Proof of manuscript published as:
Goodwin, N.R., S. Magnussen, N.C. Coops, and M.A. Wulder. 2010. Curve fitting of time series Landsat imagery for characterising a mountain pine beetle infestation disturbance. International Journal of Remote Sensing, Vol. 31, No. 12, pp. 3263-3271. (DOI: http://dx.doi.org/10.1080/01431160903186277).

# Curve fitting of time-series Landsat imagery for characterizing a mountain pine beetle infestation

NICHOLAS R. GOODWIN*†, STEEN MAGNUSSEN‡, NICHOLAS C. COOPS† and MICHAEL A. WULDER‡

†Department of Forest Resource Management, 2424 Main Mall. University of British Columbia, Vancouver, V6T 1Z4, Canada

‡Canadian Forest Service (Pacific Forestry Center), Natural Resources Canada, 506 West Burnside Road, Victoria, V8Z 1M5, Canada

(*Received 29 May 2008; in final form 12 October 2008*)

In this technical note we present a new technique using mixed linear models for characterizing a mountain pine beetle (*Dendroctonus ponderosae* Hopkins) infestation from multiyear satellite imagery. The main benefit of our approach is an ability to determine the statistical significance of each annual spectral change. Knowledge of the annual spectral change characteristics can then be used to statistically determine if a disturbance event has occurred, the timing of a given disturbance event, as well as to provide information for clustering fitted multitemporal reflectance curves (i.e. spectral trajectories) with a common shape. The spatial clustering of spectral trajectories provides insights upon the temporal process towards understanding the nature of the disturbance and recovery evident as imposed by infestation by mountain pine beetle over a 14-year period.

## 1. Introduction

Time-series analysis of satellite imagery has been demonstrated as a tool to characterise disturbance events in forested areas (Coppin *et al.* 2004, Kennedy *et al.* 2007, Roder *et al.* 2008). However, a number of challenges remain when trajectory-based methods are used to cluster fitted multitemporal reflectance curves, particularly for disturbance events where the forest canopy has not been entirely removed (e.g. mountain pine beetle, *Dendroctonus ponderosae* Hopkins) and the magnitude of spectral change is low. The major difficulty relates to the discrimination of real disturbance events from noise. Given that changes in reflectance have been found to occur both within and between image dates for reasons other than a forest disturbance event. For example, seasonality, silvicultural practices, understorey vegetation, geometric registration, radiometric normalization, atmospheric conditions and correction, and sensor calibration issues were indicated to influence reflectance values (Radeloff *et al.* 1999, Franklin *et al.* 2005, Schroeder *et al.* 2006). In addition, gaps in the data record may occur from excessive cloud cover (Ju and Roy 2008) limiting the ability to detect physiological or forest disturbance events (Wilson and Sader 2002, Jin and Sader 2005).

In this technical note we demonstrate an approach for the estimation, classification and statistical inference of spectral trajectories. We demonstrate the approach for the

*Corresponding author. Email: nicholas.goodwin@nrw.qld.gov.au

*International Journal of Remote Sensing*
ISSN 0143-1161 print/ISSN 1366-5901 online © 2010 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160903186277

purpose of extracting the area and timing of disturbance events of infestation over a 14-year period in the central interior of British Columbia, Canada.

## 2. Methods

### 2.1 *Data characteristics*

Seven Landsat-5 Thematic Mapper (TM) and one Landsat-7 Enhanced Thematic Mapper (ETM+) scenes were acquired between 26 June 1992 and 3 October 2006 over the Morice Timber Supply Area (TSA), British Columbia (Path/Row: 51/22; figure 1). This site was chosen due to the availability of aerial survey data which documented the presence and timing of infestation across the study area, with the current infestation first reported in the mid-1990s. For a more detailed study area description, refer to Nelson *et al.* (2006) and Goodwin *et al.* (2008).

