Genomic Selection in New Zealand Dual Purpose Sheep

K.G Dodds[1], B Auvray[1], M. Lee[1], S.-A.N. Newman[1] and J.C. McEwan[1].
[1]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand

ABSTRACT: Genomic selection allows prediction of genetic merit and selection of superior animals for breeding, on the basis of data from many genetic markers. Here we describe the application of this methodology in the New Zealand dual purpose sheep industry using a mixed breed (Romney, Coopworth and Perendale) training set. The gBLUP method was used, and accounting for breed structure in the genomic relationship matrix was found to be important for calculating model-based accuracies, but the predictions were similar with and without this adjustment. Realized accuracies were lower than model-based accuracies, particularly for early life traits, reflecting a selection effect in the animals chosen for genotyping. The results of this research are available commercially in New Zealand for the breed types used in the training set and for composites that include these breeds.
Keywords: breeding values; genetic markers

Introduction

The New Zealand (NZ) sheep industry is dominated by dual purpose (meat and wool) breeds and crosses. Within this sector, most sheep are predominantly Romney or one of its derivatives - Coopworth or Perendale (Dodds et al. (2013)). In recent years the Texel breed has become more prevalent. Terminal sire breed rams are commonly used in commercial flocks, but these breeds constitute less than 20% of the breeding. Even within the breeder's flocks, crossbred animals are becoming common. Both dual-purpose and terminal sire breeders record their flocks with Sheep Improvement Limited (SIL).

Genomic selection (GS; Meuwissen et al. (2001)) is a methodology developed to allow prediction of genetic merit, and subsequently selection of parents, from large numbers of genetic markers. Genomic selection has been the focus of much research (Habier (2010)) and has been quickly taken up in a number of industries (e.g. in dairy cattle; Hayes et al. (2009a)). The availability of the Illumina OvineSNP50 Beadchip (http://www.illumina.com; '50k chip') has allowed GS within sheep breeding programmes. Here we investigate if and how this resource can be used to predict genetic merit in the NZ dual purpose sheep industry.

Materials and Methods

Animals. A set of 13,364 animals were genotyped with the 50k chip where there was permission to use the associated SIL flock data. The animals were primarily sires from breeder's flocks, but also included other animals of both sexes, as well as some animals from research flocks. The breed representation in this set was 47% Romney (40% purebred Romneys), 21% Coopworth (8% purebred), 7% Texel (1% purebred), 7% Perendale (6% purebreed), 6% Primera (5% purebreed) and other breeds with less than 3% each.

Animals were assigned to a number of breed groups. Those that were recorded as at least 75% Romney, Coopworth, Perendale or Texel formed the groups R, C, P and T, respectively. Remaining animals that were recorded with more than 30% of Romney, Coopworth and Perendale combined formed the composite group (Comp) and then those still not assigned a breed group were discarded from the analyses. Recently developed 'breeds' were decomposed into their foundation breeds, as far as possible, before applying these thresholds.

Estimated breeding values (eBVs) were obtained using standard SIL procedures, except that pedigrees were cleaned to avoid genotype inconsistencies. The eBV analyses were based on a set of 756 flocks with over 4 million animals, born between 1990 and 2011.

Genotypes. There were 53,903 single nucleotide polymorphism (SNP) markers with results reported from the 50k chip. Results were obtained in Illumina's 'AB' format. SNPs that were discarded by the ovine HapMap project (http://www.sheephapmap.org), or that were denoted or appeared to be non-autosomal were discarded. SNPs with a call rate <97% or with a 10th percentile quality score (GC10) reported by the Illumina software < 0.422 or that were monomorphic were also discarded. Data were also checked for extreme departures from Hardy-Weinberg, but none discarded. This process resulted in 47,071 SNPs being available for GS.

Samples with a call rate (using all 53,903 SNPs) <0.96 were discarded. Duplicates were checked for consistency, and samples checked for consistency with recorded parentage, gender and breed. Samples were removed if inconsistencies other than pedigree could not be corrected, leaving 13,338 animals.

Genotypes were then coded as the number of A alleles (0, 1 or 2). After the data edits, missing genotypes were minimal (<0.1%) and were replaced by twice the weighted (by the animal's breed composition) average breed estimated A allele frequency.

