Assessment of Connectedness in the International Genetic Evaluation of Simmental and Montbéliard Breeds

Marie-Noëlle Fouilloux[1], Stéphanie Minery[1], Sophie Mattalia [1] and Denis Laloë[2],
[1] *Institut de l'Elevage, INRA-SGQA,*
78352 Jouy en Josas, France
[2]*Station de Génétique Quantitative et Appliquée, Département de Génétique Animale, Institut National de la Recherche Agronomique, 78352 Jouy en Josas, France*
email: marie-noelle.fouilloux@jouy.inra.fr

Abstract

Connectedness between countries is a critical point for a reliable international evaluation. Weak genetic links between countries can affect predicted genetic values and ranking of the animals. A new method is presented here to measure connectedness between countries: by simulation, a systematic difference between true genetic levels of countries is introduced, a BLUP is performed based on true pedigree and the percentage of the initial difference that can be found between the estimated genetic levels of the corresponding countries is measured. The more 2 countries exchange genes, the closer are their true genetic levels. The better 2 countries are connected, the higher is the percentage of true genetic difference re-estimated. It has then be showed that among countries involved in the Interbull Simmental and Montbéliard breed evaluation, Italy is one of the best connected countries and Slovenia probably the worst connected one. Ireland and the Netherlands use only Montbéliard French bulls. This method very easy and rapid to implement could be used in routine.

1. Introduction

Interest of BLUP genetic evaluations lies in the comparisons of genetic values among animals or groups of animals (herds, countries…). Reliabilities of these comparisons depend of genetic connectedness between those groups.

In the international genetic evaluation of the dairy cattle (Interbull), connectedness between countries is currently described by the number of common bulls between countries (=bulls with multiple evaluations in different countries). These numbers represent very partially the kind of genetic link that can exist between countries, and more particularly, it do not take into account the number of daughters of the common bulls in each country. Rekaya *et al.* (1999) introduced the concept of genetic similarity which represents the proportion of cows sired by common bulls for each pair of countries. Jorjani (1999) advocated using this criterion. Quantitative methods have also been proposed to evaluate connectedness using coefficient of determination (CD) (Laloë (1993) and Laloë *et al.* (1996), Laloë and Phocas (2003), Fouilloux and Laloë (2006)) but these methods are not well adapted to measure connectedness between countries because of either their large computational demands or some theoretical problems. The aim of this paper is to present a new method developed to quantify connectedness between countries in the Interbull evaluation context. This work is a part of the SIMOINT French project that deals with the opportunity of comparing international evaluation of the French Simmental and the Montbéliard breeds.

2. Material and Methods

2.1. Theory

Reliability of a design of *n* sires can be analysed by using the *n*-1 canonical contrasts $\{c_1,\dots,c_i,\dots c_{n-1}\}$ independently distributed under both laws of the true and predicted breeding values (BV) (Laloë, 1993; Laloë *et al.,* 1996). Main results are (i) each comparison **b** among sires is a linear combination of the canonical contrasts: $\mathbf{b}= \sum_{i=1,n-1} a_i \mathbf{c}_i$; (ii) CD of **b** is a linear combination of *CD*s of canonical contrasts: $CD(\mathbf{b}) = \sum_{i=1,n-1} a_i^2 CD(\mathbf{c}_i) \Big/ \sum_{i=1,n-1} a_i^2$;

(iii) The expectation of $\hat{\mathbf{b}}$ results from: $E(\hat{\mathbf{b}}) = \sum_{i=1,n-1} a_i CD(\mathbf{c}_i) E(\mathbf{c}_i)$. Hence, if $E(\mathbf{c}_i)$ is not null, $\hat{\mathbf{b}}$ are biased proportionately to $CD(\mathbf{c}_i)$: $E(\mathbf{b} - \hat{\mathbf{b}}) = \sum_{i=1,n-1} a_i (1 - CD(\mathbf{c}_i)) E(\mathbf{c}_i)$.

