

CONECTIVIDAD EN EVALUACIONES GENÉTICAS DE ANIMALES.

1. METODOLOGÍAS

CONNECTEDNESS IN ANIMAL GENETIC EVALUATIONS.

1. METHODOLOGIES

Fabián Magaña-Valencia¹, Rafael Núñez-Domínguez^{1*}, Rodolfo Ramírez-Valverde¹, Felipe A. Rodríguez-Almeida²

¹Departamento de Zootecnia, Postgrado en Producción Animal, Universidad Autónoma Chapingo. 56230. Carretera México-Texcoco, Km 38.5. Chapingo, Estado de México. ²Facultad de Zootecnia y Ecología, Universidad Autónoma de Chihuahua. 31031. Periférico Francisco R. Almada, Km 1. Chihuahua, Chihuahua. (rafaelndguez@yahoo.com.mx).

RESUMEN

La predicción de valores genéticos a través de las ecuaciones de modelos mixtos con el modelo animal es la herramienta más confiable para la selección de sementales y vientres, y el grado de conectividad de los datos es uno de los factores que incide en su calidad. En poblaciones donde la selección y la deriva genética repercuten en las medias genéticas de los grupos contemporáneos (GC), la predicción de valores genéticos puede ser sesgada si los GC están desconectados, disminuyendo la confiabilidad de sus comparaciones. El propósito del presente ensayo fue realizar una recopilación y descripción de los principales métodos disponibles en la literatura para evaluar la conectividad genética entre GC. Se describieron 12 métodos para determinar conectividad genética entre GC, dos cualitativos y 10 cuantitativos. Un método para determinar la conectividad idealmente debe identificar GC desconectados, y conectados directa e indirectamente. Además, debe evaluar el grado de conectividad, ser práctico, y fácil de aplicar e interpretar. Las metodologías cualitativas para evaluar la conectividad entre GC sólo permiten separar los GC que están conectados de los que no lo están, mientras que los métodos cuantitativos evalúan adicionalmente el grado de conectividad. Los métodos descritos son alternativas que se deben comparar considerando una misma estructura de los datos para escenarios comunes en evaluaciones genéticas de animales.

Palabras clave: conectividad genética, métodos de conectividad, evaluación genética.

ABSTRACT

Predicting genetic values with mixed model equations together with the animal model is the most reliable tool for selection of sires and dams, and the degree of data connectedness is one of the factors that influence selection quality. In populations where selection and genetic drift have an influence on the genetic means of contemporary groups (GC), predictions of genetic values may be biased and less reliable if the GCs are disconnected. The objective of this essay was to present a compilation and description of the main methods available in the literature for evaluating genetic connectedness among GC. Twelve methods that determine genetic connectedness among GC were described, two qualitative and 10 quantitative. A method for determining connectedness should ideally identify disconnected and directly and indirectly connected GC. It should also evaluate the degree of connectedness, and it should be practical and easy to apply and interpret. The qualitative methods for evaluating connectedness among GC only allow separating GC that are connected from those that are not, while quantitative methods additionally evaluate the degree of connectedness. The methods described are alternatives that need to be compared considering a single structure of data for common scenarios in genetic evaluations of animals.

Key words: genetic connectedness, connectedness methods, genetic evaluation.

INTRODUCTION

The mixed model equations (MME) of Henderson are used extensively with the animal model to predict genetic values (VG) of animals since they use all the additive genetic

*Autor responsable ❖ Author for correspondence.

Recibido: febrero, 2012. Aprobado: septiembre, 2012.

Publicado como ENSAYO en *Agrociencia* 46: 677-692. 2012.

INTRODUCCIÓN

Las ecuaciones de modelos mixtos (EMM) de Henderson con el modelo animal se usan extensivamente para predecir los valores genéticos (VG) de los animales, ya que utilizan todas las relaciones genéticas aditivas de la población a evaluar, maximizando la exactitud de las predicciones (Henderson, 1973). Esta metodología supone que al incorporar los efectos ambientales en el modelo, incluyendo los del grupo contemporáneo (GC), las comparaciones entre los méritos genéticos de animales nacidos en diferentes GC son apropiadas. Sin embargo, si estos GC no están conectados genéticamente de manera apropiada, existe el riesgo de cuantificar las diferencias genéticas entre GC como diferencias ambientales (Clément *et al.*, 2001; Kuehn *et al.*, 2007; Tarrés *et al.*, 2010).

La calidad de las evaluaciones genéticas depende de numerosos factores, en particular de la estructura de los datos (Hanocq *et al.*, 1999) descrita según las matrices de diseño y las matrices de (co)varianzas. Al considerar la estructura de los datos, el grado de conectividad entre los GC puede afectar la exactitud de la evaluación genética (Kennedy y Trus, 1993; Tosh y Wilton, 1994; Carneiro *et al.*, 2001b), y la confiabilidad de la comparación entre animales de diferentes GC (Mathur *et al.*, 2002a; Roso *et al.*, 2004; Fouilloux *et al.*, 2006). La exactitud de los VG se afecta por la sub o sobreestimación de parámetros genéticos, mientras que el efecto en la confiabilidad de las comparaciones puede reflejarse en la jerarquización incorrecta de animales (Fouilloux *et al.*, 2006; Kuehn *et al.*, 2009; Soga *et al.*, 2010) y por consiguiente en la respuesta a la selección (Lewis y Simm, 2000; Simm *et al.*, 2001; Kuehn *et al.*, 2008a).

Para predecir los VG, teóricamente no se requiere de conectividad (Fernando *et al.*, 1983; Kennedy y Trus, 1993; Fries y Roso, 1997) y los GC desconectados no causan predicciones sesgadas, siempre y cuando los VG de los animales base estén aleatoria e idénticamente distribuidos en toda la población (Kennedy y Trus, 1993; Tosh y Wilton, 1994). Sin embargo, esto difícilmente se cumple para poblaciones reales sujetas a fuerzas como la selección o la deriva genética, que repercuten en las medias genéticas de los GC (Kuehn *et al.*, 2007; Tarrés *et al.*, 2010) y en la validez y precisión de las estimaciones de los

relationships of the population to be evaluated, maximizing the precision of the predictions (Henderson, 1973). This methodology assumes that when environmental effects are incorporated into the model, including those of the contemporary group (GC), comparisons among the genetic merits of the animals born in different GC are appropriate. However, if these GC are not connected genetically in an appropriate way, there is a risk of quantifying the genetic differences among GC as environmental differences (Clément *et al.*, 2001; Kuehn *et al.*, 2007; Tarres *et al.*, 2010).

The quality of genetic evaluations depends on numerous factors, particularly on the structure of the data (Hanocq *et al.*, 1999) described using design matrices and (co)variance matrices. When the data structure is considered, the degree of connectedness among GC can affect the accuracy of the genetic evaluation (Kennedy and Trus, 1993; Tosh and Wilton, 1994; Carneiro *et al.*, 2001b) and the reliability of the comparison of animals from different GC (Mathur *et al.*, 2002a; Roso *et al.*, 2004; Fouilloux *et al.*, 2006). The accuracy of VG is affected by the under- or overestimation of genetic parameters, while the effect on comparison reliability may be reflected in incorrect ranking of the animals (Fouilloux *et al.*, 2006; Kuehn *et al.*, 2009; Soga *et al.*, 2010), and consequently, in the response to selection (Lewis and Simm, 2000; Simm *et al.*, 2001; Kuehn *et al.*, 2008a).

Predicting VG, theoretically, does not require connectedness (Fernando *et al.*, 1983; Kennedy and Trus, 1993; Fries and Roso, 1997) and the disconnected GC do not cause biased predictions, as long as the VG of the base animals are randomly and identically distributed throughout the population (Kennedy and Trus, 1993; Tosh and Wilton, 1994). This is, however, difficult to achieve for real populations subject to forces such as selection or genetic drift that have repercussions in genetic measurements of GC (Kuehn *et al.*, 2007; Tarres *et al.*, 2010) and in the validity and precision of the estimations of (co)variance components (Tosh and Wilton, 1994; Clement *et al.*, 2001).

The biases in VG due to disconnected GC can be reduced by including the effect of genetic group in the model to represent the mean genetic value of the sires in each group. This alternative, however, is not feasible in all scenarios. Inclusion of genetic groups

componentes de (co)varianza (Tosh y Wilton, 1994; Clément *et al.*, 2001).

