Sources of Uncertainty in Land Use Pressure Mapping

L.R. Lilburne and H.C. North
Landcare Research
P.O. Box 69, Lincoln, Canterbury 8152, New Zealand
Ph. +64 3 3256700; Fax +64 3 3252418
E-mail: lilburnel@landcareresearch.co.nz

Keywords: landuse pressure, multi-temporal, radiometric error, geometric error

Abstract
A key agricultural management pressure in Canterbury, New Zealand, is the practice of leaving land fallow during winter, because any nitrate present is likely to be leached since there is no plant uptake. Remote sensing imagery has been successfully used to identify land bare for significant periods in winter and early spring. This was achieved by use of simple rules on a temporal sequence of three Landsat 7 images. In particular, ability to correctly classify land into bare, sparsely vegetated, and fully vegetated categories according to percentage cover of vegetation was investigated using detailed field data along with the images. This paper analyses the sources of error in mapping fallow ground using remote sensing image sequences. Accurate assessment of total area and correct assessment at each spatial location were the aims. Potential sources of error include suitability of the logical model, timing of image acquisition, scale mismatch, incorrect reference data, geometric errors between images, and radiometric variation both within and between images (month to month and year to year). A probabilistic approach to providing estimates of uncertainty was applied to draw together the most important of these error sources. Radiometric error was the most significant error source. A map of uncertainty is also produced.

1. Introduction
Spatial modelling of nitrate leaching risk to groundwater at regional scale requires spatial information on soil properties, climate, land use, standard management practices, and aquifer vulnerability. A common approach is to combine this information with a simulation model of nitrate leaching to produce a map of risk (Corwin 1996).

Estimation of *actual* risk in any given period requires spatial knowledge of climate and the actual management practices that were applied (rather than assuming a standard practice). This management practice knowledge is difficult to obtain. One approach is to use agricultural statistics and questionnaires to collect the information. The difficulties with this approach include lack of geo-referencing, the level of aggregation of the statistics, non-response to the questionnaire, and incomplete or unclear answers. It is also time-consuming and thus does not lend itself to updates in subsequent years.

An alternative approach is to use remote sensing to map those relevant management practices that can be visually detected. One key cropping management practice is the sowing date of a crop and the duration over which the land is fallow in the critical winter leaching period (Lilburne et al. 2003). Our research investigates the use of multispectral satellite imagery for detecting and identifying this land. There are a number of potential sources of error in the mapping process that can have a significant impact on the utility of the map of fallow ground as an input to nitrate modelling. This paper reports the results of an assessment of these error sources and their spatial distribution.

2. Methods
2.1 Data
A sequence of Landsat 7 satellite images covering an area on the Canterbury Plains between Ashburton and Christchurch was acquired in 2002 (16 June, 19 August and 22 October). Four images were acquired in 2003 from Landsat 5, 7 and SPOT (21 July, 18 August, 15 September, 17 October). These images were all rectified to the New Zealand Map Grid using ground control points (GCPs), a digital elevation model and in-house orthorectification software. Over 100 GCPs were used for each image. Care was taken with orthorectification, since co-alignment accuracy of the image pair is crucial for direct comparison of

equivalent pixels across two or more images. Normalized Difference Vegetation Index (NDVI) is used as a proxy for percent live vegetation cover (Tucker 1979). This is a ratio of the red and infrared bands that is widely used as a simple indicator of "greenness"[1].

In this project, we are interested only in the agricultural land. A mask was constructed to remove urban areas, sea and lakes, hills, roads and rivers from consideration. The input data for this mask were national GIS layers for the most part, with river and hill outlines drawn by hand. This was applied to all images. Then for each image separately, cloud, cloud shadow and snow were classified spectrally and also masked.

Field data to create a reference dataset were collected within a few days of the August and October 2002 and August 2003 images, and within 2 weeks of the October 2003 image. Detailed notes on 130 fields were recorded in August 2002, including vegetation type, percent vegetation cover (visual estimation), state of any bare soil (moisture, cultivation type, presence of stones), interpreted land use of the field, and GPS location. Sites were selected by researchers while in the field. The GPS reference enabled us to identify each field later on the appropriate image and manually draw a polygon around it, allowing variability of reflectance within each field to be studied. The same procedure was used to record notes on field location and percent cover for the three other images.

To obtain more detailed data on vegetation percent cover, 15 paddocks being studied in another research programme were sampled in August 2003. In each paddock 10 digital photographs (2048 × 1536 pixels) were taken of a quadrat (a quarter-square-metre metal frame). A small bubble level on the top of the camera was used to make sure the photo angle was vertical. Figure 1 shows the pattern used to sample in each field. Distance between points depended on the size of the paddock but was at least 20 m. A visual estimate of the percent vegetation, soil, litter and stones was also made at the first sampling point. Farmers were contacted for crop and planting date information.

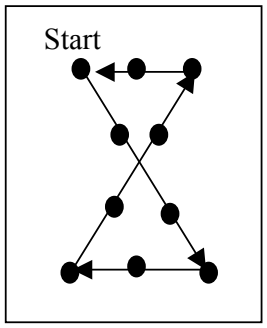

Figure 1. Sampling pattern used to take 10 photos in each paddock. Points (i.e., photos) are at least 20 m apart.

2.2 Land Pressure Model

Expert knowledge of agricultural practices and risk factors was used to derive two rules based on three categories of vegetation state: low (cover < 15%), sparse (15% < cover < 50%) and fully vegetated (cover > 50%).