Data pre-processing involved two critical steps. First, image-to-image geometric registration was undertaken using a nearest-neighbour 2$^{nd}$ order polynomial transformation. The 2001 ortho-rectified image was used as the base image due to a lack of cloud cover. All other images were rectified with a root mean square error $< 0.5$ pixels ($< 15$ m). Areas of cloud and cloud shadow were removed via manual interpretation. Second, radiometric normalization of images was undertaken to ensure that changes in spectral reflectance between years correspond to meaningful physiological events (Chen *et al*. 2005); whereby we atmospherically corrected the 2001 base image to derive surface reflectance using the COST model (corrects for cosine of the solar zenith angle) (Chavez 1996) then utilized the Multivariate Alteration Detection (MAD) algorithm (Canty *et al*. 2004, Schroeder *et al*. 2006) which utilizes canonical correlation analysis to locate invariant pixels for use in matching the remaining

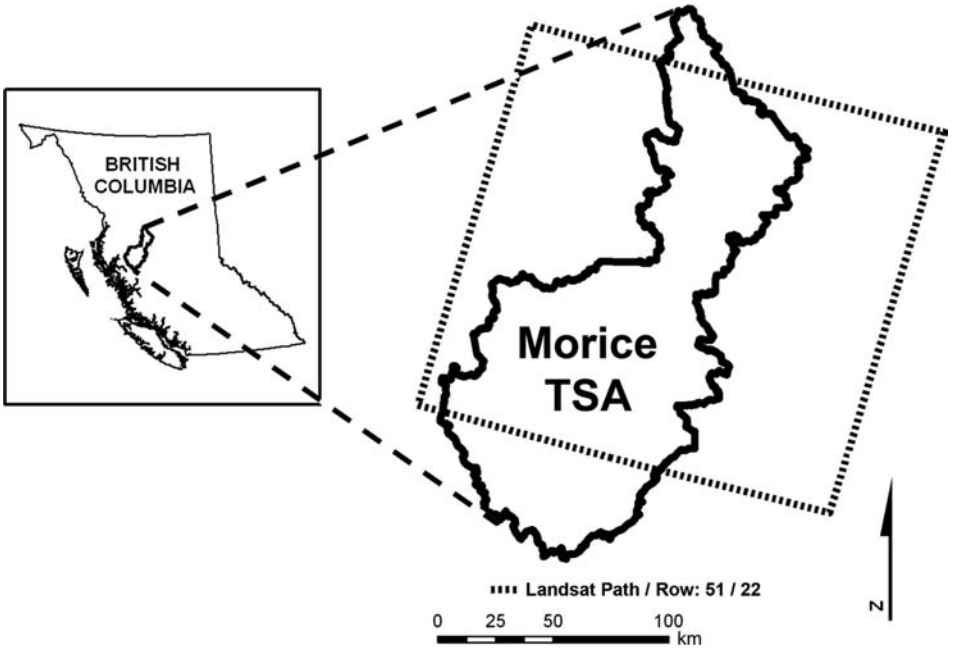



Figure 1. Study site in Morice Timber Supply Area.

images to the atmospherically corrected base image. Our implementation of the MAD algorithm, followed that of Schroeder *et al*. (2006), differing only through our use of four 1000 × 1000 pixel image subsets (instead of one) to locate invariant pixels.

Following data pre-processing, TM band 4 (near-infrared, NIR) and TM band 5 (mid-infrared, MIR) were used to calculate the Normalized Difference Moisture Index (NDMI) for each image and converted to integer format (multiplied by 10 000) using Eq (1). This metric was chosen due to its proven sensitivity to quantify forest disturbance events (Wilson and Sader 2002, Jin and Sader 2005).

$$\mathrm{NDMI} = \frac{\mathrm{NIR} - \mathrm{MIR}}{\mathrm{NIR} + \mathrm{MIR}}. \tag{1}$$

