Chapter 15

Non-additive effects and finite locus models

Julius van der Werf

Non-additive genetic effects and finite locus models

For genetic models at the locus level non-additive genetic effects are clearly defined as dominance and epistasis. At the population level, for multiple loci, we can define dominance effects and dominance variation. Assuming small contributions from many unlinked loci, genetic covariance between individuals in a non-inbred random mating population is a linear function of the genetic variance components and genetic relationships (Cockerham, 1954).

Including dominance in a mixed model can be done as a 'polygenic' dominance effects, with variance covariance matrix equal to the dominance variance multiplied by the dominance relationships matrix. Dominance relationships can be derived from the additive relationships among the parents

$$d_{xy} = 0.25(a_{s_x s_y} a_{d_x d_y} + a_{s_x d_y} a_{d_x s_y})$$

or using an algorithm for large populations (Hoeschele and Van Raden, 1991) similar to Henderson's rules.

Inbreeding has two effects in a dominance model:

1) with dominance existing, inbreeding will depress phenotypic performance (inbreeding depression)
2) inbreeding complicates the genetic covariance structure of the population. In noninbred populations, the genetic covariance is a function of additive genetic and dominance variance. In inbred populations, additional terms need to accommodate: dominance variance in a completely inbred population, covariance between additive and dominance effects in completely inbred populations and the sum (over loci) of squared effects of complete inbreeding depression (De Boer and Hoeschele, 1993).

A conceptual problem with the infinitesimal model is that it can not accommodate properly inbreeding depression. Inbreeding depression is the result of loss of dominance due to loss of heterozygous loci. We could model dominance effects, but a finite amount of inbreeding depression (decrease of mean per percentage of inbreeding) can not be explained by dominance effects at an infinite number of loci.

De Boer and Hoeschele (1993) have derived rules for exact genetic covariance matrices, including dominance effects, and accounting for inbreeding and inbreeding depression. They noticed that this method is not feasible for larger populations, and compared it with

approximate BLUP, including dominance, but ignoring inbreeding. The inbreeding depression was accommodated as a linear regression of phenotype on inbreeding coefficient. There appeared to be small differences between approximate BLUP and exact BLUP using a genetic model with 64 unlinked loci, biallelic, and with complete dominance.

The infinitesimal model versus finite locus models

The infinitesimal model can be defined as the genetic model where the genetic variation is explained by the gene action at very many different loci, each having a small effect. The consequence of this model is that genetic change due to selection is due to such small genes in allele frequencies that they do not really contribute to a change of genetic variation. Furthermore, genotypes of offspring, conditional on their parents, are independent of their sibs, and the Mendelian sampling variance is constant, and the change in genetic variance due to selection (the Bulmer effect) can be predicted. The additive genetic effects, and possibly the dominance effects are normally distributed in the base population, and often it is assumed they are in further generations after sampling of geamte3es from selected parents.

Hill (1994) says:

"The infinitesimal model continues to dominate much of the theory, not because it can actually be true, (there are not an infinitely large number of unlinked loci in the genome), but because it is mathematically tractable and leads to some simple solutions. It is a formal requirement for most of the use of BLUP in livestock improvement".

It is reasonable to assume that each trait is regulated by a finite number of loci, and for many traits, QTL's have now been detected with a moderate to major effects. A more realsitic model might be a finite locus model with the genetic variance determined with a finite number of loci. The variance determined by each locus is not likely to be constant, and there will be siomne loci that explain more variance than other. The distribution of effects of genes has been suggested to follow a geometric series. . For example, lande and Thompson (1990) assumed that when the loci are ordered, the variance at the i^{th} locus is

$$\sigma_i^2 = a^{i-1}\sigma_a^2(1-a)$$

where a is a constant between o and 1, determining the 'flatness' of the distribution of effects. Lande and Thompson (1990) defined the effective number of loci as the number of unlinked loci of equal and complete additive effects as:

$$n_e = \frac{1+a}{1-a}$$

Questions are:

- How robust is the BLUP methodology against deviations of the infinitesimal model.

