MTEDC revision history (Version 6b)

Generalized Multiple-Trait Software for EDC of Sires and Reliabilities of Animals

Pete Sullivan, Lactanet Canada

Summary of the Purposes of the Program:

1. Improve ***harmonization of EDC in MACE***, by offering a free software solution to all participating countries, based on the generalized methods described in:

   Sullivan (2007) Interbull Bulletin 37:78-81.
   MTEDC Software Available for Standardized EDC Calculations
   Sullivan, Liu, Jakobsen & Fikse (2006) Interbull Bulletin 35:112-116.
   More on Weighting Factors for Complicated Models

2. Compute EDC of sires to submit to Interbull for MACE, allowing for ***Multiple-trait*** national genetic evaluation models with any number of traits, and including any combination of ***direct*** and ***maternal*** genetic and environmental effects per trait.
3. Approximate ***national reliabilities*** for all evaluated animals, also for the generalized multiple-trait, linear mixed model, based on pedigree relationships among animals.
4. Use the ***same input files for EDC or for reliabilities***, to simplify and optimize alignment between national reliabilities generated by this software, and international reliabilities published by Interbull using EDC generated by this software.
5. Provide a simple and generalized solution for approximating pedigree-based ***reliabilities of residual polygenic*** contributions in genomic evaluation models.

The software is freely available without warranty at https://www.cdn.ca/software/mtedc.html

The most recent version and a limited number of outdated versions are available at the above URL, for comparative purposes and as alternative options in case you have any problems using the latest version with your data sets and models.

A general description of all updates added within each new version is provided below.

Revision history by Version:

6b) November 2023 : Enhancements added since version 6a:

- Updated -w ITB to match PGS1, fixing the negative EDC problem with repeated records.
- Improved code for re-direction of screen messages with -o, FSTDOUT and FSTDERR.
    - And now use --SCREEN, --CONSOLE or --STDOUT to disable the re-directions.
- Expanded the on-screen help message generated with -h.
- Added some long command-line options:
    - --HELP and --USAGE, which are equivalent to -h
    - --OPTION, which is equivalent to -O
    - --DRYRUN, a new option to quickly list all input and output files to be processed, without creating any new files, and without deleting or modifying existing files.
- Corrected the grand-progeny phenotype counts (-G) for models with maternal effects.
- Allowed Indexes with zero variance (e.g. if Gd=0 for all trait(s) in the index).
- Removed white space in variable-length delimited output files (-d).
- Allowed negative integer coding for genetic group IDs.
    - Allows use of a single input pedigree file for EBV, EDC and reliability calculations.
    - The contributions of genetic grouping to reliability are assumed equal to zero, as was always the case.

6a) June 2022 : Enhancements added since version 5g:

- Bundled the code for approximating reliabilities into the free version of MTEDC, which ***eliminates the requirement to purchase*** an add-on module to approximate reliabilities.
- Updated the heritability calculations for Sire/MGS models with maternal effects:
    - Details by ***Sullivan and Schuler (2022)*** are provided in the Appendix below.

5g) July 2020 : Enhancements added since version 5f:

- Allowed models with different combinations of Genetic versus permanent Environment effects of animals (e.g. Gm=0 & Em>0, Gd=0 & Ed>0).
    - Any variance can now be set to zero within the provided G and E matrices.
    - The zero-variance effects are excluded from computations to reduce memory, run time and approximation errors.

5f) July 2020 : Enhancements added since version 5e:

- Removed the requirement to renumber animals sequentially. Any integer numbered sequence uses similar RAM and gives identical results.
- Allowed an unlimited number of mates per sire.
- Improved the handling of ET records for multiple-trait models with direct and maternal effects, when approximating reliabilities.
- Allowed an unlimited number of fields in the pedigree file, so that any additional information of interest can be added, for example the animal's name, registration, etc.
- The program now runs without errors if zero indexes are requested.
- Removed white space when delimited output is requested (-d).

