Importance of 'connectedness' between herds for effective across herd genetic evaluation

P. K. Mathur

Canadian Centre for Swine Improvement, Central Experimental Farm,
Building 54, Maple Drive, Ottawa, ON, K1A 0C6, Canada.

1. The concept of connectedness

The term "connectedness" is commonly used in animal breeding in relation to the genetic evaluations and across herd comparison of estimated breeding values. In this context, connectedness is a measure or accuracy of across herd evaluations. It can be described as a measure of the relationships between herds or contemporary groups based on their effect on the accuracy of comparing the genetic values of animals from one herd or group to the other.

In true statistical sense, a data set is connected if the difference between effects of every pair of the same fixed factor is estimable in a model of main effects only. (Searle, 1971). In genetic evaluation models, especially using BLUP, the herds or contemporary groups constitute the pairs of fixed factors and the differences between them need to be estimable in order to compare genetic evaluations of animals from these groups. The accuracy of differences between individuals from these subclasses is affected by the degree of connectedness between subclasses.

The accuracy of the estimates of breeding values commonly referred as "repeatability" depends upon heritability of the trait and the amount of information available for the estimation, such as number of records on individual and relatives. It reflects the reliability of the estimated breeding value (EBV) as a guide to the true breeding value. However, it is not necessarily a good indicator of the accuracy of comparing EBVs from one contemporary group to the other. The accuracy of such comparisons depends upon the degree of connectedness between the contemporary groups.

This concept of connectedness applies to contemporary groups within a herd or contemporary groups from different herds. Very often, there is an interest to know if two herds have sufficient degree of connectedness between them to include them in a common genetic evaluation program. This is useful in ranking of individuals across herds and in identifying the best individuals for selection and genetic improvement. In this case each herd may be considered as a separate contemporary group. If there are several contemporary groups within a herd, one may select a particular contemporary group to represent the herd. For example, if there are several herd-year-season classes within a herd, the latest herd-year- season group may be considered as representative of the herd. These contemporary groups are henceforth referred to as herds for simplicity.

Connectedness as a statistical measure

An illustration of the concept of connectedness in genetic evaluations is given in Figure 1. Herd A and herd B are connected because of the common sire, Sire 1, is used in both herds. Herds B and C are also connected because of the common sire, Sire 2. However, herds C and D are not connected since there is no common genetic link between them. Therefore, differences between EBVs of animals from herd B and C are estimable and comparable. However, EBVs of animals from herds C and D can not be compared with sufficient accuracy as the herds are disconnected. In fact, it is not always, necessary to have direct genetic links between herds. The connectedness can also occur due to indirect genetic links. In this example, herds A and C are also connected indirectly through herd B, because of Sires 1 and 2. This is true even if the herds contain animals that are not genetically related. For example, if the animals from herd C and herd D are tested in a test station, the herds C and D would be connected and that way all the four herds A, B, C and D would be connected. It is important to note that, connectedness is a statistical measure related to the ability to have estimable comparisons. Similar views have been expressed by Kennedy and Trus (1993) and Hanocq, et al., (1999). Genetic links are not absolutely necessary for connectedness to exist. Nevertheless, they are important and useful ways of increasing the degree of connectedness.
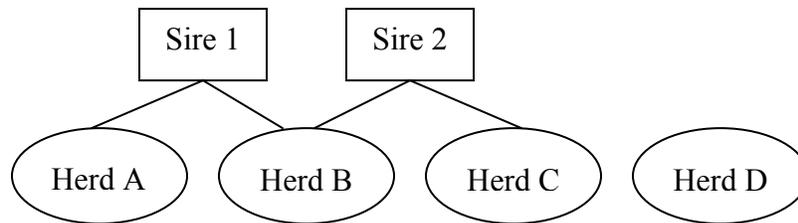
Figure 1: Connectedness among herds through common genetic links

