Statistical Analysis of Skim Milk by Direct Analysis in Real-Time

Robert Goguen[1], Crystal Hart[1] , Joseph LaPointe[1] and Brian Musselman[1]

[1] IonSense, Inc., Saugus, MA, USA

Introduction

Statistical modeling is a powerful analytical technique that allows an end user to input data and receive a specific answer back. In this application statistical modeling has been examined as a means to accurately identify whether or not skim milk is as identified and if the milk meets certain standards.

The utility of direct analysis in real time (DART)-mass spectrometry is that it permits the determination of a single spectrum for representation of the identity in seconds per sample following a simple dilution with water and application of a 3ul aliquot of material to the sample analysis screen. Generation of the DART-MS data requires less than 10 seconds per sample. The brevity of this analysis permits collection of 30 – 40 mass spectra, a statistically relevant sample set, in under 10 minutes with 96 spectra achievable in an automated manner in under 20 minutes.

Methods

Milk Samples:

Three different types of milk were used to create the models; skim, whole, and soy milk. All three milks purchased were from the same brand and had sell by dates as close to each other as possible. The milk samples were prepared by diluting the milk to 1% with water.

DART Optimization:

To obtain the richest spectrum from milk certain parameters of the DART can be optimized, the main one being temperature. For this optimization, 3µl spots of the 1% milk solution were spotted onto QuickStrip cards. These cards were then analyzed with the DART, varying the temperature as different spots passed into the ionization region. The DART rail speed and spot size were also optimized.

Optimized Parameters:

- DART Gas: He
- DART Temperature: 400°C
- Rail Speed: 0.5mm/sec.
- Spot Size: 3µl

DART analysis:

To collect enough samples to create models, the DART was used in combination with the XZ Transmission module (Figure 1). This module allowed 96 samples to be collected in about 40 minutes. This longer total run time was due to a slower rail speed determined during optimization. Final sample time was about 25 seconds per sample. Milk samples were collected in sets of 30 allowing a set of each milk to be collected in a single run with the DART.

FIGURE 1. – Screen holder assembly with steel mesh. A 12-tip pipette spotting milk samples. The assembly being lowered into the XZ transmission module

Milk Spectrum

The same major ions were present in all three milks tested but in different ratios. Skim milk and whole milk can only be differentiated visually by zooming in on specific areas where the triglycerides are present. These similarities make statistical modeling an ideal choice for skim milk identification.
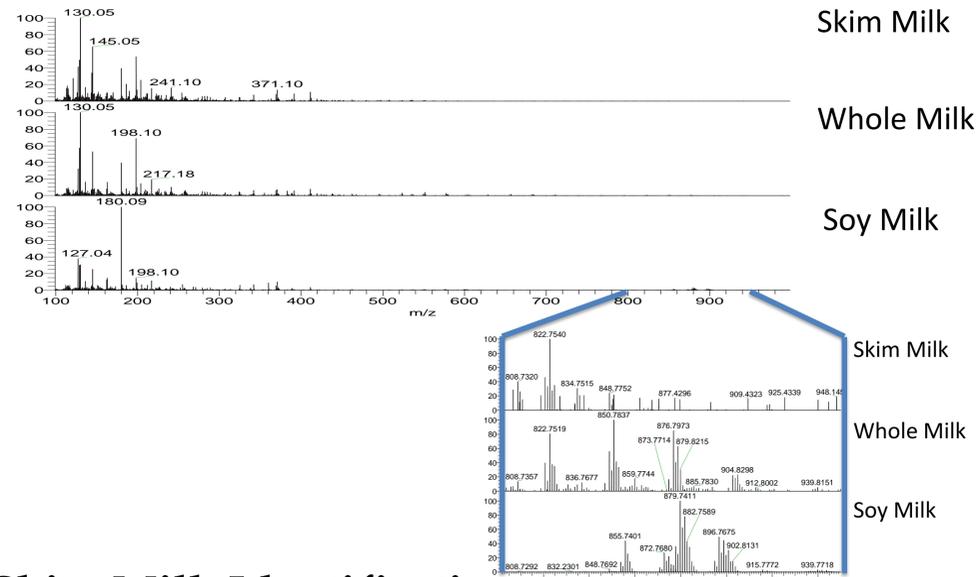
Skim Milk Identification

