Building covariance function models with reduced parameters

Martin Lidauer

MiX99 course: test-day models and single step genomic prediction

COURSE DAY 1, April 10th, 2025

Contents

- Modelling longitudinal data
- Principal component analysis
- Covariance functions with reduced rank

Modelling longitudinal data

Random regression models

- Considering a phenotypical expression is repeatedly measured over time, then

  often it is more accurate to consider these as expressions of genetically different, but correlated effects

- Random regressions allow modelling of the correlation structure over time

  → covariance functions, e.g., fitting a Legendre polynomial function of suitable order

- When modelling different phenotypical expressions simultaneously (multi-trait models), then

  → large increase in number of parameters to be estimated

  → often, high correlations between different parameters

A basic single-trait random regression model as presented by Napo

$$y_{ik} = \sum_{j=1}^{p} \phi_{ijk}\beta_{ijk} + \sum_{j=1}^{d} \phi_{jik}\alpha_{ijk} + \sum_{j=1}^{d} \phi_{jik}\gamma_{ijk} + e_{ik}$$

- y_{ik}: test-day record of cow *i* at day *k*
- ϕ_{ijk}: value of Legendre Polynomial
- α_{ijk}: RR coefficients for additive genetic effect
- γ_{ijk}: RR coefficients for nonhereditary animal effect
- e_{ik}: residual

© Luke

… the same model in matrix notation

$$\mathbf{y} = \mathbf{\Phi}_0\boldsymbol{\beta} + \mathbf{\Phi}_0\boldsymbol{\alpha} + \mathbf{\Phi}_0\boldsymbol{\gamma} + \mathbf{e}$$

- Let’s consider for the animal effects a 2nd order Legendre polynomial
- Thus, $\mathbf{\Phi}_0$ contains the 2nd order Legendre polynomial's covariables

  for instance, for 10 standard TD observations (DIM = 15, 45, …,285) of an animal *i*:

$$\mathbf{\Phi}_{0_i} = \begin{bmatrix} 0.71 & -1.16 & 1.32 \\ 0.71 & -0.95 & 0.64 \\ 0.71 & -0.75 & 0.10 \\ 0.71 & -0.54 & -0.32 \\ 0.71 & -0.34 & -0.61 \\ 0.71 & -0.14 & -0.76 \\ 0.71 & 0.07 & -0.78 \\ 0.71 & 0.27 & -0.67 \\ 0.71 & 0.48 & -0.43 \\ 0.71 & 0.68 & -0.06 \end{bmatrix}$$

- and $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the fixed and random regression coefficients

Let’s expand to a multi-trait model with 2 traits

Trait 1: $\mathbf{y}_1 = \mathbf{\Phi}_0\boldsymbol{\beta}_1 + \mathbf{\Phi}_0\boldsymbol{\alpha}_1 + \mathbf{\Phi}_0\boldsymbol{\gamma}_1 + \mathbf{e}_1$

Trait 2: $\mathbf{y}_2 = \mathbf{\Phi}_0\boldsymbol{\beta}_2 + \mathbf{\Phi}_0\boldsymbol{\alpha}_2 + \mathbf{\Phi}_0\boldsymbol{\gamma}_2 + \mathbf{e}_2$

- If records are ordered by traits, we can write the multivariate model as follows:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\end{bmatrix} + \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\gamma}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}$$

where $\mathbf{y}_1$ and $\mathbf{y_2}$ are the vectors of observations for trait 1 and 2

  - $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta_2}$ are the fixed regression coefficients for the phenotypic trajectories for trait 1 and 2
  - and $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha_2}$, and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma_2}$ are the random regression coefficients for animals’ additive genetic, and nonhereditary effect functions

