CHAPTER 5

RESULTS

The chapter starts with preliminary data analysis results obtained from the datasets introduced in Section 2.8. These results can be used to predict the performance of the different classification and change detection approaches. In Section 5.2 the inductive simulator discussed in Section 4.1.2 is validated. The final sections of the chapter give the classification and change detection accuracies and rankings of the different sequential and non-sequential hypertemporal classification and change detection algorithms discussed in Chapter 4.

5.1 PRELIMINARY DATA ANALYSIS: GAUTENG AND LIMPOPO

In this section a preliminary investigation of the datasets introduced in Chapter 2 is performed. The knowledge gained from the preliminary analysis is used to explain the classification and change detection results of Section 5.3 and Section 5.4. The data analysis is conducted under the following headings: yearly ensemble mean (Section 5.1.1), temporal Hellinger distance (Section 5.1.2), CSHO parameters (Section 5.1.3), noise correlation (Section 5.1.4) and spatial correlation (Section 5.1.5). An important data manipulation technique used in the remainder of the section is temporal grouping of multispectral observations and is expressed in mathematical notation below.

Recall from Chapter 4 that each MODIS pixel (which is denoted here explicitly by $\bar{\mathbf{x}}_c$) has eight associated time-series, such that $\bar{\mathbf{x}}_c = \{\tilde{\mathbf{x}}_c[k]\}_{k=\{1,2,\cdots\}}$, where $\tilde{\mathbf{x}}_c[k] \in \mathbb{R}^8$ is the multispectral observation at time k. To group certain time steps together, the projection operator is required and is defined as

$$\underset{i\in\mathscr{I}}{\text{pr}}\ \boldsymbol{\psi} = \{\psi_i\}_{i\in\mathscr{I}}, \tag{5.1}$$

such that $\text{pr}_i\boldsymbol{\psi}$ is the i-th component of the sequence $\boldsymbol{\psi}$. In Equation 5.1, $\mathscr{I}$ denotes the index set.

Let $\mathscr{D}_c$ denote the set of all $\bar{\mathbf{x}}_c$ belonging to land cover class c. Then, for each observation period i, $1 \leq i \leq j = 45$ (with j the number of observations in a year), and for each land cover class c, the set $\mathscr{G}_{i,c}$ is defined as [23]

$$\mathscr{G}_{i,c} = \left\{ \mathscr{X} \in \underset{i+jn}{\text{pr}} \; \bar{\mathbf{x}}_c | \bar{\mathbf{x}}_c \in \mathscr{D}_c \right\}, \; n = 1,2,\cdots,N, \; 1 \leq i \leq j,$$

such that $\mathscr{G}_{i,c}$ denotes the set of all multispectral observations for a specific time i during the year, and a particular land cover class c (with N being the number of years). Furthermore, let $\mathscr{G}_{i,c}^{\mathbf{b}}$ denote the set of observations corresponding to a particular time of the year i, a particular class c, and a selection of $|\mathbf{b}|$ spectral bands, $\mathbf{b} \subseteq \{1,\cdots,7,\text{NDVI}\}$.

5.1.1 Yearly ensemble mean

The yearly ensemble mean $\tilde{\mathbf{y}}_c(t)$ is defined by Equation 4.5. The yearly ensemble mean for each class is estimated by taking the average at each observation time step over all pixels and then over all years. In other words, inter-annual variability is not ignored but averaged to obtain the yearly ensemble mean for each class.

5.1.1.1 Yearly ensemble mean: Gauteng

The estimated yearly ensemble means $\tilde{\mathbf{y}}_v(t)$ and $\tilde{\mathbf{y}}_s(t)$ for the Gauteng data set are presented in Figures 5.1 and 5.3. The v and s subscripts respectively refer to the vegetation and settlement class. The result of fitting sinusoids on the estimated yearly ensemble means is displayed in Figures 5.2 and 5.4.

5.1.1.2 Yearly ensemble mean: Limpopo

The estimated yearly ensemble means $\tilde{\mathbf{y}}_v(t)$ and $\tilde{\mathbf{y}}_s(t)$ for the Limpopo dataset are presented in Figures 5.5 and 5.6 respectively.

5.1.1.3 Discussion of yearly ensemble mean

The average absolute distance (for each band b) between the yearly ensemble means of the vegetation and settlement classes is defined as

$$\bar{y}^b = \frac{1}{45} \sum_{k=1}^{45} |y_v^b[k] - y_s^b[k]|.$$

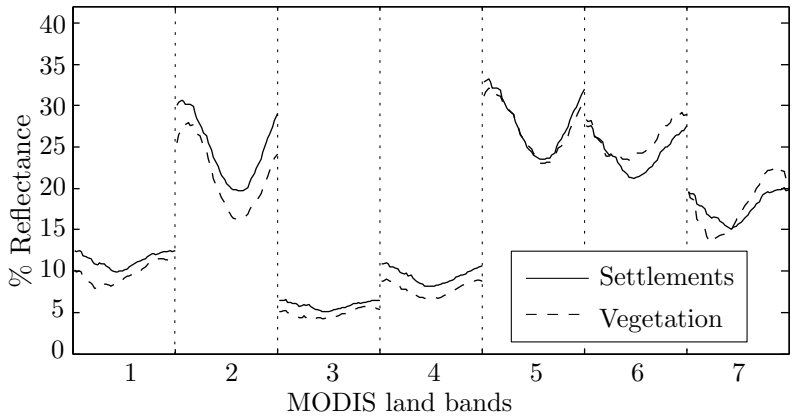


Figure 5.1: The yearly ensemble mean of the MODIS land bands for the vegetation and settlement classes (Gauteng) [2] © IEEE 2012.

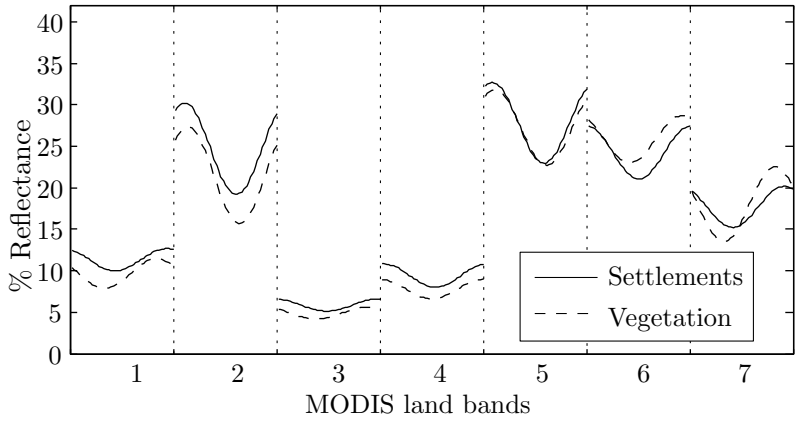


Figure 5.2: Sinusoidal fits on the yearly ensemble mean of the MODIS land bands for the vegetation and settlement class (Gauteng).

The average absolute distance between the yearly ensemble means of the vegetation and settlement classes (for both data sets) can be found in Table 5.1. The values in Table 5.1 were computed with the DN reflectance values and not the scaled values that are used in Figure 5.1 to Figure 5.6.

The average standard deviation (for each band) about the yearly ensemble mean is given in Table 5.2 (for each class and dataset). The average standard deviation is calculated by first grouping all observation time steps (over multiple years and pixels) together and then computing the standard deviation at each time step in a year. To obtain the average standard deviation, the average is then taken over all the time steps in a year of the computed standard deviations. Mathematically speaking it can be

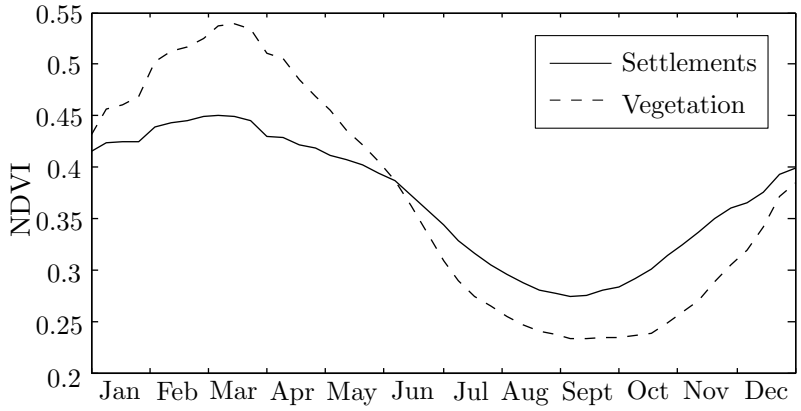


Figure 5.3: The yearly ensemble mean of NDVI for the vegetation and settlement classes (Gauteng) [2] © IEEE 2012.

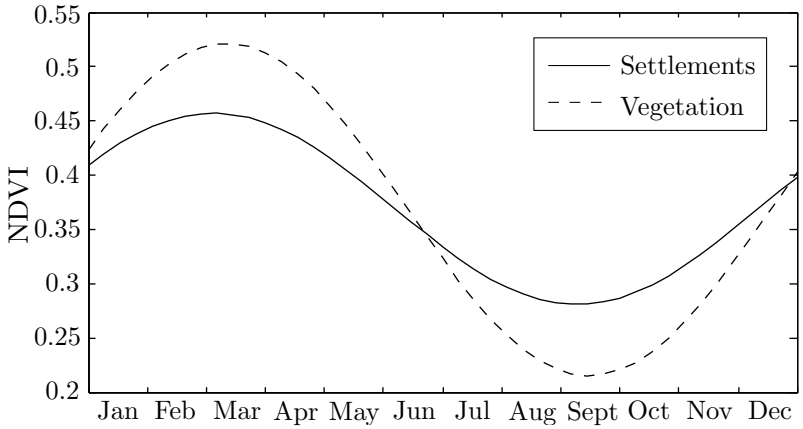


Figure 5.4: Sinusoidal fits on the yearly ensemble mean of NDVI for the vegetation and settlement class (Gauteng).

expressed as

$$\{\mathbb{E}[\{\text{std}(\mathscr{G}_{i,c}^{b})\}_{i\in\{1,\cdots,45\}}]\}_{b\in\{1,2,\cdots,7,\text{NDVI}\}}.$$

The following observations and conclusions can be made from the average yearly ensemble mean results:

1. In the case of the Gauteng dataset the average absolute distance between the yearly ensemble means of the vegetation and settlements class across all bands is equal to 144.07. For the Limpopo dataset the average absolute distance corresponds to 289.10. The absolute distance between the vegetation and settlement yearly ensemble means is therefore larger for the Limpopo dataset. This can be visually verified by using Figure 5.1 to Figure 5.6.

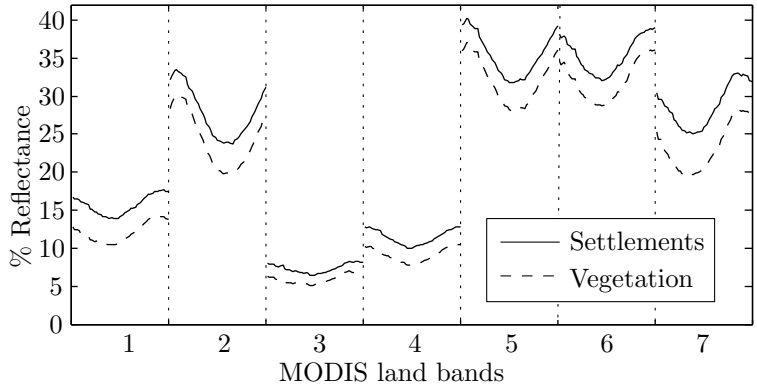


Figure 5.5: The yearly ensemble mean of the MODIS land bands for the vegetation and settlement classes (Limpopo) [2] © IEEE 2012.

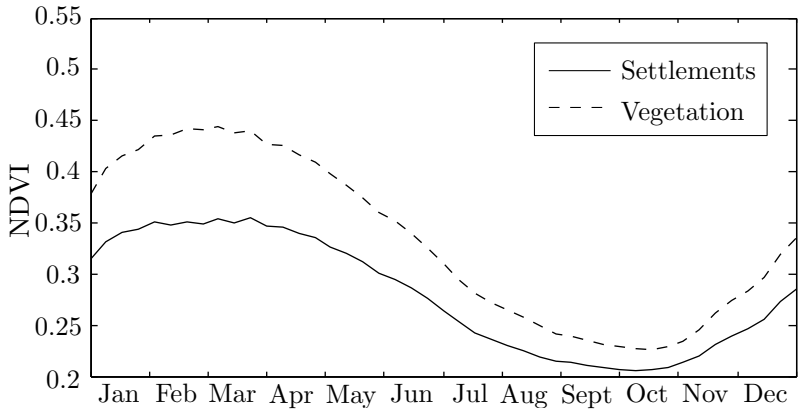


Figure 5.6: The yearly ensemble mean of NDVI for the vegetation and settlement classes (Limpopo) [2] © IEEE 2012.

2. The average standard deviation across all bands for the the vegetation class is respectively equivalent to 216.67 and 280.48 for the Gauteng and Limpopo datasets. In the case of the settlements class 223.67 and 287.35 are respectively obtained. On average the Limpopo data set has a higher standard deviation about the yearly ensemble mean when compared with the Gauteng dataset. Furthermore, in general the settlement classes also have a higher standard deviation around the yearly ensemble mean when compared with the vegetation classes.

3. The higher average absolute distance observable between the yearly ensemble means of the vegetation and settlements class for the Limpopo dataset implies that the Limpopo dataset is more separable than the Gauteng dataset. On the other hand, the higher average standard deviation about the yearly ensemble mean found in the Limpopo dataset, would suggest that the

Table 5.1: Average absolute distance between the yearly ensemble means of the vegetation and settlements class (for both datasets).

	Band							
Dataset	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**NDVI**
Gauteng	166.12	317.82	97.09	162.18	65.07	168.71	175.55	0.0474
Limpopo	357.06	363.48	142.63	232.65	341.54	340.75	534.64	0.0525

Table 5.2: Average standard deviation about the yearly ensemble mean.