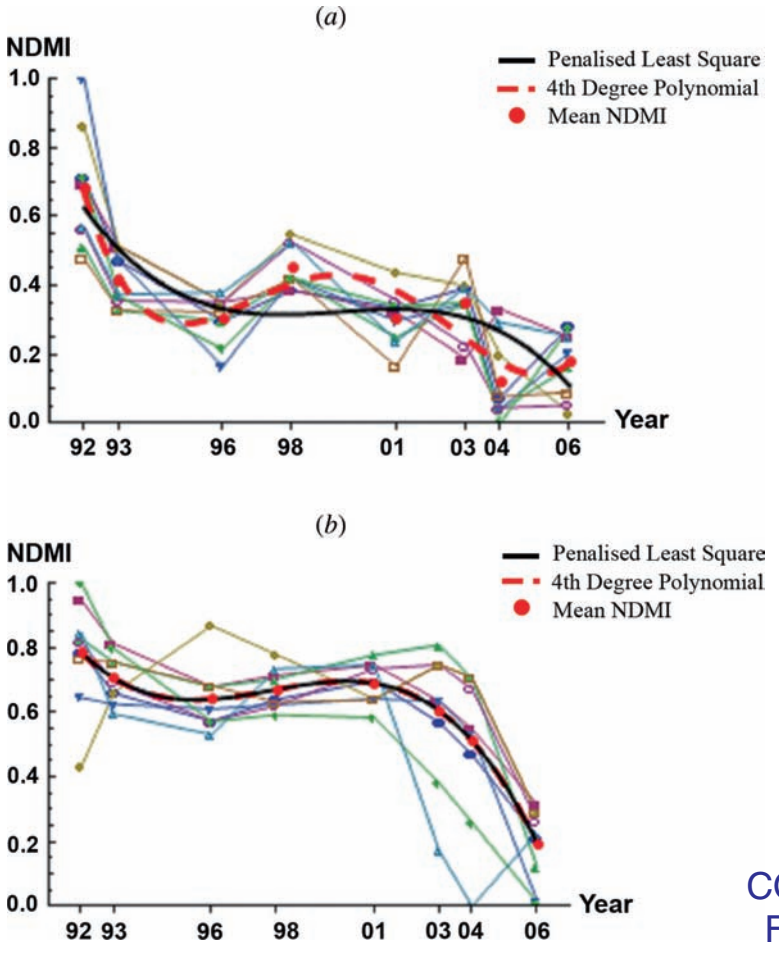
### 2.2 *Background to mixed linear modelling*

The documented change in a pixel's value over time may be considered as a sample derived from a continuous trajectory or spectral curve. Assuming enough samples are taken to adequately represent the trajectory, functions may then be used to capture and describe underlying trends in pixel trajectories. In a functional approach to classification and clustering, all trajectories must be fitted to the same function (McLachlan 1991). For a polynomial fit this means that the degree must be high enough to accommodate all curve shapes in the data. However, overfitting (i.e. too many model parameters) will lead to adverse effects on results and is a problem that needs to be addressed. A common approach to address overfitting is to add a model term that penalises curvature (Ramsay and Silverman 1997, Heckman and Ramsay 2000). When data consist of repeat observations on individual units (pixels) and the function of choice is a polynomial in time, a mixed linear model approach can accommodate the penalty term by casting units as random effects that interacts with time (McCulloch and Searle 2001, Ruppert *et al*. 2003). We adopt this mixed linear modelling approach for capturing time trends in NDMI.

### 2.3 *Mixed linear modelling*

A set of 199 beetle attack locations surveyed between 1996 and 2004 were first extracted to examine spectral trajectories of beetle attack. At each site, NDMI values were obtained for $n = 9$ pixels arranged in a 3 × 3 window and $m = 8$ years.