Statistical analysis. The data used in the analyses were based on eBVs for the following traits: weaning weight direct (WWT), weaning weight maternal (WWTM),

live weight at 8 months (LW8), greasy fleece weight at 12 months (FW12), log faecal egg count after 6-8 weeks following an initial drench (FEC1) and number of lambs born (NLB, i.e. litter size). Parental contributions were removed from the eBVs (so that only measurements on the animal and its descendants would contribute to its value), and the resulting values were deregressed by dividing by the reliability (squared correlation between true and estimated breeding value) of those values to give the 'deregressed, parent-average removed breeding values' (dpBV). These steps followed the procedures described by Garrick et al. (2009).

For a particular trait, an animal was included in the analysis only if its dpBV had a reliability of at least 0.8 times the trait heritability (approximately equivalent to having at least one trait record on itself). Breed validation sets were created by choosing the youngest animals of a breed using a birth year cut-off for each breed, to mimic the intended application to predicting merit in young animals. The cut-off for each breed was chosen so that there were at least 200 animals in a breed validation set, but using no more than 75% of the full breed resource. The remaining animals were combined in to a single mixed-breed training set. When applying these rules to the Comp group, the animals were first restricted to those that had at least 50% of Romney, Coopworth and Perendale combined and less than 25% Texel. The other Comp animals were then added to the Comp validation set (none were used in training). The Texel group was used for validation only. Numbers for each set and trait are shown in Table 1.

Table 1. Number of animals in training and validation sets for Romney (R), Coopworth (C), Perendale (P), Composite (Comp) and Texel (T) breed groups for weaning weight (WWT), WWT maternal (WWTM), live weight at 8 months (LW8), greasy fleece weight at 12 months (FW12), log faecal egg count (FEC1) and number of lambs born (NLB)

Trait	Set[§]	Breed				
		R	C	P	Comp	T
WWT	T	4643	1840	564	880	0
WWT	V	688	250	202	2277	404
WWTM	T	2308	1019	306	132	0
WWTM	V	257	287	257	848	288
LW8	T	4510	1751	417	801	0
LW8	V	649	247	293	2113	403
FW12	T	4287	1603	434	634	0
FW12	V	549	228	275	924	240
NLB	T	2317	1188	307	373	0
NLB	V	420	253	328	906	306
FEC1	T	1720	1293	248	347	0
FEC1	V	216	304	254	664	140

[§]Training (T) or validation (V)

Prediction of breeding values (giving 'molecular breeding values'; mBVs) from genotype data was undertak-

en using the 'gBLUP' method (Goddard et al. (2010)) whereby the pedigree-based numerator relationship matrix is replaced by a genomic relationship matrix (GRM), as follows. Two different versions of GRM were investigated, one ($\mathbf{G}_a$) which centers and scales using allele frequencies calculated from the whole data set (vanRaden (2008)), and another ($\mathbf{G}_b$) which centers and scales based on breed-specific allele frequencies (Harris and Johnson (2010); Auvray and Dodds (2013)), based on the proportion of R, C, P, T or 'Other'. The models for each trait included the first v principal components (PCs) of $\mathbf{G}_a$ for training set animals as covariates and a random animal effect with the covariance between animals set to $\mathbf{G}\sigma_a^2$ where $\mathbf{G}$ is a GRM. Unless specified otherwise, $v = 6$. The records in the analysis were weighted by $(1-r^2)/r^2$ where r^2 is the reliability of the dpEBV. Further details can be found in Auvray et al. (2014). The genetic variance, σ_a^2, was fixed so that the heritability (of an individual measurement) matched that used in the SIL system.

Two measures of accuracy of the mBV were calculated. The 'realized accuracy', r_A, was estimated as the correlation, in the validation set, between dpEBV and mBV divided by the square root of the 'effective heritability' (mean reliability of the dpEBV) of the observations in the gBLUP model (Lorenz et al. (2011)). The lower bound of a one-sided 95% confidence interval was also calculated. The 'model-based accuracy', r_B, was calculated as the mean individual accuracy in the validation set, where the individual accuracies are calculated from the prediction error variance from the gBLUP model (Mrode (2005); Auvray et al. (2014)).

