MILK QUALITY
BREEDING VALUE PREDICTION
BASED ON
FTIR SPECTRA

Tormod ÅDNØY, Theo ME MEUWISSEN, Binyamin DAGNACHEW
Department of Animal and Aquaculture Science (IHA),
University of Life Sciences UMB,
˚As, Norvège
Today:
  - fat%, protein% .. in milk samples found by machine prediction from infrared light (FTIR)
  - variance structure for fat%, protein%, .. estimated using relationship of animals (A)
  - blup breeding values calculated from fat%, protein%, ..

Propose:
  - covariance of FTIR spectra estimated using relationship (A)
    (Needs dimension reduction 500-1000 wavelengths --> 8 principal components)
  - blup breeding values calculated for FTIR wavelengths
  - calculate breeding values for fat%, protein% .. from heritable part of wavelengths (blup)
  - more quality measures from FTIR can be included in breeding
Milk quality

Indirect prediction (IP)

Direct prediction (DP)

FTIR phenotype

\[ \hat{\beta}_{PLS_i} \]

\[ \hat{y}_i \]

REML,BLUP

\[ \tilde{u}_i \]

\[ e_i \]

blup fat\%, ...

FTIR phenotype

FTIR genetic (blup)

FTIR environmental

PCA,REML,BLUP

\[ \hat{\beta}_{PLS_i} \]

\[ \hat{u}_i^* \]

\[ e_i^* \]

blup fat\%, ...

other milk quality

www.umb.no
MOTIVATION

- We have found genetic variability of goat milk FTIR spectra (Dagnachew & Ådnøy, 2011)

similar to

- Genetic variability of cow milk based on MIR spectra (Soyeurt et al., 2010)
  - Indicated substantial amount of genetic variation
  - Show some regions are more heritable than others
- Quality in milk is found from FTIR spectra - usually no other info used
- …Why not use the genetic part of the spectra to predict the genetic part of the traits…?
HERITABILITIES OF FTIR SPECTRA

- OH group of lactose
- Fat A Carbonyl (C=O)
- Fat B C–H of milk fat
- Amide III (protein)
- OPO asym. Stretch (mixture)
- CO stretch (mixture)

Wavenumber (cm⁻¹)

Variance ratio

www.umb.no
MATERIALS AND METHODS – SAMPLES AND GOATS

  - TINE (Norwegian dairies) has four D-labs performing routine FTIR analysis on milk samples
- Total of 28,000 milk sample spectra
  - 14,869 goats (Norwegian Dairy Goat Control)
  - 271 farms
- Of the FTIR wavelengths: 321 selected, 739 removed
  - for physicochemical reasons …
FAT, PROTEIN, LACTOSE% AND FTIR SPECTRA USED

- We used fat%, protein%, and lactose% from Dairy Goat Control (also found from FTIR spectra, but prediction details unknown to us)

- 20,000 FTIR spectra used to find PLS regression models to predict fat, protein and lactose as given above

- 8,000 FTIR spectra used in 10 cross validations of
  - Direct
  - Indirect

methods of predicting breeding values
REGRESSION FTIR-FAT%, .., PCA, VARIANCE COMPONENTS

20000 FTIR spectra used to

- find *regression coefficients* to predict future fat%, .. etc
- find principal components (PC) of spectra (by PCA)*
- and *PC multivariate (co)variance* structure (REML):
  - additive genetic
  - permanent environment
  - residual
- find *variance components of fat%, .. etc* (by back solution of principle component covariances using regression coefficients)
  - to make comparison of blup values from the two methods on equal variance basis
PRINCIPAL COMPONENT ANALYSIS OF FTIR

Available programs for variance component analysis (asreml, Wombat, DMU) will only accept up to 30-40 variables

- **Goal:** to extract a set of fewer components that explain as much variation as possible of the original variation.

- **Result:** 8 components explained >99%

\[ Y \]

\[ T \]

\[ P \]

\[ F \]

- \( Y \) are the FTIR spectra for the \( N \) samples – observations
- \( T \) is values for the new components for the \( N \) samples – score matrix
- \( P \) tells connection between new components and the spectra – loading matrix
- \( F \) – error term

‘pls package’ in R on correlation matrix of \( Y \) to find scores \( T \) considered as new traits
# Principal Components (PC) of FTIR

<table>
<thead>
<tr>
<th>Principal components</th>
<th>% variance explained</th>
<th>Variance ratios of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Genetic</td>
</tr>
<tr>
<td>1</td>
<td>49.63</td>
<td>0.253</td>
</tr>
<tr>
<td>2</td>
<td>35.74</td>
<td>0.382</td>
</tr>
<tr>
<td>3</td>
<td>5.95</td>
<td>0.330</td>
</tr>
<tr>
<td>4</td>
<td>4.74</td>
<td>0.257</td>
</tr>
<tr>
<td>5</td>
<td>1.63</td>
<td>0.371</td>
</tr>
<tr>
<td>6</td>
<td>0.56</td>
<td>0.303</td>
</tr>
<tr>
<td>7</td>
<td>0.47</td>
<td>0.198</td>
</tr>
<tr>
<td>8</td>
<td>0.33</td>
<td>0.205</td>
</tr>
</tbody>
</table>

