



A tool for monitoring genetic selection differentials in dairy herds in Canada

Bernard Ato Hagan and Roger I Cue

Department of Animal Science, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, Quebec, H9X 3V9, Canada

**A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada**

Abstract. *A software tool was developed to allow a dairy producer and/or agricultural advisor to monitor the genetic selection differentials (GSD) that a dairy farm is making. The objectives of this study were (i) to monitor GSD in individual farms, over years, so that producers can be advised as to whether or not they are achieving their selection objectives (and hence optimizing productivity and profitability); (ii) the development of a prototype software tool and visualization model to assist producers in interpreting the results for their individual farm, and to compare their farm results with suitable benchmarks. Data used for this study were the EBVs routinely calculated by the Canadian Dairy Network (CDN), for Milk Yield, Fat Yield, Protein Yield and Lifetime Profit Index. The Canadian Ayrshire breed has been used as a model, test breed; records on animals born between January 1980 and April 2016 were used. For each day between this period, the average sire EBV of all the sires available for use on that day was calculated; sires availability determined from the dates of sires' first and last usage. Sire GSD was then calculated as the EBV of the Sire of a cow minus the average sire EBV on the date the cow was conceived. The average GSD for the entire population, top and bottom 10% sires, and top and bottom 10% of herds per year of conception were computed and stored in a database. This then allows an individual producer to compare and visualize the individual animal selection he/she is making, and also to compare his/her herd against the average herd and the top 10% of herds (as a reference goal to potentially aim to also achieve). The developed software tool is updatable every time CDN releases new genetic evaluation list. The concept can be equally well applied to the other dairy breeds and livestock species for which genetic evaluations are routinely computed. This methodology is not limited to only the four named traits but can also be used for all traits genetically evaluated (currently approximately 30 traits in Canada), allowing a producer to monitor multiple traits and hence decide upon his/her selection objectives.*

Keywords. *Ayrshire, dairy, Estimated Breeding Values, Genetic selection differential, milk yield*

Introduction

Precision agriculture involves providing tailored advice and monitoring specific to the field or animal or individual farm; a prototype software tool has been developed to monitor dairy cattle genetic selection at the individual farm level. Precision dairy farming, in terms of precise selection decisions and regular monitoring of those decisions to align selection applied to the breeding objectives of producers, will improve the productivity and profitability of a dairy farm.

Dairy farmers in Canada, as in many parts of the world, seek to increase the profitability of their farming business by maximizing genetic gains for economic important traits that contribute to profit or reduction in production cost. Dairy producers make selection decisions each time they purchase semen for insemination or decide on which cows from within their herds to breed. In Canada, economic important traits include most of the traits for which genetic evaluations are generated and estimated breeding values (EBVs) produced for cows and bulls.

Animal genetic gain has a strong association with farm profitability (Thompson *et al.* 2015). Genetic gains a function of the genetic selection differential(s) and generation interval(s). Genetic selection differential (GSD) is the deviation of the mean EBV of parents from the mean EBVs of their contemporaries or an appropriate base group (Van Tassell & Van Vleck 1991; Burnside *et al.* 1992; Nizamani & Berger 1996; García-Ruiz *et al.* 2016). Besides GSD being a major determinant of annual genetic gains and indirectly related to profit, it also serves as a measure of selection practices in a population or herd and efficiency of current and historical selection practices in a herd.

Selection applied made by farmers may not always result in expected genetic gains in objective traits due to a number of reasons. These reasons include selection emphasis on traits other than their objective traits, lack of opportunity for selection due to small herd size, involuntary culling, non-random mating (Van Tassell & Van Vleck 1991; Burnside *et al.* 1992; García-Ruiz *et al.* 2016) and absence of information and feedback on farmers' selection activities. Probably the greatest limitation to achieving the expected genetic gains in dairy herds is the non-existence of tools for use by producers to monitor their selection practices and make changes if producers are going in the wrong direction. The Canadian Dairy Network (CDN) publishes national and provincial genetic trends for 16 traits of the dairy cattle breeds (Holstein, Jersey, Ayrshire, Brown Swiss, etc.) in Canada. Though this information gives an overview of the results of past selection practices in the entire population, it does not give any information on individual producers' performance so as to see what selection producers are making on their farms relative to the population.

The objectives of this study were (i) to monitor GSD in individual herds (farms), over years, for advising producers in their selection practices and (ii) develop a prototype precision agriculture software advisory tool and visualization model to assist producers in comparing the results of their farms in terms of selection differentials with suitable benchmarks.

Materials and methods

Data and animal

Data for this study were EBVs (estimates of genetic potentials of animals) routinely calculated (every 4 months) by the Canadian Dairy Network (CDN) using an animal model methodology. The EBVs used for the development of this prototype were milk yield, fat yield, protein yield and Lifetime Profit Index (LPI). Milk, fat and protein yields are the milk, fat and protein measured in kilograms from a lactating cow during a 305-day lactation.

Lifetime profit index is the expected lifetime profit of future progeny based on their genetic potential for production, longevity and udder health (Van Doormaal 2007).

All EBVs used for the development of this prototype software were from the Ayrshire breed. This breed is one of the seven dairy cattle breeds in Canada. Although the population size of this breed in Canada is not as large as that of the Holstein (93% of national dairy herds), it is adequate

enough for the exploration of the concept of this study and for demonstration of a future application to the other dairy breeds and livestock species for which genetic evaluations are routinely calculated.

There were primarily three data files which were the main inputs for this prototype software. These were bull and cow files, which contained information on animals' registration and date of birth, AI stud or herd identification, sire and dam registrations and trait EBVs, and the herd information file, contained information on individual herds (farms), the agricultural region where the farm is located, milking system being practiced by a farm and its production system.