Degree of connectedness

In a genetic evaluation model, it is possible to introduce estimability of differences between EBVs from different herds by treating the herd effects as random effects rather than as fixed effects. However, the main issue is the accuracy of comparisons even if the differences are estimable. The accuracy of such comparisons is affected by the degree of connectedness between the herds even if the herd effects are treated as random and the differences between them are estimable. Therefore, it is important and useful to refer to the degree of connectedness rather than connectedness and disconnectedness. Herds can be connected and differences between the EBVs can be estimable but if the level of connectedness is low, the accuracy of the comparisons will also be low such that it is not very different from that of completely disconnected herds. In fact, in a BLUP genetic evaluation model, connectedness between herds would occur through the genetic relationship matrix, genetic groups and several other fixed or random effects (Canon et al., 1992; Phocas and Laloë, 2004). However, it is the degree or magnitude of this connectedness that is important for accuracy of across herd comparisons.

2. Measurement of connectedness

Prediction error variance of differences between EBVs

The method of estimation depends upon the main objective of the use of the estimate. A common objective in animal breeding is the accuracy of comparing EBVs of animals from different herds or contemporary groups. Therefore, logically an ideal measure should be based on the accuracy of EBVs within herds and accuracy of their differences across herds. The accuracy of an individual EBV is often measured as the prediction error variance which is the inverse element of the mixed model equations corresponding to the animal. The accuracy of the difference between EBVs of two individuals can be estimated as the prediction error variance of their difference (PEVD) as:

$$PEVD = Var(\hat{a}_i - \hat{a}_j) = Var(\hat{a}_i) + Var(\hat{a}_j) - 2cov(\hat{a}_{ij})$$

The average accuracy of differences between EBVs of two individuals from different herds would then be the average PEV of all pair wise differences between individuals from the different herds. However, computing this statistic in a large genetic evaluation program would involve several pairs of animals between one herd and the other and several possible pairs of herds. This can be extremely time consuming; therefore an easier method is required.

Connectedness Rating

The EBV of an animal with its own record depends upon the deviation of its performance from the estimate of the contemporary group or herd effect and other factors included in the genetic evaluation model. Therefore, the PEV of the difference between the EBVs of two animals in different herds or groups can be separated into two components: one associated with the animals and their relatives. This component depends upon the prediction error variance of each EBV, the degree of relationship between the two animals and covariances between animal and herd effects. The other component is associated with the herd effects themselves which includes variances of the estimates of the herd effects and the covariance between them. In calculation of the PEVD of animals in the two herds, the component associated with the herd effects remains the same for each comparison, while that associated with the animals and their relatives is different for each comparison and tends to cancel out over all comparisons as given in the following equation:

$$VED_{ij} = Average[\, PEV(\hat{a}_{ik} - \hat{a}_{jk'})] \approx Var(\hat{h}_i - \hat{h}_j) \cong Var(\hat{h}_i) + Var(\hat{h}_j) - 2cov(\hat{h}_{ij})$$

As a result, the average PEV over all comparisons tends towards the PEV of the difference between estimates of the two herd effects. Therefore, the PEV of the difference between the herd effects can be used as a measure of the degree of connectedness between two herds. A simulation study by Kennedy and Trus (1993) suggested that this can be used as an effective estimate for measuring the degree of connectedness and there was a high correlation (0.995) between average PEVD and VED.

These variances and covariance, between estimates of herd effects can be used in the following form as Connectedness Rating (CR) to estimate of the degree of connectedness:

$$CR_{ij} = \frac{cov(\hat{h}_{ij})}{\sqrt{var(\hat{h}_i)\,var(\hat{h}_j)}} \qquad (1)$$

Where, var ($\hat{h}_i$), var ($\hat{h}_j$) and cov ($\hat{h}_{ij}$) are the variances and covariances of estimates of the respective herd effects.

Considered in this way, the CR seems similar to a correlation between the estimates of the herd effects, however, it is quite different from a simple product moment correlation.