Los sesgos en VG por GC desconectados pueden reducirse al incluir el efecto de grupo genético en el modelo para representar el valor genético medio de los sementales en cada grupo; sin embargo, esta alternativa no es viable en todos los escenarios. La inclusión de grupos genéticos debería considerarse sólo para grupos con muchos animales, con alta conectividad y alta exactitud, y en caso de detectar grandes diferencias genéticas entre las subpoblaciones de animales base (Phocas y Laloë, 2004). Para características con baja heredabilidad, estos sesgos se pueden disminuir mediante la evaluación conjunta con características de mayor heredabilidad y altamente correlacionadas, por una mejoría en la conectividad genética (Tarrés *et al.*, 2010).

Al considerar un modelo de semental, la conectividad entre GC puede conocerse por la estimabilidad de las funciones lineales entre sementales y GC. Según Hanocq *et al.* (1999), para determinar la conectividad es mejor usar un modelo de semental que considera la mayor parte de las fuentes de conexión; así, otras fuentes ignoradas en el análisis son adicionales y pueden eventualmente mejorar la conectividad. Sin embargo, con un modelo animal donde todas las relaciones de parentesco contribuyen a conectar los GC, la forma de medir la conectividad es diferente y generalmente se indica su grado.

La conectividad de los GC se puede estimar por métodos cualitativos o cuantitativos. En el primer caso sólo es posible identificar grupos conectados y desconectados, mientras que en el segundo es posible determinar el grado de conectividad ofreciendo una medida de confiabilidad o riesgo de las comparaciones entre animales en diferentes grupos (GC, hatos, regiones, países). Para evaluar el grado de conectividad genética entre GC, Kennedy y Trus (1993), Lewis *et al.* (1999) y Mathur *et al.* (2002b) proponen varios métodos, mientras que Laloë *et al.* (1996), Roso *et al.* (2004) y Soga *et al.* (2010) compararon las ventajas y desventajas de sólo algunos métodos. Además se usan algunos métodos para cuantificar el impacto de la implementación de estrategias orientadas a mejorar la conectividad genética de poblaciones reales o simuladas (Huisman *et al.*, 2006; Veselá *et al.*, 2007; Fouilloux *et al.*, 2008) y para evaluar la conectividad genética de poblaciones reales (Nakaoka *et al.*, 2009; Sun *et al.*, 2009; Pegolo *et al.*, 2012).

should be considered only for groups with many animals, with a high degree of connectedness and high accuracy, and in the case in which large genetic differences are detected among the subpopulations of the base animals (Phocas and Laloë, 2004). For low heritability traits, these biases can be reduced by means of a joint evaluation with traits of higher heritability that are highly correlated, by an improvement in genetic connectedness (Tarrés *et al.*, 2010).

When considering a sire model, connectedness among GC can be determined through the estimability of the linear functions between sires and GC. According to Hanocq *et al.*, (1999), to determine connectedness, it is better to use a sire model that considers most of the sources of connection; in this way, other sources ignored in the analysis are additional and can eventually improve connectedness. With an animal model, however, in which all of the genetic relationships contribute to connect GC, the way to measure connectedness is different and generally indicates its degree.

Contemporary group connectedness can be estimated using qualitative or quantitative methods. In the first case, it is only possible to identify connected and disconnected groups, while in the second it is possible to determine the degree of connectedness offering a measure of reliability or risk in comparing animals in different groups (GC, herds, regions, countries). To evaluate the degree of genetic connectedness among GC, Kennedy and Trus (1993), Lewis *et al.* (1999) and Mathur *et al.* (2002b) propose several methods, while Laloë *et al.* (1996), Roso *et al.* (2004) and Soga *et al.* (2010) compared the advantages and disadvantages of only a few methods. In addition, some methods are used to quantify the impact of implementing strategies aimed to improve genetic connectedness of real or simulated populations (Huisman *et al.*, 2006; Veselá *et al.*, 2007; Fouilloux *et al.*, 2008) and to evaluate genetic connectedness in real populations (Nakaoka *et al.*, 2009; Sun *et al.*, 2009; Pegolo *et al.*, 2012).

Routine national genetic evaluations in some countries use different methods: the coefficient of determination method in France for beef cattle and goats (Pegolo *et al.*, 2012), the degree of connectedness method in Canada for hogs (Mathur

En las evaluaciones genéticas nacionales rutinarias de algunos países se usan diferentes métodos: el método de coeficiente de determinación en Francia para bovinos para carne y cabras (Pegolo *et al.*, 2012), el método de grado de conectividad en Canadá para cerdos (Mathur *et al.*, 2002a), el método de correlación entre las varianzas de los errores de predicción de VG en el Reino Unido para borregos (Simm *et al.*, 2001), y el método de número total de lazos genéticos directos entre GC en Canadá para bovinos (Roso *et al.*, 2004). Para algunas poblaciones de animales (como razas de bovinos para carne, ovinos y caprinos), los hatos están dispersos en varias regiones geográficas, los tamaños de los hatos y GC muchas veces son pequeños en relación con otras especies, y la inseminación artificial se usa en forma reducida. Por tanto, un problema común para la evaluación genética nacional en este tipo de poblaciones es el bajo grado de conectividad entre los GC, lo cual requiere de métodos precisos y prácticos para su determinación.

En general, la mayoría de los métodos propuestos, comparados o utilizados usan la varianza del error de predicción (VEP) de las diferencias en VG entre animales o medidas relacionadas con ésta. Sin embargo, no hay consenso en recomendar el uso generalizado de alguno de ellos. Para facilitar la comparación de los métodos propuestos, considerando diversos escenarios y una misma estructura de datos, es importante recopilar información que describa los principios y cálculos necesarios para estimar conectividad genética. Por tanto, el propósito de este ensayo fue realizar una recopilación y descripción de los principales métodos disponibles en la literatura para evaluar la conectividad genética entre GC.

MÉTODOS PARA EVALUAR LA CONECTIVIDAD GENÉTICA

Métodos cualitativos

Método geométrico simple (MGS)

Este método, presentado por Searle (1987), se basa en el producto de las matrices $X'Z$, donde X y Z son matrices de incidencia para GC y sementales en los modelos mixtos. De manera esquemática se construye una tabla de doble entrada donde se

et al., 2002a), the method of correlation between variances of VG prediction errors in Britain for sheep (Simm *et al.*, 2001), and the method of total number of direct genetic links among GC in Canada for cattle (Roso *et al.*, 2004). For some animal populations (such as breeds of beef cattle, sheep and goats) the herds are dispersed in several geographic regions; herd and GC sizes are often small relative to other species, and artificial insemination is used on a small scale. Therefore, a common problem for national genetic evaluation in this type of populations is the low degree of connectedness among GCs, and so, precise, practical methods are required.

In general, most of the proposed, compared or utilized methods use the variance of the prediction error (VEP) of the difference in VG among animals or highly correlated measurements. However, there is no consensus in recommending the generalized use of any of them. To facilitate comparison of the proposed methods, considering diverse scenarios and a single data structure, it is important to compile information that describes the principles and calculations necessary for estimating genetic connectedness. Thus, the purpose of this essay was to present a compilation and description of the main methods available in the literature for evaluating genetic connectedness among GCs.

METHODS FOR EVALUATING GENETIC CONNECTEDNESS

Qualitative methods

Simple geometric method (SGM)

This method, presented by Searle (1987), is based on the product of the matrices $X'Z$, where X and Z are incidence matrices for GC and sires in mixed models. Schematically, a double-entry table is constructed where sires (i) and GC or herds (j) are represented, and the intersections (cells) show the progeny of the i th sire in the j th herd. Data connectedness is indicated by the continuous horizontal or vertical lines, which can change direction only if the cells contain values; if a cell does not have data, the line cannot be drawn. Cells joined by lines represent groups of connected data.

representan, por un lado a los sementales (i) y por el otro a los GC o hatos (j), y las intersecciones (celdas) muestran la progenie del i -ésimo semental en el j -ésimo hato. La conectividad de los datos se indica por el trazo de líneas continuas horizontales o verticales, las cuales pueden cambiar de dirección sólo si las celdas tienen valores; si la celda no tiene datos, el trazo no es posible. Las celdas unidas mediante líneas representan grupos de datos conectados.

Algoritmo para identificar los grupos conectados (AIGC)

Tomando como base el método geométrico simple, Fernando *et al.* (1983) propusieron un algoritmo para identificar los GC (hato-año-estación) conectados, en una clasificación de dos vías sin interacción. No requiere la absorción de efectos principales, por lo que es computacionalmente simple. El álgebra se limita a dos vectores, por lo que el algoritmo se puede aplicar cuando el número de clases de cada uno de los efectos principales sea grande. El resultado es un vector y los elementos con el mismo número corresponden a grupos conectados.

