Abstract Submission Form

Title (Mr./Mrs/Dr./Prof.)   Prof

Presenting author   Helene Soyeurt

Institute

Institute/company: University of Liège - Gembloux Agro-Bio Tech

Adress: Passage des déportés 2

ZIP/Postal code: 5030   City: Gembloux

Country: Belgium

Insert all authors and institutions

C. Nickmilder (1), J. Leblois (2), O. Christophe (3), C. Grelet (3), Holicow Consortium (4), H. Soyeurt (1)
(1) ULiège Gembloux Agro-Bio Tech, Passage des Déportés 2, 5030 Gembloux, Belgium
(2) Walloon Breeder Association, Champs Elysées 4, 5590 Ciney, Belgium
(3) Walloon Research Centre, Chée de Namur 24, 5030 Gembloux, Belgium
(4) https://holicow.nweurope.eu/

Preferred presentation

Oral

Preferred session

Session 1: WG Animal Data Exchange – Decision Support Tools of the Future – Promoting Sustainability Farm Management

Email of corresponding author   hsoyeurt@uliege.be

Title of your paper   How spectrally representative are datasets used to build MIR-based predictive models ? A data-driven study.

Insert ABSTRACT text

Milk mid infrared (MIR) spectra have been used to develop models for predicting various traits, like fatty acids (FA). To ensure a good quality of the obtained MIR predictions, the collection of samples used to create the calibration set is supposed to be representative of the studied cow population. Yet this statement is not verified. This check is crucial in Holicow Interreg Project as MIR spectra coming from different countries are aggregated. So, this study proposed a method driven by real data to solve this issue. Two databases were used. The first one included 163,657 spectra representing the cow population (DB1), the second contained 2,000 FA references and their MIR spectra (DB2). To assess if DB2 is spectrally representative of DB1, spectra were first decomposed into 3 principal components (PC). Normal distribution of data along each PC axis allow a simple comparison: is the sphere defined by DB2 records encapsulating the one defined by DB1 records? However, the actual distribution of data requires to prefer convex hulls over spheres. Volume of DB2 represented 7.43% of the DB1 volume without considering the density of DB1 records in the studied population. However, DB2 covered 99.44% of the DB1 density

weighted records, the rest being extreme cases. It indicates that FA equations derived from DB2 should be suited for the vast majority of the 41 million of milk spectra included in the Holicow DB.

Enter keywords milk, spectrum, mid-infrared, calibration, fatty acid