Sire availability and Genetic Selection Differentials

The availability periods for all bulls for used for breeding during their life time were determined. For each sire, its first and last progeny dates of birth were taken from the EBV file. From these, the dates of conception of the first and last progeny were calculated by subtracting 280 days, being the assumed gestation length of cows, from the dates of birth of progeny. For each sire, the 1st and 99th percentile dates of conception of its first and last progeny were used as the sire's first and last usage dates. The period between these two usage dates was the sire's availability period. The 1st and 99th percentiles were used to exclude the odd outliers (for example a frozen semen kept for 20 years after a bull has long died and then used for breeding later). For unused bulls (bulls that did not produce offsprings), the dates at which they attained 18 and 24 months were determined. The period between these two dates was deemed as the availability period of such unused bulls, for potential use in a young-sire testing program, with the assumption that an unused bull will be culled after two years. The average EBVs for each day from 1st January, 1980 to date was computed for all available sires and bulls.

The GSD for the sire of each cow was computed as the deviation of the sire's EBV from the mean EBV of all available sires and bulls on the day of conception of a cow. New GSDs for all parents will be computed each time new genetic evaluations are released and these stored in a database.

Development of software suites

A prototype software suite has been developed using SAS (because it is a Rapid Application Development tool) to compute average EBVs and GSDs on a routine basis and produce graphs for individual herds and benchmarks. These computations are being translated to open source tool (Linux system tools, Fortran, R). The step by step procedure involved in the development of the prototype is illustrated in the flowchart in figure 1 and explained below.

- I. Dairy herd information (DHI) together with current genetic evaluations of cows and bulls (EBVs) will be used to determine the availability period of each parent (sire) using the dplyr package and the quantile function of R.
- II. The average EBVs of all available sires on each day from 1st January, 1980 to date will be computed using a specially-written Fortran program (for speed efficiency reasons). This step will be repeated for each of the 28 traits in the EBV file and all results stored in tables for subsequent use.
- III. The GSD of each sire of any given cow will be computed for all the 28 traits using a purpose-written Fortran program. The lmerTest package of R will be used to compute the standard errors of the GSDs. All results will be stored in data tables for use in steps V and VI.
- IV. A number of benchmarks of GSDs in the population have been computed and stored in tables.
- V. On-demand visualization and reports are generated using open source software (R and gnuplot).
- VI. All the information on average EBVs, GSDs for each sire and of individual herds and the benchmarks and visuals will be delivered in a test web-server to allow individual producers to sign-in and access on-demand reports and graphs of their herds.

All information (average EBVs and GSDs) will be recomputed each time new genetic evaluations

are released by CDN.

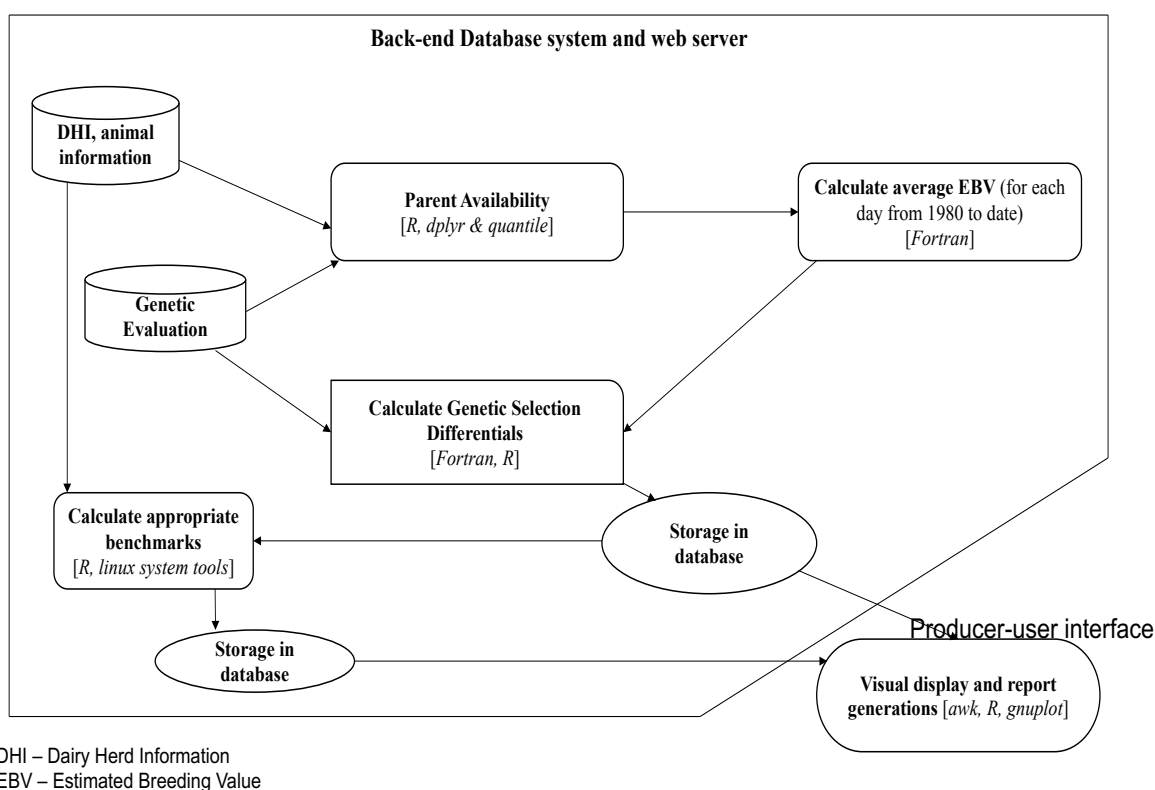


Fig. 1. Flow chart of the procedures in the development of prototype software

Table 1 illustrates the data structure for the computation of GSD for each sire. Arbitrary animal identifiers and figures are used to illustrate the computations

Table 1. Computation of genetic selection differentials (GSD) for sires of given cows

Animal Id	DOB	DOC	Sire Id	Dam Id	Sire EBV	Mean EBV	GSD
Cow001	30/12/1990	25/03/1990	Sire001	Dam001	10.0	7.8	2.2
Cow003	06/10/2000	01/01/2000	Sire005	Dam007	5.0	8.5	-3.5

DOB - Date of birth; DOC - Date of conception; EBV – Estimated Breeding Value; GSD – Genetic Selection Differential

The mean GSD for milk yield, fat yield, protein yield and LPI were computed for each year of conception from 1980 to current year, 2018 for the entire population.

