CHAPTER EIGHT

RESULTS

8.1 OVERVIEW

The first part of the chapter studies the effects of different parameter settings to determine their influence on the quality of the solutions. The second part of the chapter explores the classification accuracies of several different methods, while the last part investigates the change detection accuracies of the best performing methods. The chapter concludes with the processing of these methods on large regional scale areas and assessing the outcome.

8.2 GROUND TRUTH DATA SET

A labelled data set, offering ground truth, is required to evaluate the performance of different land cover change detection algorithms. The performance of the methods is measured with a variety of tests to assess accuracy and robustness. Two study areas were investigated in this chapter, namely the Limpopo and Gauteng provinces.

Limpopo province: The Limpopo province is located in the northern parts of South Africa and is largely covered by natural vegetation. The expansion of human settlements, often informal and unplanned, is the most pervasive form of land cover change in the province. Areas were identified where new settlements were known to have been built over the last decade.

Gauteng province: The Gauteng province is located in the highveld of South Africa and is the most urbanised province in the country. The province contributes 33% of the country's national economy. Active migration to the province from other provinces is motivated by the prospect of higher incomes and more diverse employment opportunities. An average growth of 249 310

(a) Quickbird image taken on 1 March 2004 (courtesy of Google^{TM}Earth).

(b) Quickbird image taken on 9 July 2008 (courtesy of Google^{TM}Earth).

(c) Quickbird image taken on 11 December 2009 (courtesy of Google^{TM}Earth).

FIGURE 8.1: Three high resolution images acquired over a residential area called Midstream estates located in Midrand, Gauteng, South Africa. The area was zoned for residential use in 2003 and new settlements were erected only after 9 July 2008.

persons per year within the province has been estimated over the past decade [214, 215]. It should be noted that the Gauteng province only covers 1.4% of the country's total land area, while housing over 20% of the population.

8.2.1 MODIS time series data set

The performance of different land cover change detection methods will be evaluated on a per pixel basis using a set of different spectral bands' time series, which are extracted from the MODIS land surface reflectance product. The MODIS (MCD43A4, Collection V005) 500 metre, Nadir and BRDF adjusted spectral reflectance bands were used, as these significantly reduce the anisotropic scattering effects of surfaces under different illumination and observation conditions [27, 28]. The first two spectral bands (RED and NIR spectral bands) are the only spectral bands available at a spatial resolution of 250 metre, and are not BRDF adjusted. The 500 metre resolution spectral bands were considered to illustrate the

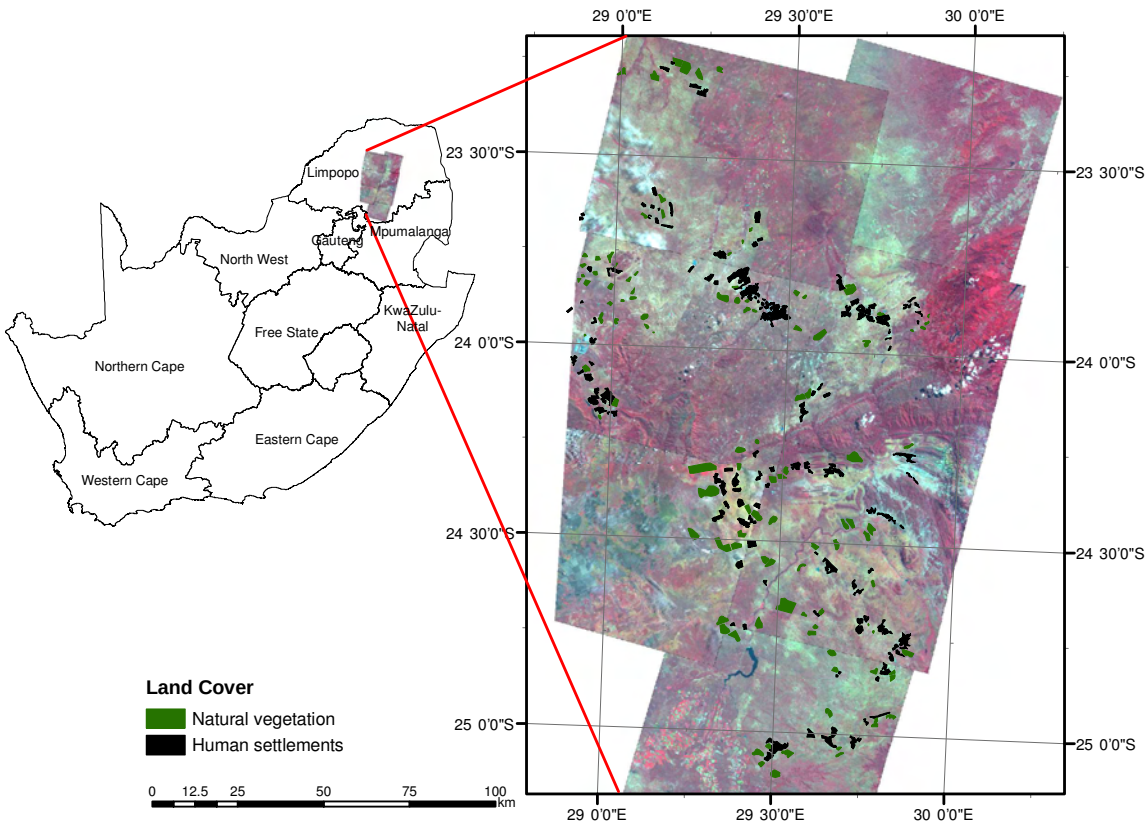


FIGURE 8.2: The Limpopo province study area has land cover types polygons overlayed using Albers projection on SPOT5 RGB 321 imagery that was acquired between March 2006 and May 2006. The SPOT2 images were acquired of the same area in May 2000 [8].

advantages of using additional spectral bands in the analysis. A time series is extracted for all 7 spectral bands from the data set (MODIS tile H20V11) for each pixel in each study area (year 2000–2008).

8.2.2 Manual inspection of study areas

Identification of no change areas: Visual interpretation of SPOT2 (year 2000) and SPOT5 (year 2006 / 2008) high spatial resolution images was used to verify that none of the areas classified as no change, experienced any form of land cover change during the study period.

Identification of change areas: This data set was captured using the same procedure explained for the no change areas, except that areas where new human settlements had formed during the study period were captured.

Even though human settlement expansion is one of the most pervasive forms of land cover change in South Africa, information on this form of land cover change is poorly documented, and vital details

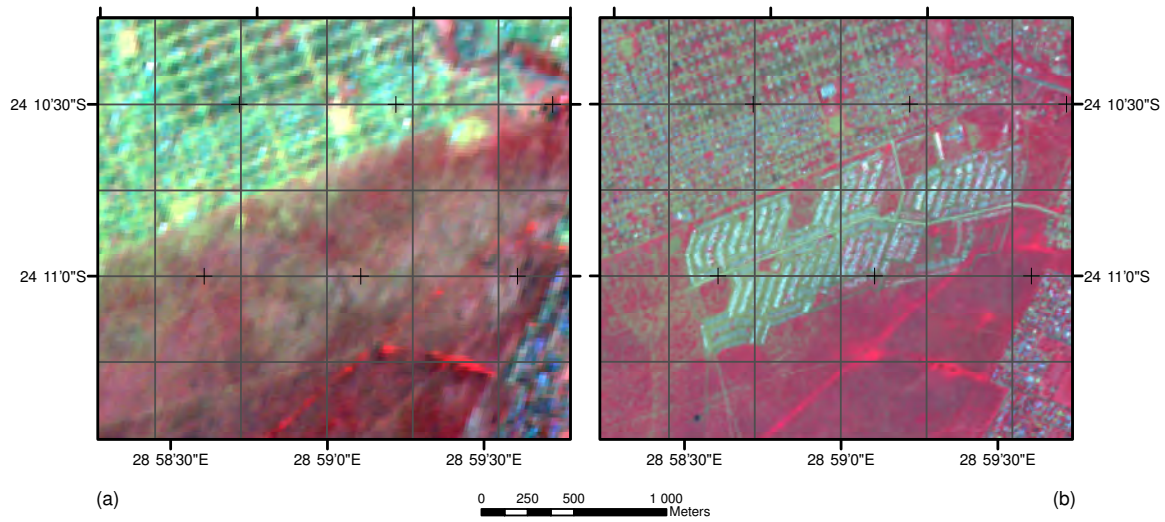


FIGURE 8.3: A land cover change of natural vegetation to human settlement in Sekuruwe. Sekuruwe is a human settlement that is located in the Limpopo province, South Africa. The SPOT2 image (RGB 321) was acquired on 2 May 2000 of the natural vegetation area (a) and a SPOT5 (RGB 321) image was acquired on 1 May 2007 of a newly developed human settlement (b). The SPOT2 and SPOT5 image is projected to a MODIS sinusoidal WGS84 projection and is overlaid with a MODIS 500 metre coordinate grid [8].

such as the date of land cover conversion cannot be determined reliably. An example of inaccurate information is shown in figure 8.1. The local municipality demarcated new roads in a suburban area for future expansion. Unfortunately, no newly developed settlements had been built until quite recently. A good estimate on the date of land cover conversion can be made if regular acquisitions are obtained for a particular area. In this example, if only the images in figure 8.1(a) and figure 8.1(c) were available, then the date of change could be somewhere between March 2004 and December 2009. The real land cover change only occurred after July 2008, which illustrates the importance of the vital statistic of knowing when change occurred.

Once the areas have been identified as change or no change, they are mapped with polygons on the geocoded SPOT imagery, as shown in figure 8.2. The SPOT images are then projected to a MODIS sinusoidal WGS84 projection and is overlaid with a MODIS 500 metre coordinate grid (Figure 8.3). The MODIS grid blocks, which contain the mapped polygons, are thus marked for extraction from the MODIS MCD43A4 product.

8.2.3 Google^{TM}Earth used for visual inspection

Google^{TM}Earth is being used more routinely in visually displaying and validating of geographical areas [216, 217]. As an additional validation step, the MODIS pixel coordinates of interest were transformed into a KML (Keyhole Markup Language) file and visually inspected in Google^{TM}Earth. The true colour of the high resolution Quickbird images available in Google^{TM}Earth made a good platform to illustrate some of the findings presented in this chapter.

Google^{TM}Earth operates on a free sharing policy of images and does not have a mandate to buy regular imagery of certain geographical areas. This means that only areas in which suitable images were acquired before and after the settlement formation could be validated using Google^{TM}Earth.

8.2.4 Simulated land cover data set

Accurate date-of-change information was not available for the ground truth data set, preventing the measurement of the delay in detecting change of the proposed methods. Land cover change events were simulated by combining data from natural vegetation and human settlement time series, with the advantage of a known date of change and transition duration [8].

Four testing data subsets were created, based on concatenating time series of different combinations of classes:

- Subset 1: natural vegetation time series (class 1) concatenated to settlement time series (class 2).
- Subset 2: settlement time series (class 2) concatenated to natural vegetation time series (class 1).
- Subset 3: settlement time series (class 2) concatenated to another settlement time series (class 2).
- Subset 4: natural vegetation time series (class 1) concatenated to another natural vegetation time series (class 1).

These four subsets were used to test if the change detection algorithm can detect change reliably on subsets 1 and 2, while not falsely detecting change for subsets 3 and 4.

8.3 SYSTEM OUTLINE

In this section an overall system outline is provided to explain how all the different methods interconnect with one another (figure 8.4) to create a change detection framework. The system starts with the input of time series extracted from the MODIS MCD43A4 land surface reflectance

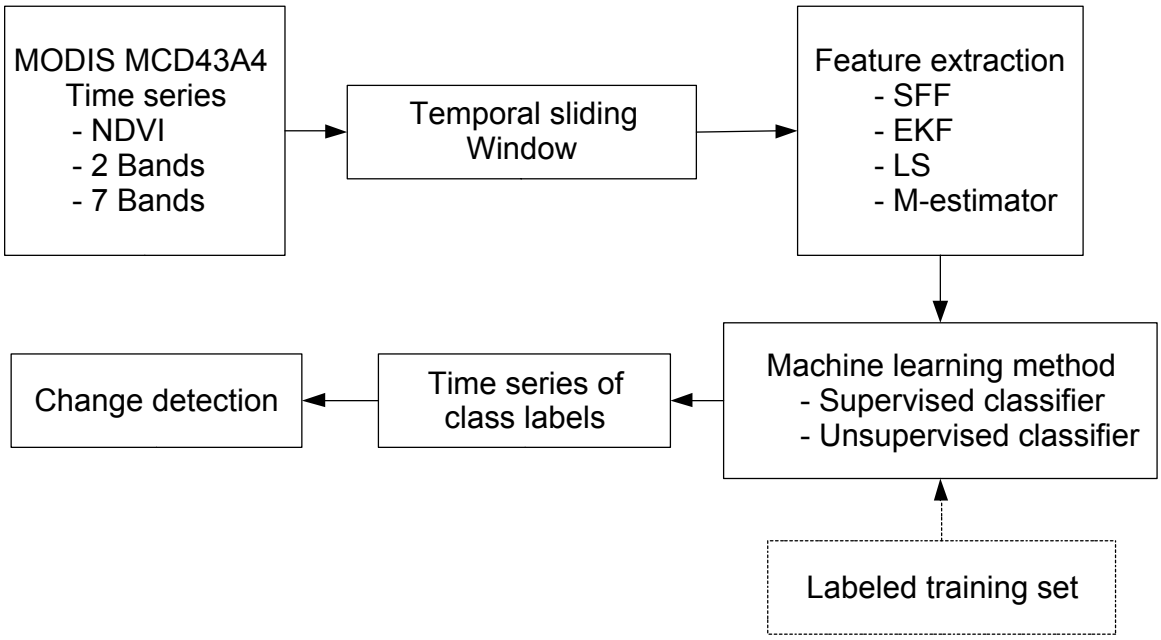


FIGURE 8.4: A flow diagram which provide a complete system outline used in this chapter in all the experiments.

product (section 2.6). The time series used as input can either be one of the following spectral band combinations as listed with the number of dimensions in the feature space as:

- NDVI (2-dimensions),
- first two spectral bands (RED and NIR spectral bands, 4-dimensions), and
- all seven spectral bands (land bands, 14-dimensions).

A temporal sliding window is used to extract sequential subsequences from the time series for analysis. The length of the temporal sliding window is varied, depending on the feature extraction method used. The feature extraction methods applied to these subsequences are listed with their corresponding temporal sliding window length as:

- SFF (6, 12, and 18 months),
- least squares (12 months, see section 8.5.3),
- M-estimator (12 months, see section 8.5.3), and
- EKF (8 days).