The spread of mBV in the validation sets was compared to what might be expected for a set of estimated breeding values with mean accuracy r'. The ratio of the expected spread to that observed was measured as

$$k = r'\sigma_A / \text{sd(mBV)}$$

where σ_A is the genetic variance of the trait.

Results and Discussion

Population Structure. The variation explained by the first 1, 2, 6, 50 and 200 PCs was 54%, 69%, 82%, 91% and 94%, respectively. Figure 1 shows that the different breeds are largely separated on the basis of the first two PCs. The Coopworth breed appears more spread than other breeds which is likely to be a reflection of its breed society allowing introgression from other breeds (Coopworth Genetics (2014)). Although the Perendales appear within the spread of Romneys, they appear as a distinct group in other dimensions (e.g. components 4 and 5). Some breed substructure is also evident from higher order (>3) components, especially within the Romney breed.
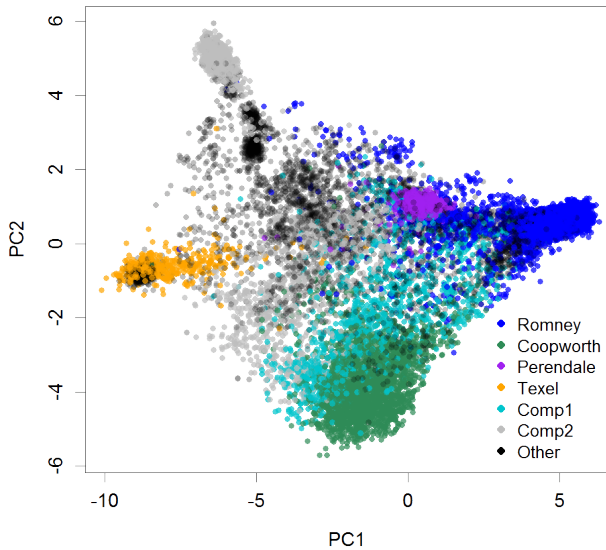
Figure 1: The first two principal components (PC1 and PC2) of $\mathbf{G_a}$, colored by breeds as defined for the genomic selection analysis. The composite group is separated into those considered for training (Comp1; ≥ 50% R+C+P, <25% T) and those only used for validation (Comp2). Other represents breed types not used for genomic selection.

Realized accuracies. Table 2 shows the r_A in the various validation sets when using GRM $\mathbf{G_a}$ and with $v= 6$. For comparison, the theoretical EBV accuracy of the parent average (PA) with a single trait record on each parent is also shown. In general, R, C and Comp r_A, are similar to the PA accuracies for liveweight traits (WWT, WWTM and LW8) and higher for the other traits. These three breed groups have the highest r_A followed by P and then T. This roughly represents the contribution of these breeds to the training sets, with low numbers of P and with T being excluded from training. The results suggest that there is some predictive power across breeds as the r_A for T were significantly positive for WWT, WWTM, LW8 and NLB. This is likely to be assisted by there being some T contribution to the training set (e.g. ≈ 1% for WWT) and some R, C and P contribution in the T validation set ((e.g. ≈ 8% for WWT). However, the results are worse than for the other breeds, and in one case the correlation between mBV and dpEBV was actually negative (for FW12), so the mBV for T do not appear to be very useful. Harris et al. (2009) and Kachman et al. (2013) also found limited prediction accuracy in breeds not in the training set, when using a bovine 50k SNP chip in dairy and beef cattle, respectively.

