



UNIVERSITY OF  
**GEORGIA**

College of Agricultural &  
Environmental Sciences

*Animal Breeding and  
Genetics Group*

# Best practices in BLUPF90

**Daniela Lourenco**

**Ignacio Aguilar**

BLUPF90 TEAM – 02/2023

*Armidale Animal Breeding  
Summer Course 2023*

*Daniela Lourenco  
Mehdi Sargolzaei*

1. Read the manual
2. Understand what the programs can do for you
3. Understand the theory
4. Read the output of the programs

```
blupf90+ renf90.par | tee blup.log
```

1. Understand the data

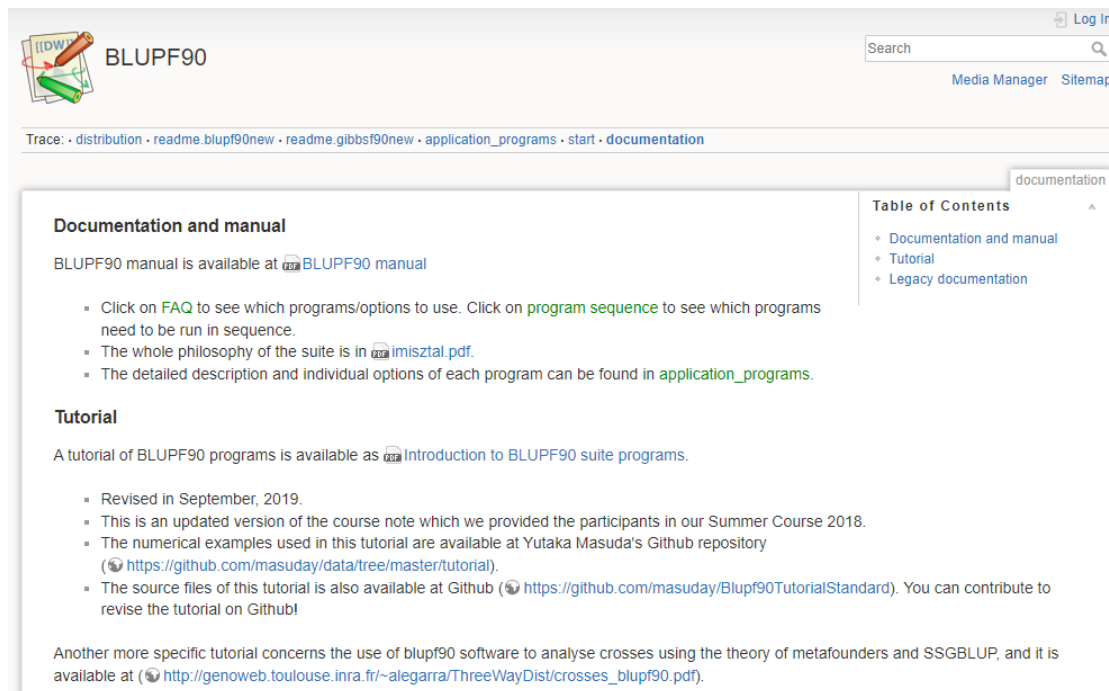
How many genotypes?

How many phenotypes?

How many geno with pheno?

Outliers?

Population structure (PCA)



The screenshot shows the BLUPF90 website documentation page. At the top left is the BLUPF90 logo, which consists of a stylized 'B' and 'F' with a red arrow and a green arrow. To the right of the logo is the text 'BLUPF90'. In the top right corner, there is a search bar and a 'Log In' link. Below the search bar are links for 'Media Manager' and 'Sitemap'. A breadcrumb trail reads: 'Trace: · distribution · readme.blupf90new · readme.gibbsf90new · application\_programs · start · documentation'. On the right side, there is a 'Table of Contents' menu with links for 'Documentation and manual', 'Tutorial', and 'Legacy documentation'. The main content area is titled 'Documentation and manual' and contains the following text: 'BLUPF90 manual is available at [BLUPF90 manual](#)'. Below this, there are three bullet points: 'Click on [FAQ](#) to see which programs/options to use. Click on [program sequence](#) to see which programs need to be run in sequence.', 'The whole philosophy of the suite is in [imisztal.pdf](#).', and 'The detailed description and individual options of each program can be found in [application\\_programs](#).'. Below this is a section titled 'Tutorial' with the text: 'A tutorial of BLUPF90 programs is available as [Introduction to BLUPF90 suite programs](#)'. This section contains three bullet points: 'Revised in September, 2019.', 'This is an updated version of the course note which we provided the participants in our Summer Course 2018.', and 'The numerical examples used in this tutorial are available at Yutaka Masuda's Github repository (<https://github.com/masuday/data/tree/master/tutorial>).'. The final bullet point states: 'The source files of this tutorial is also available at Github (<https://github.com/masuday/Blupf90TutorialStandard>). You can contribute to revise the tutorial on Github!'. At the bottom of the page, there is a paragraph: 'Another more specific tutorial concerns the use of blupf90 software to analyse crosses using the theory of metafounders and SSGBLUP, and it is available at ([http://genoweb.toulouse.inra.fr/~alegarra/ThreeWayDist/crosses\\_blupf90.pdf](http://genoweb.toulouse.inra.fr/~alegarra/ThreeWayDist/crosses_blupf90.pdf)).

# Best practices – BLUPF90

- Steps to estimate breeding values using BLUPF90 family of programs
  1. renumf90 to renumber the data
  2. Estimate variance components if they are not available
    1. If need to estimate: blupf90+ with OPTION method VCE  
The solutions file has EBV computed with the estimated VC
    1. If available: put the values into renf90.par and run blupf90+ to estimate (G)EBV

# Best practices – BLUPF90

- Remove animals from the data file that have missing phenotypes
  - They are just increasing the number of equations
  - They will get predictions if they are in the pedigree or genotype file
  - In renumf90: `OPTION remove_all_missing`
- Dealing with genomic information
  - Run preGSf90 to perform quality control and save the clean files
  - `OPTION saveCleanSNP`
  - `OPTION createGInverse 0`
  - `OPTION createA22Inverse 0`
  - `OPTION createGimA22i 0`
  - The 3 last options avoid extensive computations

# Best practices – BLUPF90

- Dealing with genomic information
  - Work with the clean file: `snp.txt_clean`
  - `OPTION SNP_file snp.txt_clean`
  - `OPTION no_quality_control`
- Investigate the output
  - What is the correlation between **G** and **A<sub>22</sub>**?
  - Diagonal = inbreeding correlation
  - All → ideal range [0.5 – 0.9]
    - Lower = incompatibility of genomic and ped information (at least one is wrong)
    - Lower = admixed population
    - Higher = limited advantage of genomic information

# Best practices – BLUPF90

- Parent-progeny conflicts?
  - preGSf90: removes genotypes for progeny with conflict
  - Seekparentf90 can search for possible parents among genotyped animals
    - Corrects the pedigree
  - Which one is better?
    - Pedigree and genotypes may be wrong
    - Knowledge about the data

# Best practices – BLUPF90

- Running any software that takes a parameter file besides renumf90?
  - Use `OPTION use_yams`
  - Speeds up several computations, especially with multi-trait models
- Estimating variance components with REML or AIREML in blupf90+?
  - Use `OPTION use_yams`
  - Do not use UPG or metafounders
    - Weird behavior if inverting the LHS of MME as in VCE
  - Need to know if convergence was reached?
    - Try blupf90+ and gibbsf90+

# Best practices – BLUPF90

- blupf90+ and gibbsf90+ can compute SE for estimates
  - `OPTION se_covar_function <label> <function>`

```
Sampling variances of covariances function of random effects (n=10000)
h2d - Function: g_2_2_1_1/(g_2_2_1_1+r_1_1)
Mean: 0.31113
Sample Mean: 0.31103
Sample SD: 0.82903E-02
elapsed time 2.1800995E-03
```

- Is it SD or SE?
  - Sample SD = SE



# Best practices – BLUPF90

- The output depends on the input

$$f(\text{trash}) = \text{trash}$$

$$f(\text{trash}) = \text{trash}^3$$

Plus  
genomics