Consequently, the comparison of BV of 2 sires (x and y) from 2 countries (1 and 2) can be written as a sum of within and between countries comparisons. Hence, Laloë and Phocas (2003) showed that in the particular case of a balanced design of n sires progeny-tested in N countries (s "national" sires with n_p progeny in only one country and t "international" sires with n_q progeny in each country), only 2 types of canonical comparisons and CD exist (Table 1): within-country ($\mathbf{v_w}$ and CD_w), and among genetic levels of countries ($\mathbf{v_b}$ and CD_b). It is worth noting that, in this context, CD_w is a function of the genetic similarity introduced by Rekaya (1999).So,

$$u_{1x} - u_{2y} = [u_{1x} - \bar{u}_1] - [u_{2y} - \bar{u}_2] + [\bar{u}_1 - \bar{u}_2]$$
$$= \sqrt{\frac{s-1}{s}} v_{w1} - \sqrt{\frac{s-1}{s}} v_{w2} + \sqrt{\frac{2}{s}} v_{b1.2}$$

It leads to:

$$CD(u_{1x} - u_{2y}) = \frac{s-1}{s} CD_w + \frac{1}{s} CD_b \qquad [1]$$

$$E\left[(u_{1x} - u_{2y}) - (\hat{u}_{1x} - \hat{u}_{2y}) \right] =$$
$$(1 - CD_w)[E(u_{1x} - \bar{u}_1) - E(u_{2y} - \bar{u}_2)] \qquad [2]$$
$$+ (1 - CD_b) E(\bar{u}_1 - \bar{u}_2)$$

Consequently, if CD_b is close to 0 (disconnection between countries) (i) the CD of comparison between BV of x and y is decreased [equation 1] but this is in fact negligible as soon as the number (s) of sires is over some dozens, (ii) the comparison among animals may be biased if the genetic levels of the countries are different [equation 2].

From these considerations, Laloë and Phocas (2003) proposed a criterion based on the minimal canonical CD of a genetic evaluation using a canonical decomposition procedure that can not be performed as soon as

n is over a thousand. Therefore, it can not be used in the Interbull context.

Based on the formula [2], which links up the minimal canonical CD to biases in the predicted values, a new method has been developed to quantify the connectedness between countries through the ability to estimate a systematic fictive bias in a real design.

Table 1. Canonical comparisons.

Type of contrasts	Number of contrasts	Canonical contrast	CD
Within country	$N(s-1)$	$v_w = \sqrt{\dfrac{s}{s-1}} [u_{1x} - \bar{u}_1]$	$CD_w = \dfrac{n_p}{n_p + \lambda}$
Between countries	$N-1$	$v_b = \sqrt{\dfrac{s}{2}} [\bar{u}_1 - \bar{u}_2]$	$CD_b = \dfrac{n_p \eta}{n_p \eta + \lambda}$

with $\eta = tn_q / (tn_q + sn_p)$, i.e. the proportion of progeny from international sires within countries.

2.2. Principle and procedure

The principle of this new method is (i) to simulate a systematic difference between the true genetic level of countries, (ii) performing a BLUP and (iii) measuring the percentage of the initial difference that can be found between the estimated genetic levels of the corresponding countries.

2.2.1. Introduction of a systematic difference

Let us considered n sires born in N different countries. As we want to quantify the connectedness between the country k (referent) and the N-1 others (pretenders), it is assumed that the BV of each sire born in k is equal to a given value (a_k) and the BV of all the others is equal to an other given value (a_o). The genetic level of the jth country (pretender or referent) is given by: $\bar{u}_j = 0.5 \left(n_{kj} \times a_k + \sum_{l=1,N}^{l \neq k} n_{lj} \times a_o \right) \Big/ \sum_{l=1,N} n_{lj}$, where n_{lj} is the number of daughters of the sires born in country l and milk recorded in country j. Genetic level of countries gives a 1st information about gene exchange between the referent (k) and the pretenders. So, $\bar{u}_j$ tends to:

2.2.2. BLUP genetic evaluation

BLUP were estimated using a successive overrelaxation iterative method. A single-trait BLUP was performed using a sire model with a heritability of 0.30 and considering all the simulated performances and only one generation of sires. The only fixed effect was the country of milk recording. The estimated BV of each sire ($\hat{u}$) was obtained and the estimated genetic level of each country was computed as follows:

$$\tilde{u}_j = 0.5 \times \sum_{i=1}^{n} n_{ij} \times \hat{u}_i \Big/ \sum_{i=1,n} n_{ij} \; ,$$

where n_{ij} is the number of daughters of the sire i with a performance in the jth country.