The extracted feature vectors are then processed by a machine learning method, which assigns a class label to each feature vector. The machine learning method can be either a supervised classifier, or

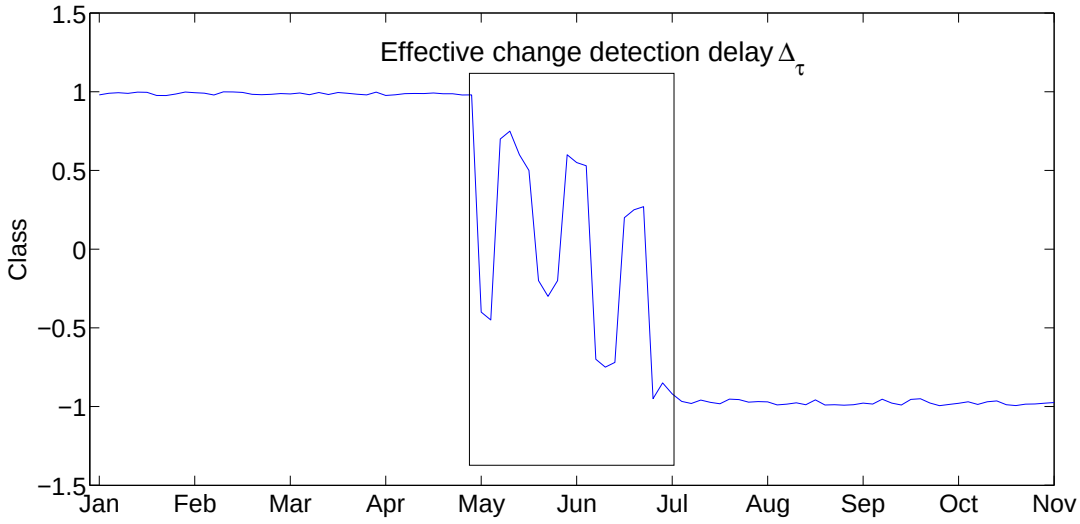


FIGURE 8.5: An illustrative example of the effective change detection delay Δ_τ, which is defined as the time duration it takes after the first acquisition of change in the MODIS time series for the land cover change detection algorithm to detect it.

an unsupervised classifier. The class labels produced by the machine learning method form a new time series, where each time index corresponds to a classification of an extracted temporal subsequence. An example of such a time series consisting of class labels is given in figure 8.5. The class labels in the time series start in the class label *1* (natural vegetation class), and transitions to the class label *-1* (human settlement class), as the position of the temporal sliding window is incremented. It is clear from the illustration that a change in the land cover has occurred in the time series.

A simulated land cover change data set was created in response to the lack of information about when the actual land cover changed (section 8.2.4). In the simulated land cover change data set, the exact position (date) of land cover change in the time series is known. This creates another dimension of evaluation, which enables the quantification of how quickly the land cover change can be detected by the land cover change detection algorithm.

This delay in detecting a change in land cover is termed the effective change detection delay Δ_τ, and is defined as the time duration in which the change detection algorithm is unable to detect the simulated land cover change in subset 1, and subset 2 after the date of change. The concatenation process (section 8.2.4) in the simulated land cover change data set produces an abrupt change in the time series, which does not necessarily represent the reality of human-induced change such as settlement expansion, which could take several months to develop. A blending period (linear blend over 12 and 24 months) from one land cover time series to another was initially considered, but it turned out that it did not affect the ability to detect the land cover change correctly, as this is a property that is exploited in the post-classification change detection approach. The blending model does not

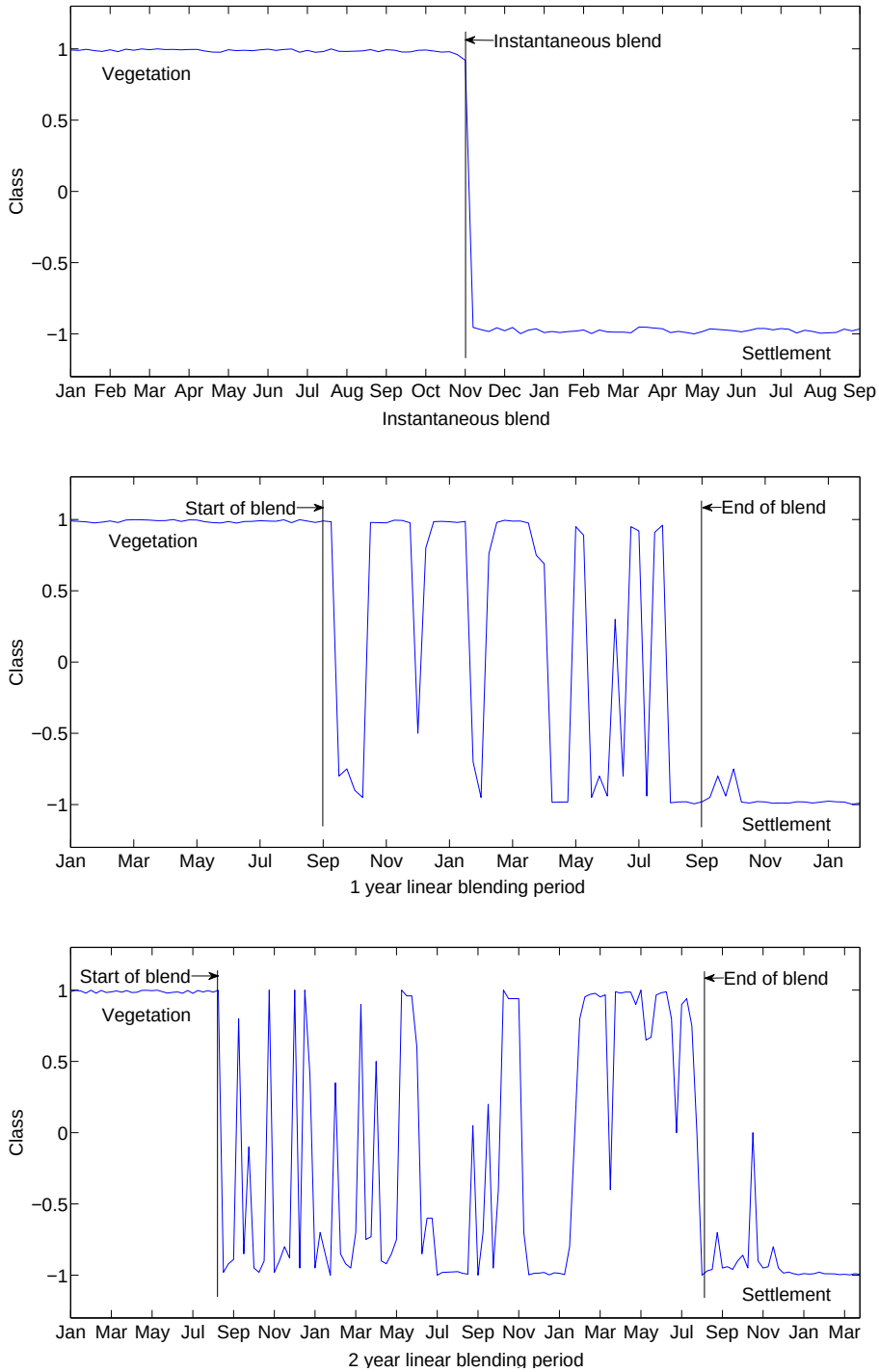


FIGURE 8.6: Class label time series for simulated land cover change from natural vegetation to human settlement. The top panel is for instantaneous simulated land cover change, the middle panel is for a land cover change over a 12 month blending period and the bottom panel is for a land cover change over a 24 month blending period.

faithfully simulate all forms of actual land cover change, but it does delay the date on which the change is declared (figure 8.6). It was concluded that only abrupt concatenation should be used when

measuring the lower limit of effective change detection Δ_τ time.

8.4 EXPERIMENTAL PLAN

In this section an overview is given of the experiments conducted in this chapter. The experiments were conducted in the Limpopo and Gauteng provinces. The number of pixels per data set in each province is given in table 8.1.

Table 8.1: Number of pixels per land cover type, per study area used for training, validation and testing data sets.

Province	Class	Number of time series
Limpopo	Vegetation - No change	1497
	Settlement - No change	1735
	Simulated land cover change	500
	Real land cover change	118
	Complete Province	590212
Gauteng	Vegetation - No change	591
	Settlement - No change	371
	Simulated land cover change	124
	Real land cover change	180
	Complete Province	78702

The experiments conducted in this chapter are grouped into four categories:

1. Parameter exploration (section 8.5),
2. Classification (section 8.6),
3. Change detection (section 8.7),
4. Provincial experiments (section 8.9).

A set of general experiments were conducted in section 8.5 to optimise the parameters which are used in the remaining sections (section 8.6 – section 8.9). The first set of experiments is used to determine the optimal network architecture for the MLP (section 8.5.1) that will minimise the generalisation error. The second set of experiments is used to explore two different training methods for the MLP (section 8.5.2): batch mode and iteratively retrained mode. The third set of experiments is used to optimise the length of the sliding window for the least squares method (section 8.5.3). The fourth set of experiments is used to compare the performance of the EKF when using the BVEP criterion (denoted by EKF_{BVEP}) and ALS methods (denoted by EKF_{ALS}, section 8.5.4). The fifth set

of experiments is used to investigate the setting of the BVEP criterion using the BVSA (section 8.5.5). The sixth set of experiments is used to investigate the performance of each of the regression methods (section 8.5.6). The seventh set of experiments is used to determine the number of clusters to use in the unsupervised classifier (section 8.5.7). The last set of experiments is used to determine the average silhouette value for different clustering algorithms (section 8.5.8).

In section 8.6, the classification accuracy is computed for each of the two classes in a range of experiments on the no change data set. In each section the average classification accuracy is reported, along with the corresponding standard deviation. Different combinations of feature extraction methods and machine learning methods are investigated in these experiments. The feature extraction methods that were explored are:

- least squares model fitting,
- M-estimator model fitting,
- SFF, and
- EKF_{BVEP}.

The classification experiments are divided into supervised classification experiments and unsupervised classification experiments. The machine learning method determines the category of the classifier. The machine learning methods that were explored are:

1. Supervised classifier:
    - Multilayer Perceptron (section 8.6.1).
2. Unsupervised classifier:
    - Hierarchical clustering, single linkage criterion (section 8.6.3),
    - Hierarchical clustering, average linkage criterion (section 8.6.3),
    - Hierarchical clustering, complete linkage criterion (section 8.6.3),
    - Hierarchical clustering, Ward clustering method (section 8.6.4),
    - Partitional clustering, K-means algorithm (section 8.6.5),
    - Partitional clustering, EM algorithm (section 8.6.6).

The objective of the classification experiments is to identify combinations of methods which have high classification accuracies and minimal corresponding standard deviations.

The change detection algorithms in this thesis are based on a post-classification approach, and are thus dependent on the classification accuracies reported in section 8.6. The classification accuracies are used to identify a set of methods that will provide acceptable change detection accuracies (section 8.7).

The first set of experiments is used to determine the change detection accuracies on the simulated land cover change data set. The number of time series blended to simulate the land cover change in each province is given in table 8.1. The true positives and false positives are reported on the simulated land cover data set in section 8.7.1.

The second set of experiments is used to determine the change detection accuracies on the real land cover change data set. The number of time series that experienced actual land cover change in the labelled data set of each province is given in table 8.1. In these experiments only the true positives are reported on the real land cover data set in section 8.7.2.

The third set of experiments is used to determine the effective change detection delay Δ_τ on the simulated land cover change data set. The number of time series blended to simulate land cover change with the exact time index known of change in each province is given in table 8.1. The effective change detection delay is reported in days in section 8.7.3.

The change detection algorithms are then applied to the complete province in section 8.9. The total number of time series in each province is given in table 8.1. The entire province is classified and areas which experienced land cover change are mapped, followed by the calculation of summary statistics.

8.5 PARAMETER EXPLORATION

8.5.1 Optimising the multilayer perceptron

The MLP comprises an input layer, one hidden layer and an output layer. All hidden and output nodes used a tangent sigmoid activation function. The input layer accepts feature 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 class 1 (natural vegetation) given the feature vector, while -1 represents a 100% certainty of class 2 (settlement).

The weights of the MLP were determined using a steepest descent gradient optimisation method in the training phase, with gradients estimated using backpropagation [130, Ch. 4 p. 140]. A validation set was used for initial MLP architecture optimisation by evaluating the generalisation error to identify overfitting of the network for each study area. The MLP architecture was optimised for different lengths of sliding window Q, number of spectral bands and training mode. In table 8.2 the number

TABLE 8.2: The number of hidden nodes used within the MLP for each experiment.

Province	Algorithm	Window length	Spectral Band		
			NDVI	2 Bands	7 Bands
Limpopo	SFF, Iteratively retrained	6 months	7	6	6
		12 months	8	10	9
		18 months	8	9	7
	SFF, Batch mode	12 months	8	10	9
	Least squares	12 months	9	8	11
	M-estimator	12 months	9	10	7
	EKF_{BVEP}	n/a	7	5	5
	EKF_{ALS}	n/a	15	13	11
Gauteng	SFF, Iteratively retrained	6 months	8	8	7
		12 months	7	7	8
		18 months	7	6	5
	SFF, Batch mode	12 months	7	7	8
	Least squares	12 months	8	10	5
	M-estimator	12 months	11	10	9
	EKF_{BVEP}	n/a	9	4	2
	EKF_{ALS}	n/a	14	6	5

of hidden nodes used in each experiment are reported. The learning rate was set to 0.01 and the momentum parameter was set to 0.9. The maximum number of epochs in each training phase was set to 10000, and used the generalisation error on the validation set as an early stopping criterion.

8.5.2 Batch mode versus iterative retrained mode

In this section the notion of an iterative retrained training mode is explored and is compared to a classical batch training mode. The change detection method extracts feature vectors sequentially from a time series using a temporal sliding window. These feature vectors must be processed to yield a class label for each feature vector.

A MLP operating on the SFFs extracted from the temporal sliding window was used to explore the difference in classification accuracies between the batch mode and iteratively retrained mode. In the batch mode [130, Ch. 7 p. 263] all the incremental sliding windows between the year 2000 and the year 2008 were used as initial training inputs to the MLP. The experiments were conducted for the 8 years without any retraining.

The iteratively retrained MLP is proposed to compensate for the inter-annual variability between years due to the rainfall variability. The iteratively retrained MLP is trained to recognise data from

Table 8.3: Classification accuracy of the batch mode and iteratively retrained MLP on the validation set. 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 in percentage for each of the classes over a temporal sliding window length of 12 months and different sets of spectral band combinations (NDVI, 2 spectral bands and all 7 spectral bands).

Province	**Spectral Band**	**Class**	**Mode**	
			Batch mode	Iteratively retrained
Limpopo	NDVI	Vegetation	67.7 ± 9.5	72.8 ± 5.3
		Settlement	83.0 ± 4.9	83.2 ± 3.7
	2 Bands	Vegetation	80.5 ± 5.6	83.1 ± 4.1
		Settlement	87.2 ± 2.0	86.8 ± 2.7
	7 Bands	Vegetation	94.5 ± 2.1	94.4 ± 1.6
		Settlement	94.8 ± 1.2	95.2 ± 1.1
Gauteng	NDVI	Vegetation	94.6 ± 4.1	96.2 ± 2.0
		Settlement	82.3 ± 8.9	88.0 ± 6.3
	2 Bands	Vegetation	96.6 ± 1.4	96.7 ± 1.6
		Settlement	92.2 ± 3.2	95.6 ± 2.3
	7 Bands	Vegetation	97.2 ± 0.4	99.8 ± 0.3
		Settlement	95.7 ± 0.4	99.3 ± 0.7

the training set within the sliding window at position p in the time series, and is then used to classify the data from the testing set within the sliding window at position p. This retraining at each time increment caused a small adaptation of the weights, and has low complexity because of the small incremental MLP weight changes over each 8 day increment of MODIS. These small MLP weight changes only required 300 epochs at each time increment for network adaptation.

