

Connectedness in the French Holstein cattle population

Eric Hanocq*, Didier Boichard

Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France

(Received 24 December 1998; accepted 12 February 1999)

Abstract – An analysis of connectedness in the French Holstein cattle population was carried out. This study was motivated by the fact that artificial insemination (AI) bulls are evaluated at the national level, whereas they are usually progeny tested only in the region of their AI stud. Connectedness among AI studs was measured by the generalised coefficient of determination (CD) of contrasts between mean breeding values of bulls from the different AI studs. Four connectedness components were distinguished. The relative influence of each component was assessed through the increase in prediction error variance (PEV) of the contrasts after this information was discarded. CDs of contrasts were always higher than 0.80. Therefore, connectedness level among AI studs was high and provided an accurate national genetic evaluation. Out of the different components of connectedness, withdrawing of proven bull connection data caused the greatest increase in PEV (+47.5 %) primarily due to the change in the connecting structure of the data. Genetic relationships among bulls were the next important source of information. In contrast, contributions from the planned use of the bulls progeny tested were quantitatively limited (8 % increase in PEV) and foreign semen had a minor contribution (2 % increase in PEV). However, in spite of its limited quantitative impact compared to the other components, planned sampling bull connectedness is recommended because it provides high quality data for model validation and bias investigations. © Inra/Elsevier, Paris

connectedness / accuracy / progeny test / genetic evaluation / dairy cattle

Résumé – **Analyse de la connexion en population bovine Holstein française.** Cet article présente une analyse de la connexion du dispositif d'évaluation génétique chez les bovins laitiers Holstein en France. Cette étude est motivée par le fait que les taureaux d'insémination artificielle (IA) sont évalués au niveau national alors qu'ils ne sont généralement testés sur descendance que dans une seule région, la

* Correspondence and reprints: Inra LGAP, domaine de Brunehaut, 80200 Estrées-Mons, France
E-mail: hanocq@mons.inra.fr

zone d'influence du centre d'IA propriétaire. La connexion entre centres est mesurée par le coefficient de détermination généralisé (CD) des contrastes entre moyennes de valeur génétique des taureaux des différents centres. Quatre sources de connexion sont identifiées et l'influence relative de chaque composante est mesurée par l'augmentation de variance d'erreur de prédiction (PEV) des contrastes lorsque cette composante est supprimée. Les CD des contrastes sont supérieurs à 0,80. La connexion entre centres est donc élevée et garantit une évaluation génétique précise à l'échelle nationale. Parmi les différentes composantes, la connexion par les taureaux de service est la plus importante, l'exclusion de ces données provoquant une augmentation de la PEV de 47,5 %, essentiellement due au changement de structure du dispositif. Les parentés entre taureaux de testage constituent la seconde source de connexion. La contribution de la connexion planifiée de testage est plus limitée (8 % d'augmentation de PEV) tandis que la contribution des taureaux étrangers est mineure (2 % d'augmentation de PEV). En dépit de son importance quantitative limitée, la connexion planifiée de testage destinée à améliorer les connexions est encouragée, car elle fournit un échantillon de qualité pour vérifier que l'évaluation génétique nationale n'est pas biaisée. © Inra/Elsevier, Paris

connexion / précision / contrôle sur descendance / évaluation génétique / bovin laitier

1. INTRODUCTION

In dairy cattle, artificial insemination (AI) has enabled breeding programmes to develop and expand geographically. As a result, exchanges of germplasm have increased, leading to intense national and even international competition between breeding companies. In most cases, exchanges are determined by the genetic level of populations evaluated in different environmental conditions. In this context, the animal model BLUP became the most widely used method in the 1990s, because it provides the most likely ranking of animals [9]. However, this ranking is correct only if the design is well connected and provides unbiased and accurate comparisons between animals and between subpopulations in different environments, i.e. herds, regions or countries. In terms of connectedness, two problems could lead to an incorrect ranking of animals distributed in different environments. On the one hand, quantitatively limited genetic ties between candidates may cause inaccurate contrast predictions, as simulated by Hanocq et al. [8]. On the other hand, connecting data may be affected by uncontrolled factors, such as preferential treatment, and produce biased estimates.

In French dairy cattle, the extensive use of artificial insemination (over 90 %) a priori suggests that there are sufficient genetic ties among herds to allow across herd genetic evaluation. A single genetic evaluation is performed for the whole country in order to provide a unique ranking of animals at the national level [1]. However, as each AI stud is specifically attached to a geographical region, sampling bulls (i.e. young bulls being progeny tested with their first crop of daughters) are mostly progeny tested within this region. Some connections are created through the exchange of proven bull semen between regions, but they are not controlled quantitatively or qualitatively. Although the quality of the design has been continuously improved since 1986 with the addition of planned connections from sampling bulls, it is worth studying the

accuracy of contrasts of sampling bulls from different AI studs and, therefore, the connectedness of the overall design.