- New option to request r_{TI} instead of r_{TI}^2 in the output file.
- New option to process only a subset of the traits, to help isolate problem trait(s) in large multi-trait systems.
- Progeny counts for indexes are now weighted sums, by index weights, of the (grand)progeny observations.

5e) September 2018 : Enhancements added since version 5d:

- Fixed bugs from 5d for models with multiple traits having direct and maternal effects, and with non-zero residual correlations.
- Improved handling of zero genetic variances for some traits (e.g. if Gd and Em are non-zero, but Gm is zero)

5d) June 2010 : Enhancements added since version 5c:

- Improved direct-by-maternal adjustments in multi-trait models.
- Improved adjustments for fixed effects in multi-trait models.
- New option to include a header row if choosing comma-separated output.

5c) December 2009 : Enhancements added since version 5b:

- Improved adjustment for fixed effects in direct + maternal effects models.
- Improved check for positive semi-definite input matrices.

5b) September 2009 : Enhancements added since version 5a:

- Allowed models with both sire and maternal grandsire incidence effects.
- Allowed permanent environmental effects for animals or dams.
- Added a requirement and check that all input matrices are positive semi-definite.
- New checks for 1-based animal IDs in the incidence file (i.e. 0 is not a valid animal).
- New checks for reasonable record weights (maximum weight = 3.0).
- New checks for reasonable coding of contemporary groups, based on group size.
- Validating all EDC are within the parameter space, and if not, no output is generated.
- Describing output file contents in the LOG file.
- New option to generate comma-separated output.
- New options to ignore some of the above checks.
- New option to define heritability within or across CG, if CG effects are random.

5a) June 2008 : Enhancements added since version 5 (released in March 2008):

- New option to include grand-progeny record counts in the output file.
- Avoiding segmentation faults due to large contemporary groups from mis-coded data.
- Reduced RAM and run time for large multi-trait models.
- Expanded the memory monitoring messages, because the original Out-of-RAM detection did not work well with 64-bit systems that allocate beyond currently available RAM if more RAM might become available later.
- Extended the Animal Model improvements from older versions to Sire Models.

Appendix: Maternal parameters derived from sire - maternal grandsire models

Pete Sullivan, Lactanet Canada
Urs Schuler, Qualitas, Switzerland
March, 2022

Direct and maternal genetic effects can be fitted to a trait by including animal and dam effects in the model, but it is sometimes difficult to apply animal-dam models, for example with survival or threshold traits. An easier alternative is to fit sire and maternal grandsire (mgs) effects, but in these models the direct and maternal genetic effects are not directly available from the estimates. The purpose of this note is to derive direct and maternal genetic effects and covariances, from the estimates and covariances for sire and mgs effects, and to estimate the direct and maternal heritabilities, and the direct-maternal genetic correlation.

Expectations for the variances and covariances between different individuals can be expressed in terms of direct and maternal covariances (Willham, 1971). More specifically, these expectations are shown for sire and mgs of phenotyped individuals in Kriese et al (1999). The expected covariances can be described by the matrix equation **a=Lb**, where:

$$\mathbf{a} = \begin{bmatrix} \boldsymbol{V(sire)} \\ \boldsymbol{C(sire, mgs)} \\ \boldsymbol{V(mgs)} \end{bmatrix}, \qquad \mathbf{L} = \begin{bmatrix} 1/4 & 0 & 0 \\ 1/8 & 1/4 & 0 \\ 1/16 & 1/4 & 1/4 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} \boldsymbol{V(direct)} \\ \boldsymbol{C(direct, maternal)} \\ \boldsymbol{V(maternal)} \end{bmatrix}$$

Matrix L is positive-definite with the following inverse:

$$\mathbf{L}^{-1} = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 4 & 0 \\ 1 & -4 & 4 \end{bmatrix}$$

It follows that $\mathbf{b=L^{-1}a}$, and this is the same set of equations used to derive elements of vector **b** in several previous studies (Thompson et al, 1981; Dwyer et al, 1986; Luo et al, 1999; Wiggans et al, 2003; Steinbock et al, 2003; Hansen et al, 2004; Heringstad et al, 2007; Eaglen and Bijma, 2009).

