

Meta-optimization of the Extended Kalman Filter's parameters through the use of the Bias Variance Equilibrium Point criterion

B.P. Salmon, W. Kleynhans, F. van den Bergh, J.C. Olivier, W.J. Marais, T.L. Grobler and K.J. Wessels

Abstract—The extraction of information on land cover classes using unsupervised methods has always been of relevance to the remote sensing community. In this paper a novel criterion is proposed which extract the inherent information in an unsupervised fashion from a time series. The criterion is used to fit a parametric model to a time series and derive the corresponding covariance matrices of the parameters for the model and estimate the additive noise on the time series. The proposed criterion uses both spatial and temporal information when estimating the covariance matrices and can be extended to incorporate spectral information. The algorithm used to estimate the parameters for the model is the Extended Kalman filter. An unsupervised search algorithm, specifically designed for this criterion, is proposed in conjunction with the criterion that is used to rapidly and efficiently estimate the variables. The search algorithm attempts to satisfy the criterion by employing density adaptation to the current candidate system. The application in this paper is the use of an Extended Kalman filter to model MODerate-resolution Imaging Spectroradiometer time series with a triply modulated cosine function as the underlying model. The results show that the criterion improved the fit of the triply modulated cosine function by an order of magnitude on the time series over all seven spectral bands when compared to the other methods. The state space variables derived from the Extended Kalman filter are then used for both land cover classification and land cover change detection. The method was evaluated in the Gauteng province of South Africa where it was found to significantly improve on land cover classification and change detection accuracies when compared to other methods.

I. INTRODUCTION

Reliable surveying of land cover and the detection of change in land cover has always been a key interest of the remote sensing community. The increase in human population is one of the major contributions to anthropogenic activities in a geographical area [1]. Several studies have since investigated the effects that anthropogenic activities have on the environment and it is estimated that more than a third of the Earth's land surface has been transformed [2]. The Gauteng province is the

study area of interest, as it is the fastest growing province in South Africa, housing more than 11.3 million people in the year 2011. This equates to 22% of South Africa's population living in the province that has land area of only 1.4% of the total land in South Africa. Proper knowledge of land cover is critical for effective allocation and management of the environmental resources. Digital classification of land cover consist mainly of spatial and spectral analysis. Few methods exploit the temporal resolution offered by coarse resolution satellites, which enables the capture of different land cover dynamics. An example of land cover change is illustrated in figure 1. The red box in figure 1 illustrates a land cover conversion from natural vegetation to newly formed human settlements. The blue box in figure 1 illustrates a seasonal variation of natural vegetation. The limitation of using two images is that similar land cover types (blue box in figure 1) can look different at various times of the year [3]. This will require the incorporation of local information to adjust the change detection algorithm's settings to reduce the number of false alarms caused by seasonal variation.

The remote sensing community's monitoring capabilities keep improving with the development and deployment of new technologies. Global data sets are becoming more accessible and computational resources are becoming more affordable [4]. These data sets come from several different sensors. The more popular satellite based sensors are: Landsat Multi-Spectral Scanner (MSS), Multi-angle Imaging SpectroRadiometer (MISR), Système Pour l'Observation de la Terre (SPOT), Advanced Very High Resolution Radiometer (AVHRR) and MODerate-resolution Imaging Spectroradiometer (MODIS). The type of land cover change that can be detected also changes with newer technologies, which requires the continual pursuit of new change detection methods [5], [6].

Lhermitte *et al.* proposed a method that separates different land cover classes using a Fourier analysis of NDVI time series [7]. It was concluded that good separation is achievable when evaluating the magnitude of the coefficients of the Fourier transform associated with the NDVI signal's mean and amplitude components. Kleynhans *et al.* proposed a method which jointly estimates the mean and seasonal components of the Fourier transform using a triply modulated cosine function [8]. Kleynhans *et al.* used an Extended Kalman filter (EKF) to estimate the parameters for the triply modulated cosine function to model NDVI time series by updating the mean (μ), amplitude (α), and phase (θ) parameters for each time increment.

B.P. Salmon and W. Kleynhans are with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

Email: brian.salmon@gmail.com. Tel: +27 12 841 3207. Fax: +27 12 841 3124

J.C. Olivier is with the School of Engineering, University of Tasmania, Australia

T.L. Grobler is with the Defense, Peace, Safety and Security Unit, Meraka Institute, CSIR, Pretoria, South Africa

W.J. Marais is with the Space Science and Engineering Center, University of Wisconsin-Madison, Wisconsin, United States

K.J. Wessels and F. van den Bergh are with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria, South Africa



Fig. 1. Land cover change detected in Protea Glen ($26^{\circ}15'52.38''S$, $27^{\circ}47'07.42''E$). The Quickbird image at the top was taken on 31 July 2001 and the bottom on 7 September 2008 (Courtesy of GoogleTM Earth). A change in land cover type is shown in the red box, while only a seasonal change is shown in the blue box.

The Extended Kalman filter (EKF) can be used as a feature extraction method based on the assumption that the parameters of the underlying model can be used to separate a set of time series into different classes. Consideration must be given when selecting the model, as it should reflect the seasonal behaviour of a specific land cover class. It follows that more separable parameters derived by the EKF makes it easier to detect changes in the assigned classes [8].

The objective of this paper is to propose a novel criterion that can be used to set the initial parameters of an EKF. An operator typically uses a training set to supervise the adjustment of the initial parameters within the EKF until acceptable performance is obtained for a set of time series. The criterion proposes an appropriately defined parameter space which uses a spatio-temporal window to enable the search for

improved parameters to use within the EKF in an unsupervised manner. An unsupervised search algorithm is provided which is used to search through this defined parameter space. The method of setting the initial parameters is tested with the feature extraction method proposed by Kleynhans *et al.* [8]. The spectral bands in the paper are modelled separately as triply modulated cosine functions, which is an extension of the method proposed in [8]. The performance of this new method is compared to a non-linear least squares algorithm and an EKF, which is set using the Autocovariance Least Squares (ALS) method [9].

The paper is organized as follows. A description of the data set is given in section II. Section III-A discusses the principle behind using an EKF to model a time series using a triply modulated cosine function. The importance of the initial

parameters used to set the Extended Kalman filter is discussed in section III-B, illustrating how the behaviour is dependent on these initial parameters. A novel criterion is proposed, called the Bias-Variance Equilibrium Point (BVEP) criterion in section III-D, which defines a desired set of initial parameters that will provide optimal performance. The BVEP criterion uses both the temporal and spatial information to design a system with desirable behaviour and it is concluded on how it can easily be extended to incorporate spectral information. A specifically designed search algorithm called the Bias-Variance Search Algorithm (BVSA) is proposed in section III-E, that will adjust the Bias-Variance Score (BVS) to best satisfy the BVEP criterion that will provide good initial parameters for the Extended Kalman filter. Section IV presents the results of optimizing the EKF, along with the land cover classification and change detection experiments in the Gauteng province. Section VI presents the conclusion on how this criterion could be used for land cover information extraction.

II. DATA DESCRIPTION

A. Study Area

An increase in anthropogenic activities is usually directly correlated to the increase in human population in a geographical area. The Gauteng province is located in the Highveld of South Africa and is currently the fastest growing province in South Africa, which is evident in the growth in population provided by the population estimates shown in Table I. These estimates are provided by Statistics South Africa in an annual report for the Gauteng province. The province is also the most urbanized province in the country and contributes more than one third of South Africa's economy. Large areas of natural vegetation still exist even with the mass expansion within the province. The method was applied to a validated study area that corresponds to a total area of approximately 285.5 km². The study area's land cover is predominantly natural vegetation and human settlements. The time series in the validated study area were verified using visual interpretation of SPOT images to map areas of no change in land cover type during the study period for the temporal component of the analysis. The proposed method was then tested on the entire Gauteng province (19676km²) to measure the growth of human settlements.

B. MODIS Time Series Data

The MODIS (MCD43A4, Collection V005) 500-meter, Nadir and Bidirectional Reflectance Distribution Function (BRDF) adjusted spectral reflectance bands were used, as it significantly reduces the anisotropic scattering effects of surfaces under different illumination and observation conditions [10], [11]. The data set provides a sample on a rolling 8 day interval based on a 16-day of MODIS surface reflectance composite period, for each of the seven spectral bands at 500 meter resolution. For each pixel in the study area, a time series was extracted for all 7 bands from the data set (tile H20V11) for the time period February 2000 to January 2011. The specifications for the 7 spectral bands are provided in table II.

TABLE I
POPULATION ESTIMATES PROVIDED BY STATISTICS SOUTH AFRICA IN AN ANNUAL REPORT FOR THE GAUTENG PROVINCE FOR THE YEAR 2000–2011

Year	Estimated Population
2000	8,571,705
2001	8,865,664
2002	9,181,751
2003	9,518,114
2004	9,643,428
2005	9,799,634
2006	10,046,871
2007	10,192,199
2008	10,467,705
2009	10,554,587
2010	11,198,051
2011	11,326,375

TABLE II
MODIS SPECTRAL BANDS' SPECIFICATIONS USED IN THE MCD43A4 (COLLECTION V005) PRODUCT.

Spectral bands	Wavelengths (nanometers)	Spectral range
Band 1	620–670	Visible (Red)
Band 2	841–876	Near infrared
Band 3	459–479	Visible (Blue)
Band 4	545–565	Visible (Green)
Band 5	1230–1250	Short infrared
Band 6	1628–1652	Short infrared
Band 7	2105–2155	Short infrared

III. METHODOLOGY

A. Extended Kalman Filter

The EKF is a non-linear estimation method that produces an estimated observation from an actual noisy observation. The estimated observations are computed using a defined parametric model. The EKF estimates the state-space parameters for the model using the noisy observations. The EKF has been used in the remote sensing community for parameter estimation of values related to physical, biogeochemical processes or vegetation dynamics models [12], [13]. The observations used in this study were obtained from the spectral bands of the MODIS sensor.

In figure 2(b), a Fourier transform is used to observe that the majority of the signal energy is contained in the mean and seasonal component of the second spectral band's time series which is the infrared spectrum (figure 2(a)). This implies that the time series is well represented in the time domain as a single cosine function with a mean offset, amplitude and phase, as shown in figure 2(c).

The EKF has also been used as a feature extraction method to model a NDVI time series for a given pixel as a triply modulated cosine function to improve land cover separation [8]. This paper proposes an extension to this model presented in [8] by modelling each spectral band separately as a triply modulated cosine function. This is expressed as

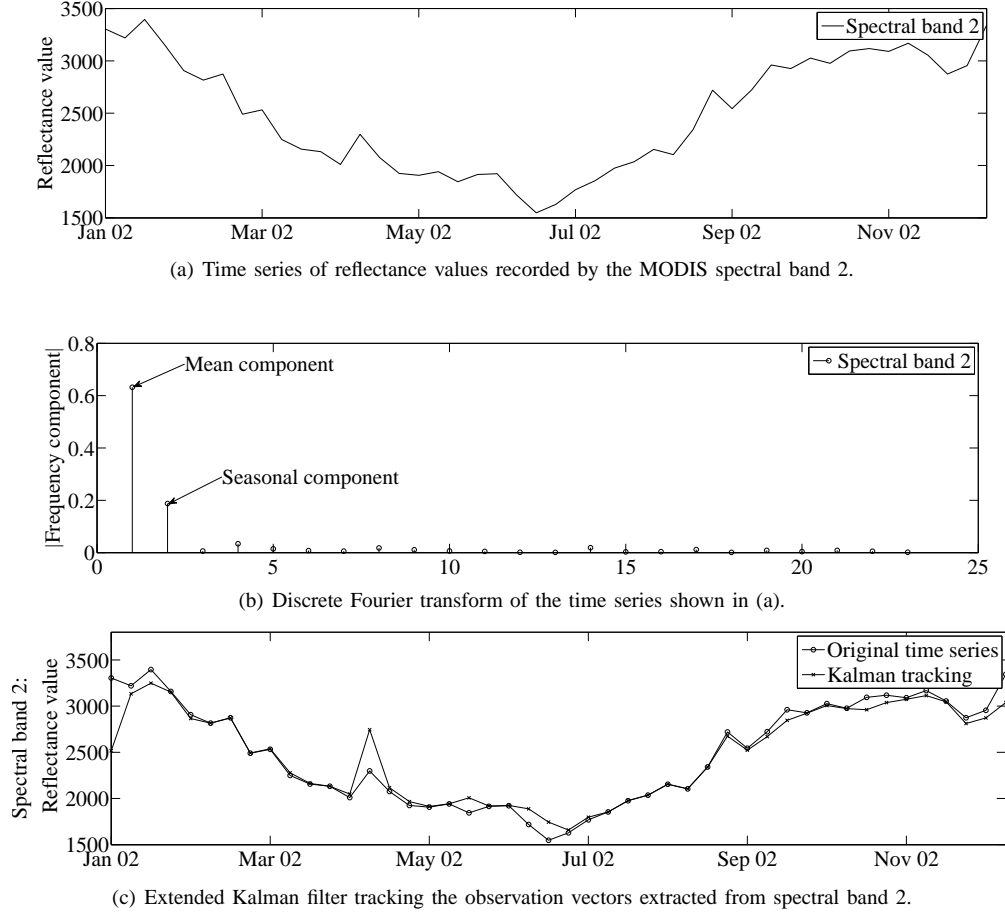


Fig. 2. The time series recorded by the second spectral band for a geographical area is shown in (a) with the corresponding magnitude of the discrete Fourier transform shown in (b). A triply modulated cosine function is fitted to the time series using an EKF in (c).

