

# MTEDC user manual (Version 6b)

---

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

Pete Sullivan, Lactanet Canada

---

### 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 multiple-trait, linear mixed model.

*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 submitted to Interbull for MACE, allowing for:
  - **Multiple-trait** national genetic evaluation models with any number of traits.
  - Traits with direct only or **direct plus maternal** genetic and/or environmental effects.
  - **Animal or sire** genetic evaluation models.
  - Index traits that are **linear functions** of the evaluated genetic effects and traits.  
Example index traits commonly submitted to Interbull for MACE include weighted or simple averages of genetic evaluations predicted separately by parity, and linear functions of direct plus maternal genetic evaluations.
3. Approximate **national reliabilities** for all evaluated animals, under the same general classes of models, genetic effects and index traits described above for EDC.
4. Use the **same input files for EDC and 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 **reliabilities of residual polygenic** contributions in genomic evaluation models, under the same general classes of models and genetic effects described above.

---

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

---

## Getting Started

1. Use **unzip** to extract contents of mtedc.zip to a desired location on your Linux filesystem.
  - a. The zip archive includes a Makefile, all source code files, documentation files, and several small test files to demonstrate different applications and to verify proper installation of the software on your system.
2. Use **make** to compile and test the program.
3. Use **./check** (provided bash file) to verify the **make** tests that ran on your system.
  - a. The Linux command **diff** is used in **./check** to confirm that all test results generated by **make** on your system are identical to the same test results generated on a Lactanet server, and that the Log summary and Message capture files are only different because of server-specific messages like the “Time and memory resources used”.

## Customized Installations

- A. The software was written entirely in C and tested using gnu C compilers natively available in Linux operating systems. The source code files are provided for portability to Linux systems installed on hardware with differing architectures (e.g. 32-bit versus 64-bit). There is no expectation that software modifications would be required, but customizations for your system are allowed and encouraged if necessary.
- B. The software was designed to run in single-processor mode, and there is currently no multi-threading or parallel-processing version available.
- C. The provided **Makefile** (upper-case **M**) can be modified or replaced as needed with your own customized **makefile** (lower-case **m**). By default, the Linux **make** command will use **makefile** instead of **Makefile** if both files exist.

## Uninstalling

All created files after unpackaging mtedc.zip, installing and testing the software (Steps 1 to 3 in Getting Started) will be in your selected **unzip** location and sub-directory tree.

1. To uninstall and delete all traces of the package from your system, delete the files extracted from mtedc\_6b.zip plus the sub-directories created in steps 1 to 3 above, which can be done with **rm -fr mtedc\_6b/**, and delete the zipfile.
2. To uninstall only the compiled executable and the test files created by steps 2 and 3 above, use the Linux command: **make clean**. This will undo steps 2 and 3 only, while leaving the zip archive mtedc\_6b.zip and the unzipped contents on your file system.

## Using the Software for Applications

### Required Input Files

1. Parameters: Specify traits, models, indexes, covariances, file names & options.
2. Pedigree: Sires/dams (or sires/mgs) of all evaluated animals (or sires).
3. Data incidence: Matrix of contemporary groupings for animals/dams or sires/mgs.

### Output Files

1. Results file: EDC or reliabilities.
2. Log summary file: Time and memory resources used, data checks and output summary statistics.
3. Screen message capture files: Optional files to capture data streams for <stdout> and <stderr> messages, which are otherwise written to the screen/console.

### Program Execution and Assistance

1. The executable file is named **crEDC\_6b**, and it is created in the Makefile directory:  
Full path to executable = <zipfile location>/mteDC\_6b/crEDC\_6b
2. All details for an application can be stored in a parameters file, which is specified at run-time using the required command-line option: **-p <parameters\_filename>**
3. All "OPTION <option>" statements in the parameters file can alternatively be requested on the command-line with: **--OPTION <option>**
4. Preferred methods are set by default, but simpler methods and different variations can be specified on the command-line, for research and debugging purposes, and in case these are helpful to narrow down data-dependent problems.
5. To print a detailed description of available options to the screen/console, in the data stream <stdout>, use the command-line option: **-h** or equivalently **--HELP**. The current screen help message is included as an Appendix at the end of this manual.
6. See the included test files provided with the software, for example applications.

