Theoretical Accuracies

**Daniela Lourenco**

05-24-2018

Understanding Accuracy

- Prediction Accuracy

  - Validation

  - $\rho = COR(U, \hat{U})$

  - $\Delta G = i\,\rho\sigma_u/L$

  - Population

  - Very popular with genomics

- Theoretical Accuracy

  - Precision of EBV

  - How much EBV changes

  - PEV = Diag($\mathbf{C^{ZZ+}}$)

  - $Acc = \sqrt{(1 - \frac{PEV}{\sigma_u^2})}$

  - Individual

Understanding Accuracy

- Under selection
  - Theoretical accuracy ≠ Prediction accuracy
  - Theoretical is NOT affected by selection
  - Prediction is affected and varies if different i
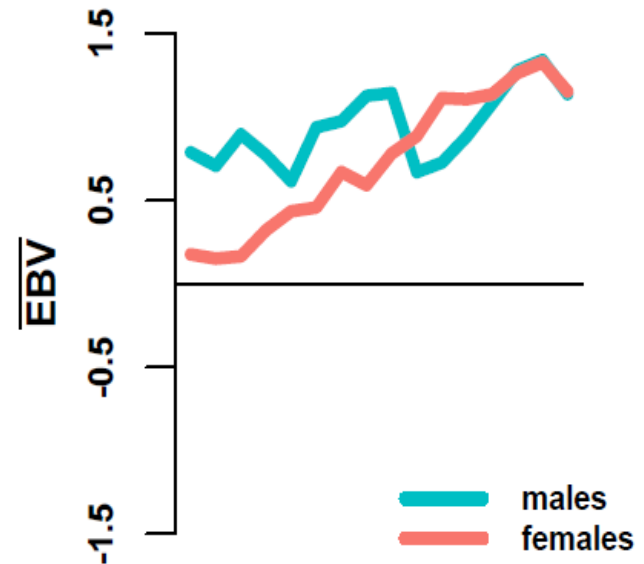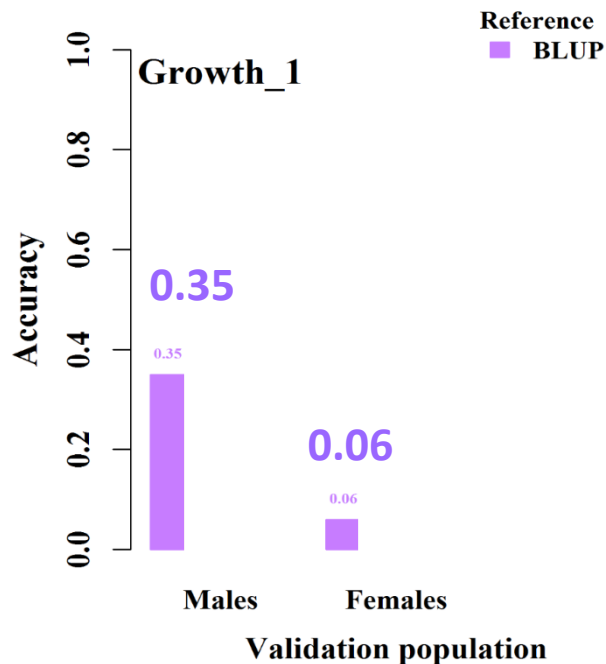
Theoretical Accuracy of EBV

- EBV are published with accuracy
  - "how much we can trust EBV"
  - amount of info used to calculate EBV

$$\text{Diag}(C^{ZZ+}) = \text{PEV}$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Records Pedigree

Theoretical Accuracy of EBV

- Relies on the inverse of $\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda$

- What if we have large datasets?