Total variance explained 99.05 %
INDIRECT PREDICTION (IP) – normal today

8000 FTIR spectra used to

- predict fat% (using found regression coefficients)
- find blup breeding values for fat% using model:

\[ \text{fat}\% = X_b + Z_u + Q_p + e \]

\(X_b\) – fixed effects: herd-testday, kidding season

\(Z_u\) – random animal breeding value with \(\text{var}(u) = A*\sigma^2_A\)

\(Q_p\) – random animal permanent environment \(\text{var}(h) = I*\sigma^2_H\)

\(e\) – random residual \(\text{var}(e) = I*\sigma^2_E\)

(Variance components \(\sigma^2_A, \sigma^2_H, \sigma^2_E\) from 20000 spectra.)

Blup breeding values for fat% predicted as \(\hat{u}\)

Blup breeding values for protein% and lactose% predicted same way (univariate)
DIRECT PREDICTION (DP) – proposed new method

8000 FTIR spectra used to

- find blup breeding values for principal components of FTIR using multitrait model:

  \[ \text{principal components} = X_b + Z_u + Q_p + e \]

  \( X_b \) – fixed effects: herd-testday, kidding season

  \( Z_u \) – random animal breeding value with \( \text{var}(u) = A \cdot \sigma_A^2 \)

  \( Q_p \) – random animal permanent environment \( \text{var}(h) = I \cdot \sigma_H^2 \)

  \( e \) – random residual \( \text{var}(e) = I \cdot \sigma_E^2 \)

Basis for variance components \( \sigma_A^2, \sigma_H^2, \sigma_E^2 \) is 20000 spectra, same as for IP, but multivariate version

- Blup breeding values for fat% found from blup values for principal components using established regression coefficients

Blup breeding values for protein% and lactose% predicted same way
COMPARISON OF IP AND DP – RESULTS

- So now we have blup breeding values for fat%, protein% and lactose% calculated by Indirect Prediction (IP) and Direct Prediction (DP)
  - from same 6000 FTIR spectra
  - based on same variance components

- **Direct Prediction is better than Indirect Prediction of breeding values for milk content when based on FTIR**

<table>
<thead>
<tr>
<th></th>
<th>Indirect prediction (IP)</th>
<th>Direct prediction (DP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fat</td>
<td>Lactose</td>
</tr>
<tr>
<td>Mean_blupvalues</td>
<td>0.0128</td>
<td>-0.0027</td>
</tr>
<tr>
<td>STD_blupvalues</td>
<td>0.2925</td>
<td>0.0844</td>
</tr>
<tr>
<td>Mean PEV</td>
<td>0.0882</td>
<td>0.0062</td>
</tr>
<tr>
<td>Reduction in mean PEV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>0.627</td>
<td>0.649</td>
</tr>
<tr>
<td>Relative genetic gain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

- Not unexpected that univariate analyses are inferior to multivariate
  
  • Multivariate analysis may be better because correlated traits support each other in predictions
  
- The 8 principal components contain more information than fat%, .. alone, and this carries over to the genetic prediction
  
- A trivariate analysis of fat%, protein% and lactose% gives better results than univariate, but not as good as with Direct Prediction.
DISCUSSION

- Finding principle components and particularly estimating their genetic variance is time consuming – but need only be done seldom
  
  • Breeding values for principle components are quick to find
  
  • And so are breeding values of derived traits when regressions are established in calibration

- Breeding values for other traits with genetic information in the FTIR spectra may be derived without estimating variance components for the new traits. Only the phenotypic regression relation is needed
  
  • Coagulation?
  
  • Fatty acids?
  
  • …
DISCUSSION

- Possibility of using spectra also to detect unwanted changes in environment? – Using deviations from predicted environment.
We invite you to:

Regional IGA conference:

Goat Milk Quality

Tromsø, Norway

4 - 6 June 2013
Milk quality
Indirect prediction (IP)

FTIR phenotype
\[ \hat{\beta}_{PLS_i} \]
\[ \hat{y}_i \]
\[ \hat{\nu}_i \]
\[ e_i \]
blup fat%, ..

Direct prediction (DP)

FTIR phenotype

FTIR genetic (blup)

FTIR environmental

\[ \hat{\beta}_{PLS_i} \]
\[ \hat{\nu}_i \]
\[ e_i \]
blup fat%, ..
other milk quality