New technologies – MIR in milk

Jennie Pryce¹,², Tingting Wang¹ and Michelle Axford³
¹Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083
²La Trobe University, Bundoora, VIC 3086
³DataGene, Bundoora, VIC 3083

Summary
Mid-infrared (MIR) spectral data is used routinely by milk recording organisations to estimate milk fat and protein composition in regular herd-testing milk samples. However, it is becoming a popular way to predict complex phenotypes such as energy balance and methane emissions in some European countries. The process is to use the absorbance of mid-infrared light through milk samples over a range of wavelengths to predict the phenotype, requiring a reference population and MIR data. The advantage with using MIR to predict these phenotypes, is that the turnaround time back to farm can be fast and at little extra cost, enabling reactive management decisions. Here we report on the accuracy of MIR prediction of energy balance, methane emissions and milk fatty acids. The accuracy ($r^2$) of predicting energy balance using MIR from Australian cows was around 0.4, which is similar to the performance of applying UK MIR prediction equations to Australian data. The prediction accuracies of methane emissions and milk fatty acids were greater than 0.5. The next phase will focus on increasing the accuracy of predicting complex traits in addition to validating them in commercial populations.

Background
Milk composition is affected by several factors (e.g. breed of cow, diet, stage of lactation and even environment). It is also indicative of not only nutritional quality but also can provide information on important phenotypes of the cow's own health and performance. Milk recording organisations worldwide provide farmers with basic composition of milk (fat and protein percentage) in addition to an estimate of the quantity of somatic cells, which is a good indicator of mastitis. However, there is an even greater opportunity to provide farmers and milk processors with even greater insights into the nutritional qualities of individual samples of cow's milk in addition to providing risk assessments for disease. This is through using mid-infrared spectral data from passing light through milk samples, which are already available commercially from milk recording organisations.

Mid-infrared (MIR) spectral data is the absorptions of infrared rays at frequencies correlated to the vibration of specific chemical bonds within a molecule (Soyeurt et al., 2011). MIR Prediction uses the absorbance of mid-infrared light through milk samples over a range of wavelengths to predict the phenotype. Calculating a prediction equation that can be applied nationally, requires a reference population that has phenotypes of interest and MIR data collected at the same time as the phenotype data. The advantage with using MIR to predict these phenotypes is that the turnaround time back to farm can be fast and at little extra cost, enabling reactive management decisions. To date, several research organisations have already done MIR prediction analysis on very detailed milk fat and protein composition with promising prediction accuracies (e.g. Soyeurt et al., 2011). Furthermore, it has also been used to predict complex phenotypes, such as energy balance (McParland et al., 2011) and methane emissions (Dehareng et al., 2012). More recently, the UK's National Milk Records, a herd-recording organisation, are publishing "at risk cows" based on their MIR predictions for energy balance.

As cows mobilise body fat in early lactation to sustain lactation (when feed requirements exceed intake), it is likely that special signatures of fatty acid composition are also observed in milk. There are several milk fatty acids in milk. Having a greater understanding of how these are associated with energy balance may help to improve MIR predictions.
and could be worth breeding for in their own right. Negative energy balance results in impaired reproductive performance (Pedernera et al., 2008) and an increased risk of health disorders, especially metabolic disorders (Pryce et al., 2016).

Early identification of cows that are at risk of negative energy balance could help farmers and dairy professionals develop strategies to reduce the impact of negative energy balance, which could in turn improve cow health and reproductive performance. However, most methods to calculate energy balance involve calculating the difference between energy intake and energy expenditure, where energy intake is calculated from actual daily feed intakes, which is both logistically difficult to manage and expensive. Although, the value of real-time prediction of methane emissions is not immediately obvious, it could be used as a proxy for efficiency, as inefficient cows are expected to be greater polluters. Consequently, cost-effective methods to predict energy balance and methane emissions are of value.