- Rule A: a field is fallow throughout the winter if the cover is in a low or sparsely vegetated state in late autumn (i.e., May or June) **and** it is still in a low or sparsely vegetated state in July or August.
- Rule B: a field is fallow for a shorter (but still risky) period if it is fully vegetated in late autumn but only has low vegetation in July or August. Satisfaction of this rule usually indicates a cultivation or herbicide application event after the first image.

The reference data from August 2002 was used to set NDVI threshold values of 0.11 for 15% cover and 0.22 for 50% cover, for the August 2002 image. Further detail can be found in North et al. (2002). Areas in either of the images that are masked due to cloud are also masked in the output risk map.

The rules were applied to the 2002 images at a pixel level (25 × 25 m). In this initial work the August NDVI thresholds were also applied to the June image. Figure 2 shows the results of both rules combined into one map. All areas bounded by a red line are predicted as being in a near fallow state over most of the critical winter period, and are therefore potentially at risk of nitrate leaching to groundwater. In total, 57,259 ha are identified as being at risk. This equates to approximately 20% of the unmasked agricultural land on the Canterbury Plains between the Ashburton and Waimakariri rivers.

[1] $NDVI = \frac{NIR - red}{NIR + red}$

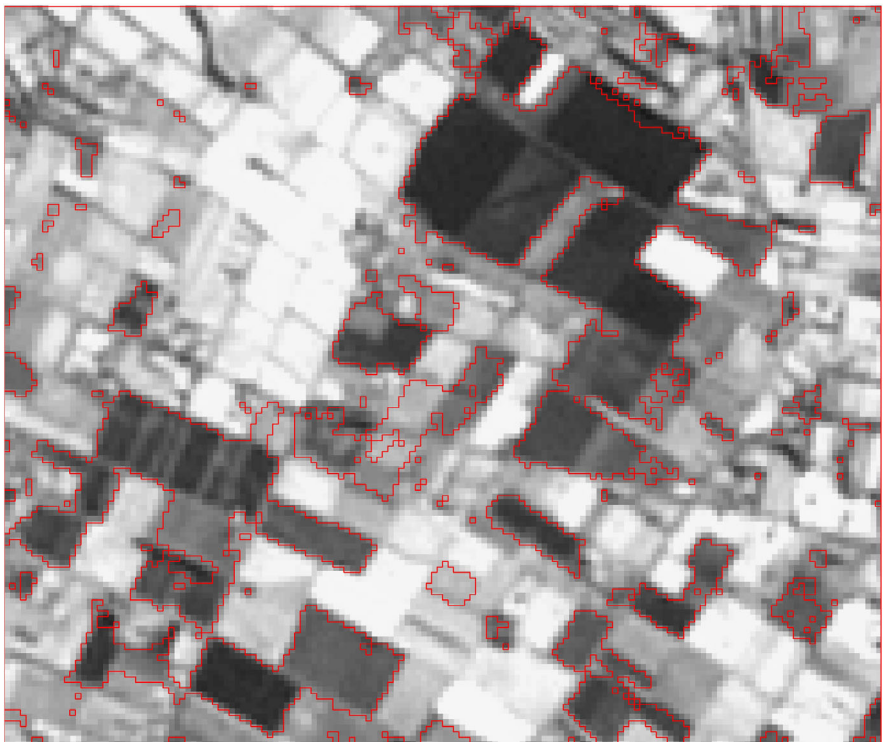

Figure 2. Fallow land in winter 2002.

3. Uncertainty Assessment

The fallow-land map is designed for use by modellers in predicting risk of nitrate leaching. Appropriate use of the map for this purpose requires that it be accompanied by information about the uncertainty in the estimate of total area of land at risk and the uncertainty at each location. Identification of the various error sources is also useful in further development of the procedure for mapping fallow land. Thus, an analysis of the various possible error sources was undertaken. The sources were compared by quantifying each in terms of the likely area of agricultural land in 2002 that might have been incorrectly classified. For each source estimates are given of the difference in the total area at risk, and the area for which pixels are incorrectly classified one way or the other (called spatial area). There are four groups of error sources:

1. Geometric error
2. Radiometric error
3. Model error
4. Reference data error.

3.1 Geometric Error

Geometric errors derive from sensor characteristics, and misregistration or distortion errors in the rectification process of registering images to a projection or to each other. Even misregistration errors as little as one-quarter of a pixel can result in significant change detection errors when multi-temporal image sequences are being analysed (Dai and Khorram 1998, Stow 1999, Foody 2002). The more spatially heterogeneous the landscape, the more significant the spatial coverage of the effect.

Ground control points (GCP) are generally used to both rectify the image and evaluate locational accuracy. Sources of error also include the DEM and location of the GCPs themselves. GCP positions measured from 1:50,000 topographic maps can include up to 100-m horizontal error. Even if higher accuracy reference points are obtained, GCP accuracy is still limited by identification of the features in the image. This is rarely more accurate than 0.5 pixel, and more often around 1 pixel. The co-location of the June and August 2002 images was tested using a multi-scale, sub-pixel chip correlation method at the GCP points where, by definition, there are strong spatial features. Misregistration at the majority of points was between 0.25 and 1.25 pixels. However, the least squares fit of many GCPs will minimize the errors to be less than that indicated by testing individual GCP points. Thus, actual misregistration error is expected to range from 0.1 to 0.25 of a pixel.