Connectedness was first an all-or-none statistical concept developed by Bose [2] followed by Eccleston and Hedayat [3]. As developed by Searle [15], a connected design is a design in which it is possible to estimate all contrasts. Foulley et al. [5, 6] proposed a continuous measure of connectedness suited to very unbalanced designs currently met in animal breeding and extended the connectedness measure to random effects. Recently, various criteria have been proposed to measure connectedness for a given contrast. To study the influence of including a factor in a model, Foulley et al. [5, 6] proposed to express the prediction error variance (PEV) of a contrast relative to that observed in a reduced model without this factor. Laloë [13] extended the concept of coefficient of determination to any contrast between random effects. Kennedy and Trus [12] described connectedness as a gene flow between subpopulations, or in terms of relationships between animals distributed across levels of fixed effect.

The purpose of this paper was to verify the assumption that connectedness among AI studs is sufficient to provide a reliable overall genetic evaluation in the French Holstein cattle population. Various components of connectedness were identified and their contribution to the overall connectedness among AI studs was determined.

2. MATERIALS

Milk yield records were extracted from the French national evaluation files of the Holstein breed. First lactation data recorded in 1993 (467 947 lactations) were analysed to study the connectedness between the five most important French AI studs. Only AI bulls were considered and classified into three groups according to their birth year and origin: proven bulls, foreign bulls or sampling bulls (i.e. young bulls being progeny tested with their first crop of daughters). Sampling (1 091), proven (150) and foreign bulls (33), were required to have at least 20, 200 and 200 daughters with performance recorded in 1993, respectively. Eighty-four per cent of the females included in this analysis were born from proven bulls, 11.0 % from sampling bulls and 5.3 % from foreign sires.

Each AI stud is attached to a geographical area defined by the insemination activity of its membership co-operatives. As a consequence, its sampling bulls were mostly progeny tested within this area. Nevertheless, AI studs were connected to each other through link records. Later on in this paper, a link record is defined as a performance of a female recorded in a region different from the region of the stud of her sire. Link records contributed 21.5 % of the whole data set. Five main sources of connectedness were identified.

- Connections due to proven bulls. As proven bull semen was widely spread over the whole country, many daughters were recorded out of the region of origin of their sire. As shown in *table I*, exchanges were not fully balanced, reflecting differences in genetic superiority and marketing policy of each AI stud. For instance, exchange was heavier between AI studs A and C and between AI studs B and D than between others. Globally, proven bull link records were quantitatively very important (15.5 % of the total data set and 71.9 % of all link records). Consequently, these link data contributed largely to AI stud comparison and to overall connectedness.

Table I. Regional distribution of the daughters with respect to the AI stud owning sires.

Region	AI stud					Imported sires (33)	All
	A (362)*	B (232)	C (190)	D (93)	E (214)		
A	106 600	7 020	2 897	209	10 094	5 145	131 965
	<i>17 966</i>	<i>488</i>	<i>309</i>	<i>120</i>	<i>374</i>		<i>19 257</i>
B	1 787	88 558	536	1 699	795	2 955	96 330
	<i>125</i>	9 805	<i>104</i>	<i>73</i>	<i>130</i>		<i>10 237</i>
C	9 439	3 460	51 835	180	4 096	10 259	79 269
	<i>124</i>	<i>248</i>	8 260	<i>82</i>	<i>130</i>		<i>8 844</i>
D	3 701	17 314	830	11 614	3 370	4 150	40 979
	<i>215</i>	<i>212</i>	<i>163</i>	3 117	<i>177</i>		<i>3 884</i>
E	2 470	1 547	442	42	61 018	2 191	67 710
	<i>175</i>	<i>114</i>	<i>123</i>	<i>109</i>	<i>8 951</i>		<i>9 472</i>
All	123 997	117 899	56 540	13 744	79 373	24 700	416 253
	<i>18 605</i>	<i>10 867</i>	<i>8 959</i>	<i>3 501</i>	<i>9 762</i>		<i>51 694</i>

* Number of sampling bulls being progeny tested in 1993. First row (normal letters): daughters of proven bulls; second row (italic letters): daughters of sampling bulls. AI: artificial insemination.