El algoritmo involucra cuatro pasos para cada subgrupo (G): 1) forma un vector $n \times 1$ (G_i) con sus j -ésimos elementos, siendo $G_{ij}=i$ para $j=1, \dots, n$, si un semental j tiene progenie en el subgrupo i , y $G_{ij}=0$ de otro modo; 2) toma $c_i=c_i-1+G_i$, donde $c_0=0$; 3) examina cada elemento de este vector iniciando con el primer elemento de c_i , c_{i1} ; para el primer elemento donde $c_{ij}>i$, cambia a "i" aquellos elementos de c_i que son iguales que c_{ij} y también los que son iguales que $c_{ij}-i$; y 4) repite el paso 3 hasta que cada n elemento de c_i es menor o igual que i . Finalmente, los elementos de c_m con el mismo número corresponden a sementales conectados a través de GC.

Métodos cuantitativos

Promedio de relaciones genéticas aditivas (PRGA)

Banos y Cady (1988) muestran que el PRGA se puede usar como una medida de conectividad entre GC. La mejor conectividad genética entre GC está

Algorithm for identifying connected groups (AIGC)

Taking the simple geometric method as the base, Fernando *et al.* (1983) proposed an algorithm to identify connected GC (herd-year-season) in a two-way classification with no interaction. It does not require the absorption of main effects, and thus it is computationally simple. The algebra is limited to two vectors, and so the algorithm can be applied when the number of classes of each of the main effects is large. The result is a vector and the elements with the same number correspond to connected groups.

The algorithm involves four steps for each subgroup (G): 1) make a vector $n \times 1$ (G_i) with its j^{th} elements, $G_{ij}=i$ for $j=1, \dots, n$, if a sire j has progeny in the subgroup i , otherwise $G_{ij}=0$; 2) take $c_i=c_i-1+G_i$, where $c_0=0$; 3) examine each of the elements of this vector beginning with the first element of c_i , c_{i1} ; for the first element where $c_{ij}>i$, change to "i" those elements of c_i that are equal to c_{ij} and also those that are equal to $c_{ij}-i$; and 4) repeat step 3 until each of the n elements of c_i is less than or equal to i . Finally, the elements of c_m with the same number correspond to sires connected through GC.

Quantitative methods

Average of additive genetic relationships (PRGA)

Banos and Cady (1988) showed that PRGA can be used as a measure of connectedness among GC. The best genetic connectedness among GC is given by the highest PRGA between animals from different GC. PRGA among animals from different GC differs from the other methods in that it is a direct measure of genetic connectedness between GC and is easy to obtain.

The PRGA is obtained from the inverse matrix of additive genetic relationships (A^{-1}). A and A^{-1} can be subdivided in the following manner:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}$$

dada por el mayor PRGA entre animales de diferentes GC. El PRGA entre animales de diferentes GC se diferencia de los demás métodos en que es una medida directa de conectividad genética entre GC y es fácil de obtener.

EL PRGA se obtiene de la inversa de la matriz de relaciones genéticas aditivas (A^{-1}). La A (y A^{-1}) se puede subdividir de la siguiente manera:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}$$

donde $A_{11}, A_{22}, \dots, A_{nn}$ es igual a A para animales del $GC_1, GC_2, \dots, GC_n$, respectivamente, y los elementos fuera de la diagonal ($A_{12}, \dots, A_{1n}$ y $A_{(n-1)n}$) son sub-matrices de relaciones genéticas aditivas (a_{ij}) para animales de dos GC. La media de a_{ij} para las sub-matrices de la diagonal (A_{xx}) para $i \neq j$ es:

$$1 / (n^2 - n) \sum_{i=1}^n \sum_{j=1}^n a_{ij}$$

donde n =orden de A_{xx} . La media de a_{ij} para las sub-matrices de fuera de la diagonal (A_{xy}) es:

$$1 / rc \sum_{i=1}^r \sum_{j=1}^c a_{ij}$$

donde r =número de filas, y c =número de columnas en A_{xy} .

Índice de conectividad (IC) e índice general de conectividad (IGC)

Foulley *et al.* (1992) presentaron el concepto de grado o nivel de no conectividad, relacionando la varianza del error de predicción (VEP) de los efectos genéticos de un modelo completo con la VEP de un modelo reducido que excluye los efectos fijos. Estos autores propusieron el índice de conectividad (IC), que se basa en el uso de un vector de coeficientes de contrastes (x):

$$IC(x) = x' C_R x / x' C x$$

where $A_{11}, A_{22}, \dots, A_{nn}$ is equal to A for animals from $GC_1, GC_2, \dots, GC_n$, respectively, and the elements outside the diagonal ($A_{12}, \dots, A_{1n}$ and $A_{(n-1)n}$) are sub-matrices of additive genetic relationships (a_{ij}) for animals from two GC. The mean of a_{ij} for the sub-matrices of the diagonal (A_{xx}) for $i \neq j$ is:

$$1 / (n^2 - n) \sum_{i=1}^n \sum_{j=1}^n a_{ij}$$

where n =order of A_{xx} . The mean for a_{ij} for the sub-matrices of values outside the diagonal (A_{xy}) is:

$$1 / rc \sum_{i=1}^r \sum_{j=1}^c a_{ij}$$

where r =number of rows, and c =number of columns in A_{xy} .

Connectedness index (IC) and general connectedness index (IGC)

Foulley *et al.* (1992) presented the concept of degree or level of non-connectedness, related to the prediction error variance (VEP) of the genetic effects of a complete model with the VEP of a reduced model that excludes fixed effects. These authors proposed the index of connectedness (IC), which is based on the use of a vector of coefficients of contrasts (x):

$$IC(x) = x' C_R x / x' C x$$

where C_R is a part of the inverse of the coefficient matrix in a reduced model, and C is part of the inverse of the same coefficient matrix in a complete model. This index reflects the reduction in VEP when the fixed effects are precisely known or do not exist. The $IC(x)$ varies between 0 and 1 and is close to 1 when the animals are well connected.

Foulley *et al.* (1992) posited the hypothesis that a well-connected data base is necessarily orthogonal (balanced). If two factors are orthogonal to each other, there is no bias in one of the factors if the other is removed from the model. If a random factor is totally orthogonal to a fixed factor eliminated from the reduced model, VEP is the same for random

donde C_R es una parte de la inversa de la matriz de coeficientes en un modelo reducido, y C es una parte de la inversa de la misma matriz de coeficientes en un modelo completo. Este índice refleja la reducción en la *VEP* cuando los efectos fijos se conocen exactamente o no existen. El $IC(x)$ varía entre 0 y 1, siendo cercano a 1 cuando los animales están bien conectados.

Foulley *et al.* (1992) plantearon la hipótesis de que una base de datos bien conectada es necesariamente ortogonal (balanceada). Si dos factores son ortogonales entre sí, no hay sesgo en uno de los factores si el otro se quita del modelo. Si un factor aleatorio es totalmente ortogonal a un factor fijo eliminado en el modelo reducido, la *VEP* es la misma para efectos aleatorios en el modelo completo y en el reducido. Por tanto, las formas cuadráticas sobre la base de la inversa de la matriz de coeficientes del modelo completo y del reducido deberían ser equivalentes, y el $IC(x)$ sería igual a uno.

Adicionalmente, Foulley *et al.* (1992) propusieron un índice general de conectividad (*IGC*) entre GC y VG, que es la media geométrica de los eigenvalores (μ_i) de $[C_R^{zz} - \mu_i C^{zz}]c_i = 0$, determinada por:

$$IGC = (|C_R|/|C|)^{1/n}$$

donde $|C|$ y $|C_R|$ son los determinantes de la inversa de las matrices de *VEP* con y sin efectos fijos en el modelo, y n es el número de animales o rango de las columnas de la matriz de incidencia de los parámetros a los que relacionan las subunidades C_R y C .