	Band							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**NDVI**
Vegetation								
Gauteng	181.47	329.06	93.08	117.83	336.67	339.28	335.88	0.0794
Limpopo	249.79	343.11	110.85	147.04	423.00	488.04	481.93	0.0734
Settlement								
Gauteng	225.74	239.93	106.97	140.07	280.65	370.95	424.97	0.0799
Limpopo	248.40	311.77	123.35	165.93	405.72	515.95	527.60	0.0617

Limpopo dataset is less separable than the Gauteng dataset. Looking at the average distance of the yearly ensemble means between classes or at the average standard deviations about the yearly ensemble means separately is not enough to predict high or low separability between classes. In contrast, the temporal Hellinger distance defined in Section 5.1.2 takes into account the average distance and the average standard deviation to determine to what extent two classes are separable.

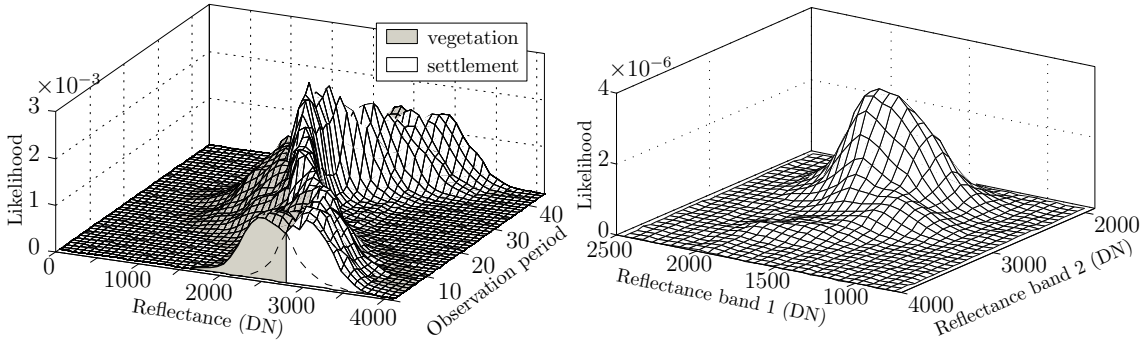
5.1.2 Temporal Hellinger distance

The Hellinger distance between probability density functions p and q is a value between 0 and 1 and is defined as

$$HD(p,q) = \sqrt{1 - \int_{-\infty}^{\infty} \sqrt{p(x)q(x)}\, dx}.$$

A Hellinger distance of $HD(p,q) \approx 0$ indicates that the densities are not separable, whereas a distance of $HD(p,q) \approx 1$ indicates that the densities are trivially separable.

The temporal Hellinger distance is the Hellinger distance between the single-band time-varying models of the settlement and vegetation classes and is a measure of the degree of separability between the two classes. In other words, the closer the temporal Hellinger distance is to 1 at a specific time step in a year the better a temporal feature classifier will be able to distinguish a settlement observation from a vegetation observation at that specific time step in the year (in theory). The time-varying model was briefly discussed in Section 4.2.3. The time-varying model is constructed by estimating the density of $\mathscr{G}_{i,c}^{\mathbf{b}}$ at each time step in a year. The densities were estimated by using Kernel Density Estimation (KDE), employing a Gaussian kernel and Silverman's rule of thumb as the bandwidth selection rule (Section A.5) [195, 196]. The densities were constructed via the *KDE toolbox for Matlab* [197]. Recall from Section 4.2.3 that the time-varying model densities are denoted with $\{q_i^{c,\mathbf{b}}\}_{i\in\{1,\cdots,45\}}$. The estimated Gauteng dataset time-varying model for both the vegetation and settlement class in land band 2 is depicted in Figure 5.7a, i.e. $\{q_i^{v,2}\}_{i\in\{1,\cdots,45\}}$ and $\{q_i^{s,2}\}_{i\in\{1,\cdots,45\}}$. The estimated multispectral density at time step $i=1$ between land bands 1 and 2 for the settlements class is displayed in Figure 5.7b.



(a) $\{q_i^{v,2}\}_{i\in\{1\cdots45\}}$ and $\{q_i^{s,2}\}_{i\in\{1\cdots45\}}$ [198] © IEEE 2012. **(b)** $q_1^{s,\{1,2\}}$.

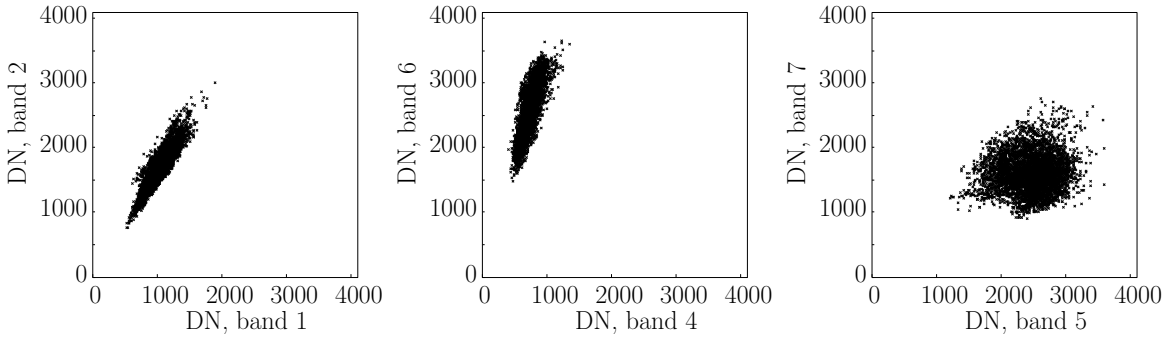
Figure 5.7: Single-band and multiband time-varying models for the Gauteng data set.

The temporal Hellinger distance in band b is defined as

$$H^b[i] = \mathrm{HD}(q_i^{v,b}, q_i^{s,b}). \tag{5.2}$$

At this point it is perhaps worthwhile to give some empirical incitement for using multispectral models. In Figure 5.8 there are three dual-band scatter diagrams at three different times of the year (for the Gauteng vegetation dataset).

Figure 5.8a indicates that there is a strong correlation at time $i=35$ between bands 1 and 2, while Figure 5.8c shows a weak correlation between bands 5 and 7 at time step $i=22$. Figure 5.8b testifies



(a) Scatter diagram ($N = 592$) between bands 1 and 2, $i = 35$.

(b) Scatter diagram ($N = 592$) between bands 4 and 6, $i = 32$.

(c) Scatter diagram ($N = 592$) between bands 5 and 7, $i = 22$.

Figure 5.8: Scatter diagrams of several spectral bands, at different times of the year.

that there is a moderate degree of correlation present between bands 4 and 6 at time $i = 32$. The degree of correlation in Figure 5.8b is less than the degree of correlation in Figure 5.8a, but more than that attested by Figure 5.8c. The reason for only giving three scatter diagrams (out of thousands of possibilities), is that the aim here is not to give a comprehensive description of the dependencies between the spectral bands, but to motivate the use of multispectral models.

5.1.2.1 Temporal Hellinger distance: Gauteng

The temporal Hellinger distance between the single-band time-varying models of the vegetation and settlement classes for the Gauteng dataset is depicted in Figure 5.9.

5.1.2.2 Temporal Hellinger distance: Limpopo

The temporal Hellinger distance between the single-band time-varying models of the vegetation and settlement classes for the Limpopo dataset is displayed in Figure 5.10.

5.1.2.3 Discussion of temporal Hellinger distance

The average and maximum temporal Hellinger distance (across all time steps in a year) between the time-varying models of the vegetation and settlement classes is given in Table 5.3.

The following comments pertain to the temporal Hellinger distance metric:

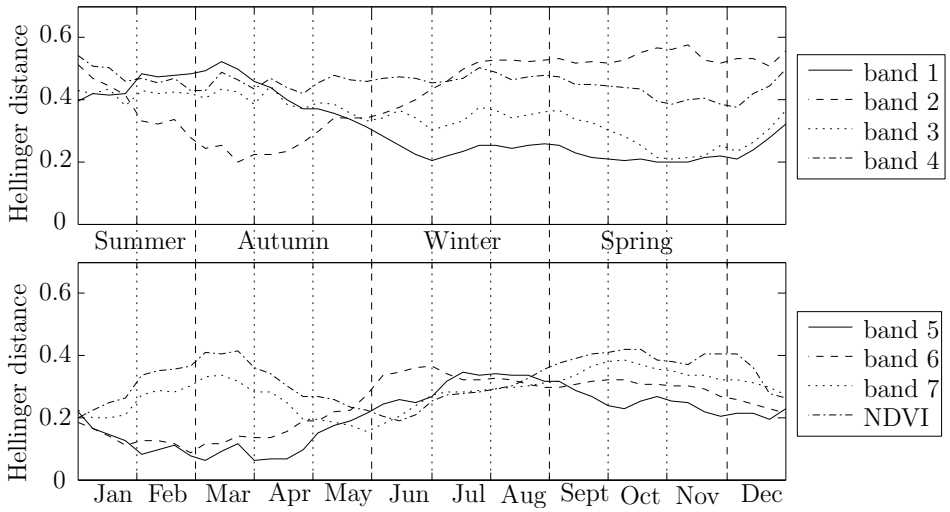


Figure 5.9: The temporal Hellinger distance between the single-band time-varying models of the vegetation and settlement classes for the Gauteng dataset [198] © IEEE 2012.

1. In the case of the Gauteng dataset the average Hellinger distance (across all bands) is equal to 0.32, while the average maximum Hellinger distance (across all bands) corresponds to 0.45.

2. In the case of the Limpopo dataset the average Hellinger distance and the average maximum Hellinger distance respectively correspond to 0.37 and 0.43.

3. During certain times of the year the time-varying models of the Gauteng dataset are more separable than the time-varying models of the Limpopo dataset (based on the average maximum temporal Hellinger distance between the time-varying models of the vegetation and settlement classes), while the time-varying models of the Limpopo dataset are on average more separable than the time-varying models of the Gauteng dataset (based on the average temporal Hellinger distance between the time-varying models of the vegetation and settlement classes). Table 5.3 predicts that the accuracy performance of a temporal classifier applied to the Gauteng or Limpopo datasets would be similar.

4. From Table 5.3 it is clear that in the case of the Gauteng dataset bands $\{1,2,3,4\}$ provide a higher degree of separability (between the time-varying models of the vegetation and settlement classes) than bands $\{5,6,7,\text{NDVI}\}$, which implies that a temporal classifier using bands $\{1,2,3,4\}$ should perform better than a temporal classifier using bands $\{5,6,7,\text{NDVI}\}$ (theoretically speaking). The vegetation and settlement classes are most separable in band 2 and least separable in band 5.

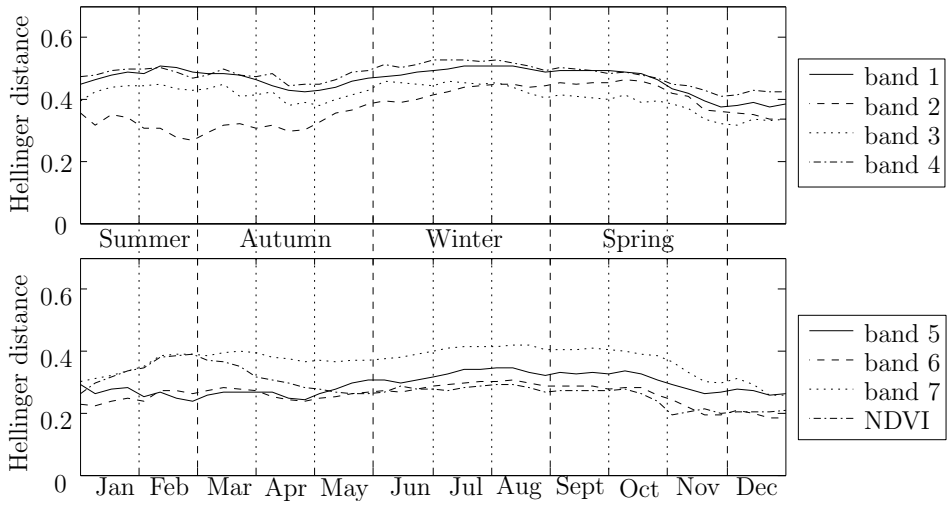


Figure 5.10: The temporal Hellinger distance between the single-band time-varying models of the vegetation and settlement classes for the Limpopo dataset.

5. It is also clear from Table 5.3 that a temporal classifier applied to the Limpopo dataset should perform better if it used data from bands $\{1,2,3,4\}$ than data from bands $\{5,6,7,\text{NDVI}\}$. The temporal classifier should produce its most accurate results if it uses data from band 1 and its most inaccurate results when it uses data from band 6.

5.1.3 CSHO model parameters

The CSHO is introduced in Section 4.1.2.2. A lot can be ascertained about the Gauteng and Limpopo datasets by looking at the parameters of the CSHO. The parameters are investigated in three different ways. Firstly, the parameter Hellinger distance between the probability density functions of the different CSHO parameters is investigated. Then a closer look is taken at the densities of $\tilde{\boldsymbol{\lambda}}$. Finally the correlation between the different parameters is investigated.

The parameter Hellinger distance between the densities of the vegetation and settlement classes for the parameter θ_i is expressed mathematically as

$$\text{HD}(f_v(\theta_i), f_s(\theta_i)),$$

where $f_c(\theta_i)$ is the marginal probability density function of $f(\tilde{\boldsymbol{\theta}})$ defined in Equation 4.11. This metric predicts which parameters would be good features to use as input to an SVM or an ANN. The parameter Hellinger distance measures the separability between the probability density functions of the parameters of the CSHO. In other words, a parameter Hellinger distance close to 1 implies that

Table 5.3: The average and maximum temporal Hellinger distance between the single-band time-varying models of the vegetation and settlement classes for the Gauteng and Limpopo datasets.

	Band							
Dataset	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**NDVI**
$\mathbb{E}[\{H^b[i]\}_{i\in\{1,\cdots,45\}}]$								
Gauteng	0.31	0.42	0.35	0.45	0.21	0.24	0.28	0.32
Limpopo	0.46	0.37	0.41	0.48	0.29	0.26	0.37	0.28
$\sup\{\{H^b[i]\}_{i\in\{1,\cdots,45\}}\}$								
Gauteng	0.52	0.57	0.44	0.54	0.34	0.36	0.38	0.42
Limpopo	0.51	0.46	0.46	0.53	0.35	0.31	0.42	0.39

the parameter is a good feature to use to differentiate between the vegetation and settlement classes, while a a parameter Hellinger distance close to 0 implies the parameter is not a good feature to use to differentiate between the vegetation and settlement classes.