– Connections due to sampling bulls. To a much smaller extent, sampling bull daughters were also spread over the whole country (*table I*). These links resulted from planned semen exchanges between AI studs, in order to voluntarily improve the connectedness of the progeny test programme and to test bulls simultaneously in several regions. Since 1986 and according to a national agreement, each AI stud exchanged with the others at least 10 % of the semen of at least 20 % of the sampling bulls. In practice, exchanges were more numerous, owing to specific collaborations between some AI studs and because of female trade. Sampling bull link records represented 0.8 % of the total data set and 3.6 % of all link records.

– Connections due to foreign bulls. Daughters sired by foreign bulls were spread over the whole country. They represented from 3 to 12 % of the females in each region, and 5.3 % on average (*table I*). They contributed 24.5 % of all link records.

– Connections arising from genetic relationships among bulls. The sampling bulls of different AI studs might have common ancestors. These genetic ties, considered through bull sires and maternal grand sires, were taken into account through the relationship matrix **A**. The mean values of the elements, within and between AI studs, are shown in *table II*. Values of within as well as between AI stud coefficients were quite high because of the reduced number of origins, particularly the number of bull sires.

– Connections due to other relationships between females and particularly dam–daughter relationships. In total, 9 825 cows with a performance in another region than their dam were found in the data set. However, these connections arising from the sales of females and embryos, were not investigated in this

Table II. Mean values of the coefficients in the numerator relationship matrix **A**.

AI stud	Within AI stud	With the rest of the population
A	0.020	0.021
B	0.024	0.021
C	0.034	0.026
D	0.042	0.026
E	0.035	0.025

AI: artificial insemination.

study because they would have required more than 1 year of data and they cannot be studied with a sire model.

3. METHODS

This study measured the degree of connectedness between AI studs and the relative contribution of each component. To reach this goal, several methods could be used, each closely associated with a statistical model. The most straightforward model is the model used in genetic evaluation, i.e. an animal model applied to three lactations and to the complete data set spanning 25 years and describing the whole selection process. However, with such a complex data set, disentangling the connecting components and measuring their respective weight appeared to be impossible. A smaller data set reduced to first lactations and to 1 year of recording (1993), analysed using a sire model, made it possible to focus on the information of interest, i.e. the major links across AI studs. Two models might be considered, with or without genetic groups. A model with fixed group effects assumes different subpopulations and would provide unbiased estimates of contrasts between AI studs. Alternatively, a model without groups assumes a random sire sampling from a unique population and, consequently, does not provide unbiased estimates of AI stud contrasts, should this assumption not hold. However, both models are equally efficient to measure the degree of connectedness of the design, but through different tools, the sampling variances of contrasts between fixed group effects in the model with groups, or the generalised CD [13] in the model without groups. Moreover, the generalised CD approach makes it possible to measure the weight of pedigree information and to compare the PEVs of contrasts between AI studs. In this paper, we present a study based on the concept of generalised CD and PEVs of contrasts applied to a sire model without groups. Each AI stud was characterised by its batch of sampling bulls with daughters in first lactation in 1993.

3.1. Model and notations

Daughters' production records were analysed with the following simple sire model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s} + \mathbf{e} \quad (1)$$

where \mathbf{y} is the vector of performances; \mathbf{b} is the vector of fixed effects of herd and proven bull; \mathbf{s} is the vector of random effects of sampling bulls and

their ancestors, assumed to be normally distributed $\mathbf{s} \sim N(\mathbf{0}, \mathbf{A}\sigma_s^2)$, with \mathbf{A} being the numerator relationship matrix, and σ_s^2 the sire variance. Genetic relationships among sampling bulls were considered through their sire and maternal grand-sire. The heritability h^2 was assumed to be 0.25; \mathbf{e} is the vector of random residuals, assumed to be normally distributed $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, with σ_e^2 being the residual variance; \mathbf{X} , \mathbf{Z} are incidence matrices.

Note that some bulls might appear as proven sire as well as sire or maternal grand-sire of sampling bulls. These bulls were considered twice in the model, once as a fixed effect, in order to study the impact of link records of their daughters, and a second time as a random effect, in order to study the impact of relationships. This simple approach easily enabled the different components of connectedness to be distinguished, although it did not provide a totally exact picture of the true situation.