A longer term solution is to select for higher energy balance, or use energy balance as a genetic predictor of other traits, such as fertility. Genomic selection could be used to predict energy balance measured in dedicated herds, that have detailed energy balance phenotypes, or using large numbers of energy balance phenotypes estimated using MIR spectra.

The objective of this study was to predict a range of traits including several milk fatty acids, energy balance and methane emissions using MIR spectral data and genomics using data from a research herd in Victoria. In addition, a MIR prediction equation for energy balance developed using UK data and available commercially by National Milk Records was validated in the Australian research herd.

Materials
Data. The research focused on three trait groups: milk fatty acids (FA), Energy Balance (EB) and Methane emission (CH₄). Phenotypes and associated MIR spectral data were available from an experiment that ran from October to December 2015 consisting of 120 Australian Holstein lactating cows that calved in the spring at the research farm of the Department of Economic Development, Jobs, Transport, and Resource (DEDJTR) in Ellinbank, Victoria, Australia. All protocols of the experiment were approved by the DEDJTR Agricultural Research and Extension Animal Ethics Committee. The cows were divided into three batches of 40 cows. Batches were balanced for days in milk (DIM), number of lactations, and body weight. Cows had ad libitum access to feed and water with the diet consisting of the cubes that (on a DM basis) were approximately contained 74% alfalfa hay, 25% crushed barley grain, 1% minerals (calcium, phosphorus, and magnesium) and provided by Multicube Ltd. (Yarrawonga, Victoria, Australia). The estimated metabolizable energy (ME) was on averaged 11.17 MJ/kg of dry matter (DM) and the crude protein was 18%.

Individual cow daily dry matter intakes (DMI, kg/d) were measured over approximately 32d using feed bins mounted on load cells that were electronically monitored by linking the intake bin weight data to electronic identification of individual cows (Gallagher Animal Management Systems, Hamilton, New Zealand). Body weight (kg) was measured once daily after the morning milking (DeLaval Automatic weigh system AWS100, Tumba, Sweden). On average there were 20 BW measurements per cow over the 32d experimental period. Cow BCS was assessed twice weekly by 4 assessors using an 8-point scale and these scores were averaged for each cow each week (Earle, 1976). Milk yield (MY, kg/d) was measured for each cow twice daily using a DeLaval Alpro milking metering system (DeLaval International, Tumba, Sweden). Three times per week, separate afternoon and morning milk samples were taken from each cow using in-line milk meters (DeLaval International) for analyses of concentration of fat (MCF, %), protein (MCP, %), and lactose (MCL, %).
Fatty acid in milk

The milk samples collected in the morning during October to December 2015 were measured with the output of 32 different types of fatty acids including saturated FA (C4:0 – C20:0), Saturated FA (10:1-18:1), Mono-unsaturated FA (18:2, 18:3), and Trans FA (CLA). Across all the fatty acids, Mahalanobis distance was calculated. Outlier samples and absorbance (with Mahalanobis distances >3) were removed.

Energy balance

Two different equations were used to calculate EB following Phuong et al (2016). The first equation of calculating EB is termed as Input-Output equation (Phuong et al. 2016):

\[ E_{Binout} = E_{intake} - (E_{maintenance} + E_{milk} + E_{activity}) \]

Where, \( E \) is energy balance, \( E_{intake} \) is feed energy intake, \( E_{maintenance} \) is energy required for maintenance, \( E_{milk} \) is energy required for MY, and \( E_{activity} \) is energy required for activity.