To a certain extent the $\tilde{\boldsymbol{\lambda}}$ parameters measure the degree of dependence between the observations of the MODIS time-series (temporal dependence). This is true since the λ parameter of the Ornstein-Uhlenbeck process regulates the coefficient of $\eta^b[i-1]$ in Equation 4.12. The influence of the previous observation on the current observation increases as $\lambda \to 0$ and decreases as $\lambda \to \infty$.

The parameter correlation matrix $\tilde{\boldsymbol{P}}_p^c$ defined in Section 4.1.2.5 measures the correlation between the parameters of the CSHO and the parameter correlation matrix $\tilde{\boldsymbol{P}}_p^c$ is thus also a measure of spectral dependence (under the Gaussian assumption). Note that $\tilde{\boldsymbol{P}}_p^c$ implies the inclusion of NDVI, whereas $\boldsymbol{P}_p^c$ does not include NDVI. Up to now the mathematical definitions of different correlation matrices were given, $\tilde{\boldsymbol{P}}_p^c, \tilde{\boldsymbol{P}}_\eta^c$ and $\tilde{\boldsymbol{\rho}}^c$. What is lacking at this point however is a computational approach for computing a correlation matrix $\boldsymbol{R}$ from a set $\{\mathscr{R}_x\}_{x=1,2,\cdots,k}$, with $\mathscr{R}_x[i]$ being observations of $\mathscr{R}_x$ and $i = 1,2,\cdots,n$. The correlation matrix $\boldsymbol{R}$ has entries

$$r_{xy} = \frac{\sum_{i=1}^{n}(\mathscr{R}_x[i]-\bar{\mathscr{R}}_x)(\mathscr{R}_y[i]-\bar{\mathscr{R}}_y)}{\sqrt{\sum_{i=1}^{n}(\mathscr{R}_x[i]-\bar{\mathscr{R}}_x)^2\sum_{i=1}^{n}(\mathscr{R}_y[i]-\bar{\mathscr{R}}_y)^2}},\ x,y = 1,2,\cdots,k. \tag{5.3}$$

5.1.3.1 CSHO model parameters: Gauteng

The parameter Hellinger distance between the probability density functions of the parameters of the CSHO for the vegetation and settlement classes of the Gauteng dataset is displayed in Figure 5.11. The probability density functions of $\tilde{\boldsymbol{\lambda}}_v$ and $\tilde{\boldsymbol{\lambda}}_s$ are given in Figure 5.12. The matrices $\tilde{\boldsymbol{P}}_p^v$ and $\tilde{\boldsymbol{P}}_p^s$ are presented in Figure 5.13.

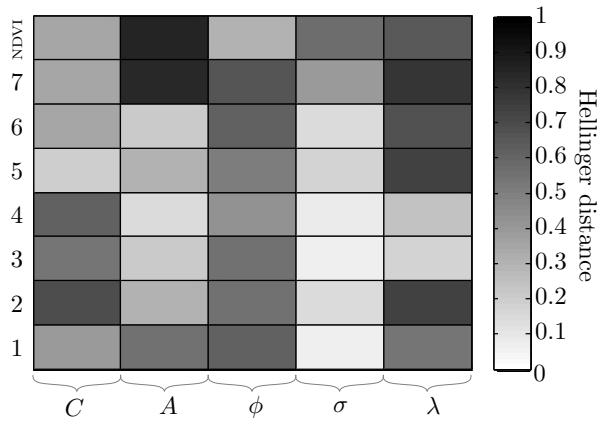
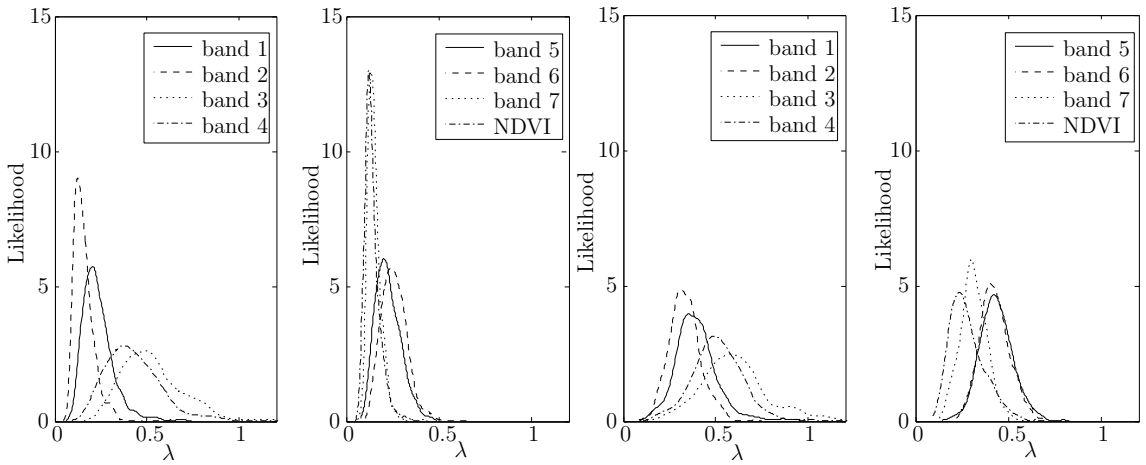


Figure 5.11: HD$(f_v(\theta_i), f_s(\theta_i))$ (Gauteng) [2] © IEEE 2012.

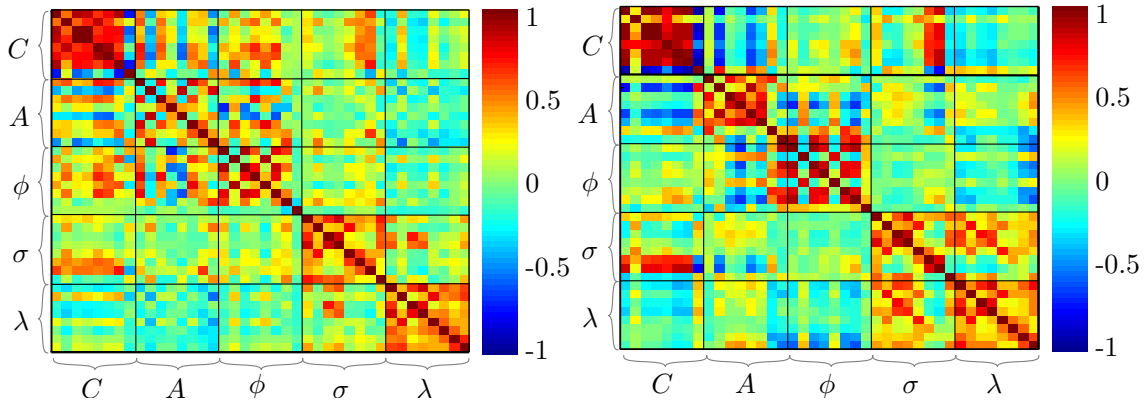


(a) $\tilde{\boldsymbol{\lambda}}_v$. **(b)** $\tilde{\boldsymbol{\lambda}}_s$.

Figure 5.12: Probability density functions of $\tilde{\boldsymbol{\lambda}}_v$ and $\tilde{\boldsymbol{\lambda}}_s$ (Gauteng).

5.1.3.2 CSHO model parameters: Limpopo

The parameter Hellinger distance between the probability density functions of the parameters of the CSHO for the vegetation and settlement classes of the Limpopo dataset is displayed in Figure 5.14.



(a) $\tilde{\boldsymbol{P}}_p^v$ [32] © IEEE 2012. **(b)** $\tilde{\boldsymbol{P}}_p^s$.

Figure 5.13: $\tilde{\boldsymbol{P}}_p^v$ and $\tilde{\boldsymbol{P}}_p^s$ (Gauteng)

The probability density functions of $\tilde{\boldsymbol{\lambda}}_v$ and $\tilde{\boldsymbol{\lambda}}_s$ for the Limpopo dataset are given in Figure 5.15. The matrices $\tilde{\boldsymbol{P}}_p^v$ and $\tilde{\boldsymbol{P}}_p^s$ are presented in Figure 5.16 for the Limpopo dataset.

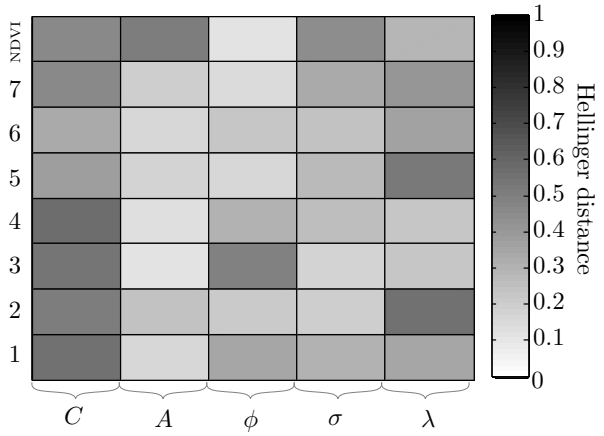
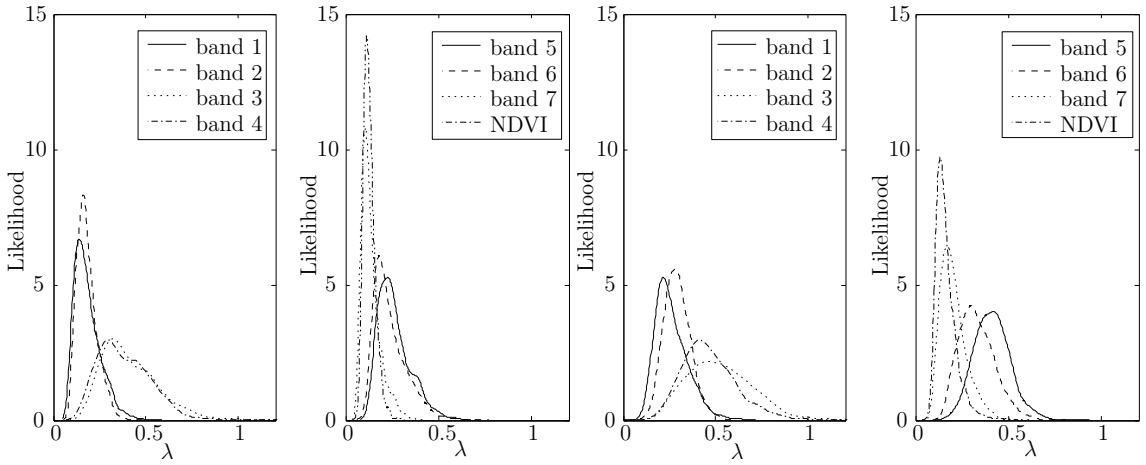


Figure 5.14: HD$(f_v(\theta_i), f_s(\theta_i))$ (Limpopo) [2] © IEEE 2012.

5.1.3.3 Discussion of CSHO model parameters

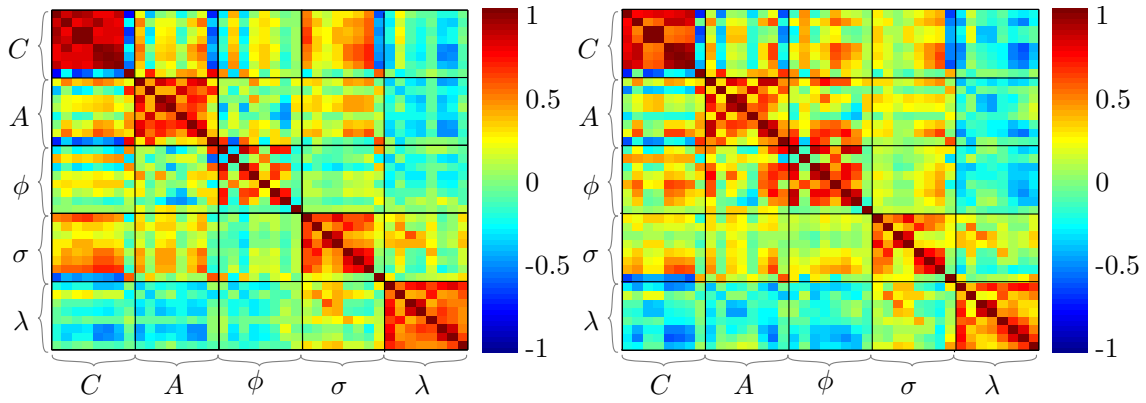
The average parameter Hellinger distances across bands and parameters are given in Table 5.4 and Table 5.5 respectively for the Gauteng and Limpopo datasets. The average values of λ for each band and class for the Gauteng and Limpopo datasets are given in Table 5.6.

The following observations and conclusion can be made from the figures and tables pertaining to the parameters of the CSHO:



(a) $\tilde{\boldsymbol{\lambda}}_v$. **(b)** $\tilde{\boldsymbol{\lambda}}_s$.

Figure 5.15: Probability density functions of $\tilde{\boldsymbol{\lambda}}_v$ and $\tilde{\boldsymbol{\lambda}}_s$ (Limpopo).



(a) $\tilde{\boldsymbol{P}}_p^v$. **(b)** $\tilde{\boldsymbol{P}}_p^v$.

Figure 5.16: $\tilde{\boldsymbol{P}}_p^v$ and $\boldsymbol{P}_p^s$ (Limpopo).

1. The average parameter Hellinger distance (across bands) can be calculated from Table 5.4 and is equal to 0.34 and 0.29 in the case of the Limpopo and Gauteng datasets. The higher average parameter Hellinger distance of the Gauteng dataset implies that the potential differentiability between the vegetation and settlement classes is higher for the Gauteng dataset than the Limpopo dataset when the parameters of the CSHO are used as classification features.

2. Based on the average parameter Hellinger distances in Table 5.4, a higher degree of differentiability between the vegetation and settlement class is possible (plausible) for the Gauteng dataset if bands $\{1,2,7,\text{NDVI}\}$ (band 7 provides the highest degree of differentiability) are used (when the parameters of the CSHO are used as classification features). The remaining

Table 5.4: Average Hellinger distance (across bands) between the densities of the parameters of the CSHO for the Gauteng and Limpopo datasets.