$$y_{i,k,b} = \mu_{i,k,b} + \alpha_{i,k,b} \cos(\omega k + \phi_{i,k,b}) + v_{i,k,b}, \quad (1)$$

where $y_{i,k,b}$ denotes the observed value of the b -th spectral band's time series, $b \in [1, 7]$, of the k -th pixel at time index i . The noise sample of the k -th pixel at time i for the b -th spectral band is denoted by $v_{i,k,b}$. The noise is additive with an unknown distribution. The cosine function is fitted to each spectral band and is based on several parameters; the frequency ω , which is the same over all the spectral bands, the nonzero mean $\mu_{i,k,b}$, the amplitude $\alpha_{i,k,b}$ and the phase $\phi_{i,k,b}$. The frequency is explicitly calculated as $\omega = 2\pi f$, where f is based on the annual vegetation growth cycle and the sampling rate of the MODIS sensor. Given the 8 daily composite period of the MCD43A4 MODIS data product, f is set to $8/365$. The values of $\mu_{i,k,b}$, $\alpha_{i,k,b}$ and the phase $\phi_{i,k,b}$ are functions of time and must be estimated for each pixel, $k, k \in [1, K_{\max}]$, given the observations $y_{i,k,b}$ for time indices $i = \{1, \dots, I_{\max}\}$ and spectral bands $b = \{1, \dots, 7\}$. The maximum number of pixels in the data set is denoted by K_{\max} , and the maximum number of observations in a time series is defined by I_{\max} . A state-space vector is estimated by the EKF at each time index for each spectral band and contains S parameters, which in this case is $S=3$ and is defined as

$$\vec{X}_{i,k,b} = [\mu_{i,k,b} \quad \alpha_{i,k,b} \quad \phi_{i,k,b}]^T. \quad (2)$$

The relation between $\vec{X}_{i,k,b}$ and $\vec{X}_{(i-1),k,b}$ for the b -th spectral band is denoted by the transition function \mathbf{f}_b . The state-space vector $\vec{X}_{i,k,b}$ is related to the observation $y_{i,k,b}$ via a non-linear measurement function \mathbf{h}_b . These relationships are expressed as

$$\vec{X}_{i,k,b} = \mathbf{f}_b(\vec{X}_{(i-1),k,b}) + \mathbf{z}_{(i-1),k,b}, \quad (3)$$

and

$$\hat{y}_{i,k,b} = \mathbf{h}_b(\vec{X}_{i,k,b}) + v_{i,k,b}. \quad (4)$$

The estimated observation is denoted by $\hat{y}_{i,k,b}$ in equation (4). The process noise is denoted by $\mathbf{z}_{(i-1),k,b}$ and the measurement noise by $v_{i,k,b}$.

For the present case of land cover classification, it is assumed that the state-space vector $\vec{X}_{i,k,b}$ does not change significantly through time; hence, the transition function is assumed linear. The measurement function, however, contains the cosine function and, as such, is evaluated via the standard Jacobian formulation, through linear approximation of the non-linear measurement function around the current state-space vector. Both these functions are possibly non-perfect, so the addition of process $\mathbf{z}_{(i-1),k,b}$ and measurement $v_{i,k,b}$ noise is required [14]. The state-space vector's update is based on

the newest observation value $y_{i,k,b}$ [14]. The estimated values of $\vec{X}_{i,k,b}$ over time i effectively results in a time series of the state-space vectors for each of the K_{\max} pixels for each spectral band.

Converting state-space vectors to land cover classes: A machine learning algorithm is used to process the estimated state-space vectors to corresponding class labels. A class label is assigned to each state-space vector for each pixel at each time index as

$$\mathcal{C}_{i,k} = \mathcal{F}_C \left(\left\{ \left\{ X_{i,k,b,s} \right\}_{s=1}^{s=S} \right\}_{b=1}^{b=7} \right) = \mathcal{F}_C \left(\left\{ \vec{X}_{i,k,b} \right\}_{b=1}^{b=7} \right). \quad (5)$$

The function \mathcal{F}_C denotes either a supervised or unsupervised classifier. The class label for the k^{th} pixel at time index i is denoted by $\mathcal{C}_{i,k}$. Land cover change is declared when a pixel k has a persistent change in class label as a function of time i . This is expressed as

$$\mathcal{C}_{i,k} \neq \mathcal{C}_{j,k}, \quad 0 < i \leq j, \forall i, j \quad (6)$$

The MODIS spectral bands are assumed to be uncorrelated and are treated independently in this method. The index b is omitted for convenience with no loss in generality in the description of the method. The importance of the initial parameters will be discussed in the next section.

B. Importance of the initial parameters

It is well known from estimation theory that many prediction results simplify when Gaussian distributions are used. The process noise and observation noise are thus assumed to be Gaussian distributed. The process noise is thus denoted by $z_{(i-1),k}, \tilde{z}_{(i-1),k} \sim \mathcal{N}(0, \mathcal{Q}_{(i-1),k})$, with $\mathcal{Q}_{(i-1),k}$ representing the process noise covariance matrix. The observation noise is denoted by $v_{i,k}, \tilde{v}_{i,k} \sim \mathcal{N}(0, \mathcal{R}_{i,k})$, with $\mathcal{R}_{i,k}$ representing the observation noise covariance matrix. The EKF recursively adapts the state-space vector for each incoming observation by predicting and updating the vector. In the prediction step the state-space vector $\vec{X}_{(i|i-1),k}$ and covariance matrix $\mathfrak{B}_{(i|i-1),k}$ is computed. The subscript $(i|i-1)$ denotes the evaluation at time index i given all the previous indices up to and including $(i-1)$. The predicted state-space vector's estimate $\vec{X}_{(i|i-1),k}$ is computed as

$$\vec{X}_{(i|i-1),k} = \mathbf{f} \left(\vec{X}_{(i-1|i-1),k} \right), \quad (7)$$

and the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ is computed as

$$\mathfrak{B}_{(i|i-1),k} = \mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-1|i-1),k} \mathbf{F}_{\text{est}}^T. \quad (8)$$

The matrix \mathbf{F}_{est} is the local linearization of the non-linear transition function \mathbf{f} . In the updating step, the posterior estimate of the state-space vector $\vec{X}_{(i|i),k}$ is computed as

$$\vec{X}_{(i|i),k} = \vec{X}_{(i|i-1),k} + \mathfrak{K}_{i,k} \left(y_{i,k} - \mathbf{h} \left(\vec{X}_{i,k} \right) \right), \quad (9)$$

using the optimal Kalman gain denoted by $\mathfrak{K}_{i,k}$ which is computed as

$$\mathfrak{K}_{i,k} = \mathfrak{B}_{(i|i-1),k} \mathbf{H}_{\text{est}}^T \mathcal{S}_{i,k}^{-1}. \quad (10)$$

The matrix \mathbf{H}_{est} is the local linearization of the non-linear measurement function \mathbf{h} . The matrix $\mathcal{S}_{i,k}$ denotes the innovation term which is computed as

$$\mathcal{S}_{i,k} = \mathbf{H}_{\text{est}} \mathfrak{B}_{(i|i-1),k} \mathbf{H}_{\text{est}}^T + \mathcal{R}_{i,k}. \quad (11)$$

The posterior estimate of the covariance matrix $\mathfrak{B}_{(i|i),k}$ is computed as

$$\mathfrak{B}_{(i|i),k} = \mathfrak{B}_{(i|i-1),k} - \mathfrak{K}_{i,k} \mathcal{S}_{i,k} \mathfrak{K}_{i,k}^T. \quad (12)$$

The tracking performance of the EKF is assessed by evaluating the stability of the state-space vector and the error in estimating the observation. The error in estimating the observation is computed as the absolute error between estimated observation $\hat{y}_{i,k}$ and the actual observation $y_{i,k}$. This is expressed as

$$\mathcal{E}_{y,i,k} = |y_{i,k} - \hat{y}_{i,k}| = \left| y_{i,k} - \mathbf{h} \left(\vec{X}_{(i|i),k} \right) \right|. \quad (13)$$

In equation (13), it is observed that the state-space vector $\vec{X}_{(i|i),k}$ determines the observation error $\mathcal{E}_{y,i,k}$. The state-space vector $\vec{X}_{(i|i),k}$ can thus be selected to minimise the observation error. The observation error can be easily minimised by significantly varying $\vec{X}_{(i|i),k}$ to accommodate the fluctuation in observations. This does not bode well if the underlying structure of the system is being analysed. A significantly varying state-space vector $\vec{X}_{(i|i),k}$ is indicative of an unstable model. The conclusion is that the state-space model must be kept stable, while also attempting to minimise the observation error in equation (13).

The initial estimates provided to the EKF will now be discussed to illustrate their importance. A stable state-space vector requires small adaptation from $\vec{X}_{(i-1|i-1),k}$ to $\vec{X}_{(i|i),k}$. The initial estimated state-space vector $\vec{X}_{(0|0),k}, \tilde{\vec{X}}_{(0|0),k} \in \mathcal{X}$, for the first observation $y_{0,k}$ is optimised using a local search method or domain knowledge which satisfies

$$\tilde{\vec{X}}_{(0|0),k} = \underset{\vec{X} \in \mathcal{X}}{\text{argmin}} \left\{ \left| y_{0,k} - \mathbf{h} \left(\vec{X} \right) \right| \right\}. \quad (14)$$

The recursive adaptation of the state-space vector's estimate $\vec{X}_{(i|i),k}$ is then calculated using the predicted step given in equation (7) and the updating step in equation (9). Equation (7) is substituted into equation (9) which yields

$$\vec{X}_{(i|i),k} = \mathbf{f} \left(\vec{X}_{(i-1|i-1),k} \right) + \mathfrak{K}_{i,k} \left(y_{i,k} - \mathbf{h} \left(\mathbf{f} \left(\vec{X}_{(i-1|i-1),k} \right) \right) \right). \quad (15)$$

The Kalman gain $\mathfrak{K}_{i,k}$ sets the magnitude of change in the estimated state-space vector for each time increment. If the observation error is large and the Kalman gain is large, then large changes will be made to the current state-space vector. If the observation error is large and the Kalman gain is small,

then the state-space's estimate $\vec{X}_{(i|i),k}$ will adapt slowly which typically leads to large observation error $\mathcal{E}_{y,i,k}$ (equation (13)) until it eventually converges. If the observation error is small and the Kalman gain is large, then the state-space vector will struggle to converge as it will continually overshoot the desired state-space vector that will minimise equation (13). Substituting the optimal Kalman gain given in equation (10) into equation (15) expands it to

$$\vec{X}_{(i|i),k} = \mathbf{f}(\vec{X}_{(i-1|i-1),k}) + \mathfrak{B}_{(i|i-1),k} \mathbf{H}_{\text{est}}^T \mathcal{S}_{i,k}^{-1} \left(y_{i,k} - \mathbf{h}(\mathbf{f}(\vec{X}_{(i-1|i-1),k})) \right). \quad (16)$$