Multivariate model assumptions for 2 longitudinal traits

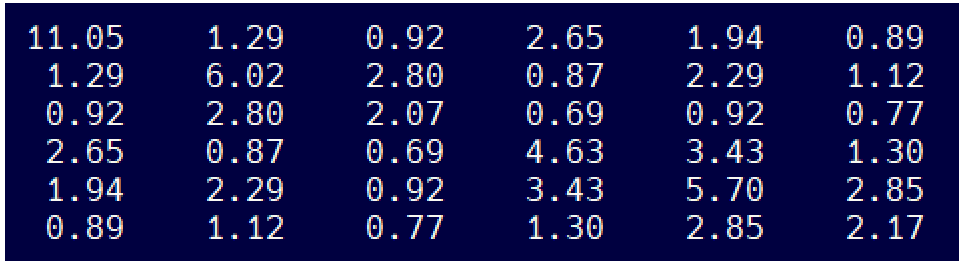
$$\begin{bmatrix}\alpha_{10}\\ \alpha_{11}\\ \alpha_{12}\\ \alpha_{20}\\ \alpha_{21}\\ \alpha_{22}\end{bmatrix} \sim MVN\left(\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}, \mathbf{K}_{\propto} = \begin{bmatrix}\sigma^2_{\alpha_{10}} & \sigma_{\alpha_{1011}} & \sigma_{\alpha_{1012}} & \sigma_{\alpha_{1020}} & \sigma_{\alpha_{1021}} & \sigma_{\alpha_{1022}}\\ & \sigma^2_{\alpha_{11}} & \sigma_{\alpha_{1112}} & \sigma_{\alpha_{1120}} & \sigma_{\alpha_{1121}} & \sigma_{\alpha_{1122}}\\ & & \sigma^2_{\alpha_{12}} & \sigma_{\alpha_{1220}} & \sigma_{\alpha_{1221}} & \sigma_{\alpha_{1222}}\\ & & & \sigma^2_{\alpha_{20}} & \sigma_{\alpha_{2021}} & \sigma_{\alpha_{2022}}\\ & & & & \sigma^2_{\alpha_{21}} & \sigma_{\alpha_{2122}}\\ \text{Sym.} & & & & & \sigma^2_{\alpha_{22}}\end{bmatrix}\right)$$

$$\begin{bmatrix}\gamma_{10}\\ \gamma_{11}\\ \gamma_{12}\\ \gamma_{20}\\ \gamma_{21}\\ \gamma_{22}\end{bmatrix} \sim MVN\left(\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}, \mathbf{K}_{\gamma} = \begin{bmatrix}\sigma^2_{\gamma_{10}} & \sigma_{\gamma_{1011}} & \sigma_{\gamma_{1012}} & \sigma_{\gamma_{1020}} & \sigma_{\gamma_{1021}} & \sigma_{\gamma_{1022}}\\ & \sigma^2_{\gamma_{11}} & \sigma_{\gamma_{1112}} & \sigma_{\gamma_{1120}} & \sigma_{\gamma_{1121}} & \sigma_{\gamma_{1122}}\\ & & \sigma^2_{\gamma_{12}} & \sigma_{\gamma_{1220}} & \sigma_{\gamma_{1221}} & \sigma_{\gamma_{1222}}\\ & & & \sigma^2_{\gamma_{20}} & \sigma_{\gamma_{2021}} & \sigma_{\gamma_{2022}}\\ & & & & \sigma^2_{\gamma_{21}} & \sigma_{\gamma_{2122}}\\ \text{Sym.} & & & & & \sigma^2_{\gamma_{22}}\end{bmatrix}\right)$$

$$\begin{bmatrix}e_1\\ e_2\end{bmatrix} \sim MVN\left(\begin{bmatrix}0\\0\end{bmatrix}, \mathbf{R} = \begin{bmatrix}\sigma^2_{e_1} & \sigma_{e_{12}}\\ \sigma_{e_{21}} & \sigma^2_{e_2}\end{bmatrix}\right)$$

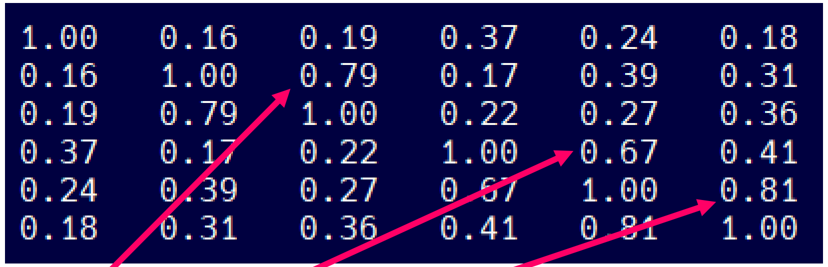
Real data Example: 2 longitudinal traits

