

TM_TRENDS – DOCUMENTATION

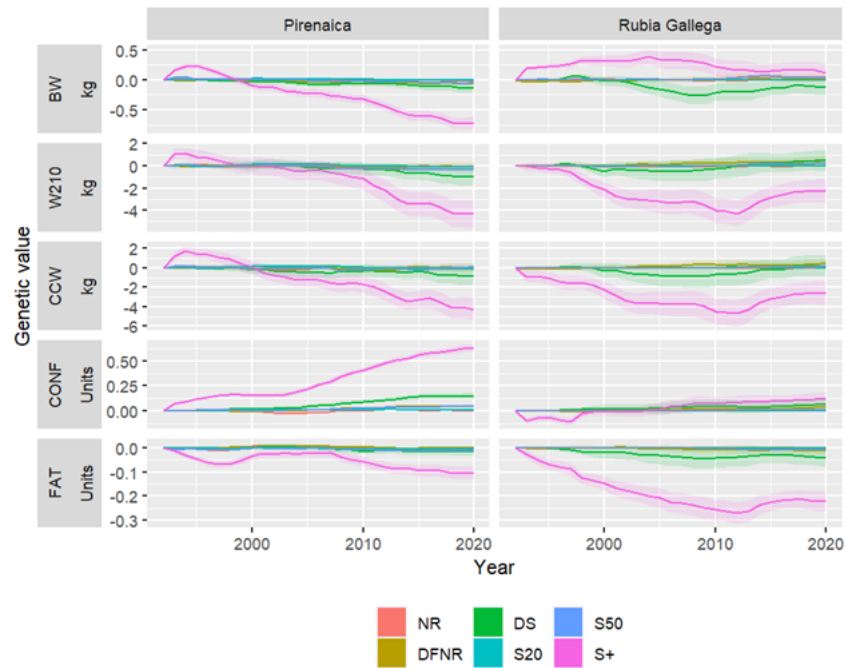
v. 0.1

David López-Carbonell¹, Andrés Legarra², and Luis Varona¹

¹Genética Cuantitativa y Mejora Animal , Universidad de Zaragoza,
Instituto Agroalimentario de Aragón (IA2) , 50013, Zaragoza, Spain

²CDCB, 4201 Northview Drive, Bowie, 20716, MD, United States

September 26, 2024



Contents

1	Introduction	3
2	Functionality	3
3	Methods	3
4	Usage	4
4.1	Parameter file	4
4.2	Pedigree file	4
4.3	Data file	5
5	Compilation and Running	5
6	Output	5
6.1	Genetic trends	5
6.2	Genetic trends samples	5
7	Results analysis	6
8	Contact	6
	References	6

1 Introduction

This is the documentation of the software **TM_TRENDS**. **TM_TRENDS** software is available on GitHub This is a tool to partition genetic trends following García-Cortés et al. (2008) method under a Bayesian approach. It was developed using TM (Legarra et al., 2008) as basis. All TM instructions and functionalities are maintained and can be consulted in the TM documentation, available on https://genoweb.toulouse.inra.fr/~alegarra/tm_folder/manualtm.pdf

2 Functionality

This software is an alternative for partitioning genetic trends. It allows do the analysis from a Bayesian perspective. In this way, uncertainty of the (co)variances componentes can be handled. Furthermore, as the standard TM software, it allows to compute:

1. multi-trait models
2. binary traits
3. censored continuous traits

3 Methods

Statistical analysis carried out by the software **TM_TRENDS** is completely equivalent to TM (Legarra et al., 2008) software. Partitioning of the genetic trends was done following García-Cortés et al. (2008) method. To develop the partition of the genetic trends, the vector of breeding values for each trait (\mathbf{u}_t) is described as linear combinations of the Mendelian sampling terms(ϕ) Therefore, it can be expressed as: $\mathbf{u}_t = \mathbf{T}\phi_t$, where T is a triangular matrix that describes the flow of the genes through pedigree (Woolliams et al., 1999).

Further, García-Cortés et al. (2008) defined a set of partitions based on k cluster of individuals.

$$\mathbf{I} = \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \dots + \mathbf{P}_k$$

. where \mathbf{P}_i . are diagonal matrices with ones in the diagonal if the individual belongs to the i th cluster and zero otherwise. Then, $\mathbf{u}_t = \mathbf{T}(\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \dots + \mathbf{P}_k)\phi_t$. As $\phi_t = \mathbf{T}^{-1}\mathbf{u}_t$:

$$\mathbf{u}_t = \mathbf{TP}_1\mathbf{T}^{-1}\mathbf{u}_t + \mathbf{TP}_2\mathbf{T}^{-1}\mathbf{u}_t + \dots + \mathbf{TP}_k\mathbf{T}^{-1}\mathbf{u}_t$$

and:

$$\mathbf{u}_t = \mathbf{u}_{t(1)} + \mathbf{u}_{t(2)} + \dots + \mathbf{u}_{t(k)}$$

where $\mathbf{u}_{t(k)} = \mathbf{TP}_k\mathbf{T}^{-1}\mathbf{u}_t$ is the vector of the k th partions of the breeding values.

