

# IMPROVING NDVI TIME SERIES CLASS SEPARATION USING AN EXTENDED KALMAN FILTER

†\*W. Kleynhans, †‡J.C. Olivier, †\*B.P. Salmon, \*K.J. Wessels, \*F van den Bergh

† Electrical, Electronic and Computer Engineering University of Pretoria, South Africa    \*Remote Sensing Research Unit Meraka Institute, CSIR, Pretoria, South Africa    ‡Defense, Peace, Safety and Security, CSIR, Pretoria, South Africa

## ABSTRACT

It is proposed that the NDVI time series derived from MODIS multitemporal remote sensing data can be modelled as a triply (mean, phase and amplitude) modulated cosine function. A non-linear Extended Kalman Filter was developed to estimate the parameters of the modulated cosine function as a function of time. It was shown that the maximum separability of the parameters for different vegetation land cover was better than that of a spectral method based on the Fast Fourier Transform (FFT). Thus it is theorized that the cosine function parameters estimated using the EKF is superior for both classifying land cover and detecting change over time when compared to methods based on the FFT. Results from two study areas in Southern Africa are provided to show the improved separability using MODIS data.

## 1. INTRODUCTION

The use of multitemporal remote sensing data in the evaluation of vegetation land cover is widely accepted to be superior to single date imagery [1, 2]. Multi-temporal coarse resolution satellite imagery such as Moderate-resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) have been widely used to map land cover at regional to global scales [3, 4]. Land cover classification methods are often based on a series of secondary metrics derived from the Normalized Difference Vegetation Index (NDVI) time series and include Principal Component Analysis (PCA) [2, 5, 6], phenological metrics [7] or Fourier (spectral) analysis [8, 9]. The Fast Fourier Transform (FFT) is often used when evaluating NDVI time series data [8, 9, 10]. In many applications where the FFT transformation of NDVI time series data is used for classification and segmentation, only the first few FFT components are considered as they tend to dominate the spectrum [8, 9]. It has been found that even when considering only the mean and seasonal FFT components [9], reliable class separation can be achieved. A drawback of using FFT-based methods is that the underlying process is assumed to be stationary. This assumption is often invalid in the case of NDVI time series data, especially if a land

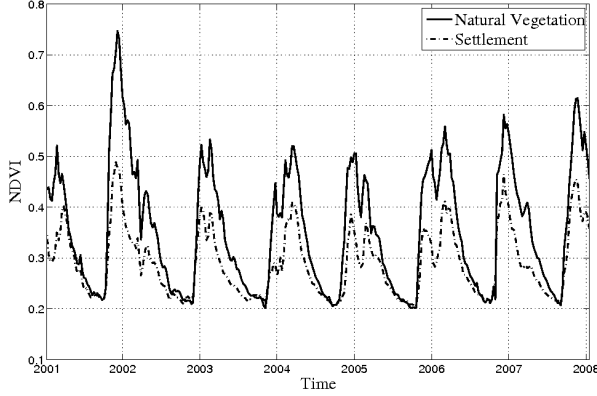
cover change is present. The extended Kalman Filter (EKF) is a non-linear estimation method that can potentially be employed to estimate unobserved parameters (process model) using noisy observations of a related measurement model. EKF techniques in remote sensing have been used for parameter estimation of values related to physical, biogeochemical processes or vegetation dynamics models [11, 12].

It is firstly proposed that the NDVI time series be modeled as a single, but triply modulated cosine function, where the mean  $\mu$ , amplitude  $\alpha$  and the phase  $\phi$  are functions of time. Secondly, it is proposed that a non-linear EKF be used to estimate these parameters as a function of time for each NDVI time series.

Using MODIS MOD43 data from two study areas, it is shown that the  $\mu$  and  $\alpha$  parameter streams over time are similar for same class land cover types and dissimilar for different land cover types representing natural vegetation and settlement land cover types in the Limpopo province of Southern Africa. The parameter sequence can thus be used to determine the level of similarity between NDVI time series belonging to different land cover types.