2.2.3. Measure of rate of the initial difference estimated

As a result of this procedure, contrasts can be calculated between the true and the estimated genetic levels of the referent (k) and each of the N-1 pretenders (j): $C_{jk} = (\overline{u}_j - \overline{u}_k)$ and $\hat{C}_{jk} = (\tilde{u}_j - \tilde{u}_k)$. As seen before, $\hat{C}_{jk}$ decreased with the connection between k and j, consequently the most interesting indicator of connectedness between k and j is the ratio: $R_{jk} = \hat{C}_{jk} \big/ C_{jk}$. Those N-1 ratios allow to graduate connectedness between k and the N-1 pretenders.

The procedure has to be run N times by changing the referent country because R_{jk} does not mean exactly the same as R_{kj}.

2.3. Data

The data analysed was the Simmental international evaluation for milk production of August 2005, where France participated for two breeds: the Simmental (referred as FRA) and the Montbéliarde (referred as FRM). The country of origin of the bulls was determined from their ID numbers. In total, 18 599 sires born in 15 populations (referred as countries) had 6 558 980 daughters milk recorded in 10 countries. The number of bulls and common bulls is given by table 2. The origin of the bulls is detailed in table 3, with their total number of daughters.

Table 2. Bulls and common bulls, Simmental milk production, August 2005.

CHE: Switzerland, CZE : Czech Republic, DEA: Germany-Austria, FRM: France Montbéliarde, FRA : France Simmental, HUN : Hungary, IRL : Ireland, ITA : Italy, NLD : Netherlands, SVN : Slovenia.

	CHE	CZE	DEA	FRM	FRA	HUN	IRL	ITA	NLD	SVN
CHE	1662	49	66	93	2	0	17	48	27	9
CZE		1597	157	99	9	8	11	74	19	16
DEA			12285	60	62	19	11	276	18	62
FRM				2475	0	1	30	106	56	15
FRA					207	5	0	36	0	13
HUN						108	0	9	1	7
IRL							34	17	16	2
ITA								796	23	36
NLD									65	1
SVN										362

Table 3. Origin of the bulls.

Country of origin	Number of bulls	Number of daughters
AUT	3591	659450
BEL	3	251
CAN	36	27615
CHE	1428	371521
CZE	1328	259662
DEU	8643	3887224
DNK	4	394
FRM	2461	1122893
FRA	138	30199
HUN	86	8546
ITA	480	56782
LUX	3	173
SVN	281	70296
USA	96	60624
IRL	3	233
NLD	18	3117

AUT : Austria, BEL : Belgium, CAN : Canada, CHE : Switzerland, CZE : Czech Republic, DEU : Germany, DNK : Denmark, FRM : France Montbéliarde, FRA : France Simmental, HUN : Hungary, IRL : Ireland, ITA : Italy, LUX : Luxemburg, NLD : Netherlands, SVN : Slovenia.

3. Results and Discussion

For this test, a_k (referent) is assumed to be 200 and a_0 (pretender), to be 400. Results were obtained in less than 10 minutes of CPU time. Time mainly depends on the degree of convergence considered in the BLUP.

Table 4 presents the results for the 10 countries participating to the Simmental Interbull evaluation. Each country has been successively considered as referent. In that

case, results are given horizontally (ie, **on the first line CHE is referent).** 3 variables are given to compare the referent to each of the 9 pretenders: (i) the genetic level of the country j (referent or pretender, $\overline{u}_j$ referred as GL), (ii) Difference of true genetic levels (C_{jk} referred as TrueD) and (iii) rate of the initial difference estimated (R_{jk} given as a percentage).