### Optional Methods for EDC and Reliability Approximation

Options are available to use either single-trait EDC methods (Interbull, 2000), or the more recently advanced multiple-trait EDC methods (Sullivan et al, 2006; 2007). The methods are equivalent for single-trait data sets, but for all other cases multiple-trait EDC should be preferred. Reliabilities can be adjusted for contemporary group (CG) effects or not, and while ignoring or accounting for animal relationships within CG. The absorption of full-sib progeny into sire by dam equations is accounted for by default, but an option is also provided to ignore sire by dam adjustments and absorb all progeny as if they were half sibs. The half-sib approach is a very common simplification in many reliability approximation algorithms.

## **INPUT 1: Parameters File (-p)**

This file is in free format, allowing users to customize as preferred for readability. Each line is interpreted as a collection of delimited “words”, where sequentially repeated delimiters (e.g. multiple spaces between two words) are treated as equivalent to a single delimiter. A word can be an alpha-numeric, an integer, or a floating-point value with any number of decimals.

Delimiters can be spaces, tabs or commas, or any combination of these three characters. A sequential combination that is space-tab-tab-comma, for example, is interpreted the same as 4 sequential spaces, 3 sequential commas, 8 sequential tabs, or a single instance of any of these delimiters, between a pair of words.

Scientific expressions using mantissa and exponent (e.g. 0.3123E2) should not be used in this file, because such values might be misinterpreted. The equivalent decimal expression should always be used instead, for example by specifying scientific values 0.3123E2 and 0.4567E-2 with the equivalent decimal expressions 31.23 and 0.004567.

### **Number of Traits, Model effects and Indexes (Line 1)**

The first line of this file must specify the number of traits included in the multiple-trait national evaluation, the subset of traits including maternal effects, and the number of index traits of interest. All traits that include maternal effects must be grouped together at the beginning of the list of traits, such that line 1 requires the following three words, as integer values:

1. The number of evaluated traits.
2. The subset number of traits with maternal effects.
3. The number of index traits.

### **Flexible Line Ordering and Commenting**

All lines after the first can be in any order, with the exception that successive rows for a covariance matrix must present the matrix in lower triangular form. Blank and commented lines can be inserted anywhere within the file to improve readability as desired. Blank lines are identified as having no words (but could have spaces or other delimiters), and comment lines are identified as having a comment character ‘#’ at the beginning of the first word. All other lines must begin with a recognized word that defines the contents of the line. All recognized words to define a line are listed below.

## Recognized Words to Define a Line in the Parameters File

TRAITS : a delimited list of names to describe the evaluated traits in output files. Note that the names themselves cannot include delimiters (i.e. spaces, commas, or tabs).

MODEL : specify if a SIRE or ANIMAL model was used (case insensitive sire/animal)

RR : a row of the residual covariance matrix among traits.

CG : a row of the covariance matrix among contemporary group effects treated as random

NOTES: CG rows are only required if random CG effects are fitted for at least one trait, in which case a full lower triangular matrix must be provided, with rows of zeros for any traits with fixed CG effects. Off-diagonal entries can take any value but CG correlations among traits are ignored with the current approximations. Adjustments for the information used to estimate CG effects are estimated independently by trait.

Ed : a direct-effects row of the covariance matrix of permanent environment effects on traits

Em : a maternal-effects row of the covariance matrix of permanent environment effects

Gd : a direct-effects row of the genetic covariance matrix among traits

Gm : a maternal-effects row of the genetic covariance matrix among traits

NOTES: The covariance matrices for permanent environment (E optional) and genetic (G required) effects must be ordered with all direct (D) effects followed by the maternal (M) effects, and with all D+M traits followed by the traits with only D. For traits with only M, these can be included as a D+M trait with a zero variance for D.