	Band							
Dataset	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**NDVI**
Gauteng	0.37	0.41	0.27	0.26	0.33	0.34	0.51	0.46
Limpopo	0.29	0.31	0.26	0.25	0.25	0.22	0.26	0.33

Table 5.5: Average Hellinger distance (across parameters) between the densities of the parameters of the CSHO for the Gauteng and Limpopo datasets.

	Parameter				
Dataset	C	A	ϕ	σ	λ
Gauteng	0.39	0.38	0.48	0.20	0.51
Limpopo	0.43	0.19	0.25	0.25	0.33

bands (band 4 provides the lowest degree of differentiability) provide less differentiability than $\{1,2,7,\text{NDVI}\}$). The Limpopo dataset has exactly the same division of its bands (as the Gauteng dataset). For the Limpopo dataset NDVI provides the highest amount of differentiability, while band 6 provides the lowest amount.

3. According to Table 5.5 the CSHO parameters can be sorted as follows: $\{\lambda,\phi,C,A,\sigma\}$ (by using the Hellinger distance values across parameters). The parameters are now in descending order (in terms of their potential usefulness as classification features). Similarly, for the Limpopo class the parameter list $\{C,\lambda,\phi,\sigma,A\}$ is constructed.

4. For the Gauteng dataset the following is observable when inspecting Figure 5.11: Firstly, the potential discerning capability (classification capability) of **C** (see Section 4.1.2.7) in bands 2 and 4 is high, since the parameter Hellinger distance is high in bands 2 and 4. Similarly, the potential classification capability that bands 7 and NDVI can provide in the case of **A** is also quite good. The phase parameters $\boldsymbol{\phi}$ are also theoretically capable of good class differentiability in all the MODIS land bands, while the estimated noise parameters provide good potential class discernment in bands 2, 5, 7 and NDVI. It is noteworthy to mention that generally band 5 cannot

Table 5.6: The average values of λ in each band and class for the Gauteng and Limpopo datasets.

	Band							
Average value of λ_v^b								
Gauteng	0.24	0.16	0.53	0.43	0.23	0.26	0.15	0.14
Limpopo	0.18	0.18	0.42	0.38	0.26	0.24	0.13	0.13
Average value of λ_s^b								
Gauteng	0.40	0.33	0.61	0.51	0.44	0.43	0.31	0.28
Limpopo	0.26	0.29	0.53	0.47	0.41	0.33	0.21	0.17

be used to provide good class differentiability except in the case of the mean reversion rate of the noise. Finally, the volatility of the noise provides poor potential class differentiability, because the volatility has an HD ≈ 0 in almost all of the MODIS bands.

5. According to Figure 5.14 the mean component is the best parameter to use as a classification feature, while the seasonal component would be the worst parameter to use, except in the case of NDVI (for the Limpopo dataset). However the most important result from Figure 5.14 is that the noise parameters can also be used to differentiate between classes.

6. According to Table 5.6 the average value of λ is equal to 0.34 (in the case of the Gauteng dataset), while the average value of λ is equivalent to 0.29 for the Limpopo dataset. As stated before, the higher the value of λ the less the dependence is between the observations in the dataset. Since $0.34 > 0.29$ it can be inferred that the dependence between the observations in the Gauteng dataset is less than for the Limpopo dataset.

7. For the Gauteng dataset, the lowest amount of dependence between the observations of the vegetation class is observable in bands $\{1,3,4,6\}$ (the lowest amount of dependence is observable in band 3), while the highest amount of dependence between the observations is observable in bands $\{2,5,7,\text{NDVI}\}$ (the highest amount is seen in the case of NDVI). For the settlement class, temporal dependence is more prominent in bands $\{3,4,5,6\}$ (temporal dependence is most prominent in band 3) than in bands $\{1,2,7,\text{NDVI}\}$ (The lowest amount of dependence is observable in the case of NDVI).

8. In the case of the Limpopo dataset, the lowest amount of dependence between the observations

of the vegetation class is observable in bands $\{3,4,5,6\}$ (the lowest amount of dependence is observable in band 3), while the highest amount of dependence between the observations is observable in bands $\{1,2,7,\text{NDVI}\}$ (the highest amount is seen in the case of NDVI). For the settlement class, temporal dependence is more prominent in bands $\{3,4,5,6\}$ (temporal dependence is most prominent in band 3) than in bands $\{1,2,7,\text{NDVI}\}$ (The lowest amount of dependence is observable in the case of NDVI).

9. From Figure 5.13 it is clear that for the Gauteng dataset the mean parameters $\mathbf{C}$ of the CSHO are the most correlated. Note that under a Gaussian assumption correlation implies dependence; however the terminology is not adopted here for the sake of generality. For the settlement class band 2 shows a significantly lower correlation for $\mathbf{C}$ than in the vegetation class. The amplitude parameters $\mathbf{A}$ are more correlated in the settlement class than in the vegetation class. There is also a high degree of correlation between $\boldsymbol{\lambda}$ and $\boldsymbol{\sigma}$ respectively; for the settlement class $\boldsymbol{\lambda}$ and $\boldsymbol{\sigma}$ are also highly correlated. The correlation profiles (patterns) of $\mathbf{C},\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are especially similar for the Gauteng dataset.

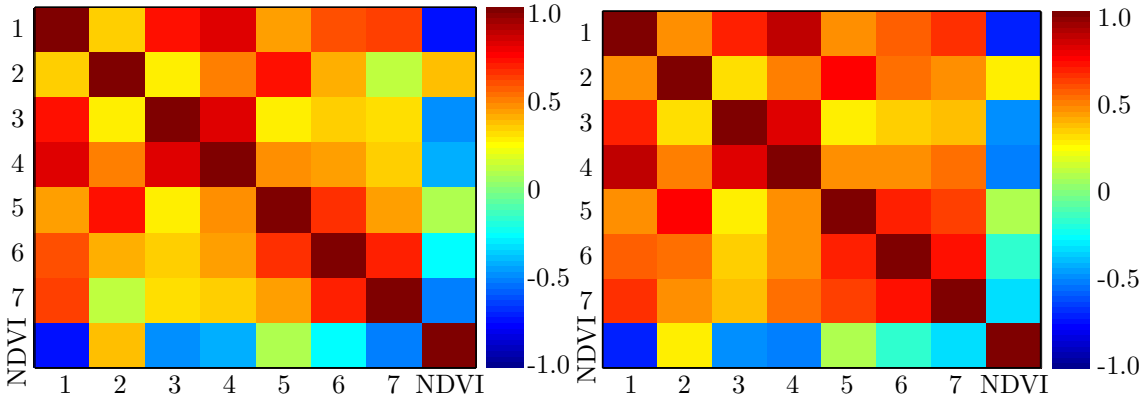
10. As can be seen in Figure 5.16, similar to the Gauteng dataset, the mean parameters of the CSHO are the most correlated. In contrast to the Gauteng dataset, the amplitude parameters $\mathbf{A}$ are also highly correlated. The correlation between $\boldsymbol{\lambda}$ and $\boldsymbol{\sigma}$ is less in the Limpopo dataset than in the Gauteng dataset. The correlation profiles of $\mathbf{C},\mathbf{A},\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are especially similar for the Limpopo dataset.

5.1.4 Noise correlation

The noise correlation matrix $\tilde{\boldsymbol{P}}_{\eta}^{c}$ is discussed in Section 4.1.2.5 and measures the degree of correlation that exists (in the noise) between the different spectral bands.

5.1.4.1 Noise correlation: Gauteng

The noise correlation matrices $\tilde{\boldsymbol{P}}_{\eta}^{v}$ and $\tilde{\boldsymbol{P}}_{\eta}^{s}$ are displayed graphically in Figure 5.17 for the Gauteng dataset.

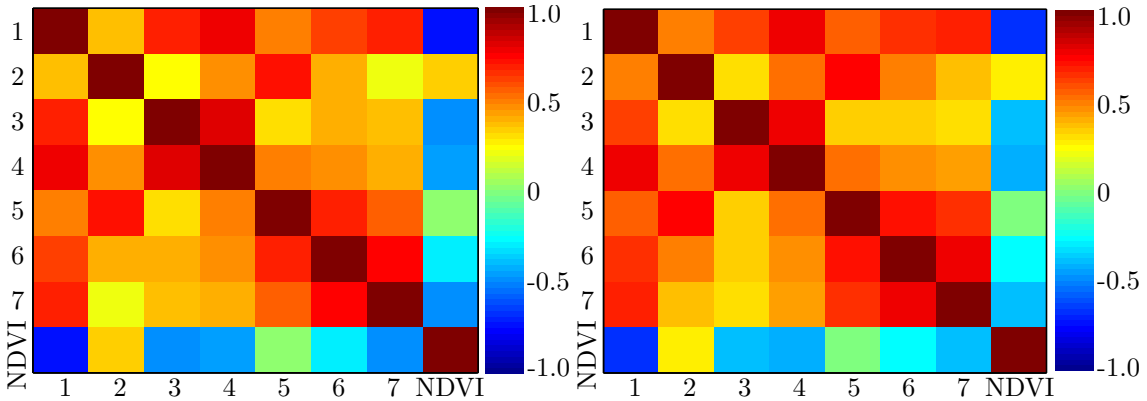


(a) $\tilde{\boldsymbol{P}}_{\eta}^{v}$ [32] © IEEE 2012. **(b)** $\tilde{\boldsymbol{P}}_{\eta}^{s}$.

Figure 5.17: $\tilde{\boldsymbol{P}}_{\eta}^{v}$ and $\tilde{\boldsymbol{P}}_{\eta}^{s}$ (Gauteng).

5.1.4.2 Noise correlation: Limpopo

The noise correlation matrices $\tilde{\boldsymbol{P}}_{\eta}^{v}$ and $\tilde{\boldsymbol{P}}_{\eta}^{s}$ are displayed graphically in Figure 5.18 for the Limpopo dataset.



(a) $\tilde{\boldsymbol{P}}_{\eta}^{v}$. **(b)** $\tilde{\boldsymbol{P}}_{\eta}^{s}$.

Figure 5.18: $\tilde{\boldsymbol{P}}_{\eta}^{v}$ and $\tilde{\boldsymbol{P}}_{\eta}^{s}$ (Limpopo).

5.1.4.3 Discussion of noise correlation

The following is conspicuous when inspecting Figures 5.17 and 5.18:

1. The parameter correlation matrices and the noise correlation matrices have similar profiles (patterns), especially in the case of the noise parameters of the CSHO.

2. The noise correlation matrices of the different classes are very similar. On average the most correlated two-band pairs are $\{1,3\},\{1,4\},\{1,6\},\{1,7\}$, $\{2,5\},\{3,4\},\{5,6\},\{6,7\}$.

3. On average the least correlated two-band pairs are $\{1,2\},\{2,3\},\{2,6\},\{2,7\}$, $\{3,5\},\{3,6\},\{3,7\},\{4,6\}$.

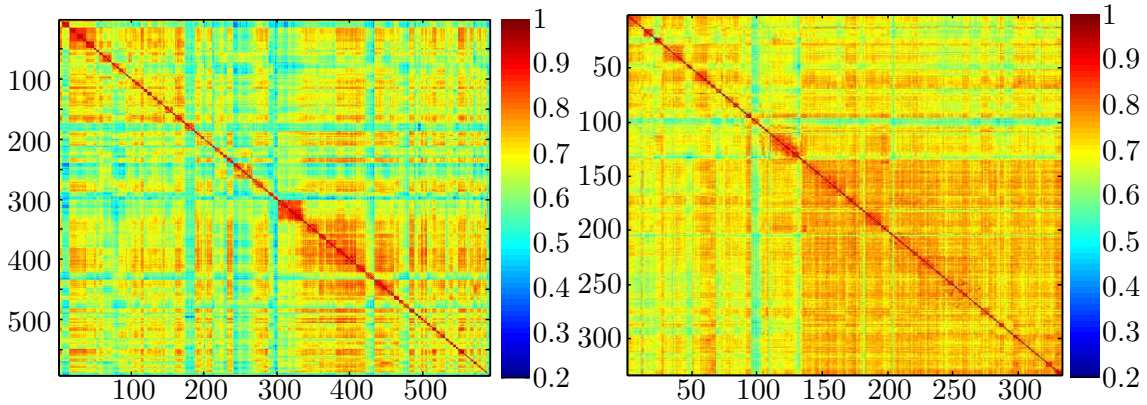
4. The following band pairs vary most between classes and datasets: $\{1,2\},\{2,6\},\{3,7\},\{5,7\}$.

5.1.5 Spatial correlation

The average spatial correlation matrix $\tilde{\boldsymbol{\rho}}^c$ is defined in Equation 4.7 and measures the average spatial correlation that exists between the pixels of a specific class. When inspecting Equation 4.7 notice that the spatial correlation is calculated by computing the average correlation between pixels and not via Euclidean distances. According to Section 4.3.3, one of the reasons that CUSUM's optimality cannot be guaranteed is because of spatial correlation (see Section 5.4.4 for more details).

5.1.5.1 Spatial correlation: Gauteng

The matrices $\tilde{\boldsymbol{\rho}}^v$ and $\tilde{\boldsymbol{\rho}}^s$ are displayed in Figure 5.19 for the Gauteng dataset.

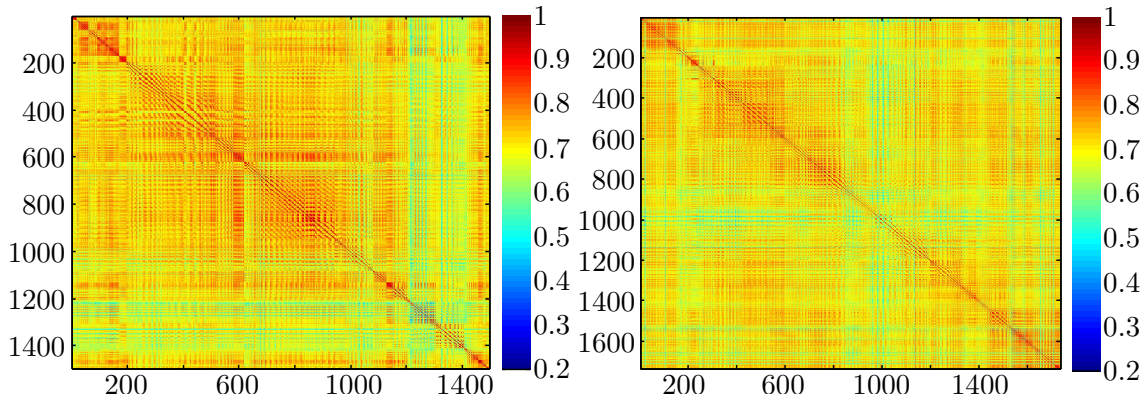


(a) $\tilde{\boldsymbol{\rho}}^v$. **(b)** $\tilde{\boldsymbol{\rho}}^s$.