The estimation of CR would also involve computation of variances and covariances of the herd estimates. These require inverse elements of mixed model equations corresponding to the herd effects. It is difficult to obtain a direct inverse of the coefficient matrix of mixed model equations for these specific elements. Hence, the following procedure was used. According to the matrix theory, the product of a matrix with its inverse is an identity matrix. This is true for the coefficient matrix (w'w) of mixed model equations as well.

$$w'w\,(w'w)^{-1} = I$$

$$\text{Therefore,} \quad w'w\,(w'w)_i^{-1} = I_i$$

where,
w'w   =   coefficient matrix of mixed model equations,
 I   =   identity matrix,
 I_i =   a vector of the identity matrix corresponding to contemporary group i (a vector with 1 for the contemporary group and zeros),
$(w'w)_i^{-1}$ =   a vector of inverse elements for the ith contemporary group.

In an iterative procedure for solving the mixed model equations, appropriate sets of the vector I_i can be used to obtain inverse elements for a row or column of the corresponding herd effects of interest. Each of these vectors can then be in the form of a block of inverse elements containing variances and covariances among the herds of interest. In fact, the method can be used more generally to obtain inverse elements for specific rows without a full direct inverse.

This computing algorithm and the estimate of Connectedness Rating were used for about 100 herds participating in the Canadian Swine Improvement Program for the past seven years. It reflects the breeding and testing procedures used. The CR has been very effective in conveying the level of connectedness to the breeders and taking necessary steps to increase it to a more desired level.

Additional methods for estimating connectedness

There are alternative methods for estimating connectedness either considering connectedness as a statistical measure or only as a measure of genetic links. Two methods for estimating connectedness were described and compared to the estimate of VED by Laloë et al (1996). These methods were: the squared correlation between prediction and true differences of genetic values (CD) and a connectedness index (IC) suggested by Foulley et al (1992) which is based on relative decrease in PEV when fixed effects are known. It is suggested that CD combines VED and genetic variability. In this way CD provides a combination of connectedness and genetic variability. The method can be useful for controlling the genetic variability, especially when there is an excessive use of a small group of sires leading to problems of inbreeding etc. However, if the primary objective is to get an estimate of the degree of connectedness and provide information to the herds to take necessary steps to increase the level of connectedness, it might be more useful to have an estimate like CR that would specifically measure the degree of connectedness. The issue of decrease in genetic variability is a common problem in dairy cattle and now in swine, due to the use of AI and availability of semen from genetically superior individuals internationally. It can be dealt with separately by providing more information on the effective progeny per sire, effective population size, level of inbreeding etc.

In a study of test groups in beef cattle, Roso et al. (2004) used four different methods to estimate the degree of connectedness. The methods used included average prediction error variance of the difference between estimated breeding values (PEVD), variance of estimated differences between test group effects (VED), total number of direct genetic links (GLT) (Fries, 1998) and Connectedness Rating (CR). The results show that the average PEVD of each of the test groups with all others could be better predicted on the basis of CR or GLT. In the data set used for test groups in beef cattle, the GLT which is based on genetic links only, also gave adequate estimates of connectedness comparable to VED and CR.

3. Program for calculating the Connectedness Rating (CR)

A program for calculating CR available on CCSI website: www.ccsi.ca /connectedness
This program is based on Animal Breeders Tool Kit (ABTK). The Animal Breeders Tool Kit is a set of programs mainly used for genetics evaluations using BLUP approach. These are a group of small programs that can be used for simple operations like matrix inversion, matrix multiplication and for building incidence matrices etc. These programs were originally developed for UNIX and Linux environments. They have been now converted into Windows programs for wider use (Mathur et al., 2002). The input to these programs is a coefficient matrix file, a file containing row numbers for herd effects and a file containing the herd identifications. The program outputs inverse of a block of the coefficient matrix and the Connectedness Ratings for each pair of herd effects.

