## BREEDING AND GENETICS: GENETIC AND GENOMIC METHODS

**0163 Evaluation of predictive ability of Cholesky factorization of genetic relationship matrix for additive and non-additive genetic effect using Bayesian regularized neural network.** H. Okut*[1], D. Gianola[2], K. A. Weigel[2], and G. J. M. Rosa[2], *[1]University of Yuzuncu Yil, Van, Turkey, [2]University of Wisconsin, Madison.*

This study aimed to explore the effects of additive and non-additive genetic effects on the prediction from using Bayesian regularized artificial neural network (BRANN). The data sets were simulated for two hypothetical pedigrees with five different fractions of total genetic variance accounted by additive ($\sigma^2_a/\sigma^2_G$ ), additive x additive ($\sigma^2_{aa}/\sigma^2_G$ ) and additive x additive x additive ($\sigma^2_{aaa}/\sigma^2_G$ ) genetic effects. A feed forward artificial neural network (ANN) with Bayesian regularization (BR) was used to assess the performance and predictive ability of different nonlinear ANNs and linear models for genetic architectures. Effective number of parameters ($\gamma$) and sum of squares error (SSE) in test data sets were used to evaluate the performance of ANNs. Distribution of weights ($w_{ij}$) and correlation between observed and predicted values in the test data set were used to evaluate the predictive ability. There were clear and significant improvements in terms of the predictive ability of linear (equivalent Bayesian ridge regression) and nonlinear models when the proportion of additive genetic variance in total genetic variance ($\sigma^2_a/\sigma^2_G$ ) increased. On the other hand, nonlinear models outperformed the linear models at each genetic architecture. The weights for the linear models were larger and more variable than for the nonlinear network, where distributions were leptokurtic, indicating strong shrinkage towards 0. In conclusion, our results showed that: a) inclusion of non-additive effects did not improved the prediction ability compared to purely additive models, b) The predictive ability of BRANN architectures with nonlinear activation function were substantially larger than the linear models for the scenarios.

**Key Words:** artificial neural networks, Bayesian regularization, additive and non-additive genetic effects

**0164 Using recursion to compute the inverse of the genomic relationship matrix.** I. Misztal*[1], A. Legarra[2], and I. Aguilar[3], *[1]University of Georgia, Athens, [2]INRA, Castanet-Tolosan, France, [3]INIA, Las Brujas, Uruguay.*

A traditional algorithm to invert the numerator relationship matrix is based on the observation that the conditional expectation for an additive effect of one animal given the effects of all other animals depends on the effects of its sire and dam only, each with a coefficient of 0.5. With genomic relationships, such an expectation depends on all other genotyped animals, and the coefficients do not have any set value. For each animal, the coefficients plus the conditional variance can be called a genomic recursion. If such recursions are known, the mixed model equations can be solved without explicitly creating the inverse of the genomic relationship matrix. Several algorithms were developed to create genomic recursions. In an algorithm with sequential updates, genomic recursions are created animal by animal. That algorithm can also be used to update a known inverse of a genomic relationship matrix for additional genotypes. In an algorithm with forward updates, a newly computed recursion is immediately applied to update recursions for remaining animals. The computing costs for both algorithms depend on the sparsity pattern of the genomic recursions. An algorithm for proven and young animals assumes that the genomic recursions for young animals contain coefficients only for proven animals. Such an algorithm generates exact genomic EBV in GBLUP and is an approximation in single-step GBLUP. That algorithm has a cubic cost for the number of proven animals and a linear cost for the number of young animals. All algorithms were evaluated with a simulated data set of 1500 genotypes and ssGBLUP. In the algorithm with sequential updates, setting very small elements in recursions to zero resulted in little sparsity. Setting larger elements to zero caused large errors in $G^{-1}$ due to accumulation of errors. However, this algorithm worked very well for inv($A_{22}$), especially when the pedigree depth was limited. When complete recursions were computed and small elements were set to 0, the accuracy of GEBVs was almost unaffected but the sparsity level was moderate. The sparsity level increased to > 60% when G was blended with 20% of $A_{22}$. In all computations involving the algorithm for proven and young animals, the correlations of GEBV with those using the regular algorithm were > 0.99. The genomic recursions can provide new insight into genomic evaluation and possibly reduce costs of genetic predictions with extremely large numbers of genotypes.

**Key Words:** genomic selection, single-step GBLUP, efficiency

**0165 Advantage of supernodal methods in restricted maximum likelihood when dense matrices are involved in a coefficient matrix of mixed model equations.** Y. Masuda*[1,2], S. Tsuruta[2], and I. Misztal[2], *[1]Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Japan, [2]University of Georgia, Athens.*

The objective of this study was to determine speed-up of an average-information (AI) REML algorithm with a supernodal sparse-matrix package. Comparisons included twenty-three models with data sets from broiler, swine, beef and dairy cattle. Models included single-trait, multiple-trait, maternal, and random regression models with phenotypic data; selected models used genomic information as a genomic relationship matrix in single-step GBLUP. The AIREMLF90 program was used to compare two sparse-matrix packages: FSPAK and YAMS; the latter package used supernodal methods for faster computing when sparse matrices contain large dense blocks. The program was compiled with the Intel Fortran Compiler 13.1 using the Intel Math Kernel Library and ran on a computer with 16-core CPUs. Computations with YAMS were on average over 10 times faster than with FSPAK and had greater advantages for large data and more complicated models including multiple traits and random regressions and with genomic effects. The highest speed-up with YAMS over FSPAK was over 20 times faster in AI REML iteration and over 80 times faster in sparse inversion. In a model with 213,297 pedigreed and 15,723 genotyped animals, a single-trait analysis with FSPAK took about 5 h and multiple-trait analyses did not converge in 1 d. With YAMS, a single-trait analysis took about 20 min and a 4-trait analysis took about 5 h. Supernodal methods dramatically improve the computing cost if the AI REML for larger and more complex analyses, especially when genomic information is included in the single-step GBLUP models.

**Key Words:** AIREML, supernodal methods, sparse-matrix package

**0166 Use of genomic recursions and APY algorithm for single-step GBLUP analyses with large number of genotypes.**

B. D. Fragomeni*[1], I. Misztal[1], D. Lourenco[1], S. Tsuruta[1], and Y. Masuda[1,2], [1]*University of Georgia, Athens,* [2]*Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Japan.*

The purpose of this study was to examine accuracy of genomic selection in single-step genomic BLUP (ssGBLUP) when the inverse of the genomic relationship matrix (G) is derived by the algorithm for proven and young animals (APY). This algorithm implements the inversion of G by genomic recursions, with recursions for young animals involving only the proven animals. With efficient implementation, the algorithm has a cubic cost for proven animals but only a linear cost for young animals. Simulated data set included 142k phenotypes in 6 generations under selection for EBV, with 170k animals in the relationship matrix. Genomic data consisted of 20k animals genotyped for 45k SNP; the simulated genomic data mimicked the bovine genome. The proven animals were 10k genotyped parents selected from the first 5 generations, and the young 10k genotyped animals were selected from the last generation. For animals treated as young, 5k had a single record and 5k had no records. Comparisons involved GEBVs obtained by ssGBLUP evaluation with either the exact G (G-REG) and the G inverted by APY algorithm (G-APY). The correlations between GEBV with the G-REG and G-APY were 0.97 overall, 0.94 for animals treated as young without records, and 0.98 for animals treated as young with records. The true accuracies for the animals with records with G-REG and G-APY were 0.57 and 0.58, respectively; for the animals without records, the accuracies for REG and APY were both 0.43. When the status of the young and proven animals was switched, the accuracies remained identical. A separate analysis involved a national data set for final score in Holsteins. Out of 74,980 genotypes for bulls, 29,552 for bulls with daughters were treated as proven and 45,428 without daughters were treated as young. The correlations of GEBV obtained with the REG and APY algorithms were > 0.99 for both groups of bulls. When the number of high-accuracy animals with genotypes is limited ( < 100k), the APY algorithm may drastically reduce the cost of the ssGBLUP evaluation without affecting the accuracy. The APY algorithm may allow using all the available genotypes in one ssGBLUP analysis to reduce biases due to preselection of young animals.

**Key Words:** single step method, genomic selection, genetic evaluation

**0167 Genomic prediction accounting for residual heteroskedasticity.**

Z. Ou*[1], R. J. Tempelman[2], J. P. Steibel[2], C. W. Ernst[2], R. O. Bates[2], and N. M. Bello[1], [1]*Kansas State University, Manhattan,* [2]*Michigan State University, East Lansing.*

Classical genomic selection (GS) models that use single-nucleotide polymorphism (SNP) marker information to predict genetic merit of animals and plants usually assume homogeneous residual variance. However, this assumption seems questionable as environmental variability can be heterogeneous and it may affect the genetic control of a given quantitative trait. This study extends classical GS models, namely RR-GBLUP, BayesA, BayesB and BayesC, to explicitly account for residual heteroskedasticity using a hierarchical Bayesian mixed-models framework implemented with Markov Chain Monte Carlo methods. Competing GS models assuming homogeneous or heterogeneous residual variances were fitted to training data under simulation scenarios reflecting a gradient of increasing residual heteroskedasticity. Model fit of competing homoskedastic and heteroskedastic GS models was compared using prediction accuracy of genomic breeding values and pseudo-Bayes factors, both computed on a validation data subset one generation removed from the training dataset. Competing models were also fitted to two quantitative traits selected from a Michigan State University swine resource population dataset, namely carcass temperature and loin muscle pH 45 min after slaughter. These traits had been pre-screened for homoskedasticity and heteroskedasticity, respectively. Using a fivefold cross-validation approach, competing GS models were compared based on predictive ability of phenotypes. Overall, under the conditions considered in this study, heteroskedastic GS models showed improved model fit and enhanced prediction accuracy compared to homoskedastic GS models under conditions of extreme residual variance

heterogeneity; however, the magnitude of the improvement was too small (approximately 1% to 2% net gain in prediction accuracy) to confer practical relevance.

**Key Words:** genomic selection model, heterogeneous residual variance, genomic breeding values.

### 0168 Are past generations contributing to evaluations on young genotyped animals?

D. Lourenco*[1], I. Misztal[1], S. Tsuruta[1], I. Aguilar[2], T. J. Lawlor[3], S. Forni[4], and J. I. Weller[5], *[1]University of Georgia, Athens, [2]INIA, Las Brujas, Uruguay, [3]Holstein Association USA Inc., Brattleboro, VT, [4]Genus Plc, Hendersonville, TN, [5]ARO, The Volcani Center, Bet Dagan, Israel.*

Datasets of US and Israeli Holsteins and pigs from PIC were used to evaluate the impact of different number of generations on ability to predict GEBV of young genotyped animals. The inclusion of only two generations of ancestors (A2) or all ancestors (Af) was also evaluated. A total of 34,506 US and 1305 Israeli Holsteins bulls, and 5236 pigs were genotyped. The evaluations were computed by traditional BLUP and single-step GBLUP, with respective computing performance recorded. For the two Holstein datasets, coefficients of determination and regression of deregressed evaluations from a full dataset with records up to 2011 on EBV or GEBV from the reduced dataset (up to 2006 for Israeli and 2007 for US) and truncations were computed. The thresholds for old data deletion were based on generation intervals of 5 yr. For the PIC dataset, correlations between corrected phenotypes and EBV or GEBV were used to evaluate the predictive ability on young animals born in 2010 and 2011. The reduced dataset contained data up to 2009 and the thresholds were based on generation interval of 3 yr. The number of generations that could be deleted without reduction in accuracy was dependent on data structure and trait. For US Holsteins, removing 3 and 4 generations of data did not reduce accuracy of evaluations for final score in Af and A2 scenarios, respectively. For Israeli Holsteins, the accuracies for milk, fat, and protein yields were the highest when only phenotypes recorded on year $\geq$ 2000 and full pedigrees were included. Of the 135 Israeli validation bulls with genotypes and daughter records only in the complete dataset, 38 and 97 were sons of Israeli and foreign bulls, respectively. While more phenotypic data increased the prediction accuracy for sons of Israeli bulls, the reverse was true for sons of foreign bulls. For PIC dataset, removing data up to five generations did not erode predictive ability for genotyped animals for litter size and number of stillborn. Given the data used in this study, truncating old data does not decrease the accuracy on young genotyped animals, while reducing computation requirements and helping to find problems due to population structure. For populations that include local and imported animals, the truncation may be beneficial for one group of animals and detrimental to another.

**Key Words:** ssGBLUP, pedigree depth, genomic selection

### 0169 Use of linear models with normal, Student-t or Slash distributed error for the analysis of quantitative traits.

B. Mestav*[1], K. Kizilkaya[2], and S. O. Peters[3], *[1]Canakkale Onsekiz Mart University, Canakkale, Turkey, [2]Adnan Menderes University, Aydin, Turkey, [3]Berry College, Mount Berry, GA*.

Some symmetric and heavy-tailed distributions, such as Student's-*t* and Slash, have been suggested for robust inference in linear mixed models. These robust models are characterized by the degrees of freedom of these distributions and include the normal distribution when the degrees of freedom approach infinity. The objective of this this study was to investigate joint estimation of degrees of freedom for the residual and all other genetic and non-genetic parameters. In a simulation study, five different populations with five replicates were simulated using multivariate linear mixed effects animal models with Normal (NOR), three (ST3) or ten (ST10) degrees of freedom Student-t, and one and half (ST1.5) or three (SL3) degrees of freedom Slash distributions. Multivariate data within each replicate were generated for 18,000 progeny from 10 sires and 20 dams mating, which is selected through three generations. Models with multivariate Student's-*t*, Slash and Normal residuals were fitted to each dataset using a hierarchical Bayesian approach. Predictive log-likelihood (PLL) values strongly favored the multivariate Student's-*t* and Slash models over the Normal models for simulated heavy-tailed datasets. Posterior mean estimates of degrees of freedom parameters seemed to be accurate and unbiased. Estimates of sire and herd variances were similar, if not identical, across fitted models. Posterior mean and 95% posterior probability interval (PPI) estimates of error variances in simulated datasets were found to be (downwardly or upwardly) biased when the fitted model was not the true model. Reliable estimates of degrees of freedom were obtained in all simulated heavy-tailed and normal datasets. The predictive log-likelihood was able to identify the correct model among the models fitted to heavy-tailed datasets. The results obtained indicated that there was no disadvantage of fitting a heavy-tailed model when the true model was normal.

**Key Words:** multivariate heavy-tailed distributions, robust linear mixed model, MCMC