The energy system chosen for use in the present study was effective energy (EE) introduced by Emmans (1994) because in that energy system the values of EE assigned to feeds are directly equivalent to the energy requirements of the animals, i.e. 1 MJ of EE supply has the same energy value as 1 MJ of lipid loss from the body. This approach is the basis to appropriately compare EBinout (MJ EE/d) with that calculated by the body reserve change method (Thorup et al., 2012). Using the estimates of metabolizable energy and crude protein of the diet fed to the cows combined with a series of equations proposed by Emmans (1994) the resulting dietary energy density in terms of effective energy was 8.33 MJ of EE/kg of DM. Consequently:

\[ E_{intake} = DMI \times 8.33 \]
\[ E_{maintenance} = 0.59 \times BW^{0.75} \] (Agnew et al., 2003)
\[ E_{milk} = MY \times (56 \times MCF + 33 \times MCP + 18 \times MCL) \] (Coffey et al., 2001)
\[ E_{activity} = 0.01 \times BW \] (Coffey et al., 2001). An amount of 2 MJ of EE was added to total energy expenditure per day to account for the energy required for walking between milking parlor and grazing paddocks (Mandok et al., 2013).

The other equation used for calculating EB used milk composition as follows:

\[ E_{Bal\_Milk} = 132.769 + 13.0675 \times MCF - 140.304 \times MCF/MCP - 95.1219 \times \text{diff}(MY) - 172.65 \times \text{diff}(MCF/MCP) + 802.306 \times \text{diff}(mPy) \]

Where, \( MY, MCF, MCP \) and \( mPy \) are milk yield, milk fat percent, milk protein percent, and milk protein yield as aforementioned; \( \text{diff}(MY), \text{diff}(MCF/MCP), \text{diff}(mPy) \) are the current minus the previous values of the milk measure.

The data used in this study were daily and individually recorded and because of this, they were highly variable, therefore a smoothing function was applied. There were two reasons for performing smoothing including:

1. there was a necessity to utilize rate of change variables and these are usually noisy when calculated using unsmoothed data and
2. when using a time-series data to make predictions it is a standard practice to use smoothed data before interpretation.

Methane emissions

On the same 120 cows as the EB, Methane emissions were measured for each cow over a 5 day period that occurred within the 32d experiment using the SF\(_6\) tracer method...
of Deighton et al (2014). Three phenotypes were calculated: 1) AvCH4, which was the mean of total methane emissions over the 5 day period; 2) AvCH4yield, which was total methane yield divided by the actual feed intake over the 5 day period; 3) AvCH4Intens, which was total methane yield divided by milk yield over the 5 day period.

**Mid-infra-red spectral data**

MIR spectral data were obtained using a Bentley FTS instrument. The spectrum for each milk sample had 899 data points (absorbance) for wavelengths ranging from 649 to 3998 cm\(^{-1}\). For all the collected MIR data across different data, Principal Component Analysis (PCA) coupled with the criteria (Mahalanobis distance between the records bigger than 3) is used to filter the outliers as shown in Figure 1.

**Figure 1.** *The PCA results for MIR data across different date.*

There were several pre-processing steps applied to the raw MIR data, using approaches developed by Grelet et al. (2015):

1. Standardizing the data from transmittance into absorbance: taking the reciprocal of transmittance. That is, the raw transmittance is randomly distributed, which might reduce prediction ability. The standardisation step would transfer the transmittance data into absorbance (the linear format).
2. Smoothing: taking the mean absorbance across every 3~5 data points.
3. Removing noise: As reported by Hewavitharana and van Brakel (1997), and De Marchi et al. (2013), two spectral regions are defined to be noise spectra, which are either water absorbance or useless chemical information. When converted from FOSS format to Bentley, two ranges of the wavelengths (from 1603 to 1682 cm\(^{-1}\); from 3006 to 3998 cm\(^{-1}\)) are treated as noise, and were therefore discarded.

After the noise filtered, there are total 620 wavelengths left for analysis when the outliers were removed (the Mahalanobis distance > 3).

Finally, MIR prediction equations developed in Europe was used to predict energy balance (McParland et al., 2011) The UK equations were developed using a different type of MIR machine (available from FOSS) machine and so the Bentley format MIR data was standardised into FOSS format according to the method described by Grelet et al. (2015).