A simulation experiment was used to quantify the effect a 1-pixel geometric error would have on the total area of fallow land, and its spatial distribution. The August image was shifted eight times (vertically, horizontally and diagonally) by 1 pixel. The number of times a pixel was classified differently as a result of each geometric shift was counted. The total area identified as being under pressure due to its fallow state decreased by 300 ha. On average, 7,700 ha had a different risk status as a result of a 1-pixel offset in the second image, i.e., 3% of the agricultural land was incorrectly identified. This is considered to be a worst case scenario, as the estimated geometric errors are mostly less than 1 pixel, and may well actually only be 0.1 of a pixel.

3.2 Radiometric Error

Radiometric errors arise from differences in reflected radiance due to sensor noise, sensor calibration, atmospheric conditions, clouds, viewing and illumination angles, topography, and surface conditions that change the target reflectance. These can be divided into sensor-, atmosphere- and ground-scene-related errors (Friedl et al. 2001). Radiometric normalization techniques can be applied to minimize atmospheric and illumination effects (Lunetta 1999). Our study area is essentially flat so we did not carry out a radiometric correction for the effect of topographic slope. Also NDVI, like many ratio indices, substantially reduces topographic effects (Walsh et al. 1997).

The images used in this study vary widely in acquisition parameters, particularly between the winter images (where the solar elevation is between 15 and 25 degrees) and the spring images (where solar elevation is between 35 and 45 degrees). This has the potential to disrupt the consistency of vegetation percent cover determination between scenes. However, most simple methods for correction for solar incidence angle apply the same factor to all bands (Jensen 1996 pp. 107–137), which therefore cancels out in an NDVI ratio. For this reason we did not apply such a correction. Atmospheric effects are seen primarily in the blue and green bands, with much less effect in red and near infrared. So in terms of relatively simple atmospheric correction methods, such as dark target correction (Sabins 1978 pp. 242–245), little can be done for the bands used in the NDVI ratio.

It is, however, likely that NDVI will be affected by the varying state of the soil and litter between scenes, due to factors like soil moisture and level of litter decay (Nagler et al. 2000). A set of bare cultivated reference points was collected for the August and October 2002 and 2003 images. These reference points were analysed for variability both within a scene (i.e., for surface effects such as surface roughness, moisture, stoniness) and between scenes (i.e., for illumination, atmospheric and general surface conditions). Figure 3 shows that while there are clear differences between images, especially with October 2002, within-image differences due to cultivation roughness are not as obvious. This is expected, as NDVI should minimize shading differences between fields having different cultivation roughness. Degree of stoniness also did not have an obvious effect on NDVI.

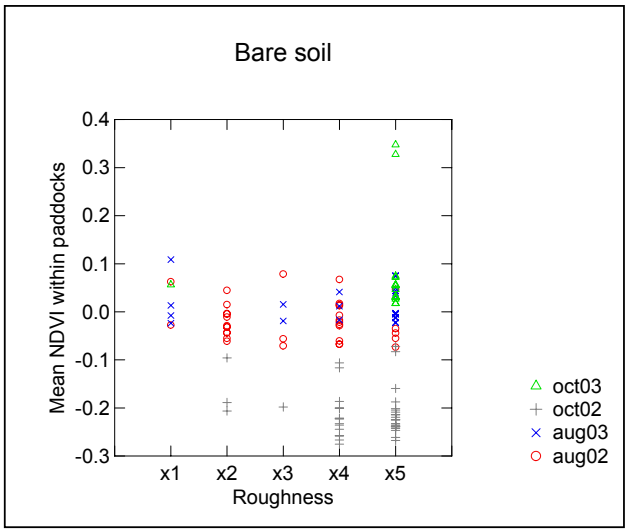

Figure 3. NDVI in bare ground paddocks, grouped by sampling period and roughness of the cultivation (x1 = smooth, x5 = very rough, cloddy).

The range of NDVI values also gives an indication of the impact of radiometric effects between images (figure 4). These histograms have been derived from the entire area of agricultural land (after masking) in each image. NDVI values are highest in September and October 2003. These are the two Landsat 5 images. The October 2002 Landsat 7 image has the lowest values. The SPOT image and the other three Landsat 7

images are grouped in the middle. At this stage we have calculated NDVI using digital numbers (DN) as supplied; we have not as yet converted these to radiance values using the satellite sensors' calibration information. We expect this will have a small contribution to the lack of alignment between the histograms.

The different ranges of the histograms do not appear to correlate with season (August vs October), or with simple rainfall data (as an indicator of soil moisture conditions), and may well be due to the combination of many influences. Ground scene effects are discussed further in the next section. Perhaps the most interesting observation from figure 4 is the characteristic bimodal shape of the NDVI histograms from the spring images versus the unimodal distribution of those from the winter images. This pattern is consistent across both years.

