Modular Open-Source Framework for Spectral Phenotyping Using Neighbourhood-Adaptive PLSR to Predict Nitrogen Use Efficiency in a Dairy Herd

Muers Ross*[1], Salavati Mazdak[1], Newbold John[1], Dewhurst Richard[1]

[1] Dairy Research & Innovation Centre, School of Veterinary Medicine and Biosciences, Scotland's Rural College (SRUC)

Mid-infrared (MIR) milk spectra are widely used for routine phenotyping in the dairy industry, yet predictive modelling approaches remain difficult to implement in practice due to technical and reproducibility barriers. Conventional calibration models, most commonly partial least squares regression (PLSR), are favoured for their mechanistic interpretability. However, global PLSR models are driven by the dominant covariance structure of the calibration datasets, predicting best for samples at the population centre. Peripheral samples are often poorly represented by a global calibration and consequently predicted with greater error, despite being of greatest biological and management interest, motivating local models built with spectrally similar observations. This study presents an open-source modelling framework to lower technical barriers to spectral phenotyping that aims to enhance translation of milk spectral data into actionable decision-support insights and introduces a promising novel neighbourhood-adaptive local PLSR approach.

We introduce the Spectral Phenotyping Prediction Framework (SPPF), a modular open-source Python package integrating spectral preprocessing, model training, and validation within a reproducible workflow. It supports both global and local modelling strategies. Neighbourhood-Adaptive PLSR (NA-PLSR) constructs sample-specific calibration subsets based on spectral similarity, optimising neighbourhood structure alongside PLS hyperparameters. The framework was applied to MIR milk spectra to predict nitrogen use efficiency (NUE), a key sustainability indicator in dairy systems.

NA-PLSR delivered modest overall gains (R2test=0.66 vs 0.63), but showed its strongest advantage at peripheral NUE ranges, reducing RMSE for outlying samples by up to 18.6%. Distributional fidelity analyses also indicate NA-PLSR predictions are closer aligned with measured NUE distributions, with 6–10% improvements across KS, Jensen-Shannon, and Wasserstein metrics. Locally adaptive spectral modelling improved prediction of a complex trait compared to a conventional global model, particularly for samples poorly represented in global calibration models. The open-source SPPF enables reproducible model development and wider adoption of advanced spectral phenotyping, supporting improved use of routine milk spectral data.