Benchmarks

Appropriate benchmarks have been calculated against which the mean GSD of individual herds would compare with. These benchmarks include the mean GSD of the population, the mean GSD of the top and bottom 10% of sires and of herds and the top 10% potential sires. These benchmark GSDs per year of conception were stored in databases. Each of these GSD for the benchmarks will also be re-computed each time new genetic evaluations are released. The mean GSDs for each year of conception were computed for each herd or farm in the data set with at least 10 years' records during the period 1980 to date.

All results of mean GSD per years of conception were displayed in graphs using SAS for visual display.

Utilization and user-interface

After delivering all the information generated as part of the development of the software to a web-server, individual producers who are end-users will benefit in a number of ways. When producers sign in with their unique passwords, producers can select their herds' mean GSD for all years (1980 to date) and compare to the number of benchmarks (population average, top and bottom 10% herds, top and bottom 10% sires, etc) calculated. The mean GSD of a producer can either be visualized in graphs or tables. Producers can save the information of the mean GSD of their herds for subsequent use. End users (producers) can also decide to specify the years they are interested, for example the mean GSD of a producer's herd for the past 10 years.

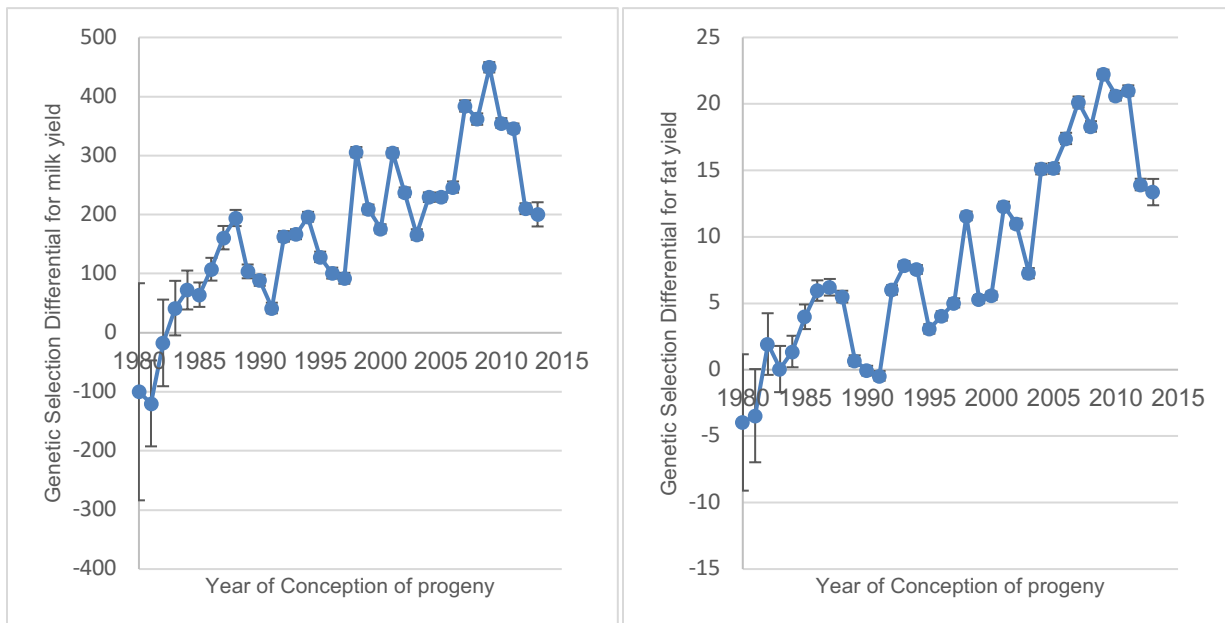
Another benefit producer stands to gain from this software is the ability to select the GSD of any of the traits in his/her breeding objectives. This will give a producer the opportunity to monitor selection pressure of a number of traits within or outside his/her breeding objectives.