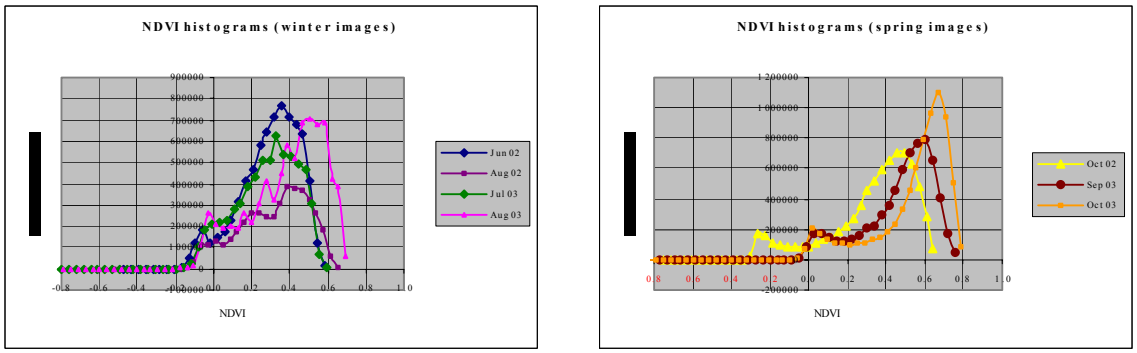

Figure 4. Graphs showing histograms of NDVI values in each image.

The maximum and minimum histogram values (excluding very low frequency outliers) are shown in table 1. The NDVI threshold values that were taken from the August 2002 reference data can be linearly translated to each image using these min–max values. Rerunning the rules using the new thresholds (i.e., corrected for radiometric effects) results in a decrease of 8,963 ha of land identified as being at.

Table 1. Minimum and maximum values of NDVI for each image (very low frequency outliers excluded).

Image	Min	Max
June 2002	–0.17	0.62
Aug 2002	–0.10	0.69
Oct 2002	–0.30	0.68
Jul 2003	–0.12	0.59
Aug 2003	–0.10	0.73
Sep 2003	–0.05	0.79
Oct 2003	–0.02	0.83

The min–max approach is dependent on the distribution of the NDVI values. As this distribution may differ with season (e.g., more green growth in spring), some alternative approaches were tested, including an "end member" approach where bare ground and fully vegetated reference data were used to identify the NDVI endpoints. The thresholds could then be interpolated between these values. It was difficult to identify an accurate NDVI value for the fully vegetated end of the range, however. This is because NDVI continues to rise with the leaf area index of the crop, beyond the point of 100% cover (LAI=1). We used, as far as possible, short-cropped grass as our fully vegetated end member, but even this can be disrupted by the vigour of growth etc. of the grass. The second alternative approach was based on aligning the histograms using the major peak and trough features. This was done successfully amongst the spring images, and amongst the winter images, and then the two sets were aligned to each other using the grouped minimum and maximum for each set.

The difference in total area between the approach leading to the smallest risk area and that leading to the largest risk area is 6,473 ha. The area of land where any one of the correction approaches gave a different risk result to the others is 10,094 ha. This gives an indication of the magnitude of the effect of uncertainty in the thresholds due to radiometric effects between images.

The number of pixels identified as being affected by shadow, cloud or snow varies from 1,049 ha to 245,279 ha in the seven images. This includes pixels that were affected by very thin cloud or haze. In 2002, shadow, cloud and snow resulted in 249,048 ha of agricultural land for which a risk status could not be determined. This is approximately 50%. One approach to obtain a total area is to assume a similar proportion

for risk vs no risk as was found in the area not affected by shadow/cloud/snow. An error or bias in this proportion of say ±2% would result in a difference of ±5,000 ha.

3.3 Model Error

This refers to uncertainty that relates to the interpretation or translation of the remotely sensed reflectance values (NDVI) into information about something in the real world (leaching risk). The translation is a "model". The model may take the form of a classification, or it may be a continuous function. The model may involve more than one image over specified time intervals. Model uncertainty results from many factors including:

- Classification error or error in a continuous model
- Appropriateness of image timing and number
- Scaling issues, including the pixel size not adequately matching the scale of variation in the real world (mixed pixels) (Friedl et al. 2001)
- The change detection algorithm (Singh 1989, Mas 1999, Foody 2002)

The first three uncertainty issues are now discussed.

3.3.1 Classification Error

Classification error may be due to an incorrect algorithm for converting reflectance to real world state, e.g., caused by a poor calibration, or missing categories in the training set. Or it might be because the reflectance is simply not a good representation of real-world state of interest. Note that this is not the same as radiometric error due to sensor, atmospheric or surface effects, but in any analysis involving non-spectral reference data (as in this study), it can be very difficult to separate these from model error.

The rule-based risk classification model as described above includes the following issues:

- Some high pressure-scenarios (due to specific practices or crop types) may not be well described by a percent-cover-based rule, or even be spectrally detectable
- Variability in the relationship between leaching risk and percent vegetation cover
- Variability in the relationship between percent vegetation cover and reflectance (NDVI) due to vegetation type as well as ground scene radiometric error
- Completeness of the rules.

Each of these is now examined.

Spectrally detectable

It is recognized that there are some management practices that would lead to high nitrate leaching are not spectrally distinguishable, e.g., very high fertilizer, stocking or irrigation rates. The output map depicting land use pressure is specifically limited to pressure as a result of land being fallow.

Relationship of nitrate leaching to vegetation percent cover

A calibrated wheat simulation model called WheatCalculator (Jamieson et al. 2003) was used to verify the relationship between percent cover and nitrate leaching, and the appropriateness of the selected thresholds used in the rules. Leaching in excess of 30 kg/ha of nitrate is considered undesirable. A number of 2002 scenarios were simulated with sowing dates at the beginning of each month from April to September for each of a coastal and an inland paddock (slower growth). A wheat crop near Lincoln planted on 1st July will have a 10% cover by 18 August. According to Lilburne et al. (2003) this is likely to have median nitrate leaching in the order of 5–45 kg/ha given soil, fertilizer and climate variability. A crop with 15% cover on 18 August will have been planted about 23 June (near Lincoln), and results in slightly less nitrate leached. However, crops planted earlier than this have much less leaching. This implies that 10 or 15% cover threshold on 18 August does match a change in risk level. A sparse–vegetated threshold of 50% implies a crop is planted in late May, which roughly equates to a median of 2–32 kg/ha nitrate leached. This reduces steadily with earlier planting dates so a 50% threshold is probably reasonable, although 60% might be more suitable (for wheat).