The iteratively retrained mode provided slightly higher mean classification accuracies when compared to the classical batch training mode. The reason why the iteratively retrained mode performed better than the batch mode (table 8.3) is that the iteratively retrained mode had the advantage of learning the most recent spectral properties of the land cover types, as time progressed. The iteratively retrained mode takes cognisance of what is within the temporal sliding window to compensate for short-term inter-annual climate variability and adapts to longer term trends in climate without confusing any of these with a particular land cover type, which has often been a problem with other regional land cover studies [218, 219]. It should be noted that these benefits of using the iteratively retrained mode comes at the cost of having shorter predictive spans, as predicting future events will require retraining with an training data set that is unavailable. The benefits of using iteratively retrained mode resulted in it being used in the remainder of this chapter.

8.5.3 Optimising least squares

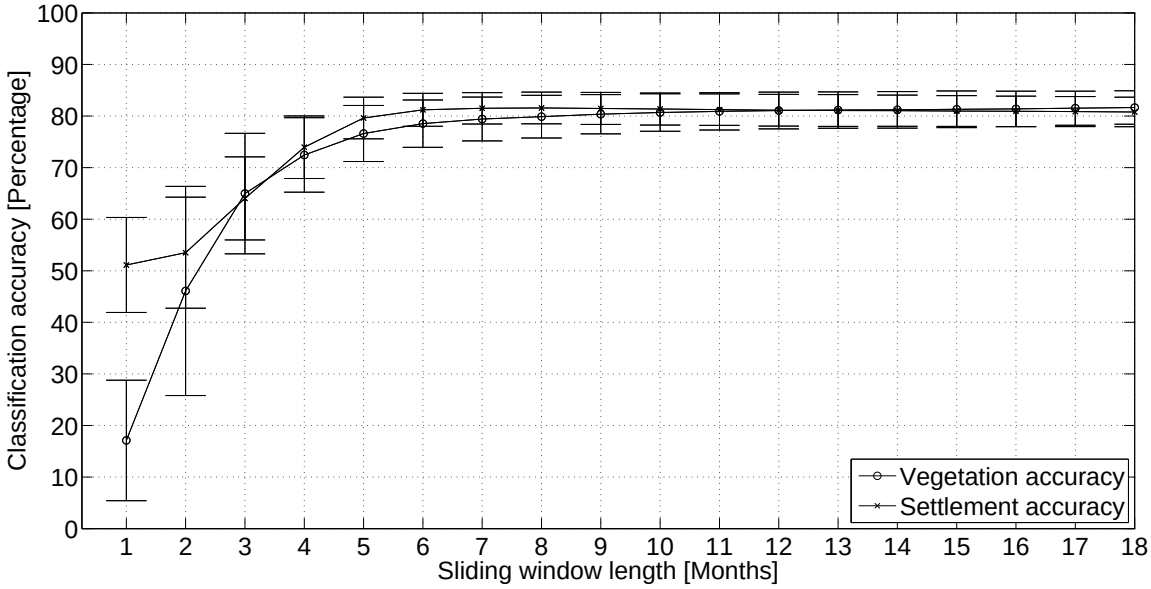
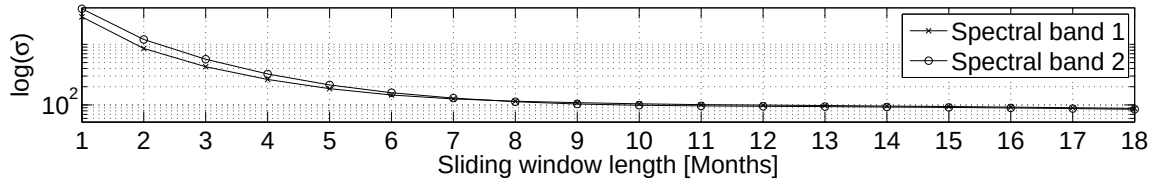


FIGURE 8.7: Classification accuracy reported by the K-means algorithm using the model fitted with a least squares model approach. The average classification accuracy is measured in percentage for each of the classes over a range of temporal sliding window length.

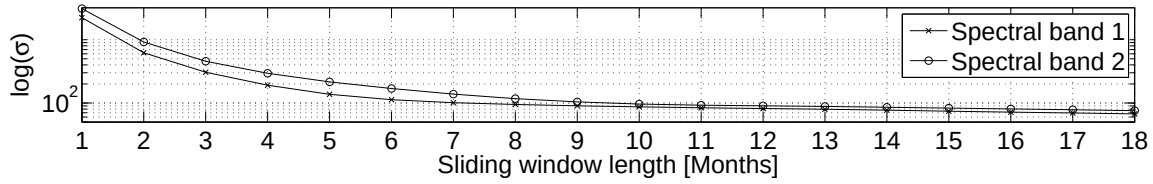
In this section an experiment was conducted to determine the optimal length of the sliding window when using the least squares approach to fit a model. The model is a triply modulated cosine model and the estimated parameters are used by a machine learning method for classification and change detection. The sliding window length was evaluated against classification accuracy, the model parameters' standard deviation and residuals of the fitted model. The classification accuracies were computed using the K-means algorithm operating on the first two spectral bands that were extracted from the Limpopo province study area. In figure 8.7, the classification accuracies are plotted as a function of the sliding window length, which is reported in the number of months.

It was observed that the settlement classification accuracy stabilised above 80% when the sliding window length surpassed the 5 month mark. The vegetation classification accuracy only stabilised above 80% after the sliding window had a length longer than 9 months. Similar classification accuracies and corresponding standard deviations were observed for both classes when the sliding window length increased beyond 11 months.

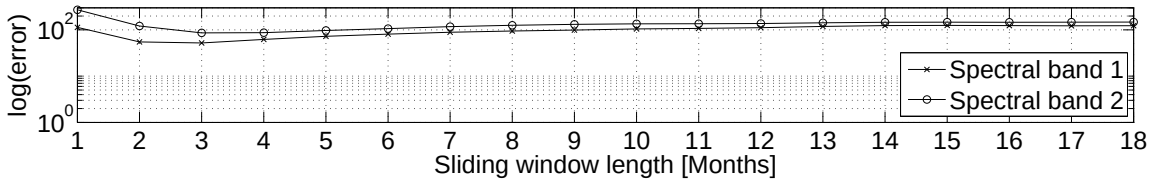
The model parameters' standard deviation for both the mean and amplitude parameters are shown in figure 8.8(a) and figure 8.8(b) respectively. It was observed that the model parameters' standard deviation for both the mean and amplitude parameters reduced as the length of the sliding window was increased. The mean parameter's standard deviation for both spectral bands started to decrease more



(a) Mean parameter



(b) Amplitude parameter



(c) Absolute error

FIGURE 8.8: The standard deviation for the mean and amplitude parameter are illustrated in (a) and (b) when using a least squares approach to fit a triply modulated cosine model to the first two spectral bands of MODIS. The absolute error between the fitted model and the actual MODIS time series is shown in (c).

slowly when the sliding window length was longer than 9 months. The amplitude parameter's standard deviation for both spectral bands started to decrease more slowly when the sliding window length was longer than 10 months.

The opposite was observed with the absolute error, which measures the difference between the fitted model and the actual MODIS time series. A shorter sliding window length had a smaller measured residuals, except if the window was too short and was severely affected by the additive noise in the MODIS time series. A sliding window of 2–3 months had the smallest measured residuals (figure 8.8(c)).

The length of the sliding window was determined based on the classification accuracies, owing to the inverse relationship between the standard deviations of the model's parameters and the absolute error. On the basis of this experiment it was decided to set the sliding window length to 12 months for all experiments using least squares to fit a model. The similarity between the results produced by the least squares and M-estimator supports the choice of a 12 month window for the M-estimator too. No significant variations in the parameter vector were found when sliding the window through the time series and using the least squares or the M-estimator.

8.5.4 BVEP versus autocovariance least squares

Table 8.4: Classification accuracy of the MLP using either the BVEP criterion or the ALS approach to fine tune the parameters of the Extended Kalman filter. 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, 2 spectral bands and all 7 spectral bands).

Province	**Spectral Band**	**Class**	**Mode**	
			EKF_{ALS}	EKF_{BVEP}
Limpopo	NDVI	Vegetation	66.6 ± 9.1	80.2 ± 4.4
		Settlement	79.2 ± 6.2	82.7 ± 3.7
	2 Bands	Vegetation	79.3 ± 2.7	87.2 ± 1.6
		Settlement	85.9 ± 2.1	89.7 ± 1.3
	7 Bands	Vegetation	86.6 ± 3.7	95.3 ± 0.7
		Settlement	90.6 ± 1.9	96.1 ± 0.6
Gauteng	NDVI	Vegetation	89.3 ± 4.8	91.4 ± 5.7
		Settlement	72.1 ± 16.9	86.9 ± 9.1
	2 Bands	Vegetation	90.6 ± 2.9	98.6 ± 1.0
		Settlement	87.6 ± 3.2	96.2 ± 1.5
	7 Bands	Vegetation	95.3 ± 1.8	99.9 ± 0.1
		Settlement	94.8 ± 2.4	99.9 ± 0.1

In this section two different methods used for setting the parameters of the EKF are investigated. The first method that is investigated is the ALS method discussed in section 7.3. The second method investigated is the BVEP criterion approach discussed in section 7.2.4.

In table 8.4, the classification accuracies for both provinces are reported when the EKF is used to extract the features. The average classification accuracy is calculated with cross-validation using 10 repeated independent experiments [127]. From these results it was concluded that the EKF_{BVEP} performed better than any experiment conducted using the EKF_{ALS}. This could be owing to the fact that the BVEP criterion utilises spatial information that is inherent in the set of time series.

8.5.5 Optimisation of Kalman filter parameters

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 space. In each epoch the algorithm attempts to minimise the standard deviation of all the state space variables while simultaneously minimising the residual between the triple modulated cosine function's output and the actual observations.

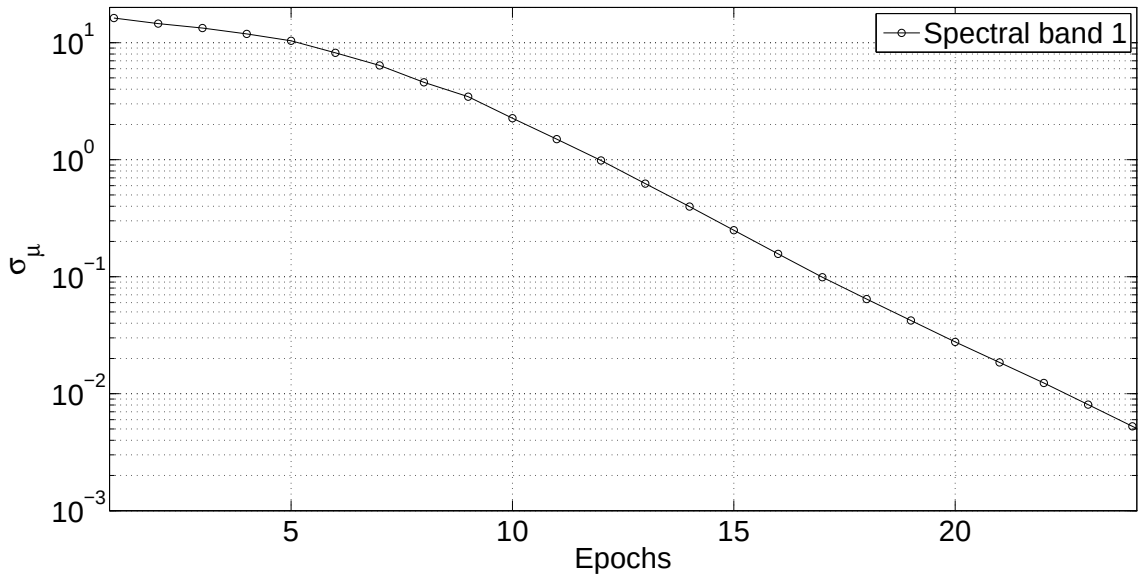


FIGURE 8.9: The expected standard deviation of the mean parameter computed for the first MODIS spectral band on the Limpopo province study area as a function of epoch.

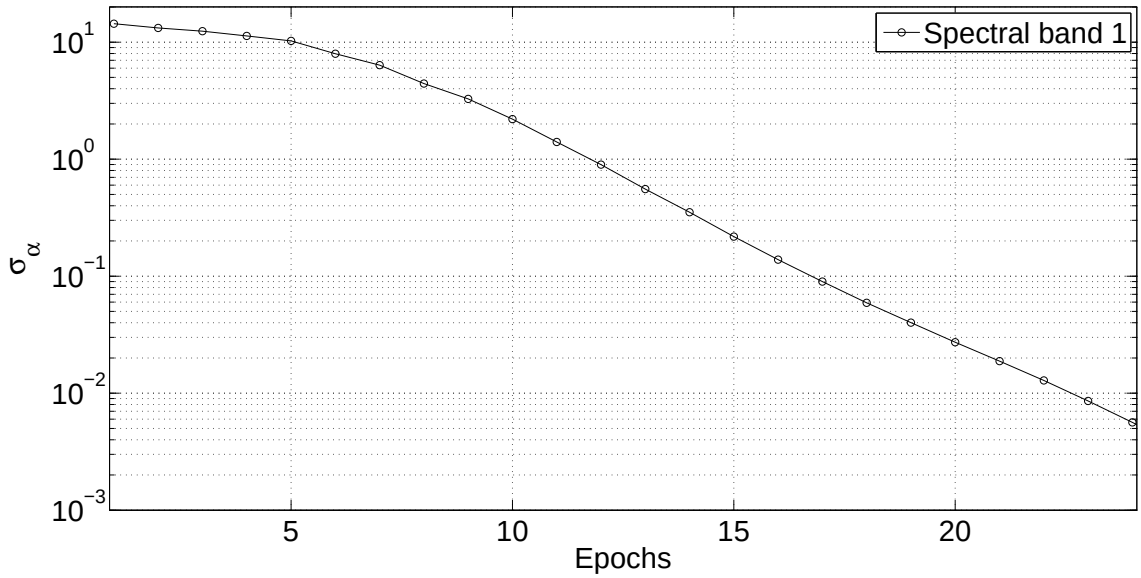


FIGURE 8.10: The expected standard deviation of the amplitude parameter computed for the first MODIS spectral band on the Limpopo province study area as a function of epoch.

In figure 8.9, the standard deviation σ_μ of the mean parameter obtained by fitting the cosine model to the first 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 Limpopo 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 first MODIS spectral band is illustrated as a function of epoch of the BVSA in figure 8.10. The standard deviation reported here is the average standard deviation found over all the time series extracted from the Limpopo 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.