- Would more correct models improve in
  - efficiency of selection and estimation of genetic effects
  - prediction of selection response
  - prediction of long term effects of selection

Furthermore, as more information is accumulating about QTL effects, genetic evaluation models that accommodate such effects are needed. Models have been described to included QTL effects (Fernando & Grossman, 1989; Meuwissen and Goddard, 1997). However, the models do not allow multiple QTL's, and do not handle very well non-additive genetic effects. Particularly, epistasis are ignored in a single QTL model. Finite locus models might be a logical framework to fit in multiple QTL effects (Goddard, 1998).

One problem with finite locus models is that there is an almost infinite number of such models. One has to define or assume:

the number of loci contributing
the number of alleles
the initial allele frequency
the allelic effects, these effects can be due to additive gene actions,
and there can be interactions within loci (dominance)
and between loci (epistasis)

It is obvious that it will be very hard to estimate all these effects or their variance components in a finite loci model, and there may be little power to choose the best fitting model given for a given data set
One can assume a given finite locus model, and study its behavior in comparison with the infinitesimal model. However, results maybe hard to generalize, because there is such a large amount of alternative models that may behave differently from the finite locus model chosen.

Using BLUP under finite locus models

Mäki -Tanila and Kennedy (1986) compared the behavior of finite locus models under mixed model methodology. Differences between a realistic finite locus model and an infinitesimal model can be expected because:

- With few loci, the distribution of genotypic effects is not normal. This causes non-linear relationships between phenotype and additive genetic value.

- Gene frequencies can be expected to change after selection, and therefore, genetic variances will change, which is not accommodated in an infinitesimal model.

- With dominance, there is a non-linear relationship between additive value and genotypic value.

Non-linearity disappears usually quickly with more loci (n), proportionally to 1/n. Results from Mäki -Tanila and Kennedy (1986) indicated that

- With no selection, estimates of genetic means are essentially unbiased, even with a 2-locus model. With complete dominance, inclusion of the dominance effect is needed to avoid bias due to ignoring inbreeding depression.
- With selection, the response to selection is overestimated with BLUP when initial gene frequencies are high, and underestimated when they are low.

Such results were also found by De Boer and Van Arendonk (1992), who simulated a finite locus model contained either 64 or 1600 loci, unlinked and biallelic, and each of equal effect. Initial frequency of the favorable allele was 0.2, 0.5 or 0.8. Phenotypic selection was used. They selected for 5 generations and used BLUP (i.e., assuming the infinitesimal model) to estimate additive genetic effects. The actual variance under the finite locus model deviated from the expected variance from the base population, after correction for the Bulmer effect, genetic covariances and inbreeding. This deviation is due to gene frequency changes. Actual genetic variance was higher when the initial frequency of the positive allele was low, and lower when the initial frequency of the positive allele was high. Due to this reason, additive genetic effects were overestimated in Generation 5 when initial frequencies were high, and underestimated when they were low. Results are in

Table 1 for generation 5, for the case where only additive genetic effects were assumed. With a 1600 loci model, gene frequency changes were very small, expected variances were equal to observed variances, and EBV's were empirically unbiased.

Table 1 Mean, observed variance of simulated additive genetic effects, expected additive genetic variance based on the infinitesimal model, and mean difference between estimated and true additive genetic effects for generation (mean of 1000 simulations) (De Boer and van Arendonk, 1992).

P_i	Mean BV	σ_a^2	$E(\sigma_a^2)$	$\hat{a} - a$	p
0.2	9.35	18.35	15.00	-0.19	0.27
0.5	10.67	22.85	23.44	0.02	0.58
0.8	7.85	11.78	15.00	0.21	0.86

A gene based model

The problems with the infinitesimal model are

- No account for changes in genetic variance due to changes in gene frequency
- Difficult to derive proper genetic covariances in the case of non-additive genetic effects and inbreeding.

Goddard (1998) added the argument that

- current models should accommodate QTL effects properly, including possible non additive gene actions at QTL and accounting for selection.