The Kalman gain is dependent on the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ and innovation term $\mathcal{S}_{i,k}$. The innovation term controls the trust region within the state-space vector's space. This is dependent on the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ and observation noise covariance matrix $\mathcal{R}_{i,k}$. Substituting the innovation term given in equation (11) into equation (16) results in

$$\vec{X}_{(i|i),k} = \mathbf{f}(\vec{X}_{(i-1|i-1),k}) + \mathfrak{B}_{(i|i-1),k} \mathbf{H}_{\text{est}}^T (\mathbf{H}_{\text{est}} \mathfrak{B}_{(i|i-1),k} \mathbf{H}_{\text{est}}^T + \mathcal{R}_{i,k})^{-1} \left(y_{i,k} - \mathbf{h}(\mathbf{f}(\vec{X}_{(i-1|i-1),k})) \right). \quad (17)$$

The last term to evaluate is the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$. The predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ is substituted to yield an updated state-space vector as

$$\vec{X}_{(i|i),k} = \mathbf{f}(\vec{X}_{(i-1|i-1),k}) + (\mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-1|i-1),k} \mathbf{F}_{\text{est}}^T) \mathbf{H}_{\text{est}}^T (\mathbf{H}_{\text{est}} (\mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-1|i-1),k} \mathbf{F}_{\text{est}}^T) \mathbf{H}_{\text{est}}^T + \mathcal{R}_{i,k})^{-1} \left(y_{i,k} - \mathbf{h}(\mathbf{f}(\vec{X}_{(i-1|i-1),k})) \right). \quad (18)$$

The transition function \mathbf{f} and measurement function \mathbf{h} are assumed to be known. The observation $y_{i,k}$ is supplied by the real system. The only variables left within equation (18) are the: (1) previous state-space vector's estimate $\vec{X}_{(i-1|i-1),k}$, (2) process noise's covariance matrix $\mathcal{Q}_{(i-1),k}$, (3) previous estimate of covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$, and (4) observation noise's covariance matrix $\mathcal{R}_{i,k}$.

The previous estimation of the covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$ will be briefly explored as it is part of equation (18). The covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$ is updated with

$$\mathfrak{B}_{(i-1|i-1),k} = \mathfrak{B}_{(i-1|i-2),k} - \mathfrak{K}_{(i-1),k} \mathcal{S}_{(i-1),k} \mathfrak{K}_{(i-1),k}^T. \quad (19)$$

The Kalman gain given in equation (10), the innovation term given in equation (11) and the predicted covariance matrix $\mathfrak{B}_{(i-1|i-2),k}$ given in equation (8) are substituted into equation (19), to yield

$$\begin{aligned} \mathfrak{B}_{(i-1|i-1),k} &= (\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) - \\ & ((\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) \mathbf{H}_{\text{est}}^T \\ & (\mathbf{H}_{\text{est}} (\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) \\ & \mathbf{H}_{\text{est}}^T + \mathcal{R}_{(i-1),k})^{-1}) (\mathbf{H}_{\text{est}} (\mathcal{Q}_{(i-2),k} + \\ & \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) \mathbf{H}_{\text{est}}^T + \mathcal{R}_{(i-1),k}) \\ & ((\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) \mathbf{H}_{\text{est}}^T \\ & (\mathbf{H}_{\text{est}} (\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}} \mathfrak{B}_{(i-2|i-2),k} \mathbf{F}_{\text{est}}^T) \\ & \mathbf{H}_{\text{est}}^T + \mathcal{R}_{(i-1),k})^{-1})^T. \end{aligned} \quad (20)$$

Equation (18) is computed for every newly obtained observation. The state-space vector's estimate $\vec{X}_{(i|i),k}$ requires the results from equation (20) to compute the current estimates. The transition function \mathbf{F}_{est} and measurement function \mathbf{H}_{est} are known, then the only variables left to compute in equation (20) are: (1) initial covariance matrix $\mathfrak{B}_{(0|0),k}$, (2) process noise covariance matrix $\mathcal{Q}_{(i-1),k}$, and (3) observation noise's covariance matrix $\mathcal{R}_{i,k}$. The conclusion from equation (18) and equation (20) is the initial parameters which are of importance are:

- 1) Initial state-space vector's estimate $\vec{X}_{(0|0),k}$,
- 2) Initial covariance matrix estimate $\mathfrak{B}_{(0|0),k}$,
- 3) Process noise covariance matrix $\mathcal{Q}_{(i-1),k}$,
- 4) Observation noise covariance matrix $\mathcal{R}_{i,k}$.

The initial state-space vector's estimate $\vec{X}_{(0|0),k}$ is initialised using equation (14). Even if an incorrect estimate is used, the state-space vector $\vec{X}_{(i|i),k}$ should converge to the correct vector as $i \rightarrow \infty$. The same is true about the initial covariance matrix $\mathfrak{B}_{(0|0),k}$. As $i \rightarrow \infty$, the covariance matrix $\mathfrak{B}_{(i|i),k}$ should tend to converge to the correct matrix. Usual operation of the EKF sets the initial covariance matrix $\mathfrak{B}_{(0|0),k}$ equal to the identity matrix.

The initial covariance matrix $\mathfrak{B}_{(0|0),k}$ will stabilise as equation (8) is a discrete Riccati equation, and under certain circumstances will converge which results in equation (20) converging to a stable state [15], [16]. The conditions for convergences of the discrete Riccati equation are:

- 1) the process noise covariance matrix $\mathcal{Q}_{(i-1),k}$ is a positive definite matrix,
- 2) the observation noise covariance matrix $\mathcal{R}_{i,k}$ is a positive definite matrix,
- 3) the pair $(\mathbf{F}_{\text{est}}, z_{(i-1),k})$ is controllable,
- 4) the pair $(\mathbf{F}_{\text{est}}, \mathbf{H}_{\text{est}})$ is observable.

Under the conditions set above, the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ converges to a stable matrix and is expressed as

$$\lim_{i \rightarrow \infty} \mathfrak{B}_{(i|i-1),k} = \mathfrak{B}_{\text{stable}}, \quad (21)$$

where $\mathfrak{B}_{\text{stable}}$ is a symmetric positive definite matrix. $\mathfrak{B}_{\text{stable}}$ is a unique positive definite solution of the discrete Riccati equation and $\mathfrak{B}_{\text{stable}}$ is independent of the initial distribution of initial state-space vector's estimate $\vec{X}_{(0|0),k}$.

The values of $\vec{X}_{(0|0),k}$ and $\mathfrak{B}_{(0|0),k}$ can also be estimated

using an offline training phase. Offline refers to observations which are stored and are used recursively for estimation. The process noise covariance matrix $\mathcal{Q}_{(i-1),k}$ and observation noise covariance matrix $\mathcal{R}_{i,k}$ are assumed to be constant throughout the recursive estimation of the observation. This is usually manually set by a system analyst in an offline training phase through successive adjustments. The initial setting of the EKF is thus defined as setting the following:

- 1) Initial state-space vector $\vec{X}_{(0|0),k}$ is estimated offline,
- 2) Initial covariance matrix $\mathfrak{B}_{(0|0),k}$ is estimated offline,
- 3) Process noise covariance matrix $\mathcal{Q}_{(i-1),k}$ is set to a fixed matrix,
- 4) Observation noise covariance matrix $\mathcal{R}_{i,k}$ is set to a fixed matrix.

The EKF will track the observations with minimum residual and have a stable internal state-space vector if all initial parameters are properly set.

C. Estimation of the noise covariance matrices

From the previous section it was shown that the initial parameters that need to be set are:

- 1) Initial state-space vector $\vec{X}_{(0|0),k}$,
- 2) Initial covariance matrix $\mathfrak{B}_{(0|0),k}$,
- 3) Process noise covariance matrix $\mathcal{Q}_{(i-1),k}$,
- 4) Observation noise covariance matrix $\mathcal{R}_{i,k}$.

The setting of the initial parameters requires reliable information about the system. The practical implementation of the EKF is usually done in the absence of this information and the initial parameters are set in an ad-hoc method by an operator to obtain reasonable performance.

Several different approaches have been formulated to solve the estimation of the initial parameters [17], [18], [19]. The Autocovariance Least Squares (ALS) method was presented by Odelson *et al* [20], where the correlation within the innovation data was explored to form a least squares problem to determine the noise covariances for the disturbances. A motivation for using this method is that it avoids a complicated non-linear estimation approach used by methods that employ a maximum likelihood estimation approach [21]. The drawback was that all these methods assumed that the noise-shaping matrix in the transition equation is known, where this information is usually not available. In the absence of information about the noise-shaping matrix the linear dynamic model is modelled as a Gaussian noise vector. Rajamani and Rawlings [9] described a method which estimates the structure of the noise-shaping matrix in an iterative approach. The ALS method assumes that:

- 1) both the measurement function \mathbf{h} and transition function \mathbf{f} are known,
- 2) enough observation vectors are available to ensure the internal covariance matrix $\mathfrak{B}_{(i|i),k}$ becomes stable, and
- 3) the residuals at different time increments are uncorrelated.

The ALS has the ability to provide unique solution to the initial parameters if the measurement function \mathbf{h} is full rank, which implies that the dimension of observation vector is equal to the dimension of states [9]. In the case of this study,

the dimension of the observation vector is smaller than the states, which leads to the condition where the ALS method will produce multiple solutions to the estimation of the initial parameters. The best performing covariance matrices produced by the ALS method were used and were obtained by applying the ALS method to various different initial estimates of the covariance matrices.

D. Bias-Variance Equilibrium Point Criterion

The general approach to estimating and initialising the state-space vectors, as well as the observation- and process noise's covariance matrices for the EKF, is usually for an analyst to determine these offline using a training data set. Proper estimation of the initial parameters through various methods leads to good filter behaviour from the EKF, while improper estimation could cause system instability which leads to delayed tracking or abnormal system behaviour.

A novel BVEP criterion is proposed that will use temporal and spatial information to design a parameter space where desirable system behaviour is expected. This is accomplished by first observing the dependencies between the initial parameters. The proposed criterion is used by an unsupervised BVSA to iteratively adjust a BVS to determine proper initial parameters for the EKF. The characteristics of the initial parameters are first explored before describing the criterion. The first parameter to investigate is the observation noise covariance matrix $\mathcal{R}_{i,k}$. The observation noise covariance matrix $\mathcal{R}_{i,k}$ is defined here as only observing the diagonals of the matrix and that only a single spectral band is evaluated as

$$\mathcal{R}_{i,k} = E[(y_{i,k} - E[y_{i,k}])^2]. \quad (22)$$

This expression holds as the spectral bands are assumed to be uncorrelated and that the MODIS sensor only produces a single reflectance value per pixel per spectral band. The second parameter is the process noise covariance matrix $\mathcal{Q}_{i,k}$ and is defined as

$$\mathcal{Q}_{i,k} = \text{diag}\{E[(X_{i,k,s} - E[X_{i,k,s}])^2]\}, \quad \forall s. \quad (23)$$

This expression holds as the state-space variables within the state-space vector are assumed to be uncorrelated. The setting of these initial parameters have a major effect on the overall system performance. The initial state-space vector $\vec{X}_{(0|0),k}$ for the first observation $y_{0,k}$ is optimised using equation (14). This can either be done using a local search algorithm or applying direct domain knowledge. The initial state-space vector in this work was set using the Fourier transform's components as proposed by Kleynhans *et al.* [8].