Our rules assume that nitrate leaching correlates with percent vegetation cover. However, what is actually more directly relevant is the depth of the plant root system and its nitrogen uptake. Some crops have a deeper root system than others, for the same percent vegetation cover, due to the soil characteristics or the plant species itself. For example, a horticultural crop may have significant amounts of bare soil between rows yet have a deeper root system and a higher uptake of nitrate, than an arable crop with a higher percentage of

vegetation cover. Our photographic estimate of a broad bean crop was 40% cover but the planting date and size of the plants would indicate low leaching risk despite satisfaction of rule B. Clover is an example of a low-growing high-ground-cover crop that may not have the same N uptake relationship with percent cover as wheat. Clover is also a nitrogen-fixing plant, which increases the risk of leaching. Grazing is another complication as new grass or oats might be grazed, making the crop look younger (i.e., less uptake) than is actually the case. To summarize, our use of percent cover as an indicator of leaching appears to be generally appropriate for grass, barley, greenfeed cereals and wheat, but this may not be the case for other crops.

Relationship of vegetation percent cover to NDVI

Examination of the August 2002 reference dataset (containing 130 data points) shows that 15 sites covering a range of wheat, litter and vegetated paddocks were incorrectly classified by the rules into low, sparse or vegetated. There are eight sites that do not fall within the range 0–15% (the low vegetation class), yet have NDVI values below the 0.11 threshold. These sites are broad bean, lucerne, litter-grass, litter-soil, clover seed and pasture sites. Most of these sites have high levels of litter. A lower threshold for NDVI will not reduce the number of errors. Three additional borderline sites (15% cover) have an NDVI >0.11. However, a higher NDVI threshold would deviate too far from the "low" vegetation class.

The sparsely vegetated class (15–50% vegetation) equates to an NDVI measure in the range of 0.11–0.22. Comparison with the reference data shows three sites (lucerne, horticulture and greenfeed) that have high NDVI but insufficient percent cover, and four sites (greenfeed, pugged and deer pasture) with high cover but low NDVI. The selected NDVI thresholds minimize the misclassifications between the three vegetation classes of low, sparse and vegetated.

Because of the uncertainty in the visual estimates of percent cover (i.e., reference data error), the relationship between percent cover and NDVI was further tested by comparing NDVI values with the digital photos from August 2003. The digital photographs were classified using supervised maximum likelihood classification (in ERDAS Imagine) with training sets for vegetation, litter and soil collected visually within each set of 10 images (10 images were acquired within each field). Once the classification was complete, it was straightforward to quantify the percent cover of green vegetative material for each field (an average from the 10 photographs). While figure 5 shows a clear relationship between NDVI and percent cover, the accuracy of the classification into low/sparse/vegetated was not as encouraging. NDVI thresholds corrected for August 2003 by the three methods described in the radiometric error section correctly identified all the low-vegetation sites, but the upper threshold did not successfully separate the sparse and fully vegetated categories.

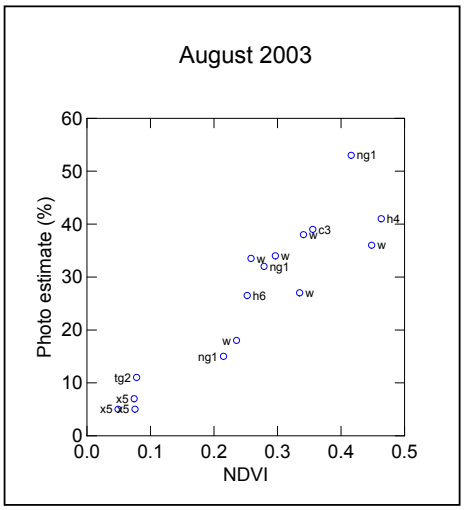

Figure 5. Plot of mean NDVI in a paddock and mean percent cover as determined from 10 photos labelled according to crop type (w=wheat, h=horticulture, ng=new grass, x=cultivated, c=pasture, tg=mix of litter and grass)

It is clear from the various analyses that due to differences in vegetation type, and presence of background litter or moisture affecting the radiance from the target (i.e., ground scene radiometric error), NDVI is an approximate predictor of percent cover, and of nitrate leached. Soil type and roughness do not appear to have a strong influence, perhaps because cropping and grazing both lead to plenty of organic material in the topsoil. Analysis of the reference data showed that approximately 10–20% of the reference points near the thresholds were incorrectly classified, largely due to different vegetation types and ground effects such as moisture and litter. The area of land that might be affected by this was calculated as 15% of the land that has an NDVI value within ±0.08 of the thresholds. This was in the order of 15,000 ha, of which 60% resulted in increased risk and 40% in less risk.