## 2. DATA DESCRIPTION

The proposed method was tested in two regions in the Limpopo province. The first study area (Region A) is centred around latitude  $24^{\circ}17'21.43''S$  and longitude  $29^{\circ}39'42.96''E$  and is 43 km south east of the city of Polokwane. Region A covers a geographic area of approximately  $190 \text{ km}^2$ ; 42 natural vegetation and 42 settlement pixels were selected for analysis. Region B is centred around latitude  $24^{\circ}19'51.50''S$  and longitude  $29^{\circ}18'04.07''E$  and is 47 km south west of the city of Polokwane. Region B covers a geographical area of  $100 \text{ km}^2$ ; 76 settlement and 52 natural vegetation pixels were selected. Each of the MODIS pixels were evaluated using SPOT5 high resolution data to ensure that none of them had experienced any land cover change during the study period. The NDVI time series data was derived from 8-daily composite MCD43 bidirectional reflectance distribution function (BRDF)-corrected, MODIS data with a spatial resolution



**Fig. 1.** Mean NDVI Time Series for Natural Vegetation and Settlement landcover in region A

of 500m [13] for the period 2001/01 to 2008/01. Figure 1 shows the mean NDVI time series for natural vegetation and settlement pixels in Region A.

### 3. METHODOLOGY

#### 3.1. FFT Method

The Fourier analysis of the NDVI time series has proved to be insightful because the signal can be decomposed into a series of cosine waves with varying amplitude, phase and frequency. The similarity of any two arbitrary NDVI time series can be evaluated by computing their FFT transformation respectively and then comparing the mean and seasonal FFT component of each FFT series [9]. A distance metric between two NDVI time series can be computed as

$$D_{\mu}^{FFT} = \|Y_0^1 - Y_0^2\|, \quad (1)$$

and

$$D_{\alpha}^{FFT} = \|2(Y_7^1 - Y_7^2)\|. \quad (2)$$

Where  $D_{\mu}^{FFT}$  and  $D_{\alpha}^{FFT}$  is the Euclidean distance between the mean and annual FFT components respectively of two NDVI time series.

#### 3.2. New EKF method

It is proposed that a NDVI time series for a given pixel can be modelled by a triply modulated cosine function given as

$$y_k = \mu_k + \alpha_k \cos(\omega k + \phi_k) + v_k, \quad (3)$$

where  $y_k$  denotes the observed value of the NDVI time series at time  $k$  and  $v_k$  is the noise sample at time  $k$ . The noise is additive but with an unknown distribution. The cosine function

is based on a number of parameters (that are not directly observable), namely the frequency  $\omega$ , the nonzero mean  $\mu$ , the amplitude  $\alpha$  and the phase  $\phi$ . The angular frequency was explicitly computed as  $\omega = 2\pi f$  where  $f$  is based on the annual vegetation growth cycle. Given the 8-day composite MCD43 MODIS data,  $f$  was equal to  $8/365$ . The values of  $\mu_k$ ,  $\alpha_k$  and  $\phi_k$  are functions of time and must be estimated given  $y_k$  for  $k \in 1, \dots, N$ . The estimation of these parameters is non-trivial and require a non-linear estimator. According to the EKF formulation, for every increment of  $k$  (the discrete time) a state vector  $\mathbf{x}_k$  is defined containing the parameters to be estimated in the form  $\mathbf{x}_k = [\mu_k \ \alpha_k \ \phi_k]^T$ . The relation between  $\mathbf{x}_k$  and  $\mathbf{x}_{k-1}$  is given by  $\mathbf{v}$ , a known but possibly non-linear function. The state vector  $\mathbf{x}_k$  is related to the observation vector  $\mathbf{y}_k$  via a non-linear measurement function  $\mathbf{h}$ . Both these models are possibly nonperfect, so the addition of process  $w_k$  and measurement  $v_k$  noise is used. This is expressed as [14]

$$\mathbf{x}_k = \mathbf{v}(\mathbf{x}_{k-1}) + \mathbf{w}_k, \quad (4)$$

and

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k. \quad (5)$$

The state vector parameters may be estimated over time  $k$  by recursive iteration [14] based on the observation data  $\mathbf{y}_k$  up to time  $k$ . In the observation equation (5),  $\mathbf{y}_k$  is the predicted measurement. Function  $\mathbf{h}$  is used to compute a measurement given the predicted state, and  $\mathbf{v}_k$  is the observation noise vector.