- Can we invert $\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda$?

$$\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda$$

$\downarrow$ $\downarrow$

Records   Pedigree

d_i^r d_i^p $\longrightarrow$

- d_i^r and d_i^p are approximated

(Misztal and Wiggans, 1988)

$$\text{Diag}(\mathbf{CZZ^+})_{ii} = \frac{1}{(\lambda + d_i^r + d_i^p)}$$

Theoretical Accuracy of EBV

$$\mathbf{Diag(CZZ^{+)}}_{ii} = 1 \Big/ (\lambda + d_i^r + d_i^p)$$

$$\mathrm{Accuracy}_i = \sqrt{1 - \mathbf{Diag(CZZ^{+)}}_{ii}}$$

- d_i^r and d_i^p are approximated

  (Misztal and Wiggans, 1988)

- Implemented in accf90

```
If (small data) then
   use blupf90 to get PEV
else
   use approximation
```

Theoretical Accuracy of GEBV

- What if we have genomic data?

```
If (small data) then
    use blupf90 to get PEV
else
    use approximation
```

- Which approximation?

Theoretical Accuracy of GEBV

- Approx. accuracy from ssGBLUP - Misztal et al. (2013)

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\left\{ \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} + \lambda \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \right\}$$

How to approximate d_i^g ?

$$d_i^r \qquad d_i^p \qquad\qquad d_i^g$$

$$\text{Diag}(\mathbf{CZZ^+})_{ii} = 1 \Big/ (\lambda + d_i^r + d_i^p + d_i^g)$$

8

Theoretical Accuracy of GEBV

- Misztal et al. (2013) – Method 1

$$\textbf{Diag(CZZ}^+)_{ii} \approx \left\{ \left[\mathbf{D}_i^r + \mathbf{D}_i^p + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \right) \boldsymbol{\alpha} \right]^{-1} \right\}_{ii}$$

Problem: $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$

- Misztal et al. (2013) – Method 2

$$\textbf{Diag(CZZ}^+)_{ii} \approx \left\{ \mathbf{D} + \left[\mathbf{I} + \mathrm{diag}\left(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \right) \right] \boldsymbol{\alpha} \right\}^{-1}$$

- Implemented in accf90GS   Problem: $\mathbf{A}_{22}^{-1}$ in APY
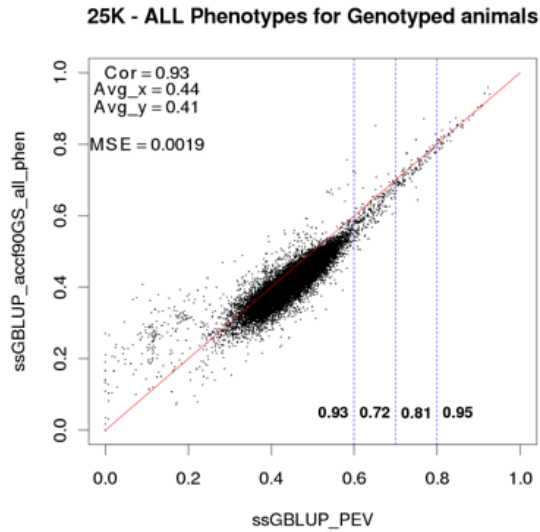
Theoretical Accuracy of GEBV

- Lourenco et al.; Tsuruta et al. (2016)

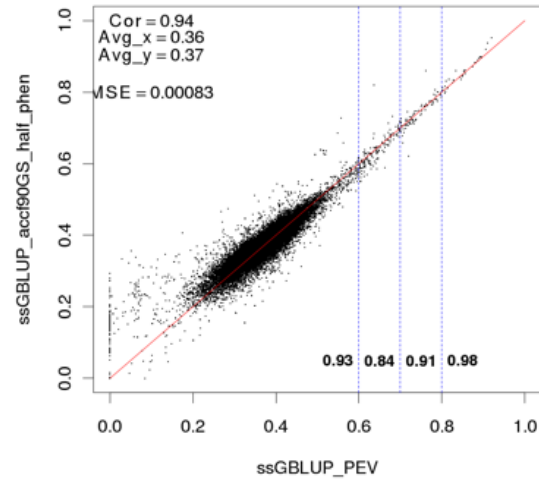$$d_i^g = var\_ratio *[\overline{Rel} + (1 - g_{ii}) + zeta * \overline{Rel - Rel_{PA}}]$$