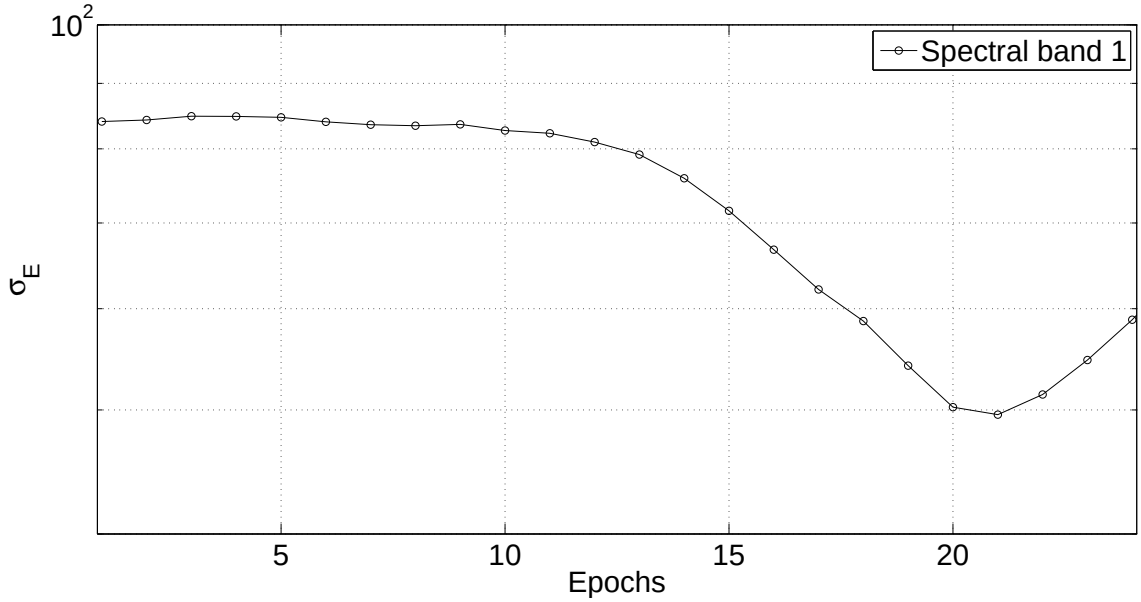


FIGURE 8.11: The expected residuals computed for the first MODIS spectral band on the Limpopo province study area as a function epoch.

In figure 8.11, the mean residual $\sigma_\mathcal{E}$ 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 decreases significantly after the 10^{th} 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 $\sigma_\mathcal{E}$ metric after the 21^{st} epoch. This overfit defines the end of the search and is used as an early stopping criterion.

Table 8.5: Parameter evaluation of two different search methods that were compared in the Limpopo province study area.

Algorithm	**Parameter evaluation**		
	σ_μ	σ_α	$\sigma_\mathcal{E}$
Simulated Annealing	14.5	12.6	94.6
BVSA	0.04	0.02	87.1

The process covariance matrix $\mathcal{Q}$ and observation covariance matrix $\mathcal{R}$ used in the 21^{st} epoch are then used to initialise the EKF for the experiments. The BVSA is applied independently to each of the

seven spectral bands and NDVI time series to obtain a process covariance matrix $\mathcal{Q}$ and observation covariance matrix $\mathcal{R}$ for each spectral band.

Table 8.6: Parameters evaluation of all four methods for the Limpopo province study area. The measurements are made on all seven MODIS spectral bands and NDVI.

Province	**Spectral Band**		**Mode**			
			Least squares	M-estimator	EKF_{ALS}	EKF_{BVEP}
Limpopo	NDVI	$\sigma_{\mathcal{E}}$	0.04	0.04	0.001	0.03
		σ_{μ}	0.02	0.01	0.04	0.02
		σ_{α}	0.02	0.02	0.05	0.001
	Band 1	$\sigma_{\mathcal{E}}$	118.6	118.7	144.0	87.1
		σ_{μ}	28.8	28.1	29.8	0.04
		σ_{α}	36.4	36.1	21.8	0.02
	Band 2	$\sigma_{\mathcal{E}}$	145.2	144.7	179.9	95.7
		σ_{μ}	38.5	37.4	29.6	0.01
		σ_{α}	56.4	57.6	25.2	0.36
	Band 3	$\sigma_{\mathcal{E}}$	58.1	58.0	62.3	47.9
		σ_{μ}	13.6	13.1	20.9	0.06
		σ_{α}	18.9	18.3	14.7	0.05
	Band 4	$\sigma_{\mathcal{E}}$	65.6	65.6	81.0	58.3
		σ_{μ}	14.2	14.1	25.5	0.05
		σ_{α}	19.7	20.8	18.0	0.04
	Band 5	$\sigma_{\mathcal{E}}$	154.6	154.3	171.1	97.3
		σ_{μ}	36.7	36.2	29.6	0.01
		σ_{α}	48.6	49.1	24.9	0.01
	Band 6	$\sigma_{\mathcal{E}}$	198.5	198.4	242.4	166.9
		σ_{μ}	46.6	45.8	33.8	0.01
		σ_{α}	67.8	68.1	27.3	0.01
	Band 7	$\sigma_{\mathcal{E}}$	232.1	232.0	302.0	201.1
		σ_{μ}	79.3	76.5	31.3	0.02
		σ_{α}	77.9	76.4	26.1	0.03

It should be noted that other optimisation algorithms were also explored, based on the objective function defined in the BVEP criterion (equation (7.50)) to evaluate the performance of the BVSA. The algorithms used to set the BVS are: (1) the interior point method [220], (2) active set method [221], and (3) simulating annealing [222]. It is observed from the active set method that larger and more aggressive step sizes are required, which supports the BVSA described on page 135. Simulated annealing (500 epochs, 5 function evaluations per epoch) produced better results than either the active set method or the interior point method. Table 8.5 compares simulated annealing to BVSA.

By evaluating the propagation direction of the simulating annealing method, it was concluded that

the method would eventually find the same solution identified by the BVSA, and yield the exact same performance. The advantage of the BVSA was the speed of convergence, which is attributed to the fact that it only requires a single function evaluation per epoch and converged in 21 epochs in this experiment.

8.5.6 BVSA parameter evaluation

Table 8.7: Parameters evaluation of all four methods for the Gauteng province study area. The measurements are made on all seven MODIS spectral bands and NDVI.

Province	**Spectral Band**		**Mode**			
			Least squares	M-estimator	EKF_{ALS}	EKF_{BVEP}
Gauteng	NDVI	$\sigma_{\mathcal{E}}$	0.04	0.04	0.001	0.003
		σ_{μ}	0.01	0.01	0.07	0.05
		σ_{α}	0.009	0.01	0.06	0.01
	Band 1	$\sigma_{\mathcal{E}}$	96.6	96.6	90.8	44.8
		σ_{μ}	17.7	17.4	21.3	0.01
		σ_{α}	22.5	22.2	17.3	15.3
	Band 2	$\sigma_{\mathcal{E}}$	156.4	155.9	204.2	123.4
		σ_{μ}	49.1	47.2	29.8	0.01
		σ_{α}	54.9	55.3	25.5	0.5
	Band 3	$\sigma_{\mathcal{E}}$	55.1	55.1	46.7	38.5
		σ_{μ}	10.2	9.8	14.9	0.03
		σ_{α}	14.0	13.5	12.2	0.02
	Band 4	$\sigma_{\mathcal{E}}$	63.3	63.3	57.0	42.7
		σ_{μ}	12.6	12.6	19.2	0.04
		σ_{α}	14.7	15.4	14.5	0.03
	Band 5	$\sigma_{\mathcal{E}}$	153.2	153.0	162.9	105.3
		σ_{μ}	47.4	46.2	26.6	0.01
		σ_{α}	54.2	53.8	22.6	0.01
	Band 6	$\sigma_{\mathcal{E}}$	157.3	157.4	130.5	87.3
		σ_{μ}	29.8	30.0	24.9	0.01
		σ_{α}	34.8	36.6	22.2	0.01
	Band 7	$\sigma_{\mathcal{E}}$	158.0	157.8	151.9	71.9
		σ_{μ}	27.8	27.0	23.0	0.02
		σ_{α}	35.0	34.3	21.7	20.5

In this section the derived parameters for each regression method are compared along with the residuals. 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 residuals $\sigma_{\mathcal{E}}$. A mean (amplitude) parameter with a small standard deviation indicates a stable variable. A small $\sigma_{\mathcal{E}}$ indicates a well-estimated output when

compared to the actual observations.

An analysis of the standard deviation of the parameters extracted from the Limpopo province data is presented in table 8.6. It was observed that the M-estimator generally performs similarly to least squares, and in some cases performed slightly better. The EKF_{ALS} method generally increased the residuals to improve the parameter stability when compared to the M-estimator. The EKF_{BVEP} outperformed all the methods in all the experiments, except for the NDVI experiments. The EKF_{BVEP} however did yield comparable results to the other methods in the NDVI experiments.

In table 8.7, the same comparison was made as in table 8.6 for the Gauteng province study area. The M-estimator again performed similar to the least squares and in a few experiments performed slightly better. The relation between the EKF_{ALS} method and M-estimator did not hold in the Gauteng province study area. The EKF_{ALS} method increased its residuals in spectral bands 2 and 5 to improve the parameters' stability when compared to the M-estimator. 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} outperformed the M-estimator in all the metrics. In the NDVI case the EKF_{ALS} decreased its residuals at the cost of parameter stability when compared to the M-estimator.

The EKF_{BVEP} outperformed all methods in all the experiments, except for the NDVI experiments. A peculiar observation was made for the EKF_{BVEP} in spectral bands 1 and 7. For the first spectral band case overfitting was observed in the amplitude parameter early in the BVSA, which is used as an 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 an overfit was reported on the residuals $\sigma_\mathcal{E}$ at the 22^{nd} epoch. If the overfit did not occur, the standard deviation σ_α of the amplitude parameter would still steadily decrease. In the remainder of the chapter only the optimised EKF using the BVEP criterion (EKF_{BVEP}) will be considered and will be referred to as the EKF method.

8.5.7 Determining the number of clusters

Determining the number of clusters is one of the most difficult design considerations. The number of clusters K must be determined that provides maximum compression of information in the feature vectors with minimal error in classification on the data set.

The average silhouette value $\mathcal{S}_{\text{ave}}$ (equation (4.31) on page 82) is the metric used to determine the number of clusters. The nature of selecting only natural vegetation and human settlement areas in the labelled time series data set, and the resolution of the MODIS sensor, suggested a strong tendency of $\mathcal{S}_{\text{ave}}$ to have a high value at lower values of K. This is due to the fact that the labelled data set contains two distinct classes. At 500 metre resolution, the MODIS pixels are quite large, and are therefore

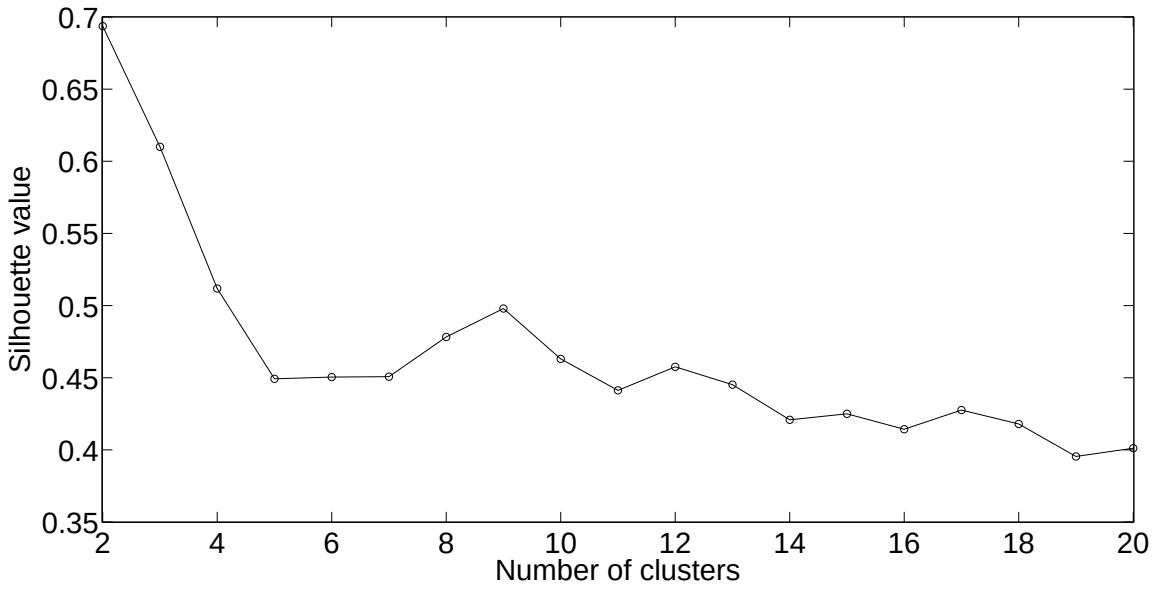


FIGURE 8.12: The average silhouette value $\mathcal{S}_{\text{ave}}$ computed over a range of different number of clusters in the Gauteng province.

likely to contain a mixture of different vegetation types. Nevertheless, it is reasonable to assume that the variability within the broader vegetation class will be large enough to justify splitting the vegetation class into subclasses. This however was not the case in the labelled data sets in this study.

In figure 8.12, an experiment was performed to compute the average silhouette value $\mathcal{S}_{\text{ave}}$ for a range of K. The experiment was conducted in Gauteng province using the EKF on the first two spectral bands. The feature vectors were then clustered using the K-means algorithm, followed by the computing of the silhouette values. The highest average silhouette value of 0.69 was recorded at two classes and steadily decreased as K increased. The experiment was repeated for all the other clustering methods, with K=2 producing the highest silhouette value in all the cases. The same experiments were conducted in the Limpopo province study area and yielded similar results.

8.5.8 Results: Cophenetic correlation coefficient

In this section the cophenetic correlation coefficient D_{cc} was computed for a range of hierarchical clustering methods: single linkage criterion (section 8.6.3), average linkage criterion (section 8.6.3), complete linkage criterion (section 8.6.3) and Ward clustering (section 8.6.4).

The cophenetic correlation coefficient evaluates how the created dendrogram retains the original placement of the feature vectors within the feature space. A high cophenetic correlation coefficient, $D_{\text{cc}} \rightarrow 1$, denotes that the distance representation is well preserved in the dendrogram. The cophenetic correlation coefficient was computed in the Limpopo province for a range of experimental

Table 8.8: The Cophenetic correlation coefficient computed for a range of hierarchical clustering methods on the Limpopo province's no change data set.

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
Single linkage criterion	SFF	6 months	0.50	0.31	0.33
		12 months	0.51	0.32	0.33
		18 months	0.52	0.32	0.33
	Least squares	12 months	0.49	0.32	0.38
	M-estimator	12 months	0.49	0.32	0.39
	EKF	n/a	0.46	0.28	0.29
Average linkage criterion	SFF	6 months	0.59	0.64	0.61
		12 months	0.59	0.65	0.61
		18 months	0.59	0.65	0.62
	Least squares	12 months	0.60	0.62	0.61
	M-estimator	12 months	0.60	0.62	0.60
	EKF	n/a	0.59	0.62	0.59
Complete linkage criterion	SFF	6 months	0.64	0.64	0.62
		12 months	0.64	0.65	0.63
		18 months	0.64	0.66	0.63
	Least squares	12 months	0.60	0.61	0.62
	M-estimator	12 months	0.60	0.62	0.62
	EKF	n/a	0.62	0.63	0.64
Ward clustering	SFF	6 months	0.69	0.71	0.68
		12 months	0.69	0.72	0.68
		18 months	0.70	0.72	0.69
	Least squares	12 months	0.67	0.73	0.69
	M-estimator	12 months	0.67	0.73	0.69
	EKF	n/a	0.68	0.74	0.69

parameters (table 8.8): hierarchical clustering methods, feature extraction methods, and spectral band combinations.

