A comparison of data-processing pipelines for methane concentration measurements from sniffers

Manzanilla-Pech Coralia*[1], van Breukelen Anouk[1], Milkevych Viktor[2], Gonzalez-Recio Oscar[3], Gredler-Grandl Birgit[1], Veerkamp Roel[1], de Haas Yvette[1]

[1] Animal Breeding and Genomics, Wageningen University and Research, [2] Center for Quantitative Genetics and Genomics, Aarhus University, [3] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria–CSIC. current affiliation:The Roslin Institute and Royal (Dick) School of Veterinary Studies R(D)SVS, The University of Edinburgh

Reducing methane emissions from dairy production is an important component of global greenhouse gas mitigation efforts. Genetic selection offers a promising long-term strategy for lowering emissions, but it requires large-scale and reliable methane phenotyping. In dairy systems, methane concentrations (CH_4c) are recorded using breath analyzers, commonly known as "sniffers," installed in automatic milking systems (AMS). These devices collect gas samples every second or every five seconds, depending on the manufacturer, throughout a milking session that typically lasts five to ten minutes. Sniffers also measure carbon dioxide concentrations (CO_2c). Because sniffers do not include animal identification capabilities, measurements must be linked to individual cows by aligning sniffer timestamps with AMS milking timestamps. However, due to the large volume of data collected and the differences between timestamp systems, this approach is challenging in practice. As a result, several research groups have developed different pipelines and software tools to align, clean, and process CH_4 and CO_2 data. There is an increasing interest in comparing the performance of these pipelines, yet no such comparison has been conducted to date. In this study, we compare two pipelines and a software program used to process CH4 and CO2 measurements in terms of alignment and editing steps.

The first pipeline (P1) was developed at Wageningen University and Research, the Netherlands, in Python. The second approach is the published GEDA program (P2), developed at Aarhus University, Denmark. The third pipeline (P3), originally written in Java, is used by different research groups (including INIA, Spain) but has not been formally published. All three pipelines process raw sniffer data to derive visit-level mean values that represent the average gas concentration recorded during each milking event. Only farms with a repeatability higher than 10 percent for both gases were retained for analysis. Repeatability was estimated by fitting a mixed model with cow as a random effect, using only cows with at least three daily records, and calculating the proportion of variation attributable to differences between cows. In total, data from 23 Dutch dairy farms were used, comprising records collected in 2022 of 2,009 cows and resulting in 49,951 daily observations, with gas concentrations sampled every five seconds during milking. Pearson correlations between the CH4c phenotypes produced by the different pipelines (P1 and P2) ranged from 0.56 to 0.86, depending on the farm. Results from the comparisons of the three pipelines will be presented and discussed at the conference.