From equation (1), the PEV matrix relative to the sampling bulls is proportional to:

$$\mathbf{C} = (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1} \quad (2)$$

where \mathbf{M} is the absorption matrix for fixed effects, equal to

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

and $\lambda = \frac{4-h^2}{h^2}$ ($= 15$). \mathbf{C} can be partitioned into four blocks as follows:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{TT} & \mathbf{C}_{TA} \\ \mathbf{C}_{AT} & \mathbf{C}_{AA} \end{pmatrix} \quad (3)$$

where \mathbf{C}_{TT} and \mathbf{C}_{AA} are blocks corresponding to the sampling bulls and their ancestors, respectively. In order to measure connectedness among the sampling bulls, analysis can be restricted to \mathbf{C}_{TT} .

For bull k , the individual CD was computed as:

$$\text{CD}_k = 1 - \lambda \frac{\{\mathbf{C}_{TT}\}_k}{\{\mathbf{A}\}_k} \quad (4)$$

where $\{\mathbf{C}_{TT}\}_k$ and $\{\mathbf{A}\}_k$ are its corresponding diagonal coefficients of the matrices \mathbf{C}_{TT} and \mathbf{A} , respectively. When relationships were voluntarily partially or completely omitted from the analysis, an incorrect relationship matrix was used (\mathbf{A}^* instead of \mathbf{A}). As a consequence, the inverse matrix corresponding to the sampling bulls in equation (3), obtained from $(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^*)^{-1} : \mathbf{C}_{TT}^*$, does not directly lead to the PEV. The PEV were obtained from

$$\mathbf{C}_{TT}^0 = \mathbf{C}_{TT}^* + \lambda\mathbf{C}_{TT}^*(\mathbf{A} - \mathbf{A}^*)\mathbf{C}_{TT}^* \quad (5)$$

instead of \mathbf{C}_{TT} [10].

3.2. Connectedness criteria

The connectedness, measured with the CD criterion proposed by Laloë [13], was expressed for a given contrast $\mathbf{x}'\mathbf{s}$ as follows:

$$\text{CD}(\mathbf{x}) = 1 - \frac{\text{Var}(\mathbf{x}'\mathbf{s}|\hat{\mathbf{s}})}{\text{Var}(\mathbf{x}'\mathbf{s})} = \frac{\mathbf{x}'(\mathbf{A}_{TT} - \lambda\mathbf{C}_{TT})\mathbf{x}}{\mathbf{x}'\mathbf{A}_{TT}\mathbf{x}} \quad (6)$$

where $\text{Var}(\mathbf{x}'\mathbf{s}|\hat{\mathbf{s}})$ is the variance of the contrast given the predicted sire effect and equivalent to the PEV of the contrast. The PEV of contrast $(\mathbf{x}'\mathbf{C}_{TT}\mathbf{x})$ is related to its maximum theoretical value in the population $(\mathbf{x}'\mathbf{A}_{TT}\mathbf{x})$. The closer to 1 the value of CD, the greater the accuracy of the comparison.

To analyse connectedness among AI studs, two kinds of contrast $\mathbf{x}'\mathbf{s}$ were defined. For AI studs i and j , with n_i and n_j sampling bulls, respectively, the vector of contrast \mathbf{x} was defined with terms $1/n_i$ and $-1/n_j$ for bulls belonging to AI stud i and j , respectively, and 0 otherwise. The vector of contrast between a particular AI stud i and all the others jointly was defined with terms $1/n_i$ for bulls belonging to AI stud i , and $-\frac{1}{\sum_{j \neq i} n_j}$ otherwise.

The effect of connectedness on the CD of contrasts was analysed by comparing these CDs under the full model and a reduced model without herd effect. Under the full model, the CD reflected both the amount of data and the connectedness of the design. Under the reduced model, the design was fully connected and the CD of contrast between AI studs, called optimal CD, reached its maximum value and was only influenced by the amount of data involved.

The impact of each source of connectedness was analysed by comparing the PEV of the contrasts, after excluding the link records of interest (PEV_R ; R for reduced data set), with its value calculated with all data included (PEV_F ; F for full data set). It was expressed through the variance ratio r :

$$r = \frac{\text{PEV}_R}{\text{PEV}_F} \quad (7)$$

In contrast to Foulley et al. [6], the notations F (for full) and R (for reduced) characterise the data set used instead of the model. By using the criterion r , emphasis is laid on the increase in prediction error variance, or equivalently, on the decrease in accuracy, relative to the reference situation, with all data. The criterion r can be expressed as a function of CD_F and CD_R computed for the full and reduced data set, respectively [7]:

$$r = \frac{\text{PEV}_R}{\text{PEV}_F} = \frac{1 - \text{CD}_R}{1 - \text{CD}_F} \quad (8)$$

However, exclusion of link data affected both the amount and the structure of information. For a fair comparison, the effect of data structure should be measured for a constant amount of data. Therefore, another analysis was carried out on a reduced data set obtained after randomly removing within each bull

a number of records equal to the number of link records. Given that the data set is large enough, the random exclusions of data affected only the quantity of information, whereas the expected structure of the design was statistically unchanged. Several replicates were performed for random exclusion for each component and results were very similar over replicates.