A small improvement in the cophenetic correlation coefficient is observed when the sliding window length is increased. It is concluded that the cophenetic correlation coefficient is highly dependent on the clustering method used, as all feature extraction methods performed similarly when using a particular clustering method.

The single linkage criterion provided the lowest cophenetic correlation coefficients among the clustering methods. The average linkage criterion provided much better cophenetic correlation coefficients than the experiments using the single linkage criterion. A small improvement is observed

in the NDVI experiments when the complete linkage criterion is compared to the average linkage criterion. Similar results were observed for the average and complete linkage criteria in the two and seven spectral band experiments. A small improvement was observed in all the experiments when Ward clustering was used instead of the complete linkage criterion.

The same trend in cophenetic correlation coefficients was observed in the Gauteng province when all the experiments were compared to the results produced in the Limpopo province. The cophenetic correlation coefficient confirms the trend, which is observed in classification accuracies through sections 8.6.3–8.6.4. This is an important experiment, as this result was derived in an unsupervised manner, meaning the class labels for each time series were not used in the cluster process. It was concluded from the experiments conducted in this section that creating spherical clusters with minimum internal variance preserves the inherent distance between feature vectors within the feature space, which results in a higher cophenetic correlation coefficient.

8.6 CLASSIFICATION

8.6.1 Classification accuracy: Multilayer perceptron

Table 8.9: Classification accuracy of the MLP using SSFs on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	Spectral Band	Class	Sliding window length		
			6 months	12 months	18 months
Limpopo	NDVI	Vegetation	69.7 ± 7.8	72.8 ± 5.3	73.9 ± 4.8
		Settlement	81.5 ± 5.0	83.2 ± 3.7	84.8 ± 3.1
	2 Bands	Vegetation	81.4 ± 4.3	83.1 ± 4.1	85.2 ± 3.7
		Settlement	86.3 ± 3.4	86.8 ± 2.7	88.1 ± 2.2
	7 Bands	Vegetation	93.1 ± 2.1	94.4 ± 1.6	94.7 ± 1.4
		Settlement	93.8 ± 1.6	95.2 ± 1.1	96.3 ± 0.9
Gauteng	NDVI	Vegetation	94.4 ± 3.7	96.2 ± 2.0	95.8 ± 2.2
		Settlement	79.5 ± 11.5	88.0 ± 6.3	88.5 ± 7.2
	2 Bands	Vegetation	95.1 ± 2.8	96.7 ± 1.6	97.2 ± 1.9
		Settlement	90.7 ± 6.7	95.6 ± 2.3	95.8 ± 2.5
	7 Bands	Vegetation	99.3 ± 0.7	99.8 ± 0.3	99.8 ± 0.3
		Settlement	98.1 ± 1.4	99.3 ± 0.7	99.6 ± 0.6

In this section the classification accuracies are evaluated for a MLP using a range of feature extraction methods. In table 8.9, the classification accuracies for both provinces are reported using SFFs. The average classification accuracy and corresponding standard deviation were calculated with

Table 8.10: Classification accuracy of the MLP using regression methods to extract features on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Method**		
			Least squares	M-estimator	EKF
Limpopo	NDVI	Vegetation	72.5 ± 5.3	72.8 ± 5.4	80.2 ± 4.4
		Settlement	83.3 ± 3.4	84.6 ± 3.4	82.7 ± 3.7
	2 Bands	Vegetation	82.2 ± 4.3	83.1 ± 4.3	87.2 ± 1.6
		Settlement	86.4 ± 2.8	87.7 ± 2.5	89.7 ± 1.3
	7 Bands	Vegetation	92.5 ± 2.3	92.5 ± 1.9	95.3 ± 0.7
		Settlement	92.6 ± 1.2	92.4 ± 1.4	96.1 ± 0.6
Gauteng	NDVI	Vegetation	92.5 ± 4.9	93.1 ± 4.4	91.4 ± 5.7
		Settlement	88.6 ± 6.4	88.8 ± 6.0	86.9 ± 9.1
	2 Bands	Vegetation	97.5 ± 1.8	97.3 ± 1.9	98.6 ± 1.0
		Settlement	95.1 ± 2.6	94.9 ± 2.9	96.2 ± 1.5
	7 Bands	Vegetation	99.8 ± 0.4	99.9 ± 0.4	99.9 ± 0.1
		Settlement	99.2 ± 0.5	99.3 ± 0.9	99.9 ± 0.1

cross-validation using 10 repeated independent experiments. The accuracy is reported for each class over a range of temporal sliding window lengths (6, 12 and 18 months) and different spectral band combinations (NDVI, 2 spectral bands and all 7 spectral bands).

It is observed that a longer sliding window has a higher classification accuracy in all the experiments, as well as a reduction in standard deviations. Overall, the trend was that the classification performance improved for a longer sliding window. Another trend that was observed was an increase in overall performance when more spectral bands were used as input to a MLP classifier. This is supported by a higher classification accuracy for the first two spectral bands when compared to the NDVI, and the highest classification accuracy was reported for all seven spectral bands.

In table 8.10, the classification accuracies for both provinces are reported using regression methods to extract the features. The regression methods attempted to fit a triply modulated cosine function to the MODIS time series. The sliding window length was set to 12 months for both the least squares and M-estimator approaches. A similar improvement is observed as in table 8.9 when more spectral bands are used in the experiments.

From all the experiments it was concluded that a significant improvement is obtained when using the first two spectral bands rather than the NDVI. A further improvement was observed when the MLP operated on all seven spectral bands. The experiments conducted in the section are repeated in the following sections using different clustering algorithms.

8.6.2 Clustering experimental setup

In the following sections (section 8.6.3–8.6.4), different clustering approaches are analysed in a range of experiments. The first set of experiments conducted in each section is the measurement of the classification accuracy of the labelled time series using SFFs. The experiments were conducted for three different lengths of sliding window: 6 months (23 MODIS samples), 12 months (46 MODIS samples), and 18 months (69 MODIS samples). The experiments also explore the use of different spectral bands: NDVI, the first two spectral bands, and all seven spectral bands. In each experiment the classification accuracy along with the standard deviation is reported for the two classes: natural vegetation and human settlement.

The class labels in the experiments are assigned to minimise the overall error. This is accomplished in the Limpopo province by assigning the cluster containing majority of the feature vectors to the settlement class, as there are more settlement class time series than vegetation class time series (table 8.1). In the experiments conducted in the Gauteng province, the cluster containing majority of the feature vectors is assigned to the vegetation class, as there are more vegetation class time series than settlement class time series (table 8.1).

The second set of experiments conducted in each section is the measurement of classification accuracies of the labelled time series using different regression methods to extract features. The experiment is conducted on three different regression methods: least squares model fitting, M-estimator model fitting, and EKF. The experiments were also conducted to explore the use of different spectral bands in the similar method as in the first set of experiments. In each experiment the classification accuracy along with the standard deviation is reported for the two classes. The class labels are again assigned to minimised the overall error.

8.6.3 Clustering accuracy: Single, Average and Complete linkage criterion

In this section the viability of using hierarchical clustering based on the single, average and complete linkage criteria are investigated. Table 8.11 shows the classification accuracy on the experiments conducted using the SFFs, which were clustered based on the single, average and complete linkage criteria.

It is clear from the experiments that the first two spectral band outperforms NDVI.The first two spectral band also offered a slight improvement over the all seven spectral band. It is important to note that the all seven spectral band feature vector already encapsulate the first two spectral band. The reason for the decrease in classification accuracy is attributed to the fact that the seven spectral band feature vector requires more clusters (number of clusters K must increase) to cater for the increase in feature dimensionality. It was observed in an independent experiment that the classification accuracy

Table 8.11: Classification accuracy of a hierarchical clustering algorithm using the single, average and complete linkage criteria with the SFFs on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation for a sliding window length of 12 months.

Province	**Spectral Band**	**Class**	**Sliding window length**		
			Single linkage	Average linkage	Complete linkage
Limpopo	NDVI	Vegetation	45.8 ± 26.7	46.2 ± 25.7	52.1 ± 28.8
		Settlement	70.3 ± 21.1	71.0 ± 18.9	67.1 ± 21.9
	2 Bands	Vegetation	72.1 ± 16.7	76.4 ± 17.6	78.8 ± 15.9
		Settlement	80.0 ± 10.1	83.5 ± 9.5	85.7 ± 11.3
	7 Bands	Vegetation	71.4 ± 17.0	76.5 ± 25.2	75.5 ± 19.1
		Settlement	77.5 ± 9.9	83.0 ± 12.8	80.6 ± 24.0
Gauteng	NDVI	Vegetation	60.9 ± 18.2	65.3 ± 11.2	64.8 ± 9.9
		Settlement	36.9 ± 25.4	40.8 ± 21.8	42.1 ± 20.0
	2 Bands	Vegetation	80.1 ± 16.1	82.8 ± 14.8	81.6 ± 11.7
		Settlement	66.4 ± 35.1	67.0 ± 33.8	69.2 ± 29.4
	7 Bands	Vegetation	79.2 ± 16.3	80.2 ± 15.1	80.5 ± 12.2
		Settlement	64.4 ± 34.2	64.8 ± 34.1	65.9 ± 30.1

rapidly improves for the seven spectral band case if K is larger than 10. The number of clusters was not increased as the objective of the use of the unsupervised classifier is to evaluate a completely unsupervised change detection method. A supervised algorithm must then be applied onto the clusters if more clusters are included.

The first two spectral band experiments offered acceptable performance in both provinces. It should be noted that these classification accuracies could only be obtained with these three hierarchical clustering methods when performing proper outlier removal. The outliers were identified by applying principle component analysis to the feature vectors and calculating the Hotellier T^2 distance between the principal components and each of the transformed feature vectors. The outliers were then selected with distances exceeding a predefined threshold. The other clustering methods did not require the removal of outliers and for this reason the single linkage, average linkage and complete linkage criteria will not be further evaluated in this chapter.

8.6.4 Clustering accuracy: Ward clustering method

In this section the viability of using the Ward clustering method is investigated. Table 8.12 and table 8.13 show the results for the experiments that were produced using the Ward clustering method.

The Ward clustering method provided no acceptable classification accuracies when clustering on the NDVI time series. The Ward clustering method did however provide reasonable classification

Table 8.12: Classification accuracy of the Ward clustering method using the SFFs on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Sliding window length**		
			6 months	12 months	18 months
Limpopo	NDVI	Vegetation	45.3 ± 19.4	45.4 ± 17.5	46.3 ± 17.2
		Settlement	64.6 ± 12.8	66.3 ± 11.9	66.6 ± 11.7
	2 Bands	Vegetation	79.0 ± 14.2	80.9 ± 13.8	81.7 ± 13.4
		Settlement	78.2 ± 11.1	77.5 ± 10.2	77.3 ± 10.3
	7 Bands	Vegetation	72.4 ± 16.5	73.8 ± 15.6	73.8 ± 15.8
		Settlement	73.6 ± 11.9	74.5 ± 11.5	74.7 ± 11.1
Gauteng	NDVI	Vegetation	66.4 ± 10.8	67.4 ± 8.8	67.5 ± 8.7
		Settlement	35.2 ± 28.9	38.7 ± 28.6	38.9 ± 29.0
	2 Bands	Vegetation	81.3 ± 14.5	86.8 ± 13.1	86.8 ± 12.7
		Settlement	68.0 ± 31.9	69.8 ± 31.8	69.9 ± 32.0
	7 Bands	Vegetation	77.4 ± 15.6	78.2 ± 17.8	76.3 ± 18.3
		Settlement	24.5 ± 19.0	26.2 ± 18.7	27.9 ± 23.1

Table 8.13: Classification accuracy of Ward clustering with the regression methods to extract features on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Method**		
			Least squares	M-estimator	EKF
Limpopo	NDVI	Vegetation	68.0 ± 16.4	68.8 ± 15.7	66.3 ± 16.5
		Settlement	78.8 ± 13.4	78.5 ± 13.4	77.5 ± 13.4
	2 Bands	Vegetation	79.9 ± 15.1	80.0 ± 15.0	85.7 ± 12.3
		Settlement	76.9 ± 11.1	76.9 ± 11.1	77.7 ± 10.9
	7 Bands	Vegetation	72.8 ± 17.5	72.8 ± 17.6	74.1 ± 14.9
		Settlement	72.8 ± 14.3	72.8 ± 14.2	75.4 ± 9.3
Gauteng	NDVI	Vegetation	94.6 ± 10.8	94.7 ± 10.9	85.1 ± 12.1
		Settlement	27.9 ± 12.5	28.1 ± 12.9	36.9 ± 23.3
	2 Bands	Vegetation	84.5 ± 14.5	84.5 ± 14.5	88.7 ± 10.2
		Settlement	68.6 ± 32.1	68.8 ± 32.0	87.9 ± 14.3
	7 Bands	Vegetation	79.6 ± 17.3	79.6 ± 17.4	78.8 ± 18.0
		Settlement	27.5 ± 22.7	27.4 ± 22.6	44.0 ± 25.2

accuracies when the first two spectral bands and the all seven spectral bands were used in the Limpopo province. Classification accuracies of above 75% were reported for the first two spectral band experiments. The EKF features using the first two spectral bands yielded classification accuracies higher than 87.9% in the Gauteng province when compared to all the other regression methods, which

yielded classification accuracies below 70%.

In the seven spectral bands experiments an interesting trend was observed in all the hierarchical clustering experiments. The classification accuracies were lower in higher dimensions (7 spectral bands) than in lower dimensions (2 spectral bands). The question that was raised was whether the feature vectors became more separable in higher dimensions. The answer was confirmed with the MLP in section 8.6.1, where the MLP reported higher classification accuracies in the seven spectral band experiments when compared to the two spectral band experiments.

This reverts back to the statement made in section 4.2.2 on page 70 that clustering in a high-dimensional feature space usually provides meaningless results if proper design considerations are not followed [197, 198]. This is usually attributed to the notion that the ratio between the nearest neighbour and average neighbourhood distance rapidly converges to one in higher dimensions.

The remedy for this reduction in classification accuracy in the seven spectral band experiments is the implementation of a more complex clustering algorithm or a more in-depth feature selection criterion. The complex clustering algorithm will create non-linear mappings as with the MLP to obtain the desired classification accuracies. The shortcoming is the need to over design the clustering algorithm for a particular data set. Feature selection is the other approach that can be used to improve clustering performance, as it is used as a dimensionality reduction procedure, which uses fewer spectral bands to improve the performance. The problem is that different combinations of spectral bands will perform better on different data sets.

Based on the impossibility theorem, the emphasis is placed on obtaining acceptable performance in the clustering algorithm. As stated previously, the Ward clustering method does provide acceptable classification accuracies when using the first two spectral bands.

8.6.5 Clustering accuracy: K-means clustering

In this section the viability of using the K-means partitional clustering method is investigated. Table 8.14 and table 8.15 illustrate the classification accuracies for the experiments conducted with the K-means clustering algorithm.

The clustering of the NDVI time series using K-means provided acceptable classification accuracies when the regression method was used in the Limpopo province (table 8.15). This however was not the case in the Gauteng province, from which it can be concluded that the performance of clustering NDVI time series with K-means was unacceptable as it is only usable in the Limpopo province.