Rule completeness

Rule completeness was assessed by consulting agronomy experts to derive a long list of land use scenarios, including crop rotations, and timing of grazing or cultivation events. The experts assessed each variation in

terms of its likely nitrate uptake. This assessment was compared with the pressure as predicted by the rule (using expert estimates for each scenario's likely percent cover). The only scenarios that failed were ones associated with a pasture cultivation, severe trampling or greenfeed-grazing event immediately preceding the second image. These scenarios were predicted to be risky by the rules but were assessed as being of low risk since in reality the fallow period over the coldest period was very short. The area covered by Rule B (i.e., land where there is a harvest or cultivation event between the two images) is 16,040 ha. If we assume an even temporal distribution of events over the 8 weeks between the two images, then 1/8th of this area (i.e., 2,005 ha) might have an event in the week preceding the second image.

3.3.2 Image Timing

Only two images are being used in the identification of fallow ground. Consequently the timing of these images is very important. Too-early and harvest events will be missed. Too far apart, or too late, and grazing/cultivation events or significant crop growth stages (e.g., from sparse to med/full vegetation) could be missed. It is also important that the critical winter period is covered – this was assumed to be July/August, thus scenes from June and August were acquired.

The approach used to assess these uncertainties was to compare the assumptions in our classification model to output from a crop simulation model. From Webb et al. (2001) it can be seen that in 1991–1994 leaching occurred between April and October depending on the rainfall, but not the same months each year. This implies that the timing of the bare ground state, e.g., mid-winter, is not as important as the length of time that the land is fallow in the wider range of critical months.

This focus on length of time land is fallow points to the third spring image being used in an additional rule that looks for no or sparse vegetation in both image 2 and 3.

- Rule C: a field is fallow in the spring if vegetation is in a low or sparse state in winter (July or August) **and** it is still in a low or sparsely vegetated state in September or October.

This resulted in an additional 10,023 ha being identified as being at risk in 2002, i.e., an increase of 20.7%. However, we note that if **any** 2-month period is generally dry then there is no risk in this period, i.e., the location of land under actual pressure in any given year is dependent on rainfall.

WheatCalculator was used to determine the likelihood of missing transitions between key crop stages, and therefore of either incorrectly identifying vegetated land as being at risk, or missing bare land due to the timing of the images and the length of the period between them. It takes anywhere from 1.25 to 4 months for a wheat crop to reach 50% cover (and therefore high nitrate uptake) depending on the sowing date and location (figure 6). So a crop planted anywhere just before image 1 will not have reached 50% cover by the time of the second image; near Lincoln it will be approximately 15% cover at this time. This means this scenario will be identified by rule A. A wheat crop cultivated and planted just after image 1 (from grass) will range up to 15% cover by the time of image 2, and will hence be covered by Rule B. A crop planted at the time of the second image will have 15–50% cover by the time of the third image, depending on the location. This all suggests that a 2-monthly spacing of the images is appropriate (for wheat).

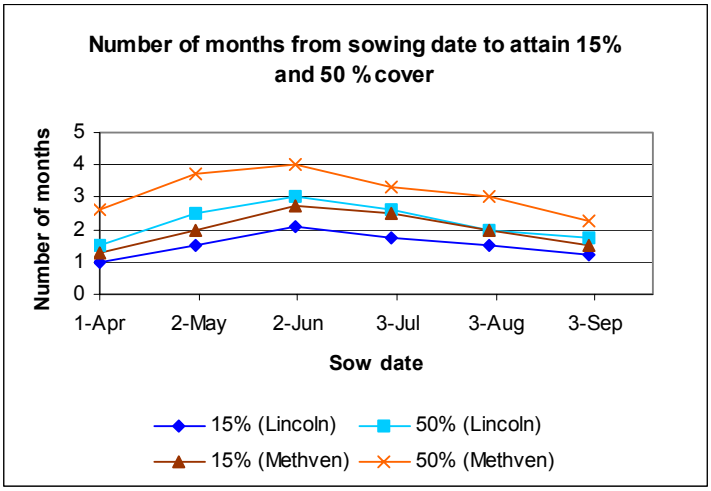

Figure 6. Number of months a wheat crop takes to reach 15% and 50% cover when planted on a given sowing date. Lincoln is near the coast, Methven is inland where it is colder.

3.3.3 Mixed Pixels

As the landscape is very fragmented and heterogeneous, there will be many mixed pixels along the borders of each field. A pixel that incorporates part of a shelterbelt or grass verge on the edge of a fallow field will have a higher NDVI value than the pure-fallow pixel. However, in terms of nitrate leaching, this pixel will be

of less risk as the shelterbelt or grass will uptake some of the nitrate. The effect of mixed pixels on accuracy is expected to be minimal, under the assumption that the reduction in risk due to partial vegetation cover is proportional to the difference in NDVI value.

3.4 Reference Data Error

Reference data spatial error was minimized by using a GPS to locate the sampling site to the nearest 5–10 m. Notes about the direction and shape of the field then allowed the relevant pixels to be easily and accurately delineated, avoiding edge effects. It is assumed that this process has led to minimal error. However, it appears that our ability to make accurate visual estimates of vegetation percent cover is not high. A comparison with the photo estimates (August 2003) shows that 50% of the visual estimates underestimate percent cover by a magnitude of 20%. This will only impact on some of the uncertainty analyses above, not on the risk output.

Similarly, there was a time lag between image acquisition and the fieldwork. Mostly this was just a few days, which would affect estimates of moisture. Any paddocks that showed evidence of very recent cultivation were not used. The October 2003 lag was several weeks. At least two paddocks must have had substantial crop growth in that period – these were dropped from the reference dataset.