There was very little difference between r_A calculated using $\mathbf{G_a}$ or $\mathbf{G_b}$ (Figure 2). This was also true for values of r_A calculated using all the validation sets, except for T, combined. There were some slight differences for the most poorly predicted combinations (FW12 and FEC1 for T). Auvray et al. (2014) also found small differences in r_A using a variety of GRM. These authors also found that single breed training sets did not give any overall advantage over mixed breed training, using the same resource described here.
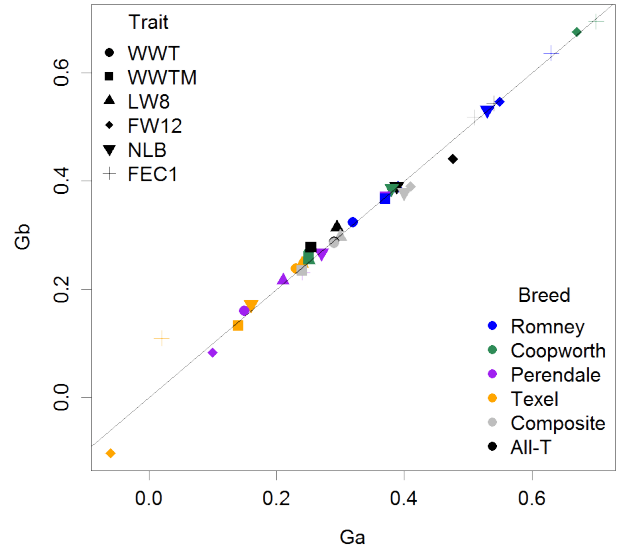

Figure 2. Comparison of realized accuracies using two different GRM ($\mathbf{G_a}$ and $\mathbf{G_b}$) for different breed sets and across all validation breed sets excluding Texel (All-T). The line of equality is also shown.

Model-based accuracies. Figure 3 shows the r_B in the various validation sets when using GRM $\mathbf{G_a}$ or $\mathbf{G_b}$. For the main breeds studied (R, C, P and their composites), the two GRM give quite similar r_B, with slightly higher values using $\mathbf{G_b}$. In contrast, the values are quite different for T, the breed with almost no contribution to the training sets. In this case the method which accounts for breed structure ($\mathbf{G_b}$) gives lower values than when ignoring breed structure ($\mathbf{G_a}$). Given that the T validation set is not closely related to the training set, it would appear that $\mathbf{G_b}$ is giving more realistic results in this case.
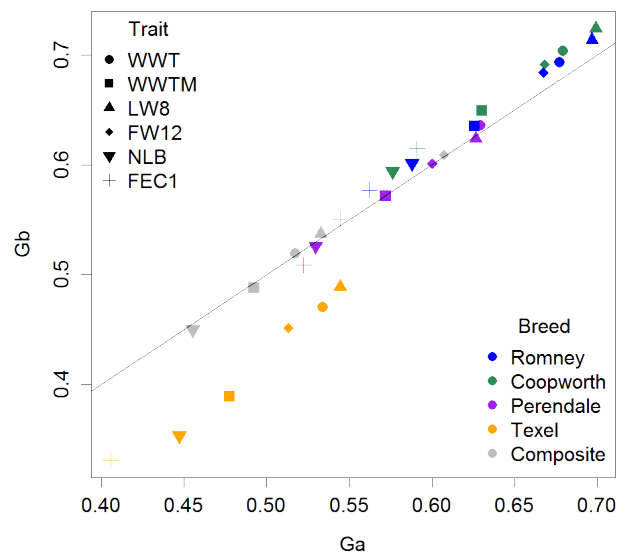

Figure 3. Comparison of model-based accuracies using two different GRM ($\mathbf{G_a}$ and $\mathbf{G_b}$) for different breed sets. The line of equality is also shown.

Realized accuracies and model-based accuracies are compared in Figure 4. For this comparison we show r_B

