**OREI Report**

**Descriptive statistical analysis**



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.

**Random forest analysis**

Table 1. Cross-validation results for 1) random forest regression models predicting the log10 concentrations of MSC and TSC and 2) random forest classification models grouping the presence of PSC in organic raw milk samples.

|  |  |  |  |
| --- | --- | --- | --- |
| Spore test | Dataset1 | Metrics | Performance |
| MSC | All | R-squared | 0.205 |
| With parlor | 0.231 |
| Without parlor | 0.190 |
| With pasture time | 0.187 |
| Without pasture time | 0.150 |
| TSC | All | 0.243 |
| With parlor | 0.382 |
| Without parlor | 0.172 |
| With pasture time | 0.167 |
| Without pasture time | 0.335 |
| PSC | All | Accuracy | 0.703 |
| With parlor | 0.675 |
| Without parlor | 0.784 |
| With pasture time | 0.701 |
| Without pasture time | 0.688 |

1 Dataset containing: All = all farms; With parlor = farms with a parlor; Without parlor = farms without a parlor; With pasture time = farms at seasons when pasture is allowed; Without pasture time = farms at seasons when pasture is not allowed.



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)



Figure 3. Random forest regression for log10 concentration of MSC, constructed using datasets that contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed



Figure 4. Random forest regression for log10 concentration of TSC, constructed using datasets that contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed



Figure 5. Random forest classification for presence of PSC, constructed using datasets that contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed



Supplementary figure 1. Weather variables strongly associated with top 5 principal components obtained from principal component analysis (PCA) performed as data preprocessing for datasets that contain all farms with spore tests for A) MSC, B) TSC, and C) PSC.

**

Supplementary figure 2. Weather variables strongly associated with top 5 principal components obtained from principal component analysis (PCA) performed as data preprocessing for datasets with MSC and contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed



Supplementary figure 3. Weather variables strongly associated with top 5 principal components obtained from principal component analysis (PCA) performed as data preprocessing for datasets with TSC and contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed



Supplementary figure 4. Weather variables strongly associated with top 5 principal components obtained from principal component analysis (PCA) performed as data preprocessing for datasets with PSC and contain A) farms with a parlor, B) farms without a parlor, C) farms at season when pasture is allowed, and D) farms at season when pasture is not allowed