

Estimating pasture quality using Landsat ETM+: application for the greenhouse gas inventory of New Zealand

A-G. Ausseil^a, J.R. Dymond^a, R. Dynes^b, J.D. Shepherd^a, B. DeVantier^c, A. Sutherland^a

^aLandcare Research, Private Bag 11052, Palmerston North, New Zealand – (AusseilA, DymondJ, ShepherdJ, SutherlandA)@LandcareResearch.co.nz

^bAgResearch, Christchurch, New Zealand – Robyn.Dynes@AgResearch.co.nz

^cAgResearch, Grassland centre, Palmerston North, New Zealand – Brian.Devantier@AgResearch.co.nz

Abstract – The greenhouse gases CH₄ and N₂O that arise from agriculture comprise 48% of New Zealand's total CO₂ equivalent greenhouse gas emissions. The majority of these emissions arise directly (enteric CH₄) or indirectly (waste CH₄ and soil N₂O) from ruminant animals. Pasture quality has a strong influence on feed intake and therefore the release of both CH₄ and N₂O. Digestibility and metabolisable energy measures influence the quantity of feed eaten – the main determinant of CH₄ emissions – while the N concentration of the diet consumed influences the quantity of N voided in faeces and urine – the main determinant of soil N₂O emissions. These pasture properties vary in space and time, and will have a strong influence on the agricultural greenhouse gas budget. In the current national greenhouse gas inventory, pasture quality measures are rudimentary, relying on one-off surveys with a limited spatial coverage. Remote sensing has the potential to provide pasture quality estimates at better temporal and spatial scales.

In this paper, we investigated the potential of remote sensing for predicting pasture quality. We sampled a range of pasture qualities around New Zealand, targeting the scheduled path of Landsat ETM+ and checking the cloudiness on the location. We applied a radiometric correction on the satellite images and modelled three pasture quality measurements (nitrogen content, metabolisable energy and digestibility) using multiple linear regression and some non-linear regression techniques. We found that the non-linear regression techniques improved the results significantly because there was a non-linear relationship in the upper range of pasture quality.

Keywords: greenhouse gas, pasture quality, LANDSAT ETM+

1. INTRODUCTION

The non-carbon dioxide (CO₂) greenhouse gases CH₄ and N₂O that arise from agriculture comprise 48% of New Zealand's total CO₂ equivalent greenhouse gas emissions (Ministry for the Environment, 2009). The majority of these emissions arise directly (enteric CH₄) or indirectly (waste CH₄ and soil N₂O) from ruminant animals. Pasture quality is a key factor in the inventory because it has a strong influence on feed intake and therefore the release of both CH₄ and N₂O. Organic Matter Digestibility (OMD) and metabolisable energy (ME) influence the quantity of feed eaten – the main determinant of CH₄ emissions – while the N concentration of the diet consumed influences the quantity of N voided in faeces and urine – the main determinant of soil N₂O emissions. For the purpose of the national greenhouse gas inventory, however, the estimates of pasture quality at the national scale for New Zealand are rudimentary, relying on pasture quality surveys taken at specific points in time and with limited spatial coverage (Ministry for the Environment, 2009).

Near-infrared spectroscopy (NIRS) is generally used for the rapid evaluation of chemical composition (Norris et al., 1976), and has been widely used to estimate forage quality parameters

(Zhao et al., 2007). Stepwise multiple linear regression is usually used for selecting wavebands that are relevant for pasture quality prediction. The key wavelengths for prediction were found in the blue range, the red-edge band (700–775 nm), and various peaks in the near infrared and short-wave infrared (Kawamura et al., 2008).