with $\mathbf{G_b}$, the GRM that seemed to give the best results, while for r_A there is little difference between the GRM so results are shown for the simpler method, $\mathbf{G_a}$. In general, the r_B are somewhat higher than the r_A. A similar effect was observed by Hayes et al. (2009b) in multi-breed populations (without accounting for breed structure in the GRM) but was not evident with single breed populations of dairy cattle. As we have seen, failing to account for breed structure can result in over-estimation of r_B, and there may be further population structure that we have not catered for with $\mathbf{G_b}$, but given the relatively minor differences in r_B between the two GRM for the main breeds, it is unlikely there would be much further change by defining breed subgroups for the calculation of $\mathbf{G_b}$. A characteristic of the GRM used here is that their elements do not have a direct identity by descent interpretation as is the case for pedigree-based numerator relationship matrix. In particular, diagonal elements may be less than one, and off diagonals can be negative. It could be that the model-based accuracy is being increased by negative relationships, which is counter-intuitive. To investigate this, a modified $\mathbf{G_b}$ was created where negative elements were set to zero ($\mathbf{G_{bz}}$). The r_B from this GRM were indeed lower than those from $\mathbf{G_b}$ (by $\approx$ 5% for higher accuracy combinations to $\approx$ 25% for lower accuracy results).
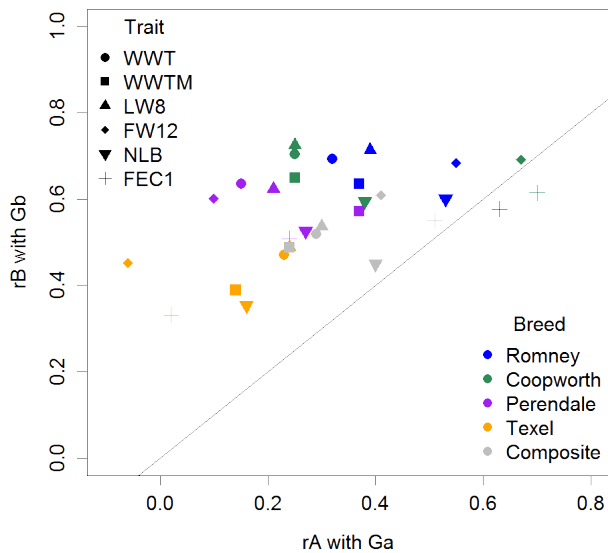

Figure 4. Comparison of model-based accuracies using $\mathbf{G_b}$ with realized accuracies using $\mathbf{G_a}$ for different breed sets. The line of equality is also shown.

Another possibility is that the r_A could be lower than the population values they are meant to represent. Deficiencies in the GRM should affect traits with similar heritability in a similar way, but e.g., WWT appears to be considerably overestimated by r_B while FEC1 does not, even though these have similar heritability (see PA values in Table 2). The traits with the greatest difference between r_A and r_B are those that are measured early in life (WWT) or are correlated with such traits (LW8, FW12). The animals used in training and validation are mostly already sires and therefore represent a selected subset, particularly for traits with reasonably accurate EBV at the time of selection, which are the ones showing the greatest discrepancy. A selection effect will lower the realized accuracy (Edel et al.

(2012)) which is consistent with these results. Further evidence that these are selection effects is that the differences between r_A and r_B are generally small for several other traits not presented here, such as disease traits (Pickering (2013); Phua et al. (2014)) or carcass composition traits.

Table 2. Realized molecular breeding value accuracies for Romney (R), Coopworth (C), Perendale (P), Composite (Comp) and Texel (T) breed groups for weaning weight (WWT), WWT maternal (WWTM), live weight at 8 months (LW8), greasy fleece weight at 12 months (FW12), log faecal egg count (FEC1) and number of lambs born (NLB)

Trait	PA[†]	Breed				
		R	C	P	Comp	T
WWT	0.30	0.32	0.25	0.15	0.29	0.23
WWTM	0.24	0.37	0.25	0.37	0.24	0.14
LW8	0.45	0.39	0.25	0.21	0.30	0.24
FW12	0.42	0.55	0.67	0.10	0.41	-0.06
NLB	0.22	0.53	0.38	0.27	0.40	0.16
FEC1	0.28	0.63	0.70	0.24	0.51	0.02

[†]The EBV accuracy that would be obtained from parent average with a single trait record on each parent.

Effect of principal component covariates. Analyses with differing numbers of PCs, namely v = 0, 6, 50, and 200, were undertaken with a subset of traits to study the effect of including these covariates when using GRM $\mathbf{G_a}$. In addition to calculating r_A from the predictions after adjusting for the PCs as has been done for the other values of r_A in this article, mBVs were also computed adding back in the values predicted by the PCs. To assist with these calculations, PCs were calculated using the full set of training and validation animals, rather than just the training animals as previously.