- Implemented in accf90GS

- Used for commercial evaluations in the US

- Problem: COR(true,approx.) < 0.90 depending on the data

- Problem: genomic contribution relies on genotyping strategy
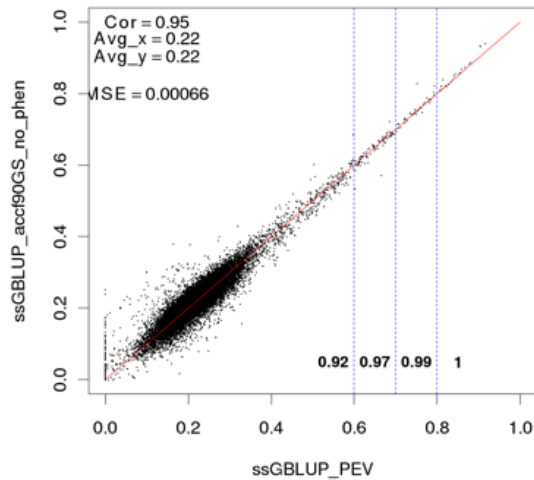
Theoretical Accuracy of GEBV

25K - ALL Phenotypes for Genotyped animals

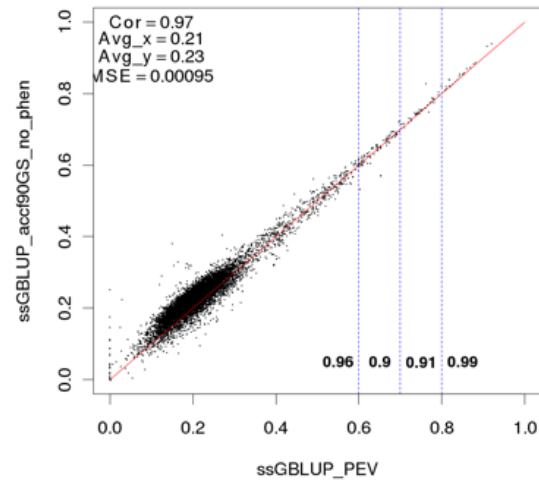Cor = 0.93
Avg_x = 0.44
Avg_y = 0.41
MSE = 0.0019

0.93 0.72 0.81 0.95

4) 14k Phen

25K - HALF Phenotypes for Genotyped animals

Cor = 0.94
Avg_x = 0.36
Avg_y = 0.37
MSE = 0.00083

0.93 0.84 0.91 0.98

3) 7k Phen

25K - NO Phenotypes for Genotyped animals

Cor = 0.95
Avg_x = 0.22
Avg_y = 0.22
MSE = 0.00066

0.92 0.97 0.99 1

2) No Phen

10K - Only genotyped animals never phenotyped

Cor = 0.97
Avg_x = 0.21
Avg_y = 0.23
MSE = 0.00095

0.96 0.9 0.91 0.99

1) No Phen

Theoretical Accuracy of GEBV

- Approximating accuracies of GEBV is still an unsolved issue

- True for all software

- Several approximations and tricks proposed and applied

- Approximation in accf90GS works, but…

- Need more accurate algorithm

Ideas

- Liu et al. (2017)

  - Based on SNP BLUP

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}^{\mu\mu} & \mathbf{C}^{\mu g} \\ \mathbf{C}^{g\mu} & \mathbf{C}^{gg} \end{bmatrix}$$

- Reliability of **DGV**

$$\mathfrak{R}_i^{SNP} = 1 - \mathbf{z}_i \mathbf{C}^{gg} \mathbf{z}_i' \sigma_u^{-2}$$

- ABGG_UGA (2018-2019)

  - Based on $\mathbf{G}_{\mathrm{APY}}^{-1}$ and YAMS

- Solving this issue = high impact research