The initial estimated covariance matrix $\mathfrak{B}_{(0|0),k}$ is usually set to the identity matrix as it is recursively estimated by the EKF. This only leaves the estimation of the observation noise covariance scalar $\mathcal{R}_{i,k}$ and process noise covariance vector $\mathcal{Q}_{i,k}$. Let the uncorrelated observation noise covariance matrix's diagonals be placed into a vector called the observation candidate vector $\Upsilon_{\mathcal{R},i,k}$, where $\Upsilon_{\mathcal{R},i,k}$ is selected from the space $\nu_{\mathcal{R}}$, and it is expressed as

$$\Upsilon_{\mathcal{R},i,k} = 10^{\zeta_{i,k}/10}, \quad (24)$$

with

$$\zeta_{i,k} = 10 \log_{10} (E[(y_{i,k} - E[y_{i,k}])^2]). \quad (25)$$

Let the uncorrelated process noise covariance matrix's diagonals be placed into a vector called the process candidate vector $\Upsilon_{\mathcal{Q},i,k}$, where $\Upsilon_{\mathcal{Q},i,k}$ is selected from space $\nu_{\mathcal{Q}}$, which is expressed as

$$\Upsilon_{\mathcal{Q},i,k} = 10^{[\zeta_{i,k,1} \dots \zeta_{i,k,S}]/10} = 10^{\bar{\zeta}_{i,k}/10}, \quad (26)$$

with

$$\zeta_{i,k,s} = 10 \log_{10} (E[(X_{i,k,s} - E[X_{i,k,s}])^2]). \quad (27)$$

It should be noted that the EKF only recursively updates the state-space vector $\vec{X}_{(i|i),k}$, and the covariance matrix $\mathfrak{B}_{(i|i),k}$. The time index of the observation noise covariance matrix $\mathcal{Q}_{i,k}$ has been left inserted to emphasise the time effect in a dynamic linear system. The EKF, however, does not alter the observation noise covariance matrix at each time index and is thus assumed constant for all time indices. This is formally stated as $\mathcal{Q} = \mathcal{Q}_i, \forall i$. The process noise covariance matrix is also retained as a constant for all time indices and is stated as $\mathcal{R} = \mathcal{R}_i, \forall i$. This concludes that the observation noise covariance matrix and process noise covariance matrix are time independent. This property allows the observation candidate vector to be rewritten as

$$\Upsilon_{\mathcal{R},k} = 10^{\zeta_k/10} \quad \forall k, \quad (28)$$

and the process candidate vector rewritten as

$$\Upsilon_{\mathcal{Q},k} = 10^{[\zeta_{k,1} \dots \zeta_{k,S}]/10} = 10^{\bar{\zeta}_k/10} \quad \forall k. \quad (29)$$

It was stated earlier that the performance of the Kalman filter is benchmarked by the residual error in tracking the observations and the internal stability of the state-space vector. A parameter space is defined here to describe the system behaviour.

The first desired behaviour is the tracking of the observation with minimal residual. This desired behaviour is expressed as the minimal achievable sum of absolute residuals $\sigma_{\mathcal{E}}$ which is computed as

$$\sigma_{\mathcal{E}} = \min_{\substack{\Upsilon_{\mathcal{R},k} \in \nu_{\mathcal{R}}, \\ \Upsilon_{\mathcal{Q},k} \in \nu_{\mathcal{Q}}}} \left\{ \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \|\hat{y}_{i,k} - y_{i,k}\| \right\}, \quad (30)$$

then

$$[\mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}_{\sigma_{\mathcal{E}}}] = \operatorname{argmin}_{\substack{\Upsilon_{\mathcal{R},k} \in \nu_{\mathcal{R}}, \\ \Upsilon_{\mathcal{Q},k} \in \nu_{\mathcal{Q}}}} \left\{ \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \|\hat{y}_{i,k} - y_{i,k}\| \right\}. \quad (31)$$

The set $[\mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}_{\sigma_{\mathcal{E}}}]$ represents the parameters required to achieve the minimal residual. The minimal residual is then computed as

$$\sigma_{\mathcal{E}} = \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \left\| \hat{y}_{i,k} - y_{i,k} \right\|_{\mathcal{R}=\mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}=\mathcal{Q}_{\sigma_{\mathcal{E}}}}. \quad (32)$$

The second criterion is based on the internal stability of the state-space vector. This can be measured as the variations in each of the state-space variables. The second desired behaviour is expressed as the minimal achievable absolute deviation in state-space variables, which is computed as

$$\sigma_s = \min_{\substack{\Upsilon_{\mathcal{R},k} \in \nu_{\mathcal{R}}, \\ \Upsilon_{\mathcal{Q},k} \in \nu_{\mathcal{Q}}}} \left\{ \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \|X_{i,k,s} - E[X_{i,k,s}]\| \right\}, \quad \forall s, \quad (33)$$

then

$$[\mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}] = \operatorname{argmin}_{\substack{\Upsilon_{\mathcal{R},k} \in \nu_{\mathcal{R}}, \\ \Upsilon_{\mathcal{Q},k} \in \nu_{\mathcal{Q}}}} \left\{ \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \|X_{i,k,s} - E[X_{i,k,s}]\| \right\}, \quad \forall s. \quad (34)$$

The set $[\mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}]$ represents the parameters required to achieve the minimal absolute deviation in the state-space variable s . The minimal absolute deviation for state-space variable s is computed as

$$\sigma_s = \sum_{k=1}^{K_{\max}} \sum_{i=1}^{\mathcal{I}_{\max}} \left\| X_{i,k,s} - E[X_{i,k,s}] \right\|_{\mathcal{R}=\mathcal{R}_{\sigma_s}, \mathcal{Q}=\mathcal{Q}_{\sigma_s}}. \quad (35)$$

The spatial information is included through the use of a set of time series all located in a specific geographical area. The set of K_{\max} time series for a geographical area is denoted by $\{\vec{y}_k\} = \{\{y_{i,k}\}_{i=1}^{\mathcal{I}_{\max}}\}$. Let $q_{i,\mathcal{E}}$ denote the probability density function derived at time index i from the residuals given over the set of observations $\{y_{i,k}\}_{k=1}^{K_{\max}}$ such that

$$P[R_1 \leq \mathcal{E} \leq R_2] = \int_{R_1}^{R_2} q(\mathcal{E}', \mathcal{R}, \mathcal{Q}) d\mathcal{E}' = \int_{R_1}^{R_2} q_{i,\mathcal{E}} d\mathcal{E}'. \quad (36)$$

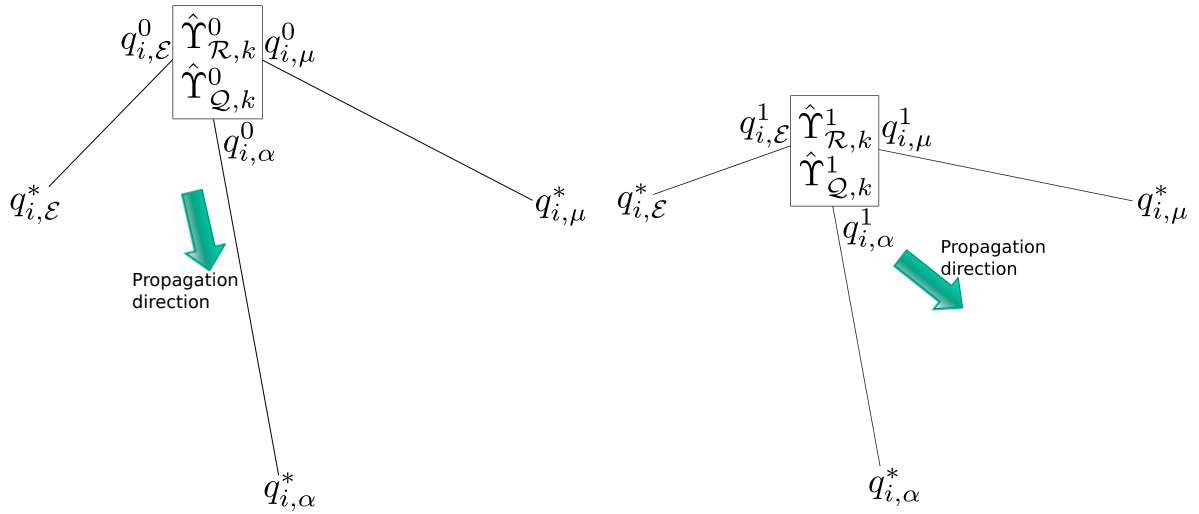
Let $q_{i,s}$ denote the probability density function for the state-space variable s derived at time index i from the deviations given over the set of state-space vectors $\{X_{i,k,s}\}_{k=1}^{K_{\max}}$ such that

$$P[R_3 \leq s \leq R_4] = \int_{R_3}^{R_4} q(s', \mathcal{R}, \mathcal{Q}) ds' = \int_{R_3}^{R_4} q_{i,s} ds'. \quad (37)$$

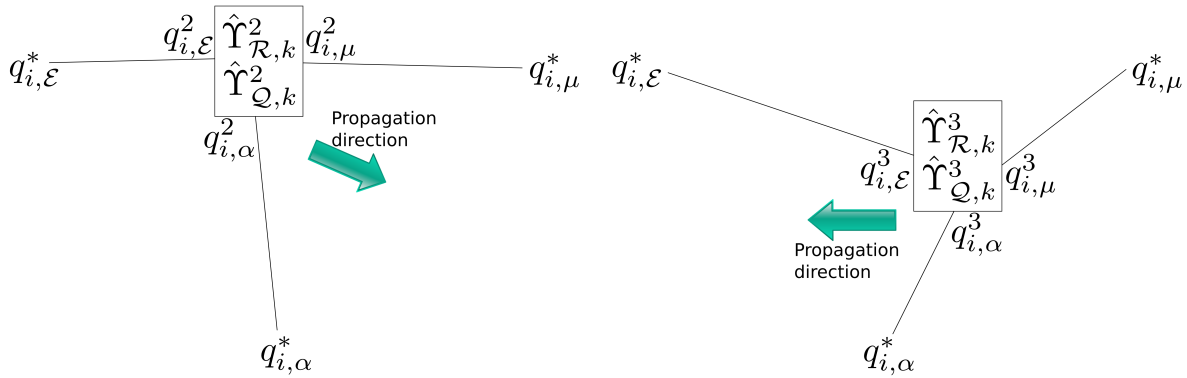
A conditioned observation probability density function $q_{i,\mathcal{E}}^*$ is defined as the probability density function $q_{i,\mathcal{E}}$, given in equation (36), which uses the set $[\mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}_{\sigma_{\mathcal{E}}}]$ to satisfy the condition given in equation (32) as

$$P[R_5 \leq \mathcal{E} \leq R_6] = \int_{R_5}^{R_6} q(\mathcal{E}', \mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}_{\sigma_{\mathcal{E}}}) d\mathcal{E}' = \int_{R_5}^{R_6} q_{i,\mathcal{E}}^* d\mathcal{E}'. \quad (38)$$

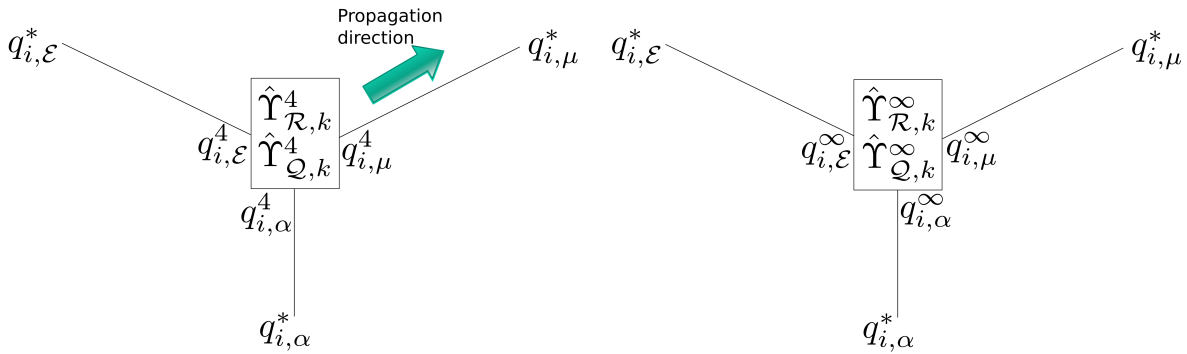
A conditioned process probability density function $q_{i,s}^*$ is defined as the probability density function $q_{i,s}$, given in equation (37), which uses the set $[\mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}]$ to satisfy the condition given in equation (35) as



(a) In the initial iteration it is determined that the amplitude parameter is the least satisfied parameter and second the mean parameter. (b) After the first iteration it is determined that the amplitude parameter and the mean parameter can be better satisfied.



(c) After the second iteration it is determined that the amplitude parameter and the mean parameter can be better satisfied. (d) After the third iteration it is determined that the observation vectors need to be tracked better.



(e) After the fourth iteration it is determined that the mean parameter can be better satisfied (f) Through the decaying variable γ^t the search algorithm terminates.