Figure 5 shows the accuracies (without adding the effects of PCs) for WWT, NLB and FEC1. On average, the r_A drop by 0.02 between v = 0 and v = 6, which is minor compared to differences observed between r_A and r_B. This is despite these 6 PCs explaining a large proportion of the variation (82%). Increasing v further tends to cause further decreases in accuracy, despite these components explaining a low proportion of variation (12% for the next 194 PCs). A similar effect was found in a mixed-breed sheep resource by Daetwyler et al. (2012) who fitted each value of v between 0 and 200. They found that the accuracies within this range plateaued between v = 50 and v = 150, depending on the trait and prediction breed. This plateau was interpreted as the contribution to prediction accuracy due to linkage disequilibrium, in contrast to that from close relationships.
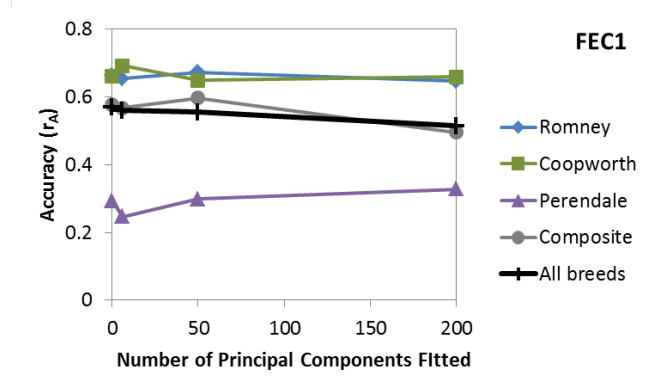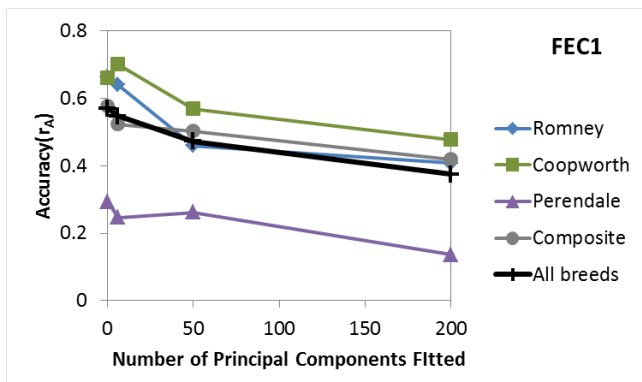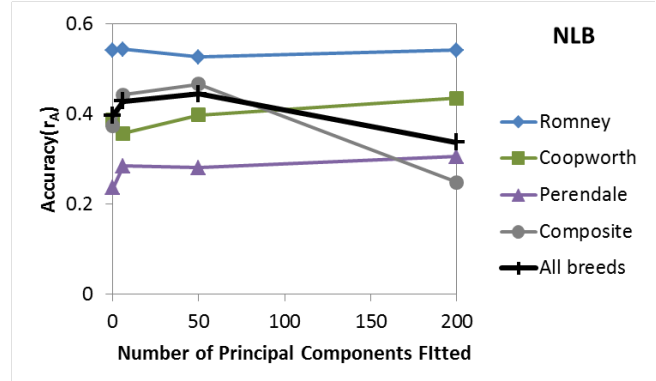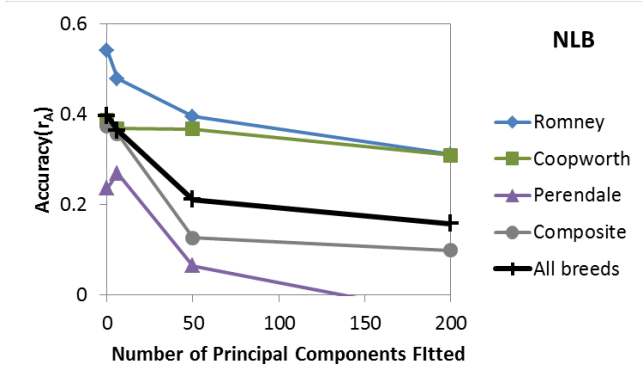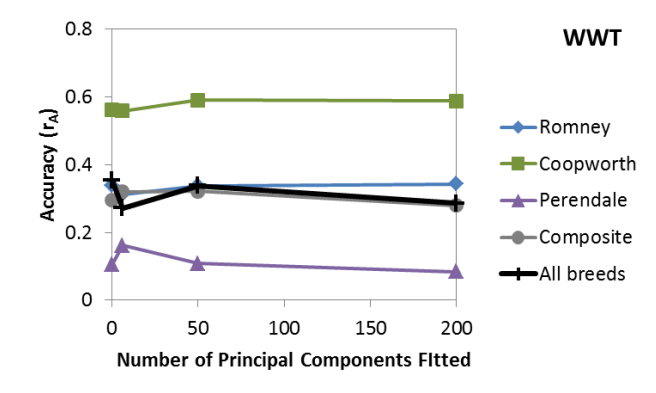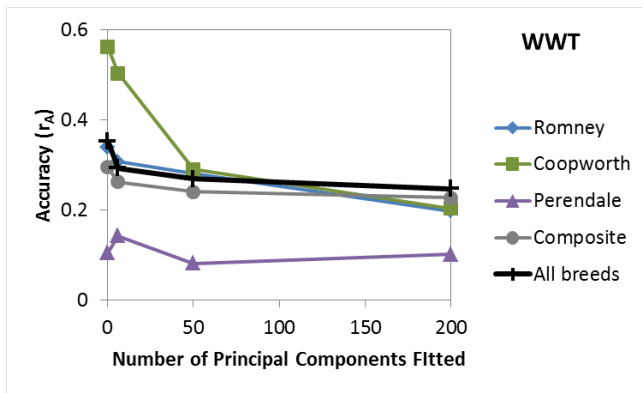
Figure 5. Effect of fitting different numbers of principal components of G_a on the realized accuracy of weaning weight (WWT), number of lambs born (NLB) and log faecal egg count (FEC1).