I1, I2, ... I<n> : relative weights on predicted genetic effects, defining index traits 1, 2, ... ,<n>, followed by the index trait heritability submitted to Interbull for MACE, followed optionally by the index trait name (spaces are allowed in index trait names)

FPED : name of the INPUT pedigree file

FOBS : name of the INPUT data incidence file

FOUT : name of the OUTPUT file for EDC or reliabilities

FLOG : name of the OUTPUT log/summary file

FSTDOUT : name of the OUTPUT capture file for <stdout> messages

FSTDERR : name of the OUTPUT capture file for <stderr> messages

OPTION <option>, where allowable values for <option> (fully described with -h) can be:

ignore\_h2\_check

ignore\_MAXSIREs\_check

ignore\_RECWTS

check\_all\_h2

ignore\_PD\_check

ignore\_SIRExMGS

h2\_within\_CG

ignore\_WTS\_check

include\_traits=1+3+4

NOTE: Any of these can also be requested with --OPTION <option> on the command line.

## An Example Parameters File

```
4 1 5
TRAITS trait1_D+M , trait2_D, tr3_D , Trait4
MODEL Animal
OPTION ignore_h2_check

# ---- INPUT files
FPED pedigree-file
FOBS incidence-file

# --- OUTPUT files
FOUT edc.outfile
FLOG edc.logfile
FSTDOUT edc.stdout
FSTDERR edc.stderr

# --- COVARIANCE MATRICES (symmetric) are in lower triangular form
RR 100.0
RR 25.1 110
RR 75.22 0.0 200
RR 0.0 0 0 500

CG 10
CG 0 0
CG 0 0 30
CG 0 0 0 0

Ed 0
Ed 0 0
Ed 0 0 0
Ed 0 0 0 0
Em 0 0 0 0 40

Gd 30
Gd 10 50
Gd 25 35 70
Gd 70 80 100 150
Gm -5 0 -10 0 10

# --- INDEX traits [k'] [MACE heritability] [Index trait name]
I5 1 0 0 0 0 0.2000 Direct Trait 1
I1 0 0 0 0 1 0.0667 Maternal Trait 1
I2 1 0 0 0 2 0.15 Total Maternal 1
I3 0 1 1 0 0 0.38 Average of 2+3
I4 0 .5 0 .5 0 0.38 Average of 2+4
```

## Index Trait Definitions for MACE

Index traits can be any that fall within the general class of linear functions ( $\mathbf{k}'$ ) of the genetic effects ( $\mathbf{g}$ ) included in the multiple-traits national evaluation model ( $\mathbf{I}=\mathbf{k}'\mathbf{g}$ ). The matrix  $\mathbf{k}'$  includes a row for each index trait of interest, and a column for each genetic (direct or maternal) effect across all traits. An example set of index traits was shown above and another example is shown below, for 6 index traits of interest from a 2-traits evaluation model, with direct and maternal effects on both traits:

I1	1	0	0	0	0.250	Trait 1 direct
I2	0	1	0	0	0.381	Trait 2 direct
I3	1	0	1	0	0.192	Trait 1 direct+maternal
I4	0	0	1	0	0.1502	Trait 1 maternal
I5	0.5	0	1		0.3115	Trait 2 maternal-grandsire effect
I6	0	1	0	2	0.3115	Trait 2 maternal-grandsire total maternal

The weights provided are always re-scaled by the software, to sums of 1 for each index. Thus, I5 and I6 are equivalent, and all EDC and reliabilities will be identical for traits I5 versus I6.