Instructions for use

Genetic level of the country (GL). On the diagonal, a GL close to 100 means that the country does not use many foreign bulls. On the contrary, a GL close to 200 means that the country uses almost only bulls from foreign origin. **Out of diagonal**, GL gives more information on the origin and the destination of sires: close to 200 the pretender uses almost no bulls coming from the referent, close to 100 the pretender uses mostly bulls coming from the referent. Hence, horizontally, GL give how bulls of the referent country are used by the others (export); vertically, it gives from which country the pretender country imports bulls.

Difference of true genetic levels (TrueD). In theory, in this test, this difference could vary from 100 when there is not exchange of genes between the 2 countries, to -100 when all the bulls of the pretender come from the referent and all the bulls of the referent come from the pretender. The value can be 0 when the 2 countries have the same origin (either referent or any of the 9 pretenders) for their bulls, in the same proportion.

Nevertheless, these 2 criteria (GL and TrueD) do not measure connectedness, but give the situation in term of exchange of bulls between countries. If country A has a lot of bulls coming from country B, but that these bulls are not used at all in country B, then connection between A and B will be poor.

Only the rate of the initial difference estimated really measures connectedness. The better 2 countries are connected, the higher is this percentage. These percentages should be read horizontally and indicate if differences of genetic level between referent and each pretender are correctly estimated.

Analyses of some special cases

a) If we except IRL and NLD which will be analysed later, **ITA** appears as the country that exchanges the most genes with the others **(diagonal GL = 144.2 and a TrueD as referent ≈ 55 with all pretenders).** From its vertical GL (ITA is pretender), we can deduce that ITA uses foreign bulls mainly from DEA (GL, ITA as pretender, DEA as referent=167.9), FRM (191.3), but also from CHE (199.7), FRA (199.9), CZE (199.9) and SVN (199.9). On the other hand, some sires born in ITA are used in FRA (GL, ITA as referent, FRA as pretender=198.3), SVN (199.3), CZE (199.9) and CHE (199.9). **These exchanges certainly contribute to the strong connectedness between ITA (referent) and each pretender with around 80.3% of bias reestimation** (one of the highest global score).

b) With a diagonal GL of 100.2 and a TrueD as referent between 91 and 99, FRM is the country that uses the least foreign bulls. Nevertheless, **percentages of bias reestimation are not that low compared to other countries, and range from 59.09 (with SVN) to 70.17 (with CHE)** if we except NLD and IRL. The foreign bulls used by FRM come from CHE (GL, FRM as pretender, CHE as referent=199.9), CZE (199.9), DEA (199.9), ITA (199.9) and SVN (199.9). FRM does not use bulls from FRA (200). On the other hand, except FRA, all countries use some FRM bulls (horizontal GL, when FRM is referent, are less than 200 except with FRA).

c) From GL we can assume that, except themselves, not **any country uses HUN or IRL bulls** (horizontal GL all equal to 200). It is almost the same with SVN. But in fact, according to the percentage of bias reestimated, these countries do not present the same profile of connection. Contrary to what one might think from the very low number of common bulls (Table 2), **HUN as referent is not in the worst position in term of connectedness,** as shown by its percentages around 66% versus around 39% for SVN as referent (Table 4).

Actually, if we compare **HUN to SVN** (Table 5), despite that both have the same percent of bulls used in other countries (18.5 and 17.4% respectively), the common bulls of

HUN represent much more than the common bulls of SVN in proportion of the daughters of the country (34.6% vs 3.9%). Moreover, **with a percentage of bias reestimation around 39% with each pretender, it can be assumed that SVN (referent) presents the weakest genetic links** (level of connectedness) with other countries.

Table 5. Common bulls and numbers of daughters for HUN and SVN.

	HUN	SVN
N° bulls in the country (n° daughters)	108 *(13 499)*	362 *(83 455)*
N° bulls with daughters only in the country (n° daughters)	88=81.5% of the bulls *(8 822=65.3% of the daughters)*	299=82.6% of the bulls *(80 218=96.1% of the daughters)*
N° bulls with daughters in other countries (n° daughters)	20=18.5% of the bulls *(4 677=34.6% of the daughters)*	63=17.4% of the bulls *(3 237=3.9% of the daughters)*