The poor comparison between our visual estimates of percent cover and that obtained from the photos leads to the following recommendations for collecting field data. All estimates should be made by a researcher on the field – not from a distance. Within-field variability is so high that more than one estimate is recommended. The content of litter, stones and moisture levels should be noted. Estimates should be made for a wide variety of crops and covers.

3.5 Uncertainty Assessment Summary

3.5.1 Summary of error sources

Each error source described above has been subjectively quantified by considering the area in 2002 that is likely to be affected, and a summary is given in table 2. In terms of the total area of land at risk, the most significant sources are the possibility of missing spring leaching and the radiometric differences between images. The dominant influences on spatial uncertainty are radiometric effects (including ground scene effects) and spring leaching. Cloud results in a very large area for which no risk status can be estimated.

Table 2. Summary of each error source and its impact on total area estimate and its spatial coverage.

Error source	Estimated total area at risk (ha)	Spatial area affected (ha)	Note
Geometric	• –500	• 7,500	• Assumes (worst case) 1-pixel offset
Radiometric			
• Differences between images	• 6,500	• 10,000	• Alternative methods for calculating the thresholds
• Cloud, shadow	• ±5,000	• 250,000	• Area lost due to cloud, shadow or snow (particularly bad in 2002). Error in total area assumes max ±2% bias
Model			
• Classification (& ground scene radiometric effects)	• +3,000	• 15,000	• Based on % error rate on points within specified threshold buffers • Assume 1 week
• Rule completeness	• –2,000	• –2,000	• Assume wet spring
• Timing (spring)	• +10,000	• 10,000	• 2-month interval is OK
• Timing (missed transitions)	• 0	• 0	• Minimal effect
• Mixed pixels	• 0	• 0	
Reference data	• 0	• 0	• Minimal effect on output

The identification of radiometric differences between images as a major error source indicates that a more detailed investigation is warranted in future into the factors influencing this.

3.5.2 Rainfall and Image Timing

The risk of missing spring leaching events (when two images only are used) also prompts consideration of timing in the early autumn period. The first image is best taken at the beginning of the leaching period once the soil has reached field capacity (generally in June). This ensures that most autumn harvest events are caught. Earlier harvest events will not make much difference to risk of leaching as drainage has not yet started. However, in a year with a wet autumn, it may well be appropriate to acquire an additional image in April. Similarly, if the spring is wet the additional spring image might be needed. If a dry period extends for several months then there will be no risk in this time, and therefore no need of images. In the case of an extreme rainfall event, all surplus nitrate will be leached regardless of the soil and crop. Consequently the need for images is very dependent on rainfall.

3.5.3 Supplying Uncertainty Information

A formal analysis of uncertainty of risk requires a reference dataset of leaching risk collected according to a randomized sampling design (Stehman and Czaplewski 1998). This would allow the generation of a confusion matrix (Congalton and Green 1998). The reference data are logistically very difficult and time consuming to obtain due to the:

- need to identify and talk to many landowners,
- need for either visual information throughout the leaching season or crop/planting date, and
- lack of scientific knowledge on leaching risk.

As providers of a map of fallow land, we intend to supply a confusion matrix that relates to the binary classification of land into fallow or not. However, we note that this confusion matrix will not be sufficient on its own to estimate uncertainty in risk of nitrate leaching. Knowing a pixel has *x*% probability of being incorrectly classified as fallow is not the same as being incorrectly classified as at risk of nitrate leaching. As shown above, an understanding of the various error sources and a detailed knowledge of the model are necessary. A confusion matrix also gives no indication of the spatial distribution of uncertainty (Foody 2002).

While not as rigorous, a supplementary approach to providing information on map accuracy is to combine what is known about likely sources of error to develop a reliability or probability map. This allows information on the possible spatial distribution of the error to be supplied. Assumptions used to derive the probabilities should be documented so that the uncertainty map can be easily revised.

Each error source was converted into a spatial map of probability. Geometric error related to the one-pixel shift simulation. This tended to be located around paddock boundaries. Timing error was a combination of the probability of rule (in)completeness (i.e., risk of missing fallow land in spring, and risk of acquiring the July image immediately after a cultivation event). Radiometric error was a combination of between- and within-image differences and model error related to crop type. This tended to be high where there was partial vegetation cover. Due to high levels of within-paddock variability, errors were more fragmented than expected. The three error sources were assumed to be independent and thus combined using equation 1. Some of the error sources, especially those related to radiometric error, may not be independent, but no information is available for estimating conditional probabilities.

$$p(A,B,C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C) \quad (1)$$

where A is geometric error, B is timing error and C is radiometric error.

One way to use the resulting probabilistic uncertainty map is to classify the risk map into four classes: likely to be at risk (risk & probability error <0.5), possibly at risk (risk & probability error >0.5), possibly not at risk (no risk & probability error >0.5), unlikely to be at risk (no risk & probability error <0.5). This is shown in figure 7. A regulatory authority may then choose to focus their environmental efforts on areas in the first category.