These same matrices can also be derived as follows, where sire and mgs effects (**e**) are expressed relative to their direct and maternal genetic contributions (**f**), such that $\mathbf{e=Kf}$ and $\mathbf{f=K^{-1}e}$:

$$\mathbf{e} = \begin{bmatrix} Sire \\ Mgs \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} direct \\ maternal \end{bmatrix} \quad : \quad \mathbf{f} = \begin{bmatrix} direct \\ maternal \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} Sire \\ Mgs \end{bmatrix}$$

Now $\mathbf{V(f)= K^{-1}V(e)(K^{-1})'}$:

$$\begin{bmatrix} \sigma_d^2 & \sigma_{dm} \\ \sigma_{dm} & \sigma_m^2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \sigma_S^2 & \sigma_{SM} \\ \sigma_{SM} & \sigma_M^2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix}$$

By multiplying this out, it is easily verified that:

$$\sigma_d^2 = 4\sigma_S^2, \quad \sigma_{dm} = -2\sigma_S^2 + 4\sigma_{SM}, \text{ and } \sigma_m^2 = \sigma_S^2 - 4\sigma_{SM} + 4\sigma_M^2$$

These are the same 3 equations as in $\mathbf{L^{-1}a}$ above.

Direct and maternal heritabilities can be derived from the covariances among sire and mgs effects. These heritabilities are the proportions of phenotypic variance in b_1=V(direct) and b_3=V(maternal). Having solved for vector **b**, we now need an estimate for the phenotypic variance V(y), but this has been inconsistently defined in the literature. In particular, the contribution of C(sire,mgs) to V(y) has ranged for 0 to twice its value. We will now clarify what this contribution should be.

The genetic effects (g) included in an animal - dam model include one direct effect (a=animal) and one maternal effect (d=dam):

$$g = D_a + M_d$$

These direct and maternal effects can be partitioned into parental contributions and Mendelian sampling terms (*):

$$g = ½\ D_s + ½\ D_d + D_{a*} + ½\ M_{mgs} + ½\ M_{mgd} + M_{d*}$$

The term D_d can also be partitioned in the same way:

$$g = ½\ D_s + ½\ (½\ D_{mgs} + ½\ D_{mgd} + D_{d*}) + D_{a*} + ½\ M_{mgs} + ½\ M_{mgd} + M_{d*}$$

The sire-mgs model includes only three of the above eight terms, leaving the remaining 5 terms as part of the model residual. We demonstrate by re-arranging the above terms and adding parentheses to separate the modeled genetic effects, followed by the genetic terms included in the residual:

$$g = (½\ D_s + ¼\ D_{mgs} + ½\ M_{mgs}) + (¼\ D_{mgd} + ½\ D_{d*} + D_{a*} + ½\ M_{mgd} + M_{d*}) = (g_u) + (g_E)$$

This equation can be expressed using matrices $\mathbf{g}=\mathbf{H}\boldsymbol{\theta}$, ordered by genetic effects within animal:

$$\mathbf{g} = \begin{bmatrix} g_u \\ g_E \end{bmatrix} = \begin{bmatrix} H_u \\ H_E \end{bmatrix} \boldsymbol{\theta} = \left[\begin{array}{cc|cc|cc|cc|cc} 0.5 & 0 & 0.25 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0.5 & 1 & 0 & 0.5 & 1 \end{array}\right] \begin{bmatrix} D_s \\ M_s \\ D_{mgs} \\ M_{mgs} \\ D_{mgd} \\ M_{mgd} \\ D_{a*} \\ M_{a*} \\ D_{d*} \\ M_{d*} \end{bmatrix}$$

