



EDITORIAL

Statistical methods in animal breeding and Bach's harpsichord concertos

Johan Sebastian Bach is one of my favourite composers, so I was pleased to have the opportunity to visit the city where the composer lived from 1723 to his death in 1750 whilst attending the 9th WCGALP in Leipzig. As a consequence, I can hardly differentiate between writing this editorial note and listening to Bach's harpsichord concertos (dating back probably to the period 1730–1738). I hope the music will introduce neither bias nor variance!

Keyboard Concerto No. 1 in D minor, BWV 1052. First movement: From all statistical methods that animal breeders have used since the early days, only those that are sound and have reasonable statistical properties have survived the 'test of time': BLUP, REML, Bayesian methods helped by MCMC algorithms. Who remembers today the intricate and completely *ad hoc* methods that appeared after somebody thought 'why if we do that ...?' Sooner or later the lack of 'reasonable' properties will send the method to the dump. One may ask what 'reasonable' means in this context, but unbiasedness *per se*, is not enough to keep us happy! We want also to look at a risk function, and we have often looked at quadratic ones (mean square error, or variance for unbiased estimators or predictors). Second movement: the music demands more from us! Again, nice properties of unbiasedness and minimum variance may not be enough, as we want our method to account for past selection. In general, methods based on likelihood have good and sound properties, but *ad hoc* 'profile likelihood' or conditional likelihood methods can lead us into turbulent waters with one or more paddles that do not move!

Third movement: *allegro!* Attending WCGALP is a feast that we have been awaiting for 4 years. Personally I expect that something will be revealed to me at the meeting: the idea behind the statistical method, its usefulness, weaknesses, restrictions, fortresses; anything that can lead me into the light! Clearly, different people have different expectations of the talks. Usually the presentations are not detailed enough for us to grasp everything we need to know. We may not be smart enough to understand, or do not have

enough courses in math-stats to get a hint of what is going on. Oftentimes the methods are ill posed but the speaker does a good job in presenting the paper. The other way around is less frequent. There were numerous presentations, so allow me to comment briefly on three subjects: two are related to models and the other one to a statistical method for estimating (co)variance components. I will begin with the latter.

Concerto II in E major, BWV 1053. First movement: *allegro*. Karin Meyer gave a contributed presentation on her recent paper with Mark Kirkpatrick (*Genetics* **85**, 1097–1110). They proposed to use a quadratic penalty on the restricted likelihood to shrink the genetic covariance matrix. The idea behind the procedure is to reduce bias in the eigenvalues of the product of the inverse of the phenotypic covariance matrix and the additive genetic covariance matrix, and to reduce sampling errors in the estimate of the latter matrix. Second movement: *siciliano*. The idea of penalizing a covariance matrix is not new, see the discussion by Yang & Berger (1994, *Ann. Stat.* **22**, 1195–1215). They observed that estimation is not invariant to the choice of the loss function, a 'rather perplexing' issue. (Third movement: *allegro*). A similar message is conveyed by Meyer and Kirkpatrick for the selection or 'tuning' parameter that controls the amount of shrinkage. They go on to propose cross-validation for the task. However, they reasonably conclude that 'further work is required to see whether suitable rules of thumb can be established'. To add all things up: a potentially useful method that may be tried for small, multi-colinear, or poorly identified data sets. I fail to see why we should use the method for the usual large data set with enough genetic information to estimate the additive genetic covariance matrix, instead of regular REML or Bayes-MCMC methods. Clearly, research on sound methods to calculate the value of the shrinkage parameter is in great need. Like the movement, *allegro ma non tanto!*, Karin was enthusiastic and entertaining as usual, and she did point out the difficulties in estimating the 'tuning' parameter.

I now move to the Concerto V in F Major, BWV 1056. First movement: *allegro moderato*. Although I have mostly worked in statistical methods applied to genetic evaluation, beef cattle production motivates much of my research. Breeders are always telling me how they appreciate bulls delivering homogeneous progeny for the market. In that regard, evaluation and selection for environmental homogeneity would be a big plus. I had read some material on the subject before Daniel Sorensen's presentation, but his gave me a good account of the state of the art dealing with estimation in such a model. Second movement: *largo*. Daniel's view of the topic is rather cautious. As most results from the literature are from field data, which are prone to possible unaccounted skewness and heavy tails in the inferential distribution, Daniel promotes a careful look at the estimates and pleads for proper selection experiments that provide the sources of information needed to estimate the parameters of the model to a certain degree of accuracy. Thus, it seems that we have a good deal of homework to do before the model for environmental homogeneity can be used for commercial applications! The third movement (*presto*) is over!

I would also like to mention 'social' genetic effects, while listening to the Concerto V in F minor, BWV 1056. Start with *allegro moderato* and notice that the word 'social' is gaining more acceptance than 'competition' to describe the genetic effects involved. Clearly, it is more general! *Andante*: my interest in the subject grew from independent work in my lab (one of my students worked in forest genetic evalua-

tion) to mainstream animal breeding approaches of Bill Muir and Peter Bijma. They both gave interesting presentations that cover the subject, either entirely or in part. Esther Allen and Laureanne Canario were faced with the problem of estimating the genetic dispersion parameters in data from poultry and pigs, respectively. Esther's data behaved well while fitting the Muir-Bijma model. In Canario's work, some sort of standardization of the variance had to be carried out in order to capture more genetic variation. We (Cantet & Cappa 2008, *J. Anim. Breed. Genet.* **125**, 371–381) pointed out the need of standardization for the additive social variance – and the covariance. I am not sure whether our *intensities of competition* (IC; values of a covariate) are well understood. However, I am sure that the ICs are difficult to implement in animal breeding without some sort of design (for example, moving some animals from one pen to another over some periods). They need to be identified: *allegro assai*, third movement. If animals are not separately identified in their additive social effects, the effects will be confounded with management (pen) effects and are likely to go astray, as a lot of estimates of the additive (co)variances have shown when adapting the model to different species.

R. J. C. Cantet
 Universidad de Buenos Aires,
 Buenos Aires, Argentina
 E-mail: rcantet@agro.uba.ar