygenic evaluation.
[2]REL obtained from Walloon genomic evaluation using only EBV_W (eq. (VII.1) and (VII.3)).
[3]REL obtained from Walloon genomic evaluation using EBV_W, EBV_M and EBV_{Wc} (eq. (VII.2) and (VII.3)).

Table VII-3. Average reliabilities (REL; SD in parentheses) associated to EBV_W, $GEBV_W$ and $GEBV_{W+M}$ for genotyped animals without MACE results and sired by genotyped bulls with MACE result for the studied traits

Trait	REL_W											
	]0.00 – 0.25[				[0.25 – 0.50[				[0.50 – 0.75[			
	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	N	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]
Milk yield	43	0.11 (0.10)	0.38 (0.08)	0.52 (0.04)	123	0.43 (0.07)	0.54 (0.05)	0.60 (0.03)	101	0.52 (0.02)	0.61 (0.02)	0.64 (0.02)
Fat yield	43	0.11 (0.10)	0.39 (0.08)	0.53 (0.04)	77	0.41 (0.08)	0.54 (0.05)	0.60 (0.04)	147	0.54 (0.03)	0.62 (0.02)	0.65 (0.02)
Protein yield	43	0.11 (0.10)	0.39 (0.08)	0.53 (0.04)	91	0.42 (0.08)	0.54 (0.05)	0.60 (0.03)	133	0.53 (0.02)	0.61 (0.02)	0.65 (0.02)
SCS	52	0.11 (0.10)	0.36 (0.08)	0.54 (0.04)	194	0.43 (0.06)	0.53 (0.04)	0.61 (0.03)	30	0.53 (0.07)	0.61 (0.06)	0.66 (0.05)
Longevity	117	0.14 (0.08)	0.30 (0.06)	0.38 (0.04)	165	0.31 (0.04)	0.40 (0.04)	0.44 (0.03)	0	--- (---)	--- (---)	--- (---)
Stature	46	0.08 (0.09)	0.36 (0.10)	0.53 (0.05)	114	0.36 (0.06)	0.51 (0.05)	0.59 (0.03)	120	0.70 (0.04)	0.74 (0.03)	0.76 (0.03)
Udder support	65	0.11 (0.09)	0.34 (0.09)	0.48 (0.06)	158	0.38 (0.07)	0.50 (0.05)	0.56 (0.04)	73	0.53 (0.03)	0.60 (0.03)	0.64 (0.03)

[1] REL obtained from Walloon polygenic evaluation.
[2] REL obtained from Walloon genomic evaluation using only EBV_W (eq. (VII.1) and (VII.3)).
[3] REL obtained from Walloon genomic evaluation using EBV_W, EBV_M and EBV_{Wc} (eq. (VII.2) and (VII.3)).

The genomic evaluation addressed another category of genotyped animals including the newborn Walloon bulls (candidate for AI bulls) and recently imported AI bulls (or with a forecasted importation), both types of animals not being yet included in the routine genetic evaluations. Therefore, these bulls had no available external information due to their absence in the pedigree file at the last official Walloon genetic evaluation. These animals were incorporated in the genomic evaluation system by only using their available information (i.e., pedigree and genotypes) and information available for their relatives. If their sires were associated with MACE EBV, REL of their GEBV were higher than the threshold defined for the considered trait to be publishable (Table VII-4). For each of the seven studied traits, ssGBayes$_{W+M-Wc}$ provided a publishable GEBV for more than two thirds of these bulls.

Table VII-4. Average reliabilities (REL; SD in parentheses) associated with GEBV$_{W+M-Wc}$[1] for genotyped bulls without external phenotype information (neither local EBV neither MACE EBV), sired by genotyped bulls with MACE results for the studied trait

Trait	Publication rule: REL ≥	No. of bulls	Averaged REL$_{W+M-Wc}$ (SD)	No. of publishable GEBV$_{W+M-Wc}$
Milk yield	0.50	17	0.53 (0.05)	13
Fat yield	0.50	17	0.53 (0.06)	13
Protein yield	0.50	17	0.53 (0.05)	13
SCS	0.45	20	0.54 (0.05)	19
Longevity	0.35	23	0.38 (0.05)	18
Stature	0.50	21	0.54 (0.06)	15
Udder support	0.50	21	0.47 (0.07)	15

[1] GEBV$_{W+M-Wc}$ and REL$_{W+M-Wc}$ from Walloon genomic evaluation using EBV$_W$, EBV$_M$ and EBV$_{Wc}$

Currently, the system is not yet optimized by genotyping additional related animals with information (e.g., maternal grand-sires, brothers, half-brothers) in order to increase the links between these candidate animals and the genotyped animals with information. An appropriate strategy will be implemented to detect the most important animals to be also genotyped which should increase the proportion of publishable GEBV even further.

The Walloon genomic evaluation system was used and results tested inside the GEBV tests of Interbull. Results passed the tests for several traits in April and July 2013. Currently, research is undertaken to optimize the formation of the modified genomic relationships matrix. Indeed, several tests showed that the weighting used has a large influence and that the optimal proportion between raw genomic and pedigree

relationships directly reflects the critical partitioning of total genetic variance in variances explained by SNP effects or polynomial residuals.

CONCLUSIONS

The ssGBayes method, through its Bayesian approach, integrated well MACE results into ssGBLUP and allowed recovering indirectly a large amount of phenotypic information. All available external sources of information were correctly propagated avoiding double counting of contributions due to relationships and due to own records. Therefore, the ssGBayes method proved to be a good choice for the Walloon genomic evaluation system integrating Walloon and MACE EBV. Additional optimizations are currently under development by genotyping important sires and by adapting the correct partitioning of additive total variance for a given trait in order to increase the number of traits that pass the Interbull GEBV test. The ssGBayes method used in the Walloon genomic evaluation system can also be adapted to a multi-trait setting allowing the genomic evaluation of only locally available traits (e.g., fine milk composition, methane emissions) using external information from correlated traits (e.g., traits evaluated by Interbull).

Finally, the ssGBayes approach has the potential to improve current genomic prediction strategies as it can be used in other settings (e.g., beef cattle and pigs) where the combination of different sources of information is required.

ACKNOWLEDGEMENTS

The authors acknowledge the Ministry of Agriculture of Walloon Region of Belgium (Service Public de Wallonie, DGO3) for financial support through several grants, i.e. research projects D31-1233 and D31-1274. Jérémie Vandenplas, as a research fellow, and Nicolas Gengler, as an honorary senior research associate, acknowledge the former and ongoing support of the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium) for their positions and through several grants. Pierre Faux acknowledges the financial support of National Research Fund, Luxembourg through project NextGenGES-PHD-09-118, as part of a private public partnership between CONVIS (Ettelbruck, Luxembourg) and Gembloux Agro-Bio Tech of the University of Liege. Jérémie Vandenplas acknowledges I. Misztal for hosting him at Animal and Dairy Sciences Department of University of Georgia (Athens, USA), G. Gorjanc for hosting him at Animal Science Department of University of Ljubljana. The help of S. Tsuruta

(University of Georgia, Athens, USA) and I. Aguilar (Instituto Nacional de Investigación Agropecuaria, Las Brujas, Urugay) concerning the BLUPF90 family of programs, on which the developed computing programs are based, is acknowledged. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI) funded by the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium) under Grant No. 2.5020.11.

REFERENCES

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Auvray, B., and N. Gengler. 2002. Feasibility of a Walloon test-day model and study of its potential as tool for selection and management. *Interbull Bull.* 29:123–127.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.

Croquet, C., P. Mayeres, A. Gillon, S. Vanderick, and N. Gengler. 2006. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89:2257–2267.

Henderson, C.R. 1984. Applications of linear models in animal breeding. 2nd ed. University of Guelph, Guelph, ON, Canada.

Misztal, I. 2013. BLUPF90 family of programs. Accessed Jul. 3, 2012. http://nce.ads.uga.edu/wiki/doku.php.

Misztal, I., and G.R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models? *J. Dairy Sci.* 71(Suppl. 2):27–32.

Vandenplas, J., F.G. Colinet and N. Gengler. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. *Genet. Sel. Evol.* 46:59

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

VanRaden, P.M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. *Interbull Bull.* 45:29–33.

Chapter VIII. GENERAL DISCUSSION

Based on previous research, Chapter III to Chapter VII presented improvements and implementations of Bayesian approaches that integrate several sources of external information into an internal genetic or single-step genomic evaluation. Although the Bayesian approaches were developed to integrate external information correlated to the internal phenotypic traits into a multivariate genetic evaluation (e.g., Chapter VI), external information was always considered to be the same trait as the internal phenotypic trait from Chapter III to Chapter VII. Therefore, in this Chapter, a general discussion was first initiated by investigating the integration of correlated external information into a multivariate genetic evaluation. Second, a comparative study among the different approaches that combine simultaneously external information and internal data was detailed.

Bayesian approaches that integrate external information, i.e., estimated breeding values (EBV) and associated reliabilities (REL), into an internal genetic evaluation were proposed by several authors (e.g., Quaas and Zhang, 2006; Legarra et al., 2007). Different issues were identified and, thereby, improvements and implementations of algorithms and methods, as well as extension of their scope to genomics were proposed in Chapter III until Chapter VII. This included the reduction of computational burdens (Chapter III), and the consideration of double counting of contributions due to relationships (Chapter III) and due to records (Chapter VI). In Chapter VI, modified mixed models equations (MME) were presented to integrate several sources of external information into genetic and genomic evaluations. To our knowledge, such an integration of several sources of external information has not been showed by previous authors in the context of Bayesian approaches. Also, the proposed improvements and algorithms were tested with real data in the context of the Belgian genetic evaluation for jumping horses (Chapter IV) and in the context of the Walloon single-step genomic evaluation system for Holstein cattle (Chapter VII).