Figure 5.19: $\tilde{\boldsymbol{\rho}}^v$ and $\tilde{\boldsymbol{\rho}}^s$ (Gauteng)

5.1.5.2 Spatial correlation: Limpopo

The matrices $\tilde{\boldsymbol{\rho}}^v$ and $\tilde{\boldsymbol{\rho}}^s$ are presented graphically in Figure 5.20 for the Limpopo dataset.



(a) $\tilde{\boldsymbol{\rho}}^v$. **(b)** $\tilde{\boldsymbol{\rho}}^s$.

Figure 5.20: $\tilde{\boldsymbol{\rho}}^v$ and $\tilde{\boldsymbol{\rho}}^s$ (Limpopo).

5.1.5.3 Discussion of spatial correlation

The following observations can be made about Figures 5.19 and 5.20:

1. In the case of the Gauteng dataset, the average spatial correlation (computed by taking the average of $\tilde{\boldsymbol{\rho}}^c$) is equivalent to 0.66 and 0.71 for the vegetation and settlement class respectively. In the case of the Limpopo dataset the average spatial correlation is equal to 0.71 and 0.70 for the vegetation and settlement class respectively.

2. By using Figures 5.19 and 5.20 as a visual aid, it can be confirmed that on average the spatial correlation is higher in the Limpopo dataset than in the Gauteng dataset.

5.2 SIMULATOR RESULTS: GAUTENG AND LIMPOPO

The algorithm for simulating a MODIS pixel is discussed in detail in Section 4.1.2.7. As mentioned in Chapter 1, the main purpose of the inductive simulator is to augment datasets for the data-intensive sequential algorithms, especially CUSUM. The inductive simulator is used together with the CUSUM algorithm in Section 5.4.4. Figure 5.21 contains a true vegetation pixel from the Gauteng dataset and its recreated (simulated) counterpart.

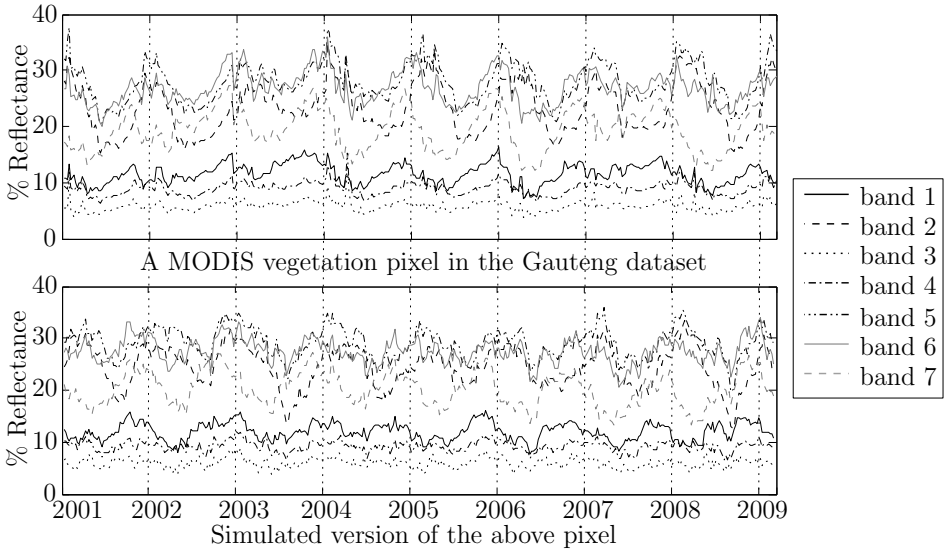


Figure 5.21: A true vegetation pixel belonging to the Gauteng dataset as well as its recreated (simulated) counterpart.

The following observations can be made from Figure 5.21:

1. From the real vegetation pixel graph it is clear that there is correlation between spectral bands (if the reflectance goes up in band 1 it usually also goes up in band 4) and spectral dependence (consider, for instance the mean of each spectral band).

2. The simulated pixel does not replicate the temporal behaviour exactly.

3. The long-term mean and seasonal components of the real and simulated pixels are however similar.

4. The real vegetation pixel has increment difference outliers, while the simulated pixel does not.

As mentioned in Chapter 4, the spectral signature for each class is encapsulated by $f(\boldsymbol{\theta}_c)$ (see Section 4.1.2.4) and $\boldsymbol{P}_{\eta}^{c}$ (see Section 4.1.2.5). The metrics introduced in this section are used to determine if the class signature is replicated adequately by the simulator.

The experimental procedure used to validate the simulator is discussed in Section 5.2.1. The different metrics and the reason for selecting each metric are discussed in Section 5.2.3. There are two main metric types. The details of each type are given in Section 5.2.4 and Section 5.2.5 respectively. The results obtained via the experimental procedure discussed in Section 5.2.1 are presented in

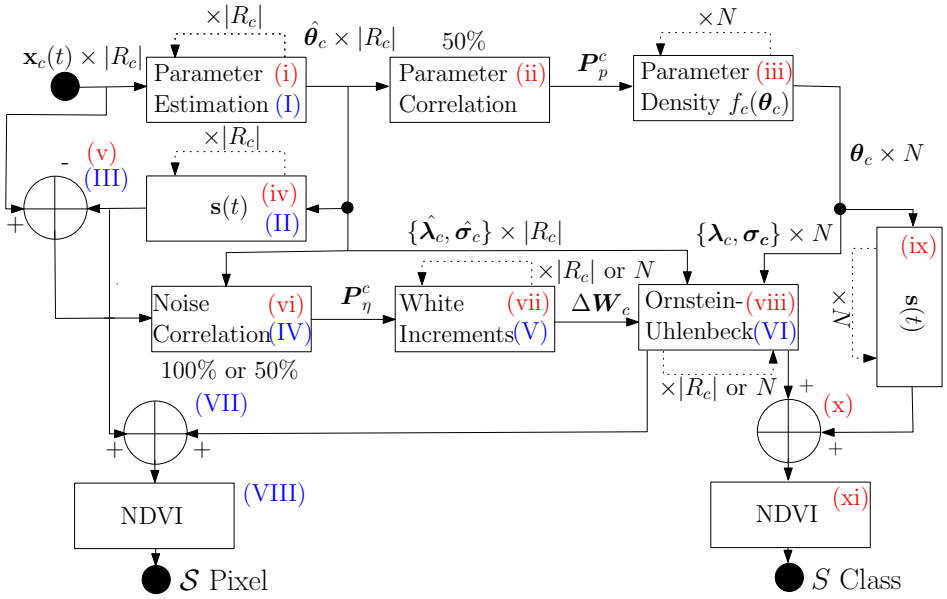


Figure 5.22: Flow diagram illustrating how S and $\mathscr{S}$ are generated. When there are two possibilities at a block, the first option relates to the generation of the pixel set $\mathscr{S}$, while the second option is used to create the class set S. The capital roman numerals are the steps needed to create the pixel set $\mathscr{S}$, while the lowercase roman numerals are the steps required to create the class set S [32] © IEEE 2012.

Section 5.2.6.

5.2.1 Simulator validation

The simulator is validated by using *class* and *pixel* metrics. The class metrics are used to determine whether the simulated dataset has the same statistical attributes as the original dataset and are important, since class attributes are used by classifiers to distinguish between classes [2, 23]. The pixel metrics, in contrast, are used to verify that the simulator can also reproduce any given pixel accurately by comparing every real world pixel to its simulated counterpart. The construction procedures of the simulated datasets to which the two types of metrics are applied differ and are illustrated in Figure 5.22. The construction procedures are an extension of the algorithm presented in Section 4.1.2.7.

The steps required to generate the dataset S ($|S| = N$) to which the class metrics are applied are summarised below:

i Estimate the parameters of R_c (all the pixels in R belonging to class c).

ii Select a random 50% of the estimated parameters to construct $\boldsymbol{P}_p^c$ (discussed in Section 4.1.2.5). The pixels associated with the selected parameters form the training set. The remaining pixels in R_c belong to the validation set.

iii Create $f_c(\boldsymbol{\theta}_c)$ from $\boldsymbol{P}_p^c$ (actually the parameters are used directly) using Equation 4.11 and draw $N \times \boldsymbol{\theta}_c$ from it.

iv Calculate $\mathbf{s}(t)$ with Equation 4.1, by using the harmonic parameters of step i.

v Determine the residual by subtracting $\mathbf{s}(t)$ from $\mathbf{x}(t)$.

vi Compute $\boldsymbol{P}_\eta^c$ from the residual, by using the same training set as in step ii and Equation 4.12.

vii Calculate N time-series of correlated increments $\boldsymbol{\Delta W}_c$ using $\boldsymbol{P}_\eta^c$, Equation 4.13 and Equation 4.15.

viii Generate correlated noise by using the noise parameters of $\boldsymbol{\theta}_c$ (drawn in step iii), $\boldsymbol{\Delta W}_c[i]$ and Equation 4.12.

ix Create the simulated harmonic component by using the harmonic parameters of $\boldsymbol{\theta}_c$ and Equation 4.1.

x Add the correlated noise to the harmonic component.

xi Generate NDVI from the simulated data by using band 1 and 2.

The pixel metrics simulated dataset $\mathscr{S}$ is constructed by using a different approach. The steps required to generate the dataset $\mathscr{S}$ ($|\mathscr{S}| = |R_c|$) to which the pixel metrics are applied are summarised below:

I-III. Follow steps i, iv and v of the class generation algorithm.

IV. Execute step vi of the class generation algorithm, but use all of the pixels in R_c.

V. Perform step vii of the class generation algorithm, but generate $|R_c|$ time-series instead of N.

VI. Generate correlated noise by using the estimated noise parameters derived in step i instead of the noise parameters of $\boldsymbol{\theta}_c$.

VII. Add the correlated noise to the harmonic signal generated in step ii.

VIII. Generate NDVI from the simulated data by using bands 1 and 2.

5.2.2 Preliminary validation results

At this point it would be useful to provide initial visual evidence of the validity and usefulness of the CSHO simulator. In Figure 5.23 the temporal Hellinger distance between the single-band time-varying models of S_v and S_s are displayed for the Gauteng dataset. The datasets S_v and S_s have 1000 elements and are single random instances generated with a modified version of the class metric simulated dataset algorithm presented in Section 5.2.1. The modification entails that all the data of R_c are used (not only 50%) as training data. The subscripts v and s are only added to S if their absence causes ambiguity.

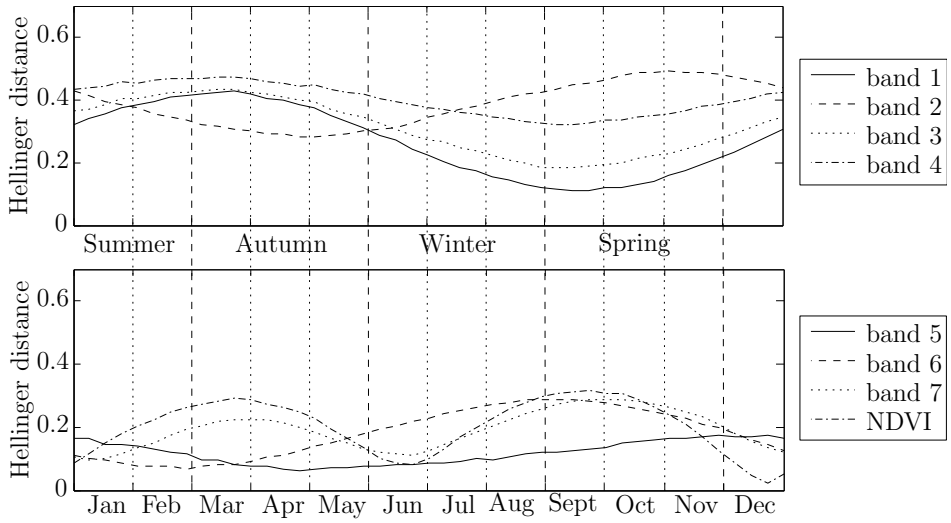


Figure 5.23: The temporal Hellinger distance between the single-band time-varying models of S_v and S_s (Gauteng).

Another important visual aid is Figure 5.24, which displays the temporal Hellinger distance between the single-band time-varying models of R_v and S for the Gauteng dataset. To avoid repetition the other combinations of R_c and S are not displayed. Figure 5.24 is very important, as it shows how accurately the simulated dataset replicated the original dataset in terms of the yearly average temporal behaviour.

Similar to Figure 5.23, Figure 5.25 displays the temporal Hellinger distance between the single-band

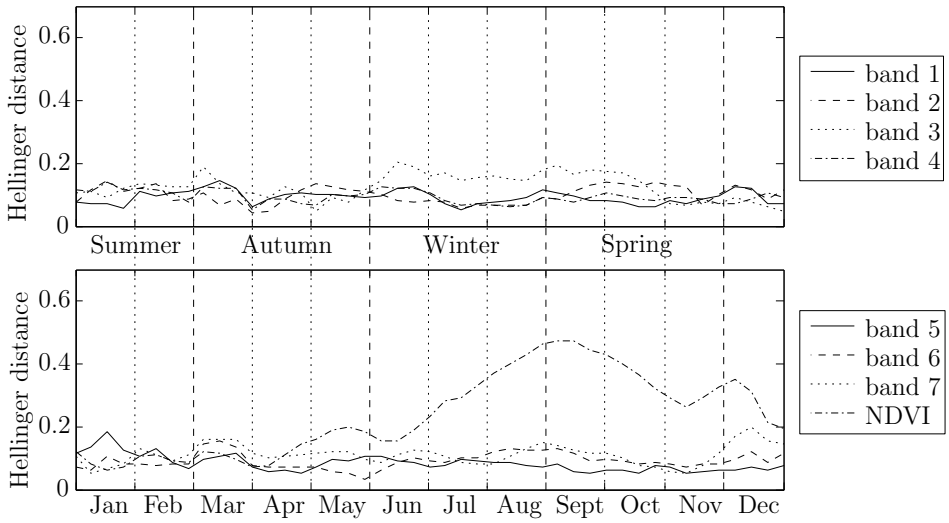


Figure 5.24: The temporal Hellinger distance between the single-band time-varying models of R_v and S (Gauteng).

time-varying models of S_v and S_s for the Limpopo dataset.