Coefficiente de determinación (CD) e índices generales de precisión

Aunque la ortogonalidad de los datos propuesta por Foulley *et al.* (1992) es deseable, Laloë (1993) menciona que una medida de precisión es más apropiada para determinar si los animales de diferentes ambientes pueden ser comparados, y propone el coeficiente de determinación (*CD*) de un vector de contrastes (x) de VG, que es la forma general del *CD* individual o exactitud de la predicción del VG. El $CD(x)$ corresponde al cuadrado de la correlación entre la diferencia predicha y verdadera de valores genéticos; también varía de cero a uno y se interpreta igual (mide la exactitud de la diferencia

effects in both the complete and the reduced model. Therefore, the quadratic forms on the base of the inverse of the coefficient matrix of the complete model and of the reduced model should be equivalent, and $IC(x)$ would be equal to one.

Additionally, Foulley *et al.* (1992) proposed a general index of connectedness (*IGC*) between GC and VG, which is the geometric mean of the eigenvalues (μ_i) of $[C_R^{zz} - \mu_i C^{zz}]c_i = 0$, determined by:

$$IGC = (|C_R|/|C|)^{1/n}$$

where $|C|$ and $|C_R|$ are determinant of the inverse of the *VEP* matrices with and without fixed effects in the model, and n is the number of animals or column rank of the incidence matrix of parameters to which the subunits C_R and C are related.

Coefficient of determination (CD) and general precision indexes

Although data orthogonality, proposed by Foulley *et al.* (1992), is desirable, Laloë (1993) mentions that a measure of precision is more appropriate to determine whether animals from different environments can be compared and proposes the coefficient of determination (*CD*) of a vector of contrasts (x) of VG, which is the general form of individual *CD* or accuracy of VG prediction. $CD(x)$ corresponds to the square of the correlation between the predicted and true differences of the genetic values; it also varies between zero and one and is interpreted in the same way (it measures the accuracy of the difference in genetic value between two animals, or GC). Thus, the accuracy or reliability of the comparison of VG between animal i and j is given by:

$$CD(x_{ij}) = 1 - \frac{Var(\hat{a}_i - \hat{a}_j)}{Var(a_i - a_j)} = 1 - \lambda \frac{C_{ii}^{zz} + C_{jj}^{zz} - 2C_{ij}^{zz}}{A_{ii} + A_{jj} - 2A_{ij}}$$

where x_{ij} is the contrast between animal i and j , $\lambda = \sigma_e^2 / \sigma_a^2$, A_{ii} , A_{ij} and A_{jj} are elements of the matrix of additive genetic relationships, C_{ii}^{zz} , C_{ij}^{zz} , and C_{jj}^{zz} are elements of the inverse of the coefficient matrix of the mixed model equations. In matrix form, *CD* is

del valor genético entre dos animales o GC). Así, la exactitud o confiabilidad de la comparación del VG entre el animal i y el j , está dada por:

$$CD(x_{ij}) = 1 - \frac{Var(\hat{a}_i - \hat{a}_j)}{Var(a_i - a_j)} = 1 - \lambda \frac{C_{ii}^{zz} + C_{jj}^{zz} - 2C_{ij}^{zz}}{A_{ii} + A_{jj} - 2A_{ij}}$$

donde x_{ij} es el contraste entre el animal i y el j , $\lambda = \sigma_e^2 / \sigma_a^2$, A_{ii} , A_{ij} y A_{jj} son los elementos de la matriz de relaciones genéticas aditivas, C_{ii}^{zz} , C_{ij}^{zz} , y C_{jj}^{zz} son los elementos de la inversa de la matriz de coeficientes de las ecuaciones de modelos mixtos. En forma matricial, el CD es:

$$CD(x) = 1 - \frac{\lambda(x'C^{zz}x)}{x'Ax} = \frac{x'(A - \lambda C^{zz})x}{x'Ax}$$

donde x es un vector cuyos elementos suman cero, $C^{zz} = (Z'MZ + \lambda A^{-1})^{-1} \sigma_e^2$, y $M = I - X(X'X)^{-1}X'$.

Laloë (1993) también propone dos índices generales de precisión usando las formas cuadráticas del CD , y relacionándolas con los eigenvalores (μ_i) y los eigenvectores (c_i) resultantes de la solución al problema generalizado de los eigenvalores:

$$[(A - \lambda C^{zz}) - \mu_i A]c_i = 0$$

El número de eigenvalores es igual al número de VG predichos. El eigenvalor menor siempre será cero; el resto corresponde a todos los contrastes independientes posibles. Los μ deben ser ordenados en forma ascendente. Los dos estadísticos propuestos están en función de esos eigenvalores:

$$p_1 = \sum_{i=2}^n \frac{\mu_i}{n-1} \quad p_2 = \left[\prod_{i=2}^n \mu_i \right]^{1/(n-1)}$$

El primer índice (p_1) es la media aritmética de los $n-1$ eigenvalores, y el segundo índice (p_2) es la media geométrica de los mismos eigenvalores; p_2 es más sensitivo a eigenvalores bajos. Un valor nulo de un eigenvalor da un p_2 nulo, indicando que existe por lo menos un contraste sin información suplementaria (desconexión).

$$CD(x) = 1 - \frac{\lambda(x'C^{zz}x)}{x'Ax} = \frac{x'(A - \lambda C^{zz})x}{x'Ax}$$

where x is a vector whose elements add up to zero, $C^{zz} = (Z'MZ + \lambda A^{-1})^{-1} \sigma_e^2$, and $M = I - X(X'X)^{-1}X'$.

Laloë (1993) also proposes two general precision indexes using the quadratic forms of CD and relating them to the eigenvalues (μ_i) and the eigenvectors (c_i) resulting from the solution of the generalized problem of the eigenvalues:

$$[(A - \lambda C^{zz}) - \mu_i A]c_i = 0$$

The number of eigenvalues is equal to the number of predicted VG. The lowest eigenvalue will always be zero; the rest correspond to all possible independent contrasts. The μ should be arranged in ascending order. The two proposed statistics are in function of these eigenvalues:

$$p_1 = \sum_{i=2}^n \frac{\mu_i}{n-1} \quad p_2 = \left[\prod_{i=2}^n \mu_i \right]^{1/(n-1)}$$

The first index (p_1) is the arithmetic mean of the $n-1$ eigenvalues, and the second index (p_2) is the geometric mean of the same eigenvalues; p_2 is more sensitive to low eigenvalues. A null value of an eigenvalue gives a null p_2 , indicating that there is at least one contrast with no supplementary information (disconnection).

For large databases, Boichard *et al.* (1992) obtained a reasonable approximation (bias < 4 %) of $tr(A^{-1}\Omega)$ for an animal model with one class of fixed effects and one of random effects, where tr =trace and $\Omega = C^{-1} = (Z'MZ + \lambda A^{-1})^{-1}$. Since p_1 is a function of $tr(A^{-1}\Omega)$ ($p_1 = [1 - \lambda tr(A^{-1}\Omega)] / (n-1)$), this method can be used to approach p_1 in large databases.

VEP of the differences in VG between animals (VEP(x))

Genetic connectedness between GCs reduces VEP of the comparisons between animals from

Para bases de datos grandes, Boichard *et al.* (1992) obtuvieron una aproximación razonable (sesgo $< 4\%$) de $tr(A^{-1}\Omega)$ para un modelo animal con una clase de efectos fijos y una de efectos aleatorios, donde $tr = \text{traza}$ y $\Omega = C^{-1} = (Z'MZ + \lambda A^{-1})^{-1}$. Ya que p_1 es una función de $tr(A^{-1}\Omega)$ ($p_1 = [1 - \lambda \text{tr}(A^{-1}\Omega)] / (n-1)$), este método se puede usar para aproximarse a p_1 en bases de datos grandes.