Although NIRS techniques have great advantages compared with the laboratory chemical method, they are still time-consuming as they require collection and processing (drying, grinding) of vegetation samples. Hand-held spectro-radiometers were therefore investigated to improve real-time prediction (Mutanga et al., 2004; Starks et al., 2004). Skidmore et al. (2010) tested the use of hyperspectral remote sensing to estimate forage quality of savannas (trees and grass). They mapped the spatial patterns of foliar nitrogen and polyphenols using a neural network model and demonstrated the influence of fire and soil type on foliar nitrogen concentration. Beerli et al. (2007) found that spectral estimates were within 8% between field spectrometer and crude protein measurements, and applied their algorithm to hyperspectral imagery.

Field spectrometry is suitable for paddock-scale estimation of pasture, but would not be appropriate for providing on-going information at national scale. In this context, remote sensing by satellite images has the ability to repeatedly cover large areas for monitoring trends of pasture condition. Studies on the estimation of forage quality variables using remote sensing are, however, quite limited. Starks et al. (2006) used six broad wavebands similar to Landsat Thematic Mapper and found that reflectances in the red and near infrared bands were most highly correlated with forage quality measures. They also found that the ratios blue/red were most highly correlated with crude protein ($r^2 = 0.71$). Phillips et al. (2006; 2009) studied the use of satellite images to estimate forage quality and quantity using the Modified Soil Adjusted Vegetation Index (MSAVI), which is computed with Red and NIR bands, to minimise soil background influences. Their results showed that the estimation of forage quality and quantity had an error of less than 20%.

In this paper, we test the ability of remote sensing images to predict three pasture quality parameters (digestibility, ME and N content) using Landsat ETM+. We fitted linear and non-linear regressions and assessed their predictive capability.

2. METHODS

2.1. Pasture quality sampling

We targeted the scheduled path of Landsats 5 and 7 and checked the cloudiness on that day before sampling paddock sites across New Zealand. Since Landsat has a repeat coverage interval of 16 days, we had only one to two images per month available on average for each location. We allowed up to 7 days of deviation from the actual pasture sampling date, to give technical staff sufficient time to access the farm location. We assumed that pasture quality varied little within that time interval.

A database was compiled for regression analysis using data from this sampling campaign supplemented with historical

datasets for which spatially explicit referenced measurements of pasture quality were examined to match archived Landsat 5 TM and Landsat 7 ETM+ images (Ausseil et al., 2009). These pasture quality estimates were collected between 2001 and 2008.

The database used for regression comprised 389 samples evenly distributed between the North Island and the South Island. The regression dataset covered a broad range of ME, OMD and N content. OMD values ranged from 30.1% to 90.3%; ME values ranged from 4.8 to 13.3 MJME/kg; and N content ranged from 0.2 to 5.1% (Table 1). The histogram for OMD, ME and N content are shown on figure 1.

Table 1 Statistics of OMD, ME and N content

Variable	Minimum	Mean	Maximum	Standard deviation
OMD	30.1	68.6	90.3	10.3
ME	4.8	10.18	13.3	1.4
N content	0.2	2.7	5.1	0.8