d) NLD and IRL are in a particular situation. For both of them as referent, diagonal GL is close to 200 and TrueD with each pretender is around 0, meaning that they use almost only foreign bulls and consequently have almost the same true genetic levels as the pretenders. Moreover, when FRM is referent, TrueD are also around 0 with these 2 countries, such as if NLD and IRL have only bulls coming from FRM. That can be confirmed by GL where FRM is referent and NDL or IRL pretender that are respectively equal to 106.0 and 104.1. Further analyses show that among the 34 bulls with daughters in IRL, 31 come from FRM, and among the 65 Dutch bulls, 60 come from FRM. In these particular cases, percentages of bias re-estimation are difficult to interpret because of very small values for TrueD and estimated genetic differences. **Those cases illustrate that the new method is able to detect easily this kind of very strong connection between countries**. **And that percentage of bias re-estimation should be interpreted regarding TrueD value.**

Discussion

The main disadvantage of this method is that there is not a unique criterion for a countries pair. For example, percent of true difference re-estimated between CHE and CZE is 51.91% when CHE is referent and 58.73 when CZE is referent.

Several advantages are clearly identified: this procedure is easy to run, there is no need for extra information, the classical files provided by Interbull can be used, and the different criteria contains more information than a simple number of common bulls, without multiplying the number of variables. Above all, it measures disconnectedness through its main harmful consequence: the ability or not to compare genetic level of animals from different origins or groups.

4. Conclusion

This method is novel and efficient to analyse exchange of bulls between countries and to measure genetic connection created by the common bulls. It can help some joint progenies testing program, like EUROTEST (Dodenhoff *et al.,* 2003) for the Simmental, to decide which countries should first exchange semen or animals to increase genetic ties.

Because it is easy to run, it would be possible to implement it in routine. It could then complete the number of common bulls currently presented to describe the connection between countries involved in the international evaluations of Interbull.

Some improvements of the method could be done in a near future. First of all, it could be of interest to consider genetic relationships between the sires by adding at least one generation of ancestors to take into account indirect genetic links between countries.

Other studies could be done such as (i) evolution analyses of connectedness could be considered either by dividing the cows per age within countries or by applying this method on old Interbull file or (ii) applying that method on other breeds involved in the Interbull evaluations.

5. Acknowledgements

6. References

Dodenhoff, J., Egger-Danner, Ch. & Mattalia, S. 2003. Results from a joint progeny testing program in Simmental. *Interbull Bulletin 31,* 137-142.

Fouilloux, M.N. & Laloë, D. 2006. A method for measuring connectedness among herds in mixed linear models. *Proc. 8th World Congr. Gen. Appl. Livest. Prod.,* Belo Horizonte, Brazil (submitted).

Jorjani, H. 1999. Connectedness in dairy cattle populations. *50th Meeting of the EAAP,* Zurich, Switzerland, 22-26 August 1999.

Laloë, D. 1993. Precision and information in linear models of genetic evaluation. *Genet. Sel. Evol. 25,* 557-576.

Laloë, D., Phocas, F. & Ménissier, F. 1996. Considerations about measures of precision and connectedness in mixed linear models of genetic evaluation. *Genet. Sel. Evol. 28*, 359-378.

Laloë, D. & Phocas, F. 2003. A proposal of criteria of robustness in genetic evaluation. *Livest. Prod. Sci. 80(3),* 241-256.

Rekaya, R., Weigel, K.A. & Gianola, D. 1999. Bayesian estimation of a structural model for genetic covariances for milk yield in five regions of the USA. *50th Meeting of the EAAP,* Zurich, Switzerland, 22-26 August 1999.

Table 4. Difference of estimated/true genetic levels between countries.

