Applied Statistical Methods In Animal Science

Peter von Rohr

20.02.2023

Administration

- Course: 2 hours of lecture (2 V)
- Plan: 2 V $\rightarrow$ 1 U + 1 V (i.e., 1 hour of lecture intersperced with time to do exercises)
- Exercises: Work on problems in R
- Material: course notes, slides, solution to exercises
- Exam: written, date: 22.05.2023, 08:15-10:00

Objectives

The students

- are familiar with the properties of **fixed linear effects models**
- are able to analyse simple data sets
- know why least squares cannot be used for genomic selection.
- know the statistical methods used in genomic selection, such as
  - BLUP-based approaches,
  - Bayesian procedures and
  - LASSO.
- are able to solve simple exercise problems using the statistical framework R.

Lecture Program

Week	Date	Topic
1	20.02	Introduction to Applied Statistical Methods
2	27.02	Linear Regression Models
3	06.03	Linear Fixed Effect Models
4	13.03	Model Selection
5	20.03	Pedigree BLUP
6	27.03	Variance Components
7	03.04	GBLUP - Marker-Effects Models
8	10.04	**Easter Monday** (Ostermontag)
9	17.04	GBLUP - Breeding Value Models
10	24.04	Lasso
11	01.05	**Labor Day** (Tag der Arbeit)
12	08.05	SVM
13	15.05	Bayesian Approaches in Linear Mixed Effects Models
14	22.05	Exam
15	29.05	Pentcote Monday (Pfingstmontag)

Information

- Website: https://charlotte-ngs.github.io/asmss2023
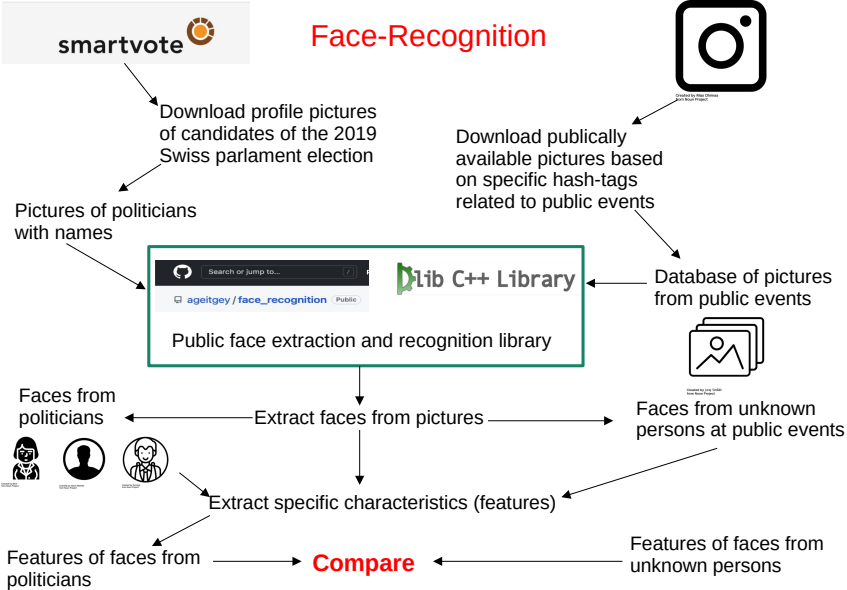- Topics for master thesis: will follow
- Exam: 22.05.2023 08:15 – 10:00

This Course

▶ Use dataset that is used to predict genomic breeding values and introduce four methods

1. Fixed Linear Effects Models - Least Squares
2. GBLUP - genomic version of BLUP
3. LASSO - still fixed linear effects model, but modified parameter estimation
4. Bayesian approach to estimate unknown parameter

Significance

- ▶ Why is this important?
- ▶ Is this only relevant for animal breeding?
- ▶ What about the rest of animal science?
- ▶ General trend of collecting data has led to development of `Big Data`
- ▶ Examples
  - ▶ Presidential campains in the US
  - ▶ Health care
  - ▶ Face recognition
  - ▶ Agriculture: Precision Farming
  - ▶ Animal Science: Precision Livestock Farming