After a location specific scaling of NDMI to the unit interval [0,1] and a corresponding scaling of time (year) we found that a fourth degree ordinary least-squares polynomial fitted to the 9 (pixels) × 8 (years) observations captured the overall temporal trends in the data quite well (relative errors on a mean of nine pixels $< 5\%$) (figure 2). As may be expected, there are examples of apparent overfitting where changes in NDMI – not considered as beetle attack (*i.e.* noise) – were being represented by the model. To incorporate the 3 × 3 window structure of the data and to combat overfitting we considered linear polynomial models with a quintic, quartic, and cubic truncated power basis along with random pixel and pixel × time interactions (Ruppert *et al*. 2003). Based on Akaike's Information Criterion (AIC) adapted to model comparison for 3 × 3 window data (Vaida and Blanchard 2005, Gurka 2006) we found that a cubic truncated power basis to be adequate. Specifically, for a single location the model trajectory is:




Q5

Figure 2. Examples of observed pixel-level temporal trends in NDMI (thin coloured lines) and average pixel penalized least-squares third degree polynomial (PLS3, thick black line). The average of NDMI at years of observation is indicated by red dots. A fourth degree polynomial fitted by ordinary least-squares (thick red dashed line) illustrates both overfitting (top) and a nearly perfect match (bottom, hidden behind PLS3).

$$\mathrm{NDMI}_{ij} = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 t_j^3 + \alpha_i + \sum_{k=1}^{8} \left(t_j - (K_k)\right)_+^3 + \varepsilon_{ij}, \tag{2}$$

where $NDMI_{ij}$ is the scaled $NDMI$ of pixel $i$ ($i = 1,\ldots, n = 9$) at (scaled) time $t_j$ ($j = 1,\ldots, m = 8$), $\beta_{0,\ldots,}\beta_3$ are fixed effects regression coefficients, $\alpha_i$ is a random pixel effect, $\alpha_{ik}$ is a random pixel $\times$ time interaction (slope) term, $K_k$ is a time (knot) placed at midpoints between observation years and half a year before the first observation i.e.

$$K_k = 0.5(t_{k-1} + t_k), \tag{3}$$

where $k = 2,\ldots, 8$ and $K_1 = t_1 - \frac{1}{30}$, $(t - K)_+$ is the positive part of $(t - K)$, and $\varepsilon_{ij}$ is a random zero mean error term with (assumed) constant variance. Estimates of fixed ($\beta$) effects were obtained via restricted maximum likelihood methods while random effects were best linear unbiased predictions (Ruppert *et al.* 2003). The location

specific estimates of the coefficients $\beta_{0,\ldots,}\beta_3$ define the average time-trajectory of NDMI. The estimate of the variance covariance matrix of $\hat{\beta}_0, ..., \hat{\beta}_3$ is obtained as outlined in Pinheiro and Bates (2000). Application of the delta-technique (Kendall and Stuart 1969) provides the following estimate of the variance of the penalized cubic polynomial for NDMI at time $t \in [0, 1]$:

$$\mathrm{var}\left(\overline{\mathrm{NDMI}}_t\right) = \frac{1}{n-4}\sum_{i=0}^{3}\sum_{j=0}^{3}\hat{\sigma}\left(\beta_{ij}\right)t^i t^j + \frac{\hat{\sigma}_\alpha^2}{n} + \frac{\hat{\sigma}_{\alpha_k}^2}{n}\sum\nolimits_k (t-K_k)_+^{2\times 3} + \frac{\hat{\sigma}_\varepsilon^2}{nm}, \quad (4)$$

where $\hat{\sigma}\left(\beta_{ij}\right)$ is the estimated covariance between regression parameters $\beta_i$ and $\beta_j$ $i,j = 0,\ 1,\ 2,\ 3$, $\hat{\sigma}_\alpha^2$ is the estimate of the among-pixel variance, $\hat{\sigma}_{\alpha_k}^2$ is pixel × time interaction variance, and $\hat{\sigma}_\varepsilon^2$ is the estimate of the variance of the residual errors in equation (2).