Fig. 3. An visual example of the operations of the unsupervised search algorithm attempting to satisfy the BVEP criterion.

$$P[R_7 \leq s \leq R_8] = \int_{R_7}^{R_8} q(s', \mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}) ds' = \int_{R_7}^{R_8} q_{i,s}^* ds'. \quad (39)$$

The performance of the current estimate $\Upsilon_{\mathcal{R},k}$ and $\Upsilon_{\mathcal{Q},k}$ is defined by a criterion that evaluates how well the conditions stated in equation (30) and equation (33) are satisfied. The current estimates are recursively updated and are denoted by $\hat{\Upsilon}_{\mathcal{R},k}^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell$, where ℓ denotes the iteration number. The current estimates $\hat{\Upsilon}_{\mathcal{R},k}^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell$ are used to derive the set of probability density functions $\{\hat{q}_{i,\mathcal{E}}^\ell\}$, $\forall i$, and $\{\hat{q}_{i,s}^\ell\}$, $\forall i$.

A f-divergent distance known as the Hellinger distance [22] is used to measure the similarity between the probability density functions $\hat{q}_{i,\mathcal{E}}^\ell$ and $q_{i,\mathcal{E}}^*$. The modified Hellinger distance $\mathcal{H}_{i,\mathcal{E}}$, $\mathcal{H}_{i,\mathcal{E}} \in [0, 1]$, is computed as

$$\mathcal{H}_{i,\mathcal{E}} = 1 - \sqrt{1 - \sqrt{\int_{-\infty}^{\infty} \hat{q}_{i,\mathcal{E}}^\ell q_{i,\mathcal{E}}^* d\mathcal{E}'}}}, \quad (40)$$

where a value of $\mathcal{H}_{i,\mathcal{E}} \rightarrow 1$ means high similarity between $\hat{q}_{i,\mathcal{E}}^\ell$ and $q_{i,\mathcal{E}}^*$, while $\mathcal{H}_{i,\mathcal{E}} \rightarrow 0$ means low similarity. The modified Hellinger distance is also used to measure the similarity for the state-space variables. The modified Hellinger distance $\mathcal{H}_{i,s}$, $\mathcal{H}_{i,s} \in [0, 1]$, is computed as

$$\mathcal{H}_{i,s} = 1 - \sqrt{1 - \sqrt{\int_{-\infty}^{\infty} \hat{q}_{i,s}^\ell q_{i,s}^* ds'}}}, \quad (41)$$

where a value of $\mathcal{H}_{i,s} \rightarrow 1$ means high similarity between $\hat{q}_{i,s}^\ell$ and $q_{i,s}^*$, while $\mathcal{H}_{i,s} \rightarrow 0$ means low similarity. The BVS is defined to encapsulates all the similarity metrics as

$$\Gamma_i = \min \left(\{\mathcal{H}_{i,s}\}_{s=1}^{s=S} \cup \{\mathcal{H}_{i,\mathcal{E}}\} \right). \quad (42)$$

Finding optimal estimates for $\hat{\Upsilon}_{\mathcal{R},k}^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell$ requires a stable covariance matrix $\mathfrak{B}_{(i|i),k}$. Equation (21) states that the predicted covariance matrix $\mathfrak{B}_{(i|i),k}$ should converge to a stable matrix under certain prerequisite conditions. Let \mathcal{I}_T denote the number of time steps which are required to ensure the predicted covariance matrix $\mathfrak{B}_{(\mathcal{I}_T|\mathcal{I}_T-1),k}$ converges to ensure a stable covariance matrix $\mathfrak{B}_{(\mathcal{I}_T|\mathcal{I}_T),k}$. The BVS is deemed accurate at \mathcal{I}_T which is defined as

$$\Gamma_{\mathcal{I}_T} = \min \left(\{\mathcal{H}_{\mathcal{I}_T,s}\}_{s=1}^{s=S} \cup \{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\} \right). \quad (43)$$

The BVEP criterion is defined as the BVS which optimally maximises the conditions. The BVEP criterion is thus formally defined as

$$\Gamma_{\mathcal{I}_T}^* = \max_{\Upsilon_{\mathcal{R},k}^\ell \in v_{\mathcal{R}}, \Upsilon_{\mathcal{Q},k}^\ell \in v_{\mathcal{Q}}} \{\Gamma_{\mathcal{I}_T}\}. \quad (44)$$

If the reflectance values of the spectral bands are correlated, then the BVS is expanded to compensate for this as

$$\Gamma_{\mathcal{I}_T} = \min \left(\left\{ \{\mathcal{H}_{\mathcal{I}_T,b,s}\}_{s=1}^{s=S} \right\}_{b=1}^{b=B} \cup \{\mathcal{H}_{\mathcal{I}_T,b,\mathcal{E}}\}_{b=1}^{b=B} \right). \quad (45)$$

E. Bias-Variance Search algorithm

The Bias-Variance Search Algorithm is proposed that will attempt to estimate $\hat{\Upsilon}_{\mathcal{R},k}^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell$ to satisfy the BVEP criterion using the BVS given in equation (43). The BVSA starts by creating ideal operating conditions for each parameter in the EKF to compute $q_{i,\mathcal{E}}^*$ and $q_{i,s}^*$; followed by using a hill-climbing algorithm approach to search for a set of $\hat{\Upsilon}_{\mathcal{R},k}^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell$ that will satisfy at best the ideal operating conditions for all the parameters in the EKF.

The first ideal condition is a system that employs perfect tracking of the observations. This ideal condition is used to create the probability density function $q_{i,\mathcal{E}}^*$, which is obtained by

$$q_{i,\mathcal{E}}^* = \{q_{i,\mathcal{E}} : \{\zeta_k\} \rightarrow -\infty; \{\varsigma_{k,s}\} \rightarrow \infty, \forall s\}. \quad (46)$$

Under perfect condition the probability density function $q_{i,\mathcal{E}}^*$ should tend to be an impulse of unity power situated around the zero position, meaning zero error residual is measured. The second ideal condition is a system that employs a stable state-space variable. This ideal condition is used to create the probability density function $q_{i,s}^*$. This is obtained by

$$q_{i,s}^* = \{q_{i,s} : \{\zeta_k\} \rightarrow \infty; \{\varsigma_{k,\{s'\}_{s'=1}^{s'=S} \setminus s}\} \rightarrow \infty; \{\varsigma_{k,s}\} \rightarrow -\infty\}. \quad (47)$$

This condition creates an environment which attempts to track the state-space variable s with the smallest variation. After the ideal observation conditions' probability density functions $q_{i,\mathcal{E}}^*$ and $q_{i,s}^*$ are computed, a hill-climbing search algorithm is applied to find a set of initial parameters that will best satisfy all these ideal conditions. The BVSA iteratively searches the parameter space and is shortly described below in the following steps (a flow diagram is also provided in figure 4):

- 1) The BVSA starts with the initial parameters set as $\zeta_k^0 = 0\text{dB}$, $\forall k$, and $\varsigma_{k,s}^0 = 0\text{dB}$, $\forall k, s$.
- 2) Compute the state-space vector $\vec{X}_{(\mathcal{I}_T|\mathcal{I}_T),k}$ at time \mathcal{I}_T using the same $\hat{\Upsilon}_{\mathcal{R},k}^\ell = \zeta_k^\ell$ and $\hat{\Upsilon}_{\mathcal{Q},k}^\ell = \{\zeta_k^\ell\}_{s=1}^{s=S}$ for every time series in the set $\{\vec{y}_k\}_{k=1}^{k=K_{\max}}$.
- 3) Obtain the probability density function of the residual errors $q_{\mathcal{I}_T,\mathcal{E}}^\ell$ over the K_{\max} time series at time index \mathcal{I}_T .
- 4) Obtain the probability density function of the variable distribution $q_{\mathcal{I}_T,s}^\ell$ of the state-space variable s over the K_{\max} time series at time index \mathcal{I}_T .
- 5) Compute the modified Hellinger distance $\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}$ as shown in equation (40).
- 6) Compute the modified Hellinger distance $\mathcal{H}_{\mathcal{I}_T,s} \forall s$ as shown in equation (41).
- 7) Determine the best performing condition $\mathcal{H}_{\text{best}}$ as

$$\mathcal{H}_{\text{best}} = \max \left\{ \{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\} \cup \{\mathcal{H}_{\mathcal{I}_T,s}\}_{s=1}^{s=S} \right\}. \quad (48)$$

- 8) Determine the worst performing condition $\mathcal{H}_{\text{worst}}$ as

$$\mathcal{H}_{\text{worst}} = \min \left\{ \{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\} \cup \{\mathcal{H}_{\mathcal{I}_T,s}\}_{s=1}^{s=S} \right\}. \quad (49)$$

- 9) Adjust the new ζ_k^ℓ according to its relative position to the best and worst performing parameters using a threshold

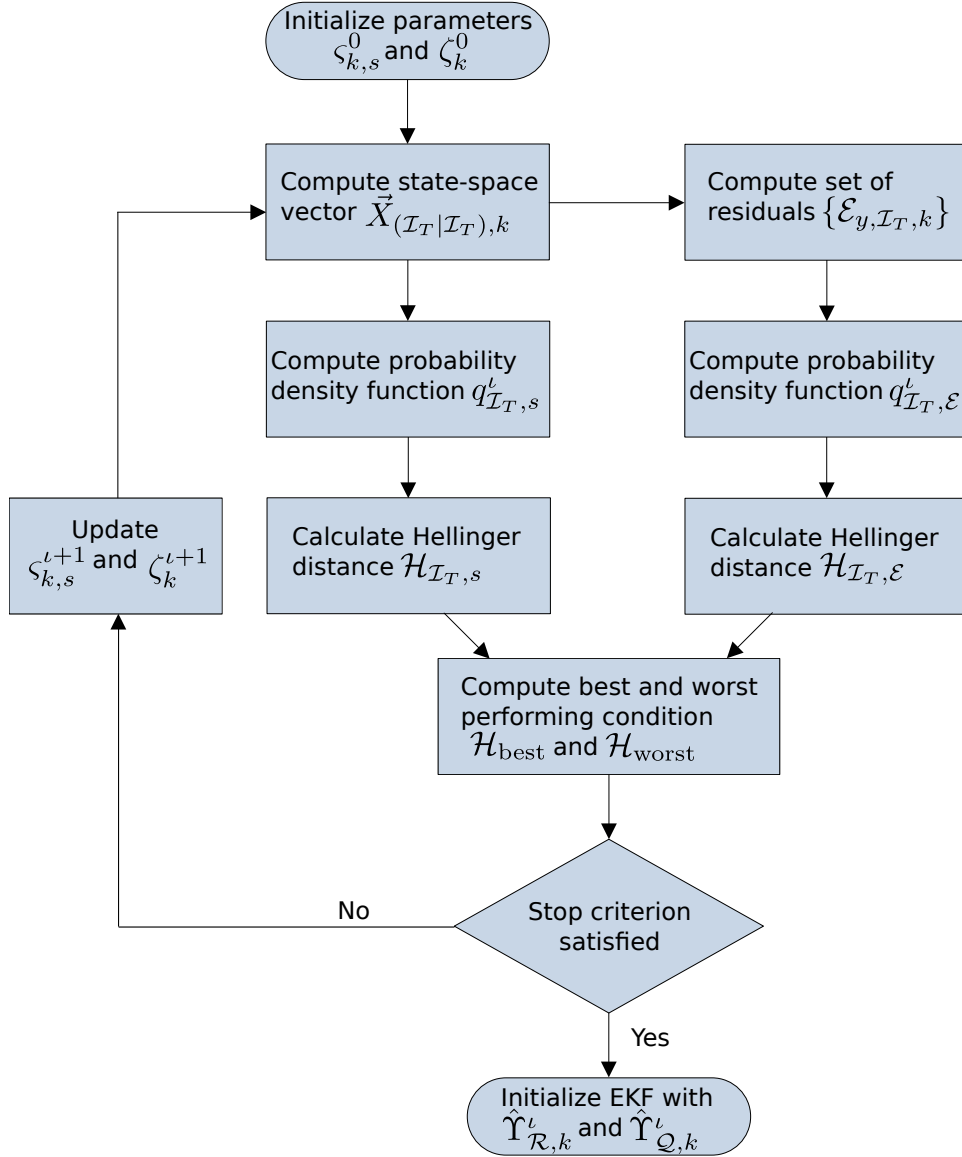


Fig. 4. A flow diagram that describes the various steps of the unsupervised search algorithm.