4. Level of connectedness needed for across-herd genetic comparisons

Very often the question is what level of connectedness is required for effective across herd comparisons. In a simplified approach, the magnitude of connectedness required depends upon the magnitude of accuracy desired. This can be determined based on a t test as follows:

$$t_\alpha < \frac{d}{\sqrt{VED(\hat{h}_i - \hat{h}_j)}}$$

where,

t_α = value of the Student-t statistic for the desired level of significance, α

d = the average difference between EBVs that needs to be detected

$VED(\hat{h}_i - \hat{h}_j)$ = variance of estimated difference between herds h_i and h_j

The relationship between Connectedness Rating and VED can be described as follows according to Mathur et al. (1998):

$$VED(\hat{h}_i - \hat{h}_j) = var(\hat{h}_i) + var(\hat{h}_j) - 2\,cov(\hat{h}_{ij})$$

$$= var(\hat{h}_i) + var(\hat{h}_j) - 2\,CR_{ij}\sqrt{var(\hat{h}_i)\,var(\hat{h}_j)}$$

Hence the level of Connectedness Rating required is:

$$CR_{ij} > \frac{\left[var(\hat{h}_i) + var(\hat{h}_j)\right] - \dfrac{d^2}{t_\alpha^2}}{2\sqrt{var(\hat{h}_i)\,var(\hat{h}_j)}}$$

A rough approximation of the $VED(\hat{h}_i - \hat{h}_j)$ based on management group size is:

$$VED(\hat{h}_i - \hat{h}_j) \cong \sigma_e^2\left(\frac{1}{n_i} + \frac{1}{n_j} - 2\frac{CR_{ij}}{\sqrt{n_i.n_j}}\right)$$

where, σ_e^2 = residual variance for the trait, and

n_i and n_j are the number of records in the management groups or herds.

Defined in this way, the t statistic required to detect a significant difference d, at a probability level α is:

$$t_\alpha < \frac{d}{\sqrt{\sigma_e^2\left[\dfrac{1}{n_i} + \dfrac{1}{n_j} - 2\dfrac{CR_{ij}}{\sqrt{n_i n_j}}\right]}}$$

The required level of connectedness can be obtained by:

$$CR_{ij} > \sqrt{n_i n_j} \frac{1}{2} \left[\left(\frac{1}{n_i} + \frac{1}{n_j} \right) - \frac{d^2}{\sigma_e^2 (t_\alpha)^2} \right]$$

In this way the level of connectedness required depends upon the size of the herds or contemporary groups (n_i and n_j), level of significance (α), the average difference between EBVs that needs to be detected (d), and the residual variance σ_e^2.

If the herd sizes are equal ($n_i = n_j = n$):

$$CR_{ij} > 1 - \frac{n}{t_\alpha^2} \left(\frac{d}{\sigma_e} \right)^2$$

The required Connectedness Rating is a function of the sample size, the level of significance and the difference to be detected expressed as a fraction of the residual standard deviation.

Considering a typical situation in pig breeding, where the average index is standardized to a mean of 100 and standard deviation is 25 index points, it might be of interest to detect the differences greater than 5 index points i.e. about 20% of the standard deviation. Let us further assume that the level of significance required is P<0.05 and size of the contemporary groups from each of the two herds is 100. Then the level of connectedness should be greater than 48%. The magnitude of Connectedness Rating required depending upon the sample size ($n_i = n_j = n$) and difference to be detected expressed as percentage of the residual standard deviation assuming p < 0.05 is given in Figure 2.
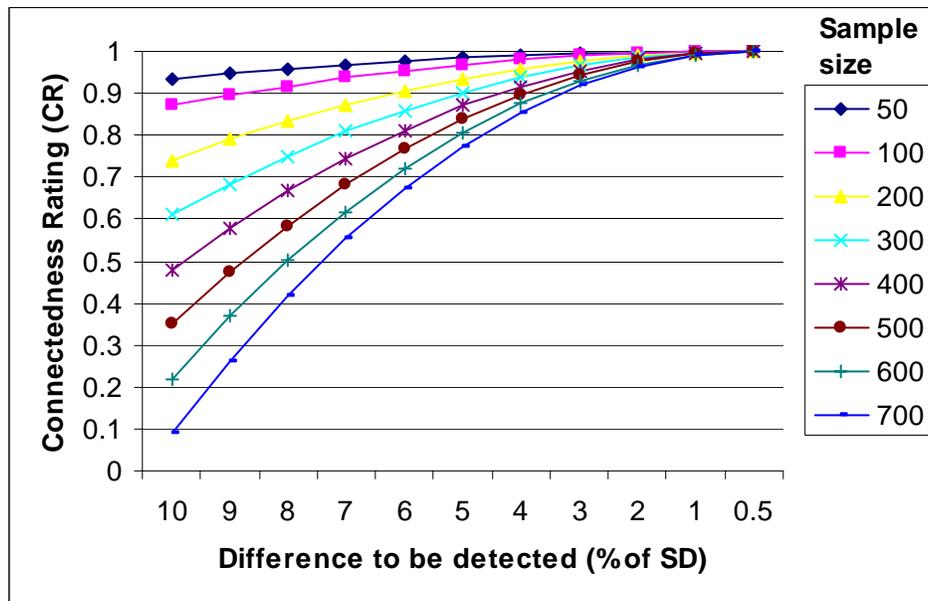