Estimates of the annual change in the NDMI value of a location specific average pixel are obtained as simple differences $\overline{\mathrm{NDMI}}_{t+\Delta t} - \overline{\mathrm{NDMI}}_t$, where $\Delta t$ is the time interval in units of $t$ corresponding to a calendar year. The variance of this difference was estimated as:

$$\begin{aligned}\mathrm{var}\left(\overline{\mathrm{NDMI}}_{t+\Delta t} - \overline{\mathrm{NDMI}}_t\right) &= \frac{1}{n-4}\sum_{i=0}^{3}\sum_{j=0}^{3}\hat{\sigma}\left(\beta_{ij}\right)(t+0.5\Delta t)^i(t-0.5\Delta t)^j \\ &+ \frac{1}{n}\sum_k\sum_{\mathrm{K}}\hat{\sigma}_{\alpha_k}\hat{\sigma}_{\alpha_{k'}}(t+0.5\,\Delta t-\kappa_k)_+^3(t-0.5\,\Delta t-\kappa_{k'})_+^3 + \frac{2\hat{\sigma}_\varepsilon^2}{nm},\end{aligned} \quad (5)$$

where it is assumed that the estimate of change and its variance are centred in the interval for which the change occurred. We can use also use equation (5) to gauge the significance of an annual change.

## 3. Results

The estimate one year change for the years of observation is indicated by circles and an approximate 95% confidence interval has been added (dotted line) (figure 3). A nominal (pointwise) 95% confidence band entirely above or below the zero-line

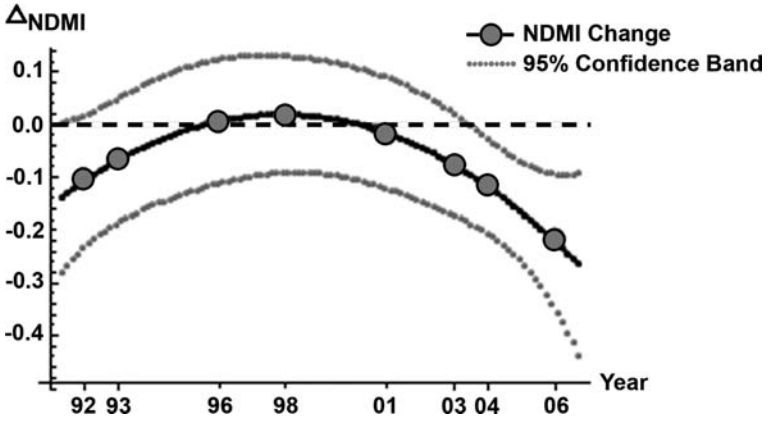



Figure 3. Illustration of an annual change trajectory with an approximated 95% confidence band.

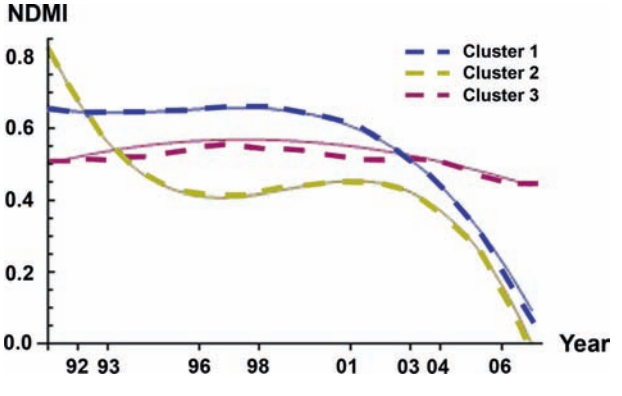
Table 1. Cluster averages and standard error of the mean (SE) of the estimates of the three regression coefficients in the penalized least-squares model in equation (2).

| | Cluster 1 | | Cluster 2 | | Cluster 3 | |
|---|---|---|---|---|---|---|
| | Mean | SE | Mean | SE | Mean | SE |
| $\beta_1$ | 0.64 | 0.20 | 0.68 | 0.23 | 0.96 | 0.26 |
| $\beta_2$ | 1.97 | 0.44 | 1.77 | 0.52 | 2.34 | 0.55 |
| $\beta_3$ | 1.81 | 0.29 | 1.55 | 0.36 | 1.90 | 0.35 |

suggests a statistically significant change in NDMI at the 5% level. Our example in figure 2 indicates a significant NDMI decline in 2004. Of relevance to mountain pine beetle management, knowledge of the year of attack data is instrumental for salvage harvesting, timber supply forecasting, and fire suppression activities.