It is worth noting that external information was always considered to belong to the same trait as the considered internal phenotypic trait from Chapter III to Chapter VII. However, the Bayesian approaches were also developed for integrating external information correlated to the internal phenotypic traits (hereafter called correlated external information) into multivariate genetic evaluations (e.g., Chapter VI; Quaas and Zhang, 2006). These Bayesian approaches need an approximation of the inverse of the prediction error (co)variance matrix associated with external EBV obtained from external multivariate MME. An approximation of the inverse of this matrix was proposed in Chapter VI. As stated in this Chapter VI, the proposed approximation is different from the approximation proposed by Quaas and Zhang (2006). The difference between the two approximations is linked to the approximation of the least squares part of the left hand side (LHS) of the external multivariate MME that is needed for the approximation of the inverse of the prediction error (co)variance matrix. Therefore, our approximation (Chapter VI) and the one presented by Quaas and Zhang (2006) were tested and compared based on simulated data. Results are given in the subsequent section. Finally, a comparison of the three types of approaches that combine simultaneously internal data and external information for univariate analyses and multivariate genetic evaluations is presented in the last section of this Chapter.

INTEGRATION OF CORRELATED EXTERNAL INFORMATION

Thanks to correlations among different traits, some advantages of multivariate genetic evaluations are firstly the prediction of EBV for traits of interest for which phenotypes could be difficult, or impossible to collect internally and, secondly, the improvement of accuracy of EBV for the different traits, under some conditions (e.g., Schaeffer, 1994; Mrode, 2005). Thereby, integration of external information correlated to the internal phenotypic traits into an internal multivariate genetic evaluation could be interesting to solve different issues. Firstly, accuracy of internal evaluations may be still low for some traits of interest, while accurate external evaluations for similar traits or for correlated traits are routinely performed, potentially at an international level (e.g., MACE evaluations for commonly evaluated traits). Therefore, integration of correlated external information provided by external evaluations into a multivariate evaluation could improve the accuracy of the traits of interest. Secondly, external information can be expressed on other scales or units of measurement, or it can be associated with different heritabilities and genetic parameters than the internal traits of interest. Therefore, like MACE, integration of correlated external information could be an optimal approach to evaluate genetic merits of animals without the use of conversion equations and without the dependence of the internal genetic evaluation to these equations. In the context of the Bayesian evaluations integrating external information into a multivariate genetic evaluation, two different approximations of the least squares part of the LHS of the external MME, involved in the approximation of the inverse of the prediction error (co)variance matrix associated with external EBV (Chapter III; Chapter VI), were proposed. The first approach to approximate the external least squares part was proposed by Quaas and Zhang (2006) and used in Chapter III and Chapter IV. This approach involves the matrix of additive genetic (co)variances among traits. The second approach was proposed in Chapter VI and involves the matrix of residual (co)variances among traits, instead of the additive genetic (co)variance matrix. Therefore, based on simulated data, the aim of this study was to test the two proposed approximations through their use in a Bayesian approach that combines internal pedigree and phenotypes for a trait of interest with external information, i.e., EBV and REL provided by an external genetic evaluation for a trait genetically correlated to this trait of interest.

MATERIAL AND METHODS

The context of the simulation was a population where only females were

associated with phenotypes. Females were assigned to different herds. Phenotypes were observed for two genetically correlated traits. It was also assumed that phenotypes were observed for only one of the two traits in each herd. Therefore, females were associated with phenotypes for only one of the two traits and the residual correlation was assumed to be equal to zero. For pedigree, one hundred replicates were first generated by the QMSim program (Sargolzaei and Schenkel, 2009). The parameter file considered 40 male founders and 200 female founders to generate 10 generations of animals. The litter size and the proportion of male progeny were assumed to be 1 and 50%, respectively. Matings and selection were random. It was also assumed that 40% of sires and 10% of dams were replaced in all generations. For each pedigree, females were randomly attributed to one of the five assumed herds under the assumption that each herd included on average 1/5 of the total amount of females. Phenotypes for the two traits, hereafter called “trait of interest” and “correlated trait”, were simulated for each female following Van Vleck (1982). Females attributed to the two first herds were only associated with phenotypes related to the trait of interest and females attributed to the three last herds were only associated with phenotypes related to the correlated trait. Heritabilities of 0.10 and 0.35 were considered for the trait of interest and the correlated trait, respectively. Corresponding phenotypic variances were 80000 u_t^2 and 100 u_c^2 where u_t^2 and u_c^2 are the squares of the units of measurement for the trait of interest (u_t) and the correlated trait (u_c), respectively. Genetic correlations between traits (r_g) equal to 0.10, 0.25, 0.50, 0.75, and 0.90 were considered. As explained previously, the residual correlation between traits (r_e) was assumed to be equal to 0.00 because there was no environmental covariance among the traits as the two traits were not observed on same animals. Hereafter, phenotypic data for the trait of interest and related to the two first herds will be considered as internal data while phenotypic data for the correlated trait and related to the three last herds will be considered as external data. The simulation of phenotypes was replicated for each of the 100 pedigree and for each of the 5 considered genetic correlations.

Using simulated data, three conventional genetic evaluations and two Bayesian evaluations were performed for each genetic correlation. All evaluations were based on the same pedigree that includes all animals. (1) Joint evaluations ($EVAL_J$) were performed as bivariate Best Linear Unbiased Prediction (BLUP) evaluations using the system of equations (VI.2) and based on external and internal data. These evaluations

were assumed to be the reference. (2) Internal evaluations ($EVAL_I$) were performed as bivariate BLUP evaluations using the system of equations (VI.2) and based only on internal data. External data was ignored by $EVAL_I$. (3) External evaluations ($EVAL_E$) were performed as bivariate BLUP evaluations using the system of equations (VI.2) and based on external data. Internal data was ignored by $EVAL_E$.

Then, two bivariate Bayesian evaluations integrating external information, i.e., EBV and associated REL provided by $EVAL_E$, for all external sires having daughters with phenotypes for the correlated trait, were also performed using the following system of equations, with the same compact notation for both evaluations (e.g., Chapter III):

$$\begin{bmatrix} \mathbf{X'_I R_I^{-1} X_I} & \mathbf{X'_I R_I^{-1} Z_I} \\ \mathbf{Z'_I R_I^{-1} X_I} & \mathbf{Z'_I R_I^{-1} Z_I + D_E^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_\mathbf{I} \\ \hat{\mathbf{u}}_\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{X'_I R_I^{-1} y_I} \\ \mathbf{Z'_I R_I^{-1} y_I + D_E^{-1} \hat{u}_E} \end{bmatrix}$$

where I and E refer to $EVAL_I$ and $EVAL_E$, respectively, $\mathbf{X_I}$ and $\mathbf{Z_I}$ are incidence matrices relating internal phenotypes in $\mathbf{y_I}$ to the vector of fixed effects $\hat{\boldsymbol{\beta}}_\mathbf{I}$ and the vector of random additive genetic effects $\hat{\mathbf{u}}_\mathbf{I}$, respectively, $\mathbf{R_I^{-1}}$ is the inverse of the residual (co)variance matrix, and $\mathbf{D_E^{-1}}$ is the inverse of prediction error (co)variance matrix associated to the vector of external EBV ($\hat{\mathbf{u}}_\mathbf{E}$).

As explained in the previous Chapters, the vector of external EBV for all internal animals and external sires is estimated as $\hat{\mathbf{u}}_\mathbf{E} = \begin{bmatrix} \mathbf{G_{IE} G_{EE}^{-1} \hat{u}_{E(E)}} \\ \hat{\mathbf{u}}_\mathbf{E(E)} \end{bmatrix}$ where $\hat{\mathbf{u}}_\mathbf{E(E)}$ is the vector of external EBV for the external sires obtained from $EVAL_E$ and $\mathbf{G} = \begin{bmatrix} \mathbf{G_{II}} & \mathbf{G_{IE}} \\ \mathbf{G_{EI}} & \mathbf{G_{EE}} \end{bmatrix}$ is the additive genetic (co)variance matrix.

Although the systems of equations related to the two Bayesian evaluations had the same compact notation, the difference between both evaluations concerned the approximation of $\mathbf{D_E^{-1}}$, and especially the computation of the approximation of the least squares part of the LHS of $EVAL_E$, i.e., the computation of the block diagonal matrix $\boldsymbol{\Lambda}$, needed for the approximation of $\mathbf{D_E^{-1}}$, that is $\mathbf{D_E^{-1}} = \mathbf{G^{-1}} + \boldsymbol{\Lambda}$. (4) For the Bayesian evaluation based on the approximation of Quaas and Zhang (2006) ($EVAL_Q$), block diagonals of $\boldsymbol{\Lambda}$, hereafter noted $\boldsymbol{\Lambda_Q}$, related to external animals are equal to $\boldsymbol{\Lambda_{Qi}} \mathbf{G_0^{-1}} \boldsymbol{\Lambda_{Qi}}$ for i = 1, 2, …, N external animals where the matrix $\mathbf{G_0}$ is a matrix of additive genetic

(co)variances between the 2 traits and $\mathbf{\Delta_{Qi}} = \begin{bmatrix} \sqrt{REL_{1i}/(1-REL_{1i})} & 0 \\ 0 & \sqrt{REL_{2i}/(1-REL_{2i})} \end{bmatrix}$ with REL_{1i} and REL_{2i} being REL for the *i*th animal for the first and second trait, respectively. (5) For the Bayesian evaluation based on the approximation proposed in Chapter VI ($EVAL_V$), block diagonals of $\mathbf{\Lambda}$, hereafter noted $\mathbf{\Lambda_V}$, related to external animals were equal to $\mathbf{\Delta_{Vi}R_0^{-1}\Delta_{Vi}}$ for i = 1, 2, …, N external animals where the matrix $\mathbf{R_0}$ is a matrix of residual (co)variances between the two traits and $\mathbf{\Delta_{Vi}} = \begin{bmatrix} \sqrt{RE_{1i}} & 0 \\ 0 & \sqrt{RE_{2i}} \end{bmatrix}$ with RE_{1i} and RE_{2i} being the record equivalents for the *i*th animal for the first and second trait, respectively. For internal animals, associated block diagonals were equal to 0 for both $EVAL_Q$ and $EVAL_V$. It was noted that no phenotype for the correlated trait was considered by $EVAL_I$, $EVAL_Q$ and $EVAL_V$. Only $EVAL_E$ and $EVAL_J$ considered phenotypes for the correlated trait.