VEP de las diferencias en VG entre animales (VEP(x))

La conectividad genética entre GC reduce la VEP de las comparaciones entre los animales en diferentes GC, porque la VEP de las comparaciones entre animales emparentados es menor que la obtenida entre animales no emparentados, y porque las conexiones genéticas reducen la variabilidad de las estimaciones de las diferencias entre los efectos de GC. Así, Kennedy y Trus (1993) propusieron el promedio de la VEP de las diferencias en VG ($\hat{a}$) entre animales con registro en diferentes GC, o la VEP de la comparación entre animales, como una medida estadística lógica y apropiada del grado de conectividad de los datos. La VEP de la diferencia entre dos animales (VEPD), uno del i -ésimo y otro del j -ésimo grupo, está dada por:

$$VEPD_{ij} = Var(\hat{a}_i) + Var(\hat{a}_j) - 2Cov(\hat{a}_i, \hat{a}_j)$$

Ambas, el promedio de la VEP y la VEP de las diferencias en VG entre animales de diferentes GC, pueden obtenerse para un contraste (x):

$$VEP(x) = x'C^{zz}x\sigma_e^2$$

donde C^{zz} es una parte de la inversa de la matriz de coeficientes de las ecuaciones de modelos mixtos. Esta medida no tiene un rango definido y está muy relacionada con el CD, que es más fácil de interpretar. Además, la obtención de esta varianza para bases de datos grandes, como ocurre en muchas evaluaciones genéticas, es prácticamente imposible (Carneiro *et al.*, 2001a; Mathur *et al.*, 2002a), por lo que la aplicación rutinaria de este método es poco factible (Roso *et al.*, 2004).

different GC because the VEP of the comparisons between related animals is lower than that obtained when the animals are not related and because the genetic connections reduce the variability of the estimations of the differences among GC effects. Therefore, Kennedy and Trus (1993) proposed the average VEP of the differences in VG ($\hat{a}$) between animals with records in different GC, or the VEP of the comparison between animals, as a logical statistical measure of the degree of connectedness of the data. The VEP of the difference between two animals (VEPD), one from the i^{th} and the other from the j^{th} group, is given by:

$$VEPD_{ij} = Var(\hat{a}_i) + Var(\hat{a}_j) - 2Cov(\hat{a}_i, \hat{a}_j)$$

Both, the average VEP and the VEP of the differences in VG between animals of different GC, can be obtained for a contrast (x):

$$VEP(x) = x'C^{zz}x\sigma_e^2$$

where C^{zz} is part of the inverse of the coefficient matrix of the mixed model equations. This measure does not have a defined range and is highly related to the CD, which is easier to interpret. Moreover, obtaining this variance for large databases, as occurs in many genetic evaluations, is practically impossible (Carneiro *et al.*, 2001a; Mathur *et al.*, 2002a), and so the routine application of this method is not feasible (Roso *et al.*, 2004).

Gene flow (FG)

This method is an extension, together with the animal model, of the $X'Z$ matrix used with the sire model to measure connectedness, and simply traces the flow of genes from one GC to the next. FG was proposed by Kennedy and Trus (1993), and its main virtue is its computational ease. The method measures historical connectedness (through ancestors) by multiplying the matrices $X'ZTQ$. Gene flow can be applied as a geometric measurement of connectedness or as a quantitative measurement between two GC, by taking the absolute difference between the sum of the elements of the diagonal and the sum of the elements outside the diagonal (a small value indicates high connectedness). The Q matrix identifies founder animals with respect to its

Flujo de genes (FG)

Este método es una extensión, con el modelo animal, de la matriz $X'Z$ usada para medir conectividad con el modelo de semental, y simplemente traza la transferencia de genes de un GC al siguiente. El FG fue propuesto por Kennedy y Trus (1993) su virtud principal es la facilidad de cómputo y mide conectividad histórica (a través de ancestros) mediante la multiplicación de las matrices $X'ZTQ$. El FG puede aplicarse como una medida geométrica de conectividad o como una medida cuantitativa entre dos GC, al tomar la diferencia absoluta entre la suma de los elementos de la diagonal y la suma de los elementos fuera de la diagonal (un valor pequeño indica alta conectividad). La matriz Q identifica animales fundadores con respecto a su GC de origen, y T es una matriz triangular inferior que traza el flujo de genes de una generación a la siguiente; así $A = T'WT$; donde W es la diagonal de la matriz de las varianzas de muestreo mendeliano. Para obtener la matriz T es necesario ordenar los animales por generación y seguir las instrucciones: 1) todos los elementos de la diagonal son uno; 2) todos los elementos a la derecha de la diagonal son cero; 3) los elementos a la izquierda de la diagonal que corresponden a las filas de los animales base son cero; y 4) los elementos de las filas a la izquierda de la diagonal de los animales con padres conocidos son el promedio de los elementos correspondientes a la fila del padre y de la madre. No es necesario usar toda la matriz T generada, normalmente sólo se usan las columnas que pertenecen a los animales fundadores. Los elementos dentro de todas las filas para animales con padres conocidos, muestran la fracción de genes de los animales fundadores y suman uno.

Varianza de la deriva genética (VDG)

Este método se basa en el promedio de las relaciones genéticas entre y dentro de los GC y no requiere la identificación de animales base o ancestros (Kennedy y Trus, 1993). Considera el cómputo de la matriz A o partes de ella, y aunque es más difícil que la obtención de A^{-1} , su cómputo es factible para bases de datos grandes.

La matriz $X'ZAZ'X$ mide la suma de las relaciones genéticas entre y dentro de GC. El promedio de relaciones genéticas, entre y dentro de GC, se obtiene

GC of origin, and T is a lower triangular matrix that traces gene flow from one generation to the next, so that $A = T'WT$, where W is the diagonal of the matrix of variances from Mendelian sampling. To obtain matrix T , it is necessary to order the animals by generation and follow these instructions: 1) all the elements of the diagonal are one; 2) all the elements to the right of the diagonal are zero, 3) the elements to the left of the diagonal that correspond to the rows of base animals are zero, and 4) the elements of the rows to the left of the diagonal of animals with known parents are the average of the elements corresponding to the row of the sire and the mother. It is not necessary to use the entire T matrix generated, normally only the columns that belong to founder animals are used. The elements in all the rows for animals with known parents show the fraction of genes from founder animals and that add up to one.

Genetic drift variance (VDG)

This method is based on the average of the genetic relationships between and within GC and does not require identification of base animals or ancestors (Kennedy and Trus, 1993). It considers the computation of matrix A or parts of it and, although it is more difficult than obtaining A^{-1} , its computation is feasible for large databases.

The matrix $X'ZAZ'X$ measures the sum of genetic relationships among and within GC. The average of genetic relationships, among and within GC, is obtained by dividing each element of the diagonal by the number of records squared of each GC and the elements outside the diagonal by the product of the number of records in the GC involved. The resulting matrix can be interpreted as the genetic components of (co)variance of the genetic drift among GC, thus, the VDG between groups would be equal to the sum of the elements of the diagonal minus the sum of the elements outside the diagonal, where a small value indicates a high degree of connectedness (Kennedy and Trus, 1993).

Variance of the differences between effects of GC (VdGC)

This method, proposed by Kennedy and Trus (1993), is based on the computation of the

al dividir cada elemento de la diagonal entre el número de registros al cuadrado de cada GC, y los elementos fuera de la diagonal entre el producto del número de registros en los GC involucrados. La matriz resultante se puede interpretar como los componentes genéticos de la (co)varianza de la deriva genética entre los GC; así, la VdG entre grupos sería igual a la suma de los elementos de la diagonal menos la suma de los elementos fuera de la diagonal, donde un valor pequeño indica un alto grado conectividad (Kennedy y Trus, 1993).

Varianza de las diferencias entre efectos de GC (VdGC)

Este método, propuesto por Kennedy y Trus (1993), se basa en el cómputo de la matriz de (co) varianzas de los efectos de GC, representada por $\{XX - X'Z(Z'Z + \lambda A^{-1})^{-1} Z'X\}^{-1} \sigma_e^2$. Los elementos positivos fuera de la diagonal son el resultado de las relaciones genéticas entre animales de diferentes GC e indican la conectividad genética; para grupos no conectados estos elementos son cero. De acuerdo con Mathur *et al.* (2002a) y Roso *et al.* (2004), la varianza de la diferencia entre el i -ésimo y el j -ésimo grupo (gc) está dada por:

$$\begin{aligned} VdGC_{ij} &= Var(\hat{gc}_i - \hat{gc}_j) \\ &= Var(\hat{gc}_i) + Var(\hat{gc}_j) - 2Cov(\hat{gc}_i, \hat{gc}_j) \end{aligned}$$

La $VdGC$ mide la precisión con que son estimadas las diferencias en las medias genéticas entre GC (Kuehn *et al.*, 2008b); sin embargo, para poblaciones grandes con muchos GC, el cálculo de la matriz de (co)varianzas de los efectos de GC puede ser difícil (Kennedy y Trus, 1993). Además del desafío computacional, la VEP de la diferencia entre las estimaciones de efectos de GC tiene otra desventaja; es dependiente del tamaño y la estructura de los GC, así como de la naturaleza de las conexiones entre ellos (Mathur *et al.*, 2002a).