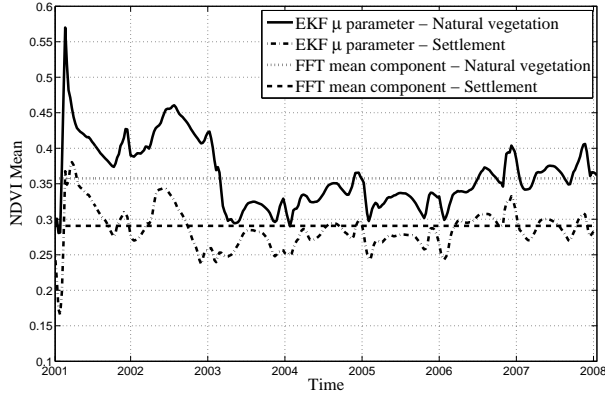
The estimated values for  $\mathbf{x}_k = [\mu_k \ \alpha_k \ \phi_k]^T$  over time  $k$  effectively results in a time series for each of the three parameters. The next step was to define metrics to measure class separability corresponding to land cover type. As shown in [9], substantial separability can be achieved when comparing mean and annual FFT components of NDVI time series of different land cover types. In the proposed EKF framework, the mean FFT component (i.e FFT component 0) corresponds to  $\mu$  and the annual FFT component (i.e FFT component 7 for the seven year NDVI time series) corresponds to  $\alpha$ . Hence it is proposed that within the EKF framework a separability or distance metric between two NDVI time series be defined as

$$D_{\mu}^{EKF} = \max\{\mu_{k,1} - \mu_{k,2}\}, \quad 1 \leq k \leq N \quad (6)$$

and

$$D_{\alpha}^{EKF} = \max\{\alpha_{k,1} - \alpha_{k,2}\}, \quad 1 \leq k \leq N. \quad (7)$$

$D_{\mu}^{EKF}$  is the maximum distance between the first ( $\mu_1$ ) and second ( $\mu_2$ ) parameter sequence over time  $k$ .  $D_{\alpha}^{EKF}$  is calculated in a similar manner finding the maximum distance between the annual amplitude parameter sequence. For the present case it was assumed that the state vector  $\mathbf{x}$  does not change significantly when time is advanced by one, hence  $v = 1$  and the process model is linear. The measurement model, however, contains the cosine term, and as such is evaluated



**Fig. 2.** Comparing the EKF derived  $\mu$  parameter with the FFT mean component for natural vegetation and settlement for region A

via the standard Jacobian formulation, thereby linearising the non-linear measurement model around the current state vector [14].

#### 4. RESULTS

Taking the FFT of two NDVI time series and comparing the Euclidean distance between their first and annual FFT components respectively, produces a scalar quantity in each case. When using the EKF to estimate the  $\mu$  and  $\alpha$  parameter sequence for each NDVI time series, the difference between the parameter sequences fluctuates over the 7 year period. This is illustrated in figure 2 where the  $\mu$  sequence estimated using the EKF is shown along with the FFT mean component for two typical NDVI time series belonging to each of the two classes in region A. The  $\mu$  sequence for the settlement and natural vegetation time series clearly vary in similarity (Figure 2). This is to be expected as land cover classes tend to be more similar during certain parts of the season than others. This characteristic was exploited by only considering the maximum distance between each parameter sequence as given in (6) and (7).

To evaluate the performance of the proposed EKF algorithm, an all-to-all comparison was made between the NDVI time series of each pixel in the natural vegetation class with each pixel in the settlement class. The distance metrics  $D_{\mu}^{EKF}$  and  $D_{\alpha}^{EKF}$  were recorded and the consequent distributions of the values of these distance metrics were calculated. The distribution and values of the FFT distance metrics  $D_{\alpha}^{FFT}$  and  $D_{\mu}^{FFT}$  were also calculated for comparison. Figure 3 shows the distribution of the FFT and EKF  $D_{\alpha}$  values for region A.