Because external sires could be related among them, double counting of contributions due to relationships could exist. Therefore, contributions due to relationships were estimated through the two-step algorithm (TSA; Chapter III) for external sires. The different REL_{1i}, REL_{2i}, RE_{1i} and RE_{2i} estimated by TSA and used for the computation of $\mathbf{\Lambda}$ were therefore assumed free of contributions due to relationships among external animals and due to correlations among traits.

RESULTS AND DISCUSSION

For univariate analyses, computations of $\mathbf{\Lambda_Q}$ and $\mathbf{\Lambda_V}$ are equivalent. Following Chapter III and Chapter VI, it can be written for the *i*th external animal for a univariate genetic evaluation:

$$\mathbf{\Delta_{Vi}R_0^{-1}\Delta_{Vi}} = RE_{1i}\left(\sigma_e^2\right)^{-1} = \frac{REL_{1i}}{1-REL_{1i}} * \frac{\sigma_e^2}{\sigma_a^2}\left(\sigma_e^2\right)^{-1} = \frac{REL_{1i}}{1-REL_{1i}}\left(\sigma_a^2\right)^{-1} = \mathbf{\Delta_{Qi}G_0^{-1}\Delta_{Qi}}$$

where σ_a^2 and σ_e^2 are the additive genetic variance and the residual variance, respectively. However, the equality is not observed for multivariate analyses. Indeed, the block diagonal matrix $\mathbf{\Delta_{Qi}G_0^{-1}\Delta_{Qi}}$ could add unobserved contributions to the elements of the LHS corresponding to the *i*th external animal because the matrices $\mathbf{G_0^{-1}}$ and $\mathbf{R_0^{-1}}$ have not the same structure. For example, if residual covariances are assumed to be equal to zero,

the block diagonal matrix $\mathbf{\Lambda_{Qi}G_0^{-1}\Lambda_{Qi}}$ would add unobserved contributions to the elements of the LHS corresponding to the *i*th external animal, especially through the genetic correlations among traits. Consequences of additional unobserved contributions could lead to overestimated REL and less accurate evaluations. Furthermore, zero residual correlations could be a common situation in the context of integration of correlated external information into a multivariate genetic evaluation because phenotypes used to compute external information would not be observed on internal animals and, therefore, would not be available for internal evaluations. Therefore, zero residual covariances were assumed for the data simulations to illustrate such situations and consequences of the computations of $\mathbf{\Lambda_Q}$ and $\mathbf{\Lambda_V}$ on the Bayesian evaluations.

Regarding the simulated datasets, the 100 replicates included each 2240 animals and all results presented hereafter concern only the trait of interest. The following parameters were explored. Firstly, Spearman's rank correlation coefficients (r) of $EVAL_J$ with $EVAL_I$, $EVAL_Q$ or $EVAL_V$ are presented for external sires (i.e., sires having daughters with records for the correlated trait; Table VIII-1) and for females having records for the trait of interest and sired by the external sires (hereafter called "female progeny"; Table VIII-2). Secondly, the average REL computed from prediction error variances obtained from the inverse of the LHS of $EVAL_J$ and $EVAL_I$, $EVAL_Q$ and $EVAL_V$ (no corrections were performed for inbreeding) are reported for external sires (Table VIII-1) and for female progeny (Table VIII-2). Finally, mean squared errors (MSE) expressed as a percentage of average internal MSE for external sires (Table VIII-1) and for female progeny (Table VIII-2) are also presented. The average internal MSE was expressed on a base of 100. All reported parameters were the averages and the associated standard deviations of the 100 replicates.

On average, 183.0 (±1.1) sires were associated with external information. Firstly, whatever the computation of $\mathbf{\Lambda}$ or the considered r_g, the integration of correlated external information for external sires led to rankings that were more similar to the rankings of $EVAL_J$. For the external sires, r between $EVAL_J$ and $EVAL_I$ varied from 0.987 for r_g = 0.10 to 0.563 for r_g = 0.90. The r of $EVAL_J$ with $EVAL_Q$ or $EVAL_V$ increased to at least 0.990 after integrating correlated external information. Furthermore, MSE also showed that the integration of correlated external information led to better predictions of $EVAL_J$ for the external sires, leading to MSE lower than 3.12% (±0.61), whatever r_g (Table VIII-1). Secondly, when comparing $EVAL_Q$ and $EVAL_V$ for the external sires, no

difference can be observed between both evaluations on average across the 100 replicates for r_g = 0.10 or r_g = 0.25. These results could be expected because r_g, as well as the absolute differences between r_g and $r_{e,}$ were low (Schaeffer, 1984). Differences were observed only from r_g = 0.50. The r of $EVAL_J$ with $EVAL_Q$ or $EVAL_V$ were similar (≥0.990) from r_g = 0.50 to r_g = 0.90. Differences of rankings between $EVAL_Q$ and $EVAL_V$ were lower than 0.008 points for r_g = 0.90. Thereby, differences between $EVAL_Q$ and $EVAL_V$ for external sires associated with external information were mostly observed through MSE and REL. Regarding REL, REL for $EVAL_Q$ were overestimated from 8.98% for r_g = 0.50 to 76.21% for r_g = 0.90 (Table VIII-1). REL for $EVAL_V$ were overestimated from 1.80% for r_g = 0.50 to 2.74% for r_g = 0.90 (Table VIII-1). As shown in Table VIII-1, overestimation of REL for $EVAL_Q$ increased with increasing r_g, and therefore, with the increase of the absolute differences between r_g and r_e. Overestimation of REL for $EVAL_V$ could be attributed to a double counting of contributions due to relationships since the TSA approximated these different contributions (Chapter III; Chapter VI). However, double counting of contributions due to relationships cannot explain the large overestimation observed for $EVAL_Q$. Contrariwise, this large overestimation can be mainly attributed to unobserved contributions added by the product $\mathbf{\Delta_{Qi}G_0^{-1}\Delta_{Qi}}$ for each *i*th external animal. For beef cattle and based on simulated data, Zhang et al. (2002) studied the effects of integration of external information related to three traits into an internal multivariate genetic evaluation evaluating the same three traits. The Bayesian approach was the approach proposed by Quaas and Zhang (2006). External information included sires' EBV and their associated accuracies for the three traits obtained from an external multivariate genetic evaluation. Variance components were assumed to be equal among all evaluations. Unlike our simulations, residual covariances were not equal to zero because the three traits could be observed internally, and also externally, on the same animals. Nevertheless, results of this study also showed that average accuracies associated with the Bayesian evaluation were higher than average accuracies associated with a joint evaluation based on internal and external data, for the three traits. Such an overestimation of accuracies showed that additional and unobserved contributions were integrated into the internal multivariate genetic evaluation. In addition to double counting of contributions due to relationships that was not taken into account in the study of Zhang et al. (2002), the difference between the residual correlations among the three traits (in the range between -0.08 and 0.23) and the genetic correlations (in the

range between 0.31 and 0.50) could also explain these overestimations of accuracies. Furthermore, regarding MSE obtained in the present study, it can be observed that additional and unobserved contributions led to a less accurate, or biased, Bayesian evaluation with a difference of MSE between $EVAL_V$ and $EVAL_Q$ equal to 0.02% for r_g = 0.50 until 2.64% for r_g = 0.90, although average REL for $EVAL_Q$ was higher than average REL for $EVAL_V$ (Table VIII-1).