The trend in the combined R, C, P and Comp validation sets (shown as 'All breeds' in Figure 5) is similar to the within breed trends, so these trends are not an artefact of calculating the accuracies within breed. We have chosen to retain $v = 6$ to safeguard against simply predicting differences associated with population substructure (e.g. breed differences) and because there appears to be only a minor sacrifice in accuracy. For these analyses, the PC analysis used data from training plus validation, rather than just training animals, as above. In most cases this had little effect, with accuracies differing by less than 0.01 for 8 of the 15 comparable ($v = 6$) cases. However, using training plus validation improved the accuracy of WWT mBVs in C by 0.26. It is not clear why this has occurred for this one combination, but appears to be related to a subset of flocks.

Figure 6. Effect of fitting different numbers of principal components of G_a on the realized accuracy of weaning weight (WWT), number of lambs born (NLB) and log faecal egg count (FEC1) if principal component effects are added to the predictions.

Adding the effects of PCs back into these analyses gives accuracies that are largely independent of v (Figure 6) with the main exception being for NLB in composites, where the accuracies still drop with increasing v. These results suggest that adding back PC effects does not have any advantage over fitting zero or a few PCs.

The spread of mBV. The values of k (ratio of the expected spread in mBV to that observed) are shown in Table 3. In dairy cattle, it is common to assess 'genomic inflation' by looking at the slope of EBV based on a progeny test with the mBV. This does not account for changes in spread of these variables due to the accuracy of their selection, however in dairy cattle applications, both accuracies are usually high and so this is not an issue. The k values given here are designed as a measure of genomic inflation which can be used more generally. The values are all less than one (Table 3), indicating that the mBV are more

spread than expected. The k values vary considerably between trait-breed combinations.

Table 3. The ratio, k, of expected (assuming accuracies r_A) spread to observed spread of molecular breeding values (adjusted for PCs), calculated using G_a, for Romney (R), Coopworth (C), Perendale (P), and Composite (Comp) breed groups for weaning weight (WWT), WWT maternal (WWTM), live weight at 8 months (LW8), greasy fleece weight at 12 months (FW12), log faecal egg count (FEC1) and number of lambs born (NLB)

Trait	Breed			
	R	C	P	Comp
WWT	0.43	0.32	0.23	0.47
WWTM	0.56	0.45	0.66	0.45
LW8	0.67	0.45	0.45	0.61
FW12	0.81	0.77	0.16	0.64
NLB	0.64	0.54	0.48	0.61
FEC1	0.80	0.90	0.51	0.86

Commercial application. The results of this research and similar calculations from other industry EBV have been commercialized via a third party for use in the NZ industry. Initially mBV were calculated by the third party, and IP constraints meant that they required prediction equations, rather than direct calculation from the research data set. G_a is the only GRM that allows such direct calculation, via the use of 'SNP coefficients' in a regression prediction equation (vanRaden (2008)). Individual model-based accuracies could not directly be calculated under this scenario, so breed-trait combinations of accuracies were given. Because there appear to be deficiencies in both r_A and r_B, the average of these two methods was given (r_A from using G_a as in the predictions, and r_B from using G_{bz} as it appeared to be more conservative). The traits which are mostly penalized by this approach are those that showed a selection effect for r_A, but these are the traits which will benefit the least from GS, as they have phenotypic information early in life. The k values were also recalculated based on the above average accuracy, and these values used to rescale the mBV so they had the expected spread.