**Methods**

Partial least squares regression method is commonly used in the analysis of MIR spectral data (De Marchi et al., 2012). An example is the SimPLS algorithm in R which
was implemented for MIR prediction. To avoid the over-fitting problems, 10 fold cross validation (10fold-CV) was used to assess the accuracy of the MIR prediction, as the correlation between the prediction and phenotype in the validation set; where the data sets are randomly separated into 10 subsets, and then each subset is treated as validation set while the other nine sets are combined as reference set. To evaluate the prediction accuracy, the coefficient of determination between reference set and validation set is used.

**Results**

The correlations between energy balance and milk fatty acids are shown in Figure 1:

![Figure 1. Pearson correlations between milk fatty acids and energy balance derived using individual cow data and the prediction accuracy using MIR data.](image)

We found a considerable number of significant correlations between energy balance and milk fatty acids, especially C:18 and C:20 (Figure 2). The high proportion of C:18 and C:20 during periods of negative energy balance is related to a high uptake of long chain fatty acids released from the mobilisation of body fat reserves (Bastin et al., 2011). Most of the correlations between energy balance and milk fatty acids were quite small. An explanation for these findings is that the cows that participated in the experiment had passed peak lactation.

There were 32 milk fatty acids measured on 95 cows with MIR samples. The accuracy of all the traits as determined by the coefficient of determination was higher than 0.50, with relatively lower errors (Root Mean Square Error; RMSE). For some fatty acids, for example, C4:0-C19:0, the prediction accuracy reached around 0.80.

**Table 1.** MIR prediction of two Energy Balance (predicted as the difference between energy intake and output; EBinout; energy balance using an equation applied to milk production data; EBalMilk) and three methane emission traits (average methane; AVCH4; AVCH5; AVCH6).
methane yield AVCH4\text{yield} and methane intensity AVCH4\text{intense}).

<table>
<thead>
<tr>
<th>Phenotypes</th>
<th>Country</th>
<th>$r^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy Balance</td>
<td>AU</td>
<td>0.42</td>
<td>31.27</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>0.39</td>
<td>-</td>
</tr>
<tr>
<td>Methane Emission</td>
<td>AU</td>
<td>0.45</td>
<td>31.05</td>
</tr>
<tr>
<td>AVCH4</td>
<td>AU</td>
<td>0.51</td>
<td>21.48</td>
</tr>
<tr>
<td>AVCH4\text{yield}</td>
<td>AU</td>
<td>0.49</td>
<td>31.09</td>
</tr>
<tr>
<td>AVCH4\text{Intense}</td>
<td>AU</td>
<td>0.52</td>
<td>20.45</td>
</tr>
</tbody>
</table>

*The accuracy was assessed as the coefficient of determination ($r^2$) and root mean square error of calibration (RMSE) calculated as the mean of 10 fold cross validation applied to prediction equations developed using Australian (AU) and UK reference datasets.

The MIR prediction accuracy on energy balance is 0.42 (Table 1) with the root mean square error of calibration (RMSE) is 31.27. On the same trait, the equation derived from the UK (estimated with a much larger reference dataset and a similar energy balance trait) has a similar accuracy (around 0.39). Similarly, the MIR prediction accuracy of energy balance based on milk composition is around 0.45.

An important finding is that the prediction accuracy of energy balance using MIR exceeded the prediction of energy balance using herd-test results. Applying the equation EBalMilk to milk components from a herd-test (i.e. without data smoothing) had a correlation of 0.22 with EBinout.

In addition, the accuracy of MIR prediction of the three methane emissions traits ranged from 0.49 to 0.52 with the RMSE ranging between 20.45 and 31.09, which are higher errors than the prediction on milk fatty acid.

**Discussion**

Monitoring cow energy status at the individual level has been shown to offer benefits for early identification of cows at risk of health and reproductive problems (Soyeurt et al., 2008; Rutten et al., 2009). Consequently, the use of MIR spectral data to predict this phenotype has positive management and genetic selection implications.

