Comparison of different strategies in calculating calibration models:

example of milk analysis

Stefania Barzaghi1\*, Giusi Poerio1,

1 CREA- Centro di Zootecnia e Acquacoltura, [stefania.barzaghi@crea.gov.it](mailto:stefania.barzaghi@crea.gov.it); giusi.poerio@gmail.com

\*Corresponding author

In 2009, Filtzmoser et al., proposed the “Repeated double cross validation” (rdCV), in calculating calibration models, to make a realistic estimate of the prediction errors of any future sample.

The aim of the present study was to compare the performances of calibration models, from milk samples’ NIR spectra, calculated using different strategies: A) the optimal number of latent variables was obtained with the rdCV, using the “Chemometrics” package of R; B) the calibration model was calculated by sorting the samples into two sets, one for calibration and one for validation, chosen with the Kennard-Stone algorithm, using the PLS Toolbox software (Eigenvector Research, USA)

105 milk samples taken both from single cows and from bulk milk, were collected in the months of March and April 2021 to be used in the calculation of the calibration models. The performance of the models obtained, was evaluated on 35 bulk milk samples, from different farms, collected in a subsequent period, between May and September 2021. The milk samples were analyzed with Milkoscan (Foss Italia, Padua) to quantify the content in fat, protein, lactose and total solids, and scanned with a ProxiMate TM spectrometer (Büchi Labortechnik AG, Flawil (CH)) in the spectral range from 400 to 1700nm. NIR spectra were recorded in transflectance mode, and the samples (1mm thick) were placed in rotating Petri dishes.

Applying both strategies, in quantifying the different components of milk, the optimal number of latent variables was the same, except for the lactose content. In the prediction of independent samples, the best performances were obtained for the models calculated with strategy B. For the fat content, the RMSEP of the model from A was 0.55 while for the model B was 0.13. For the protein content the RMSEPs were respectively equal to 0.39 for A and 0.29 for B.

**Keywords:** Repeated double cross validation, calibrations, milk, grasso, fat, proteins, lactose

**Acknowledgements:** This study was carried out within the Project AGRI HUB Sviluppo ed integrazione tecnologica di una piattaforma high-throughput per il miglioramento sostenibile dei processi produttivi delle filiere dell’agroalimentare – Sub-task 3.2 Influenza delle innovazioni sul benessere animale e sulla qualità del latte.

REFERENCES

P. Filzmoser, B. Liebmann, K. Varmuza. 2009 Repeated double cross validation, J. Chemom. 23, 160-171. https://doi.org/10.1002/cem.1225