- 800 cows with repeated milk yield and dry matter intake observations from first lactation
- Let's apply same model as explained previously:
  - Trait 1, daily milk yield: $\mathbf{y}_1 = \mathbf{X}\mathbf{b}_1 + \mathbf{\Phi}_0\boldsymbol{\beta}_1 + \mathbf{\Phi}_0\boldsymbol{\alpha}_1 + \mathbf{\Phi}_0\boldsymbol{\gamma}_1 + \mathbf{e}_1$
  - Trait 2, daily dry matter intake: $\mathbf{y}_2 = \mathbf{X}\mathbf{b}_2 + \mathbf{\Phi}_0\boldsymbol{\beta}_2 + \mathbf{\Phi}_0\boldsymbol{\alpha}_2 + \mathbf{\Phi}_0\boldsymbol{\gamma}_2 + \mathbf{e}_2$
- Let's look at the estimated variance components for the nonhereditary animal effects
- $\mathbf{K}_\gamma$=



```
11.05     1.29     0.92     2.65     1.94     0.89
 1.29     6.02     2.80     0.87     2.29     1.12
 0.92     2.80     2.07     0.69     0.92     0.77
 2.65     0.87     0.69     4.63     3.43     1.30
 1.94     2.29     0.92     3.43     5.70     2.85
 0.89     1.12     0.77     1.30     2.85     2.17
```

and corr. $\mathbf{C}_\gamma$=



```
1.00    0.16    0.19    0.37    0.24    0.18
0.16    1.00    0.79    0.17    0.39    0.31
0.19    0.79    1.00    0.22    0.27    0.36
0.37    0.17    0.22    1.00    0.67    0.41
0.24    0.39    0.27    0.67    1.00    0.81
0.18    0.31    0.36    0.41    0.81    1.00
```

we find moderate to high correlations
between regression coefficients....

Modelling considerations

- Large number of animal-specific regression coefficients will increase number of equations to be solved
- High correlations between coefficients may lower speed of convergence
- Large number of animal equations may prohibit to use more efficient preconditioner
- Can we develop an equivalent model with better solving properties?

Factor analytical model

  - have fast convergence
  - not necessarily reduction in number of equations

Principal component analyses

  - the variance-covariance structure is described by an equivalent set of eigenvalues and eigenfunctions
  - uninformative eigenfunctions can be discarded → reduced fit → reduces number of equations
  - improves convergence

Principal component analysis

Coefficient matrix of covariance function

Considering for the previous 2-traits example:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad \text{and } \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_0 \end{bmatrix}$$

then, $var(\mathbf{y}) = \boldsymbol{\Phi}\ \mathbf{A} \otimes \mathbf{K}_{\propto}\boldsymbol{\Phi}' + \boldsymbol{\Phi}\ \mathbf{I} \otimes \mathbf{K}_{\gamma}\boldsymbol{\Phi}' + \mathbf{R}$

were $\mathbf{K}_{\propto}$ and $\mathbf{K}_{\gamma}$ are the coefficient matrices of the two covariance functions

Let's decompose the 6x6 $\mathbf{K}_{\gamma}$ coefficient matrix:

- Decomposition should be done on the correlation matrix, NOT on the variance-covariance matrix
- First, we build a diagonal matrix with the standard deviations: $\mathbf{S} = \left(diag(\mathbf{K}_{\gamma})\right)^{0.5}$
- and the correlation matrix: $\mathbf{C}_{\gamma} = \mathbf{S}^{-1}\mathbf{K}_{\gamma}\mathbf{S}^{-1}$

© Luke

Principal component decomposition

- We decompose the correlation matrix into principal components: $\mathbf{C}_{\gamma} = \mathbf{VDV}'$

  where **V** is the eigenfunction matrix and **D** is the matrix with the eigenvalues

- For our example we obtain:

$$\mathbf{V} = \begin{bmatrix} 0.02 & 0.06 & 0.41 & 0.85 & 0.19 & 0.26 \\ -0.40 & 0.59 & -0.09 & 0.02 & -0.58 & 0.39 \\ 0.39 & -0.59 & -0.10 & 0.09 & -0.58 & 0.39 \\ -0.29 & -0.18 & -0.73 & 0.15 & 0.41 & 0.40 \\ 0.64 & 0.41 & 0.04 & -0.28 & 0.30 & 0.50 \\ -0.45 & -0.33 & 0.53 & -0.40 & 0.19 & 0.46 \end{bmatrix} \text{ and } \mathbf{D} = \begin{bmatrix} 0.06 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.51 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.93 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.31 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.92 \end{bmatrix}$$