Grado de conectividad (GrC)

Para separar los efectos de tamaño de GC y estructura en la conectividad, Mathur *et al.* (2002a,b) definieron este método como la correlación entre los

matrix of (co)variances of GC effects, represented by $\{XX - X'Z(Z'Z + \lambda A^{-1})^{-1} Z'X\}^{-1} \sigma_e^2$. The positive elements outside the diagonal are the result of the genetic relationships between animals from different GC and indicate genetic connectedness; for non-connected groups these elements are zero. According to Mathur *et al.* (2002a) and Roso *et al.* (2004), the variance of the difference between the i th and the j th group (gc) is given by:

$$\begin{aligned} VdGC_{ij} &= Var(\hat{gc}_i - \hat{gc}_j) \\ &= Var(\hat{gc}_i) + Var(\hat{gc}_j) - 2Cov(\hat{gc}_i, \hat{gc}_j) \end{aligned}$$

$VdGC$ measures the precision with which the differences in genetic means between GC are estimated (Kuehn *et al.*, 2008b). However, for large populations with many GC, calculation of the matrix of (co)variances of GC effects can be difficult (Kennedy and Trus, 1993). Besides the computational challenge, the VEP of the difference between estimations of GC effects has another disadvantage: it is dependent on the size and structure of the GC, as well as on the nature of the connections among them (Mathur *et al.*, 2002a).

Degree of connectedness (GrC)

To separate the effects of GC size from structure in connectedness, Mathur *et al.* (2002a,b) defined this method as the correlation between the estimated effects of two GC. Thus, the GrC between the i th and the j th GC is given by:

$$GrC_{ij} = \frac{Cov(\hat{gc}_i, \hat{gc}_j)}{\sqrt{Var(\hat{gc}_i)Var(\hat{gc}_j)}} \times 100$$

where the components of (co)variances with a simple animal model are obtained from the VEP matrix, which is given by the product of the error variance and the sub-matrix C^{zz} of the EMM.

According to Mathur *et al.* (2002a), to overcome the problem in obtaining the direct inverse of the coefficient matrix in large databases, the following procedure can be used:

efectos estimados de dos GC. Así, el GrC entre el i -ésimo y el j -ésimo GC está dado por:

$$GrC_{ij} = \frac{Cov(\hat{g}c_i, \hat{g}c_j)}{\sqrt{Var(\hat{g}c_i)Var(\hat{g}c_j)}} \times 100$$

donde los componentes de (co)varianzas con un modelo animal simple se obtienen de la matriz de VEP , que está dada por el producto de la varianza del error y la submatriz C^{zz} de las EMM.

De acuerdo con Mathur *et al.* (2002a), para superar el problema en la obtención de la inversa directa de la matriz de coeficientes en bases de datos grandes se puede usar este procedimiento:

$$CC^{-1} = I \quad \text{así,} \quad CC_i^{-1} = I_i$$

donde, C =matriz de coeficientes de las EMM, I =matriz identidad, I_i =un vector de la matriz identidad correspondiente al i -ésimo GC (un vector con uno para el GC y ceros), y C_i^{-1} =un vector de elementos de la inversa para el i -ésimo GC.

El vector C_i^{-1} se puede obtener para un GC a la vez por iteraciones. Esos vectores se combinan y el bloque de elementos de la inversa correspondientes al GC más reciente es extraído. Los elementos de la inversa son las (co)varianzas del error de predicción de los efectos de GC de interés. Este método puede usarse para obtener elementos de la inversa en las filas o columnas de cualquier matriz grande donde la inversa directa no es posible.

Correlación entre las varianzas de los errores de predicción de VG (r_{ij})

Tomando en cuenta que la covarianza del error de predicción (CEP) entre los VG de dos animales debería ser cero si no están conectados, Lewis *et al.* (1999) propusieron como criterio de conectividad una correlación estadística similar al GrC , definida como la correlación entre las varianzas de los errores de predicción de VG (r_{ij}):

$$r_{ij} = \frac{CEP(\hat{a}_i, \hat{a}_j)}{\sqrt{VEP(\hat{a}_i)VEP(\hat{a}_j)}}$$

$$CC^{-1} = I \quad \text{thus,} \quad CC_i^{-1} = I_i$$

where C =coefficient matrix of the EMM, I =identity matrix, I_i =a vector of the identity matrix corresponding to the i^{th} GC (one vector with one for the GC and zeros), and C_i^{-1} =a vector of elements of the inverse for the i^{th} GC.

The vector C_i^{-1} can be obtained for one GC at a same time through iterations. These vectors combine and the block of elements of the inverse corresponding to the most recent GC is extracted. The elements of the inverse are the (co)variances of the prediction error of GC effects of interest. This method can be used to obtain elements of the inverse in the rows or columns of any large matrix where the direct inverse is not possible.

Correlation between variances of VG (r_{ij}) prediction errors

Taking into account that the covariance of the prediction error (CEP) between the VG of two animals should be zero if they are not connected, Lewis *et al.* (1999) proposed, as a criterion of connectedness, a statistical correlation similar to GrC , defined as the correlation between the variances of the prediction errors of VG (r_{ij}):

$$r_{ij} = \frac{CEP(\hat{a}_i, \hat{a}_j)}{\sqrt{VEP(\hat{a}_i)VEP(\hat{a}_j)}}$$

where $\hat{a}_i$ is the predicted VG for the i^{th} animal. The authors suggest averaging this statistic for all pairs of animals in different GC to evaluate their connectedness.

Total number of direct genetic links among GC ($TLGD$)

This method was developed by Fries and Roso (1997), Fries (1998) and Roso *et al.* (2004) to measure the degree of connectedness among GC through direct genetic links due to common sires and dams. The basic steps for the algorithm and the criterion used for computing $TLGD$ are 1) calculate the number of direct genetic links between pairs of GC due to common sires (LGs) and dams (LGb);

Cuadro 1. Escala de los métodos[†] para determinar conectividad, valores que representan mejor conectividad (VRMC), requerimientos de matrices (C^{zz} y A) y capacidad de identificar grupos contemporáneos desconectados (GCD) y conectados indirectamente (GCCl).

Table 1. Scale of the methods[†] for determining connectedness, values that represent better connectedness (VRMC), matrix requirements (C^{zz} and A) and capacity for identifying disconnected (GCD) and indirectly connected (GCCl) contemporary group.

Métodos	GCD	GCCl	C^{zz}	A	Escala	VRMC	Autor
PRGA	✓	-	-	-	≥ 0	Mayores	Banos y Cady, 1988
IC	-	✓	✓	-	0 a 1	Mayores	Foulley <i>et al.</i> , 1992
CD	-	✓	✓	✓	0 a 1	Mayores	Laloë, 1993
VEP(x)	-	✓	✓	-	≥ 0	Menores	Kennedy y Trus, 1993
FG	-	-	-	-	≥ 0	Menores	Kennedy y Trus, 1993
VDG [§]	✓	-	✓	✓	0 a 1	Mayores	Kennedy y Trus, 1993
VdGC	-	✓	-	-	≥ 0	Menores	Kennedy y Trus, 1993
GrC	✓	✓	✓	-	0 a 100	Mayores	Mathur <i>et al.</i> , 2002b
r_{ij}	✓	✓	✓	-	-1 a 1	Extremos	Lewis <i>et al.</i> , 1999
TLGD	✓	✓	-	-	≥ 0	Mayores	Fries y Roso, 1997

[†] PRGA=promedio de las relaciones genéticas aditivas, IC=índice de conectividad, CD=coeficiente de determinación, VEP(x)=varianza del error de predicción de las diferencias en valores genéticos, FG=flujo de genes, VDG=varianza de la deriva genética, VdGC=varianza de las diferencias entre efectos de grupos contemporáneos, GrC=grado de conectividad, r_{ij} =correlación de las varianzas de los errores de predicción entre valores genéticos, TLGD=número de lazos genéticos directos totales debidos a sementales y hembras comunes. [§] C^{zz} =inversa de la matriz de coeficientes en un modelo animal. [§] La VDG para grupos contemporáneos (GC) desconectados (elementos fuera de la diagonal=cero) indica cierto grado de conectividad, por lo que la medida se modificó de la siguiente manera: VDG_2 entre el GC_i y $GC_j = 1 - [((DGC_i + DGC_j) - 2(FGC_{ij})) / (DGC_i + DGC_j)]$, donde DGC = elemento de la diagonal y FGC =elemento fuera de la diagonal. Con esta modificación se obtiene una medida proporcional de la varianza de la deriva genética entre dos GC, donde valores cercanos a uno indican mayor conectividad y valores de cero indican GC desconectados. [†] PRGA=average of additive genetic relationships; IC=connectedness index, CD=coefficient of determination; VEP(x)=variance of the prediction error of the differences in genetic values; FG=gene flow; VDG=genetic drift variance; VdGC=variance of the differences between effects of contemporary groups; GrC=degree of connectedness; r_{ij} =correlation of the variances of prediction errors among genetic values; TLGD=total number of direct genetic links due to common sires and dams; [§] C^{zz} =inverse of the coefficient matrix in an animal model. [§] VDG for disconnected (elements outside the diagonal zero) contemporary groups (GC) indicates a certain degree of connectedness; thus, the measure was modified in the following manner: VDG_2 between GC_i and $GC_j = 1 - [((DGC_i + DGC_j) - 2(FGC_{ij})) / (DGC_i + DGC_j)]$, where DGC =element of the diagonal and FGC =element outside the diagonal. With this modification, a proportional measure of the genetic drift variance between two GC is obtained, where values near one indicate greater connectedness, and values of zero indicate disconnected GC.

donde $\hat{a}_i$ es el VG predicho para el i -ésimo animal. Los autores sugieren promediar este estadístico para todos los pares de animales en diferentes GC con el fin de evaluar su conectividad.