3.3. Comparison to a theoretical design

In practice, typical values of CD of contrast were illustrated by the definition of a theoretical and simple design providing the same CD values. This simple design involved only unrelated sampling bulls in two AI studs attached to their own region. Numbers of bulls per AI stud, total number of daughters per bull and distribution of daughters across the regions were the three parameters influencing the CDs of contrasts. Numbers of bulls per AI stud were known (*table I*). The total number of daughters per bull (n) was the number of daughters required to reach the optimal CD of contrast derived from the real population assumed the AI stud effect known (*table III*). The distribution of daughters across the regions, assuming sampling bulls having n_1 daughters in the region of their own AI stud and n_2 daughters in the other one ($n_1 + n_2 = n$), was then deducted from the true CD of contrast. The proportion of link records was computed as n_2/n .

Table III. Theoretical effective number of daughters and their distribution across regions according to the CDs of contrasts comparing each AI stud and all the others jointly.

	Number of sampling bulls ^a	CD of contrasts	Optimal CD of contrast	Effective number of daughters ^b	Distribution of daughters across regions ^c
AI stud A	362 – 729	0.88	0.91	155	116/39 (25 %)
AI stud B	232 – 859	0.87	0.90	136	91/45 (33 %)
AI stud C	190 – 901	0.84	0.87	105	83/22 (21 %)
AI stud D	93 – 998	0.85	0.87	100	78/22 (22 %)
AI stud E	214 – 877	0.86	0.90	130	90/40 (31 %)

^a In AI stud of interest – in the other AI studs; ^b number of daughters corresponding to the optimal CD; ^c theoretical distribution of daughters across regions providing the real CD value (% link records). CD: coefficient of determination. AI: artificial insemination.

4. RESULTS

Table IV presents the average individual sampling bull CDs within AI stud. They varied from 0.71 to 0.76 under the full model, and from 0.74 to 0.78 under the reduced fully connected model. These individual CDs were obtained with 38 (AI stud D) to 51 (AI stud A) actual daughters per bull, whereas they corresponded to 34–47 effective daughters only. This loss in accuracy was in agreement with the expected cost of herd effect estimation, owing to the limited

contemporary group size, and was not related to connectedness. Because of the rather large size of sampling bull batches (93–362, according to the AI stud), individual CDs would not have been greatly affected even in a fully disconnected situation.

Table IV. Within AI stud mean individual CD of sampling bulls and mean number of daughters per bull.

	AI stud A	AI stud B	AI stud C	AI stud D	AI stud E
Individual CD	0.76	0.75	0.75	0.71	0.74
Optimal CD ^a	0.78	0.76	0.77	0.74	0.76
Actual number of daughters per bull	51	47	47	38	46
Effective number of daughters	47	43	42	34	41

^a Individual CD in a fully connected design. CD: coefficient of determination. AI: artificial insemination.

CDs of contrasts calculated for each AI stud pair (*table V*) varied from 0.81 to 0.88 under the full model and from 0.84 to 0.91 under the reduced model. The CDs of contrasts between each AI stud and all the others jointly illustrated the connectedness level of each AI stud in the overall national context. They varied from 0.84 to 0.88 with the full model, whereas corresponding optimal CDs varied from 0.87 to 0.91. The effective number of daughters per sampling bull which would provide such values in a theoretical design varied from 100 to 155 (*table III*), i.e. much more than the actual number of daughters. This result is due to the multiple sources of connections in the true situation, whereas only sampling bull link records are involved in the theoretical design.

Table V. Coefficients of determination for contrasts among AI studs.

	AI stud A	AI stud B	AI stud C	AI stud D	AI stud E	Rest of the population
AI stud A		0.86 (0.89)	0.88 (0.91)	0.85 (0.88)	0.87 (0.91)	0.88 (0.91)
AI stud B	0.86 (0.89)		0.84 (0.88)	0.88 (0.90)	0.88 (0.91)	0.87 (0.90)
AI stud C	0.88 (0.91)	0.84 (0.88)		0.81 (0.84)	0.82 (0.86)	0.84 (0.87)
AI stud D	0.85 (0.88)	0.88 (0.90)	0.81 (0.84)		0.84 (0.87)	0.84 (0.87)
AI stud E	0.87 (0.91)	0.88 (0.91)	0.82 (0.86)	0.84 (0.87)		0.86 (0.90)

In parenthesis, CD of contrast in a fully connected design. AI: artificial insemination.