Figure 2: Required levels of Connectedness Rating for different sample size and desired accuracy

For more accurate comparisons where the difference to be detected is very small fraction of the standard deviation of the trait, a higher level of connectedness is required. This is particularly the case when the sample size is very small.

A practical way of looking at the desired level of connectedness is examining the existing levels of connectedness between different pairs of herds and their effect on the standard error of difference between estimates of EBVs. The relationship between the Connectedness Ratings of these pairs and the standard error of difference (square root of VED) between the herd effects for the herds in a large population is given in Figure 3.
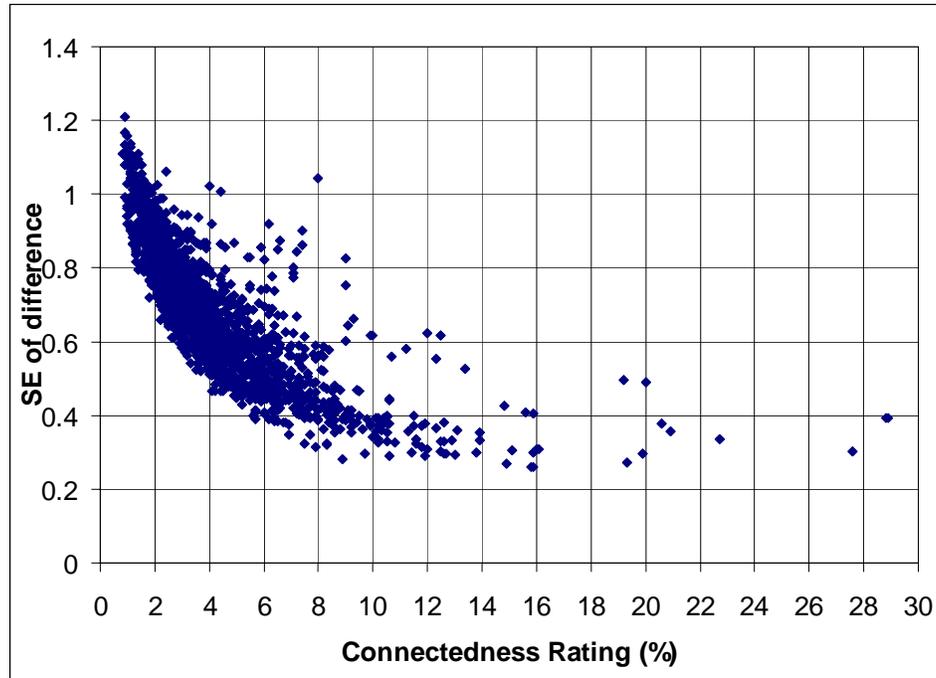

Figure 3: Relationship between standard error of the difference and Connectedness Rating for litter size

The figure shows a nonlinear relationship between the two variables. The standard error of difference decreases as the Connectedness Rating increases. The Connectedness Rating is very low for many pairs of herds. The error is especially large when connectedness is below 5%. Therefore, this can be considered as the minimum level of connectedness required for reasonable across herd comparisons.