Table 1 gives the Bayesian decision error for both the FFT

**Table 1.** Error Probability Comparison of the FFT and EKF method

Region	$P_e^{FFT}(\mu)$	$P_e^{EKF}(\mu)$	$P_e^{FFT}(\alpha)$	$P_e^{EKF}(\alpha)$
A	18%	13%	4%	3%
B	32%	34%	35%	21%

and EKF distance metrics for each region defined as

$$P_e^{FFT}(\alpha) = \int_{D_{\alpha}=0}^{D_{\alpha}^{f*}} p(D_{\alpha}|d_f) + \int_{D_{\alpha}=D_{\alpha}^{f*}}^{\infty} p(D_{\alpha}|s_f), \quad (8)$$

$$P_e^{EKF}(\alpha) = \int_{D_{\alpha}=0}^{D_{\alpha}^{k*}} p(D_{\alpha}|d_k) + \int_{D_{\alpha}=D_{\alpha}^{k*}}^{\infty} p(D_{\alpha}|s_k), \quad (9)$$

$$P_e^{FFT}(\mu) = \int_{D_{\mu}=0}^{D_{\mu}^{f*}} p(D_{\mu}|d_f) + \int_{D_{\mu}=D_{\mu}^{f*}}^{\infty} p(D_{\mu}|s_f), \quad (10)$$

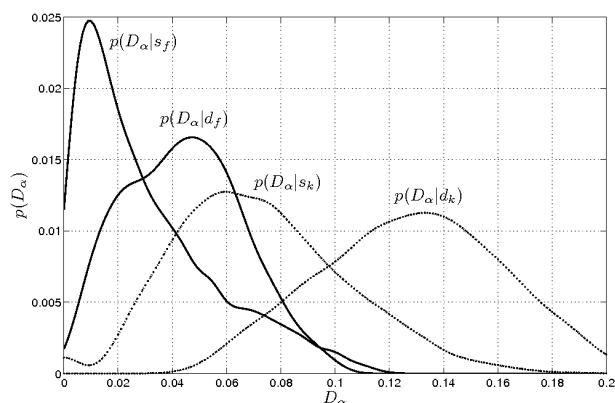
$$P_e^{EKF}(\mu) = \int_{D_{\mu}=0}^{D_{\mu}^{k*}} p(D_{\mu}|d_k) + \int_{D_{\mu}=D_{\mu}^{k*}}^{\infty} p(D_{\mu}|s_k). \quad (11)$$

Where the value of  $D_x^{y*}$ ,  $x \in \{\alpha, \mu\}$ ,  $y \in \{f, k\}$  is the optimal decision threshold minimizing the probability of error  $P_e$  in each instance. It can be seen that the Bayesian decision error in region A using the EKF was reduced by 5% over the FFT method when considering  $D_{\mu}$  and by 1% when considering  $D_{\alpha}$  (Table 1). In region B the Bayes error of the EKF method was increased with 2% over the FFT method when considering  $D_{\mu}$ , but a significant reduction of 14% was achieved when considering  $D_{\alpha}$ . Thus, overall it may be concluded that the EKF formulation has a reduced probability of error when compared to the FFT-based approach on the same data. This implies that the EKF formulation offers improved separability of land cover classes for regions A and B.

The phase parameter  $\phi$  was found to provide negligible additional separability in the classes and thus was excluded from the results. The EKF was found to accurately track the model parameters after 100 samples, this relates to approximately the first two years of the time series. The parameter sequence for the first two years were consequently disregarded.