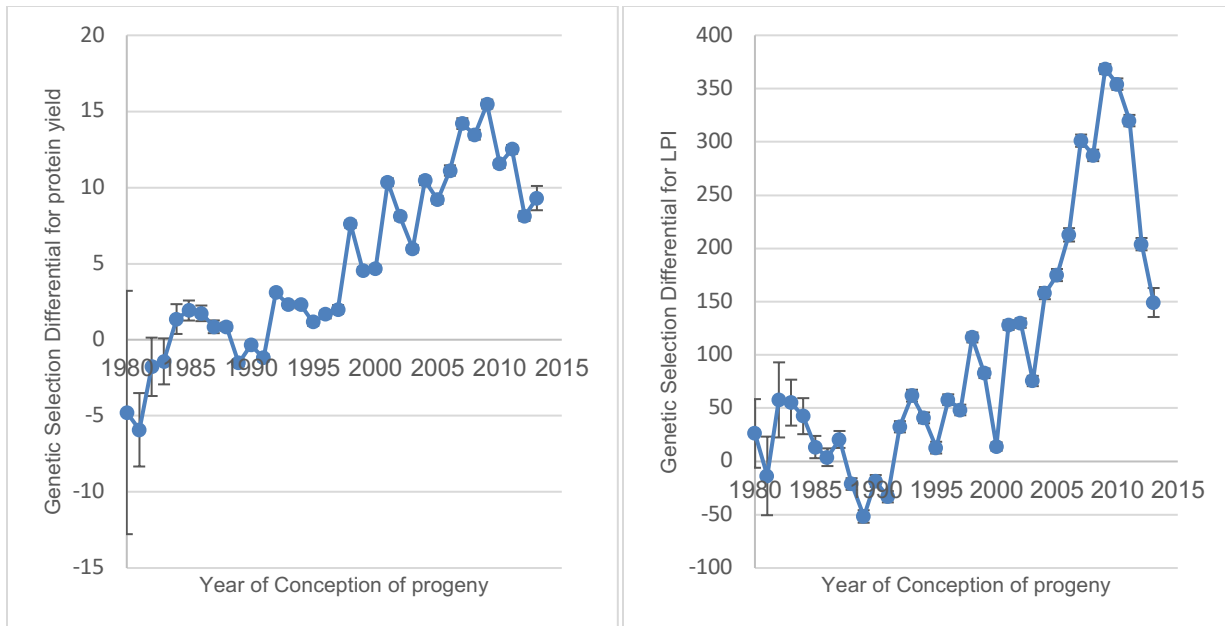
Producers can again have access to information on the GSD for both the sire-cow and dam-cow pathways for all traits captured in the software. Lastly, producers have information on the standard errors for the mean GSD in their herds for each year. This will give them information of the variability of sires they are using for breeding each year.

Results and discussion

Benchmarks

For precision dairy production, it is important for dairy producers to have knowledge of past selection applied on economic important traits so that these producers can make changes if they are moving in an unintended direction. The dairy industry on the whole also needs this knowledge of past selection applied so that it can make changes in selection emphasis when new traits gain prominence in their contribution to the profitability of the dairy production business. A prototype precision agriculture software has been developed to allow dairy producers monitor past selection applied and current selection being applied in their herds. The mean GSD per year of conception for milk yield, fat yield, protein yield and LPI for sires of cows are shown in figures 2, 3, 4 and 5 respectively. These are population benchmarks. They are useful for both describing selection applied in the population in the past and suitable benchmarks individual producers can compare with. The population mean GSD for milk, fat and protein yields generally showed increasing trends with occasional drops before peaking in 2009 after which there were declines.

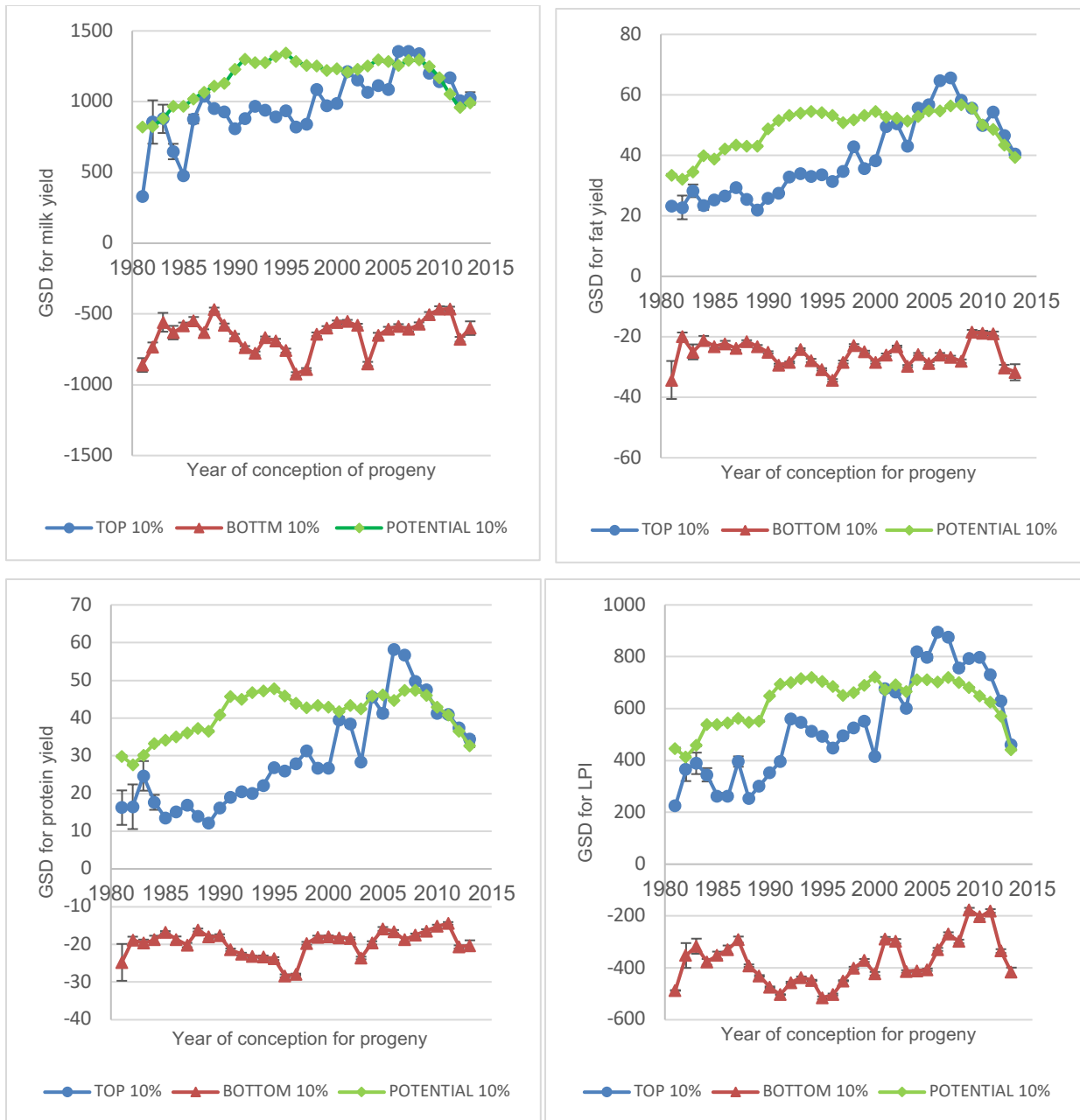




Figs. 2, 3, 4 and 5. Population mean genetic selection differentials for milk, fat, protein yields and LPI by year of conception for the Ayrshire sires between 1980 and 2013