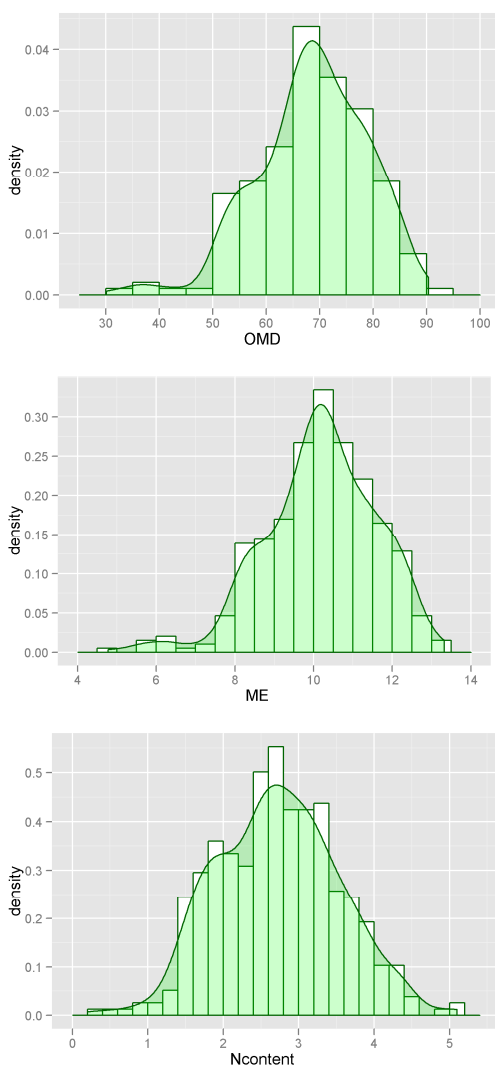


Fig 1. Histogram of input OMD (%), ME (MJME/kg) and N content (%).

The native tussock grasslands have very low values for all parameters, and sheep and beef farms have generally lower OMD, ME and N content than dairy farms, due to the differences in soils, slope, environment and in farm management practices. Two binary factors were therefore added: *IsTussock* and *IsSheepBeef*, which are categorical

variables to account for the different groups (Tussock = 1, non-tussock = 0, sheep and beef farms = 1, not sheep and beef farm = 0). There were five observations with tussock, and 258 with sheep and beef farms.

2.2. Satellite data

The radiometric correction of the Landsat 5 and 7 imagery followed the algorithms developed by Dymond and Shepherd (2004). Each sampled paddock extracted from the satellite image was processed to standardised spectral reflectance (spectral reflectance at nadir view for the satellite and 50° elevation for the sun). Processing of satellite imagery to standardised reflectance involves physical modelling of radiation: from the sun and sky through the atmosphere; reflection of the light from the vegetation canopy; and the transmission of the reflected light through the atmosphere to the satellite sensor. We used the 6S code (Vermote et al., 1997) to model irradiance and transmission of light through the atmosphere, and WAKII (Dymond et al., 2001) to model bidirectional reflectance of vegetation.

2.3. Regression fitting

Multiple linear and non-linear regressions were fitted to predict OMD, ME and N content using the R software version 2.12.0 (R Development Core Team, 2010) and a selection of packages. Linear regressions were tested including multiple linear, robust regression (package *robust*), and least trimmed squared (LTS) regression (package *robustbase*). The robust and LTS regression puts less weight on influential observations with high errors. The non-linear regressions tested were generalised additive models (GAM) (package *mgcv*), multivariate adaptive splines models (package *earth*), random forest (package *RandomForest*), and Gradient Boosting model (package *gbm*). GAM and Multivariate Adaptive Splines are both methods that use non-parametric functions to allow flexible function description of complex response to variables. GAM uses smoothing splines, and multiple adaptive spline regression uses piecewise linear basis functions. Random Forest and Gradient boosting models are regression trees techniques that both build a large collection of regression trees and optimise the result. In order to compare all the models equally, we used the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{SSE}{df}}$$

where $SSE = \sum_n (obs_i - pred_i)^2$ the sum of square errors

(obs_i being the observed value and $pred_i$ the predicted value for sample i) and df the degrees of freedom.

R-squared was also computed. Note that we haven't provided adjusted R-squared to enable comparison between models.

3. RESULTS

Tables 2, 3 and 4 show results for linear and non-linear regression fitting for ME, OMD and N content respectively.

Table 2 Performance of multiple regressions models for ME

Regression type	SSE	RMSE (in MJME/kg)	R ²
Multiple linear	322	0.92	0.59
Robust regression	323	0.92	0.58
LTS regression	323	0.92	0.58
Generalised additive model	267	0.84	0.66
Multivariate adaptive spline	293	0.88	0.62
Random Forest	134	-	0.85
Gradient boosting model	240	-	0.69

Table 3 Performance of multiple regressions models for OMD

Regression type	SSE ($\times 10^3$)	RMSE (in %)	R ²
Multiple linear	16.27	6.53	0.60
Robust regression	16.29	6.54	0.60
LTS regression	16.40	6.57	0.60
Generalised additive model	13.10	5.86	0.68
Multivariate adaptive spline	14.40	6.15	0.64
Random Forest	7.10	-	0.85
Gradient boosting model	13.00	-	0.67