Connectedness Rating and sample size

Connectedness, because it is a statistical measure based on the accuracy of the estimates of the herd effects, depends on the herd size. The effect of herd size on connectedness is especially large when the size is less than 50 (Mathur et al., 2002). As the sample size increases the level of connectedness depends more on the genetic links rather than the sample size.

5. Recommendations for increasing the level of connectedness

One of the simplest ways to increase connectedness is to use common sires across herds and contemporary groups. The use of AI boars that have progeny in well connected herds can accelerate the rate of increase in connectedness. It is important to consider the number of progeny, the number of herds in which these progeny are found and a weighted connected rating over several herds (Mathur et al.,1998). Usually about 15% of progeny from well connected sires in a herd are enough to reach the minimum level of connectedness. One may also use young AI boars that do not have any progeny, provided they come from well connected herds.

The use of central test stations where pigs from several herds are tested under common environmental and management conditions is also a useful way to increase connectedness, especially when there are very few genetic exchanges.

It is also important to have sufficient size of the contemporary groups in which the EBVs have been calculated. If a contemporary group has less than 10 animals of the same breed, the accuracy of the EBVs is substantially lower. In practical circumstances, it is therefore important to have at least 10 pigs in each contemporary group or herd-year-season class which is rather a base minimum. These 10 or more pigs should come from at least 3 different sires to have sufficient connectedness within herds and proper accuracy of the evaluations.

In order to have a continuous stream of well-connected boars of sufficient genetic merit, it is necessary that breeders participating in the program make some of their superior genetics available or form a pool of superior AI boars. This type of participation allows the structure of a national breeding program similar to large nucleus with the possibility of good across herd genetic variability and connectedness. This will allow good across herd comparisons, availability of large selection differentials and more genetic progress.

6. References

Canon, J., J. Gruand, J. P. Gutierrez, L. Ollivier (1992) A repeat-sire mating selection experiment on lean tissue growth in the pig : genetic trends of the traits selected. Gen. Sel. Evol. 24: 449-462

Foulley, J.L., E. Hanocq, D. Boichard (1992) A criterion for measuring the degree of connectedness in linear models of genetic evaluation. Genet Sel Evol 24: 315-330

Fries, L.A. (1998) Connectability in beef cattle genetic evaluation: the heuristic approach used in MILC.FOR. Proc. 6th World Cong. Genet. Appl. Livest. Prod., Armidale, NSW, Australia. 27:449-500.

Hanocq, E., L. Tiphine, B. Bibe (1999) An overview of connectedness in animal breeding. Productions Animales. 12:(2)101-111

Kennedy, B. W. and D. Trus (1993) Considerations on genetic connectedness between management units under an animal model.. J. Anim. Sci. 71: 2341-2352.

Laloe, F., F. Phocas, and F. Menissier (1996) Considerations on measures of precision and connectedness in mixed linear models of genetic evaluation. Genet Sel Evol 24: 315-33

Mathur, P. K., B.P. Sullivan and J.P. Chesnais (1998). Estimation of the degree of connectedness between herds or management groups in the Canadian swine population. Canadian Centre for Swine Improvement, Ottawa, Canada. (Mimeo).

Mathur, P. K., B.P. Sullivan and J.P. Chesnais (2002) Measuring connectedness: Concept and application to a large industry breeding program. Proc. 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, Communication No: 20-13.

Phocas, F. and D. Laloë (2004) Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation Genet. Sel. Evol. 36: 325-345

Roso, V.M., F.S. Schenkel, and S.P. Miller (2004) Degree of connectedness among groups of centrally tested beef bulls. Can. J. Anim. Sc. 34: 37-47

Searle, S.R. (1971) Disconnectedness and variance component estimation. Biometrics. 31:969.