Given the effective numbers of daughters computed above, the theoretical distribution of the daughters across regions which provided a similar decrease was computed. Thus, the simplified design which had the same characteristics as the real population was a design where the percentage of daughters with performance out of the region of their sire varied from 33 % (AI stud B) to 21 % (AI stud C). As a conclusion, the accuracy of the contrast between AI

studs corresponds to the accuracy of a progeny test with 100–155 daughters, out of which 21–33 % are link records.

The various sources of connectedness were successively or simultaneously omitted. Their overall influence, including the effect of both the amount and structure of data, were measured by the impact of their omission on the PEVs of genetic difference estimates between AI studs. Results, shown in *table VI*, were very similar for all AI studs. The exclusion of proven bull connections resulted in the highest increase in PEV (47.5 %). Withdrawing sampling bull connections resulted in an increase in PEV of 8 %, whereas the increase was limited to 2 % when foreign semen data were excluded. When all connections between the AI studs (including pedigree) were excluded, the PEV increase was found to be 223 %. Of course, in this situation, the PEV reached its maximum theoretical value and corresponded to a zero CD. Nevertheless, computation of PEV was feasible because of the prior information on sampling sires, accounted for in the model as a random effect. Rao [14] showed that $\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1}$ is always a positive-definite matrix and, therefore, invertible.

Table VI. Contribution of the different components of connectedness to overall connectedness in the French Holstein population in 1993 (PEV in the complete data set, fixed at 100 in this table, is taken as the reference).

	PEV of contrast ^a after excluding some link records	PEV of contrast after omitting an equal amount of data at random	Proportion attributed to change in connecting structure ^b
Full data set	100	100	
Without proven bulls connections (71 928 rec. omitted)	147.5 ^c	105.0	89
Without sampling bulls connections (3 595 rec. omitted)	108.2	105.6	32
Without imported semen (24 700 rec. omitted)	102.1	101.2	43
All connections excluded (100 223 rec. and \mathbf{A} matrix omitted)	323.6	137.6	83

^a Mean value for contrasts between a given AI stud and the rest of the population; ^b $100 \times (\text{column 2} - \text{column 3}) / (\text{column 2} - 100)$; ^c $100 \times r$ (equation 6) = $100 \times \frac{\text{PEV}_R}{\text{PEV}_F}$. PEV: prediction error variance.

This increase in PEV was then attributed either to changes in the quantity or the connecting structure of the data. Results were fairly similar from one AI stud to another. A proportion of 89 % of the overall increase in PEV, observed when proven bull connections were omitted, was attributed to the change in the connecting structure of the design and only 11 % of the increase was due

to a reduction in the amount of information. For sampling sires, the results were completely opposite. Of the overall effect of the omission of these data, 32 % were attributed to the change in the structure and 68 % to the change in the quantity of data. Of the overall increase in PEV, found by simultaneously excluding all components of connectedness, 83 % were due to the change in the connecting structure of the data.

The study of connectedness resulting from genetic relationships among bulls was carried out in two successive stages. First, only relationships between sampling bulls owned by different AI studs were removed from the analysis, while genetic relationships within AI organisation were still considered. As shown in *table VII*, an increase in PEV of around 39 % was observed, varying from 35 to 45 %. In contrast, when all between and within AI stud genetic relationships were excluded, PEV increased by only 7 to 15 % according to the AI stud, and 10 % on average.

5. DISCUSSION AND CONCLUSION

5.1. Connectedness in the Holstein breed in France

Most individual CDs were at least 0.70. The effective numbers of daughters corresponding to the individual CDs were quite close to the real values and these differences (four daughters on average) could be almost entirely attributed to the cost of contemporary group effect estimation. Theoretically, individual CDs also depend on connectedness, i.e. on distribution of daughters over regions. However, owing to the quite high number of sampling bulls per AI stud, each bull is compared to a reasonably large subpopulation and the practical effect of connectedness on individual CDs is very limited. As a consequence, the individual CD is not a useful criterion to assess connectedness level.

Every CD calculated for contrasts between AI stud pairs and for contrasts between a single AI stud and the others jointly was at least 0.80. In terms of connectedness, the actual design is equivalent to a theoretical design with 100–150 daughters per bull, of which 21–33 % are link records. It is interesting to note that the effective numbers of daughters corresponding to these CDs of contrasts between AI studs are much greater than the actual numbers of daughters. This emphasises that this high accuracy of contrasts depends on the whole design and all connectedness components, and not only on sampling bull link records.