The first two spectral band experiments provided better classification accuracy performance when compared to any similar hierarchical clustering method. The EKF approach was deemed the best

Table 8.14: Classification accuracy of K-means with the SFFs on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Sliding window length**		
			6 months	12 months	18 months
Limpopo	NDVI	Vegetation	53.2 ± 12.8	54.4 ± 8.3	54.8 ± 9.2
		Settlement	58.7 ± 7.1	59.9 ± 5.3	59.7 ± 7.3
	2 Bands	Vegetation	81.7 ± 4.7	82.9 ± 3.7	83.4 ± 3.5
		Settlement	81.4 ± 2.2	82.0 ± 2.4	81.8 ± 2.2
	7 Bands	Vegetation	75.8 ± 5.0	76.2 ± 4.6	76.3 ± 4.3
		Settlement	74.9 ± 2.8	75.2 ± 2.3	75.2 ± 2.1
Gauteng	NDVI	Vegetation	61.3 ± 8.0	63.1 ± 5.3	65.5 ± 6.7
		Settlement	42.3 ± 28.3	39.8 ± 30.2	38.9 ± 29.9
	2 Bands	Vegetation	85.1 ± 9.1	90.0 ± 7.3	90.4 ± 7.2
		Settlement	72.6 ± 19.4	70.9 ± 21.3	71.2 ± 21.7
	7 Bands	Vegetation	76.5 ± 13.2	77.3 ± 13.1	77.3 ± 13.4
		Settlement	38.7 ± 7.6	41.2 ± 6.8	41.6 ± 6.3

Table 8.15: Classification accuracy of K-means with the regression methods to extract features on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Method**		
			Least squares	M-estimator	EKF
Limpopo	NDVI	Vegetation	69.9 ± 5.7	71.4 ± 5.7	70.5 ± 6.8
		Settlement	79.3 ± 3.5	81.2 ± 3.4	79.1 ± 4.7
	2 Bands	Vegetation	81.5 ± 3.5	81.5 ± 3.6	84.4 ± 0.2
		Settlement	80.7 ± 3.1	80.6 ± 3.0	82.3 ± 0.2
	7 Bands	Vegetation	76.7 ± 3.8	76.7 ± 3.7	76.3 ± 0.2
		Settlement	74.3 ± 2.8	74.5 ± 2.7	75.1 ± 0.1
Gauteng	NDVI	Vegetation	94.4 ± 5.2	94.4 ± 5.2	68.3 ± 14.2
		Settlement	29.2 ± 2.7	29.3 ± 2.6	39.9 ± 32.2
	2 Bands	Vegetation	87.2 ± 7.6	87.2 ± 7.6	92.3 ± 0.4
		Settlement	73.9 ± 20.1	73.9 ± 20.2	84.7 ± 2.2
	7 Bands	Vegetation	75.9 ± 12.5	76.0 ± 12.4	75.9 ± 1.9
		Settlement	24.5 ± 6.6	24.5 ± 6.6	33.2 ± 0.7

performing feature extraction method in view of the small standard deviation in classification accuracy.

A similar observation was made for the partitional clustering as for the hierarchical clustering when clustering in higher dimensions. A small decrease of 6% was measured in classification accuracy when the first two spectral band experiments were compared to the all seven spectral band experiments in

the Limpopo province. A large decrease of over 30% was measured in classification accuracy when comparing the same experiments in the Gauteng province. This suggested that the same approach as described in section 8.6.4 must be followed.

8.6.6 Clustering accuracy: Expectation-Maximisation

In this section the viability of using the EM clustering algorithm is investigated. Table 8.16 and table 8.17 illustrate the results for the experiments conducted with the EM clustering algorithm. It was concluded from the experiments that the K-means clustering algorithm and EM clustering algorithm perform similarly, as the experimental results were almost exactly the same.

Table 8.16: Classification accuracy of EM algorithm with the SFFs on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Sliding window length**		
			6 months	12 months	18 months
Limpopo	NDVI	Vegetation	51.3 ± 12.8	52.4 ± 8.5	52.9 ± 11.7
		Settlement	58.7 ± 7.1	58.8 ± 6.5	57.7 ± 7.3
	2 Bands	Vegetation	80.7 ± 4.6	81.9 ± 3.7	81.4 ± 3.6
		Settlement	81.4 ± 2.2	81.1 ± 2.2	80.6 ± 2.1
	7 Bands	Vegetation	75.8 ± 5.0	76.3 ± 4.5	76.3 ± 4.3
		Settlement	75.0 ± 2.9	75.2 ± 2.3	75.2 ± 2.1
Gauteng	NDVI	Vegetation	61.3 ± 8.0	63.1 ± 5.3	65.5 ± 6.7
		Settlement	42.3 ± 28.3	39.8 ± 30.2	39.0 ± 29.9
	2 Bands	Vegetation	85.1 ± 9.1	90.0 ± 7.4	90.4 ± 7.2
		Settlement	72.6 ± 19.4	70.9 ± 21.1	71.2 ± 21.7
	7 Bands	Vegetation	76.5 ± 13.2	77.3 ± 13.2	77.3 ± 13.4
		Settlement	38.7 ± 7.6	41.2 ± 6.8	41.6 ± 6.3

The EM clustering algorithm did however have a slightly lower classification accuracy at a negligible increase in standard deviation in a few of the experiments. For this reason the K-means clustering algorithm was chosen for its lower computational complexity.

8.6.7 Summary of classification results

In this section the results of the classification accuracies for section 8.6 are summarised. The first classifier that was considered in this section was the supervised MLP, which had the advantage of modelling a non-linear relationship between the input and output vectors.

The prospect of detecting land cover change 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

Table 8.17: Classification accuracy of EM algorithm with the regression methods to extract features on the no change data set. Each entry gives the average classification accuracy in percentage along with the corresponding standard deviation.

Province	**Spectral Band**	**Class**	**Method**		
			Least squares	M-estimator	EKF
Limpopo	NDVI	Vegetation	69.9 ± 5.9	71.3 ± 5.7	69.5 ± 6.9
		Settlement	79.3 ± 3.5	81.3 ± 3.4	79.0 ± 4.7
	2 Bands	Vegetation	81.5 ± 3.5	81.5 ± 3.5	84.3 ± 0.2
		Settlement	80.7 ± 3.1	80.6 ± 3.1	81.3 ± 0.2
	7 Bands	Vegetation	76.7 ± 3.8	76.8 ± 3.8	76.3 ± 0.2
		Settlement	74.5 ± 2.4	74.4 ± 2.5	75.0 ± 0.1
Gauteng	NDVI	Vegetation	94.4 ± 5.2	94.4 ± 5.2	68.3 ± 14.2
		Settlement	29.2 ± 2.6	29.3 ± 2.9	40.1 ± 31.2
	2 Bands	Vegetation	87.2 ± 8.4	87.2 ± 8.3	92.2 ± 0.4
		Settlement	73.1 ± 22.0	73.1 ± 22.0	83.9 ± 2.1
	7 Bands	Vegetation	75.8 ± 12.3	75.9 ± 12.5	75.8 ± 1.9
		Settlement	24.5 ± 6.8	24.4 ± 6.6	33.2 ± 0.7

the results in [223]. The classification accuracies produced by the MLP were however found to be the highest when using all seven spectral bands.

The MLP was deemed to be the best classifier in this chapter when the feature vectors were extracted with the EKF. Classification accuracies of 95.3% with a standard deviation of 0.7% for the vegetation class, and 96.1% with a standard deviation of 0.6% for the settlement class were reported in the Limpopo province. In the Gauteng province classification accuracies of 99.9% with a standard deviation of 0.1% for the vegetation class and 99.9% with a standard deviation of 0.1% for the settlement class were reported.

It should be noted that the MLP classifier can be replaced with a variety of other classifiers. The MLP performed the best of all the classifiers in this thesis, but like most other supervised machine learning methods, the MLP is dependent on a training set and is required to be robust to any errors occurring within the training set [14]. The drawback in the remote sensing field is that the training data set has to be created with the aid of high spatial resolution imagery, and because of the temporal component must be updated periodically. These periodic updates are a costly endeavour, which justifies the consideration of unsupervised classification methods.

An unsupervised classifier is usually designed by *learning from example*. Thus several clustering methods were evaluated to make deductions about the nature of the feature vectors in the feature space.

Acceptable performance was only obtained with the single, average and complete linkage criteria with proper outlier removal. The other clustering methods did not require the removal of outliers and

for this reason was not explored further.

Ward's clustering method produced the best results of all the hierarchical clustering methods. It was concluded from the experiments conducted that creating spherical clusters with minimum internal variance preserves the inherent distance between feature vectors in the feature space. The algorithm provided acceptable performance for all experiments conducted in the Limpopo province, with the exception that acceptable performance was only observed for the first two spectral band experiments in the Gauteng province.

K-means and EM clustering algorithms were investigated as representative partitional clustering methods, with both methods performing very similarly. The experiments showed empirically that the partitional clustering methods outperformed all the hierarchical clustering methods in the Limpopo province. The partitional clustering methods had the same outcome as the Ward clustering method in the Gauteng province, with similar poor performances in the NDVI- and seven spectral band experiments. The partitional clustering methods were deemed to be better than the Ward clustering method, as they presented classification accuracies with lower standard deviations. The K-means algorithm was the preferred partitional clustering method for its reduced computational complexity.

In the next section the change detection capabilities of the algorithms are explored. Only a few methods were explored, since the change detection in this chapter is based on a post-classification approach. The algorithms that provided acceptable classification performance, which will be explored in the next section, are:

1. the Multilayer perceptron,
2. the Ward clustering method, and
3. the K-means algorithm.

8.7 CHANGE DETECTION

8.7.1 Simulated land cover change detection

A simulated land cover change data set was created to assess the land cover change detection algorithm objectively. The time series data set is used to ensure that the change detection algorithm is able to detect a transition between classes, while analysing the transition.

In table 8.18, the first set of change detection experiments are shown that were conducted in the Limpopo province. All the viable classification approaches that yielded acceptable performance in section 8.6 are shown in these experiments. Each entry in table 8.18 gives the average change detection

Table 8.18: The land cover change detection accuracies are given on the simulated land cover change data set in the Limpopo province. Each entry gives the true positives in percentage (false positives in parentheses).

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
MLP	SFF	6 months	69.2 (30.0)	77.6 (22.4)	90.5 (9.6)
		12 months	70.2 (29.5)	78.2 (21.3)	90.8 (9.4)
		18 months	71.9 (29.2)	78.7 (20.7)	91.0 (8.9)
	Least squares	12 months	68.4 (31.8)	77.5 (22.3)	90.0 (10.1)
	M-estimator	12 months	69.0 (31.1)	77.2 (23.4)	90.2 (10.0)
	EKF	n/a	70.0 (30.3)	79.8 (20.2)	91.7 (8.7)
Ward clustering	SFF	6 months	51.2 (50.5)	71.1 (25.7)	68.3 (30.5)
		12 months	52.4 (48.5)	71.6 (25.5)	68.7 (30.3)
		18 months	52.6 (42.8)	72.2 (24.5)	69.2 (30.1)
	Least squares	12 months	65.4 (33.7)	69.8 (27.9)	67.6 (32.1)
	M-estimator	12 months	65.8 (33.7)	70.1 (28.0)	67.7 (32.3)
	EKF	n/a	59.8 (38.1)	73.0 (22.2)	66.6 (30.8)
K**-means**	SFF	6 months	50.0 (46.8)	71.3 (26.8)	64.3 (33.7)
		12 months	52.7 (46.1)	72.6 (26.5)	65.0 (33.0)
		18 months	53.5 (40.4)	72.9 (24.5)	65.7 (33.7)
	Least squares	12 months	63.4 (36.1)	70.4 (29.8)	65.4 (35.8)
	M-estimator	12 months	63.5 (36.3)	70.6 (29.5)	65.4 (35.8)
	EKF	n/a	57.9 (42.0)	72.8 (22.7)	64.8 (33.8)

accuracies, with the corresponding false alarm rate in parentheses. The change detection accuracies (true positives) are measured on subset 1 and subset 2, which were discussed in section 8.2.4, and the false alarm rates (false positives) are measured on subset 3 and subset 4.

The worst performing experiment was the method that employs the NDVI time series. The overall change detection accuracies were well below 70%, with a reported false alarm rate higher than 30%. In the first two spectral band experiments, acceptable performance was measured across all the methods, with overall change detection accuracies of above 70%, and a reported false alarm rate usually below 26%.

The seven spectral band experiment yielded similar behaviour when compared to the results observed in the classification accuracies. The MLP (supervised classifier) performed exceptionally by reporting overall change detection accuracies above 90% and a false alarm rate below 10%. The unsupervised classifiers, Ward clustering and K-means, reported change detection accuracies which are lower in the higher dimensions (7 spectral bands) than in the lower dimensions (2 spectral bands).

Table 8.19: The land cover change detection accuracies are given on the simulated land cover change data set in the Gauteng province. Each entry gives the true positives in percentage (false positives in parentheses).

Algorithm	**Feature extraction**	**Window length**	**Spectral Band** NDVI	2 Bands	7 Bands
MLP	SFF	6 months	81.2 (16.3)	89.7 (11.1)	97.3 (2.7)
		12 months	83.8 (16.3)	91.8 (10.5)	98.5 (1.5)
		18 months	83.9 (16.4)	92.0 (8.9)	98.5 (1.4)
	Least squares	12 months	78.1 (20.2)	90.0 (13.4)	97.5 (3.4)
	M-estimator	12 months	80.1 (18.9)	90.2 (13.0)	97.6 (3.2)
	EKF	n/a	82.5 (14.0)	93.2 (8.4)	98.4 (1.3)
Ward clustering	SFF	6 months	27.7 (28.8)	77.6 (25.4)	32.6 (31.6)
		12 months	33.2 (31.5)	80.0 (21.6)	36.9 (35.1)
		18 months	35.6 (34.6)	81.1 (19.8)	39.3 (35.4)
	Least squares	12 months	24.5 (17.4)	78.9 (19.7)	33.5 (28.6)
	M-estimator	12 months	24.5 (17.0)	79.2 (19.4)	33.4 (28.7)
	EKF	n/a	25.1 (17.2)	86.1 (7.2)	42.7 (26.0)
K**-means**	SFF	6 months	37.2 (42.9)	77.2 (26.6)	50.4 (41.3)
		12 months	43.8 (41.6)	80.3 (23.4)	51.2 (46.9)
		18 months	45.9 (46.7)	80.4 (24.6)	55.8 (38.7)
	Least squares	12 months	28.6 (21.3)	74.6 (28.5)	50.6 (45.7)
	M-estimator	12 months	28.6 (21.3)	75.0 (28.3)	51.3 (45.4)
	EKF	n/a	36.1 (37.8)	83.8 (5.9)	50.7 (40.8)

The reduction in change detection accuracies can be attributed to the reduction in classification accuracies shown in section 8.6.4 and section 8.6.5. The remedy for this reduction in change detection accuracy in the seven spectral band experiment is again either a more complex clustering algorithm or a more detailed selection of features. The more complex clustering algorithm typically requires a non-linear clustering region to obtain higher change detection accuracies. It is reported in the literature that this shortcoming can typically be solved by over designing the clustering algorithm for a particular data set. The second approach to remedy this reduction is to apply dimensionality reduction, which implies selecting different combinations of spectral bands. The potential risk is that different combinations of spectral bands will perform better on different data sets.