Table VIII-1. Parameters[1] (SD in parentheses) averaged on 100 replicates and obtained for external sires

Genetic correlations	Parameters	$EVAL_J$	$EVAL_I$	$EVAL_Q$	$EVAL_V$
0.10	r	1.000 (0.000)	0.987 (0.004)	>0.999 (0.000)	>0.999 (0.000)
	REL	0.103 (0.003)	0.100 (0.003)	0.103 (0.003)	0.103 (0.003)
	MSE	-	100.00 (17.28)	0.23 (0.05)	0.23 (0.05)
0.25	r	1.000 (0.000)	0.927 (0.020)	>0.999 (0.000)	>0.999 (0.000)
	REL	0.117 (0.003)	0.100 (0.003)	0.118 (0.003)	0.118 (0.003)
	MSE	-	100.00 (17.09)	0.24 (0.05)	0.24 (0.05)
0.50	r	1.000 (0.000)	0.777 (0.053)	0.999 (0.000)	0.999 (0.000)
	REL	0.167 (0.003)	0.100 (0.003)	0.182 (0.003)	0.170 (0.003)
	MSE	-	100.00 (17.17)	0.31 (0.06)	0.29 (0.06)
0.75	r	1.000 (0.000)	0.634 (0.079)	0.998 (0.001)	0.999 (0.000)
	REL	0.255 (0.003)	0.100 (0.003)	0.344 (0.004)	0.262 (0.003)
	MSE	-	100.00 (17.58)	0.87 (0.19)	0.39 (0.07)
0.90	r	1.000 (0.000)	0.563 (0.091)	0.990 (0.003)	0.998 (0.000)
	REL	0.328 (0.004)	0.100 (0.003)	0.578 (0.005)	0.337 (0.004)
	MSE	-	100.00 (17.65)	3.12 (0.61)	0.48 (0.09)

[1]r = rank correlation between a joint evaluation ($EVAL_J$) and an internal evaluation ($EVAL_I$), a Bayesian evaluation proposed by Quaas and Zhang (2006; $EVAL_Q$) or a Bayesian evaluation proposed in Chapter VI ($EVAL_V$); REL = average reliability; MSE = mean squared error expressed as a percentage of the average internal mean squared error.

As shown in the previous Chapters for univariate Bayesian evaluations, one of the advantages of Bayesian approaches is that external information is propagated to internal animals through the additive genetic (co)variance matrix **G**. This advantage is still observed for multivariate Bayesian evaluations, in addition to the propagation of external information from one trait to the other one, and it can be observed through female progeny of external sires that have records for the trait of interest. On average, external sires sired 241.15 (± 47.06) daughters across the 100 replicates. For the trait of interest, results for 100 replicates showed that r of $EVAL_J$ with $EVAL_Q$ and with $EVAL_V$ ranged from 0.997 for r_g = 0.10 to 0.892 for r_g = 0.90 for female progeny. The rank correlations of $EVAL_J$ with $EVAL_I$ ranged from 0.992 to 0.652. These results showed that integration

of external correlated information for external sires led to rankings of sires' progeny more similar to the rankings of $EVAL_J$, even if r_g is low. Effects of the integration was also observed through MSE which decreased of about 65% for r_g = 0.10 to r_g = 0.90. Comparisons of r and MSE between $EVAL_Q$ and $EVAL_V$ showed that unobserved contributions added by $EVAL_Q$ led to similar r and MSE. However, these unobserved contributions led to higher average REL obtained from $EVAL_Q$ compared to average REL obtained from $EVAL_V$ (Table VIII-2).

Table VIII-2. Parameters[1] (SD in parentheses) averaged on 100 replicates and obtained for female progeny sired by external sires and having records for the trait of interest

Genetic correlations	Parameters	$EVAL_J$	$EVAL_I$	$EVAL_Q$	$EVAL_V$
0.10	r	1.000 (0.000)	0.992 (0.002)	0.997 (0.001)	0.997 (0.001)
	REL	0.138 (0.002)	0.136 (0.002)	0.137 (0.002)	0.137 (0.002)
	MSE	-	100.00 (13.49)	36.48 (5.75)	36.48 (5.75)
0.25	r	1.000 (0.000)	0.954 (0.009)	0.983 (0.003)	0.983 (0.003)
	REL	0.144 (0.002)	0.136 (0.002)	0.141 (0.002)	0.141 (0.002)
	MSE	-	100.00 (13.31)	36.34 (5.77)	36.34 (5.77)
0.50	r	1.000 (0.000)	0.844 (0.029)	0.947 (0.010)	0.946 (0.010)
	REL	0.169 (0.002)	0.136 (0.002)	0.157 (0.002)	0.154 (0.002)
	MSE	-	100.00 (13.46)	36.03 (5.88)	36.05 (5.88)
0.75	r	1.000 (0.000)	0.721 (0.048)	0.910 (0.017)	0.910 (0.017)
	REL	0.212 (0.002)	0.136 (0.002)	0.197 (0.003)	0.177 (0.002)
	MSE	-	100.00 (14.31)	35.74 (6.05)	35.65 (6.04)
0.90	r	1.000 (0.000)	0.652 (0.057)	0.892 (0.020)	0.892 (0.021)
	REL	0.247 (0.003)	0.136 (0.002)	0.255 (0.004)	0.195 (0.003)
	MSE	-	100.00 (15.18)	36.46 (6.16)	35.34 (6.10)

[1]r = rank correlation between a joint evaluation ($EVAL_J$) and an internal evaluation ($EVAL_I$), a Bayesian evaluation proposed by Quaas and Zhang (2006; $EVAL_Q$) or a Bayesian evaluation proposed in Chapter VI ($EVAL_V$); REL = average reliability; MSE = mean squared error expressed as a percentage of the average internal mean squared error.

Based on these results, the proposed Bayesian approaches seem to well integrate the correlated external information into a multivariate genetic evaluation, especially if the block diagonal matrix was $\mathbf{\Lambda_V}$ (Chapter VI) instead of $\mathbf{\Lambda_Q}$ (Quaas and Zhang, 2006). Better predictions of EBV and improvement of REL were observed for external animals and for their progeny in both approaches. However, the better results obtained for $EVAL_V$, in comparison to $EVAL_Q$, must be confirmed with real data and other approximations of $\mathbf{\Lambda}$ (e.g., Schaeffer, 2001) must be tested. Furthermore, it is worth noting that different units of measurement were assumed between the two traits, as well as different heritabilities and different variance components. Therefore, like MACE, the proposed

Bayesian approach allows the combination of different sources of information associated to different scales, different heritabilities and variance components, or different units of measurement. Also, multi-trait Bayesian evaluations allow to consider genotype by environment interactions and different trait definitions associated with the internal and external evaluations. This latter characteristic permits to avoid conversion of external information (e.g., expressed as indices or standardized estimates) to the expression of the random animal effects considered by the internal evaluations. For example, in the context of dairy cattle and for numerous countries, milk yield is a trait published as an average yield on 305 days for three lactations while models are test-day models, possibly considering random regressions for additive genetic effects (Gengler and Vanderick, 2008; Liu et al., 2014). In cases of random regression test-day models, Gengler and Vanderick (2008) proposed to convert external indices which express average yields on 305 days for three lactations to (one of) the random regression additive genetic effects for one or several lactations. Alternatively, external information can be considered as an additional trait which has no observed phenotype internally and which is genetically correlated to all the random regression additive genetic effects for all lactations, similarly to the present study. Such a strategy could be considered for integrating MACE information into single-step genomic evaluations.

Regarding the correlations, it was assumed that residual and genetic variances of the traits as well as the residual and genetic correlations between external information and internal data were known. However, in practice, variances and correlations are not known and must be estimated. Therefore, approaches to estimate variances and correlations associated with internal data and external information must be studied and developed. Also, it is worth noting that, although a diagonal residual (co)variance matrix was assumed for this simulation, off-diagonal elements different from zero could be considered, for example, if phenotypes for the different traits were observed on the same internal animals, as shown by Zhang et al. (2002).

CONCLUSIONS

Based on these results, the proposed Bayesian approaches integrated well correlated external information into a multivariate genetic evaluation for simulated data, especially if the block diagonal matrix was $\mathbf{\Lambda_V}$ (Chapter VI) instead of $\mathbf{\Lambda_Q}$ (Quaas and Zhang, 2006). Therefore, the proposed multivariate Bayesian approaches have the potential to integrate correlated external information into a multivariate genetic evaluation

allowing for different heritabilities, variance components, units of measurement or models between external and internal traits. However, while the multivariate Bayesian approaches seem promising, their implementation could be difficult due to availability or estimation of correlations between external information and internal data. Hence, further research on real data is needed to confirm these first results.

COMPARISON OF APPROACHES THAT COMBINE INTERNAL DATA AND EXTERNAL INFORMATION

Three types of approaches allow to combine simultaneously internal data and external information, i.e., EBV and associated REL:

(1) absorption based approaches (e.g., Henderson, 1975; Van Vleck, 1982; Bolgiano et al., 1983),

(2) pseudo-records based approaches (e.g., Bonaiti and Boichard, 1995; VanRaden et al., 2014), and

(3) Bayesian approaches (Chapter III, Chapter VI).

These approaches were described in the previous Chapters and some similarities among some of them were already noted (Chapter II). Therefore, the aim of the present section is to make pairwise comparisons of the three approaches to identify similarities and differences among them. As reviewed in Chapter II, different implementations were proposed for all the three approaches, and from those, the implementations proposed by Bolgiano et al. (1983), by VanRaden et al. (2014) and in Chapter VI were chosen.