$\rho_{\mathcal{H}}, \rho_{\mathcal{H}} \in [0, 1], \rho_{\mathcal{H}} \in \mathbb{R}$. The adjustment is made as

$$\zeta_k^{\ell+1} = \begin{cases} \zeta_k^{\ell} + \gamma^{\ell} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} > \rho_{\mathcal{H}} \right) \\ \zeta_k^{\ell} - \gamma^{\ell} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} \leq \rho_{\mathcal{H}} \right) \end{cases}. \quad (50)$$

The variable γ^{ℓ} is a decreasing scalar measured in decibels and is a non-negative real number.

- 10) Adjust the new ζ_k^{ℓ} according to its relative position to the best and worst performing parameters using a threshold $\rho_{\mathcal{H}}, \rho_{\mathcal{H}} \in [0, 1], \rho_{\mathcal{H}} \in \mathbb{R}$. The adjustment is made as

$$\zeta_{k,s}^{\ell+1} = \begin{cases} \zeta_{k,s}^{\ell} + \gamma^{\ell} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,s} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} > \rho_{\mathcal{H}} \right) \\ \zeta_{k,s}^{\ell} - \gamma^{\ell} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,s} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} \leq \rho_{\mathcal{H}} \right) \end{cases}. \quad (51)$$

The variable γ^{ℓ} is a decreasing scalar measured in decibels and is a non-negative real number.

Repeat steps 2–10 until one of the parameters ζ_k or $\zeta_{k,s}$

stabilises, or a stopping criterion is satisfied. This process of adapting the state parameters $\hat{Y}_{\mathcal{R},k}^{\ell}$ and $\hat{Y}_{\mathcal{Q},k}^{\ell}$ are illustrated in a visual example in figure 3. It should be noted that figure 3 is a visual example and not a mathematical description of the unsupervised search algorithm. In this illustrative example the search algorithm converges after a set number of iterations. After the search algorithm converges, the estimates $\hat{Y}_{\mathcal{R},k}^{\ell}$ and $\hat{Y}_{\mathcal{Q},k}^{\ell}$ are used to initialise the EKF.

IV. EXPERIMENTAL RESULTS

A. Optimizing using the BVEP criterion

In this section the results obtained by using the BVSA are discussed. The BVSA is an iterative algorithm that moves the BVS through a defined parameter space. In each epoch the algorithm attempts to minimise the standard deviation of all the state-space variables while simultaneously minimising the

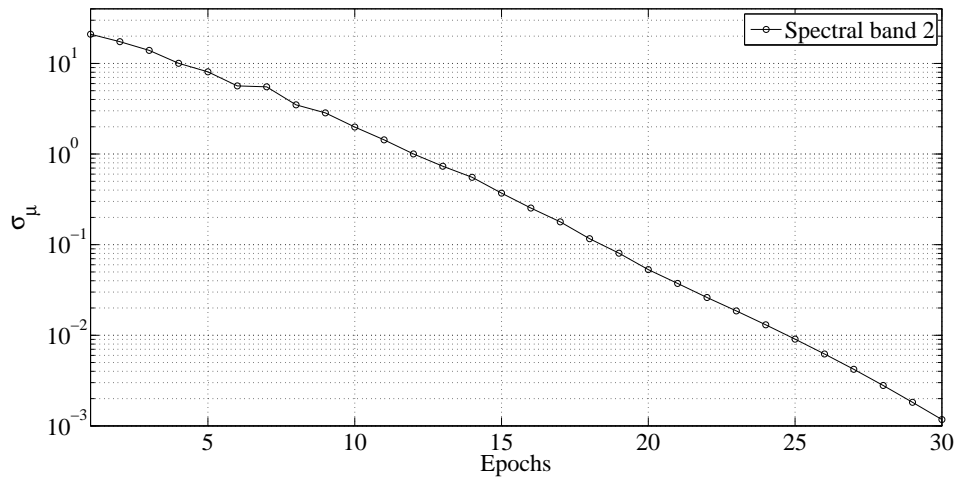


Fig. 5. The expected standard deviation of the mean parameter computed over all the time series for the second MODIS spectral band on the Gauteng province study area as a function of epoch.

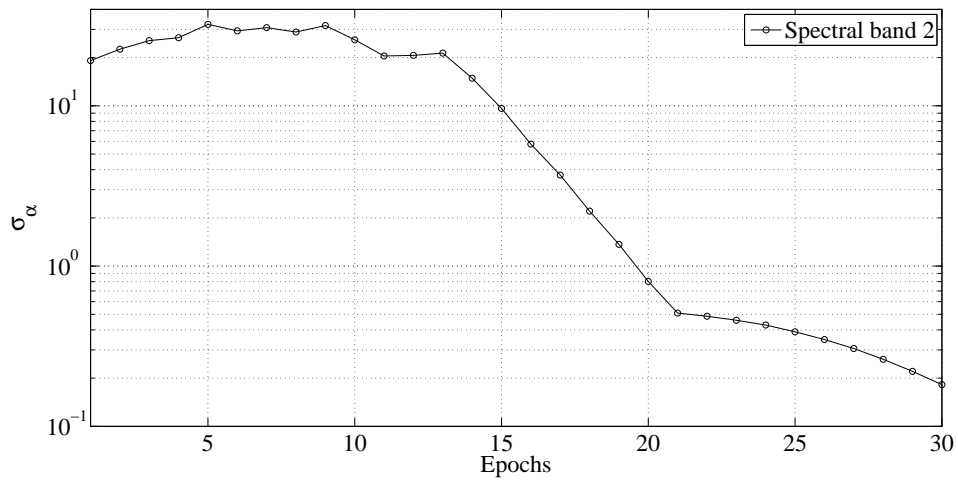


Fig. 6. The expected standard deviation of the amplitude parameter computed over all the time series for the second MODIS spectral band on the Gauteng province study area as a function of epoch.

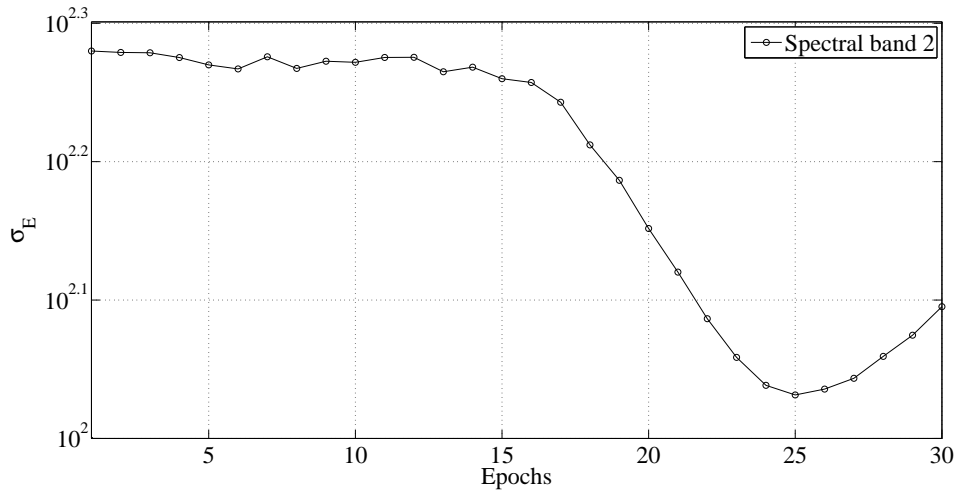


Fig. 7. The expected residuals computed over all the time series for the second MODIS spectral band on the Gauteng province study area as a function epoch.

observation error. The experiments were conducted on labelled time series extracted from the Gauteng province.

In figure 5, the standard deviation σ_μ of the mean parameter obtained by fitting the cosine model to the second MODIS spectral band is illustrated as a function of epoch in the BVSA. The standard deviation reported here is the average standard deviation found over all the time series extracted from the Gauteng province study area. It is clear from the graph that the standard deviation decreases as more epochs are processed, which implies that the mean parameter appears to become more stable with each iteration.

The standard deviation σ_α of the amplitude parameter that is used to fit the second MODIS spectral band is illustrated as a function of epoch of the BVSA in figure 6. The standard deviation reported here is the average standard deviation found over all the time series extracted from the Gauteng province study area. It is clear from the graph that the standard deviation decreases as more epochs are processed, implying increasing stability with further iterations.

In figure 7, the mean residual σ_ε over all the time series' difference between the actual observations and EKF output is illustrated as a function of epoch in the BVSA. It is observed that the residual decreased significantly after the 15th epoch. Overfitting appears towards the end of the optimisation process. This overfit can occur on any metric and in this experiment the overfit is observed on the σ_ε metric after the 25th epoch. This overfit defines the end of the search and is used as an early stopping criterion.

The process noise covariance vector \mathcal{Q} and observation noise covariance scalar \mathcal{R} used in the 25th epoch are then used to initialise the Extended Kalman filter for the experiments. The BVSA is applied independently to each of the seven spectral bands and NDVI time series to obtain a process noise covariance vector \mathcal{Q} and observation noise covariance scalar \mathcal{R} for each spectral band.

B. Parameter evaluation

In this section the measured parameters of three different methods are compared, namely the least squares, the EKF using the ALS method and the BVEP criterion. The comparison is based on the standard deviation σ_μ of the mean parameter, the standard deviation σ_α of the amplitude parameter, and the observation error σ_ε . A mean (amplitude) parameter with a small standard deviation indicates a stable variable. A small σ_ε indicates a well-estimated output when compared to the actual observations.

An analysis of the standard deviation of the parameter evaluated from the Gauteng province's data set is presented in table III. The EKF using the ALS method is denoted by EKF_{ALS} and the EKF using the BVEP criterion is denoted by EKF_{BVEP} . The EKF_{ALS} method increased its residuals in spectral bands 2 and 5 to improve the parameters stability when compared to the non-linear least squares method. In spectral bands 1, 3, and 4 the mean parameter's standard deviation σ_μ was increased to improve the other two metrics. In spectral bands 6 and 7, EKF_{ALS} performed better than the non-linear least squares in all the metrics. In the NDVI

TABLE III
PARAMETERS EVALUATION OF ALL THREE METHODS FOR THE GAUTENG PROVINCE STUDY AREA. THE MEASUREMENTS ARE MADE ON ALL SEVEN MODIS SPECTRAL BANDS AND NDVI.

Spectral Band		Mode		
		Least squares	EKF_{ALS}	EKF_{BVEP}
NDVI	σ_ε	0.04	0.002	0.003
	σ_μ	0.01	0.07	0.05
	σ_α	0.01	0.06	0.01
Band 1	σ_ε	96.6	90.8	44.8
	σ_μ	17.7	21.3	0.01
	σ_α	22.5	17.3	15.3
Band 2	σ_ε	156.4	204.2	123.4
	σ_μ	49.1	29.8	0.01
	σ_α	54.9	25.5	0.5
Band 3	σ_ε	55.1	46.7	38.5
	σ_μ	10.2	14.9	0.03
	σ_α	14.0	12.2	0.02
Band 4	σ_ε	63.3	57.0	42.7
	σ_μ	12.6	19.2	0.04
	σ_α	14.7	14.5	0.03
Band 5	σ_ε	153.2	162.9	105.3
	σ_μ	47.4	26.6	0.01
	σ_α	54.2	22.6	0.01
Band 6	σ_ε	157.3	130.5	87.3
	σ_μ	29.8	24.9	0.01
	σ_α	34.8	22.2	0.01
Band 7	σ_ε	158.0	151.9	71.9
	σ_μ	27.8	23.0	0.02
	σ_α	35.0	21.7	20.5

case, the EKF_{ALS} decreased its observation error at the cost of parameter stability when compared to the non-linear least squares.

The EKF_{BVEP} performed better than all methods in all the spectral band experiments, except for the NDVI experiments, where the EKF_{BVEP} improves its tracking by increasing the variation in the mean parameters. A peculiar observation was made for the EKF_{BVEP} in spectral bands 1 and 7. For the first spectral band, overfitting was observed in the amplitude parameter early in the BVSA, which is used as a early stopping criterion. For the seventh spectral band case the standard deviation σ_α of the amplitude parameter slowly monotonically decreased for each epoch of the BVSA until on overfit was reported on the residuals σ_ε at the 22nd epoch. If the overfit did not occur, then the standard deviation σ_α of the amplitude parameter would still have steadily decrease.