The three regression coefficients $(\hat{\beta}_0, ..., \hat{\beta}_3)$ in equation (2) provide the basis for a functional discrimination of the *NDMI* trajectories to unique change-class types (table 1) or a clustering of similar trajectory shapes (Tarpey 2007, Wang *et al.* 2007). We demonstrate the latter by a clustering of coefficient triplets that minimises the within cluster dissimilarity and maximises the inter-cluster dissimilarity (Everitt *et al.* 2001). The dissimilarity measure we used was the Mahalanobis distance between the coefficients from two locations (McLachlan 1991), which takes into consideration both the variance and covariance of the coefficients. The clustering algorithm was a modified *k*-means procedure with the number of groups determined by optimizing a Silhouette test statistic (Kaufman and Rousseeuw 1990).

Within the Morice set of trajectories, three clusters of size 92, 59 and 49 were identified. The mean and median trajectory of each cluster is shown in figure 4; the close match between the mean and median trajectory confirms a regular distribution of trajectories within each cluster. Large declines in NDMI are likely to indicate a forest that is experiencing beetle attack. Possible scenarios for the three trajectory types are that areas experienced: (1) low levels of beetle attack as shown by the limited change in NDMI over the years of observation, (2) two distinct disturbance events




Q5

COLOUR FIGURE

Figure 4. Mean (full line) and median (dashed lines) trajectories in three clusters identified by a modified *k*-means procedure with the number of groups determined by optimizing a Silhouette test statistic (Kaufman and Rousseeuw 1990).