Número total de lazos genéticos directos entre GC (TLGD)

Este método fue desarrollado por Fries y Roso (1997), Fries (1998) y Roso *et al.* (2004) para medir el grado de conectividad entre GC a través de lazos genéticos directos debidos a sementales y hembras comunes. Los pasos básicos para el algoritmo y el criterio usado para el cómputo de TLGD son: 1)

then, for each GC the total number of genetic links (LGT) is calculated with all the other GC; 2) the GC with the largest LGT is called “main group”, and the sires and dams with progeny in this group form the initial connected core; 3) all the GC are converted to zeros (LGT) and are analyzed recurrently counting the LGs and LGh generated by the initial connected core, at the end of each run the GC with $LGT > 9$ and at least three progenitors (sires or dams) are considered connected with the “main group”; 4) the rest of the GC (disconnected) are converted to zeros and the process is repeated until the number of connected GC is equal to that of the previous run.

calcular el número de lazos genéticos directos entre pares de GC debido a sementales (*LGS*) y hembras (*LGH*) comunes; luego, para cada GC se calcula el número total de lazos genéticos (*LGT*) con todos los otros GC; 2) el GC con mayor *LGT* es llamado “grupo principal”, y los sementales y hembras con progenie en este grupo forman el núcleo inicial conectado; 3) todos los GC se ponen en ceros (*LGT*) y son analizados recurrentemente contando los *LGS* y *LGH* generados por el núcleo inicial conectado; al final de cada corrida los GC con $LGT > 9$ y por lo menos tres progenitores (sementales o hembras) son considerados conectados con el “grupo principal”; y 4) el resto de los GC (desconectados) se ponen en ceros y se repite el proceso hasta que el número de GC conectados sea igual al de la corrida anterior.

Las parejas de GC que no tienen algún lazo genético directo pueden estar conectadas indirectamente a través de otros GC y consecuentemente tener alta exactitud de comparación de *VG* estimados entre ellos. Por esta razón, *TLGD* es inadecuado para indicar el grado de conectividad entre pares de GC. La media aritmética también es inadecuada, ya que un par de GC con igual media aritmética puede tener un grado de conectividad diferente. Potencialmente se sugiere el uso de la media armónica de los $LGT \left(2 / \left[(1/LGT_i) + (1/LGT_j) \right] \right)$ (Roso *et al.*, 2004). Para superar estas deficiencias, Roso y Schenkel (2006) desarrollaron el programa AMC - A Computer Program to Assess the Degree of Connectedness Among Contemporary Groups para medir el grado de conectividad entre los GC. El total de lazos genéticos es función del número y tipo de relaciones de parentesco entre animales en los diferentes GC.

Principales atributos de los métodos cuantitativos para determinar conectividad

Un método para determinar conectividad idealmente debe identificar GC desconectados, y conectados directa e indirectamente; además, debe medir el grado de conectividad, ser práctico, y fácil de aplicar e interpretar. En el Cuadro 1 se muestra la escala de los métodos de conectividad, requerimientos de matrices (C^{zz} y *A*), y su capacidad para identificar GC desconectados y conectados indirectamente. De los 10 métodos cuantitativos descritos, cinco identifican GC desconectados, de los cuales tres estiman conectividad en forma directa e indirecta (*GrC*, r_{ij} y

The pairs of GC that do not have a direct genetic link can be connected indirectly through other GC and consequently have high comparative precision of estimated *VG* between the pair members. For this reason, *TLGD* is not suitable for indicating the degree of connectedness between pairs of GC. The arithmetic mean is also unsuitable since a pair of GC with the same arithmetic mean can have a different degree of connectedness. Potentially, the use of the harmonic mean of the *LGT* is suggested $LGT \left(2 / \left[(1/LGT_i) + (1/LGT_j) \right] \right)$ (Roso *et al.*, 2004). To overcome these deficiencies, Roso and Schenkel (2006) developed the software AMC - A Computer Program to assess the degree of connectedness among contemporary groups. The total number of genetic links is a function of the number and type of genetic relationships among animals from different GC.

Main attributes of the quantitative methods for determining connectedness

A method for determining connectedness should identify disconnected and directly and indirectly connected GC; it should also measure the degree of connectedness, be practical and be easy to apply and interpret. Table 1 shows the scale of the connectedness methods, matrix requirements (C^{zz} and *A*), and their capacity to identify disconnected and indirectly connected GC. Of the ten quantitative methods described, five identify disconnected GC, and three of these estimate connectedness directly and indirectly (*GrC*, r_{ij} and *TLGD*). For large databases, the most complex matrices are the inverse of the coefficient matrix (C^{zz}) and occasionally the *A* matrix; of the 10 methods considered here, six require C^{zz} and two require the *A* matrix. Of the three methods that can identify disconnected and indirectly connected GC, only the *TLGD* method does not require the matrices. The interpretation of connectedness values depends on the measurement scale; for five methods, this interpretation is facilitated since their scale is a defined interval, while for the rest, the scale is greater than or equal to zero, with no upper limit, making interpretation of connectedness values difficult.

Choosing the best method for evaluating genetic connectedness among GC depends on the size and

TLGD). Para bases de datos grandes, las matrices más complejas para obtener son la inversa de la matriz de coeficientes (C^{zz}) y en algunas ocasiones la matriz A , y de los 10 métodos considerados seis requieren C^{zz} y dos la matriz A . De los tres métodos que pueden identificar GC desconectados y conectados indirectamente, sólo el método *TLGD* no requiere las matrices complejas mencionadas. La interpretación de los valores de conectividad depende de la escala de medición; para cinco métodos esta interpretación se facilita dado que su escala es un intervalo definido, mientras que para el resto la escala es mayor o igual que cero, sin un límite superior, lo que dificulta la interpretación de los valores de conectividad.

La elección del mejor método para evaluar la conectividad genética entre GC depende del tamaño y estructura de los datos de poblaciones animales, por lo que se requiere comparar los métodos descritos, considerando una misma estructura de datos para escenarios representativos en evaluaciones genéticas.

CONCLUSIONES

Un método para determinar la conectividad, idealmente debe identificar grupos contemporáneos desconectados, y conectados directa e indirectamente; además, debe evaluar el grado de conectividad, ser práctico, y fácil de aplicar e interpretar. Las metodologías cualitativas para evaluar la conectividad entre grupos contemporáneos, sólo permiten separar los grupos que están conectados de los que no lo están y son útiles cuando se utiliza el modelo de semental para la predicción de los valores genéticos. Debido al uso generalizado de las ecuaciones de modelos mixtos de Henderson con el modelo animal, se utilizan métodos cuantitativos que evalúan adicionalmente el grado de conectividad, acercándose a las características de un método ideal. Los métodos cuantitativos descritos en este ensayo son alternativas que se deben comparar considerando una misma estructura de los datos, para escenarios comunes en evaluaciones genéticas de animales.

AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Ciencia y Tecnología y al Consejo Mexiquense de Ciencia y Tecnología por el financiamiento otorgado para los estudios de Maestría en Ciencias del primer autor.

structure of the animal population data. This requires comparing the described methods, considering the same data structure for representative scenarios in genetic evaluations.