The context of the comparisons is an internal univariate evaluation integrating one source of external information associated with external animals that have no observations at the internal level. This assumption concerning external animals and internal records was taken for simplification. Nevertheless, results of the comparison would be identical without this assumption. Such a context can be observed, for example, for a dairy cattle genetic evaluation for milk production traits that integrates external information for bulls only (e.g., Bolgiano et al., 1983). Also, all animals are assumed to be partitioned between internal animals (I) and external animals (E). Thereby, in this context, a system of equations for the internal genetic evaluation combining internal data and external information can be written for the three approaches with the same compact notation as follows:

$$\begin{bmatrix} \mathbf{X_I'X_I} & \mathbf{X_I'Z_I} & \mathbf{0} & \mathbf{X_I'Z_I} \\ \mathbf{Z_I'X_I} & \mathbf{Z_I'Z_I'} + \mathbf{A^{II}}\lambda & \mathbf{A^{IE}}\lambda & \mathbf{Z_I'Z_I} \\ \mathbf{0} & \mathbf{A^{EI}}\lambda & \mathbf{A^{EE}}\lambda + \mathbf{\Psi} & \mathbf{0} \\ \mathbf{Z_I'X_I} & \mathbf{Z_I'Z_I} & \mathbf{0} & \mathbf{Z_I'Z_I} + \mathbf{I}\alpha \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_\mathbf{I} \\ \hat{\mathbf{u}}_\mathbf{I} \\ \hat{\mathbf{u}}_\mathbf{E} \\ \hat{\mathbf{p}}_\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{X_I'y_I} \\ \mathbf{Z_I'y_I} \\ \boldsymbol{\theta} \\ \mathbf{Z_I'y_I} \end{bmatrix}, \qquad \text{(VIII.1)}$$

where $\mathbf{y_I}$ is the vector of internal observations associated with the internal animals, $\boldsymbol{\beta}_\mathbf{I}$ is the vector of fixed effects, $\mathbf{u_I}$ and $\mathbf{u_E}$ are the vectors of random additive genetic effects of internal and external animals, respectively, $\mathbf{p_I}$ is the vector of random permanent environmental effects and $\mathbf{e}$ is the vector of residuals. $\mathbf{X_I}$ and $\mathbf{Z_I}$ are incidence matrices relating internal observations to fixed effects and to random effects, respectively, $\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A^{II}} & \mathbf{A^{IE}} \\ \mathbf{A^{EI}} & \mathbf{A^{EE}} \end{bmatrix}$ is the inverse of the relationships matrix, $\lambda = (1-r)/h^2$ and $\alpha = (1-r)/(r-h^2)$ where r and h^2 are repeatability and heritability, respectively. The matrix $\mathbf{\Psi}$ and the vector $\boldsymbol{\theta}$ must be defined following the considered approach, that are the absorption based approach (Bolgiano et al., 1983), the pseudo-records based approach (VanRaden et al., 2014) and the Bayesian approach (Chapter VI).

ABSORPTION BASED APPROACHES AND PSEUDO-RECORDS BASED APPROACHES

Concerning the absorption based approach, the system of equations (VIII.1) has the same compact notation as the system of equations [2] proposed by Bolgiano et al. (1983). Following their definitions and after some simplifications, the diagonal element of the diagonal matrix $\mathbf{\Psi}$ for the *i*th animal, $\mathbf{\Psi}_{ii}$, is equal to $\mathbf{\Psi}_{ii} = n_i * \lambda/\lambda_s$ where n_i is the number of daughters of the *i*th external animal and which provide equivalent information with one record per daughter and all in the same herd-year-season. The element of the vector $\boldsymbol{\theta}$ corresponding to the *i*th animal, $\boldsymbol{\theta}_i$, is equal to $\boldsymbol{\theta}_i = \frac{(1-r)}{(4-h^2)h^2}(4+(n_i-1)h^2)\boldsymbol{\mu}_{\mathbf{E}_i} = (n_i * \lambda/\lambda_s + \lambda)\boldsymbol{\mu}_{\mathbf{E}_i}$ where $\boldsymbol{\mu}_\mathbf{E}$ is the vector of external EBV associated with the external animals (Henderson, 1975; Table VIII-3).

Also, the system of equations for univariate analyses (VIII.1) can be applied for the pseudo-records based approach proposed by VanRaden et al. (2014). Following their definitions and after some simplifications, the diagonal element of the diagonal matrix $\mathbf{\Psi}$ for the *i*th external animal is equal to $\mathbf{\Psi}_{ii} = \mathbf{RE}_i$ where $\mathbf{RE}_i$ is the element for the *i*th animal of the vector of RE associated with external animals. The element of the vector $\boldsymbol{\theta}$

for the *i*th external animal for the pseudo-records based approach, $\boldsymbol{\theta}_i$, is equal to $\boldsymbol{\theta}_i = \mathbf{RE}_i * \mathbf{DRP}_i$ where $\mathbf{DRP}_i$ is the element for the *i*th animal of the vector of the deregressed proofs (DRP) associated with external animals (VanRaden et al., 2014; Table VIII-3).

Therefore, considering these definitions, the two approaches are equivalent under some assumptions. Firstly, the diagonal elements of $\boldsymbol{\Psi}$ are equivalent assuming $n_i = DE_i$ where $DE_i = RE_i * \lambda_s / \lambda$ is the daughter equivalents which are defined as the number of daughters of the *i*th external animal which provide equivalent information with one record per daughter, all daughters having an infinite number of management group mates and the other parent with perfect REL (VanRaden and Wiggans, 1991). Secondly, the elements of $\boldsymbol{\theta}$ are equivalent assuming $\mathbf{DRP}_i = (1 + \lambda / \mathbf{RE}_i)\boldsymbol{\mu}_{\mathbf{E}_i}$. It is noted that the equivalence between absorption based approaches and pseudo-records based approaches, under some assumptions, was already shown by Bonaiti and Boichard (1995). However, in practice, differences may appear between results of both approaches if the previous assumptions are not verified (e.g., VanRaden et al., 2014).

Table VIII-3. Non-zero elements[1] of the matrix $\boldsymbol{\Psi}$ and the vector $\boldsymbol{\theta}$ for the *i*th external animal that is not associated with records at the internal level in the context of a univariate evaluation, following the absorption based approach (Bolgiano et al. 1983), the Bayesian based approach (Chapter VI) and the pseudo-records based approach (VanRaden et al. 2014)

Approaches	$\boldsymbol{\Psi}_{ii}$	$\boldsymbol{\theta}_i$
Absorption based approach	$n_i * \lambda / \lambda_s$	$(n_i * \lambda / \lambda_s + \lambda)\boldsymbol{\mu}_{\mathbf{E}_i}$
Bayesian based approach	$\boldsymbol{\Lambda}_{\mathbf{E}_{ii}} \sigma_e^2$	$\left(\boldsymbol{\Lambda}_{\mathbf{E}_{ii}} \sigma_e^2 + \mathbf{A}^{-1}_{\mathbf{EE}_{ii}} \lambda\right)\boldsymbol{\mu}_{\mathbf{E}_i} + \lambda \sum_{j=1, j \neq i}^{N} \mathbf{A}^{-1}_{\mathbf{EE}_{ij}} \boldsymbol{\mu}_{\mathbf{E}_j}$
Pseudo-records based approach	$\mathbf{RE}_i$	$\mathbf{RE}_i * \mathbf{DRP}_i$

[1] n_i = number of daughters of the *i*th external animal and which provide equivalent information with one record per daughter and all in the same herd-year-season; λ = ratio of residual to genetic variance ; λ_s = ratio of error to sire variance; $\boldsymbol{\mu}_{\mathbf{E}}$ = vector of external estimated breeding values associated with the external animals; $\boldsymbol{\Lambda}_{\mathbf{E}}$ = diagonal matrix; σ_e^2 = residual variance; $\mathbf{A}^{-1}_{\mathbf{EE}}$ = inverse of the relationships matrix among external animals; **RE** = vector of record equivalents associated with external animals; **DRP** = vector of deregressed proofs associated with external animals.

ABSORPTION BASED APPROACHES AND BAYESIAN APPROACHES

In the context defined previously, based on the assumptions specified for the

Bayesian approach defined in Chapter VI and by multiplying the system of equations (IV.3) by the residual variance σ_e^2, the system of equations (IV.3) has an equivalent compact notation than the system of equations (VIII.1). Thereby, following the definitions given in Chapter VI and after some simplifications, the diagonal element for the *i*th animal of the diagonal matrix $\mathbf{\Psi}$, $\mathbf{\Psi}_{ii}$, is equal to $\mathbf{\Psi}_{ii} = \mathbf{\Lambda}_{\mathbf{E}_{ii}}\sigma_e^2 = \mathbf{RE}_i$ for the Bayesian approach, with $\mathbf{\Lambda}_{\mathbf{E}}$ being a diagonal matrix with a diagonal element for each *i*th animal equal to $\mathbf{\Lambda}_{\mathbf{E}_{ii}} = \mathbf{RE}_i\left(\sigma_e^2\right)^{-1}$ (Chapter VI). This demonstrated the equivalence between the absorption based approach (Henderson, 1975; Bolgiano et al., 1983) and the Bayesian approach (Chapter VI) for the computation of $\mathbf{\Psi}$, assuming $n_i = DE_i = \mathbf{RE}_i * \lambda_s / \lambda$. Considering the vector $\boldsymbol{\theta}$ for the Bayesian approach, after some developments, the element of $\boldsymbol{\theta}$ for the *i*th external animal is equal to (Chapter VI; Table VIII-3):

$$\boldsymbol{\theta}_i = \left(\mathbf{RE}_i + \mathbf{A}^{-1}_{\mathbf{EE}_{ii}}\lambda\right)\boldsymbol{\mu}_{\mathbf{E}_i} + \lambda \sum_{j=1,\, j\neq i}^{N} \mathbf{A}^{-1}_{\mathbf{EE}_{ij}}\boldsymbol{\mu}_{\mathbf{E}_j} \qquad \text{(VIII.2)}$$

where $\mathbf{A}^{-1}_{\mathbf{EE}}$ is the inverse of the relationships matrix among external animals and j = 1, 2, …, N refers to the *j*th external animal different from the *i*th animal.