#### 5. CONCLUSION

Previous research has found that class similarity can be evaluated by considering the difference in FFT components, in particular the mean and annual FFT components, as they tend to carry the majority of signal energy [9, 8]. In this paper, the mean and annual frequency components were estimated for each time step using an EKF. Having iterative estimates of these components allows one to exploit the fact that the mean and annual frequency dissimilarity is more prevalent during certain parts of the seasonal cycle than other parts, an effect



**Fig. 3.** Probability distribution functions of  $D_\alpha$  using the FFT and EKF method in region A where  $p(D_\alpha|s_f)$  is the distribution of the  $D_\alpha^{FFT}$  values for NDVI time series of the same class,  $p(D_\alpha|d_f)$  is the distribution of the  $D_\alpha^{FFT}$  values for NDVI time series of different classes,  $p(D_\alpha|s_k)$  is the distribution of the  $D_\alpha^{EKF}$  values for NDVI time series of the same class and  $p(D_\alpha|d_k)$  is the distribution of the  $D_\alpha^{EKF}$  values for NDVI time series of different classes.

that is merely averaged out using the traditional FFT over the entire NDVI time series. The total probability of decision error was reduced for both of the regions considered in this study. A further possible application of this method is towards land cover change detection. By following the changes of the cosine parameters through time and comparing them with neighboring pixels, a change detection method can be formulated and this is currently being researched. Finally, note that the EKF is well suited to multiple observations so that the EKF can capitalize on all seven bands as opposed to just the NDVI time series data to further improve estimates of the underlying process model. This possibility is also currently being researched.

## 6. REFERENCES

- [1] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, Jan. 2007.
- [2] J.R.G Townshend, T.E Goff, and C.J. Tucker, "Multitemporal dimensionality of images of normalized difference vegetation index at continental scales," *IEEE Geoscience and Remote Sensing*, vol. GE-23, no. 6, pp. 888–895, Nov. 1985.
- [3] K. J. Wessels et al., "Mapping regional land cover with MODIS data for biological conservation: Examples from the greater Yellowstone ecosystem, USA and Para state, Brazil," *Remote Sensing of Environment*, vol. 92, no. 1, pp. 67–83, July 2004.
- [4] M. A. Friedl et al., "Global land cover mapping from MODIS: algorithms and early results," *Remote Sensing of Environment*, vol. 83, no. 1/2, pp. 287–302, Nov. 2002.
- [5] Mryka Hall-Beyer, "Comparison of single-year and multiyear NDVI time series principal components in cold temperate biomes," *IEEE Geoscience and Remote Sensing*, vol. 41, no. 11, pp. 2568–2574, Nov. 2003.
- [6] R. Lasaponara, "Estimating interannual variations in vegetated areas of Sardinia island using SPOT/VEGETATION NDVI temporal series," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 481–483, Oct. 2006.
- [7] D. Alcaraz, J. Paruelo, and J. Cabello, "Identification of current ecosystem functional types in the Iberian peninsula," *Global Ecology and Biogeography*, vol. 15, no. 2, pp. 200–212, Mar. 2006.
- [8] R.I.N Juarez and W.T. Liu, "FFT analysis on NDVI annual cycle and climatic regionality in northeast Brasil," *International Journal of Climatology*, vol. 21, no. 14, pp. 1803–1820, Nov. 2001.
- [9] S. Lhermitte et al., "Hierarchical image segmentation based on similarity of NDVI time-series," *Remote Sensing of Environment*, vol. 112, no. 2, pp. 506–512, Feb. 2008.
- [10] Sanghoon Lee and Melba M. Crawford, "Multitemporal classification of image series with seasonal variability using harmonic components," in *IEEE International Geoscience and Remote Sensing Symposium*, Toulouse, France, July 2003, vol. 5, pp. 3353–3355.
- [11] O. Samain, J. Roujean, and B. Geiger, "Use of a Kalman filter for the retrieval of surface BRDF coefficients with a time-evolving model based on the ECOCLIMAP land cover classification," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1337–1346, Apr. 2008.
- [12] M. Chen, S. Liu, L.L. Tieszen, and D.Y. Hollinger, "An improved state-parameter analysis of ecosystem models using data assimilation," *Ecological Modelling*, vol. 219, no. 3-4, pp. 317–326, Dec. 2008.
- [13] C. Schaaf et al., "First operational BRDF, albedo nadir reflectance product from MODIS," *Remote Sensing of Environment*, vol. 83, no. 1/2, pp. 135–148, Nov. 2002.
- [14] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle Filters for Tracking Applications*, Artech House, London, 1 edition, 2004.