- Inspecting the eigenvalues we find:

  → the smallest eigenvalue, $d_1 = 0.06$, explains only 0.98% of the total variance

Luke

Considering only significant eigenvalues

- Often, when assessing eigenvalues of high dimensional variance-covariance matrices, many very small eigenvalues can be found
- Removing very small eigenvalues (and associated eigenfunctions) improves the quality of the model
- Let’s consider that the CF needs to explain at least 99% of the original total variance,

  then we can drop one eigenvalue of our $\mathbf{K}_\gamma$ matrix:

$$\mathbf{V}^* = \begin{bmatrix} 0.06 & 0.41 & 0.85 & 0.19 & 0.26 \\ 0.59 & -0.09 & 0.02 & -0.58 & 0.39 \\ -0.59 & -0.10 & 0.09 & -0.58 & 0.39 \\ -0.18 & -0.73 & 0.15 & 0.41 & 0.40 \\ 0.41 & 0.04 & -0.28 & 0.30 & 0.50 \\ -0.33 & 0.53 & -0.40 & 0.19 & 0.46 \end{bmatrix} \text{ and } \mathbf{D}^* = \begin{bmatrix} 0.27 & 0 & 0 & 0 & 0 \\ 0 & 0.51 & 0 & 0 & 0 \\ 0 & 0 & 0.93 & 0 & 0 \\ 0 & 0 & 0 & 1.31 & 0 \\ 0 & 0 & 0 & 0 & 2.92 \end{bmatrix}.$$

Thus, the original correlation matrix can be described as: $\mathbf{C}_\gamma^* = \mathbf{V}^*\mathbf{D}^*\mathbf{V}^{*\prime}$

Covariance function with reduced rank

Deriving covariance function (CF) with reduced rank

- Further, we can manipulate the reduced rank eigenfunction $\mathbf{C}^*_\gamma = \mathbf{V}^*\mathbf{D}^*\mathbf{V}^{*\prime}$ to get in place of $\mathbf{D}^*$ an identity matrix: $\mathbf{C}^*_\gamma = \mathbf{V}^*\mathbf{D}^*\mathbf{V}^{*\prime} = \mathbf{V}^*\mathbf{D}^{*0.5}\mathbf{I}\mathbf{D}^{*0.5}\mathbf{V}^{*\prime} = \mathbf{V}^*_\mathrm{d}\mathbf{I}\mathbf{V}^{*\prime}_\mathrm{d}$
- And putting all together we get: $\mathbf{\Phi}\mathbf{K}_\gamma\mathbf{\Phi}' = \mathbf{\Phi}\mathbf{S}\mathbf{C}_\gamma\mathbf{S}\mathbf{\Phi}' = \mathbf{\Phi}\mathbf{S}\mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{S}\mathbf{\Phi}'$

$$\sim \mathbf{\Phi}\mathbf{S}\mathbf{V}^*\mathbf{D}^*\mathbf{V}^{*\prime}\mathbf{S}\mathbf{\Phi}' = \mathbf{\Phi}\mathbf{S}\mathbf{V}^*_\mathrm{d}\mathbf{I}\mathbf{V}^{*\prime}_\mathrm{d}\mathbf{S}\mathbf{\Phi}' = \mathbf{\Phi}\mathbf{E}^*\mathbf{E}^{*\prime}\mathbf{\Phi}' = \mathbf{U}\mathbf{U}'$$

- Deriving also covariance function for $\mathbf{\Phi}\mathbf{K}_\alpha\mathbf{\Phi}' \sim \mathbf{\Lambda}\mathbf{\Lambda}'$ we can reformulate the original model:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\end{bmatrix} + \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\gamma}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}$$

$$\sim \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{\Phi}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \mathbf{\Lambda}\boldsymbol{\delta} + \mathbf{U}\boldsymbol{\nu} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}$$

Note, an animal has the same set of random regression coefficients for both traits