C. Classification

In this paper a Multilayer Perceptron (MLP) was used for land cover classification. The MLP is a feedforward Artificial Neural Network (ANN) model that uses multiple layers of neurons to distinguish inputs that are not linearly separable. The case for using the MLP for urban land use classification rather than a maximum likelihood method was made in [23]. It was shown that the MLP learned the complex non-linear interdependencies of the multi-dimensional time series data

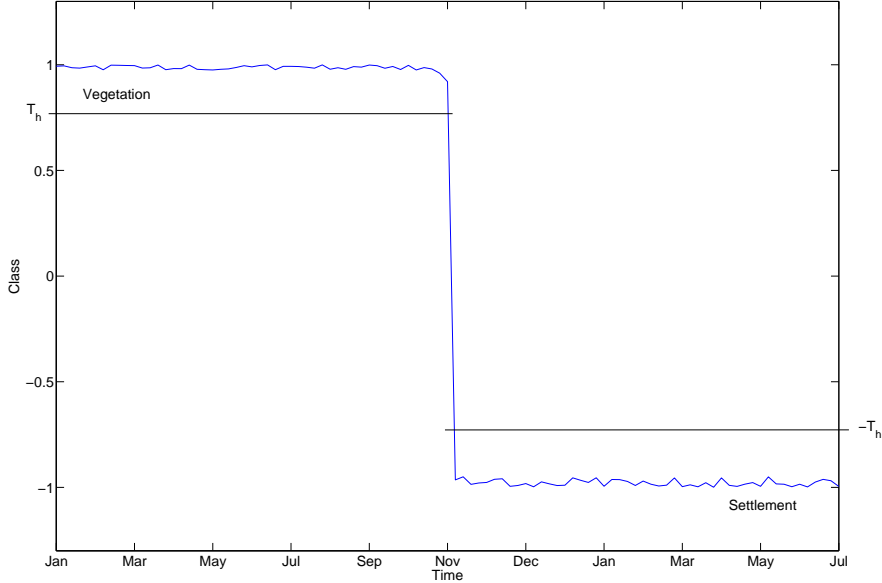


Fig. 8. An example of persistent change in land cover class label over a time period of several months.

TABLE IV

CLASSIFICATION ACCURACY OF MLP USING THREE REGRESSION METHODS. EACH ENTRY GIVES THE AVERAGE CLASSIFICATION ACCURACY FOR EACH MODE, CALCULATED OVER 10 REPEATED INDEPENDENT EXPERIMENTS ALONG WITH THE CORRESPONDING STANDARD DEVIATION. THE AVERAGE CLASSIFICATION ACCURACY IS GIVEN AS A PERCENTAGE FOR EACH OF THE CLASSES OVER A NUMBER OF SPECTRAL BAND COMBINATIONS (NDVI, FIRST 2 SPECTRAL BANDS (RED AND NEAR INFRARED SPECTRAL BANDS) AND ALL 7 SPECTRAL BANDS).

Province	Spectral Band	Class	Mode		
			Least squares	EKF _{ALS}	EKF _{BVEP}
Gauteng	NDVI	Vegetation	92.5 ± 4.9	89.3 ± 4.8	91.4 ± 5.7
		Settlement	88.6 ± 6.4	72.1 ± 16.9	86.9 ± 9.1
	2 Bands	Vegetation	97.5 ± 1.8	90.6 ± 2.9	98.6 ± 1.0
		Settlement	95.1 ± 2.6	87.6 ± 3.2	96.2 ± 1.5
	7 Bands	Vegetation	98.8 ± 0.4	95.3 ± 1.8	99.9 ± 0.1
		Settlement	98.2 ± 0.5	94.8 ± 2.4	99.9 ± 0.1

derived from multiple spectral bands. Another example is the implementation of a MLP using a sliding window to classify informal settlement in a MODIS time series [24]. The method employs an iteratively retrained MLP described in [24] to capture all local patterns and to compensate for the time-varying climate change in the geographical area.

The MLP comprises an input layer, one hidden layer and an output layer. All hidden and output nodes used a tangent sigmoid activation function in each node. The input layer accepts input vectors for classification, while the output layer represents the likelihood that an input belongs to a specific class. The MLP output was in the range (-1;1), where 1 represents a 100% certainty of class membership to natural vegetation given the input vector, while a -1 represents a 100% certainty of settlement. For correct classification it was required that the MLP correctly classified the time series according to a threshold T_h . This threshold was used to impose a strict evaluation on the output class membership stream to ensure that coherent classification was achieved. This means that the MLPs tangent sigmoid activation function output was

classified as vegetation in the range $[T_h, 1]$, settlement in the range $[-1, -T_h]$, and uncertain in the range $(-T_h, T_h)$. The weights in the training phase of the MLP were determined using a steepest descent gradient optimization method, with gradients estimated using backpropagation [25]. A validation set was used for initial MLP architecture optimization by testing the generalization error to identify overfitting of the network for each study area. The retraining at each time increment caused a small adaptation of the weights, and has low complexity due to the small incremental MLP weight changes over each 8 day increment. These small MLP weight changes only required 300 epochs at each time increment for network adaptation.

The combination of different spectral band's state-space vectors used as input vectors to the MLP were adopted from the work presented in [24]. The NDVI, first two spectral bands (Red and Near Infrared spectral bands) and all seven land bands were investigated. In table IV, the classification accuracies are reported for the three methods when the EKF is used to extract the features. The average classification

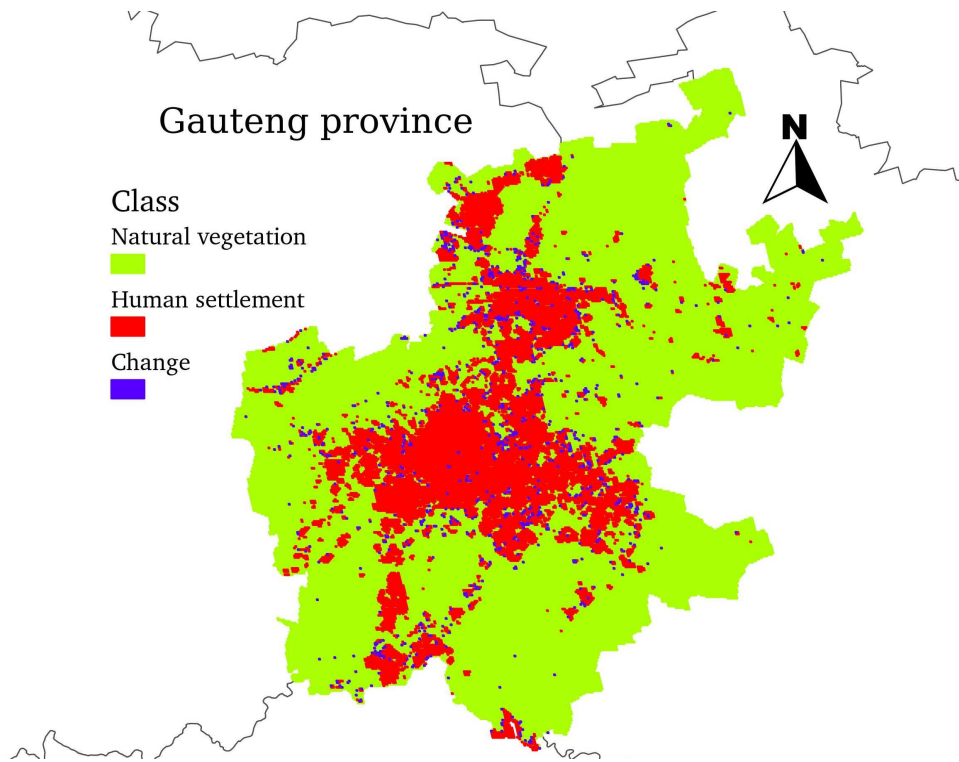


Fig. 9. A classification/ change detection map of the entire Gauteng province (19676 km²).

accuracy is calculated with cross validation using 10 repeated independent experiments [26].

From these results it was concluded that the EKF_{BVEP} performed better than any experiment conducted on the spectral bands using the EKF_{ALS} and the least squares. This could be owing to the fact that the BVEP criterion utilises spatial information that is inherent in the set of time series. In the case of the NDVI time series, the least squares method did provide higher classification accuracies when compared to the EKF initialize with either the BVEP or the ALS.

A general improvement trend was observed when using more spectral bands in the experiments. The prospect of acceptable land cover classification was confirmed as possible by either using the NDVI time series or the first two spectral bands time series of the MODIS data, as this was supported by the results in [27]. The classification accuracies produced by the MLP were however found to be the highest when using all seven spectral bands.

D. Change detection

As stated in section III-A, a class label is assigned to each pixel at each time index as shown in equation 5. Land cover change is declared when a pixel's time series has a persistent change in class label as a function of time (equation (6)). This type of change detection is grouped as a post-classification change detection algorithm [6]. The MLP's output was used to monitor if any change was detected in the time series. Land cover change is declared in this section when the majority of the class labels in the first year (2001) is different when compared to the majority of the class labels in the last

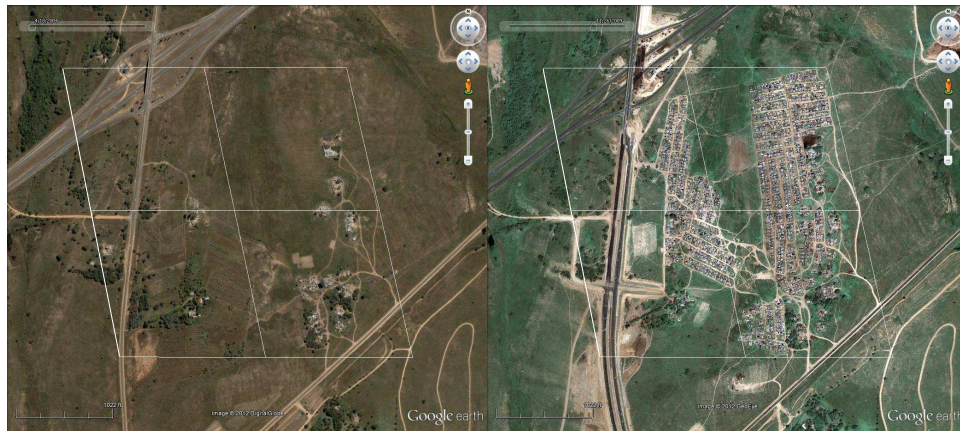
year (2011). In figure 8, an illustrative example is shown of persistent change in class labels from natural vegetation to human settlements.

The change detection accuracy is based on the performance of the classification accuracy reported in table IV. The change detection accuracies for the labelled data set is presented in table V along with the false alarms in parentheses. The best performing classification experiment was the EKF_{BVEP} and was benchmarked to the least squares experiment. The EKF_{BVEP} experiment's results perform better than the least squares in every spectral band combination.

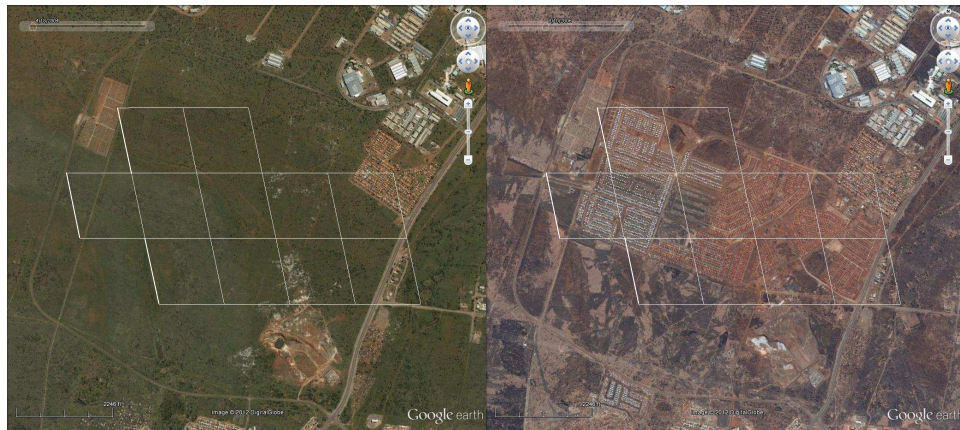
A similar trend observed in the classification accuracies in section IV-C is also observed in this section. An improvement in the experimental results is reported when more spectral bands are used.