The total variance of genetic effects contributing to V(y) is $\boldsymbol{V_{H\theta}} = H_u\boldsymbol{\theta}(H_u)' + H_E\boldsymbol{\theta}(H_E)'$, and the latter term is not required specifically, because it is included in the estimated residual variance from the sire-mgs model, which is V(E+e), but we do expand it below to compare with the animal-dam model. If the sire, mgs and mgd of the phenotyped individual are unrelated, then these terms are computed as:

$$H_u\boldsymbol{V_\theta}(H_u)' = \tfrac{1}{4}\mathrm{V}(D_s) + \tfrac{1}{16}\mathrm{V}(D_{mgs}) + \tfrac{2}{8}\mathrm{C}(D_{mgs}, M_{mgs}) + \tfrac{1}{4}\mathrm{V}(M_{mgs})$$

$$H_E\boldsymbol{V_\theta}(H_E)' = \tfrac{1}{16}\mathrm{V}(D_{mgd}) + \tfrac{2}{8}\mathrm{C}(D_{mgd}, M_{mgd}) + \tfrac{1}{4}\mathrm{V}(M_{mgd}) + V(D_{a*}) + \tfrac{1}{4}\mathrm{V}(D_{d*}) + \tfrac{2}{2}\mathrm{C}(D_{d*}, M_{d*}) + V(M_{d*})$$

If we further ignore inbreeding of these ancestors, we can substitute:

$$V(D_i) = \sigma_d^2,\ V(M_i) = \sigma_m^2,\ C(D_i, M_i) = \sigma_{dm},\ and\ V[D_{i*} \quad M_{i*}] = \tfrac{1}{2}V[D_i \quad M_i] = \tfrac{1}{2}\mathbf{G},$$

$$\text{such that } H_u\boldsymbol{V_\theta}(H_u)' = (\tfrac{1}{4}+\tfrac{1}{16})\sigma_d^2 + (\tfrac{1}{4})\sigma_{dm} + (\tfrac{1}{4})\sigma_m^2 \text{ and}$$

$$H_u\boldsymbol{V_\theta}(H_u)' + H_E\boldsymbol{V_\theta}(H_E)' = (\tfrac{1}{4}+\tfrac{1}{16}+\tfrac{1}{16}+\tfrac{1}{2}+\tfrac{1}{8})\sigma_d^2 + (\tfrac{1}{4}+\tfrac{1}{4}+\tfrac{1}{2})\sigma_{dm} + (\tfrac{1}{4}+\tfrac{1}{4}+\tfrac{1}{2})\sigma_m^2 = \sigma_d^2 + \sigma_{dm} + \sigma_m^2$$

This last equation is the same expectation as from an animal-dam model, where the relationship between animal and dam is $a_{ad} = \frac{1}{2}$. To be more precise, the above expansion assumes that:

$$\boldsymbol{V_\theta} = \begin{bmatrix} \mathbf{I_3} \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{I_2} \otimes \mathbf{G} \end{bmatrix} \quad \textbf{... [1]}$$

In the presence of inbreeding, and additionally if sire and mgs are related, then the following is more correct, but I believe it is rarely used to derive heritabilities:

$$\boldsymbol{V_\theta} = \begin{bmatrix} \mathbf{A_3} \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{D_2} \otimes \mathbf{G} \end{bmatrix} \quad \textbf{... [2]}$$

The $\mathbf{A_3}$ is a matrix of additive genetic relationships among the sire, mgs and mgd, and $\mathbf{D_2}$ is a diagonal matrix accounting for the impact of inbred parents of the phenotyped animal and the dam, on the Mendelian sampling variances of the animal and the dam. For an expression of average heritability for a population, average matrices could be used in [2], but this is only needed for $\overline{\mathbf{A_3}}$, since $\overline{\mathbf{D_2}}$ is for genetic contributions that do not need to be partitioned from V(E+e).