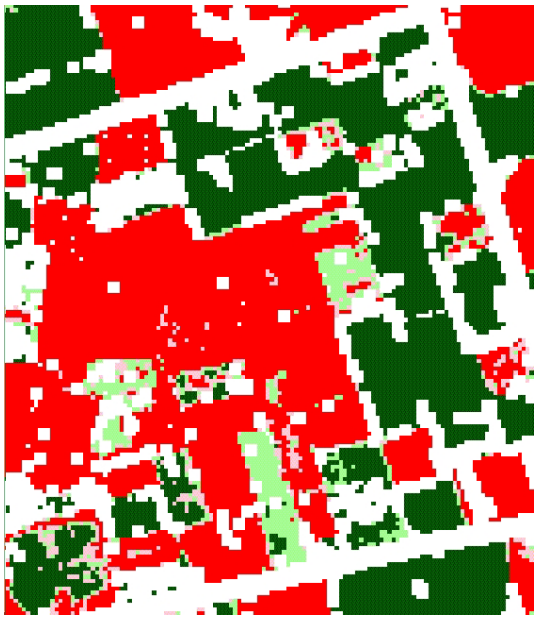

Figure 7. Areas in red are likely to be at risk of nitrate leaching, pink areas are those that might be at risk, light green pixels might be of no risk, and dark green represent those areas that likely to be of no risk. Masked pixels are white.

4. Conclusion

The analyses have shown there are many sources of error that impact on any estimate of land at risk from nitrate leaching on account of its fallow state. A combination of nitrogen cycling crop models, expert advice as to the possible scenarios, error simulation, and reference data has been used to compare the various sources of error. Estimates of the area affected by each source allowed the most significant sources to be identified, and the derivation of a probabilistic map of uncertainty. Radiometric errors were identified as being a significant source.

This approach allows confidence in risk predictions to be represented spatially. It will also guide future developments in the fallow-land mapping methodology. The map of land at risk will eventually be combined with soil, aquifer, and climate maps, and a mechanistic model of nitrate leaching. This error assessment will be compared with assessments of the effect of climate and environmental variability as well as simulation model errors.

Acknowledgements

This work was supported by the New Zealand Foundation for Research, Science and Technology, contract number C09X0306. David Pairman, James Barringer and Trevor Webb are thanked for their helpful comments, as are Stephen Burgham who assisted with the field data and Stella Belliss who prepared the images.

References

Congalton, R. and Green, K. (1998) *Assessing the accuracy of remotely sensed data: principles and practices,* Lewis Publishers, Boca Raton.

Corwin, D. L. (1996) GIS applications of deterministic solute transport models for regional-scale assessment of non-point source pollutants in the vadose zone. In *Application of GIS to the modeling of non-point source pollutants in the vadose zone.* Soil Science Society of America, Madison, Wisconsin, pp. 69-100.

Dai, X. L. and Khorram, S. (1998) The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing,* **36,** 1566-1577.

Foody, G. M. (2002) Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* **80,** 185-201.

Friedl, M. A., McGwire, K. C. and McIver, D. K. (2001) An overview of uncertainty in optical remotely sensed data for ecological applications. *Spatial uncertainty in ecology : implications for remote sensing and GIS applications,* 258-283.

Jamieson, P. D., Armour, T. and Zyskowski, R. F. (2003) On-farm testing of the Sirius Wheat Calculator for N fertiliser and irrigation management. Proceedings of 11th Australian Agronomy Conference, Geelong, Victoria, 2-6 Feb,Australian Society of Agronomy.

Jensen, J. R. (1996) *Introductory digital image processing: a remote sensing perspective,* Prentice-Hall, New Jersey, USA.

Lilburne, L. R., Webb, T. H. and Francis, G. S. (2003) Relative effect of climate, soil, and management of nitrate leaching under wheat production in Canterbury, New Zealand. *Australian Journal of Soil Research,* **41,** 699-709.

Lunetta, R. S. (1999) Applications, project formulation, and analytical approach. In *Remote sensing change detection* (Eds, Lunetta, R. S. and Elvidge, C. D.) Taylor and Francis, London, pp. 1-19.

Mas, J.-F. (1999) Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing,* **20,** 139-152.

Nagler, P. L., Daughtry, C. S. T. and Goward, S. N. (2000) Plant litter and soil reflectance. *Remote Sensing of Environment,* **71,** 207-215.

North, H., Lilburne, L., Pairman, D., Webb, T., Burgham, S., Belliss, S. and McNeill, S. (2002) Mapping agricultural land-use practices using multi-temporal satellite remote sensing. Proceedings of Image & Vision Computing New Zealand, University of Auckland, New Zealand, 26-28 November. pp. 287-292.

Sabins, F. F. (1978) *Remote sensing: principles and interpretation,* Freeman & Co.

Singh, A. (1989) Review Article: Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing,* **10,** 989-1003.

Stehman, S. V. and Czaplewski, R. L. (1998) Design and analysis for thematic map accuracy assessment - fundamental principles. *Remote Sensing of Environment,* **64,** 331-344.

Stow, D. A. (1999) Reducing the effects of misregistration on pixel-level change detection. *International Journal of Remote Sensing,* **20,** 2477-2483.

Tucker, C. J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment,* **8,** 127-150.

Walsh, S. J., Moody, A., Allen, T. R. and Brown, D. G. (1997) Scale dependence of NDVI and its relationship to mountainous terrain. In *Scale in remote sensing and GIS* (Eds, Quattrochi, D. A. and Goodchild, M. F.) Lewis Publishers, Boca Raton, pp. 27-55.

Webb, T. H., Lilburne, L. R. and Francis, G. S. (2001) Validation of the GLEAMS simulation model for estimating net nitrogen mineralisation and nitrate leaching under cropping in Canterbury. *Australian Journal of Soil Research,* **39,** 1015-1025.