The **TM_TRENDS** software implemented this partition at each iteration of the Gibbs Sampler and computes the average of each year partitions to calculate of the posterior mean and the posterior standard deviation of the partitioned genetic trends for each trait. Additionally, it generates a text file that contains the samples of the partitioned genetic trends that can be used to calculate the posterior distributions

4 Usage

4.1 Parameter file

Two additional lines are added to TM original parameter file. They are located after the number of effects definition line (8 and 9). The first one reflect the number of temporal units (i.e. yeards) in which genetic trends have to divided. The second one indicates the number of clusters in which the population is split. Both must be an integer numeric value. An example parameter file would be:

```
Data file
data.txt
Pedigree file
genea.txt
Model
animal
6 Number of effects (including animal)
30 Number of unit times
2 Number of groups
1 Number of covariates
1 Number of genetic groups
2 Number of traits
1 Number of threshold traits
2 Categories for the threshold traits
0 Number of random environmental effects
1 Number of animal effects
1 10 20 100 20 1000 Levels for each effect (do not include genetic group)
1 1 1 1 1 1 Model for trait 1
0 1 1 1 1 1 Model for trait 2 ... repeat as many lines as traits
Task
VCE
Total number of iterations
100000
Burn-in (discarded only in the results and solutions file)
30000
Thin interval (samples are taken every...)
100
Genetic variance
1 0
0 1
Permanent (keep always this title)
Residual
1 0
0 1
```

4.2 Pedigree file

Two additional columns must be added to pedigree file. First additional column is the cluster in which that individual is assigned. It has to be codified as an integer number from 1 to

the number of clusters. Second one is the time unit in which that individual was assigned (born). It has to be codified as an integer number from 1 to the number of time units.

4.3 Data file

Data file doesn't suffer any change compared to TM necessities

5 Compilation and Running

There is not differences compared to TM software.

6 Output

Screen printing doesn't differ from TM software screen printing.

6.1 Genetic trends

A file `trends.txt` containing genetic trends is produced every 100 · thin iterations after burn-in. This file is built by a total of 7 columns. Header is not provided. Each column contains:

- 1 Trait number
- 2 Animal effect number
- 3 Cluster number
- 4 Time unit
- 5 Mean estimate of contribution to the genetic trend
- 6 Standard error of the genetic trend estimate
- 7 Number of individuals in that genetic cluster

It is important to highlight two aspects of this file results:

- Contribution of the individuals in the time unit 1 is settled to 0 for all clusters. To avoid this, a previous phantom generation can be added to the pedigree file.
- Cluster *number of clusters* + 1 indicates the whole population genetic trend.

6.2 Genetic trends samples

All samples from the posterior distribution of the genetic trends estimates (each cluster and whole population) are printed. It is done for each trait and animal effect separately. Then, the number of files is variable. File name is `trends_trait_n_a.txt`, being *n* the number of the trait and *a* the number of animal effect.

Each file is form by a number of *number of time units* + 2 columns and *number of iterations* · (*number of clusters* + 1) rows. Each column, contains:

- 1 Iteration number, i .
- 2 Cluster number, k .
- 3 Estimation in iteration i of the contribution to the genetic trend of the cluster k at unit time 1.
- 4 Estimation in iteration i of the contribution to the genetic trend of the cluster k at unit time 2.
- ...
- 2 + n Estimation in iteration i of the contribution to the genetic trend of the cluster k at unit time n.

It is important to highlight two aspects of this file results:

- Contribution of the individuals in the time unit 1 is settled to 0 for all clusters. To avoid this, a previous phantom generation can be added to the pedigree file.
- Cluster *TOT* indicates the whole population genetic trend.

7 Results analysis

Multiple examples and their analysis and interpretation using R R Core Team et al. (2013) can be found on **TM_TRENDS** GitHub.

8 Contact

All questions, doubts and comments about this software are always welcome. We would appreciate feedback and problems report to improve **TM_TRENDS** software. It can be done by e-mail to davidlc@unizar.es

References

- García-Cortés, L. A., J. Martínez-Ávila, and M. Toro (2008). Partition of the genetic trend to validate multiple selection decisions. *Animal* 2(6), 821–824.
- Legarra, A., L. Varona, and E. Lopez de Maturana (2008). Tm threshold model. *Retrieved on 26*.
- R Core Team, R. et al. (2013). R: A language and environment for statistical computing.
- Woolliams, J., P. Bijma, and B. Villanueva (1999). Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics* 153(2), 1009–1020.