The selection pattern for LPI generally started on a downward trend (from 1982 to 1989) before increasing and peaking in 2009. The similarity in selection patterns for milk, fat and protein yields in the Ayrshire populations is partly attributed to the strong positive genetic correlations (0.71-0.93) among these traits (Welper & Freeman 1992; Dematawewa & Berger 1998). The selection pattern for LPI was slightly different from the yield traits especially in the initial years (between 1981 and 1989). This is due to the fact that LPI was not in existence until after 1990 hence farmers were not selecting on this index during that period. After 1990, the pattern of selection for LPI in the Ayrshire population remained similar to those of the production traits due to the heavy weight the Ayrshire LPI formula placed on production, especially between 2007 and 2017 (50 to 54%). Also, the weights of the production traits in the Ayrshire LPI index formula have remained stable between 2007 and 2014 resulting in the similar patterns observed among the production traits and LPI. The standard error bars in the initial years were larger, suggesting fewer sires were used for breeding in the population.

Another group of benchmarks computed and stored in databases were the mean GSDs for the top and bottom 10% sires for milk, fat, protein yields and LPI shown graphically in figures 6, 7, 8 and 9. Yet another benchmark for comparison were the mean GSDs for the potential top 10% sires per year of conception.



Figs. 6, 7, 8 and 9. Mean GSD for top and bottom 10% sires per year of conception and the top 10% potential sires for milk, fat, protein yields and LPI for the Ayrshire population

To precisely monitor whether the dairy industry has maximised selection for a given trait in any given year, the ratio of the GSD of the top 10% potential sires to the top 10% sires can be monitored. A ratio of less than 1 is an indication that selection applied in the population had not been maximised. The GSD of the top 10% sires used were lower than the top 10% potential sires available prior to the year 2000. This was an indication that within that period, selection applied had not been maximised. After 2000, it appears the GSD of the top 10% sires were similar to the possible top 10% potential sires available. This suggests that in recent times, producers are utilizing the top sires available for the traits under consideration in servicing their cows. The mean GSD of the top 10% of sires used in the population between 1980 and 2013 were between 2.4 - 4.2 and 23.1 – 164.4 times higher than the population average for milk, fat, protein and LPI. This means that producers using the average sires for breeding each year can increase their selection pressure substantially by selecting sires from the top 10%. The mean GSD of the top 10% of sires each year appears to be increasing for all traits until after 2006/07 where there was an apparent decline. On the other hand, the mean GSD of the bottom 10% of sires remained fairly stable.

Packaging these analyses into monitoring tools where outputs can be visualized in graphs could aid to describe the population dynamics in terms of selection directions over the years and make the necessary changes if needed.

The increasing trends in the population mean GSD (figs. 2, 3, 4 and 5) are encouraging, but are all herds driving these increases? The software tool detailed here will enable individual producers to monitor their herd's performances relative to the population average or other suitable benchmarks produced. To illustrate the use of this precision animal improvement software, two herds were selected from the Ayrshire population and their individual selection patterns over the years were compared to those of the population average and top and bottom 10% herds. Individual farmers could pull out their herd's performance at any given time to see what they have been doing or selecting for.

In figure 10, it can be observed that the selection pressure of herd A for milk yield was similar to that of the population mean whilst that of herd B was consistently below the population mean except for 1983 and 1985. The mean GSDs of herd B were most of the years within the bottom 10% of herds. The standard error bars in the graphs give an indication of the variability in sire usage. Generally, there seems to be greater variability in the choice of sires used in herd A compared to herd B except for years 1993 to 1996 where herd A was relatively less variable in its choice of sires for breeding. For instance, in 1994 herd A used 10 different sires for the 19 matings made in the herd with semen from one particular bull being used for inseminating 8 different replacement cows. On the other hand, 11 different sires were used for 18 matings in herd B during the same period.