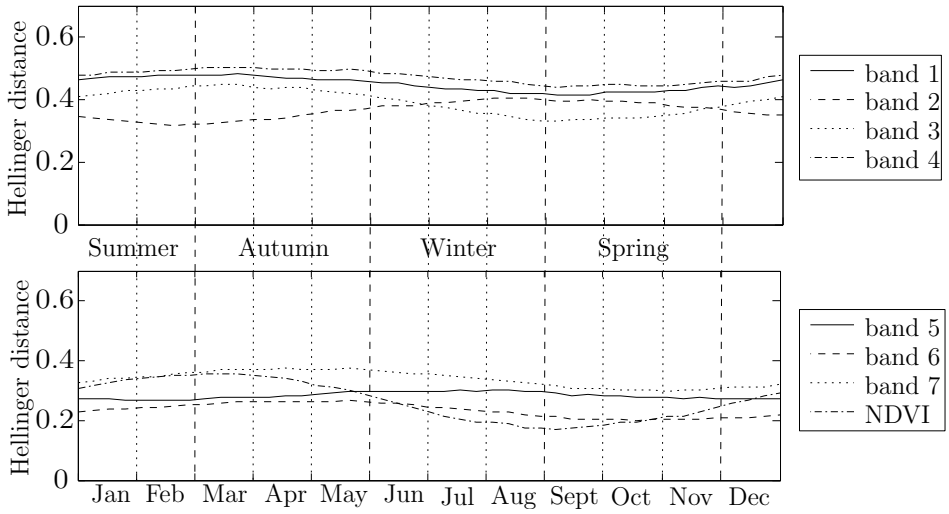


Figure 5.25: The temporal Hellinger distance between the single-band time-varying models of S_v and S_s (Limpopo).

The following observations can be made when inspecting Figure 5.23, Figure 5.24 and Figure 5.25:

1. The curves in Figure 5.23 and Figure 5.9 have similar profiles. The same can be said for Figure 5.25 and Figure 5.10.

2. The Hellinger distance for most bands in Figure 5.24 stay below 0.2. The exception is NDVI, which stays below 0.5. For most of the year however the Hellinger distance of NDVI is close to zero. The small Hellinger distance implies that the yearly temporal behaviour is replicated sufficiently by the CSHO simulator. The other possible combinations of R_c and S (although not presented) exhibit similar behaviour than that of Figure 5.24. Simular results are obtained for the Limpopo dataset.

5.2.3 Discussion of metric selection

The metrics in this section are based on the metrics proposed in [199]. Two underlying metrics are used, namely the Sum of Squared Error (SSE) and Hellinger distance (except for the power spectral density metric that measures power). In both cases a value *close to zero is desirable*. When "Hellinger" is not part of the metric name, it indicates that the SSE was used as the base metric. Each metric was chosen to verify that the simulator reproduces three important characteristics, namely temporal dynamics, spectral behaviour and accurate noise.

5.2.3.1 Temporal dynamics

There are two types of temporal dynamics to account for, namely intra-annual and inter-annual variation. The main reason for intra-annual variation is due to seasonality, which is caused by a wide range of factors including plant phenology. Inter-annual variation can be caused by many factors, including a drought or a flood. The *total model error* metric is a first-order statistic and is used to verify whether the average seasonal behaviour is replicated correctly. The *average temporal Hellinger distance* is probably the most important metric from the perspective of Section 4.2.3 and Section 4.3.3, as it measures the difference between the first-order statistical description of the CSHO and the true dataset. The average temporal Hellinger distance therefore measures whether the CSHO sufficiently replicates intra-annual variation and how effective the CSHO is in compensating for inter-annual variation. The *autocorrelation* metric is a second order statistic which measures whether the CSHO also models the temporal behaviour of any given pixel properly.

5.2.3.2 Spectral behaviour

As discussed in Section 4.1.2, the main aim of the simulator is to replicate spectral behaviour. It is well know that each class has a unique spectral signature within a certain allowable margin of varia-

tion [39]. The proposed simulator encapsulates and models the spectral signature for each class by using Equation 4.11. Equation 4.11 enforces the class-specific statistical restrictions imposed by the different CSHO model parameters of each spectral band on one another. The *parameter correlation metric* measures how effective the simulator is in reproducing spectral dependence (under the assumption of Gaussianity), while the *average parameter Hellinger distance* measures how trustworthy the joint Gaussian assumption of $f_c(\tilde{\boldsymbol{\theta}}_c)$ is.

Furthermore, the model also enforces noise correlation by using the approach presented in Sections 4.1.2.5 and 4.1.2.6. The *noise correlation metric* measures how well the noise correlation is modelled.

5.2.3.3 Accurate noise

A widely used assumption for remotely sensed time-series noise is that it is white [74, 78] if all information-carrying frequency components have been extracted [31]. The different *power spectral density* metric values reveal whether a white or coloured assumption is more appropriate when using an SHO as the underlying deterministic model. The *average noise increment Hellinger distance* determines whether the noise increments of each pixel are similar to the increments of the Ornstein-Uhlenbeck process.

5.2.4 Class metrics

The total model error, the average parameter Hellinger distance, the parameter and noise correlation and the average temporal Hellinger distance are respectively discussed in Section 5.2.4.1, Section 5.2.4.2, Section 5.2.4.3 and Section 5.2.4.4. A few figures are presented in this section to aid the reader in understanding the different class metrics. These figures were generated by comparing mostly R_v and S. The dataset S used by the figures is constructed by using the same procedure detailed in Section 5.2.1. Again, to avoid repetition, only the graphs for the Gauteng vegetation class are presented.

5.2.4.1 Total model error

The equation for the total model error is given by

$$\int_0^I \|\tilde{\mathbf{y}}_c^{R_c}(t) - \tilde{\mathbf{y}}_c^{S}(t)\|_2^2 dt, \tag{5.4}$$

where $\tilde{\mathbf{y}}_c(t)$ is the yearly ensemble mean of c and is defined in Equation 4.5. To determine the SSE of each time step in the year Equation 5.4 needs to be divided by 45. To give some insight into the total model error metric, Figure 5.26 presents the yearly ensemble means of R^v and S for the Gauteng dataset.

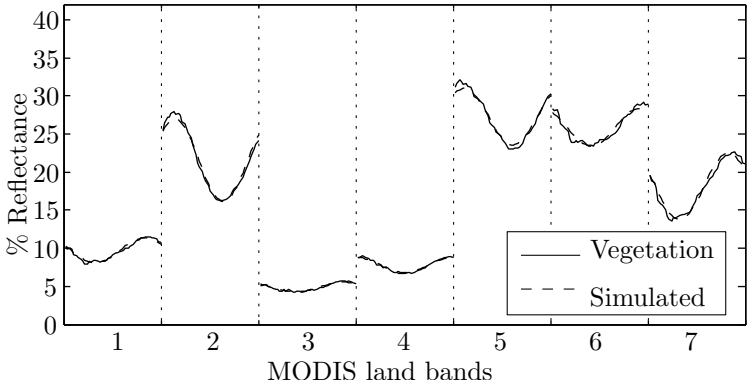


Figure 5.26: The yearly ensemble mean of R^v and S (Gauteng).

It is clear from Figure 5.26 that the total model error between R^v and S (for Gauteng) is close to zero.

5.2.4.2 Average parameter Hellinger distance metric

The equation for the average parameter Hellinger distance is equal to

$$\frac{1}{|\tilde{\boldsymbol{\theta}}_c|} \sum_{k=1}^{|\tilde{\boldsymbol{\theta}}_c|} \mathrm{HD}(f_c^{R_c}(\theta_k), f_c^{S}(\theta_k)),$$

where $f_c(\theta_k)$ is the marginal probability density function of $f_c(\tilde{\boldsymbol{\theta}}_c)$ and $\mathrm{HD}(f_c^{R_c}(\theta_k), f_c^{S}(\theta_k))$ represents the Hellinger distance between $f_c^{R_c}(\theta_k)$ and $f_c^{S}(\theta_i)$. Figure 5.27 ought to make the definition of the average parameter Hellinger distance clearer. According to Figure 5.27 the parameter Hellinger distance is greatest for **A**.

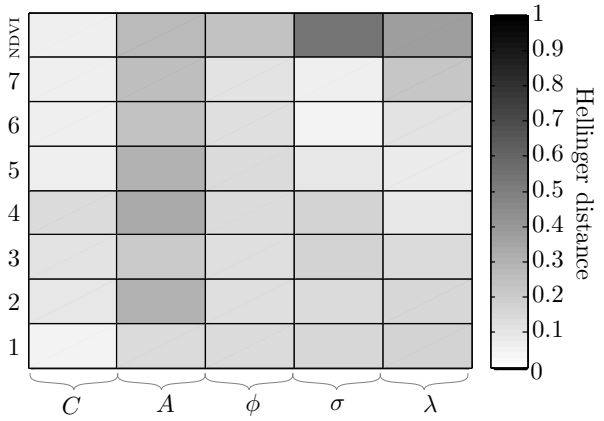


Figure 5.27: $\mathrm{HD}(f_v^{R_v}(\theta_i), f_v^{S}(\theta_i))$ (Gauteng).

5.2.4.3 Parameter and noise correlation metrics

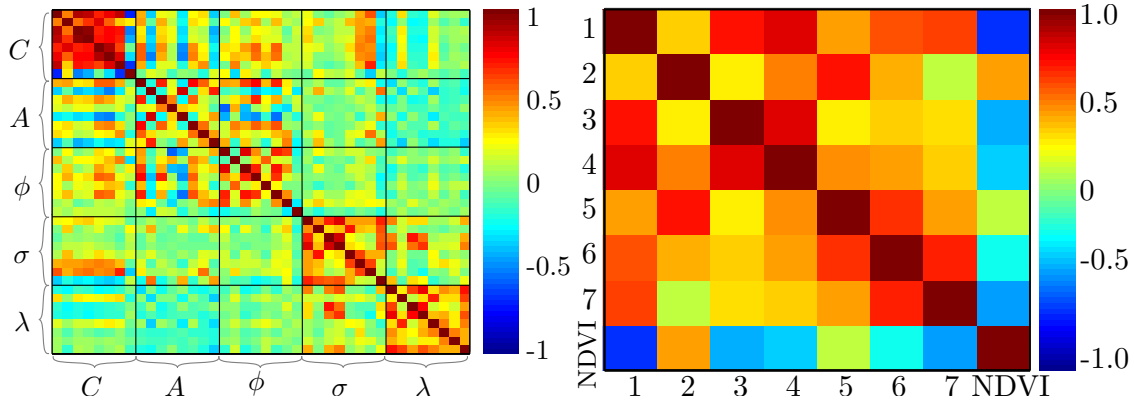
The equations for the noise and parameter correlation metrics are given by

$$\|\tilde{\boldsymbol{P}}^c_{p_{R_c}} - \tilde{\boldsymbol{P}}^c_{p_S}\|_2^2$$

and

$$\|\tilde{\boldsymbol{P}}^c_{\eta_{R_c}} - \tilde{\boldsymbol{P}}^c_{\eta_S}\|_2^2.$$

The noise correlation metric needs to be divided by 8×8 (NDVI was added for completeness), while the parameter correlation metric needs to be divided by 40×40 to determine the average SSE. For convenience the matrices $\tilde{\boldsymbol{P}}^v_{p_S}$ and $\tilde{\boldsymbol{P}}^v_{\eta_S}$ are displayed in Figure 5.28. There is not much difference between Figure 5.28a and Figure 5.13a. The same goes for Figure 5.28b and Figure 5.17a.



(a) $\tilde{\boldsymbol{P}}^v_{p_S}$. **(b)** $\tilde{\boldsymbol{P}}^v_{\eta_S}$.

Figure 5.28: $\tilde{\boldsymbol{P}}^v_{p_S}$ and $\tilde{\boldsymbol{P}}^v_{\eta_S}$ (Gauteng).

5.2.4.4 Average temporal Hellinger distance metric

The equation for the average temporal Hellinger distance is expressed as

$$\frac{1}{8}\sum_{b=1}^{\text{NDVI}}\frac{1}{I}\int_0^I \text{HD}(f^{R_c}_{x_c^b(t)}, f^{S}_{x_c^b(t)})\,dt,$$

where $f_{x_c^b(t)}$ is the probability density function in band b at time step t. Figure 5.29 and Figure 5.30 are visual aids to help explain the average temporal Hellinger distance metric. The reflectance probability density functions at time step 3 (of 45) for all eight years in MODIS band 1 for R_v and S (Gauteng) are displayed in Figure 5.29. The probability functions of S seem to be symmetrical about the mean (and almost identical) of the mean values of the densities of R_v. The temporal Hellinger distance for time steps $3+45n,\ n=\{0,\cdots,8\}$, in band 1 is calculated by determining the Hellinger distance between the probability density functions in Figure 5.29. The Hellinger distance between the probability density functions of the observation time steps of MODIS band 4 for R_v and S (Gauteng) is displayed in Figure 5.30. The mean and variance for the curve in Figure 5.30 are respectively 0.1926 and 0.0051, which is closer to zero than one.

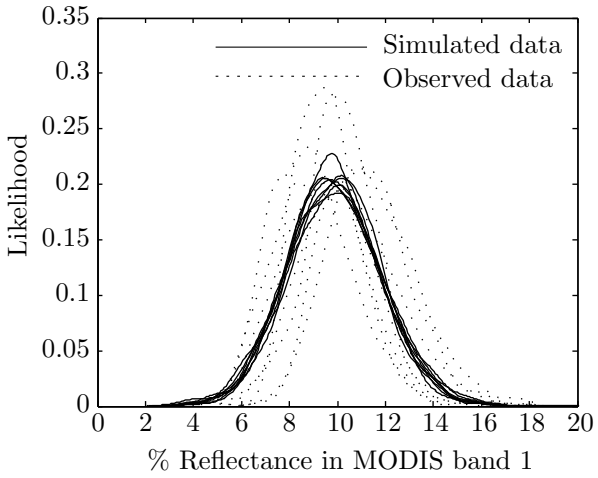


Figure 5.29: The reflectance probability density functions at time step period 3 (of 45) for all eight years in MODIS band 1 for R_v and S (Gauteng).