TABLE V
THE LAND COVER CHANGE DETECTION ACCURACY ON THE LAND COVER CHANGE DATA SET IN THE GAUTENG PROVINCE. EACH ENTRY GIVES THE TRUE POSITIVES IN PERCENTAGE (FALSE POSITIVES IN PARENTHESES).
THE LAND COVER CHANGE DETECTION ACCURACY IS GIVEN AS A PERCENTAGE OVER A NUMBER OF SPECTRAL BAND COMBINATIONS (NDVI, FIRST 2 SPECTRAL BANDS (RED AND NEAR INFRARED SPECTRAL BANDS) AND ALL 7 SPECTRAL BANDS).

Province	Spectral Band	Mode	
		Least squares	EKF_{BVEP}
Gauteng	NDVI	80.0 (16.7)	83.4 (17.0)
	2 Bands	87.7 (11.8)	92.1 (9.9)
	7 Bands	94.3 (2.5)	95.5 (1.6)



(a) Land cover change detected in Nietgedacht ($25^{\circ}59'15.58''S$, $27^{\circ}55'46.46''E$). The Quickbird image on the left was taken on 18 April 2004 and the right on 26 November 2009 (Courtesy of GoogleTM Earth).



(b) Land cover change detected in Ga-Rankuwa ($25^{\circ}34'21.91''S$, $27^{\circ}58'48.96''E$). The Quickbird image on the left was taken on 18 April 2004 and the right on 10 August 2006 (Courtesy of GoogleTM Earth).



(c) Land cover change detected in Devon ($26^{\circ}20'53.34''S$, $28^{\circ}46'29.50''E$). The Quickbird image on the left was taken on 23 October 2004 and the right on 22 August 2011 (Courtesy of GoogleTM Earth).

Fig. 10. Example of three geographical areas where land cover change was detected in the Gauteng province. The Quickbird imagery is overlaid with a MODIS 500 meter coordinate grid.

TABLE VI
THE CLASSIFICATION AND CHANGE DETECTION RESULTS PRODUCED FOR THE ENTIRE GAUTENG PROVINCE. THE RESULTS ARE PRESENTED IN PERCENTAGE COVER OF TOTAL AREA IN THE PROVINCE.

Feature extraction	Algorithm	Spectral Band	Class allocation [%]		
			Natural vegetation	Human settlement	Land cover change
EKF	MLP	7 Bands	76.97	21.86	1.17

V. REGIONAL EXPERIMENTS

In this section the optimized EKF was evaluated at a regional scale using a MLP operating on all seven spectral bands. Using all seven spectral bands was preferred based on the classification accuracies presented in table IV and the change detection accuracies presented in table V. The results were produced by processing the entire Gauteng province into three defined categories: natural vegetation, human settlement and land cover change. The results for the experiments are presented in table VI. An illustration of this experiment's output is shown in figure 9, which illustrates the entire Gauteng province. The natural vegetation class covered 76.97% of the province, while the human settlements covered 21.86%. This result supports the concept that Gauteng is a heavily urbanised province. The land cover that changed was flagged at 1.17% of the total area in the province. This is a significant large area that has changed in the study period, which equates to a total land area of 230 km².

Three examples of where land cover change was identified is shown in figure 10. The first illustration in figure 10(a) shows 4 detected MODIS pixels that were subjected to land cover change in the Nietgedacht area (25°59'15.58"S, 27°55'46.46"E). The second illustration in figure 10(b) shows a large area of 11 detected MODIS pixels that were subjected to land cover change in the Ga-Rankuwa area (25°34'21.91"S, 27°58'48.96"E). The last illustration in figure 10(c) shows 3 detected MODIS pixels that were subjected to land cover change in the Devon area (26°20'53.34"S, 28°46'29.50"E).

VI. CONCLUSION

It was demonstrated in this paper that each of the seven spectral bands could be successfully modelled with a triply modulated cosine function for the study area located in the Gauteng province, South Africa. The model was fitted using an EKF, where the parameter derived could be used to separate natural vegetation and human settlements (Table IV).

A novel BVEP criterion was proposed in this paper, which is used to derive initial parameters for an EKF to ensure that it is operating in a desirable state. The EKF was defined as operating in a desirable state when it was tracking the observations with minimum residual and have a stable internal state-space vector. The criterion is used to compute the process noise covariance matrix and observation noise covariance matrix using spatial and temporal information with an unsupervised search algorithm. The resulting parameters are then used to initialize the EKF which is used as a feature extraction method. It was shown that the criterion can be extended to include the spectral information if the spectral bands are assumed to be correlated.

The BVSA is also proposed in this paper, which is a specifically designed unsupervised search algorithm used to adjust the BVS in an attempt to satisfy the BVEP criterion. The BVSA provides covariance matrices that could be used for a variety of different applications.

The parameters extracted from the EKF were encapsulated in an input vector and presented to a MLP for classification. The classification accuracies in most of the experiments were

above 90% for the first two spectral band and all seven spectral band combinations. From the results presented in section IV-C it is clear that better classification accuracies were obtained if all seven spectral bands of the MODIS sensor were used. The feature extracted from the EKF can also be applied in combination with a variety of other classifiers or machine learning methods.

REFERENCES

- [1] G. Daily and P. Ehrlich, "Population, sustainability, and Earth's carrying capacity," *Bioscience*, vol. 42, no. 10, pp. 761–771, Nov. 1992.
- [2] P. Vitousek, H. Mooney, J. Lubchenco, and J. Melillo, "Human domination of Earth's ecosystems," *Science*, vol. 277, pp. 494–499, July 1997.
- [3] R. Lunetta, J. Knight, J. Ediriwickrema, J. Lyon, and L. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sensing of Environment*, vol. 105, no. 2, pp. 142–154, November 2006.
- [4] R. DeFries and J. Chan, "Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data," *Remote Sensing of Environment*, vol. 74, no. 3, pp. 503–515, December 2000.
- [5] 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, January 2007.
- [6] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: a review," *International Journal of Remote Sensing*, vol. 25, no. 9, pp. 1565–1596, May 2004.
- [7] 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.
- [8] W. Kleynhans, J. Olivier, K. Wessels, F. van den Bergh, B. Salmon, and K. Steenkamp, "Improving land cover class separation using an Extended Kalman Filter on MODIS NDVI Time-Series Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 381–385, Apr. 2010.
- [9] M. Rajamani and J. Rawlings, "Estimation of the disturbance structure from data using semidefinite programming and optimal weighting," *Automatica*, vol. 45, no. 1, pp. 142–148, January 2009.
- [10] 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.
- [11] W. Wanner *et al.*, "Global retrieval of bidirectional reflectance and Albedo over land from EOS MODIS and MISR data: theory and algorithm," *Journal of Geophysical Research*, vol. 102, no. D14, pp. 17 143–17 162, July 1997.
- [12] M. Chen, S. Liu, L. Tieszen, and D. Hollinger, "An improved state-parameter analysis of ecosystem models using data assimilation," *Ecological Modelling*, vol. 219, no. 3–4, pp. 317–326, December 2008.
- [13] 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, April 2008.
- [14] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle Filters for Tracking Applications*, 1st ed. London: Artech House, 2004.
- [15] R. Nikoukhan, A. Willsky, and B. Levy, "Kalman filtering and Riccati equations for descriptor systems," Massachusetts Institute of Technology, Laboratory for information and decision systems, Tech. Rep. 0704-0188, January 1991.
- [16] W. Arnold and A. Laub, "Generalized Eigenproblem algorithms and software for algebraic Riccati equations," *Proceedings of the IEEE*, vol. 72, no. 12, pp. 1746–1754, December 1984.
- [17] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 12, pp. 175–184, April 1970.
- [18] B. Carew and P. Belanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariances," *IEEE Transactions on Automatic Control*, vol. 18, no. 6, pp. 582–587, December 1973.
- [19] G. Noriega and S. Pasupathy, "Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data," *IEEE Transactions on Geoscience Remote Sensing*, vol. 35, no. 5, pp. 1146–1159, September 1997.

- [20] B. Odelson, M. Rajamani, and J. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," *Automatica*, vol. 42, no. 2, pp. 303–308, February 2006.
- [21] R. Shumway and D. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, July 1982.
- [22] M. Nikulin, N. Limnois, N. Balakrishnan, W. Kahle, and C. Huber-Carol, *Advances in degradation modeling: Applications to reliability, survival analysis, and finance*, 1st ed. 233 Spring street, New York, USA: Springer, 2010.
- [23] J. Paola and R. Schowengerdt, "Detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 4, pp. 981–996, July 1995.
- [24] B. Salmon, J. Olivier, W. Kleynhans, K. Wessels, F. van den Bergh, and K. Steenkamp, "The use of a Multilayer Perceptron for detecting new human settlements from a time series of MODIS images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 6, pp. 873–883, December 2011.
- [25] C. Bishop, *Neural Networks for Pattern Recognition*, 2nd ed. New York, USA: Oxford University Press, 1995.
- [26] S. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," *Data mining and knowledge discovery*, vol. 1, no. 3, pp. 317–328, September 1997.
- [27] M. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, "MODIS collection 5 global land cover: algorithm refinement and characterization of new datasets," *Remote Sensing of Environment*, vol. 114, no. 1, pp. 168–182, January 2010.



J.C. Olivier received the Ph.D. degree in 1990 from the University of Pretoria in Electrical Engineering. He is currently the head of the School of Engineering at the University of Tasmania, Australia. He was with Bell Northern Research (BNR) in Canada, and with Nokia Research Center in the United States. His research interests are in Estimation and Detection theory, as well as applications of Machine Learning. He serves as an Editor for the IEEE Trans. on Wireless Communications.



W.J. Marais received the B.Eng in computer engineering from the University of Pretoria, Pretoria, South Africa in 2005. He is currently doing the M.Sc. in electrical engineering at the University of Wisconsin, Madison. His research interest include signal processing in the domain of remote sensing.



B.P. Salmon received a M.Eng degree in electronic engineering and a Ph.D. degree in electronic engineering from the University of Pretoria, South Africa in 2008 and 2012 respectively. He is presently a senior researcher in the Remote Sensing Research Unit at the Council for Scientific and Industrial Research. He was the lead author of the paper that won the IEEE Geoscience and Remote Sensing Society 2012 Symposium prize paper award. His research interests are information theory, coding theory, machine learning and graph theory.



T.L. Grobler received his B.Eng. and M.Eng. degrees from the University of Pretoria, South Africa, in 2005 and 2008 respectively. He is currently a Ph.D. student at the University of Pretoria, South Africa. His research interests include remote sensing, time-series analysis, wireless communications, statistical detection and machine learning.



W. Kleynhans received a B.Eng., M.Eng. and Ph.D. (Electronic Engineering) from the University of Pretoria, South Africa as well as an MBA from Heriot-Watt University, Scotland. He is currently a senior researcher with the Remote Sensing Research Unit at the Council for Scientific and Industrial Research in Pretoria, South Africa and is also affiliated with the University of Pretoria. His research interests include remote sensing, time-series analysis, wireless communications, statistical detection and estimation theory as well as machine learning.



K.J. Wessels received an M.Sc. in Landscape Ecology and Conservation Planning from the University of Pretoria (South Africa) in 1997 and a Ph.D. in Geography from University of Maryland (US) in 2005. He was a research associate at NASA Goddard Space Flight Center, Hydrospheric and Biospheric Laboratory (2006-2006). He is presently a chief researcher and leads the Remote Sensing Research Unit within the CSIR Meraka Institute in Pretoria, South Africa. His research interests include time-series analysis of satellite data for monitoring environmental change and the estimation ecosystem state variables and services with remote sensing.



F. van den Bergh received the M.Sc. degree in computer science (machine vision) and the Ph.D. degree in computer science (particle swarm optimization) from the University of Pretoria, Pretoria, South Africa, in 2000 and 2002, respectively. He is currently a principal researcher at the Council for Scientific and Industrial Research. His research interests include automated feature extraction from high-resolution satellite images, as well as automated change detection. He maintains an active interest in particle swarm optimization and machine

learning.