CONCLUSIONS

A method for determining connectedness should ideally identify disconnected and directly or indirectly connected contemporary groups. Moreover, it should evaluate the degree of connectedness; it should be practical and easy to apply and interpret. Qualitative methodologies for evaluating connectedness among contemporary groups only permit separating groups that are connected from those that are not; they are useful when the sire model is used to predict genetic values. Because of the generalized use of Henderson's mixed model equations with the animal model, quantitative methods are used to evaluate, additionally, the degree of connectedness, thus coming closer to the characteristics of an ideal method. The methods described in this essay need to be compared using the same data structure and for common scenarios in animal genetic evaluations.

—End of the English version—



LITERATURA CITADA

- Banos, G., and R. A. Cady. 1988. Genetic relationship between the United States and Canadian Holstein bull populations. *J. Dairy Sci.* 71: 1346-1354.
- Boichard, D., L. R. Schaeffer, and A. J. Lee. 1992. Approximate restricted maximum likelihood and approximate prediction error variance of the Mendelian sampling effect. *Genet. Sel. Evol.* 24: 331-343.
- Carneiro S., A. P., R. A. Torres, R. F. Euclides, M. A. Silva, P. S. Lopes, P. L. S. Carneiro, e R. A. T. Filho. 2001a. Efeito da conexidade de dados sobre o valor fenotípico médio e a variancia genética aditiva. *Rev. Bras. Zoot.* 30: 336-341.
- Carneiro S., A. P., R. A. Torres, R. F. Euclides, M. A. Silva, P. S. Lopes, P. L. S. Carneiro, e R. A. T. Filho. 2001b. Efeito da conexidade de dados sobre a acurácia dos testes de prole e performance. *Rev. Bras. Zoot.* 30: 342-347.
- Clément V., B. Bibé, E. Verrier, J. M. Elsen, E. Manfredi, J. Bouix, and E. Hanocq. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Genet. Sel. Evol.* 33: 369-395.

- Fernando, R. L., D. Gianola, and M. Grossman. 1983. Identifying all connected subsets in a two-way classification without interaction. *J. Dairy Sci.* 66: 1399-1402.
- Fouilloux, M. N., S. Minery, S. Mattalia, and D. Laloë. 2006. Assessment of connectedness in the international genetic evaluation of Simmental and Montbéliard breeds. *Interbull Bulletin* 35: 129-135.
- Fouilloux, M. N., B. Clément, and D. Laloë. 2008. Measuring connectedness among herds in mixed linear models: from theory to practice in large-sized genetic evaluations. *Genet. Sel. Evol.* 40: 145-159.
- Foulley, J. L., E. Hanocq, and 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., e V. M. Roso. 1997. Conectabilidade em avaliações genéticas de gado de corte: uma proposta heurística. *In: Anais XXXIV Reun. Soc. Bras. Zoot., Juiz de Fora, Brasil.* pp: 159-161.
- Fries, L. A. 1998. Connectability in beef cattle genetic evaluation: the heuristic approach used in MILC.FOR. *In: Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia* 27: 449-450.
- Hanocq E., D., L. Tiphine, et B. Bibë. 1999. Le point sur la notion de connexion en génétique animale. *INRA Prod. Anim.* 12(2): 101-111.
- Henderson, C. R. 1973. Sire evaluation and genetic trends. *In: Proceedings of the Animal Breeding in Genetic Symposium in Honor of Dr. Jay L. Lush. American Society of Animal Science. Champaign, IL.* pp: 10-41.
- Huisman, A. E., B. Tier, and D. J. Brown. 2006. On assessing contrasts between groups of animals. *Livest. Sci.* 104: 254-267.
- Kennedy, B. W., and D. Trus. 1993. Considerations on genetic connectedness between management units under an animal model. *J. Anim. Sci.* 71: 2341-2352.
- Kuehn, L. A., R. M. Lewis, and D. R. Notter. 2007. Managing the risk of comparing estimated breeding values across flocks or herds through connectedness: a review and application. *Genet. Sel. Evol.* 39: 225-247.
- Kuehn, L. A., D. R. Notter, and R. M. Lewis. 2008a. Assessing genetic gain, inbreeding, and bias attributable to different flock genetic means in alternative sheep sire referencing schemes. *J. Anim. Sci.* 86: 526-535.
- Kuehn, L. A., D. R. Notter, G. J. Nieuwhof, and R. M. Lewis. 2008b. Changes in connectedness over time in alternative sheep sire referencing schemes. *J. Anim. Sci.* 86: 536-544.
- Kuehn, L. A., R. M. Lewis, and D. R. Notter. 2009. Connectedness in Targhee and Suffolk flocks participating in the United States national sheep improvement program. *J. Anim. Sci.* 87: 507-515.
- Laloë, D. 1993. Precision and information in linear models of genetic evaluation. *Genet. Sel. Evol.* 25: 557-576.
- Laloë, D., F. Phocas, and F. Ménissier. 1996. Considerations on measures of precision and connectedness in mixed linear models of genetic evaluation. *Genet. Sel. Evol.* 28: 359-378.
- Lewis, R. M., R. E. Crump, G. Simm, and R. Thompson. 1999. Assessing connectedness in across-flock genetic evaluations. *In: Proc. Br. Soc. Anim. Sci., Scarborough, England.* p. 121.
- Lewis R. M., and G. Simm. 2000. Selection strategies in sire referencing schemes in sheep. *Livest. Prod. Sci.* 67: 129-141.
- Mathur, P. K., B. Sullivan, and J. Chesnais. 2002a. Estimation of the degree of connectedness between herds or management groups in the Canadian swine population. http://www.ccsi.ca/include/docs/connectedness/connec-tedness_asds_1998.PDF. (consultado: marzo, 2007).
- Mathur, P. K., B. Sullivan, and J. Chesnais. 2002b. Measuring connectedness: concept and application to a large industry breeding program. *In: Proc. 7th World Congr. Genet. Appl. Livest. Prod., August 19-23. Montpellier, France.* p. 13 (abstract).
- Nakaoka, H., C. Gaillard, K. Fujinaka, N. Watanabe, M. Ito, K. Kawada, T. Ibi, Y. Sasae, and Y. Sasaki. 2009. The use of link provider data to improve national genetic evaluation across weakly connected subpopulations. *J. Anim. Sci.* 87: 62-71.
- Pegolo, N. T., D. Laloë, H. N. de Oliveira, R. B. Lôbo, and M. N. Fouilloux. 2012. Trends on the genetic connectedness measures among Nelore beef cattle herds. *J. Anim. Breed. Genet.* 129: 20-29.
- 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., R. S. Schenkel, and S. P. Miller. 2004. Degree of connectedness among groups of centrally tested beef bulls. *Can. J. Anim. Sci.* 84: 37-47.
- Roso, V. M., and R. S. Schenkel. 2006. AMC – A computer program to assess the degree of connectedness among contemporary groups. *In: Proc. 8th World Congr. Genet. Appl. Livest. Prod., August, 13 to 18, Belo Horizonte, Brazil.* Communication no. 27-26.
- Searle, S. R. 1987. *Linear Models for Unbalanced Data.* John Wiley & Sons, New York. USA. 536 p.
- Simm, G., R. M. Lewis, J. E. Collings, and G. J. Nieuwhof. 2001. Use of sire referencing schemes to select for improved carcass composition in sheep. *J. Anim. Sci.* 79 (E. Suppl): E255-E259.
- Soga, N., M. L. Spangler, C. R. Schwab, P. J. Berger, and T. J. Baas. 2010. Comparison of connectedness measures and changes in connectedness of the U. S. Duroc population. *In: Proc. 9th World Congr. Genet. Appl. Livest. Prod., July 31 to August 6, Leipzig, Germany.* 4 p.
- Sun, C. Y., C. K. Wang, Y. C. Wang, Y. Zhang, and Q. Zhang. 2009. Evaluation of connectedness between herds for three pig breeds in China. *Animal* 3(4): 482-485.
- Tarrés, J., M. Fina, and J. Piedrafit. 2010. Connectedness among herds of beef cattle bred under natural service. *Genet. Sel. Evol.* 42: 6-15.
- Tosh, J. J., and J. W. Wilton. 1994. Effect of data structure on variance of prediction error and accuracy of genetic evaluation. *J. Anim. Sci.* 72: 2568-2577.
- Veselá, Z., J. Přibyl, L. Vostrý, and L. Štolc. 2007. Stochastic simulation of the influence of insemination on the estimation of breeding value and its reliability. *Czech J. Anim. Sci.* 52(8): 236-248.