These results suggest that accuracy is sufficient to allow comparisons of the genetic level of the AI studs in France. None of the AI studs appears to stand out in any way from the rest of the population. Therefore, although the national breeding programme involves several AI organisations, the connectedness of the design is sufficient to run a unique sire evaluation with a high theoretical accuracy.

For a more accurate description of connectedness among AI studs, the various sources of connectedness were analysed separately. From a quantitative point of view, proven bull connections provide the highest contribution to overall connectedness. Foreign semen is of minor importance. The contribution of sampling bull connections is of intermediate value. The weight of each source

Table VII. Effect of the relationships within and between AI studs on the PEV of comparisons between a given AI stud and the rest of the population.

AI stud	With all relationships	Without relationships between AI studs	Without relationships
A	100	135.0 ^a	106.8
B	100	138.1	108.9
C	100	145.1	114.9
D	100	141.6	114.8
E	100	137.4	107.1

^a $100 \times r$ (equation 6) = $100 \times \frac{PEV_R}{PEV_F}$. AI: artificial insemination. PEV: prediction error variance.

is not directly related to its amount of data. For instance, the relative influence of sampling bull link records on overall connectedness is higher than suggested by the limited number of data involved.

The impact of the different sources of connectedness is due to the quantity of the data, and to the connecting structure of these data. In these terms, proven and sampling bull connections play completely opposite roles. As shown previously, influence on accuracy was essentially due to the connecting structure of data in the former case, and to the amount of data in the latter case. There are two main reasons for this difference. Excluding proven bull link records does not affect the primary information to evaluate the sampling bulls who still have the same number of daughters, whereas excluding sampling bull link records logically decreases their number of daughters. Second, semen exchanges are essentially bilateral for sampling bulls, whereas proven bull semen is distributed at a national level, allowing for a better connecting structure. Depending on the AI stud examined, the exclusion of foreign semen data had a variable but limited impact on the connecting structure of data. Indeed, use of foreign semen is very unbalanced from one AI stud to another and is generally restricted to a class of breeder who contributes little to progeny test.

Genetic relationships among bulls within and between AI studs had opposite effects. Connectedness increases with relationships between groups, whereas it decreases when the within group relationship increases. This conclusion is in agreement with Kennedy and Trus [12] and Hofer [11]. Relationships between sires of a given group mean that genetic values of animals are correlated and lead to a decrease in the effective group size.

The fact that the computation of the PEV was feasible when excluding all connections between the AI studs (including pedigree) illustrates that the connectedness may have a different significance when the effect of interest (sampling bull effect in this paper) is considered as random or fixed. In the case of a random effect, prior information (on expectation, variance and distribution of breeding values) is taken into account in addition to the amount and the structure of data, and as explained in detail by Foulley et al. [5], connectedness in the sense of estimability is always ensured, which can be a source of inaccuracy or bias if results are interpreted without caution [8].

From this study it can be said that in the Holstein breed in France, quantitative connectedness among AI studs is sufficient to run a national genetic evaluation covering all AI organisations. Genetic differences between AI studs are estimated with a high theoretical accuracy. This is absolutely necessary to ensure the overall efficiency of a selection scheme [8]. As a matter of fact, a high accuracy in estimation allows the highest overall selection intensity and therefore optimal gene flows across AI studs. Despite guarantees of quantitative aspects of the connectedness among AI studs, any evaluation of animals from different origins needs to be validated through post-analysis of the quality of the connecting data. In order to study this aspect, Hanocq [7] measured on the same data set the impact on the estimates of excluding sources of connectedness. He showed that none of the components studied, including foreign or French proven bulls whose use is not controlled for regional effect or preferential treatments, generated biased evaluations of sampling bulls. This finding agrees with the analyses of the distribution of the residuals by regions within sire origin.

In a more general way, the international situation can be considered as an extension of a national situation, by replacing AI studs or regions by countries. The increase in the international trade of genetic material from dairy cattle has intensified competition between countries and it is based on a unique ranking of bulls of different origins. The methods used in the present study could be readily applied to the international level.

5.2. Overall considerations on studies of connectedness

Connectedness was measured using two complementary approaches: i) based on the CD criterion characterising the accuracy of a contrast [13]; and ii) based on the comparison of PEV in two situations (complete or reduced by the factor studied) in a similar way to that of Foulley et al. [6]. The choice of criteria used to measure connectedness depends upon the purpose of the study and the structure of the population. In this paper, two criteria were used to describe first the absolute level of accuracy and second the relative impact on accuracy of a given component. It is very important to consider both these aspects simultaneously as illustrated by Hanocq et al. [8]. The CD approach was used to describe the first aspect because it is a very well-known concept providing easily interpretable values. The absolute level of accuracy being characterised, the PEV appears to be very convenient to measure a relative change in accuracy.