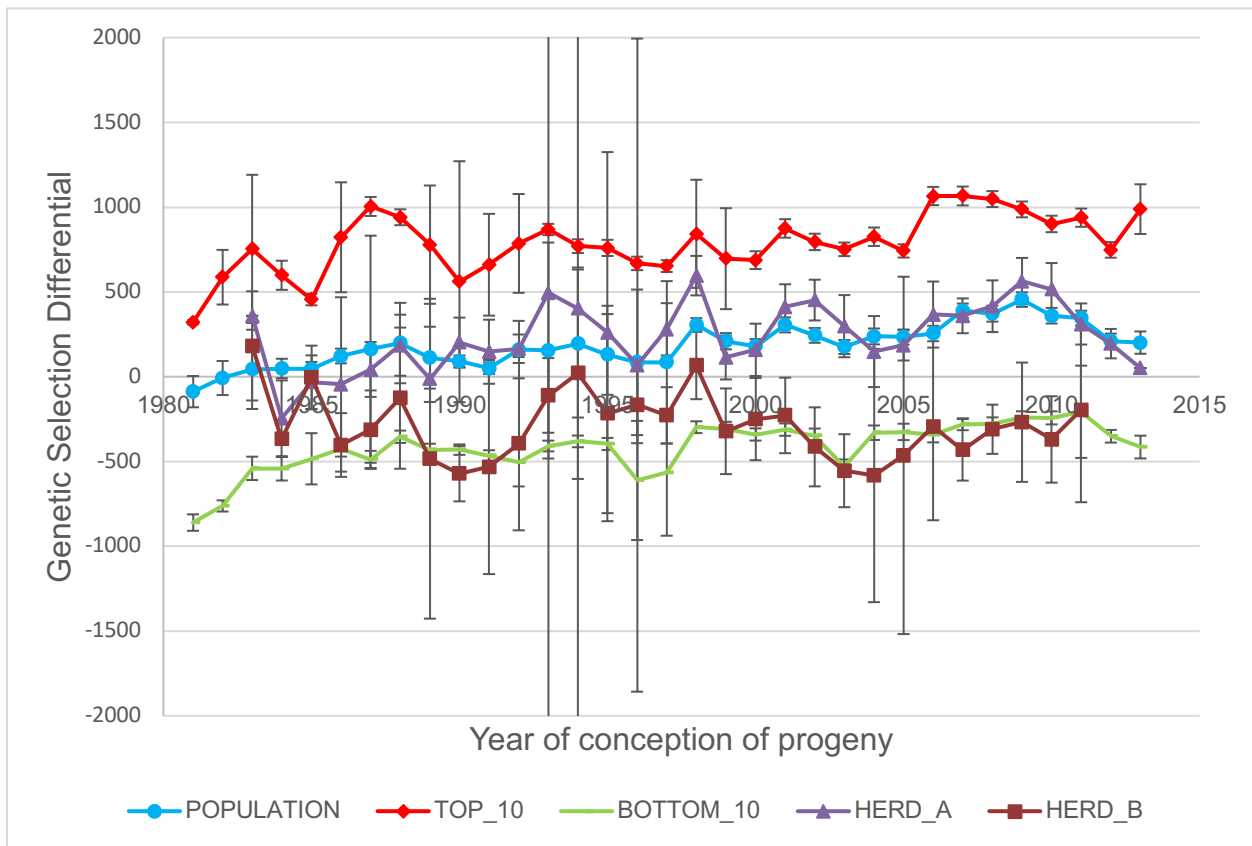


Fig. 10. Mean GSD for population, top and bottom 10% herds and herds A and B by year of conception of their progeny for milk yield

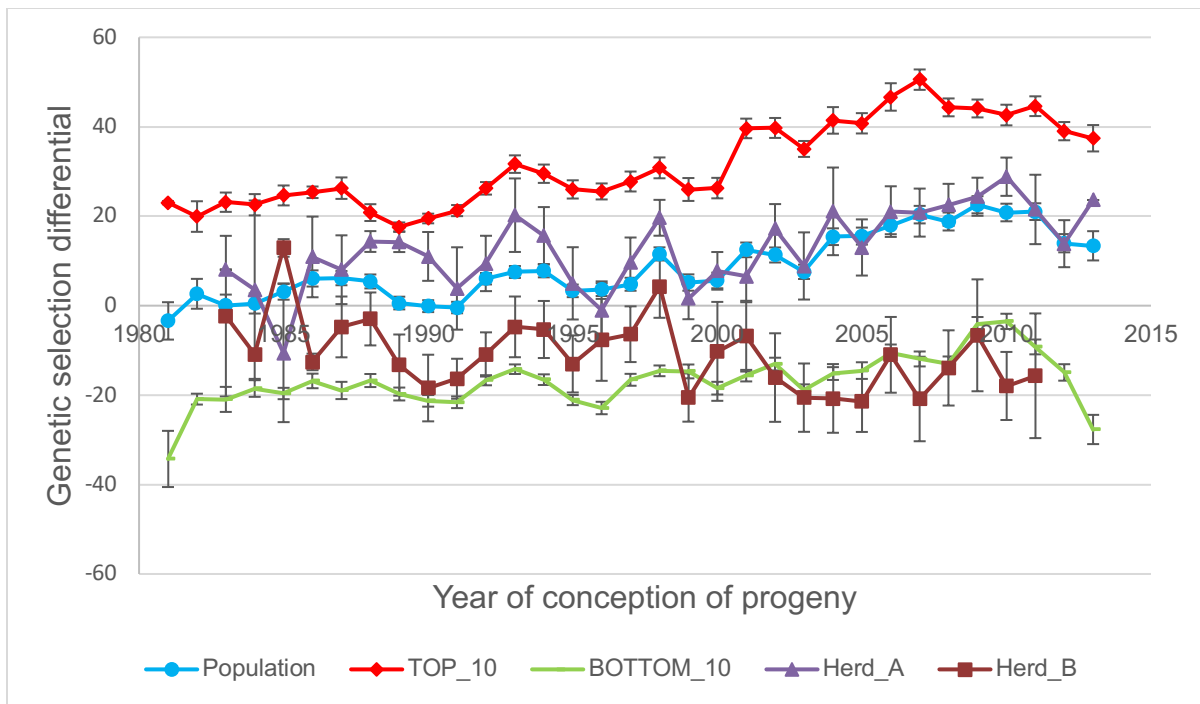


Figure 11: Mean GSD for population, top and bottom 10% herds and herds A and B by year of conception of their progeny for fat yield

The selection pressures for fat yield and protein yields (figs. 11 and 12) were similar to those of milk yield. This can be attributed to the strong positive genetic correlations among milk, fat and protein yields (Welper & Freeman 1992; Dematawewa & Berger 1998); hence selection for one of these traits implies indirect selection for the other traits.