Referents		Pretenders CHE	CZE	DEA	FRM	FRA	HUN	IRL	ITA	NLD	SVN
CHE	GL	123.8	199.5	199.8	199.9	198.4	200	200	199.7	196.7	199.8
	TrueD		75.68	76.00	76.15	74.54	76.18	76.18	75.85	72.87	76.03
	%		51.91	50.54	58.77	50.23	49.74	58.96	51.62	58.03	49.18
CZE	GL	199.9	116.0	199.9	199.9	200	200	200	199.9	200	199.9
	TrueD	84.03		84.02	84.04	84.05	84.05	84.05	83.90	84.05	83.90
	%	58.73		58.60	62.06	58.22	58.56	62.32	58.98	63.51	56.88
DEA	GL	198.2	194.2	115.0	199.9	175.2	169.5	200	167.9	200	188.2
	TrueD	83.23	79.28		85.01	60.21	54.56	85.05	52.98	85.05	73.26
	%	56.10	61.45		58.71	84.66	73.56	59.25	75.84	59.59	54.45
FRM	GL	196.8	194.9	199.7	100.2	200	199.5	104.1	191.3	106.0	199.4
	TrueD	96.62	94.73	99.54		99.85	99.31	3.94	91.17	5.89	99.23
	%	70.17	66.40	60.53		60.21	59.75	57.36	63.06	41.15	59.09
FRA	GL	200	199.9	199.9	200	130.3	198.0	200	199.9	200	200
	TrueD	69.70	69.69	69.68	69.71		67.66	69.71	69.56	69.71	69.71
	%	80.48	80.56	81.07	80.49		81.11	80.50	81.15	80.51	80.45
HUN	GL	200	200	200	200	200	136.7	200	200	200	200
	TrueD	63.31	63.31	63.31	63.31	63.31		63.31	63.31	63.31	63.31
	%	65.90	66.03	66.28	65.95	66.20		65.95	66.17	65.96	65.90
IRL	GL	200	200	200	200	200	200	195.9	200	200	200
	TrueD	4.09	4.09	4.09	4.09	4.09	4.09		4.09	4.09	4.09
	%	82.81	82.75	82.57	83.48	82.56	82.56		82.71	83.44	82.55
ITA	GL	199.9	199.9	199.9	199.9	198.3	200	200	144.2	200	199.3
	TrueD	55.72	55.69	55.77	55.76	54.07	55.80	55.80		55.80	55.12
	%	79.48	79.63	81.17	80.90	81.39	80.32	80.97		81.13	80.15
NLD	GL	199.4	199.9	199.9	200	200	200	200	200	198.4	200
	TrueD	0.99	1.50	1.56	1.56	1.56	1.55	1.56	1.56		1.56
	%	118.16	92.02	89.54	93.40	89.32	88.91	93.52	89.97		88.22
SVN	GL	200	199.9	199.9	199.9	200	200	200	199.9	200	116.84
	TrueD	83.16	83.13	83.14	83.14	83.16	83.16	83.16	83.09	83.16	
	%	38.64	38.69	38.91	38.75	38.82	38.74	38.76	39.08	38.75	

GL: True genetic level from simulation
TrueD: Difference of true genetic levels between the 2 countries
%: Percentage of true difference re-estimated

Table 6. Distribution of the daughters, by country of milk recording and by origin of sire.

Origin of the sire	CHE	CZE	DEU	FRM	FRA	HUN	ITA	SVN	IRL	NLD	Total
				Country of milk recording							
AUT	0	12968	639464	194	853	488	2600	2883	0	0	659450
BEL	148	0	0	0	0	0	0	0	0	103	251
CAN	26054	0	1561	0	0	0	0	0	0	0	27615
CHE	359813	1535	8408	351	674	0	321	128	0	291	371521
CZE	72	258119	1125	90	0	0	138	118	0	0	259662
DEU	8574	17716	3805975	396	10189	4116	30421	9837	0	0	3887224
DNK	0	394	0	0	0	0	0	0	0	0	394
FRM	15259	15711	13814	1055564	0	73	8236	520	5461	8255	1122893
FRA	0	55	1130	0	28596	276	142	0	0	0	30199
HUN	0	0	0	0	0	8546	0	0	0	0	8546
ITA	389	350	1415	418	711	0	52929	570	0	0	56782
LUX	173	0	0	0	0	0	0	0	0	0	173
SVN	0	99	584	150	0	0	64	69399	0	0	70296
USA	59102	0	1522	0	0	0	0	0	0	0	60624
IRL	0	0	0	0	0	0	0	0	233	0	233
NLD	2708	168	104	0	0	0	0	0	0	137	3117
Total	472292	307115	4475102	1057163	41023	13499	94851	83455	5694	8786	6558980