The term $\lambda \sum_{j=1,\, j\neq i}^{N} \mathbf{A}^{-1}_{\mathbf{EE}_{ij}}\boldsymbol{\mu}_{\mathbf{E}_j}$ of the equation (VIII.2) counts for the non-zero off-diagonal elements for the *i*th external animal and, thereby, takes the contributions due to relationships of other external animals into account. Equivalent vectors $\boldsymbol{\theta}$ are then computed by the absorption based approach and the Bayesian approach if external animals are assumed unrelated, i.e., $\mathbf{A}_{\mathbf{EE}} = \mathbf{I}$, and if $n_i = \mathbf{RE}_i * \lambda_s / \lambda$.

PSEUDO-RECORDS BASED APPROACHES AND BAYESIAN APPROACHES

From the previous comparisons, it is noted that the pseudo-records based approach (VanRaden et al., 2014) and the Bayesian approach (Chapter VI) estimate equivalently the matrix $\mathbf{\Psi}$ for univariate cases, i.e., for the *i*th external animal, $\mathbf{\Psi}_{ii} = \mathbf{RE}_i$ for the pseudo-records based approach and $\mathbf{\Psi}_{ii} = \mathbf{\Lambda}_{\mathbf{E}_{ii}}\sigma_e^2 = \mathbf{RE}_i$ for the Bayesian approach (Table VIII-3). The equivalence between the two approaches can also be shown for the vector $\boldsymbol{\theta}$ under some assumptions. Let the vector **DRP** be the vector including DRP associated with the external animals and computed following the procedure proposed by Jairath et al. (1998). To be in agreement with the system of equations (VIII.1), the procedure for computing DRP is hereafter extended to an animal model and the phantom group effect is

ignored (e.g., Calus et al., 2014). Therefore, after convergence and following the second equation of the system of equations proposed by Jairath et al. (1998) extended to an animal model, the vector $\mathbf{DRP}$ is equal to:

$$\mathbf{DRP} = \mathbf{D_d^{-1}}\left(\mathbf{D_d}\mathbf{1}m + \left(\mathbf{A^{EE}}\lambda + \mathbf{D_d}\right)\left(\boldsymbol{\mu}_\mathbf{E} - \mathbf{1}m\right) + \mathbf{A^{EI}}\lambda\boldsymbol{\mu}_\mathbf{I}\right) \qquad \text{(VIII.3)}$$

where m is the overall mean, $\mathbf{D_d}$ is a diagonal matrix with elements equal to the RE associated with the external animals (i.e., $\mathbf{D_d} = diag\left(\mathbf{RE}\right) = \boldsymbol{\Lambda}_\mathbf{E}\sigma_e^2$) and $\boldsymbol{\mu}_\mathbf{I}$ is the vector of external EBV for internal animals estimated through the Jairath's procedure.

From the third equation of the system of equations proposed by Jairath et al. (1998) extended to an animal model, it can be shown that $\boldsymbol{\mu}_\mathbf{I} = -\left(\mathbf{A^{II}}\right)^{-1}\mathbf{A^{IE}}\left(\boldsymbol{\mu}_\mathbf{E} - \mathbf{1}m\right)$ and the equation (VIII.3) can be simplified as $\mathbf{DRP} = \mathbf{D_d^{-1}}\left(\mathbf{D_d}\mathbf{1}m + \left(\mathbf{A_{EE}^{-1}}\lambda + \mathbf{D_d}\right)\left(\boldsymbol{\mu}_\mathbf{E} - \mathbf{1}m\right)\right)$. Therefore, assuming $m = 0$ and after simplifications, the element of the vector $\boldsymbol{\theta}$ for the *i*th external animal is equal to:

$$\boldsymbol{\theta}_i = \mathbf{RE}_i * \mathbf{DRP}_i = \left(\mathbf{RE}_i + \mathbf{A}^{-1}_{\mathbf{EE}_{ii}}\lambda\right)\boldsymbol{\mu}_{\mathbf{E}_i} + \lambda\sum_{j=1,\, j\neq i}^{N}\mathbf{A}^{-1}_{\mathbf{EE}_{ij}}\boldsymbol{\mu}_{\mathbf{E}_j}\,,$$

which demonstrates the equivalence between the pseudo-records based approach and the Bayesian approach for univariate analyses, under some assumptions. However, differences among the results of both approaches may be observed in practice, especially by using another procedure for the computation of DRP (e.g., VanRaden et al., 2014).

Unlike the absorption based approach described in the previous subsections, both the pseudo-records based approach (VanRaden et al., 2014) and the Bayesian approach (Chapter VI) were extended to multivariate analyses. While it was not discussed in Chapter V, conceptual equivalence between the pseudo-records based approaches and the Bayesian approaches for multivariate analyses can be developed from Chapter V by assuming that $\mathbf{R_P} = \boldsymbol{\Lambda}_\mathbf{E}^{-1}$ at the step 4) of the approach computing the pseudo-records and the user-supplied (co)variance matrix in Chapter V. For general cases, it was originally proposed in Chapter V to compute $\mathbf{R_P}$ as $\mathbf{R_P} = \mathbf{Z_P}\left(\mathbf{I} \otimes \mathbf{R_0}\right)\mathbf{Z_p}$ where $\mathbf{I}$ is an identity matrix of size k equal to the number of records for each trait and $\mathbf{R_0}$ is the residual (co)variance matrix between traits for 1 record. The computation of $\mathbf{R_P}$ as $\mathbf{R_P} = \boldsymbol{\Lambda}_\mathbf{E}^{-1}$ rises that the steps 2) and 5) can be considered as a procedure that deregresses EBV, similarly to Jairath et al. (1998) and Schaeffer (2001), and needed by the pseudo-records based approach (e.g., VanRaden et al., 2014). However, in practice, both approaches

would lead to different results, mainly because they compute differently the block diagonal matrix $\boldsymbol{\Psi}$. VanRaden et al. (2014) proposed to compute the block diagonals of $\boldsymbol{\Psi}$ for the *i*th external animal as $\boldsymbol{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}}\mathbf{G}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\Delta}_{\mathbf{VRi}}^{\mathbf{0.5}}$ where $\boldsymbol{\Delta}_{\mathbf{VR}}$ is a diagonal matrix with a diagonal element for each *j*th trait equal to $REL_{ij}/\left(1-REL_{ij}\right)$ and $\mathbf{G_0}$ is the matrix of genetic (co)variances among traits. Nevertheless, it was proposed in Chapter VI to compute the block diagonals of $\boldsymbol{\Psi}$ (i.e., the matrix $\boldsymbol{\Lambda}_{\mathbf{V}}$ in the previous subsection entitled "Integration of correlated external information"), for the *i*th external animal, as $\boldsymbol{\Delta}_{\mathbf{Vi}}\mathbf{R}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\Delta}_{\mathbf{Vi}}$, where the matrix $\mathbf{R_0}$ is a matrix of residual (co)variances among traits and $\boldsymbol{\Delta}_{\mathbf{V}}$ is a diagonal matrix with a diagonal element for each *j*th trait equal to RE_{ij}. Because the approaches proposed by VanRaden et al. (2014) and by Quaas and Zhang (2006) to compute $\boldsymbol{\Psi}$ are identical, the consequences of the different computations of $\boldsymbol{\Psi}$ between the pseudo-records based approach and the Bayesian approach were already discussed in the previous subsection entitled "Integration of correlated external information".

CONCLUSIONS

The three approaches that combine simultaneously external information and internal phenotypic and pedigree data were compared. Equivalences among the three approaches, especially for univariate analyses, were observed under some assumptions. These assumptions concern, e.g., the definitions of the weights associated with external information (e.g., n_i, DE_i, $\mathbf{RE}_i$), their consideration by the genetic evaluations, and the used deregression steps. However, most of these assumptions are not fulfilled in practice and results may differ among the three approaches. Also, with regard to their implementation, it is worth noting that the absorption based approaches could be difficult to generalize for complex models (Quaas and Zhang, 2006), while pseudo-records based approaches and Bayesian approaches propose easy adaptations to complex models, such as multivariate models (e.g., Chapter VI; VanRaden et al., 2014). Also, the pseudo-records based approach proposed by VanRaden et al. (2014) can be easily applied with software packages available in animal breeding community. However, this approach requires a deregression step, which is not a trivial problem (Chapter II), as well as the explicit computation of external information free of internal information (VanRaden et al., 2014) when external information includes internal information. The Bayesian approaches proposed in Chapter VI avoid deregression steps as well as explicit computations of

external information free of internal information. Computational burden is then simplified and risks of potential computational errors propagated through the different steps performed before the evaluations are avoided. Also, Chapter V proposed a method to perform the Bayesian approaches using currently available software packages.

REFERENCES

Bolgiano, D.C., R.L. Quaas, and L.D. Van Vleck. 1983. Variances of prediction errors for differences of estimated genetic values of cows in the same and in different herds. *J. Dairy Sci.* 66:1370–1376.

Bonaiti, B., and D. Boichard. 1995. Accounting for foreign information in genetic evaluation. *Interbull Bull.* 11:4pp.

Calus, M., G. Schopen, and R. Veerkamp. 2014. Simultaneous deregression of cow and bull breeding values. *Interbull Bull.* 48:63–67.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Henderson, C.R. 1975. Use of all relatives in intraherd prediction of breeding values and producing abilities. *J. Dairy Sci.* 58:1910–1916.

Jairath, L., J.C.M. Dekkers, L.R. Schaeffer, Z. Liu, E.B. Burnside, and B. Kolstad. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81:550–562.

Legarra, A., J.K. Bertrand, T. Strabel, R.L. Sapp, J.P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295.

Liu, Z., M.E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* 97:5833–5850.