Covariate matrix and (co)variance matrix for nonhereditary animal effect

where **U** for an animal with 10 milk yield and 10 dry matter intake observations is:

$$\begin{array}{r} \text{Milk}\left\{\vphantom{\begin{matrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\end{matrix}}\right. \\ \text{Intake}\left\{\vphantom{\begin{matrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\end{matrix}}\right. \end{array} \mathbf{U} = \begin{bmatrix} -1.37 & 0.73 & 2.05 & 1.15 & 0.39 \\ -0.92 & 0.77 & 1.97 & 1.45 & 0.08 \\ -0.53 & 0.79 & 1.92 & 1.64 & -0.11 \\ -0.19 & 0.80 & 1.87 & 1.70 & -0.16 \\ 0.09 & 0.80 & 1.85 & 1.65 & -0.11 \\ 0.31 & 0.78 & 1.84 & 1.46 & 0.07 \\ 0.47 & 0.75 & 1.84 & 1.14 & 0.40 \\ 0.58 & 0.71 & 1.87 & 0.71 & 0.84 \\ 0.63 & 0.65 & 1.91 & 0.14 & 1.41 \\ 0.61 & 0.58 & 1.96 & -0.54 & 2.10 \\ -1.07 & -0.12 & 0.23 & 0.18 & 0.20 \\ -0.79 & -0.49 & 0.48 & 0.14 & -0.15 \\ -0.55 & -0.79 & 0.66 & 0.13 & -0.37 \\ -0.34 & -1.01 & 0.76 & 0.17 & -0.43 \\ -0.16 & -1.16 & 0.79 & 0.24 & -0.35 \\ -0.02 & -1.23 & 0.75 & 0.36 & -0.12 \\ 0.09 & -1.23 & 0.62 & 0.53 & 0.29 \\ 0.17 & -1.16 & 0.43 & 0.73 & 0.82 \\ 0.21 & -1.01 & 0.16 & 0.98 & 1.53 \\ 0.22 & -0.79 & -0.18 & 1.26 & 2.37 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{I}_\gamma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right)$$

Breeding values

Calculation of lactation yield breeding values for animal *i*:

- Original model
  - $BV_{305d\_milk_i} = 30.5 \times \mathbf{I}'\boldsymbol{\Phi}_0\hat{\boldsymbol{\alpha}}_{1_i}$
  - $BV_{305d\_drymatterintake_i} = 30.5 \times \mathbf{I}'\boldsymbol{\Phi}_0\hat{\boldsymbol{\alpha}}_{2_i}$
- Covariance function model
  - $BV_{305d\_milk_i} = 30.5 \times \mathbf{I}'\boldsymbol{\Lambda}_{\mathrm{milk}}\hat{\boldsymbol{\delta}}_i$
  - $BV_{305d\_drymatterintake_i} = 30.5 \times \mathbf{I}'\boldsymbol{\Lambda}_{\mathrm{drymatterintake}}\hat{\boldsymbol{\delta}}_i$
- Converting CF coefficients to original RR coefficients for the nonhereditary animal effect
  - $\begin{bmatrix} \hat{\boldsymbol{\gamma}}_{1_i} \\ \hat{\boldsymbol{\gamma}}_{2_i} \end{bmatrix} \sim \begin{bmatrix} \hat{\boldsymbol{\gamma}}^*_{1_i} \\ \hat{\boldsymbol{\gamma}}^*_{2_i} \end{bmatrix} = \mathbf{E}^* \ \hat{\boldsymbol{\nu}}_i$

© Luke

Correlations between breeding values

Covariance functions

- No rank reduction

		Original model		CF model	
		Milk	Intake	Milk	Intake
Original model	Milk	1.000	0.901	**1.000**	0.901
	Intake		1.000	0.901	**1.000**
CF model	Milk			1.000	0.901
	Intake				1.000

Covariance function with reduced rank

- Smallest eigenvalue of $\mathbf{K}_\gamma$ $(d_1 = 0.06)$ discarded

		Original model		CF model	
		Milk	Intake	Milk	Intake
Original model	Milk	1.000	0.901	**0.998**	0.887
	Intake		1.000	0.906	**0.995**
CF model	Milk			1.000	0.898
	Intake				1.000

Thank you