The accuracy of MIR prediction of energy balance were promising (~0.3), but are currently lower than other comparable studies, where the prediction accuracies are close to 0.6-0.7 (McParland et al., 2011). The most likely explanation is that the McParland et al (2011) had a large reference population of 6,665 test days from 465 lactations of 277 cows that had energy balance phenotypes matched to MIR spectra. The Australian reference population in comparison is small. Options to increase the accuracy include: 1) use equations already developed overseas or 2) increase the size of the domestic reference population. In fact, we attempted strategy “1” and found similar accuracies of prediction using an equation developed in the UK. The reasons why the accuracy in Australia was lower than observed in studies such as McParland et al (2011) include the fact that the reference population was not optimised for Australia; the phenotypes were subtly different to the energy balance prediction that was used for the UK and it is possible that genotype by environment interactions exist. The second strategy is to increase the reference population and could include expanding the number of phenotypes from research herds, or devising ways in which to measure energy balance on commercial farms, such as regular body condition score and body weight records in the early part of lactation. Although these phenotypes are considerably cheaper than obtaining individual cow feed intake data, they are still difficult to obtain routinely in commercial herds. However, in the next phase of the MIR for Profit project, we are investigating ways in
which more energy balance phenotypes can be collected from research and commercial dairy herds.

In this study there was only data from 355 animals for energy balance and 120 animals for methane emission for genomic prediction, obviously this is a very small dataset for genomic prediction and consequently the accuracies are low. However, if MIR prediction equations could be validated in larger populations of commercial cows, which is occurring in the next phase of the project, then the prospects for generating large reference populations of MIR estimated energy balance phenotypes for genomic prediction are promising. For commercial purposes, the International Committee on Animal Recording advise a prediction accuracy of greater than 75%. However, for indicative use (e.g. at risk groups), a lower accuracy may be acceptable.

In contrast, the MIR predictions of milk fatty acid and methane emissions were encouraging. Here, the prediction accuracies are reasonably good, up to 0.8 in milk fatty acid with 0.5 in methane emission. In fact, the phenotypes and corresponding MIR spectra in the Australian dataset is large by international standards (e.g. Dehareng et al., 2012) used 11 cows in predicting methane emissions using MIR. However, the practical implications of real-time prediction of methane emissions are not immediately clear, although it could be seen as a proxy of feed efficiency, as cows that are higher emitters for a given level of yield are less efficient.

Conclusions
Our analyses show MIR prediction of energy balance, fatty acids and methane emissions is promising but needs further improvement. Provided accuracies are improved (i.e. larger reference populations), then MIR prediction methods could also be applied to other phenotypes which are only measurable in comparatively small populations. Prediction equations could then be applied to all herd-tested cows, which opens up more traits for management and genetic improvement purposes.

What’s next?
The next part of the project is to validate the MIR predictions in commercial herd data. In late 2016 MIR spectral data from around 2000 genotyped cows in 6 herds were collected twice per month in addition to pregnancy testing of these cows (using the IDEXX method of pregnancy confirmation). Regular milk recording, calving, health and fertility data will continue to be collected and stored at DataGene for future reference. The prediction equations that have been developed so far in the project will be applied to the MIR spectral data from these cows and compared to estimates of energy balance calculated using smoothed early lactation data (EbalMilk). We will also quantify if MIR predictions of energy balance are higher in cows that become pregnant versus those that do not. A final experiment of 10,000 cows is planned for 2017/18. An overview of MIR for Profit activities is described in Figure 3.

![Figure 3. Overview of MIR for Profit project activities.](image)

Acknowledgements
This work was part of a project “MIRprofit: integrating very large genomic and milk midinfrared data to improve profitability of dairy cows” funded by the Commonwealth of Australia. The authors thank the staff of DEDJTR Ellinbank research farm for cow feeding and husbandry.

References