To create a robust model, many milk samples were needed. 30 samples were collected every week for each milk (270 total) over a three week period. Initially the degradation of the milk over time was also going to be looked at, but it was determined that no major changes in the milk spectrum occurred and all three weeks could be used to create one model.

The set of 270 samples were collected and input into the AnalyzeIQ software(Analyze IQ Ltd., Galway, Ireland). Many different models were tested using a 5-run 10-fold cross-validation method. The table below shows models that had a training error rate less than 5%. The models were tested against a set of unknowns collected on a different day to give the testing error rate.

Target	Model	Parameters	Training Error Rate	Testing Error Rate
Skim Milk	SAV	Default	267/270	7/9
	SVM – Polynomial	Exponent 1	258/270	9/9
	Linear Regression	---	268/270	8/9
Soy Milk	SAV	Default	268/270	9/9
	SVM - Polynomial	Exponent 1	266/270	9/9
	SVM - Weighted Spectral Linear	Default	266/270	8/9
	Linear Regression	---	262/270	7/9
	kNN	k = 3	269/270	9/9
Whole Milk	SAV	Default	268/270	8/9
	kNN	k = 3	265/270	7/9
	Linear Regression	---	265/270	6/9

Milk Antibiotics

To maintain the health of their livestock, farmers may give their animals antibiotics to thwart viral or bacterial threats. These antibiotics may find their way into the milk of these animals. These contaminated milks must be monitored to prevent them from entering a consumer market.

To simulate a contaminated milk sample, three antibiotics were spiked into skim milk – thiabendazole, fenbendazole, and mebendazole. 90 samples of the spiked milk were run along with 90 samples of unadulterated milk. These samples were used to construct models and then tested against spiked milk samples run on a different day.
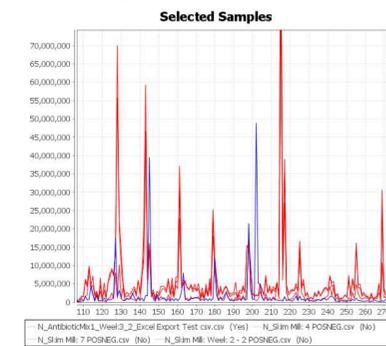
FIGURE 2. Shows a mass spectral overlay of a selected mass region of four skim milks. The three in red are skim milk and the one in blue is skim milk spiked with antibiotics. The thiabendazole(m/z=202) peak can clearly be seen in the blue spectrum.

Model	Parameters	Training Error Rate	Testing Error Rate
kNN	k = 3	179/180	9/9
Linear Regression	---	174/180	9/9
SVM – Polynomial	Exponent 1	173/180	9/9
SAV	---	179/180	9/9
SVM – Weighted Spectral Linear	Default	174/180	9/9

Conclusions

- With statistical modeling in combination with DART-MS, it is possible to quickly identify skim milk from other milk varieties. Some models had a 0% error rate when tested against unknowns, such as the SVM – Polynomial Skim Milk model.
- Statistical modeling allows the quality of milk to be maintained by indicating if a sample matches the uncontaminated milk model.

Future Research

The next step in this process would be to build a more robust milk contamination model. This would include improving the training error rate of the overall model and then improving the output of the model. Instead of just saying the milk is contaminated, the model could give what polarity caused the sample to fail or automatically display the area that caused the failure.

The milk identification model can also be improved. Through our testing it was proven that different brands of skim milk give different spectra. The model could be adjusted to not only identify if a milk is skim but also what brand of milk a sample is.