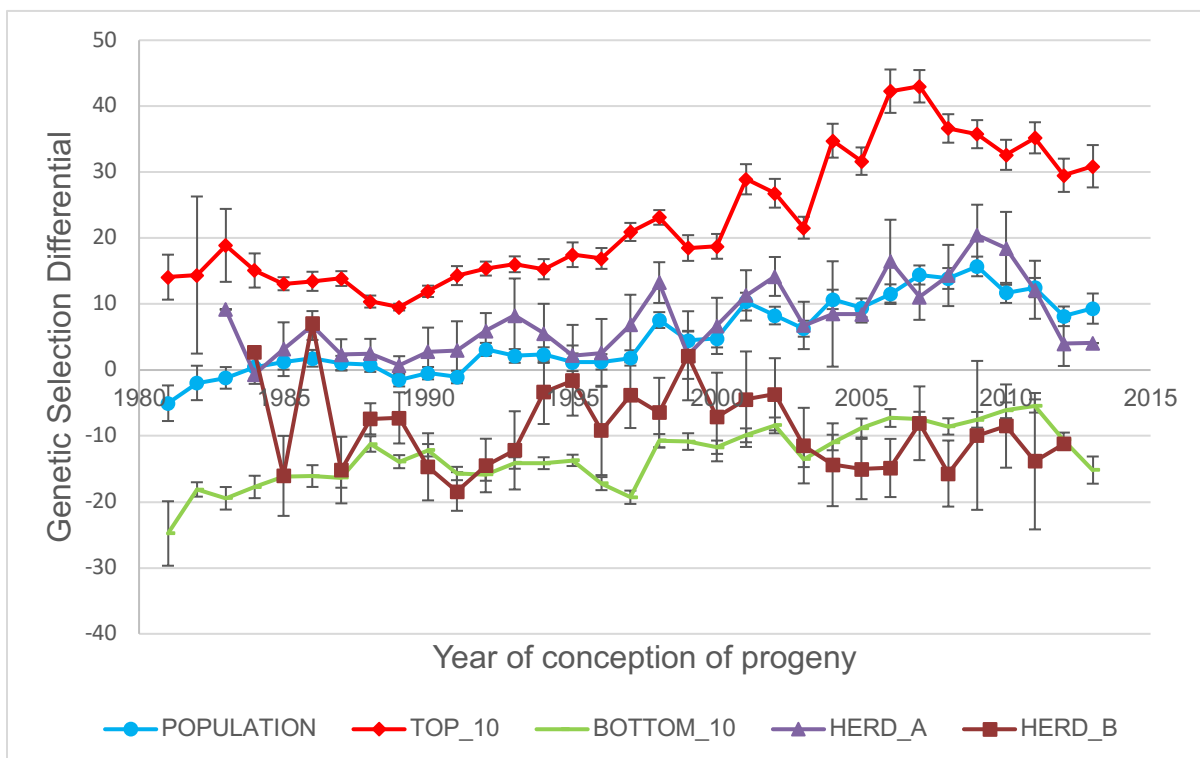


Figure 12: Mean GSD for population, top and bottom 10% herds and herds A and B by year of conception of their progeny for protein yield

For LPI, the selection pressures for herds A and B were similar to that of the population mean between 1980 and 2000 (fig. 13). Whilst the selection pressure of herd A remained similar to the population average even after 2000, that of herd B began to decline. What is not clear is whether the farmer of herd B is aware his selection pressure on LPI has been relaxed or whether this was a conscious, deliberate decision. When the software tool is made available to dairy producers to monitor on a regular basis their selection practices, it will offer producers the opportunity to make changes when they are moving in an unintended direction. With the use of this precision agriculture tool, producers can determine and change the direction of selection applied in their herds to meet more precisely their breeding objectives. This will indirectly result in increased farm profit as genetic selection results in genetic gain which is associated with profit (Thompson *et al.* 2015).

