

BY THE NUMBERS

by Andre Garcia, Angus Genetics Inc.

The Evaluation Machine

From cars to genomic evaluations, machines require maintenance.

The statistical machinery used to calculate expected progeny differences (EPDs) can be likened to other machines we use in our daily lives. Take a car, for example. A car is a collection of parts, mechanical and electronic, that work together to take us where we need to go. In a similar way, the “single-step machine” that runs weekly at the American Angus Association is a collection of statistical methods and algorithms that calculate EPDs. This analogy, although not perfect, is helpful to understand several aspects involved in the process of combining all of the data submitted by members and computing EPDs.

In order to work properly, a car, or any other machine for that matter, has a few basic requirements. First, all of its parts need to be working in synergy for the whole machine to function. Second, it needs inputs such as engine oil and gas. Third, it needs regular maintenance to make sure it will be there, fully functional, when we most need it.

Similarly, in order to calculate EPDs accurately, our genetic evaluation also has a few requirements. First, sound statistical methods and algorithms need to be combined into a software that calculates the EPDs. Second, a good

genetic evaluation needs inputs such as high-quality phenotypes, pedigrees and genotypes. Third, genetic evaluations also need maintenance and updates from time to time.

This maintenance aspect, of both cars and genetic evaluations, can sometimes be overlooked as we tend to go with the approach, “if it isn’t broke, don’t fix it.” Being proactive and identifying components that need updates before they become outdated or break is a good strategy to avoid bigger issues in the future.

We all know regular checkups can make a difference in the longevity and performance of a car. Likewise, for genetic evaluations, regular maintenance of models, genetic parameters and the algorithms under the hood is necessary to ensure our evaluation machine is up to date and properly applying cutting-edge technology to calculate EPDs.

Under the hood

The single-step genomic best linear unbiased prediction (ssGBLUP) machine allows us to combine information from pedigrees, phenotypes and genotypes into a simple, accurate and efficient genomic evaluation. In single-step, the genomic information is used to calculate more accurate relationships

among animals by building a genomic relationship matrix (GRM). These relationships are then combined with the traditional pedigree relationships, connecting all the information and increasing the accuracy of the evaluations.

Although ssGBLUP provides a simple and efficient framework to implement genomic evaluations, when the number of genotyped animals surpasses 150,000, there is simply not enough computing power to directly create the inverse of the GRM, which is needed to calculate EPDs.

However, genomic evaluations with far more than 150,000 genotyped animals have been conducted using single-step for years, clearly illustrating researchers have managed to overcome this barrier.

In fact, this was accomplished in 2014, when researchers at the University of Georgia developed a procedure called “Algorithm for Proven and Young” (APY) to facilitate the use of ssGBLUP for large genotyped populations.

Simply speaking, this algorithm works by setting a group of reference animals to represent the entire genotyped population, and then calculates the inverse of the GRM for

Continued on page 24

all genotyped animals as a function of this group.

In the early developments of APY, highly accurate (proven) animals were used as this reference. However, further research showed if the reference group was large enough, the choice of animals did not affect the accuracy of the predictions, and “proven and young” became “core and non-core,” though the procedure is still referred to as APY.

In the context of genomic evaluations, APY is not a new concept, and it is certainly not new to the Association. In 2017, after extensive research and testing by Angus Genetics Inc. (AGI) and the University of Georgia, single-step with APY was implemented in the Angus evaluation and has run on a weekly basis ever since. Thanks to these sophisticated algorithms, our single-step genomic evaluation currently runs with millions of animals and more than 950,000 genomic profiles. Additionally, research using data from the U.S. dairy industry has shown it is feasible to include more than 2.3 million genotyped animals in ssGBLUP using APY.

Powered by Angus breeders, the uptake of genomic evaluation and genotyping technology has been incredible. With genotyping at full speed, it is time for APY to receive a core update.

Why update?

Since the implementation of single-step in 2017, more than 600,000 animals have been genotyped, which is more genomic testing than in all previous years combined. As such, the majority of genotyped animals were not yet born when the core group was established. To ensure the core animals in APY are a good representation of the current

genotyped population, an update is needed. It is important to note, regardless of which core is being used in APY, all animals are included in the GRM and all relationships are accounted for in the weekly evaluation.

The unprecedented milestones of 20 million registered animals achieved in March, and one million genotyped animals, which will be achieved by the Association in 2021, are signs of the prosperity of the Angus breed. These milestones put the Association in a unique position to leverage millions of data points to compute EPDs for its members.

However, such a large amount of data brings the great challenge of crunching all of the numbers in a limited amount of time. Members may be surprised to learn it takes days, not hours or minutes, to perform all of the calculations.

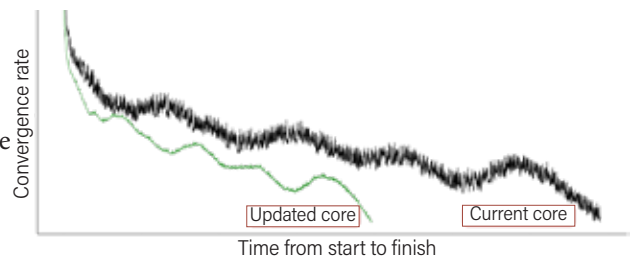
At the start of single-step in 2017, weekly evaluations began running on our Association’s computers on Friday evening and finished before Monday morning. As more and more information has been added, this completion time has stretched into late Tuesday evening, an increase of almost two days.

A core update leads to better representation of the genotyped population, which improves the overall fit of the model (making the model better) and greatly reduces computing times. Figure 1 shows a representation of the computing time needed to finish an evaluation with the current core (black line) and with the updated core (green line).

The bottom line


Good data, sophisticated algorithms and regular maintenance

Figure 1: Run time of evaluations with current and updated core in APY.



of the evaluation machine, including periodic core updates, are essential to ensure a good fit of our models and maintain the weekly delivery of EPDs. A series of preliminary tests with growth data have resulted in correlations of .99 between EPDs calculated using current and updated cores. However, some reranking of animals will occur as relationships are better defined. Outliers with changes exceeding +/- 10 pounds have been observed in test runs of weaning weight EPD.

In the age of genomics, the uptake of genotyping has been rapid and is being driven by progressive Angus breeders pushing the boundaries of genomic technologies. To keep up with this fast pace and provide the best genetic evaluation possible in a timely manner, the AGI team is committed to leveraging the power of the Angus database by conducting research and taking good care of our evaluation machine.

With the same amount of data, better modeling can reduce computing time and increase model fit. 

Andri D. Garcia
agarcia@angus.org

Editor's note: For questions, contact the AGI customer service team at 816-383-5100.