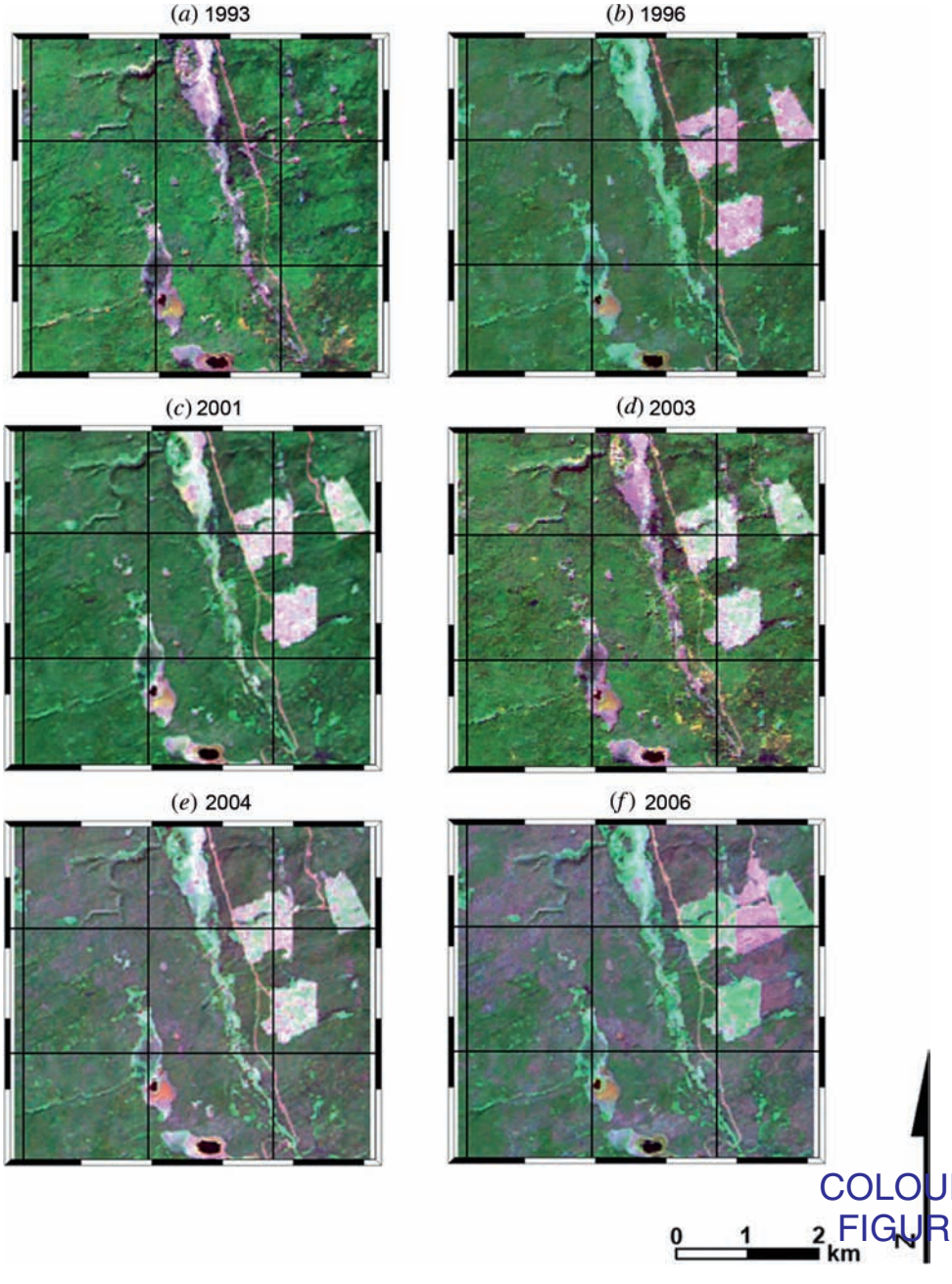



Figure 5. Multiyear observations of mountain pine beetle attack between 1993 and 2006 in the Morice Timber Supply Area. The RGB band combination used was Landsat TM bands 5, 4, and 3 with a 2 SD stretch. Dark green areas represent healthy forest, reddish areas MPB, and harvested areas light green or magenta.

with NDMI decline occurring between 1992 and 1996, and between 2003 and 2006, and (3) a single disturbance between 2003 and 2006. These interpretations are consistent with earlier assessments at Morice which indicated the onset of a beetle attack in the mid-1990's that continued with an increased intensity after 1999 (Nelson *et al.*

2006). This trend of progressive beetle attack is illustrated in figure 5 with expanding orange to brown areas associated with beetle attack, particularly in the years 2004 and 2006.

## 4. Discussion and conclusions

When applying the approach, two key issues present themselves: (1) the spatial scale of disturbance, and (2) temporal resolution of available data. In this study, a ~ 1 ha window (3 × 3 pixels with a resolution of 28.5 m) was used as registration errors greater than one pixel may exist within the data set, as discussed by (Kennedy *et al.* 2007); therefore, we acknowledge that strong signals suggesting defoliation or some other disturbance from a minority of pixels in a ~1 ha window size can be suppressed. Missing observation years will also negatively impact change detection and trajectory analysis (James and Sugar 2003, Franklin *et al.* 2005, Fisher *et al.* 2006). A de facto interpolation of missing observations can seriously distort a trajectory causing a signal to be missed or mis-represented. Additionally, using a multiyear set of images acquired from the same month ('anniversary dates') is likely to improve the accuracy of results by limiting spectral variation due to seasonality effects.

The mixed linear modelling approach appears suitable for extracting information on disturbance events from Landsat-based multitemporal spectral trajectories. The proposed spatial and temporal decomposition of the observed variance expands our inference options. More simple approaches often ignore the dependence structure in the data with a risk of biased inference (Pinheiro and Bates 2000).

## Acknowledgements

This project is funded by the Government of Canada through the Mountain Pine Beetle Initiative, a six-year, $40 million programme administered by Natural Resources Canada, Canadian Forest Service. Additional information on the Mountain Pine Beetle Initiative may be found at: http://mpb.cfs.nrcan.gc.ca.

## References

Canty, M.J., Nielsen, A.A. and Schmidt, M., 2004, Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment*, **91**, pp. 441–451.

Chavez, P.S., 1996, Image-based atmospheric corrections – revisited and improved. *Photogrammetric Engineering and Remote Sensing*, **62**, pp. 1025–1036.