Goddard (1998) proposed a 'gene based model' fitting a finite number of loci (4). The model was simply for animal i

$$y_i = \mu + \sum_j g_{ij} + e_i$$

whre j=1,..4) and g_{ij} could have values +a, d or –a. The model was compared with 'approximate BLUP' containing only additive genetic effects and a regression on inbreeding coefficient. Gibbs sampling was used to estimate parameters in the gene-model. A 'high line' and a 'low line' were simulated, based on phenotypic selection, and with a genetic model with 10 unlinked biallelic loci and full dominance. BLUP underestimated response in the highline, and overestimated it in the low line, due to not accounting for gene frequency changes. The BLUP analysis overestimated inbreeding depression. The gene model was able to estimate responses much better, and also accounted for non-additive genetic effects: the means of the crossbred groups were estimated more precisely (Table 2).

Table 2 Comparison of BLUP and gene based estimation (with Gibbs Sampling) of genetic values (from Goddard, 1998)

Generation	Line	Simulated Mean	BLUP Mean	Gene-Model Mean
4	H	1.95	2.18	1.93
4	L	-1.07	-0.93	-1.00
5	F1	1.51	1.24	1.45
6	F2	0.89	0.99	0.92

Stricker and Fernando (1998) summarized some properties of finite locus models and infinitesimal models, in the context of mixed inheritance models and segregation analysis:

- In the infinitesimal model, the conditional variance of the genotypic value of an offspring given the genotypic values of parents is a constant and does not depend on whether parents are selected. In finite locus models, the conditional variance of the progeny, given the parents, is lower when parents have extreme genotypes, because such parents are more likely to be homozygous at many loci.
- The genotypic distribution is not symmetric in a finite locus model if allelic frequencies deviate from 0.5. In the infinitesimal model, distributions are always symmetric.
- In the infinitesimal model, response to selection is constant, and change in the variance is zero after equilibrium is reached. In a finite locus model, allele frequencies change and the genetic mean and genetic variance change (not at a constant rate) until all alleles are fixed.

Fernando et al. (1994), have proposed a finite locus model and showed that it provided similar likelihood values than an exact likelihood for the mixed inheritance model for a simple pedigree. For larger and more complicated pedigrees, the exact likelihood can not

be calculated for the mixed inheritance model, and the finite locus model might be useful here.

In conclusion, finite locus models have been proposed and they have shown to work. However, it is very hard to generalize whether are of much better value than the infinitesimal model or the mixed inheritance model. Its superiority over BLUP depends on gene frequency changes, and their genetic action, and these are unknown for real data in livestock. However, given the need to model QTL effects, epistatic effects, and the increased amount of information becoming available at DNA level, will enhance increased use of such models and increased opportunity to test them on actual data.

References:

Caballero and Santiago (1998). Proc 6th World Congress on Genet. Applied to Livest. Production. Armidale. Vol 26:5.

Cockerham (1954). Genetics 39:859.

De Boer and Van Arendonk (1992). Theor. Appl. Genet. 84:451.

De Boer and Hoeschele (1993). Theor. Appl. Genet. 86:245.

Gowe.(1983). Proc. 32nd Annual. Poultry Breeders Roundtable. : 22.

Fernando, Stricker and Elston (1994). Theor. Appl. Genet. 88:573.

Goddard (1998). Proc 6th World Congress on Genet. Applied to Livest. Production. Armidale. Vol 26:33.

Hill (1994) Proc 5th World Congr. on Genet. Appl. to Livest. Prod. Guelph: Vol 19:125.

Kimura (1957). Ann. Math. Statst. 28:882.

Lande and Thompson (1990). Genetics 124:743.

Maki-Tanilla and Kennedy (1986) Proc 4th World Congress on Genet. Applied to Livest. Production. Nebraska. Vol 12:443.

Stricker and Fernando (1998)Proc 6th World Congress on Genet. Applied to Livest. Production. Armidale. Vol 26:25