Mrode, R.A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publishing, Wallingford, UK.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Sargolzaei, M., and F.S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25:680–681.

Schaeffer, L.R. 1984. Sire and cow evaluation under multiple trait models. *J. Dairy Sci.* 67:1567–1580.

Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.

Schaeffer, L.R. 2001. Multiple trait international bull comparisons. *Livest. Prod. Sci.* 69:145–153.

VanRaden, P.M., M.E. Tooker, J.R. Wright, C. Sun, and J.L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952-7962.

VanRaden, P.M., and G.R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737–2746.

Van Vleck, L.D. 1982. Derivation of Henderson's method of incorporating artificial insemination sire evaluations into intraherd prediction of breeding values. *J. Dairy Sci.* 65:284–286.

Zhang, Z.W., R.L. Quaas, and E.J. Pollak. 2002. Simulation study on the effects of incorporating external genetic evaluation results. CD–ROM Comm. 20–14 in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Chapter IX. IMPLICATIONS, FUTURE RESEARCH TOPICS AND GENERAL CONCLUSION

The last Chapter of this thesis presents implications of the research presented in the previous Chapters and introduces four topics for future research. A general conclusion closes this Chapter.

IMPLICATIONS

Research undertaken during this thesis led to the development of a Bayesian approach that integrates several sources of external information into a genetic or a single-step genomic evaluation. This Bayesian approach can be easily adapted to complex models and considers double counting of contributions due to relationships and due to records. Therefore, these results allowed for different advances in genetic and genomic evaluations.

Another implication of the present thesis is the development of a genomic evaluation system for Holstein cattle in the Walloon Region of Belgium. The Walloon genomic evaluation presents several advanced features and combines simultaneously all available genotypes, pedigree, Walloon and multiple across country evaluation (MACE) information (i.e., estimated breeding values (EBV) and reliabilities (REL) provided by Walloon evaluations and MACE) for the milk, fat and protein yields, as well as for other traits, like somatic cell score. Contributions considered by both Walloon evaluations and MACE are also considered by the Walloon genomic evaluation in order to avoid their double counting. The development of the Walloon genomic evaluation system allowed the Walloon Region of Belgium to participate to the international genomic evaluations performed by International Bull Service (Interbull, Uppsala, Sweden).

FUTURE RESEARCH TOPICS

Several important research topics were identified during this thesis and shall be explored in the future. These topics address different issues:

ON THE ESTIMATION OF CONTRIBUTIONS DUE TO RELATIONSHIPS

An iterative algorithm that estimates the contributions due to relationships included in the external information (so-called two-step algorithm; TSA) was proposed in Chapter III. Several applications on different simulated and real data (Chapter III, Chapter IV, Chapter VI and Chapter VII) showed that the TSA performed well. However, an issue of the TSA is that each iteration needs the inversion of a matrix that is a function of the relationship matrix that accounts for the relationships between external animals and their ancestors. For the Walloon genomic evaluations, a Fortran implementation of the TSA using the multithreaded Intel® Math Kernel Library (Intel® MKL) allows the estimation of contributions due to relationships for around 21 000 external animals using 12 threads

and around 50 GB of RAM during less than 5 hours on a cluster made of 26 computing nodes, each with two Intel Sandy Bridge 8-cores E5-2670 processors at 2.6 GHz and with 128 GB of RAM. However, such an inversion could lead to a difficult implementation of the TSA if a higher number of animals is associated with external information or if an adequate cluster is not available. Because the TSA is a fixed point algorithm, an acceleration procedure could be used to accelerate the convergence (Brezinski and Chehab, 1998) and to avoid some iterations and, therefore, also some matrix inversions. Although tested only on simulated data with the software package GNU Octave (Eaton et al., 2011), the implementation of the method of Lemaréchal (Brezinski and Chehab, 1998) in the TSA allowed for a faster convergence. However, such acceleration procedures do not solve the issue of the matrix inversion when a high number of animals are associated to external information. Therefore, an algorithm that avoids matrix inversions and that estimates the diagonal elements of an unknown matrix $\mathbf{D}$ involved in the following equation should be developed:

$$\left(\mathbf{G}^{-1} + \mathbf{D}\right)^{-1} = \mathbf{P}$$

where $\mathbf{G}^{-1}$ is the inverse of a known positive-definite symmetric matrix, $\mathbf{D}$ is an unknown diagonal matrix (for the first step of the TSA) or an unknown positive-definite symmetric matrix (for the second step of the TSA) and $\mathbf{P}$ is a positive-definite symmetric matrix for which only the diagonal elements are known.

Finally, pending the development of an algorithm estimating the diagonal elements of $\mathbf{D}$ without matrix inversions, the results from the simulation in Chapter III as well as from the Walloon example in Chapter VI suggest that contributions due to relationships could be ignored. However, it was worth noting that effects of double counting of contributions due to relationships should be tested before ignoring them.

ON THE INTEGRATION OF CORRELATED EXTERNAL INFORMATION

A simulation study on the integration of correlated external information into a multivariate genetic evaluation was proposed in Chapter VIII. These first results showed that the proposed Bayesian approaches integrated well correlated external information, based on rank correlations, mean squared errors and average REL. It was also shown that the approximation of the least squares part of the left hand side (LHS) of the external evaluation proposed by Quaas and Zhang (2006) led to overestimated REL because unobserved contributions were considered. However, all these observations were based on

a simulation study and may be influenced by the different assumptions, as the consideration of known (co)variances or low average REL. Indeed, residual and genetic (co)variances among traits are usually unknown and covariances among EBV increase from residual to approach genetic covariances as REL increase (VanRaden et al., 2014). Therefore, studies on real data should be performed to confirm these first results and other approximations of the least squares part of the LHS of the external evaluation (e.g., Schaeffer, 2001) should be tested.

ON THE ESTIMATION OF CORRELATIONS BETWEEN EXTERNAL INFORMATION AND INTERNAL DATA

One strong assumption was taken in Chapter VIII: all variance components were assumed to be known. In practice, this is usually not the case and (co)variances must be estimated. For linear mixed models that assume expectations equal to zero for random effects, Gibbs sampling (e.g., Sorensen and Gianola, 2002) and restricted maximum likelihood (Patterson and Thompson, 1971) are the most popular methods for the estimation of variance components. However, current programs do not allow to estimate of variance components for linear mixed models that assume expectations different from zero for random effects. Therefore, approaches should be developed to estimate variance components for (multivariate) mixed models considering both internal data and external information. Because the equivalence between pseudo-records based approaches and Bayesian approaches was shown, at least under some assumptions (Chapter VIII), an approach could consist of estimating the variance components for a mixed model including external information as weighted pseudo-records. Another approach could consist of estimating variance components with current software modified to consider expectations different form zero and non-conventional (co)variance matrices (e.g., the sum of (co)variance matrices (e.g., Chapter III, Chapter VI) instead of the conventional genetic (co)variance matrix) associated with random effects.

ON THE INTEGRATION OF GENOMIC INFORMATION

With the development of genomic selection in many species and the increase of available SNP data and information derived from this data source (e.g., direct genomic values (DGV) and their associated REL) approaches and algorithms that combine sources of phenotypic, genealogical and genomic data and information are needed. In this context, Chapter VII proposed a Bayesian approach, based on the single-step genomic evaluation

(Aguilar et al., 2010; Christensen and Lund, 2010) to combine genomic data with internal and external information. However, this thesis does not propose approaches to integrate external genomic information, e.g., DGV and associated REL, into an internal genetic evaluation, while it was previously suggested and studied in a Bayesian context (Gengler and Vanderick, 2008; Hyde et al., 2013). Such situation may arise in different emerging situations as, for example, through the wish to integrate externally generated DGV for novel phenotypes (e.g., dry matter intake, methane). However, several issues should be further explored, such as the double counting of the same information used in both the genetic evaluation and in the estimation of genomic prediction equations, the scaling of DGV, or the consideration that DGV follow, or does not follow, the same distribution as internal EBV (Mäntysaari and Strandén, 2010; Hyde et al., 2013).

GENERAL CONCLUSION

The aim of this thesis was to develop algorithms to combine phenotypic, genealogical and genomic data as well as information originating from diverse sources and to test them on simulated and real data. After a review of the various proposed approaches to combine different sources of information, it was chosen to focus on the Bayesian approaches, based on a Bayesian view of the linear mixed models. Research presented in this thesis solved different issues to finally develop equations for (multivariate) genetic and single-step genomic evaluations that integrate and blend simultaneously several sources of information and that avoid double counting of contributions due to relationships and due to records. Computational burden was also considered during this research. The performance of the developed algorithms and equations were evaluated using simulated and real datasets. The different results showed that:

- the developed equations integrated and blended several sources of information in a proper way into a genetic or a single-step genomic evaluation,
- more reliable EBV were obtained for external animals after integration of external information,
- relatives of external animals benefited from integrated external information, also leading to more reliable EBV, because all available external sources of information were correctly propagated,
- double counting of contributions due to relationships and due to records were (almost) avoided, and

- correlated external information was properly integrated following the approach proposed to approximate the prediction error (co)variance matrix associated with multivariate EBV.

The developed equations were applied to develop a genomic evaluation system for Holstein cattle in the Walloon Region of Belgium that combines simultaneously all available genotypes, pedigree, Walloon and MACE information (i.e., EBV and REL provided by Walloon evaluations and MACE) for the production traits, as well as for other traits, like somatic cell scores. However, in spite of these developments, further research should be carried out, especially, on the estimation of contributions due to relationships, on the integration of correlated external information, on the estimation of correlations between external information and internal data, and on the integration of genomic information, such as DGV.