Chen, X., Vierling, L. and Deering, D., 2005, A simple and effective radiometric correction method to improve landscape change detection across sensors and across time. *Remote Sensing of Environment*, **98**, pp. 63–79.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004, Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, **25**, pp. 1565–1596.

Everitt, B., Landau, S. and Leese, M., 2001, *Cluster Analysis* (New York: Arnold).

Fisher, J.I., Mustard, J.F. and Vadeboncoeur, M.A., 2006, Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment*, **100**, pp. 265–279.

Franklin, S.E., Jagielko, C.B. and Lavigne, M.B., 2005, Sensitivity of the Landsat enhanced wetness difference index (EWDI) to temporal resolution. *Canadian Journal of Remote Sensing*, **31**, pp. 149–152.

Goodwin, N.R., Coops, N.C., Wulder, M.A., Gillanders, S., Schroeder, T.A. and Nelson, T., 2008, Estimation of insect infestation dynamics using a temporal sequence of Landsat data. *Remote Sensing of Environment*, **112**, pp. 3680–3689.

GURKA, M.J., 2006, Selecting the best linear mixed model under REML. *American Statistician*, **60**, pp. 19–26.

HECKMAN, N.E. and RAMSAY, J.O., 2000, Penalized regression with model-based penalties. *Canadian Journal of Statistics – Revue Canadienne De Statistique*, **28**, pp. 241–258.

JAMES, G.M. and SUGAR, C.A., 2003, Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, **98**, pp. 397–408.

JIN, S.M. and SADER, S.A., 2005, Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment*, **94**, pp. 364–372.

JU, J. and ROY, D.P., 2008, The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment*, **112**, pp. 1196–1211.

KAUFMAN, L. and ROUSSEEUW, P., 1990, *Finding Groups in Data* (New York: Wiley).

KENDALL, M. and STUART, A., 1969, *The Advanced Theory of Statistics* (London: Charles Griffin).

KENNEDY, R.E., COHEN, W.B. and SCHROEDER, T.A., 2007, Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, **110**, pp. 370–386.

MCCULLOCH, C.E. and SEARLE, S.R., 2001, *Generalized, Linear, and Mixed Models* (New York: Wiley).

MCLACHLAN, G.J., 1991, *Discriminant Analysis and Statistical Pattern Analysis* (New York: Wiley).

NELSON, T., BOOTS, B. and WULDER, M.A., 2006, Large-area mountain pine beetle infestations: Spatial data representation and accuracy. *Forestry Chronicle*, **82**, pp. 243–252.

PINHEIRO, J.C. and BATES, D.M., 2000, *Mixed-Effects Models in S and S-Plus* (New York: Springer).

RADELOFF, V.C., MLADENOFF, D.J. and BOYCE, M.S., 1999, Detecting jack pine budworm defoliation using spectral mixture analysis: Separating effects from determinants. *Remote Sensing of Environment*, **69**, pp. 156–169.

RAMSAY, J.O. and SILVERMAN, B.W., 1997, *Functional Data Analysis* (New York: Springer).

RODER, A., HILL, J., DUGUY, B., ALLOZA, J.A. and VALLEJO, R., 2008, Using long time series of Landsat data to monitor fire events and post-fire dynamics and identify driving factors. A case study in the Ayora region (eastern Spain). *Remote Sensing of Environment*, **112**, pp. 259–273.

RUPPERT, D., WAND, M. and CARROLL, R., 2003, *Semiparametric Regression* (Cambridge: Cambridge University Press:).

SCHROEDER, T.A., COHEN, W.B., SONG, C.H., CANTY, M.J. and YANG, Z.Q., 2006, Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment*, **103**, pp. 16–26.

TARPEY, T., 2007, Linear transformations and the k-means clustering algorithm: Applications to clustering curves. *American Statistician*, **61**, pp. 34–40.

WANG, X.H., RAY, S. and MALLICK, B.K., 2007, Bayesian curve classification using wavelets. *Journal of the American Statistical Association*, **102**, pp. 962–973.

WILSON, E.H. and SADER, S.A., 2002, Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, **80**, pp. 385–396.