Table 4 Performance of multiple regressions models for N content

Regression type	SSE	RMSE (in %)	R ²
Multiple linear	145	0.62	0.43
Robust regression	146	0.62	0.43
LTS regression	148	0.62	0.45
Generalised additive model	119	0.56	0.53
Multivariate adaptive spline	133	0.59	0.48
Random Forest	57	-	0.83
Gradient boosting model	106	-	0.59

Since ME and OMD are highly correlated variables (correlation = 0.98), the results presented in tables 2 and 3 are similar in performance of the regression models. For those two variables, a stepwise linear regression selected bands -7 and the two qualitative variables *IsTussock* and *IsSheepBeef*. Band 1 was therefore dropped out for all the linear and non-linear regressions tested. The results show an RMSE between 0.59 and 0.92 MJME/kg and for ME 4.3% and 6.57% for OMD. The robust or LTS regressions did not significantly improve the multiple linear regression. The best fitting results were obtained with the random forest model for both ME and OMD. Figure 2 shows the observed versus predicted OMD. Similar results were obtained for ME (figure not shown). There is an over-prediction in the lower range (below 50% of OMD) and an under-prediction in the higher range (over 80% OMD) that was reduced with the Random forest model.

For nitrogen content, a stepwise regression selected bands 1, 2, 3, 4, 7 and one qualitative variable *IsSheepBeef*. Band 5 and *IsTussock* were therefore dropped for all the linear and non-linear regressions tested. The results show an RMSE between 0.39 and 0.62 %. As for ME and OMD, the robust, least trimmed or generalised least square regressions did not significantly improve the multiple linear regression. The best fitting results were obtained with the random forest model, as for ME and OMD. Figure 3 shows the observed versus predicted N content. There is an over-prediction in the lower range (below 2%) and an under-prediction in the higher range (over 3.5%) that was reduced with the random forest model.

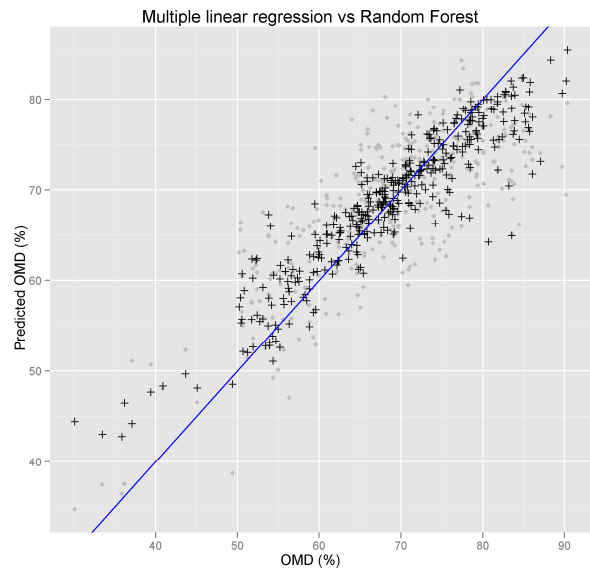


Fig 2. Observed vs predicted OMD (%) with multiple linear regression (grey dots) and random forest (black crosses).