REFERENCES

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Brezinski, C., and J. Chehab. 1998. Nonlinear hybrid procedures and fixed point iterations. *Numer. Funct. Anal. Optim.* 19:465–487.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.

Eaton, J.W., D. Bateman, S. Hauberg, and R. Wehbring. 2011. GNU Octave. A high-level interactive language for numerical computations. 3rd ed. Free Software Foundation, Inc., Boston, MA.

Gengler, N., and S. Vanderick. 2008. Bayesian inclusion of external evaluations into a national evaluation system: Application to milk production traits. *Interbull Bull.* 38:70–74.

Hyde, L.R., W.R. Shafer, S.C. McGuire, M. Saatchi, and D.J. Garrick. 2013. ASA experience with incorporating genomics into genetic evaluation. Pages 10–14 in Proceedings of the 10th Genetic Prediction Workshop. Beef Improvement Federation, Kansas City, MO.

Mäntysaari, E.A., and I. Strandén. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluations. CD–ROM Comm. 353 in

Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554.

Quaas, R.L., and Z. Zhang. 2006. Multiple-breed genetic evaluation in the US beef cattle context: Methodology. CD–ROM Comm. 24–12 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

Schaeffer, L.R. 2001. Multiple trait international bull comparisons. *Livest. Prod. Sci.* 69:145–153.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer-Verlag, New York, NY.

VanRaden, P.M., M.E. Tooker, J.R. Wright, C. Sun, and J.L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952–7962.

LIST OF ABBREVIATIONS

AI	Artificial insemination
BLUE	Best Linear Unbiased Estimate
BLUP	Best Linear Unbiased Prediction
DE	Daughter equivalent
DGV	Direct genomic value
DNA	Deoxyribonucleic acid
DRP	Deregressed proof
DYD	Daughter yield deviation
EBV	Estimated breeding value
EDC	Effective daughter contribution
EGM	Estimate of genetic merit for animals (e.g., EBV) or for what they transmit to their progeny (e.g., PTA)
FBE	First version of modified Bayesian evaluation
GEBV	Genomically enhanced breeding value
GMACE	Genomic multiple across country evaluation
$\mathbf{h^2}$	Heritability
LBE	Legarra-type Bayesian evaluation
LHS	Left hand side
MACE	Multiple across country evaluation
MB	Mean biais
MCE	Multiple-country evaluation
MME	Mixed model equations
MSE	Mean squared error
MSEP	Mean squared errors of prediction
MT-MACE	Multiple-trait multiple across country evaluation
MVN	Multivariate normal
PA	Parent average
PC	Progeny contribution
PEV	Prediction error variance
PL	Productive life
PTA	Predicted transmitting ability

QBE	Quaas-type Bayesian evaluation
RE	Records equivalent
$\mathbf{r_e}$	Residual correlation between traits
REL	Reliability
$\mathbf{r_g}$	Genetic correlation between traits
RHS	Right hand side
$\mathbf{r_{y:\hat{y}}}$	Pearson correlation coefficient between observed and estimated performances
SBE	Second version of modified Bayesian evaluation
SD	Standard deviation
SI	Selection index
SNP	Single nucleotide polymorphism
ssGBayes	Single-step Genomic Bayesian Prediction
ssGBLUP	Single-step genomic BLUP
TSA	Two-step algorithm

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS, ORAL PRESENTATIONS, AND POSTERS

PUBLICATIONS

Colinet, F., J. Vandenplas, P. Faux, S. Vanderick, R. Renaville, C. Bertozzi, X. Hubin, and N. Gengler. 2013. Walloon single-step genomic evaluation system integrating local and MACE EBV. *Interbull Bull.* 47:203–210.

Hammami, H., J. Vandenplas, M.J. Carabaño, B. Logar, J. Bormann, C. Bertozzi, and N. Gengler. 2014. Strategies to combine novel traits across countries: example of heat stress. *Interbull Bull.*, 48:79–84.

Soyeurt, H., C. Bastin, F. Colinet, V. Arnould, D. Berry, E. Wall, F. Dehareng, N. Nguyen, P. Dardenne, J. Schefers, J. Vandenplas, K. Weigel, M. Coffey, L. Theron, J. Detilleux, E. Reding, N. Gengler, and S. McParland. 2012. Mid-Infrared Prediction of Lactoferrin Content in Bovine Milk: Potential Indicator of Mastitis. *Animal.* 6:1830–1838.

Vandenplas, J., C. Bastin, N. Gengler, and H.A. Mulder. 2013. Genetic variance in micro-environmental sensitivity for milk and milk quality in Walloon Holstein cattle. *J. Dairy Sci.* 96:5977–5990.

Vandenplas, J., O.F. Christensen, and N. Gengler. 2014. Short communication: Alteration of priors for random effects in Gaussian linear mixed models. *J. Dairy Sci.* 97:5880–5884.

Vandenplas, J., F.G. Colinet, and N. Gengler. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. *Genet. Sel. Evol.* 46:59.

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526.

Vandenplas, J., S. Janssens, N. Buys, and N. Gengler. 2013. An integration of external information for foreign stallions into the Belgian genetic evaluation for jumping horses. *J. Anim. Breed. Genet.* 130:209–217.

Vanrobays, M.-L., J. Vandenplas, H. Hammami, E. Froidmont, and N. Gengler. 2014. Short communication: Novel method to predict body weight of primiparous dairy cows throughout the lactation. *J. Dairy Sci*. In press.

ORAL PRESENTATIONS

ON INVITATION

Bastin, C., J. Vandenplas, and N. Gengler. 2014. Improving dairy cow fertility using milk-based indicator traits. CD–ROM Comm. 155 in Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada.

Bastin, C., D.P. Berry, J. Vandenplas, H. Soyeurt, and N. Gengler. 2014. Selection of milk fatty acid composition for improved dairy products from more fertile and healthier cows. Proceedings of the 69th British Cattle Conference, Telford, UK.

ON A PERSONAL PROPOSAL

Bastin, C., J. Vandenplas, A. Laine, and N. Gengler. 2013. Genetics of mastitis in the Walloon Region of Belgium. Page 523 in Book of Abstracts of the 64th Annual Meeting of the European Federation of Animal Science, Nantes, Frances. Wageningen Academic Publishers, Wageningen, the Netherlands.

Colinet, F.G., A. Vanlierde, S. Vanden Bossche, M. Sindic, F. Dehareng, G. Sinnaeve, J. Vandenplas, H. Soyeurt, C. Bastin, and N. Gengler. 2012. Mid-infrared prediction of milk titratable acidity and its genetic variability in first-parity cows. Page 17 in Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., C. Bastin, N. Gengler, and H.A. Mulder. 2012. Genetic variance in environmental sensitivity for milk and milk quality in Walloon Holstein cattle. Page 15 in Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., F.G. Colinet, P. Faux, S. Vanderick, and N. Gengler. 2013. Direct use of MACE EBV in the Walloon single-step Bayesian genomic evaluation system. *J. Dairy Sci.* 96(E-Suppl.):ii.

Vandenplas, J., F.G. Colinet, and N. Gengler. 2013. Combination of genotype, pedigree, local and foreign information. Page 452 in Book of Abstracts of the 64th Annual Meeting of the European Federation of Animal Science, Nantes, France. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., and N. Gengler. 2012. Extension of Bayesian procedures to integrate and

to blend multiple external information into genetic evaluations. *J. Dairy Sci.* 95(Suppl. 2):449–450.

Vandenplas, J., and N. Gengler. 2014. Bayesian approach integrating correlated foreign information into a multivariate genetic evaluation. Page 317 in Book of Abstracts of the 65th Annual Meeting of the European Federation of Animal Science, Copenhagen, Denmark. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., I. Misztal, P. Faux, and N. Gengler. 2012. Bayesian integration of external information into the single step approach for genomically enhanced prediction of breeding values. *J. Dairy Sci.* 95(Suppl. 2):682.

Vanrobays, M.-L., C. Bastin, F.G. Colinet, J. Vandenplas, T. Troch, H. Soyeurt, and N. Gengler. 2012. Estimation of Myostatin gene effects on production traits and fatty acid contents in bovine milk. Page 78 in Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovakia. Wageningen Academic Publishers, Wageningen, the Netherlands.

POSTERS

Colinet, F.G., J. Vandenplas, P. Faux, S. Vanderick, C. Bertozzi, X. Hubin, and N. Gengler. 2013. Development of a genomic evaluation for milk production for a local bovine breed. Page 233 in Book of Abstracts of the 64th Annual Meeting of the European Federation of Animal Science, Nantes, France. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., and N. Gengler. 2012. Integration of external information into genetic evaluations by a Bayesian procedure. Communication in the Proceedings of the 17th Symposium on Applied Biological Sciences, Leuven, Belgium.

Vandenplas, J., and N. Gengler. 2014. Use of high performance computing in animal breeding. Page 326 in Book of Abstracts of the 65th Annual Meeting of the European Federation of Animal Science, Copenhagen, Denmark. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vandenplas, J., S. Janssens, and N. Gengler. 2011. Integration of foreign breeding values for stallions into the Belgian genetic evaluation for jumping horses. Page 140 in Book of Abstracts of the 62nd Annual Meeting of the European Federation of Animal Science, Stavanger, Norway. Wageningen Academic Publishers, Wageningen, the Netherlands.

Vanrobays, M.-L., J. Vandenplas, H. Hammami, and N. Gengler. 2014. Prediction of body weight of primiparous dairy cows throughout lactation. CD-ROM comm. 572 in Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada.