Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Single Step Methods with a View towards Poultry Breeding

A. Legarra[1], I. Aguilar[2], I. Misztal[3]
[1]INRA, UMR1388, Toulouse, France, [2]INIA, Las Brujas, Uruguay, [3]University of Georgia, Athens, GA

ABSTRACT: Single Step methods combine pedigree relationships and marker genotypes in a single conceptual and streamlined framework based on extending animal model BLUP to include marker genotypes. The extension can be done because genotypes at markers follow covariances described by the pedigree. This results in a relationship matrix that combines pedigree and markers, called **H**. Fitting $\mathbf{H}^{-1}$ in Henderson's Mixed Model Equations gives the Single Step Genomic BLUP, which is a single estimator of breeding values that includes all available information. It is easily generalizable to multiple trait and many different models, including Bayesian regressions. Extensions to crosses are largely untested and may need theoretical developments. Single step is increasingly used in poultry breeding, for its ease of use, generality, and better accuracy than any competing method. Some results in poultry are discussed.
Keywords: genomic; evaluation; BLUP

Introduction: Why Single Step

"I am Mr. Wolf. I solve problems."
Reservoir Dogs

Animal breeders are problem solvers. They have a long tradition of solving practical problems while at the same time keeping an eye on most recent technical developments. The paramount example is BLUP (and its cousin REML) that allows accommodating genetic evaluation for almost any species and set of traits, across very unbalanced data sets and possibly very different relatives. This is due to the use of state of the art theory (linear models, relationship matrices, quantitative genetics), and also because Henderson's formulation of BLUP leads to the existence of very efficient algorithms.

Genomic evaluation methods had been thoroughly tested in the last five years. Two main conclusions emerge. The first is that parametric methods (Bayesian regressions and GBLUP, which is a particular case of Bayesian regression) are the most efficient ones, in spite of their implicit assumptions such as linkage equilibrium across markers or additive marker action. The second is that accuracy in genomic predictions comes in two ways (Habier, 2010): better genetic relationships (VanRaden 2008; Hayes et al., 2009), as in GBLUP, and capture of major genes (DGAT1, IGF2) by markers (e.g., VanRaden 2009). GBLUP can capture these major genes but Bayesian regressions do an even better job. The existence of major genes varies by trait and species (e.g., in dairy cattle they only seem to exist for production traits). Increasing the marker density does not therefore improve the capture of these genes, because they do not seem to exist (e.g., VanRaden et al., 2013).

Genomic selection is costly, in particular for small species. Key animals are typically genotyped. This breaks down the carefully constructed relationship matrix used in BLUP, and the Animal Model can no longer be used. Using sire or reduced animal models seems inadequate since selection on females' side will not be accounted for. Therefore the current solution consists in projecting data of closely related animals to create pseudo-data for genotyped animals. These close animals can be offspring (VanRaden and Wiggans, 1991; Garrick et al., 2009) or own performance, parents and offspring (Ricard et al., 2013). The projection usually involves a regular BLUP Animal Model evaluation (with pedigree) with post processing of results (e.g. Garrick et al 2009, VanRaden et al. 2009).. Later, the pseudo-data are feed to a genomic evaluation model.

This process is called multiple step and is a success in dairy cattle because pseudo-data for progeny-tested bulls are very precise. But this is not the case in most species and this leads to several problems. First, it is clumsy. Clumsiness *per se* is not a problem, but induces to errors. For instance, a genomic evaluation for dairy bulls involves: (a) running the regular genetic evaluation and extracting the solutions, (b) computing for each record a record "corrected" by estimates of environmental effects and cow's dam, (c) averaging for each bull these corrected records across daughters, (d) giving a weight (or precision) to these averages (and usually this weight is an approximation), (e) running a genomic evaluation and sometimes (f) combining its results with (a). Getting all programs, each one with its intricacy, to run without fails is delicate.

Second, and more important, it ignores or grossly approximates information. Many relatives of genotyped individuals are ignored (for instance the dam). Correction of records is done "as if" effects were estimated exactly. Precisions of pseudo-records are approximations, and covariances across pseudo-records are ignored. For instance, if two sires have their offspring in the same farm this information is ignored. Also, the genetic trend is tricky to be accounted for.

Third, it does not propagate to non-genotyped animals. For instance, candidates to selection are selected among offspring of elite males and females. Males benefit from genomic evaluation but females (which are themselves daughters of genomic males) do not. However, large selection pressures can be exerted on females. Fourth, it is affected by selection (Patry and Ducrocq, 2012). Pseudo-

data will be calculated "as if" there is no genomic selection, and therefore true genetic trend will be underestimated. This will also hamper comparison of EBV's across generations.

Fifth, and perhaps the most important one in the long term, the multiple step process cannot be generalized. Each of the animal model BLUPs for multiple traits, maternal effects, competition models, random regression, threshold models, and so forth, needs *ad hoc* adjustments to fit into multiple step methods. For instance, there is no easy way to create pseudo-data for competition effect models, because pseudo-data for each sire will be related across sires due to their progeny sharing pens. The random regression pseudo-data exists (Liu et al., 2004) but is quite complex. Multiple trait de-regression for traits recorded in different sets of candidates (e.g. food conversion rate and daily growth) becomes very complicated, although doable (Liu et al., 2004).

The joint pedigree and genomic relationship matrix H

For these reasons, researchers have tried to combine Animal Model BLUP with genomic evaluations in a streamlined, conceptually coherent, framework. VanRaden (2008) GBLUP fits naturally into existing BLUP software and methods (including threshold models, multiple traits, etc.). Therefore, a common idea was to extend genomic relationship matrices to all individuals (Misztal et al., 2009; Christensen et al., 2010). First attempt (Legarra et al., 2009) was naïf: substitute pedigree relationships $\boldsymbol{A}$ by genomic relationships $\boldsymbol{G}$, if they are available. In mathematical form (denoting subindex 1 for un-genotyped): $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{G} \end{pmatrix}$ but this matrix makes little biological or mathematical sense. Two twins identified thanks to $\boldsymbol{G}$ would seem to produce offspring that are cousins and not sibs. Also, it can be shown that a dairy cattle evaluation with genotyped bulls would give the same results as without genotypes with regular $\boldsymbol{A}$ (Aguilar et al., 2010). A better way of considering the problem projects genomic relationships into pedigree relationships (Legarra et al., 2009) or, in other words, works imputing missing genotypes, and this imputation is based on pedigree relationships (Christensen and Lund, 2010).

The last derivation deserves a further look. Linear imputation based on pedigree relationships is akin to consider genotype as a quantitative trait. Matrix $\boldsymbol{G}$ is formed by a matrix with genotypes:

$$\boldsymbol{G} = \boldsymbol{Z}_2 \boldsymbol{D} \boldsymbol{Z}_2'$$

where $\boldsymbol{Z}_2$ for the i-th animal consists in one row with, for the j-th loci, the genotype coded as $\{-2p_j, 1-2p_j, 2-2p_j\}$ for genotypes $\{AA, Aa, aa\}$ and p_j is the frequency of allele a. However, the genotype so coded is a quantitative trait (animals receive half their parents plus a Mendelian sampling) and therefore genotype of the non genotyped animals can be predicted: $\widehat{\boldsymbol{Z}}_{\boldsymbol{1}} = \boldsymbol{A}_{\boldsymbol{12}} \boldsymbol{A}_{\boldsymbol{22}}^{-\boldsymbol{1}} \boldsymbol{Z}_{\boldsymbol{2}}$ **.** However this prediction is inaccurate, in particular for animals far apart: $\boldsymbol{Var}(\widehat{\boldsymbol{Z}}_{\boldsymbol{1}} | \boldsymbol{Z}_{\boldsymbol{2}}) = (\boldsymbol{A}_{\boldsymbol{11}} - \boldsymbol{A}_{\boldsymbol{12}} \boldsymbol{A}_{\boldsymbol{22}}^{-\boldsymbol{1}} \boldsymbol{A}_{\boldsymbol{21}}) \boldsymbol{V}$ where $\mathbf{V}$ contains $2p_j(1-p_j)$ in the diagonal. The genomic relationship matrix across non-genotyped individuals is something like $\widehat{\boldsymbol{Z}}_{\boldsymbol{1}} \boldsymbol{D} \widehat{\boldsymbol{Z}}_{\boldsymbol{1}}'$ **,** *but* the fact that imputation is not perfect and has error as above needs to be taken into account. Putting everything together results in

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{G}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{G} \\ \boldsymbol{G}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{G} \end{pmatrix}$$

with inverse

$$\boldsymbol{H}^{-\boldsymbol{1}} = \boldsymbol{A}^{-\boldsymbol{1}} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^{-\boldsymbol{1}} - \boldsymbol{A}_{\boldsymbol{22}}^{-\boldsymbol{1}} \end{pmatrix}$$

(Christensen and Lund., 2010; Aguilar et al. 2010). The inverse is particularly simple and rather sparse. It is composed of three parts: regular relationships $\mathbf{A}^{-1}$, genomic relationships $\mathbf{G}^{-1}$, and a correction $\mathbf{A}_{22}^{-1}$ to avoid double counting of genotyped animals. This $\boldsymbol{H}^{-\boldsymbol{1}}$ is particularly convenient because it can be fitted in BLUP in a straightforward manner: $\boldsymbol{A}^{-1}$, although in the hundreds of thousands or millions, is very sparse; and dense submatrices $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ are of size number of genotyped individuals, which is typically rather small (in the thousands except for dairy cattle).

A word needs to be said about imputation. Genotypes of non-genotyped individuals can be imputed using other methods that consider linkage and Mendelian coherence (AlphaImpute, FImpute, etc.). Typically this results in a single genotype for each individual. Then the genomic evaluation method proceeds as if all imputed animals are imputed exactly, ignoring the uncertainty of the imputation both *within* an individual and *across* individuals. This will result in severe biases.

An example may clarify this point. Assume a very long complex pedigree and the final generation genotyped for one locus, with allelic frequency $p = freq(a)$. Due to only having one generation with genotypes and to the long and complex pedigree, best guesses of genotypes in the base animals will be nearly identical and equal to *2p*, for <u>all</u> base animals. Therefore, using "best guess" of genotypes without taking uncertainty into account, all base population individuals will be treated by the genomic evaluation as identical (even if they have different phenotypic records), which will force them to have the same estimated breeding value, which is paradoxical.

In practice, using imputed genotypes for all animals will bias and reduce the accuracy of EBV's of individuals "far" from those being genotyped, and this is the reason why Hickey et al. (2011) obtained worse accuracies by imputing all individuals than using the Single Step. In addition, it seems a waste of resources imputing millions of females to evaluate thousands of males (an example will be shown later).

Matrix $\boldsymbol{H}$ has interesting aspects. It can be verified that is (semi)positive definite if $\boldsymbol{G}$ is (semi)positive definite. If all animals are genotyped, then $\boldsymbol{H} = \boldsymbol{G}$. If none is genotyped, $\boldsymbol{H} = \boldsymbol{A}$. Also, it accommodates genomic information by tweaking pedigree information to get closer to genomic one. For instance, two non genotyped sibs will be made more related than 0.25 if their offspring is more related than 0.125.

Single Step GBLUP

Because **H** is the covariance matrix of *all* individuals, there is no need to use any pseudo-data. Regular derivation of BLUP holds and the mixed model equations are, in the usual notation:

$$\begin{pmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}W} \\ \boldsymbol{W'R^{-1}X} & \boldsymbol{W'R^{-1}W + H^{-1} \otimes G_0} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{W'R^{-1}y} \end{pmatrix}$$

This is very desirable. Computation is very much streamlined (one set of equations for the whole process). Compared to regular BLUP, the only additional work is the construction of $\boldsymbol{G^{-1} - A_{22}^{-1}}$, which can be easily encapsulated either in an external program or within the genetic evaluation. In addition, the general framework of BLUP holds and any model fit in BLUP (e.g., random regression or multiple trait models) can be fit in SSGBLUP. Variances can also be estimated by REML or Gibbs sampler procedures.

As discussed, $\boldsymbol{H^{-1}}$ is very sparse. For instance, the study of Chen et al. (2011) considered 290,000 chicken, 4,000 of then genotyped. Imputing all genotypes would need 50,000 markers x 290,000 animals $\approx$ 14.5 Gbytes. However, $\boldsymbol{H^{-1}}$ contains 9 x 290,000 coefficients (from the $\boldsymbol{A^{-1}}$ part) and $4{,}000^2$ coefficients (from $\boldsymbol{G^{-1} - A_{22}^{-1}}$), with a total of $\approx$ 0.019 Gbytes. This also lowers the computational time. For this example, running the genetic evaluation (after imputing all animals) would have an associated cost proportional to $50{,}000^2$ operations, versus $4{,}000^2$ for Single Step GBLUP.

Single Step versus Reduced Animal models. Wolc et al. (2011a,b) proposed the use of a Reduced Animal Model. The Reduced Animal Model can be used if there is only one generation of non-genotyped individuals after the genotyped individuals. This method assigns performance of the (nogenotyped) progeny of genotyped individuals to their genotyped parents, and is equivalent to a Single Step in which pedigree starts at the genotyped individuals (Legarra et al., 2009). The authors also derived a Reduced Animal version of Bayesian Regressions, in which average phenotype of the offspring is a function of average *genotype* of the parents, which is more convenient than assigning this phenotype to either parent as with pseudo-data.

Compatibility of pedigree and markers. Quality control is of utmost importance, i.e., to avoid discrepancies between **G** and **A**. Even then, VanRaden (2008) and Vitezica (2011) showed that markers and pedigree relationships need to be compatible for the genomic evaluations to be unbiased; this was empirically verified in chicken by Chen et al. (2011). This is true for the Single Step or any other genomic evaluation method, often implicitly. The kernel of the problem is that typically pedigrees go back more generations than markers, and therefore markers cannot "see" the effect of drift or selection. Also, discrepancies between **A** and **G** exists if pedigrees are incomplete. They seem unimportant when proven genotyped animals have large information (e.g., in dairy) but can cause biases and convergence problems otherwise (Misztal et al., 2013).

There are a few solutions to this problem. The first is to cut off irrelevant information, and consider pedigrees and phenotypes from the last few generations, even increasing accuracy (Lourenco et al., 2014). If we are interested in genetic evaluations now, information (pedigree and also phenotypes) ten generations back is useless and only adds noise to the evaluation. The second is to force unbiasedness of the genomic relationship matrix by creating ("tuning") a modified matrix $\boldsymbol{G^*} = a + b\boldsymbol{G}$; where a describes the existing average (pedigree) relationship at the time of genotyping, and b the reduction in genetic variance due to drift and selection. This results in more accurate and unbiased predictions, both in simulated and real data (Vitezica et al., 2011; Chen et al., 2011; Christensen et al., 2012). The correction implicitly assumes a population mating at random and Christensen (2012) has suggested to "tune" $\boldsymbol{A}$ (which makes accepted but false assumptions) to match $\boldsymbol{G}$ constructed with 0.5 allelic frequencies. This **G** is independent of pedigree quality.

Crosses. This is one of the most active areas of research in genomic evaluations. First, there is a decision to be made concerning whether the marker effects are the same across the pure lines and their crosses (Ibanez-Escriche et al., 2009; Zeng et al., 2013). This can be estimated (Karoui et al., 2012). If they are the same, a relationship matrix can accommodate all genotyped animals. In pure GBLUP methods, the way to construct this matrix is largely irrelevant (Strandén and Christensen, 2011). However, for Single Step the correct procedure to set up **G** matrices (possibly with tuning) and **H** matrices is still far from being clear in crosses. Lourenco et al. (unpublished) working with simulated data obtained the best results with **G** constructed as for a single population. Alternatively, one could fix **G** at allelic frequencies of 0.5 and tune **A**, extending Christensen's (2012) methodology. Recently, Christensen et al. (in press) suggested separating the effects in the F1 in pure breed gametes, and fit separate **H** and **G** matrices within breed. This assumes that marker effects are different across breeds. At any rate, a general model accommodating crosses is sought, and testing in real data sets is much needed. Note that the same problems apply to multiple step methods, which proceed by assuming that pure lines are completely unrelated.

Bayesian regressions. Bayesian regressions (BayesA,B,C and R; Bayesian Lasso; VanRaden (2008) nonlinear A and B; etc.), in which marker effects are explicitly fit, sometimes give more accurate estimates than GBLUP like methods, in particular in the presence of major genes. There are ways to accommodate them in Single Step. The first one is to use expressions in Legarra and Ducrocq (2012). One of them iterates on regular BLUP evaluation and Bayesian regressions; the other one is explicit on marker effects. Both are cumbersome to program but not of very large size. For instance, the Chen et al. (2011) data set would be analyzed fitting a Gibbs sampler for an animal model with 290,000 unknowns, and another Gibbs sampler for 50,000 markers in 4,000 animals. These approaches are untested for two reasons. First, this is not felt as an urgent need; second, it needs some programming and iteration may be very long. A simpler approach uses the fact that marker effects can be backsolved from EBVs: $\widehat{\boldsymbol{a}}|\widehat{\boldsymbol{u}}_2 = \boldsymbol{DZ}_2'\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}}_2$ (Wang et al., 2012) to give more weight to markers with larger effect. This is also useful for GWAS, and about the only alternative for complex traits where pseudo-data are hard to construct (Dikmen et al., 2013).

Computational aspects. Straightforward application of Single Step GBLUP is most adequate for populations with up to 100,000 genotyped individuals. Beyond this number, matrices $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ become very large (and difficult to store and handle). The computation of matrices $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ and their inverses is done by efficient algorithms and, preferably, using parallel computations (Aguilar et al., 2011). With an average workstation, timing is about 1h for 100k animals, rising cubically. Higher limits can be obtained either by indirect predictions (predictions of SNP effects from GBLUP and subsequent use of these effects for prediction of young animals) or by alternate computing algorithms (Misztal et al., 2009; Legarra and Ducrocq, 2012). Misztal et al. (2014) proposed a special $\boldsymbol{G}^{-1}$ based on genomic recursions and a decomposition into proven and young (or low accuracy) animals. Costs are linear for young animals and preliminary tests indicate accuracy similar to regular $\boldsymbol{G}^{-1}$.

Convergence problems have been reported (VanRaden, unpublished; Harris and Johnson, 2013; Aguilar et al., 2013) which seem related to inclusion of cows or non tested bulls (with little information) in $\boldsymbol{G}$. This seems a problem only for very large dairy cattle data sets.

Unknown parent groups. This is delicate in genomic evaluations. Pseudo-data is supposed to be already corrected for unknown parent groups but in our experience, pseudo-data can be biased (unpublished). Introduction of unknown parent groups in Single Step is most efficiently done using the untransformed Thompson-Westell equations $= \boldsymbol{Xb} + \boldsymbol{Qg} + \boldsymbol{u} + \boldsymbol{e}$ with explicit genetic groups *before* the QP transformation (Misztal et al., 2013). This can dramatically improve the accuracy of evaluations. Use of $\boldsymbol{A}^{-1}$ with unknown parent groups, albeit approximate, is usually a good compromise if groups are well defined (Tsuruta et al., 2013).

Process

In this section we will describe the streamlined process of genomic evaluation using Single Step, as implemented in the suite of programs Blupf90 (http://nce.ads.uga.edu) (Misztal et al., 2002). Pedigree and phenotypes are typically obtained from some kind of data base, and converted into text files, possibly with alphanumeric identifiers. Genotypes are typically stored in another database and output as text files with identifier and genotype coded as 0/1/2. If raw files from Illumina are used, `Illumina2preGS` can be used to generate the needed files.

Renumbering. Because most softwares do not accept alphanumeric fields, programs such as `renumf90` can recode and verify the three files (pedigree, data and markers). This includes creating crosslinks across genotypes and identifiers used in pedigree and data file.

Quality control. A number of quality control checks need to be done. The most common ones are call rate, Hardy-Weinberg equilibrium (*within* population), parent-offspring discordances (i.e., a sire cannot be AA and its son aa), and sample duplicates (same genotype, different identifiers, or same identifier, two genotypes). Most errors are due to mislabeling but they are unfortunately hard to track down. In addition, discordances between matrices $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ are checked, although this check can (should) be disabled if the pedigree is of low quality. Also, matrices $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ can be different if the depth of the pedigree is much longer than the genotypes (case of swine and chicken) or if the pedigree has many missing parentships (case of ruminants). Typically, $\mathbf{G} - \mathbf{A}_{22}$has a SD of 0.04 for complete pedigrees. Correlation is ~0.8 for off-diagonal elements but can be as low as 0.3 for inbreeding coefficients (the reason is that realized inbreeding depends on many more events than across-individuals relationships). In this case, it is better to cut pedigrees, not checking, or both. These checks can be done by program `preGSf90` and, in part, by others such as `Mendelsoft` or `Plink`.

Computation of $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$. This is typically done by an external program, that can be `preGSf90`, and matrices are verified. After computation, these two matrices are stored as binary files and read. This computation can be also be done from other software (`blupf90, remlf90`, etc).

Solving, and backsolving marker effects. Afterwards, the solvers (`blupf90, remlf90`, etc) set up the mixed model equations with regular $\boldsymbol{A}^{-1}$, include $\boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1}$ to create $\boldsymbol{H}^{-1}$, and solve. If needed, `postGSf90` can be used to obtain marker effect estimates and (potentially) iterate Bayesian regressions as described above. With marker effects computed, `predf90` can predict breeding values based on genotypes only.

Practical experiences

Whom to genotype. Creating a large reference population *per se* is pointless. Most of the information comes from close relationships, and therefore what is needed is to genotype a good representation of the recent genetic background of the entire population. This can be achieved by genotyping, e.g., the last ten years males. Phenotyping each generation is needed, because of the decline of accuracy, even with Bayesian regression methods (Habier et al., 2010; Wolc et al., 2011b). Research in dairy cattle has shown little extra accuracy by including females' genotypes and phenotypes. So a way to go in ruminants seems to continuously genotype key individuals: males. However, unpublished work by Lourenco, Fragomeni et al. shows that in real poultry data, males' genotypes benefit males, females' genotypes benefit females, but joint genotypes benefit both.

Species were the Single Step has been extensively tested include dairy cattle, dairy sheep, swine and poultry.

Dairy cattle. Single Step GBLUP can be *less* accurate than multiple step methods for fat and protein contents, where a major gene exists. However "Single Step Bayesian Regressions" have not been seriously attempted so far in large dairy cattle data sets. Otherwise, for most traits it is as accurate as other methods and often slightly less biased. The reason why it is not more accurate is because all bulls are genotyped, and also because pseudo-data for each bull is extremely accurate.

Dairy sheep. Single Step is more accurate (~0.05-0.20 increase in accuracy) and less biased than GBLUP. Genetic evaluation is quite straightforward and includes up to a few million animals and ~4,000 genotyped rams in Lacaune. Single Step provides higher accuracies than regular BLUP even for small breeds such as Basco-Bearnaise, with 500 genotyped males, provided that all recent cohorts of males are completely genotyped. Inclusion of unknown parent groups is important and done via the untransformed Thompson-Westell model, which results in correct genetic trends.

Swine. The most extensive testing of Single Step in (pure line) pigs was done by Christensen et al. (2012). They showed a higher accuracy (~ extra 0.10) of Single Step methods than pure GBLUP methods. In particular, Single Step allowed much higher accuracy for the scarcely recorded Food Conversion Rate, which benefitted from genotypes and from the joint prediction with the massively recorded daily gain. This analysis is easily doable with Single Step, but difficult with any other prediction method.

Chicken. Extensive analysis have been done, either with Single Step (Chen et al., 2011a,b; Simeone et al., 2011) or its simplified version the Reduced Animal Model (Wolc, 2011a,b). The emerging trend is that both alternatives are more efficient than regular genomic predictions which do not consider ungenotyped relatives (Chen, 2011a). In the extreme, Bayesian Regression with genotyped animals alone can do a worse work than pedigree BLUP (Chen et al., 2011a). Reduced Animal Model GBLUP is as accurate as Reduced Animal Model Bayesian regressions, which simplifies computations and opens the door for multiple trait analysis, which are more accurate than single trait ones (Chen et al., 2011a). Joint analysis of two lines (but without any cross) using crude models did not increase accuracy versus separate analysis, and across-line rankings where dependent of assumptions of the model (Simeone et al., 2011).

Conclusion

Poultry breeders do not need to change (much) traditional practices (Fulton, 2012) to implement genomic selection. Current abundant information on pedigree and multiple trait phenotypes can be conveniently integrated in the Single Step. Genotyping of candidates to selection (young males) and key individuals (males) will become a routine process, with females being genotyped only if economically profitable. Attention to detail and good handling and quality control of genotyping, imputation, and matching pedigree, records and phenotypes will be of utmost importance. This will require well trained people. As for academia, theoretical developments in genomic evaluation including multiple lines and their crosses are of great importance, as well as testing with real data sets.

Table 1. Some examples of accuracy of Single Step GBLUP / Reduced Animal Model (RAM) versus pedigree or "pure" genomic evaluations*

Paper / Trait[§]	Method	Pedigree BLUP	"Pure" genomic evaluation
Chen et al., 2011a	Single Step GBLUP		
LS, Line 2	0.73	0.43	0.11
BM; Line 2	0.51	0.33	0.47
Wolc et al., 2011a	RAM		
ePD	~0.33	~0.18	
ICO	~0.65	~0.45	

*excluding phenotypes from non-genotyped individuals
[§]LS: leg score; BM: ultrasound measure of breast meat; ePD: egg production; ICO: egg color

Acknowledgements

This work has been financed by X-Gen and GenSSeq actions from SelGen metaprogram (INRA).

Literature Cited

Aguilar, I., Legarra, A., Tsuruta, S. et al. (2013) Interbull Bulletin 47

Aguilar, I., Misztal, I., Johnson, D.L. et al. (2010) J Dairy Sci 93:743-752

Aguilar, I., Misztal, I., Legarra, A. et al. (2011) J Anim Breed Genet 128:422-428

Chen, C.Y., Misztal, I., Aguilar, I. et al. (2011a). J. Anim. Sci. 89:23-28

Chen, C.Y., Misztal, I., Aguilar, I. et al. (2011b). J. Anim. Sci. 89:2673-2679

Christensen, O., Madsen, P., Nielsen, B. et al.(2012) Animal 6:1565-1571
Christensen, O.F. (2012) Gen Sel Evol 44:37
Christensen, O.F., Lund, M.S. (2010) Gen Sel Evol 42:2
Dikmen, S., Cole, J.B., Null, D.J. et al.(2013) PLoS ONE 8:e69202
Fulton, J.E. (2012) Animal Frontiers 2:30-36.
Karoui, S., Carabano, M.J., Diaz, C. et al. (2012) Gen Sel Evol 44:39.
Garrick, D.J., Taylor, J.F., Fernando, R.L. (2009) Gen Sel Evol 41:44.
Harris, B.L., Winkelman, A.M., Johnson, D.L. (2013) Interbull Bulletin 47.
Hickey, J.M., Kinghorn, B.P., Tier, B. et al. (2012) Gen Sel Evol 44:9
Legarra, A., Aguilar, I., Misztal, I. (2009) J Dairy Sci 92:4656-4663.
Legarra, A., Ducrocq, V. (2012) J Dairy Sci 95:4629-4645.
Lourenco, D.A.L., Misztal, I., Tsuruta, S., et al. (2014) J Dairy Sci 97:1742-1752.
Misztal, I., Legarra, A., Aguilar, I. (2009) J Dairy Sci 92:4648-4655.
Misztal, I., Vitezica, Z., Legarra, A. et al (2013) J Anim Breed Genet 130:252-258.
Quaas, R.L. (1988) J Dairy Sci 71:1338-1345.
Ricard, A., Danvy, S., Legarra, A. (2013) J Anim Sci 91:1076-1085.
Simeone, R., Misztal, I., Aguilar, I. et al. (2011) J. Anim. Breed. Genet. 129:3-10
Strandén, I. and Christensen, O.F. (2011) Gen Sel Evol 43:25.
Tsuruta, S., Lourenco, D.A.L., Misztal, I. (2014) Book of Abstracts EAAP meeting, 231.
VanRaden, P.M. (2008) J. Dairy Sci. 91:4414-4423.
VanRaden, P.M., Tassell, C.P.V., Wiggans, G.R. et al (2009) J Dairy Sci 92, 16-24.
Vitezica, Z., Aguilar, I., Misztal, I. et al (2011) Gen. Res. 93:357-366.
Wang, H., Misztal, I., Aguilar et al. (2012) Gen. Res. 94:73-83.
Wolc, A., Stricker, C., Arango, J. et al. (2011a) Gen Sel Evol 43:5
Wolc, A., Arango, J., Settar, P. et al. (2011b) Gen Sel Evol 43:23
Zeng, J., Toosi, A., Fernando, R.L., et al (2013). Gen Sel Evol 45:11