Face Recognition

Face-Recognition

Download profile pictures of candidates of the 2019 Swiss parlament election

Download publically available pictures based on specific hash-tags related to public events

Pictures of politicians with names

Dlib C++ Library

Public face extraction and recognition library

Database of pictures from public events

Faces from politicians

Extract faces from pictures

Faces from unknown persons at public events

Extract specific characteristics (features)

Features of faces from politicians

Compare

Features of faces from unknown persons

Smart Farming

- Automated Milking Systems (AMS): Collection of data
- Sensor systems: SESAM
- CFIT: automated gathering of feed intake data

SESAM
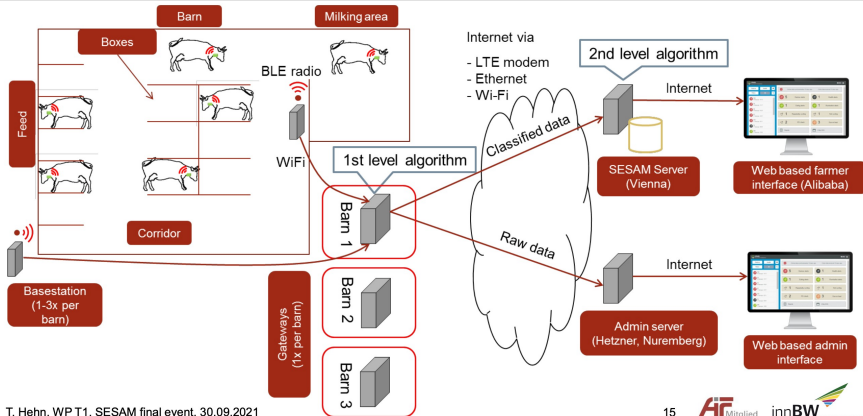
CFIT

Use video image analysis based on time of flight (TOF) approach for

- cow identification
- collection of feed intake data
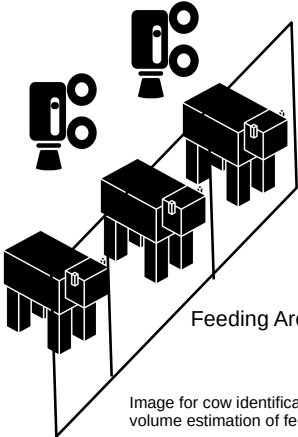
Cow Identification

Associate Image with Ear Tag ID
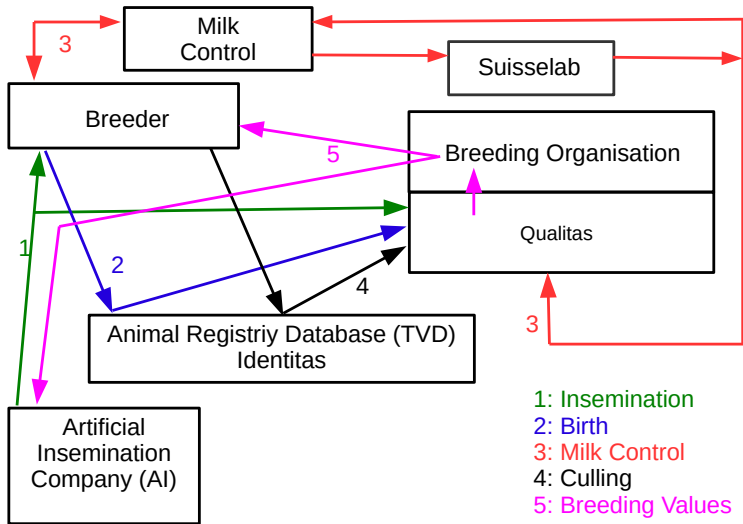
3D camera

ear tag reader

Image plus ear tag ID

Feeding Area

Image for cow identification and volume estimation of feed

Traditional Animal Breeding

- Before 2006
- Data collected for other purposes were used to predict breeding values
- Predicted breeding values as side-product

Data Logistics

Genomic Selection

- ▶ Same goal as in traditional breeding: Find animals with best genetic potential as parents of next generation
- ▶ New: use additional source of information
- ▶ **Genomic** information
  - ▶ spread accross whole genome
  - ▶ single nucleotide polymorphisms (SNP)
- ▶ Introduction:

"> Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829"
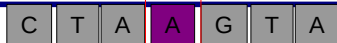
- ▶ Popularisation:

"> L. R. Schaeffer. Strategy for applying genome-wide selection in dairy cat- tle. Journal of Animal Breeding and Genetics, 123(4):218–223, 2006. ISSN 09312668. doi: 10.1111/j.1439-0388.2006.00595.x."
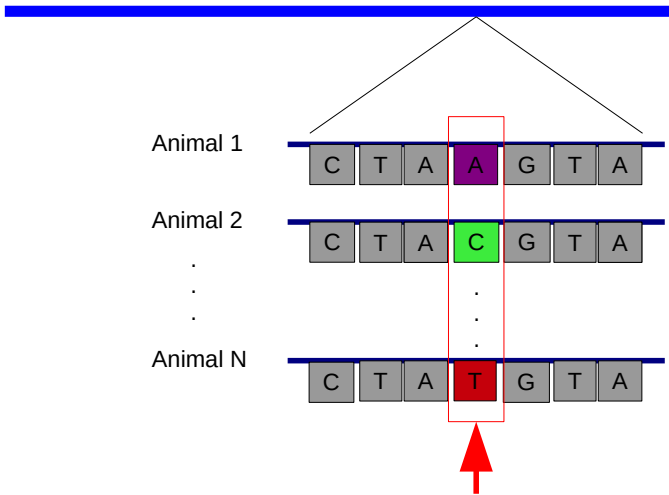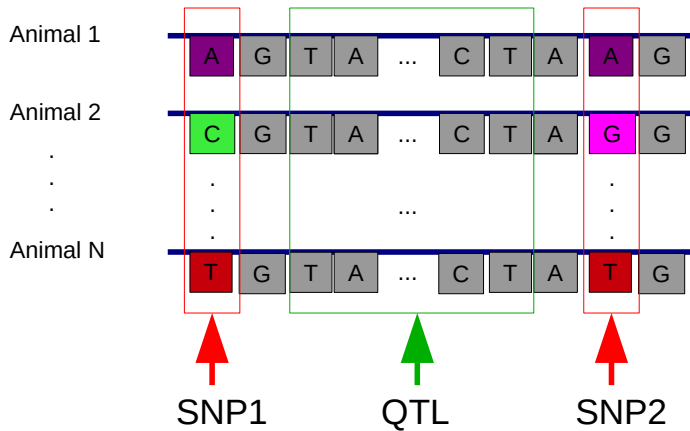
QTL

Linkage

- ▶ Flanking SNPs and QTL not independent passed on from parents to progeny
- ▶ Favorable QTL-allele linked with a given SNP-allele
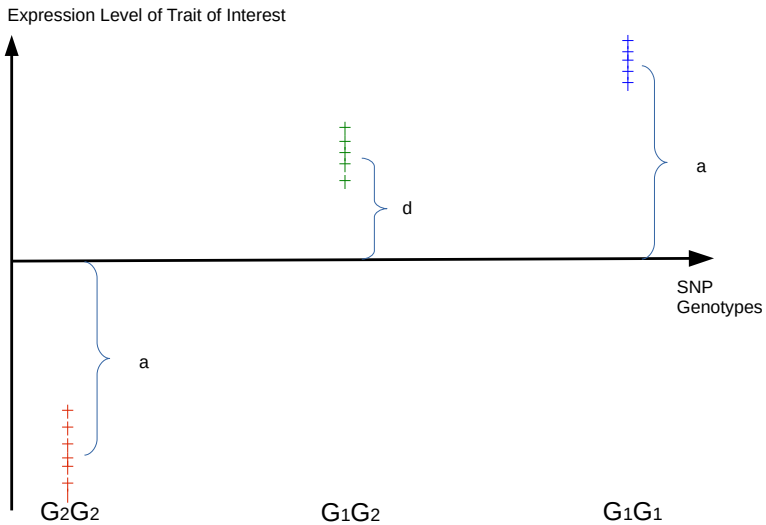- ▶ QTL is unknown, but use SNPs close to QTL as information for selection

Monogenic Model

- Assume quantitative trait is influenced by one locus only
- Locus is bi-allelic $\rightarrow$ two alleles (G_1 and G_2) and three genotypes
- Look at Distribution of trait values for three different genotypes

Distribution No Effect
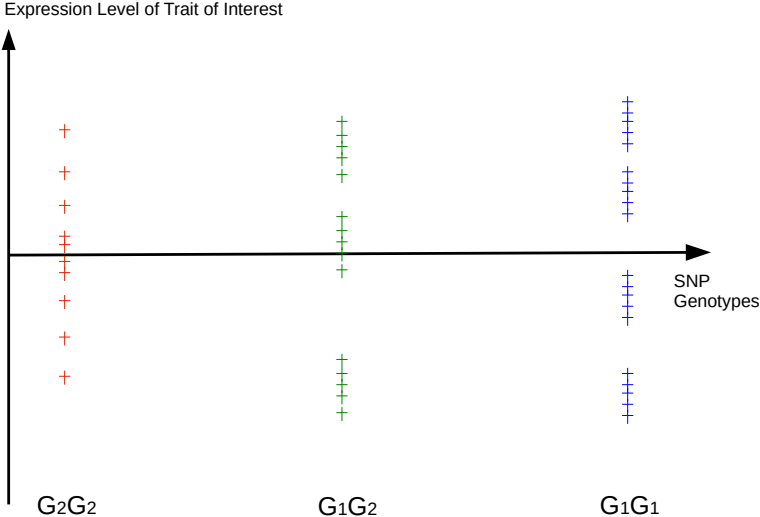
Distribution With Effect

Breeding Value

- Definition: Two times deviation from large number of offspring from population mean
- Assume: Hardy-Weinberg equilibrium
- Compute population mean as expected value of genotypic values
- Compute expected genotypic value of offspring for each of the three parental genotypes
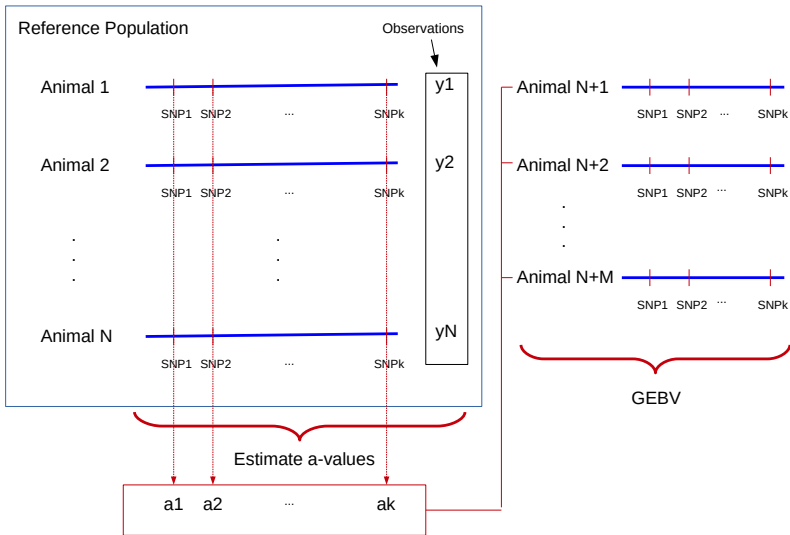- Assume purely additive loci, hence $d = 0$

Genomic Breeding Value

- Take into account many loci
- Approximate unknown QTL with linked SNP
- Estimate a-effects from monogenic model
- Compute genomic breeding values for all loci based on a effects

Two Approaches

1. Two Step Procedure (used currently in Swiss Dairy Cattle)
2. Single Step

Two Step

Single Step

- Combine all information into one single BLUP-based analysis
- Problem: Determine covariance between animals with and without genomic information

Summary: Traditional versus Genomic Selection

Animal Model

Genomic Selection

Population

Infinite number of loci with unknown position with infinitely small effect ==> **Infinitesimal Model**

Finite number of loci with estimated effect ==> **polygenic model**