5.2.5 Pixel metrics

The autocorrelation metric, the average noise increment Hellinger distance and the power spectral density metric are respectively discussed in Section 5.2.5.1, Section 5.2.5.2 and Section 5.2.5.3. A few figures are displayed in this section to help the reader comprehend the different pixel metrics.

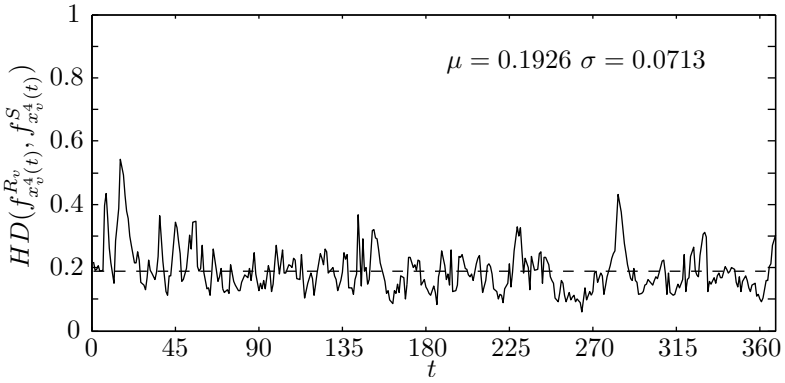


Figure 5.30: The Hellinger distance between the probability density functions of the observation time steps of MODIS band 4 for R_v and S (Gauteng).

These figures were created by comparing mostly random pixels from R_v and $\mathscr{S}$. The dataset $\mathscr{S}$ used to create the figures is constructed by using the same procedure described in Section 5.2.1. Again, to avoid repetition, only the graphs for the Gauteng vegetation class are presented.

5.2.5.1 Autocorrelation metric

The autocorrelation metric is expressed mathematically as

$$\frac{1}{|R_c|}\sum_{p=1}^{|R_c|}\int_0^I \|\tilde{\mathbf{R}}_c^{R_c(p)}(\tau) - \tilde{\mathbf{R}}_c^{\mathscr{S}(p)}(\tau)\|_2^2 \, d\tau, \tag{5.5}$$

where $\tilde{\mathbf{R}}_c^{R_c(p)}$ is the autocorrelation (defined in Equation 4.6) of the p-th pixel in R_c, while $\tilde{\mathbf{R}}_c^{S(p)}$ is defined similarly. To determine the average SSE per lag value Equation 5.5 needs to be divided by 368 (number of observations). Figure 5.31 displays the curves of a random vegetation pixel in Gauteng, as well as its replicated counterpart (belonging to $\mathscr{S}$). Clearly the two curves in Figure 5.31 are very similar.

5.2.5.2 Average noise increment Hellinger distance metric

The equation for the average noise increment Hellinger distance is defined as

$$\frac{1}{8}\sum_{b=1}^{\text{NDVI}}\frac{1}{|R_c|}\sum_{p=1}^{|R_c|}\text{HD}(f_{\Delta\eta^b}^{R_c(p)}, f_{\Delta\eta^b}^{\mathscr{S}(p)}),$$

where $f_{\Delta\eta^b}^{R_c(p)}$ is the density function of the noise increments $\eta^b[t+1]-\eta^b[t]$ for pixel p in dataset R_c, while $f_{\Delta\eta^b}^{\mathscr{S}(p)}$ is defined similarly.

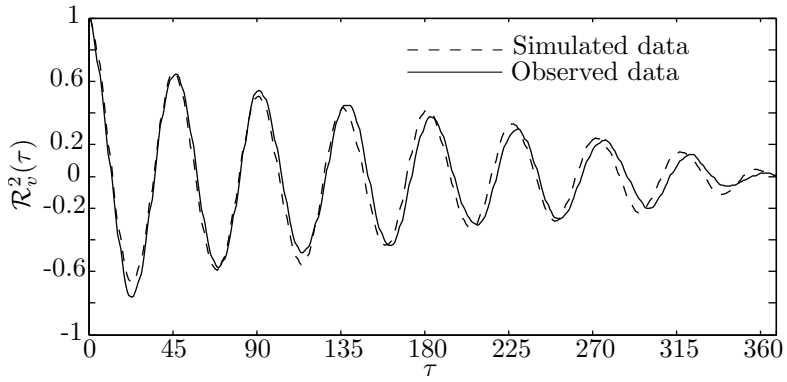


Figure 5.31: The autocorrelation function of a real and simulated vegetation pixel in band 2 (Gauteng).

5.2.5.3 Power spectral density metric

The equation for the power spectral density metric is equal to

$$\frac{1}{8}\sum_{b=1}^{\text{NDVI}}\frac{1}{|R_c|}\sum_{p=1}^{|R_c|}\int_0^{0.1} D_b^{R_c^\eta(p)}(f)\,df,$$

where $D_b^{R_c^\eta(p)}(f)$ is the power spectral density of the estimated noise of pixel p in dataset R_c in band b. The same metric can be applied to $\mathscr{S}^\eta$ and W^η, where $\mathscr{S}^\eta$ is the coloured noise (Ornstein-Uhlenbeck) representation of R_c^η (R_c after subtracting the SHO) and W^η is the white noise representation of R_c^η. Figure 5.32 displays $D_2^{R_v^\eta(p)}(f)$, $D_2^{\mathscr{S}^\eta(p)}(f)$ and $D_2^{W^\eta(p)}(f)$ for a random vegetation pixel p in the Gauteng dataset. From Figure 5.32 it is clear that $\int_0^{0.1} D_2^{R_v^\eta(p)}(f)\,df \approx \int_0^{0.1} D_2^{\mathscr{S}^\eta(p)}(f)\,df$, while $\int_0^{0.1} D_2^{R_v^\eta(p)}(f)\,df \neq \int_0^{0.1} D_2^{W^\eta(p)}(f)\,df$.

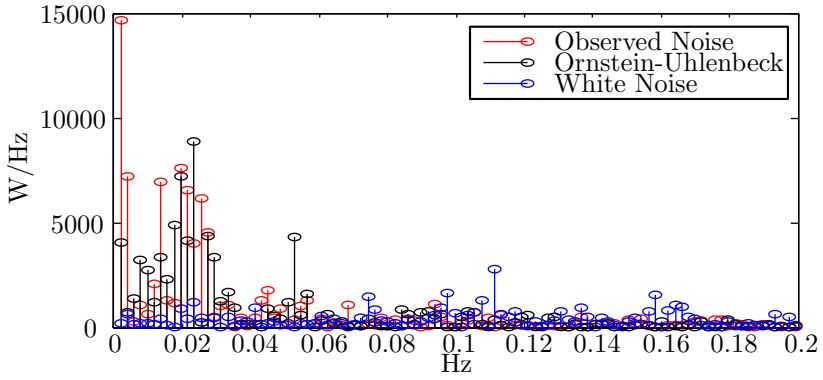


Figure 5.32: The power spectral density of the estimated noise of a vegetation pixel in MODIS band 2 along with its coloured and white representations.

5.2.6 Discussion of simulator results

Random split cross-validation was performed to create 50 different class metric simulated data sets, with the algorithm presented in Section 5.2.1 and $N = 1000$. No cross-validation was used for the pixel metrics, since all the pixels in R_c were used. For the pixel metrics 50 independent experiments were also conducted. Each data set for each experiment was independently created via the algorithm in Section 5.2.1. The class and pixel metrics were then applied to each experiment to produce the results in Table 5.7 and Table 5.8.

Table 5.7: Difference metrics between R_c, S and $\mathscr{S}$ (50 experiments) for the Gauteng dataset. The v index indicates vegetation, while the s index indicates settlement.

Metric Name	R_v	R_s
Model Error	58.3917 ± 8.1991	52.8838 ± 22.9432
Parameter Hellinger Distance	0.1817 ± 0.0070	0.2563 ± 0.0081
Parameter Correlation	19.6224 ± 3.6121	32.5042 ± 5.7380
Noise Correlation	0.0371 ± 0.0085	0.0629 ± 0.0101
Temporal Hellinger Distance	0.2349 ± 0.0042	0.2266 ± 0.0094
Autocorrelation	31.0137 ± 0.2431	34.3992 ± 0.4749
Noise Hellinger Distance	0.1674 ± 0.0003	0.1755 ± 0.0003
Power in R_η^c	92.7649	38.3202
Power in S_η	94.6440 ± 0.6564	46.0178 ± 0.2901
Power in W_η	22.9897 ± 0.0642	11.7971 ± 0.0542

The following observations can be made from Table 5.7 and Table 5.8:

1. Relative to the definitions of the SSE and Helinger distance metrics, the results in Table 5.7 and Table 5.8 are close to zero, implying that the simulator accurately replicates the temporal dynamics and spectral characteristics of the MODIS datasets. The small variances in Table 5.7 and Table 5.8 imply that the metric results are stable and reliable.

2. Relative to the other classes the Gauteng settlement class has a higher standard deviation on its metric results, which can be explained by the fact that the Gauteng settlement dataset is much smaller than the other datasets.

Table 5.8: Difference metrics between R_c, S and $\mathscr{S}$ (50 experiments) for the Limpopo dataset. The v index indicates vegetation, while the s index indicates settlement.

Metric Name	R_v	R_s
Model Error	68.5056 ± 11.7454	56.8919 ± 11.3777
Parameter Hellinger Distance	0.2045 ± 0.0054	0.2097 ± 0.0041
Parameter Correlation	15.4586 ± 2.6663	12.9429 ± 1.8421
Noise Correlation	0.0422 ± 0.0063	0.0421 ± 0.0064
Temporal Hellinger Distance	0.2147 ± 0.0047	0.1939 ± 0.0033
Autocorrelation	31.6738 ± 0.1975	30.4443 ± 0.1531
Noise Hellinger Distance	0.1630 ± 0.0002	0.1582 ± 0.0002
Power in R_η^c	110.6438	80.5593
Power in S_η	111.0334 ± 0.45182	85.4707 ± 0.3112
Power in W_η	26.8485 ± 0.0513	21.7033 ± 0.0368

3. Relative to the metric definitions the autocorrelation metric has the largest value, which is understandable since the non-stationarity that is present in the MODIS data shows up in the autocorrelation metric. The error incurred due to parameter estimation also affects the autocorrelation metric.

4. For the power metric in general $R_\eta^c \approx S_\eta$, while $R_\eta^c \neq W_\eta$, which implies that a coloured noise model is more appropriate for the current datasets.

5. There is no standard deviation for the power in R_η^c as the power of R_η^c obviously only needs to be calculated once.

5.3 CLASSIFICATION RESULTS: GAUTENG AND LIMPOPO

The focus of this section is on the classification performance of the noise-harmonic feature group $\boldsymbol{\theta}$ derived from Equation 4.4 and discussed in Section 4.2.4.2. The feature group $\boldsymbol{\theta}$ under investigation extends the feature group $\boldsymbol{\iota}$ proposed in [5]. The feature group $\boldsymbol{\theta}$ is also compared with the temporal feature group $\boldsymbol{\zeta}$, which is discussed in Section 4.2.4.2 [15]. The different feature groups are used as inputs to SVM classifiers (discussed in Section 4.2.4). The SVM classifiers are also compared

with two benchmarking techniques, namely the minimum distance classifier and the time-varying maximum likelihood classifier, which are discussed in Section 4.2.2 and Section 4.2.3 respectively. The time-varying maximum likelihood classifier is especially important, as it is based on sequential analysis. This section starts by introducing the different classification accuracy metrics employed, followed by the presentation of classification results. The classification results of the Gauteng dataset are presented in Section 5.3.3, Section 5.3.4 and Section 5.3.5. The classification results of the Limpopo dataset are presented in Section 5.3.6.

5.3.1 Classification accuracy metrics

Two classification accuracy metrics are used in this section, namely *Overall Accuracy (OA)* and the κ-*coefficient*. The κ-coefficient is especially useful, since it can determine whether the values contained in an error matrix represent a result significantly better than random [200, 201]. The two metrics in question are computed by using an error matrix, which is a matrix containing the number of pixels classified correctly and incorrectly for each class under consideration. An example error matrix is presented in Table 5.9.

Table 5.9: Error matrix used to explain the definition of OA and the κ-coefficient.

	Class 0	**Class** 1	
Class 0	x_{11}	x_{12}	x_{1*}
Class 1	x_{21}	x_{22}	x_{2*}
	x_{*1}	x_{*2}	N

OA is a percentage (obviously the closer the metric is to 100 the better the classifier) and is defined as

$$100 \times \sum_{i=1}^{r} x_{ii}/N, \tag{5.6}$$

while the κ-coefficient is computed with

$$\kappa = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i*} \times x_{*i}}{N^2 - \sum_{i=1}^{r} x_{i*} \times x_{*i}}. \tag{5.7}$$

In Equation 5.6 and Equation 5.7, N is the total number of pixels in the error matrix, r is the number of rows in the matrix, x_{ii} is the number in row i and column i, x_{i*} is the total for row i, and x_{*i} is the total for column i (see Table 5.9 for more details). Note that x_{11} refers to the number of pixels belonging to class 0 classified correctly and x_{22} refers to the number of pixels belonging to class

1 which were classified correctly, while x_{21} refers to the number of pixels belonging to class 0 which were not classified correctly and x_{12} refers to the number of pixels belonging to class 1 which where not classified correctly. The value of κ can be smaller than or equal to 1. The magnitude guidelines for κ published in [202] are adopted here. According to [202], a larger κ indicates that the classifier can easily discern the different classes, while a smaller (or even negative) κ indicates that the classifier cannot easily discern the different classes.

5.3.2 Structure used for accuracy metrics

In this section a three-dimensional structure is presented that is used to organise the cross-validation classification results of Section 5.3.3 to Section 5.3.7.