When estimating V(y) from the sire-mgs model, we need to consider that V(E+e) includes some of the genetic variances and covariances between direct and maternal effects, and we therefore add only the genetic portion of variance explained by the sire and mgs, which is $H_u\boldsymbol{V_\theta}(H_u)'$. Thus, if we are using [1], with the assumption that $a_{s,mgs} = 0$:

$$V(y) = V(sire) + V(mgs) + V(E+e) = \tfrac{5}{16}\sigma_d^2 + \tfrac{1}{4}\sigma_{dm} + \tfrac{1}{4}\sigma_m^2 + V(E+e) \quad \text{... [3]}$$

If we are using [2], allowing for $a_{s,mgs} > 0$, but ignoring inbreeding in $\mathbf{A_3}$:

$$V(y) = V(sire) + V(mgs) + 2 * a_{12}C(sire, mgs) + V(E+e) \quad \text{... [4]}$$

To additionally consider inbreeding, it might be easiest to set up the upper 4x4 block of $\boldsymbol{V_\theta}$, for an average sire-mgs combination, based on [2], and then:

$$V(y) = H_u\boldsymbol{V_\theta}(H_u)' + V(E+e) \quad \text{... [5]}$$

Any additional random effects in the model (e.g. random contemporary groups, etc) would also be included in V(y), but these do not affect derivation of the estimated genetic variances described above.

Dwyer, D.J., L.R. Schaeffer and B.W.Kennedy. 1986. Bias due to corrective matings in sire evaluations for calving ease. J. Dairy Sci. 69:794-799. https://www.journalofdairyscience.org/article/S0022-0302(86)80468-4/pdf

Eaglen, S.A.E. and P. Bijma. 2009. Genetic parameters of direct and maternal effects for calving ease in Dutch Holstein-Friesian cattle. J. Dairy Sci. 92:2229-2237. https://doi.org/10.3168/jds.2008-1654

Hansen, M., M.S. Lund, J. Pedersen and L.G. Christensen. 2004. Genetic parameters for stillbirth in Danish Holstein cows using a Bayesian threshold model. J. Dairy Sci. 87:706-716. https://doi.org/10.3168/jds.S0022-0302(04)73214-2

Heringstad, B., Y.M. Chang, M.Svendsen and D. Gianola. 2007. Genetic analysis of calving difficulty and stillbirth in Norwegian red cows. J. Dairy Sci. 90:3500-3507. https://doi.org/10.3168/jds.2006-792

Kriese, L.A, J.K. Bertrand and L. Benyshek. 1991. Age adjustment factors, heritabilities and genetic correlations for scrotal circumference and related growth traits in Hereford and Brangus bulls. J. Anim. Sci. 69:468-489. https://doi.org/10.2527/1991.692478x

Luo, M.F., P.J. Boettcher, J.C.M. Dekkers and L.R. Schaeffer. 1999. Bayesian analysis for estimation of genetic parameters of calving ease and stillbirth for Canadian Holsteins. J. Dairy Sci. 82:August. (on-line only). https://doi.org/10.3168/jds.S0022-0302(99)75416-0

Steinbock, L., A. Näsholm, B. Berglund, K. Johansson and J. Philipsson. 2003. Genetic effects on stillbirth and calving difficulty in Swedish Holsteins at first and second calving. J. Dairy Sci. 86:2228-2235. https://doi.org/10.3168/jds.S0022-0302(03)73813-2

Thompson, J.R., A.E. Freeman and P.J. Berger. 1981. Age of dam and maternal effects for dystocia in Holsteins. J. Dairy Sci. 64:1603-1609. https://doi.org/10.3168/jds.S0022-0302(81)82732-4

Willham, R.L. 1971. The role of maternal effects in animal breeding: III, Biometrical aspects of maternal effects in animals. Presented at ASAS symposium on “The Role of Maternal Effects in Animal Breeding”, California, August 3. https://doi.org/10.2 527/jas1972.3561288x

Wiggans, G.R., I. Misztal and C.P. Van Tassell. 2003. Calving ease (co)variance components for a sire-maternal grandsire threshold model. J. Dairy Sci. 86:1845-1848. https://doi.org/10.3168/jds.S0022-0302(03)73771-0