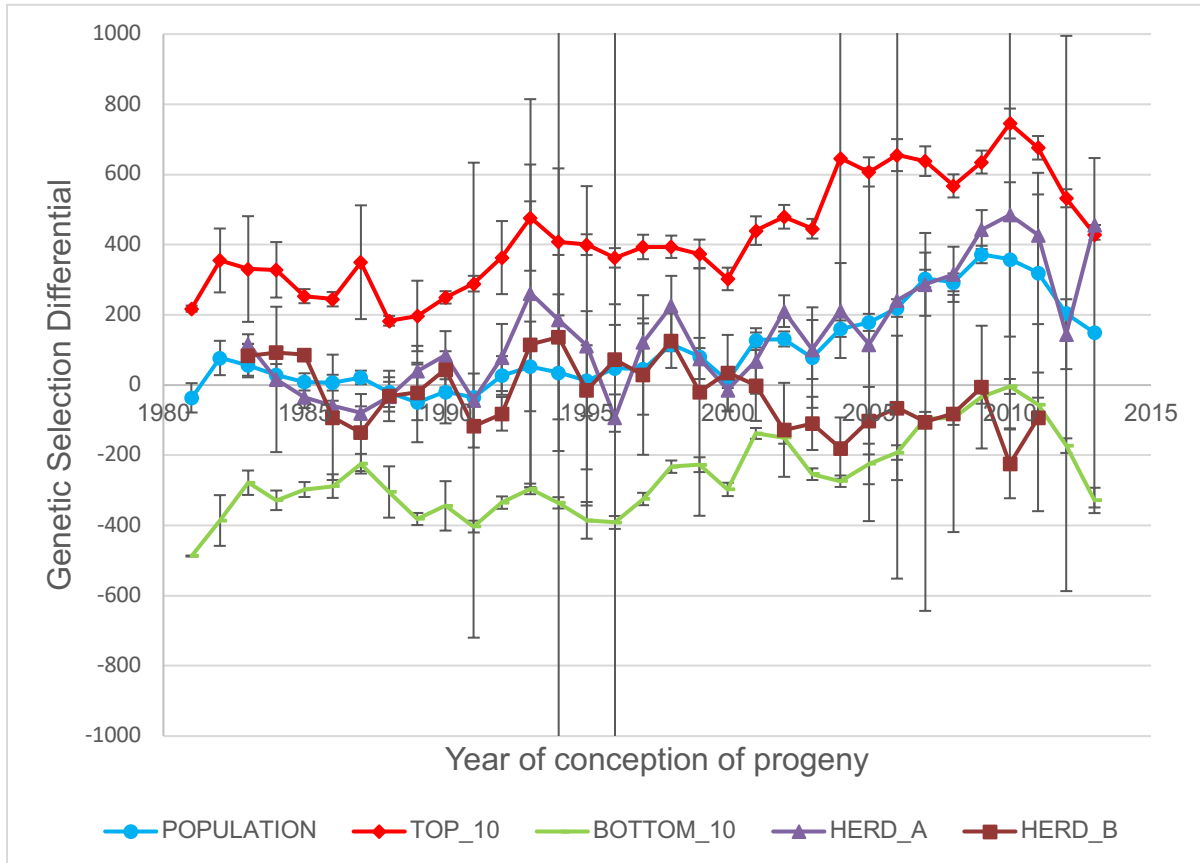


Figure 13: Mean GSD for population, top and bottom 10% herds and herds A and B by year of conception of their progeny for LPI

Results by (Brown & Cue 1992) indicated that there were significant differences between herd types (owner sampler vs officially supervised dairy recorded herds) in their selection for sires, hence further benchmarks of the mean GSD for both herds types, tie stalls and free stalls and the top and bottom 10% herds within each type would be established. This will increase the number of benchmarks for which dairy farmers could properly compare their herds with in order to have a more appropriate assessment of their herds' performance.

Conclusion

A prototype software tool has been developed for monitoring genetic selection differential applied by individual farmers over the years. With this tool, dairy farmers have the opportunity of monitoring their selection practices over the years and making changes to their selection decisions if producers are not moving in the right direction. Dairy producers will also have an idea

of the best bulls and cows available in terms of bulls/cows' genetic superiority as compared to their available contemporaries in the short term (at least for the next 4 months before the release of new genetic evaluations) hence dairy producers can choose the best bulls available to maximise the genetic gains in the producers' herds. Though this prototype has been developed using the Ayrshire dairy breed, it is equally applicable to other dairy breeds, such as the Holstein, Brown Swiss and Jersey, and other livestock breeds and species for which genetic evaluation are routinely calculated (e.g. beef cattle, swine, sheep, goats). Similarly, while the benchmarks shown here are the population averages, top and bottom 10% of herds and top and bottom 10% sires, the concepts can easily be adapted to use other appropriate benchmarks, e.g. type of housing system (Free-Stall or Tie-Stall), or agricultural region where the farm is located, or production system (Conventional or Organic), or other user-defined benchmarks. Genetic improvement in the dairy cattle involves four pathways of selection; sire-bull, dam-bull, sire-cow and dam-cow paths. Though this paper has demonstrated development of a tool for monitoring selection in only the sire-cow pathway, the software tool includes all the four pathways and 28 traits producers select for. The tool will therefore be useful to both dairy producers and artificial insemination (AI) studs who together control these four paths of selection. The prototype software has been developed using SAS (for rapidity and convenience in a University research environment); the prototype software is being translated to an Open Source tool (R, Linux system tools, Fortran), for portability.

Acknowledgement

The authors thank Dr. Filippo Miglior and the Canadian Dairy Network for providing the genetic evaluation data for this work, Valacta for the ancillary herd information data, and the Ghana Education Trust Fund (GETFund) for financial support of the first author.

References

- Brown T.P. & Cue R.I. (1992) Differences in production and type traits selection differentials between bulls used in official and owner/sampler dairy herds. *Canadian Journal of Animal Science* 72, 783-9.
- Burnside E.B., Jansen G.B., Civati G. & Dadati E. (1992) Observed and theoretical genetic trends in a large dairy population under intensive selection. *Journal of Dairy Science* 75, 2242-53.
- Dematawewa C. & Berger P. (1998) Genetic and Phenotypic Parameters for 305-Day Yield, Fertility, and Survival in Holsteins¹. *Journal of Dairy Science* 81, 2700-9.
- García-Ruiz A., Cole J.B., VanRaden P.M., Wiggans G.R., Ruiz-López F.J. & Van Tassell C.P. (2016) Changes in genetic selection differentials and generation intervals in US dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences*.
- Nizamani A. & Berger P. (1996) Estimates of Genetic Trend for Yield Traits of the Registered Jersey Population¹. *Journal of Dairy Science* 79, 487-94.
- SAS Institute Inc., SAS 9.4 Help and Documentation, Cary, NC: SAS Institute Inc., 2016.
- Thompson B., Stevens D., Bywater A., Rendel J. & Cox N. (2015) Impacts of animal genetic gain on the profitability of three different grassland farming systems producing red meat. *Agricultural Systems* 141, 36-47.
- Van Doormaal B. (2007) Genetic Evaluation of Dairy Cattle in Canada
- Van Tassell C.P. & Van Vleck L.D. (1991) Estimates of genetic selection differentials and generation intervals for four paths of selection. *Journal of Dairy Science* 74, 1078-86.
- Welper R. & Freeman A. (1992) Genetic Parameters for Yield Traits of Holsteins, Including Lactose and Somatic Cell Score¹. *Journal of Dairy Science* 75, 1342-8.