**OREI Report**

Descriptive analysis showed that the mean (± standard deviation) concentrations for aerobic plate count (APC), mesophilic spore count (MSC), thermophilic spore count (TSC), and psychrotolerant spore count (PSC) are 3.18 (±0.67), 0.77 (±0.69), -1.51 (±0.67), 0.41 (±0.70) log10CFU/mL, suggesting similar variability across all types of microbial counts between farms (**Figure 1**). To investigate the effects of farm management and weather on the spore levels, we continued to develop three random forest models for all spore types (i.e., MSC, TSC, PSC). Prior to the model development, principal component extraction was performed on all weather-related variables to minimize the collinearity and reduce the dimensionality to prevent model overfitting. Near-zero variance filter was applied as a regularization technique to all variables to remove highly imbalanced predictor variables. Using pre-processed data, random forest regressions were performed for log10MSC and log10TSC while random forest classification was performed for presence of PSC, due to a large proportion of negative results (i.e., approximately 28%) from Most Probable Number (MPN) for PSC. The results from random forest analysis showed that only a minor proportion of variance in the spore levels can be explained by the farm management and weather variables, as indicated by the R2 and classification accuracy from cross-validated training results. Specifically, random forest models for predicting log10MSC and log10TSC have R2 of 0.21 and 0.24, respectively while the random forest model for classifying the detectable PSC has an accuracy of 0.70. Overall, these results suggest that current approach for describing farm management styles was not able to sufficiently capture the dynamics between farm practices and spore levels, and therefore incapable of predicting the spore levels for practical use. To further extrapolate utility from the model outcome, we proceeded to construct the variable importance plots (VIPs) from all three models (**Figure 2**). All three VIPs showed that farm size (e.g., cow number, full employee number), farm location, certification year, APC variability within a farm, and principal component 1 (i.e., highly correlated with temperature) were consistently ranked high for predicting spore levels. Some practices were also shown important to spore levels. For example, clipping/flaming udder and using machine to dry towels were important for predicting log10MSC. However, due to the marginal variance each predictor variable can explain, while ranked high in VIP, changes in these variables alone could not lead to substantial change in the spore levels. This was confirmed by partial dependence plot (PDP). For example, although clipping/flaming udders is highly important for predicting log10MSC, applying this practice is expected to reduce the spore concentration by less than 0.1 log10CFU/mL (**Figure 3**). Further analysis is needed to confirm if any interactions between two variables will lead to a larger impact.

**Figure 3**. Partial dependence plot showing the relationship between udder clipping/flaming and log10MSC

Chart, histogram

Description automatically generated

**Figure** 1. Spore concentration by test (expressed as log10 CFU/mL for APC, MSC and TSC, and log10 MPN/mL for PSC) summarized over all farms and sampling time points. The farthest left bin for MSC, TSC, and TSC represents the samples that have spore concentrations below detection limits.

Chart

Description automatically generated

**Figure 2.** Variable importance plots for cross-validated models constructed using data from all farms (A = random forest regression for log10 concentration of MSC; B = random forest regression for log10 concentration of TSC; C = random forest classification for presence of PSC)