The emphasis in this thesis is placed on obtaining acceptable performance with the clustering algorithm based on the impossibility theorem. Acceptable performance is reported for all methods employing the first two spectral bands, and exceptional performance is reported for the MLP employing all seven spectral bands.

In table 8.19, the second set of change detection experiments are shown that were conducted in the Gauteng province. The same setup is used in these experiments as in the experiments conducted in the Limpopo province. The best performing algorithms were the methods that employ the MLP. The overall change detection accuracies were above 80% with a false alarm rate below 17%. A significant increase in change detection accuracy is observed when the two spectral bands are evaluated when compared to the NDVI. Both the NDVI and two spectral bands' experiments uses the same spectral bands, which implies that using the two spectral bands separately is better.

The worst performing experiments were the methods that employed either the NDVI or all seven spectral bands with an unsupervised classifier. It was observed that experiments conducted with the first two spectral bands along with an unsupervised classifier yielded acceptable performance. The reported overall change detection accuracies were above 75% with a false alarm rate below 30%.

8.7.2 Real land cover change detection

Table 8.20: The land cover change detection accuracy on the real land cover change data set in the Limpopo province. Each entry gives the true positives in percentage (false positives in parentheses).

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
MLP	SFF	6 months	65.4 (32.5)	75.1 (19.5)	84.8 (9.3)
		12 months	66.1 (28.2)	75.3 (18.9)	85.3 (7.9)
		18 months	68.0 (28.7)	76.0 (18.8)	85.3 (8.2)
	Least squares	12 months	64.8 (28.6)	73.8 (23.1)	84.3 (10.1)
	M-estimator	12 months	64.7 (29.9)	73.4 (22.8)	84.3 (9.9)
	EKF	n/a	64.2 (24.6)	78.6 (16.7)	86.8 (8.7)
Ward clustering	SFF	6 months	38.8 (44.7)	67.3 (26.7)	58.7 (35.5)
		12 months	40.3 (52.1)	70.7 (25.9)	63.0 (32.9)
		18 months	40.5 (50.3)	70.0 (25.2)	63.3 (32.6)
	Least squares	12 months	57.6 (36.8)	65.4 (29.0)	62.8 (32.8)
	M-estimator	12 months	57.0 (36.3)	65.4 (28.5)	62.2 (32.8)
	EKF	n/a	52.8 (41.7)	71.8 (26.4)	63.5 (31.1)
K**-means**	SFF	6 months	44.8 (41.1)	70.2 (25.8)	59.8 (29.8)
		12 months	46.0 (42.0)	70.5 (25.4)	60.6 (31.1)
		18 months	46.9 (42.3)	70.5 (25.4)	61.0 (31.4)
	Least squares	12 months	59.8 (37.3)	68.4 (31.1)	61.0 (32.0)
	M-estimator	12 months	59.0 (36.5)	69.0 (30.3)	61.5 (33.4)
	EKF	n/a	51.7 (40.1)	72.0 (24.4)	63.0 (29.9)

In this section, the real land cover change data set (section 8.2.2) is used to measure the performance of the land cover change detection algorithms. This data set is used to test the validity of the algorithms for real world applications [127].

In table 8.20, the first set of change detection experiments are reported that were conducted in the Limpopo province. In these experiments all the viable classifiers identified in section 8.6.7 are explored. Each entry in table 8.20 gives the change detection accuracies (true positives), with corresponding false alarm rates (false positives) in parentheses.

The worst performing methods were those that employed the NDVI spectral band. Overall change detection accuracies in these experiments were observed to be well below 70%. On the other hand, acceptable performance was reported across all the methods using the first two spectral bands, except for the unsupervised classifiers operating on the features extracted with the least squares, and M-estimator.

Table 8.21: The land cover change detection accuracy on the real land cover change data set in the Gauteng province. Each entry gives the true positives in percentage (false positives in parentheses).

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
MLP	SFF	6 months	82.3 (20.5)	86.5 (9.8)	94.3 (2.2)
		12 months	82.3 (16.8)	90.0 (8.8)	95.1 (1.1)
		18 months	83.7 (15.3)	90.4 (8.9)	95.1 (1.0)
	Least squares	12 months	80.0 (16.7)	87.7 (11.8)	94.3 (2.5)
	M-estimator	12 months	80.0 (17.5)	87.7 (10.9)	92.9 (2.8)
	EKF	n/a	83.4 (17.0)	92.1 (9.9)	95.5 (1.6)
Ward clustering	SFF	6 months	15.8 (24.2)	80.1 (21.2)	28.7 (29.8)
		12 months	20.7 (27.0)	80.3 (21.5)	31.3 (30.1)
		18 months	21.2 (28.8)	80.3 (21.4)	31.3 (30.3)
	Least squares	12 months	18.8 (18.0)	78.0 (23.1)	29.7 (29.4)
	M-estimator	12 months	18.1 (17.7)	75.5 (22.2)	30.5 (29.6)
	EKF	n/a	17.8 (17.5)	82.3 (11.3)	38.8 (24.8)
K**-means**	SFF	6 months	32.9 (34.4)	79.2 (24.2)	40.9 (38.9)
		12 months	38.3 (35.1)	79.2 (24.1)	44.7 (42.0)
		18 months	36.0 (34.7)	80.8 (22.7)	46.2 (40.4)
	Least squares	12 months	24.3 (23.9)	75.1 (26.6)	42.3 (40.1)
	M-estimator	12 months	22.8 (23.1)	75.1 (26.2)	44.7 (42.0)
	EKF	n/a	33.3 (29.8)	80.6 (9.8)	43.5 (43.2)

The MLP performed better, by reporting overall change detection accuracies above 84% when using all seven spectral bands. The unsupervised classifiers performed better on the first two spectral

bands than on all seven spectral bands. This was expected, as a similar trend was observed in the classification accuracies.

In table 8.21, the same set of experiments for the real land cover change data set were conducted in Gauteng results are reported. The best performing set of experiments is again the methods that employ the MLP. The overall change detection accuracies are above 80% with false alarm rates below 20%. A significant increase in change detection accuracy is observed when the two spectral spectral bands are evaluated when compared to the NDVI. Because both the NDVI and two spectral bands' experiments uses the same spectral bands, it can be concluded that using the two spectral bands separately is better. This claim is supported by all the previous experiments in this chapter.

The worst performing methods are those that employ either the NDVI or all seven spectral bands with an unsupervised classifier. Meanwhile, similar experiments conducted with the first two spectral bands with an unsupervised classifier yielded acceptable performance. The reported overall change detection accuracies were above 75%, with a false alarm rate below 25%.

The conclusion from both sets of experiments is that using the first two spectral bands with any change detection methods yields acceptable performance. At the same time, experiments using all seven spectral bands with a supervised classifier offered the best reported performance.

8.7.3 Effective change detection delay

In this section, the effective change detection delay Δ_τ is reported. The results of the experiments are presented in table 8.22 for the Limpopo province, and table 8.23 for the Gauteng province. The experiments' results are reported in the average number of days (1 MODIS sample = 8 days) for the ensemble of time series in the simulated land cover change data set.

The MLP was deemed the best performing classifier, as it achieved the shortest effective change detection delay. The MLP's effective change detection delay improved as more spectral bands were included. The best performing feature extraction method was the SFF with a temporal sliding window length of 6 months. The overall trend was that a shorter temporal sliding window length had a shorter effective change detection delay. This is intuitive as fewer data points contribute to the current state of the output class membership. The SFFs outperform the least squares and M-estimator using a similar temporal sliding window length of 12 months.

The unsupervised classifiers (Ward clustering method and K-means) reported an overall increase in effective change detection delay when compared to the MLP classifier. A similar observation is made here as in the discussion of classification accuracy in section 8.6.7. The first two spectral bands outperformed the NDVI and all seven spectral band combinations. This is due to the improved

Table 8.22: Effective change detection delay for simulated land cover change conducted in the Limpopo province. Each entry gives the average number of days for each study area, calculated over 10 repeated independent experiments.

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
MLP	SFF	6 months	88	76	73
		12 months	117	101	92
		18 months	178	120	106
	Least squares	12 months	130	109	102
	M-estimator	12 months	146	118	109
	EKF	n/a	110	96	91
Ward clustering	SFF	6 months	132	92	116
		12 months	177	113	160
		18 months	253	176	218
	Least squares	12 months	185	130	166
	M-estimator	12 months	189	125	186
	EKF	n/a	163	104	151
K**-means**	SFF	6 months	127	94	119
		12 months	169	107	154
		18 months	233	164	216
	Least squares	12 months	186	127	165
	M-estimator	12 months	186	123	179
	EKF	n/a	166	105	151

classification accuracies reported in section 8.6.3–8.6.6 for the first two spectral bands.

Most experiments conducted in the Limpopo province had the K-means algorithm producing shorter effective change detection delays than the Ward clustering method, while no distinguishing difference was observed in the Gauteng province. In these experiments a clear improvement in the effective change detection delay is observed when the SFF is compared to the least squares and M-estimator with a similar sliding window length.

8.7.4 Summary of change detection results

In this section the results of the change detection experiments are summarised. In section 8.7.1, true positives and false positives were reported for the experiments conducted on the simulated land cover change data set. In section 8.7.2, the true positives were reported for the experiments conducted on the real land cover change data set. In section 8.7.3, the average effective change detection delays were reported for the experiments conducted on the simulated land cover change data set.

Table 8.23: Effective change detection delay for simulated land cover change conducted in the Gauteng province. Each entry gives the average number of days for each study area, calculated over 10 repeated independent experiments.

Algorithm	**Feature extraction**	**Window length**	**Spectral Band**		
			NDVI	2 Bands	7 Bands
MLP	SFF	6 months	84	69	65
		12 months	111	87	81
		18 months	153	114	109
	Least squares	12 months	122	98	94
	M-estimator	12 months	127	99	97
	EKF	n/a	108	89	81
Ward clustering	SFF	6 months	117	84	102
		12 months	146	103	139
		18 months	168	140	168
	Least squares	12 months	155	120	146
	M-estimator	12 months	164	123	154
	EKF	n/a	151	97	138
K**-means**	SFF	6 months	118	88	110
		12 months	139	112	143
		18 months	172	157	189
	Least squares	12 months	153	126	149
	M-estimator	12 months	157	128	153
	EKF	n/a	137	106	134

The MLP was considered the best classifier used for change detection. The MLP had better change detection accuracies and effective change detection delays when using more spectral bands. It was also found that a trade-off existed in the length of the temporal sliding window when comparing the difference between change detection accuracy and effective change detection delay. A longer temporal sliding window length improves the classification accuracy at the cost of a longer effective change detection delay. A shorter temporal sliding window length reacts faster to change in the time series at the loss in change detection accuracy.

Poor performance with the unsupervised methods used for clustering on the NDVI time series and all seven spectral bands' time series indicated that classes could not be well encapsulated in the clusters. The first two spectral bands, on the other hand, resulted in acceptable performance across all the change detection experiments and effective change detection delay's experiments.

The K-means algorithm and Ward clustering method performed similarly in all the experiments, except that the Ward clustering method had slightly higher change detection accuracies while the

Table 8.24: A list of different combinations of change detection algorithms that will be tested at a regional scale.

Feature extraction	**Sliding window length**	**Spectral band**	**Machine learning method**
SFF	12 months	2 Bands, 7 Bands	MLP
	12 months	2 Bands	Ward clustering method
	12 months	2 Bands	K-means algorithm
EKF		2 Bands, 7 Bands	MLP
		2 Bands	Ward clustering method
		2 Bands	K-means algorithm

K-means algorithm had a shorter effective change detection delay. This observation could be attributed to the K-means classification experiments, which yielded a very small standard deviation when compared to the Ward clustering method. In all the experiments conducted in this section (section 8.7), it was observed that the SFFs and EKF features outperformed the least squares and M-estimator features in the performance metrics. It is concluded from these experiments that the combinations given in table 8.24 yielded the best performance and will be evaluated on a regional scale.

8.8 CHANGE DETECTION ALGORITHM COMPARISON

In this section the change detection accuracies measured in section 8.7 are compared to other change detection algorithms found in the literature. The change detection methods used for comparison are:

- the annual NDVI differencing method (denoted by NDVI_{CDM}) [19],
- the EKF change detection method (denoted by EKF_{CDM}) [120], and
- the ACF change detection method (denoted by ACF_{CDM}) [121].

All three these methods listed above are supervised in nature, as a training data set is required to set a threshold, which is used to declare change. These three methods are compared in table 8.25 to a few methods listed in table 8.24.

The worst performing method was the NDVI_{CDM} method, having a change detection accuracy of 69% with a false alarm rate of 13% in the Limpopo province, and a change detection accuracy of 57% with a false alarm rate of 14% in the Gauteng province. A possible explanation for this poor performance is given in [224], which is that the method assumes that the annual NDVI difference between years is normally distributed, which could imply that it has difficulty in detecting land cover

Table 8.25: Comparison of the change detection accuracies in percentage (false alarm rate in parentheses) of the proposed change detection algorithms to other change detection algorithms found in the literature.

Algorithm	**Province**	
	Limpopo province	Gauteng province
EKF_{CDM} [19]	89% (13%)	75% (13%)
ACF_{CDM} [120]	81% (12%)	92% (15%)
NDVI_{CDM} [121]	69% (13%)	57% (14%)
EKF_{BVEP}, MLP, 7 spectral bands	87% (9%)	96% (2%)
EKF_{BVEP}, MLP, 2 spectral bands	79% (23%)	92% (10%)
EKF_{BVEP}, K-means, 2 spectral bands	72% (24%)	81% (10%)
EKF_{BVEP}, Ward clustering, 2 spectral bands	72% (26%)	82% (11%)

change in heterogeneous areas. The method performed the poorest in the Gauteng province owing to the land cover diversity [224].

The EKF_{CDM} had the highest change detection accuracy of 89% in the Limpopo province, with a false alarm rate of 13%. This was attributed to the fact that most of the province is covered by natural vegetation, which is the result of the high correlation between the parameter sequences of the neighbouring pixels in the spatio-temporal window [224]. The relative difference between the change and no change parameter streams was high enough to detect change. The EKF_{CDM} method's performance was lower in the Gauteng province, which was attributed in [224] to the land cover diversity.

The ACF_{CDM} exploits the non-stationary property of the change time series when compared to the no change time series. The method was applied to the 4^{th} spectral band of MODIS, as it offered the best performance [224]. The method reported a higher change detection accuracy in the Gauteng province when compared to the Limpopo province.

The performance of the two unsupervised classifiers (K-means and Ward clustering) operating on the first two spectral bands was similar. Both methods had better change detection accuracies and false alarm rates when compared to the NDVI_{CDM} method. The methods had a 6% higher change detection accuracy when compared to the EKF_{CDM} in the Gauteng province, but a 17% decrease in the Limpopo province.

The MLP operating on the EKF_{BVEP} features computed from the first two spectral bands had the same change detection accuracy as the ACF_{CDM} in the Gauteng province, but had the advantage of having a 5% lower false alarm rate. The reverse was observed in the Limpopo province, as the MLP operating on the first two spectral bands had a 2% lower change detection accuracy and 11% higher false alarm rate when compared to the ACF_{CDM} method.