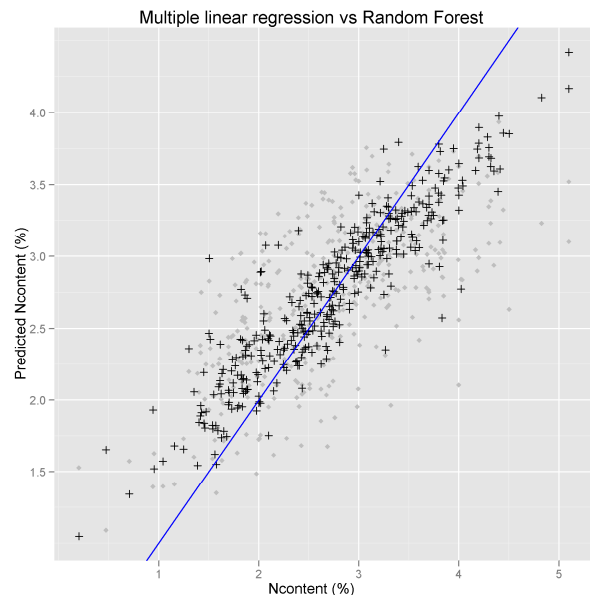


Fig 3. Observed vs predicted Ncontent (%) with multiple linear regression (grey dots) and random forest (black crosses).

4. DISCUSSION

The histogram of values for ME and OMD covered the full range of pasture quality in New Zealand. The histograms were left-skewed, showing some very low values of ME and OMD. The correlations for the regressions were improved from previous research (R-squared around 0.5 for OMD, ME and N content) (Dymond et al., 2002; Dymond et al., 2006). This is due to the use of Landsat ETM+ to locate the paddocks within 30-m pixels rather than the 1-km pixel size of SPOT VEGETATION used previously.

The predictions for ME, OMD and N content all showed some under-prediction in the higher range of pasture quality. This can be an issue for the inventory as it might be hard to assess any improvement in pasture over time. Non-linear regressions improved model fitting compared with the multiple linear regression. This is due to the ability to fit non-linearities,

especially in the upper range of pasture quality. Machine learning techniques such as Random Forest and stochastic gradient boosting showed the best fit of models, although there was still some remnant over-prediction in the lower range and under-prediction in the upper range, suggesting a lack of informative power in the predictors. These models however have a tendency to over-fit on a sample of data, thus reducing their predictive power. The over-fitting may occur as the measurement errors in the sample (due to variability in quality within the paddock) and in the spectral reflectance (due to the atmospheric correction and spatial resolution) are unknown. Choosing the right model will therefore depend on the ability of the model to predict pasture quality over space and time in a robust manner.

N content has the lowest model fitting results. This may be due to the growth of pasture that influences the greenness of the pasture, and time between satellite measurement and sample taken. This hypothesis is supported by Lamb et al. (2002), who found a significant influence of canopy biomass for low-level LAI on red-edge determination of leaf nitrogen concentration of ryegrass.

The regression equations can be used in an operational system by applying the equations on MODIS images. The MODIS sensor provides images at a regular interval (every 10 days) but a lower spatial resolution (250–500 m). This is appropriate to get a single image of New Zealand at regular time stamps. An overlay of GIS layers can therefore be used as inputs (MODIS data, land-use type and vegetation type), producing pasture quality maps as outputs. It will then be possible to monitor pasture quality trends over time in different bioclimatic regions of New Zealand, and also get a better estimate of greenhouse gas emissions over time.

5. CONCLUSIONS

The estimation of pasture quality is an important component of the greenhouse gas inventory of New Zealand, since the methane and nitrous oxide emissions are mainly driven by livestock dry matter intake. Although we have found limited agreement between spectral broad bands and pasture quality parameters, the prediction would still be valuable to monitor trends over time.

The non-linear regression techniques have improved the fitting results compared to multiple linear regression. These models better represent the non-linearities especially in the upper range of pasture quality. However, machine learning techniques such as random forest or stochastic gradient boosting make the interpretation difficult and have limited abilities in a predictive mode.

The next step of this research will be to test the various regression models in a predictive mode, and discard models that have over-fitted data and made the predictions unreliable over the full spatio-temporal dataset.

ACKNOWLEDGEMENTS

This work was supported by the New Zealand Ministry of Agriculture and Forestry. AgResearch, Lincoln University, Fonterra and DairyNZ provided measurements of pasture quality. We would like to thank Donna Giltrap for review and Anne Austin for editing assistance.