In this paper, the model used to describe connectedness was highly simplified in comparison to the animal model used in the official evaluation, but for the purpose of this study, this simplicity did not affect the results to any great extent. The model would not have provided reliable estimates of breeding values but it allowed a good measure of the real connectedness between AI studs. In fact, the use of the animal model would make it possible to account for a larger number of relationships, but many of these additional relationships are within herds. In terms of connectedness, the most important relationships are already accounted for in a sire model. The study focused on connectedness among the sampling bulls being tested, for which accurate selection is a key factor for the efficiency of the breeding programme. In this paper, bulls were grouped only through the definition of the contrasts used to measure connectedness. Thus, no

assumption was made on connectedness between animals when allocating them to the same AI stud. This allowed a more realistic description of the effective connectedness. As in Hanocq [7], this flexible strategy could be extended to any contrast definition, for instance within AI stud, across herds, years, small regions or levels of fixed effects in the model. In contrast, if the bulls had been grouped through groups explicitly included in the model of analysis, the study of connectedness between AI studs would have been carried out under the unverified assumption of a perfect connection of the animals within a group.

ACKNOWLEDGEMENTS

The authors are grateful to J.L. Foulley and D. Laloë (Inra-SGQA, Jouy-en-Josas) for their helpful advice and suggestions to realise the study and improve the manuscript and to the anonymous referees for their valuable comments.

REFERENCES

- [1] Bonaïti B., Boichard D., Verrier E., Ducrocq V., Barbat A., Briend M., La méthode française d'évaluation génétique des reproducteurs laitiers, *Inra Prod. Anim.* 3 (1990) 83–92.
- [2] Bose R.C., Presidential Address, in: *Proc. 34th Indian Sci. Congress*, 1947.
- [3] Eccleston J.A., Hedayat A., On the theory of connected designs; characterization and optimality, *Ann. Stat.* 2 (1974) 1238–1255.
- [4] Foulley J.L., Bouix J., Goffinet B., Elsen J.M., Comparaison de pères et connexion, in: Elsen J.M., Foulley J.L. (Eds.), *Insémination artificielle et amélioration génétique : Bilan et perspectives critiques*, Les colloques de l'Inra 29, 1984, pp. 131–176.
- [5] Foulley J.L., Bouix J., Goffinet B., Elsen J.M., Connectedness in genetic evaluation, in: Gianola D., Hammond K. (Eds.), *Advances in Statistical Methods for Genetic Improvement of Livestock*, Springer-Verlag, Heidelberg, 1990, pp. 277–308.
- [6] Foulley J.L., Hanocq E., Boichard D., A criterion for measuring the degree of connectedness in linear models of genetic evaluation, *Genet. Sel. Evol.* 24 (1992) 315–330.
- [7] Hanocq E., *Etude de la connexion en sélection animale*, Ph.D. thesis, Institut National Agronomique Paris Grignon, Paris, France, 1995.
- [8] Hanocq E., Boichard D., Foulley J.L., A simulation study of the effect of connectedness on genetic trend, *Genet. Sel. Evol.* 28 (1996) 67–82.
- [9] Henderson C.R., Sire evaluation and genetic trends, in: *Proc. Anim. Breed. Genet. Symp. in Honor of Dr J.L. Lush*, Blacksburg, Virginia, 1972, 1973, pp. 10–41.
- [10] Henderson C.R., Comparison of alternative sire evaluation methods, *J. Dairy Sci.* 41 (1975) 760–770.
- [11] Hofer A., Precision of comparisons of estimated breeding values of centrally tested pigs across herds of origin, in: *5th World Congress of Genetics Applied to Livestock Production*, University of Guelph, Ontario, Canada, 7–12 August, 1994, pp. 447–450.
- [12] Kennedy B.W., Trus D., Considerations on genetic connectedness between management units under an animal model, *J. Anim. Sci.* 71 (1993) 2341–2352.
- [13] Laloë D., Precision and information in linear models of genetic evaluation, *Genet. Sel. Evol.* 25 (1993) 557–576.
- [14] Rao C.R., *Advanced Statistical Methods in Biometric Research*, J. Wiley, New York, 1952.
- [15] Searle S.R., *Linear Models*, J. Wiley, New York, 1986.