The MLP operating on the EKF_{BVEP} features computed on all seven spectral bands was deemed the best change detection method in this section. The method had the highest change detection accuracy and lowest false alarm rate in the Gauteng province. It had the second highest change detection accuracy (2% lower than the highest) and the lowest false alarm rate in the Limpopo province.

8.9 PROVINCIAL EXPERIMENTS

A list of the best performing change detection algorithms is given in table 8.24, which is to be evaluated on a regional scale. The areas that will be evaluated are the entire Limpopo and Gauteng provinces.

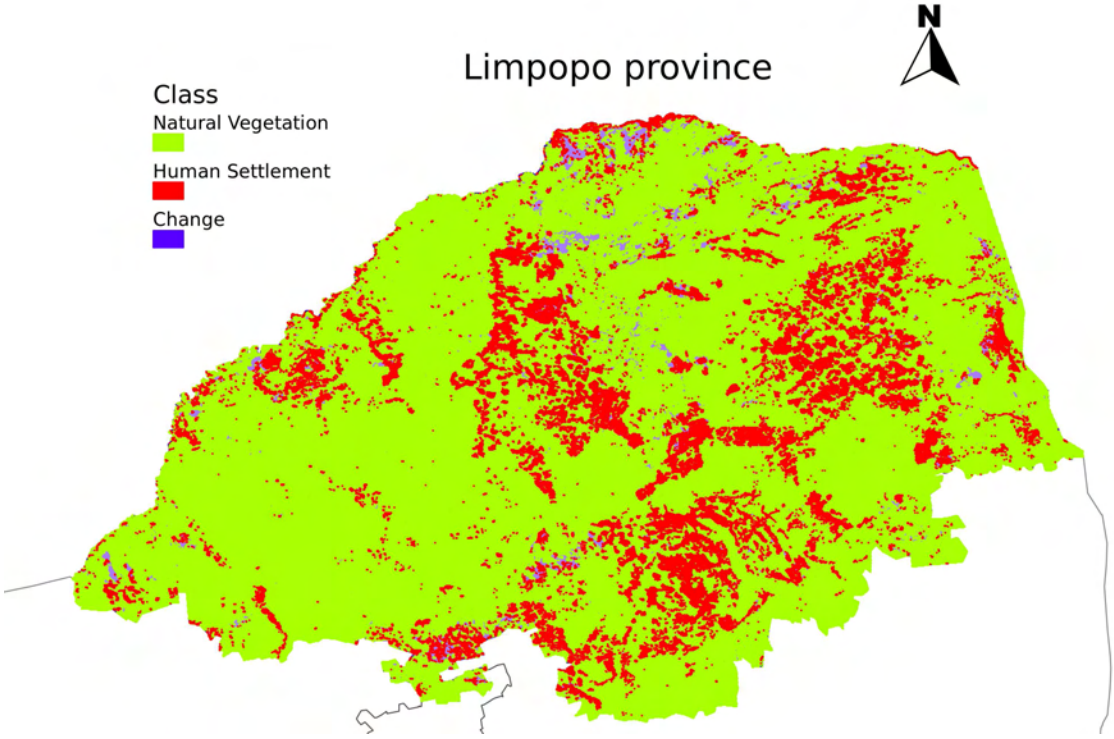


FIGURE 8.13: A classification/ change detection map of the entire Limpopo province.

The results obtained from processing the entire Limpopo province are presented in table 8.26. The table divides the results into three categories: natural vegetation, human settlements, and change. An illustration of one of these experiments is shown in figure 8.13, which represents the Limpopo province. The overall trend throughout all the methods was that natural vegetation covered 85%–88% of the province, and that human settlement covered 9%–12% of the province. This signifies that majority of the province is still largely covered by natural vegetation. The land cover change that is reported here is the transformation of natural vegetation to human settlement. The land cover change that was reported ranged from 1%–4% of the total area in the province. This is a significant area that has changed in

the province over the past decade, since the total human settlement class has expanded by 12%–40% in the study period. This suggests that some of the algorithms might be too sensitive towards change events or that the labelled data set should be expanded to incorporate a larger variety of classes. On the other hand, it should be noted that the controlled experiments that were conducted on the labelled data set involved land cover that transformed from natural vegetation to human settlement. This did not include any examples of other land cover transformations, which could exist in the province. This could be rectified, as the algorithms are versatile enough to include other classes to improve the classification, and in turn change detection accuracies. Future expansion of the work could entail collecting agricultural land cover information in each of the provinces.

Table 8.26: The classification and change detection results produced for the entire Limpopo 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
SFF	MLP	2 Bands	86.94	10.31	2.75
		7 Bands	87.69	10.61	1.70
	Ward clustering	2 Bands	86.33	9.64	4.03
	K-means	2 Bands	86.05	10.02	3.93
EKF	MLP	2 Bands	85.74	11.57	2.69
		7 Bands	86.33	12.11	1.56
	Ward clustering	2 Bands	86.20	10.32	3.48
	K-means	2 Bands	85.81	10.90	3.29

Closer inspection of table 8.26 allows the deduction of some interesting trends. These trends cannot be confirmed, as no ground truth exists for the current results, which are only based on observations. The MLP consistently detected more human settlement than the unsupervised classifiers, while indicating a reduced number of detected land cover changes. This puts emphasis on the classification at the beginning of the time series, as both the detected land cover change class and the human settlement class agree that the time series ends in the human settlement class. This could be attributed to the fact that the province experienced a rainfall shortage in 2001/2002 (beginning of the study period).

The unsupervised classifiers detected more land cover change when compared to the MLP. In some experiments the size of changed areas that were reported almost doubled. Another observation among the unsupervised classifiers is that the Ward clustering method flagged more land cover changes than the K-means algorithm. This trend was also observed in the controlled experiments and was deduced

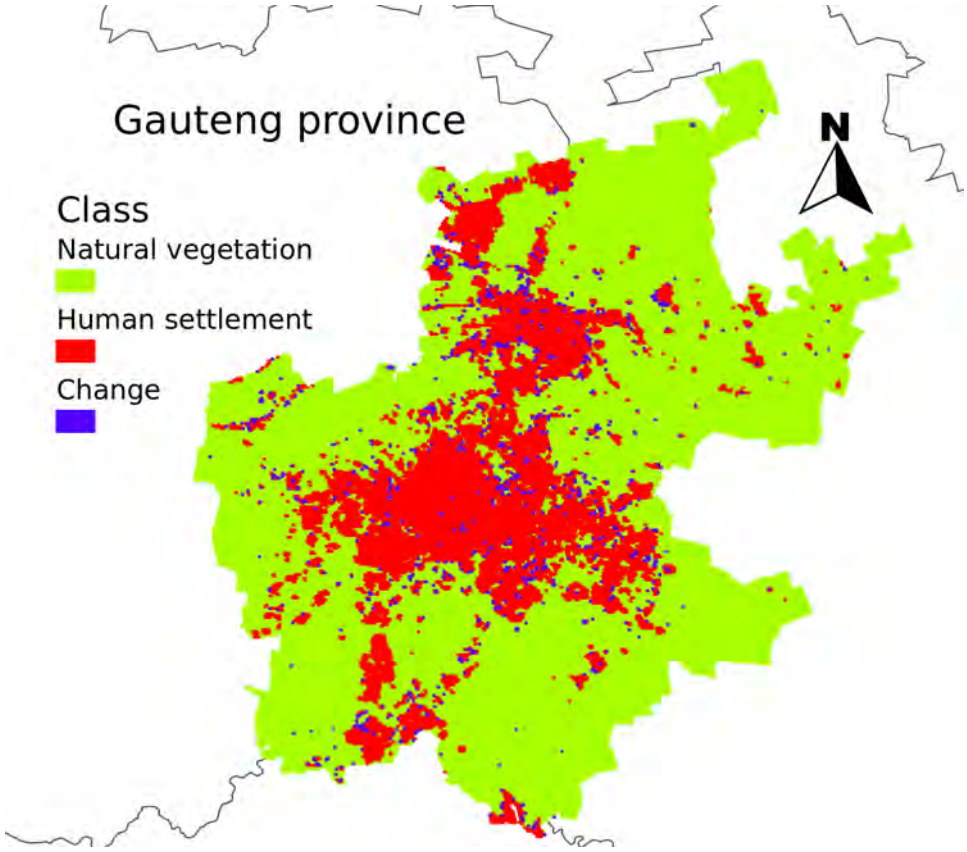


FIGURE 8.14: A classification/ change detection map of the entire Gauteng province.

from the observation that the Ward clustering method had a wider standard deviation in its classification accuracies than the K-means.

Table 8.27: 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
SFF	MLP	2 Bands	76.65	20.12	3.23
		7 Bands	77.33	21.39	1.28
	Ward clustering	2 Bands	75.53	19.90	4.57
	K-means	2 Bands	75.43	20.46	4.11
EKF	MLP	2 Bands	76.01	20.92	3.07
		7 Bands	76.89	21.46	1.17
	Ward clustering	2 Bands	76.22	19.56	4.22
	K-means	2 Bands	76.08	19.96	3.96

The same experiment was conducted in the Gauteng province and its results are presented in

table 8.27. The results were produced by processing the entire Gauteng province into the three defined categories. An illustration of one of these experiments is shown in figure 8.14, which represents the Gauteng province. The overall trend in this province was significantly different from the results produced in the Limpopo province, as this province is mostly urbanised. The natural vegetation class covered 75%–78% of the province, while human settlements covered 19%–22%. This result supports the concept that Gauteng is a heavily urbanised province.

The land cover change which was flagged ranged from 1%–5% of the total area in the province. This is a significant large area that has changed in the study period, as the total human settlement class has expanded by 5%–23% in the province. The same trends that were observed in the results produced for the Limpopo province with regard to the nature of the change detection algorithm were observed in the Gauteng province.

8.10 COMPUTATIONAL COMPLEXITY

In this section a comparison is made of the complexity of extracting the EKF features and the SFFs. A time series $\mathbf{x}$ of length $\mathcal{I}$, is defined as

$$\mathbf{x} = [\vec{x}_1\ \vec{x}_2\ \ldots\ \vec{x}_{\mathcal{I}}], \tag{8.1}$$

with

$$\vec{x}_i = [x_{i,1}\ x_{i,2}\ \ldots\ x_{i,T}]. \tag{8.2}$$

The variable T denotes the number of elements in vector $\vec{x}_i$. If the state-space vector $\vec{W}_i$ used in the EKF has S elements, then the complexity of filtering a single time series is at least $\mathcal{O}(\mathcal{I}S^2)+\mathcal{O}(\mathcal{I}T^{2.4})$. In the case of the EKF features extracted from a triply modulated cosine function on uncorrelated spectral bands, S=3 and T=1.

The complexity of extracting the SFF is based on the complexity of the FFT algorithm and the length of the temporal sliding window. If the time series is length $\mathcal{I}$ and the length of the temporal sliding window is Q, then the processing of a single time series is equal to $\mathcal{O}((\mathcal{I}-Q)Q\log_2 Q)$, with $Q \ll \mathcal{I}$.

A timing experiment was conducted on a cluster node to calculate the computational time of both feature extraction methods and the results are reported in table 8.28. The computer's specifications used for this experiment are:

- Dell PowerEdge 1955 blade, Intel Xeon 5355 (Quad-Core) 2.66 GHz, 8 GB RAM, 1333 MHz

Table 8.28: The computational time required to extract features from 25000 time series using either the EKF feature extraction method or SFF extraction method. The results is reported in milliseconds per time series.

Feature	Millisecond per time series
SFF	0.47
EKF	22.81

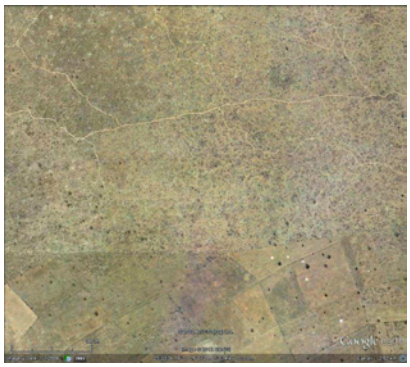
FSB, Gigabit Ethernet, 4x 2.1 kW redundant power supplies (3+1), 2x Gigabit Switch Modules, 1x Avocent Digital Access KVM switch, Software Debian Testing AMD64 with MATLAB R2012a.

The experiment was conducted over 25000 time series and it was concluded that the SFF could be extracted from the time series 48.5 times faster than the EKF features. The next requirement addressed is the time required to optimise the EKF features using the BVEP criterion. The BVSA is an iterative search algorithm that sets the EKF parameters within the BVS in an attempt to best satisfy the BVEP criterion. If the BVSA requires E_{BVSA} iterations to set the EKF parameters, the the extraction of $\mathrm{EKF}_{\mathrm{BVEP}}$ features takes at least $48.5E_{\mathrm{BVSA}}$ times longer than the SFF. The typical range of iterations used for E_{BVSA} in these experiments were between 20 and 30.

8.11 SUMMARY

In this section a summary is provided of the observarions made in this chapter. It was found that the supervised classifier outperformed the unsupervised methods. The downside was the costs involved in producing a labelled training data set. The best performance was obtained when the MLP was optimally set to operate on all seven spectral bands of MODIS. The training method adopted was the iteratively retrained mode, which compensates for the inter-annual variability. A temporal sliding window length of 12 months used on either the SFF, least squares, or M-estimator offered the best trade-off between parameter variability, effective change detection delay and change detection accuracy. Similar gains were obtained in the trade-off with the EKF features if the parameters were optimised with the BVEP criterion.

The change detection algorithms yielded better performance in the Gauteng province than the Limpopo province. This could be attributed to the more dense natural vegetation found in the Gauteng province. Figure 8.15 illustrates a difference between the informal settlements and natural vegetation found in both provinces. The Gauteng province houses more compact informal settlements and more dense natural vegetation when compared to the Limpopo province.



(a) Natural vegetation located in the Limpopo province.

(b) Informal settlements located in the Limpopo province.

(c) Natural vegetation located in the Gauteng province.

(d) Informal settlements located in the Gauteng province.

FIGURE 8.15: Four high resolution images acquired in the two provinces; Limpopo and Gauteng. (a) A natural vegetation area located in the Limpopo province. (b) An informal settlement located in the Limpopo province. (c) A natural vegetation area located in the Gauteng province. (d) An informal settlement located in the Gauteng province. (courtesy of Google^{TM}Earth).

A general trend of performance improvement was observed when the first two spectral bands (Red and NIR spectral bands) were used instead of the NDVI. The use of the first two spectral bands as input was deemed superior, as the same spectral bands are used to compute the NDVI. Further improvement was observed when using all seven spectral bands with a supervised classifier.

The SFFs and EKF features yield better performance in detecting land cover change when compared to the features extracted using least squares and M-estimator methods. The EKF features only provided better separation between classes than the SFFs when the BVEP criterion was used to set the EKF parameters. The consequence of this is that the SFF was deemed the better approach when compared to the EKF features, as the EKF-extracted features required the computation of the covariance matrices using the BVEP criterion. This improvement into separation in classes was not significant, and the SFF was deemed better owing to its lower computational time (section 8.10).