Linear functions can exactly describe many traits submitted to Interbull for MACE, and if there are no additional random effects in the national model other than the ones listed in the MTEDC parameters file, the trait heritability submitted to Interbull should exactly match the heritability derived from  $\mathbf{k}'\mathbf{g}$  by the MTEDC software. If there are notable discrepancies between a submitted versus derived index trait heritability, the software will issue warnings and the EDC output file will not be created. There is an available OPTION ignore\_h2\_check, to allow creation of an EDC output file despite these warnings, if exact heritability matches cannot be expected for all indexes, and this can be the case when:

- a) A trait submitted for MACE can only be approximated with a linear function, or
- b) The national model includes additional random effects (e.g. random sire-by-herd) that are not accommodated as separate input matrices to the MTEDC software.

The heritability derived from  $\mathbf{k}'\mathbf{g}$  is based on the linear definitions of each index in  $\mathbf{k}'$  and the genetic and non-genetic covariance matrices provided as RR, CG, Ed, Em, Gd and Gm. When an exact match to the MACE heritability cannot be expected, it is still desirable that the MACE heritability would be reasonably close to the derived heritability from  $\mathbf{k}'\mathbf{g}$ . Increasing the variances in matrix RR is a way to incorporate additional information loss due to the estimation of additional random effects in the national model. This approach can improve alignment between the heritability derived from  $\mathbf{k}'\mathbf{g}$  and the MACE heritability provided to Interbull.

To ensure a proper weighting of national EBV in MACE, the generated EDC are scaled to align with MACE heritability, as needed for properly scaled MACE reliabilities. The national reliabilities from MTEDC are consistent with  $\mathbf{k}'\mathbf{g}$  heritability and properly aligned with MACE reliabilities, but not with MACE heritability if it differs significantly from the  $\mathbf{k}'\mathbf{g}$  heritability.

## INPUT 2: Pedigree File (FPED)

Animals can be numbered in any order, and sequential numbering is not required. Use positive integer values for known individuals, and zero or negative integers for unknown individuals (i.e. unknown parents or unknown parent groups). In many cases, it is expected that the same pedigree files used for genetic evaluation could also be used as input for the MTEDC software.

If necessary, animals will be re-numbered chronologically by the software to ensure proper execution of the approximation algorithms. The chronological ordering will fail, however, if there are pedigree loops, for example if an animal is an immediate or distant ancestor to itself. Pedigree loops must therefore be avoided when creating the input pedigree file.

Each line in the pedigree file must have at least 3 words, while any number of words greater than three is allowed. Animal descriptions could therefore be added after the third word, for example the animal's registration, name, birth date, or any other information of interest. Only the first 3 words are processed by the software, which expects these words to represent:

ANIMAL SIRE DAM (when specifying MODEL ANIMAL)  
SIRE Sire\_of\_SIRE MGS\_of\_SIRE (when specifying MODEL SIRE)

## INPUT 3: Data Incidence File (FOBS)

This file provides incidence matrices for the contemporary group, animal, dam, sire, and MGS effects included in the model. The MODEL specification in the parameters file informs the program which of these effects are included. For example, the Dam (or MGS) is only expected if there are traits with maternal effects. The Animal (or Sire) is always required, even if direct genetic variance is 0 for all traits, in which case the direct genetic effect predictions will be equal to zero for all animals.

The data incidence file should include a line for every individual with phenotype(s) for at least one of the traits. Repeated records are allowed for the same individual in national repeatability models, and records can be sorted in any order. The expected words in each line are:

- 1) Integer ID of the individual (animal or sire) contributing a direct effect to the observation(s).
- 2) (If maternal effects) Integer ID of the individual (dam or MGS) contributing maternal effects.
- 3) Integer IDs for contemporary group of every trait (non-zero for only the observed traits).
- 4) An optional set of record weights for each trait (allowing heterogeneous residual variances).

Record weights can be either excluded completely or provided for all traits. If heterogeneous variances apply to only a subset of the traits, use record weights of 1 for the unweighted traits



with homogeneous residual variances. A given record weight of  $S$  corresponds with a residual variance of  $1/S$  times the residual variance for a record with weight equal to 1.

The numbering of individuals must be the same in the incidence file as in the pedigree file.

The order of traits in the incidence file, for contemporary group IDs and record weights, must match the trait order in all covariance matrices specified in the parameters file.

Example records below are for an animal model with maternal effects, where animals 3 and 4 are maternal sibs of dam 2, with observations for traits 1,3,4 and 1,2,4 respectively. Individual 4 has observations with higher residual variances (e.g. they could be partial lactation records)

```
3 2 1 0 4 3 1.000 , 0.000 1.000 1.000
4 2 , 1 1 0 3 0.750 0.750 0.000 0.600
```

## Embryo Transfer Animals

For models with maternal effects, the maternal environment of an embryo transfer calf is provided by the recipient cow (e.g. dam in FOBS), while the genetic contributions to maternal breeding value of the calf are from the embryo donor (i.e. dam in FPED). The approximation algorithms used in MTEDC software will accommodate ET calves with a proper allocation of environmental versus genetic maternal contributions, to recipient and donor cows respectively, when the dam of a calf is different between the FOBS and FPED files.

Regardless of this feature to properly separate maternal contributions of donor versus recipient dams of ET calves, it is important to fit ET data for EDC and reliabilities in a way that is consistent with the EBV. The pedigree and data incidence codes used as input to MTEDC should therefore be the same as were used to compute the EBV, thus ensuring EDC and reliabilities reflect properly how maternal contributions were apportioned in the EBV model.

The approximation methods are slightly more complicated when allowing multiple maternal dams to affect a single performance record of a calf. The addition of this feature in MTEDC software was extensively tested, during which time an option (`--noET`) was used to ignore and thereby exclude ET calf records when processing the input file FOBS. This option simplified examination of the ET data impacts on performance of the program and resulting outputs. The `--noET` option is still available in the current version, should users wish to examine the impacts of ET records with their own data sets.

## **OUPUT 1: Log File (FLOG)**

The log file contains summary information about the input data and parameters provided, the command-line options used, the computing resources that were required, and the generated results. Many of the warnings and error messages captured in FSTDOUT and FSTDERR are also included in FLOG.

## **OUPUT 2: Results File (FOUT)**

The format and contents of the results file (FOUT) are dictated by the chosen options for the application. The output file formats are therefore fully described in the FLOG output file. The results file can be either a fixed-length format by default, or a variable-length comma-delimited format with the option -d. If requesting -d, then a header row can also be added with -H, to describe the contents of each field within FOUT. When using -dH (equivalent to -d -H), the header row can be made more informative by providing a list of trait names, using TRAITS in the parameters input file (-p).

By default, the EDC are output as integer-rounded values. Option -R will instead write the EDC as real values rounded to the fourth decimal.

With EDC output, FOUT includes both EDC and progeny phenotype counts by default. Grand-progeny phenotype counts can also be added with -G.

## **OUPUT 3+4: Screen capture (FSTDOUT & FSTDERR)**

Screen messages showing the steps of program execution, timings and resources used, etc, are saved in FSTDOUT. Error messages are captured in FSTDERR. Under normal execution, and when there are no detected data problems, FSTDERR should be an empty file, and FSTDOUT will include the final message “Program completed successfully”, followed by a summary of the total RAM and time that was used. The summary of total RAM and time used is also written to the end of the FLOG file.

## Appendix: Screen help message (crEDC\_6b -h)

Usage for crEDC\_6b

crEDC\_6b [-h] [-p \*] ...

-h : Usage (this message)  
--HELP : Usage (this message)

-p \* : Use parameters file \* (REQUIRED OPTION)  
--DRYRUN : List all files to be processed, then quit ... no files will be created, modified or deleted  
--OUTDIR \* : Write ALL output files in directory \* ignoring output directories in parameters file  
-d : Format output file as comma-delimited variable length  
-dH : Include a header row in the comma-delimited output file  
-R : Output EDCs as Real (F12.4) instead of integer (I7)  
-A : Add EDC-based reliabilities (EDC/(EDC+vratio)) to EDC output file  
NOTE: -A reliabilities are interesting, but they ignore EDC of male relatives.  
USE: -w REL4 for reliabilities that include information from all relatives.

-G : Add grand-progeny record counts to EDC output file

NOTE: the output filename (FOUT in parameters file) will be appended if requesting an option below...

--EDC : Generate EDCs using the preferred method (PGS4) ... written to FOUT appended with \_EDC  
--REL : Generate Reliabilities using the preferred method (REL4) ... written to FOUT appended with \_REL

Additional EDC/REL options are available using -w ...

-w \* : Use weighting factor method \* for EDC. Options for \* are:  
ND => Number of Daughters with a record ... FOUT appended with \_ND  
Ne => Simple adjustment (1-1/n) for CG effect ... FOUT appended with \_Ne  
ITB => Interbull (2000) method with PGS1 correction ... FOUT appended with \_ITB  
PGS1 => ITB method corrected for sire models (Soussse 2004) ... FOUT appended with \_PGS1  
PGS4 => MT-absorption of progeny (Kuopio 2006) ... FOUT appended with \_PGS4  
(-w PGS4 and --REL use the same methods, but the output file names differ)  
-w \* : Use reliability method \* for REL. Options for \* are:  
REL4 (consistent with -w PGS4 for EDCs) ... FOUT appended with \_REL4  
RELM (uses Karin Meyer's general approximation strategy) ... FOUT appended with \_RELM  
ref: Tier and Meyer (2004) J. Anim. Breed. Gen. 121:77-89.  
ref: Meyer (1989) Livest. Prod. Sci. 21:87-100.  
=> Xb adjustment accounts for # animals in CG  
=> Absorptions ignore some mate information  
ACC4 => output accuracies (sqrt(REL4)) instead of reliabilities ... FOUT appended with \_ACC4  
ACCM => output accuracies (sqrt(RELM)) instead of reliabilities ... FOUT appended with \_ACCM  
(-w REL4, -w ACC4 and --REL use the same methods, but the output file names differ)

NOTE: --EDC is the default if no request for --EDC, --REL or -w (FOUT is not appended in this case)

-O \* : Use OPTION \* (e.g. -O ignore\_h2\_check)  
--OPTION \* : Use OPTION \* (e.g. --OPTION ignore\_h2\_check)  
ignore\_h2\_check : do not stop processing due to h2\_check warnings.  
check\_all\_h2 : summarize h2 checks for all traits, not just the traits with h2 warnings.  
ignore\_MAXSIREs\_check : do not stop processing due to CG with a very large number of sires.  
ignore\_PD\_check : do not stop processing due to a covariance matrix that is not p.d.  
ignore\_WTS\_check : allow any value for record weights (otherwise WT>3.0 is considered a data error).  
ignore\_RECWTS : ignore record weights provided in the data file (same as -X).  
h2\_within\_CG : do not include V(CG) in V(y) when computing heritability.  
ignore\_SIRExMGS : assume unrelated SIRE-MGS in a SIRE model with maternal effects.  
include\_traits=1+3 : process a subset of the traits, in this example traits 1 and 3 only.  
-o : Redirect <stdout> and <stderr> to default-named files (e.g. parameters\_filename.stdout)  
... OR specify preferred names for these files with FSTDOUT and FSTDERR in the parameters file.  
... OR use Linux re-direction (e.g. >out\_file 2>err\_file in bash)  
--SCREEN : Over-ride and ignore: -o, FSTDOUT and FSTDERR (--CONSOLE and --STDOUT are equivalent)

Features to help debug data-dependent problems

-S : Selection index, ignore estimation of CG effects ... FOUT appended with \_SI  
-B : Assume 1 big CG includes all animals (approximately == -S) ... FOUT appended with \_1CG  
-N : include records for natural progeny only, i.e. ignore ET records ... FOUT appended with \_noET  
-X : ignore the record weights in FOBS file  
-z : Ignore non-zero residual correlations (then EDC should only increase if rG>0)