In general, mBV are not reported if the r_A for that particular breed-trait combination was not significantly positive. In some cases r_A changed from significant to non-significant in an updated analysis. These fell just below the significance cut-off and may reflect changes in validation sets rather than true loss of predictive ability, therefore these combinations continued to be reported allowing continuity of product. This was the case for WWT and FW12 in P among results reported here.

Genomic calculations have now moved to within the national evaluation system, SIL, which allows more flexibility in their calculation, including the calculation of individual accuracies. Results (scaled by k as above) are blended using the method of Harris and Johnson (2010) with EBV from an across flock evaluation encompassing all client flocks as well as other flocks to provide better connectedness. Young animals are evaluated using a lower density (5000-6000 SNPs) chip (at a lower price than the 50K chip) and genotypes imputed to the 50K prediction set. An ad-hoc small adjustment is made to the individual accuracies calculated to compensate for imputation.

Conclusion

Genomic selection in the NZ dual purpose sheep industry is possible using a mixed breed training set, allowing predictions in purebreds and composites. It is important to account for breed structure in the genomic relationship matrix when calculating model-based accuracies, but the predictions were similar with and without this adjustment. Realized accuracies may be biased downwards due to selection of genotyped animals. The results of this research are available commercially in NZ.

Literature Cited

Auvray B. and Dodds K.G. (2013) Proc. Assoc. Adv. Ani. Breed. Genet. 20:257-260

Auvray B., McEwan J.C., Newman S.-A.N. et al. (2014) J. Anim. Sci. (submitted)

Coopworth Genetics (2014) http://www.coopworthgenetics.co.nz/page.php?page=Breed+Characteristics Accessed March 2014

Daetwyler H.D., Kemper K.E., van der Werf J.H.J. and Hayes B.J. (2012) J. Anim. Sci. 90:3375-3384

Dodds K.G., Newman S.-A.N., Auvray B. et al. (2013) Proc. Assoc. Adv. Ani. Breed. Genet. 20:274-277

Edel C., Neuner S., Emmerling R. and Götz K.U. (2012) Interbull Bulletin 46:16-19

Garrick D., Taylor J. and Fernando R. (2009) Genet. Sel. Evol. 41:55

Goddard M.E., Hayes B.J. and Meuwissen T.H.E. (2010) Genet. Res. 92:413-421

Habier D. (2010) J. Anim. Breed. Genet. 127:336-337

Harris B., Johnson D. and Spelman R. (2008) ICAR Tech. Series 13:325-330

Harris B.L. and Johnson D.L. (2010) J. Dairy Sci. 93:1243-1252

Hayes B.J., Bowman P.J., Chamberlain A.J. et al. (2009a) J. Dairy Sci. 92:433-443

Hayes B.J., Bowman P.J., Chamberlain A.C. et al. (2009b) Genet. Sel. Evol. 41:51

Kachman S., Spangler M., Bennett G. et al. (2013) Genet. Sel. Evol. 45:30

Lorenz A.J., Chao S., Asoro F.G. et al. (2011) Adv. Agronomy 110:77-123

Meuwissen T.H.E., Hayes B.J. and Goddard M.E. (2001) Genetics 157:1819-1829

Mrode R.A. (2005) Linear models for the prediction of animal breeding values. 2nd ed. CAB International, Wallingford, U.K.

Phua S.H., Hyndman D.L., Baird H.J. et al. (2014) Anim. Genet. (in press)

Pickering N.K. (2013) Ph.D. thesis, Massey University, Palmerston North, New Zealand

VanRaden P.M. (2008) J. Dairy Sci. 91:4414-4423