REFERENCES

- Aussein, A.G.E., Schierlitz, C., Dynes, R., Shepherd, J.D., Dymond, J.R., 2009. Updated estimates of pasture digestibility, ME and N content. Landcare Research Ltd, p. 32 p.
- Beeri, O., Phillips, R., Hendrickson, J., Frank, A.B., Kronberg, S., 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment* 110 216-225.
- Dymond, J.R., Shepherd, J.D., 2004. The spatial distribution of indigenous forest and its composition in the Wellington region, New Zealand, from ETM+ satellite imagery. *Remote Sensing of Environment* 90 116-125.
- Dymond, J.R., Shepherd, J.D., Clark, H., Andrew, R., 2002. Spatial and temporal estimates of herbage quality. Landcare Research Ltd, p. 56 p.
- Dymond, J.R., Shepherd, J.D., Clark, H., Litherland, A., 2006. Use of VEGETATION satellite imagery to map pasture quality for input to a methane budget of New Zealand. *International Journal of Remote Sensing* 27 1261-1268.
- Dymond, J.R., Shepherd, J.D., Qi, J., 2001. A simple physical model of vegetation reflectance for standardising optical satellite imagery. *Remote Sensing of Environment* 77 230-239.
- Kawamura, K., Watanabe, N., Sakanoue, S., Inoue, Y., 2008. Estimating forage biomass and quality in a mixed sown pasture based on partial least squares regression with waveband selection. *Grassland Science* 54 131-145.
- Lamb, D.W., Steyn-Ross, M., Schaare, P., Hanna, M.M., Silvester, W., Steyn-Ross, A., 2002. Estimating leaf nitrogen concentration in ryegrass (Lolium spp.) pasture using the chlorophyll red-edge: Theoretical modelling and experimental observations. *International Journal of Remote Sensing* 23(18) 3619-3648.
- Ministry for the Environment, 2009. New Zealand's greenhouse gas Inventory 1990-2007: Wellington, New Zealand.
- Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment* 89 393-408.
- Norris, K.H., Barnes, R.F., Moore, J.E., Shenk, J.S., 1976. Predicting forage quality by infrared reflectance spectroscopy. *Journal of Animal Science* 43 889-897.
- Phillips, R., Beeri, O., Liebig, M., 2006. Landscape estimation of canopy C:N ratios under variable drought stress in Northern Great Plains rangelands. *Journal of Geophysical Research* 111(G02015).
- Phillips, R., Beeri, O., Scholljegerdes, E., Bjergaard, D., Hendrickson, J., 2009. Integration of geospatial and cattle nutrition information to estimate paddock grazing capacity in Northern US prairie. *Agricultural Systems* doi:10.1016/j.agsy.2009.01.002.
- R Development Core Team, 2010. R: A language and Environment for Statistical Computing. Vienna, Austria. Available at: <http://www.R-project.org>
- Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Wieren, S.E.V., Peel, M., Grant, R.C., Prins, H.H.T., Bektas Balci, F., Venus, V., 2010. Forage quality of savannas - Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment* 114 64-72.
- Starks, P.J., Coleman, S.W., Phillips, W.A., 2004. Determination of forage chemical composition using remote sensing. *Journal of Range Management* 57(6) 635-640.
- Starks, P.J., Zhao, D., Phillips, W.A., Coleman, S.W., 2006. Development of Canopy reflectance algorithms for real-time prediction of bermudagrass pasture biomass and nutritive values. *Crop Science* 46 927-934.
- Vermote, E., Tanre, D., Deuze, J.L., Herman, M., Morcrette, J.J., 1997. Second Simulation of the Satellite Signal in the Solar Spectrum (6S): an overview. *IEEE Transactions of Geosciences and Remote Sensing* 35 675-686.
- Zhao, D., Starks, P.J., Brown, M.A., Phillips, W.A., Coleman, S.W., 2007. Assessment of forage biomass and quality parameters of bermudagrass using proximal sensing of pasture canopy reflectance. *Grassland Science* 53 39-49.