The κ-coefficients and the OA percentages generated by the cross-validation experiments in Section 5.3.3 to Section 5.3.6 can be organised into a three-dimensional irregular structure $\Xi_{\upsilon,\mu}$, where

$$\upsilon \in \{\text{`}\kappa\text{'}, \text{`OA'}\}$$

and

$$\mu \in \{\text{Min Dist,TVML}, \boldsymbol{\theta}, \boldsymbol{\iota}, \boldsymbol{\zeta}\} \tag{5.8}$$

with elements

$$\xi^{\mu}_{x,y(x),z} \in \{\kappa, \text{OA}\},$$

where

$$\begin{aligned} x &\in \{1,2,\cdots,8\}; \\ y(x) &\in \left\{1,2,\cdots,\binom{x}{8}\right\}; \\ z &\in \{1,2,\cdots,e\}. \end{aligned} \tag{5.9}$$

In Equation 5.8, Min Dist,TVML, $\boldsymbol{\theta}, \boldsymbol{\iota}$ and $\boldsymbol{\zeta}$ are respectively associated with the minimum distance classifier, the time-varying maximum likelihood classifier, the noise-harmonic feature group, the harmonic feature group and the temporal feature group defined in Section 4.2.4.2. When μ is omitted it implies that the classification procedure that generated Ξ_{υ} is unknown.

In Equation 5.9, x represents the band restriction value (the amount of bands that may be used for classification), while each y is associated with a unique band combination given the restriction of x (which explains the combination notation used to define y). The z index points to a specific cross-validation experiment and $e \in \{1,2,\cdots\}$ denotes the amount of experiments performed.

Recall that each MODIS pixel consists of a time-series. The notation $\Xi_\upsilon[n]$ should be interpreted as representing a structure similar to Ξ_υ, with the only difference being that it consists of classification accuracy metric elements obtained by using truncated ($n \in \{1,2,\cdots,368\}$) MODIS pixels. It should be clear that if $[n]$ is omitted from $\Xi_\upsilon[n]$ then it implies that no truncation was performed before classification commenced.

The notation $\Phi_\upsilon = \mathbb{E}_\omega[\Xi_\upsilon]$ should be interpreted as the sample mean of Ξ_υ along the dimension $\omega \in \{x,y,z\}$ (which is similar to the Matlab *mean* command). The resulting structure Φ_υ has only two dimensions, since $\mathbb{E}_\omega[]$ eliminated a dimension of Ξ_υ. The notation $\Theta_\upsilon = \mathbb{E}_{\eta\neq\omega}[\mathbb{E}_\omega[\Xi_\upsilon]]$ should be interpreted as the sample mean of Ξ_υ along the dimension ω, followed by the sample mean of Φ_υ along the dimension $\eta \in \{x,y,z\}$. Note that the resulting structure Θ_υ has only one dimension. The structures $\sigma_\omega\{\Xi_\upsilon\}$ and $\sigma_{\eta\neq\omega}\{\sigma_\omega\{\Xi_\upsilon\}\}$ should be interpreted in a similar way, except for the fact that σ indicates that standard deviation (which is similar to the Matlab *std* command) should be used instead of the sample mean when deriving $\sigma_\omega\{\Xi_\upsilon\}$ and $\sigma_{\eta\neq\omega}\{\sigma_\omega\{\Xi_\upsilon\}\}$. The notation

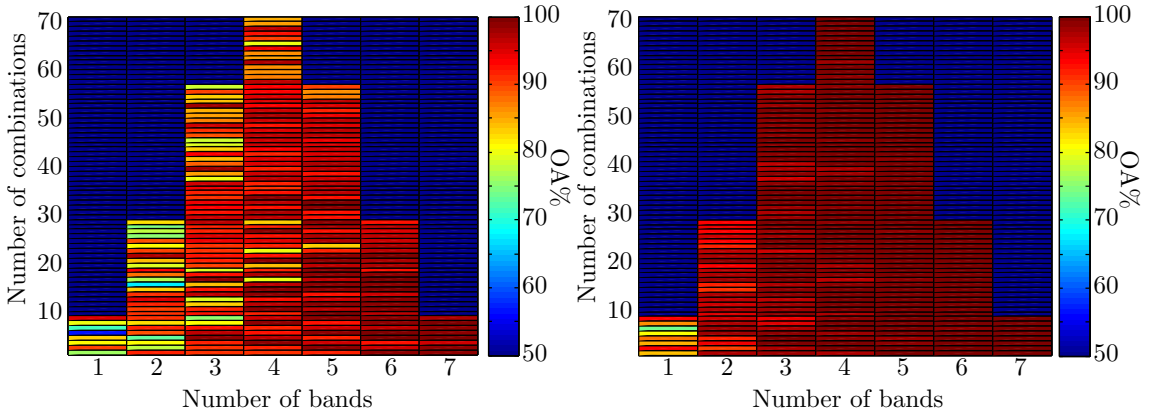
$$\Sigma_\upsilon = \Xi_\upsilon^{\{1,2,\cdots,7\},*,1}$$

should be interpreted as a substructure of Ξ_υ which only contains the elements in Ξ_υ for which $x = \{1,2,\cdots,7\}$ and $z = 1$. The $*$ is a wild card which indicates all valid values y can ascertain. If one of the dimensions of a structure Σ_υ is equal to one, then $(\Sigma_\upsilon)^@$ should be interpreted as the structure obtained after the redundant dimension is removed. When two structures Ξ_{υ,μ_1} and Ξ_{υ,μ_2} are compared, then classification approach μ_1 performs better than μ_2 if $\xi^{\mu_1}_{x,y(x),z} > \xi^{\mu_2}_{x,y(x),z}$ for more than 50% of all the possible index values (see Equation 5.9).

5.3.3 Preliminary benchmark classification results: Gauteng

In this section the results of the benchmark classification approaches are presented. The benchmark classifiers that were used are the minimum distance classifier and the time-varying maximum likelihood classifier presented in Section 4.2.2 and Section 4.2.3 respectively [23]. These two approaches were selected, as they are frontier hypertemporal approaches. As the algorithms are only used for

benchmarking, the full vegetation and settlement datasets were used for training and validation (implying that $e = 1$). The estimation procedure for the yearly ensemble means used by the minimum distance classifier is discussed in Section 5.1.1, while the estimation of the time-varying model used by the time-varying maximum likelihood classifier is discussed in Section 5.1.2. Moreover, no sequential thresholds were used for the time-varying maximum likelihood classifier and $\pi = 0.5$. The OA classification results of the benchmarking approaches for each possible band combination are found in Figure 5.33.



(a) $\left(\Xi^{\{1,2,\cdots,7\},*,1}_{\text{'OA'},\text{Min Dist}}\right)^{@}$. **(b)** $\left(\Xi^{\{1,2,\cdots,7\},*,1}_{\text{'OA'},\text{TVML}}\right)^{@}$.

Figure 5.33: The classification results (OA) of the benchmarking approaches of the Gauteng dataset.

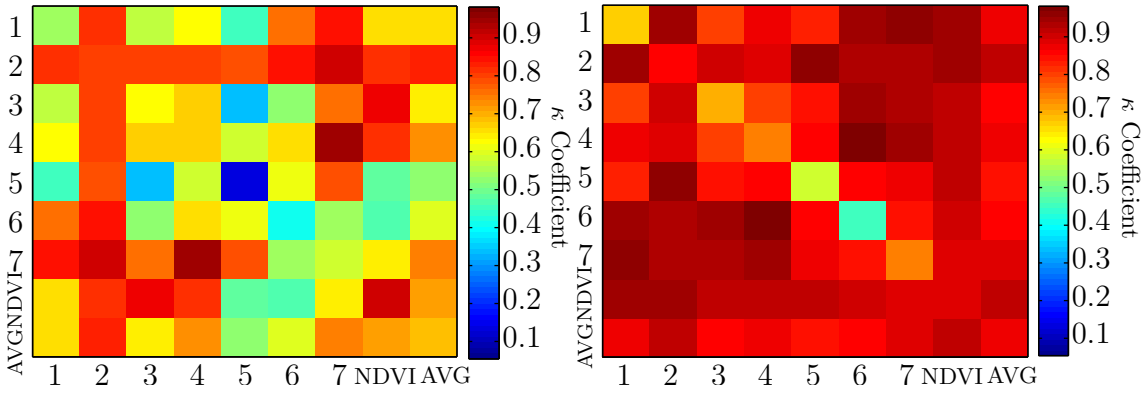
Each possible two-band combination κ-coefficient for the benchmarking approaches is presented in Figure 5.34.

The following observations and conclusions can be made from Figure 5.34:

1. With respect to the minimum distance classifier, the bands that in combination perform best are $\{2,4,7,\text{NDVI}\}$, while bands $\{1,3,5,6\}$ perform worst (in terms of classification capability). The band that in combination performs best is band 2, while band 5 performs worst. The band combination that separates the two classes best is $\{4,7\}$, while the lowest κ value is produced by $\{1,5\}$.

2. The results are exactly the same for the time-varying maximum likelihood classifier, except that $\{4,6\}$ is the band combination that performs best, while band combination $\{1,3\}$ performs worst.

3. For both approaches NDVI performs best (single-band). The worst performing bands are res-



(a) $\left(\Xi_{\text{`}\kappa\text{'},\text{Min Dist}}^{\{1,2\},*,1}\right)^{@}$ packed into a symmetrical matrix (plus a mean row and column). **(b)** $\left(\Xi_{\text{`}\kappa\text{'},\text{TVML}}^{\{1,2\},*,1}\right)^{@}$ packed into a symmetrical matrix (plus a mean row and column).

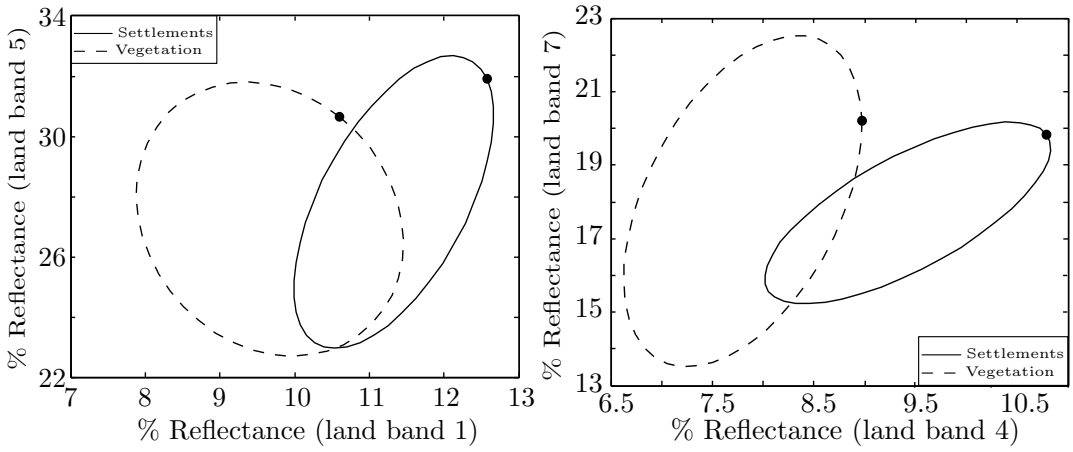
Figure 5.34: The two-band classification results (κ) of the benchmarking approaches of the Gauteng dataset.

pectively band 5 and 6 in the case of the minimum distance classifier and the time-varying maximum likelihood classifier.

4. Recall from Section 5.1.2.3 that the temporal Hellinger metric predicted that a temporal classifier using bands $\{1,2,3,4\}$ would perform better than a temporal classifier using bands $\{5,6,7,\text{NDVI}\}$. As seen in Figure 5.34, this prediction is reasonably close to the actual observed behaviour of the two temporal benchmarking classifiers investigated in this section. A possible reason for the discrepancy between the observed behaviour and the predicted behaviour is discussed in Section 5.3.4.

5. Generally the time-varying maximum likelihood classifier outperforms the minimum distance classifier, as it uses a superior metric, namely the posterior sequence, which incorporates the yearly ensemble mean as well as the inter-class variance.

As an interesting side note, a closer look is taken at bands $\{4,7\}$ and $\{1,5\}$ from Figure 5.34a. The yearly ensemble means (after fitting appropriate sinusoids) of $\{4,7\}$ and $\{1,5\}$ are displayed in Figure 5.35a and Figure 5.35b, respectively. It is clear from Figure 5.35 that, in the case of bands 4 and 7, the yearly ensemble means of the settlement and vegetation classes have a greater distance between them than in bands 1 and 5. The increased distance observable when inspecting the yearly ensemble means of $\{4,7\}$ and $\{1,5\}$ helps to explain why the minimum distance classifier using bands $\{4,7\}$ outperforms the minimum distance classifier using bands $\{1,5\}$.



(a) Yearly ensemble mean models of bands 1 and 5. **(b)** Yearly ensemble mean models of bands 4 and 7.

Figure 5.35: Two-dimensional yearly ensemble mean models of Gauteng.

In the last part of this section the focus will shift to the time-varying maximum likelihood classifier. The time-varying maximum likelihood classifier is unique in the sense that it is based on sequential analysis (discussed in Section 3.4) and was first introduced (to the remote sensing field) in [23]. As the time-varying maximum likelihood classifier is based on sequential analysis it can be truncated, which allows the classifier to make a decision after each observation is received by using all the observations received up to that point. The focus of [23] was on the trade-off between classification accuracy and classification delay. The classification study in Section 5.3 however focused solely on increasing classification accuracy. One of the aims of this thesis, which was stated in Chapter 1, is to verify the sequential results presented in [23]. The most important result of [23] was therefore reproduced and can be found in Figure 5.36. Figure 5.36 presents the mean single-band classification κ-coefficient or average single-band classification performance of the time-varying maximum likelihood classifier as a function of time or sample size. If the notation from Section 5.3.2 is used then the mean single-band classification κ-coefficient (as a function of sample size) can be expressed as

$$\mathbb{E}_{y(1)}\left[\left(\Xi^{1,*,1}_{\text{`}\kappa\text{'},\text{Min Dist}}[n]\right)^{@}\right].$$

The delay and accuracy measures discussed in Section 4.2.3 are not calculated, as the datasets under consideration have finite sizes and Figure 5.36 thus displays an alternative fixed-sample-size performance measure.

Generally the average κ-coefficient in Figure 5.36 increases with time. A steep increase is observable in the first year and a smaller increase during the second year. After the second year however the increase of κ is small. The temporal dynamics of the average κ-coefficient in Figure 5.36 suggests that

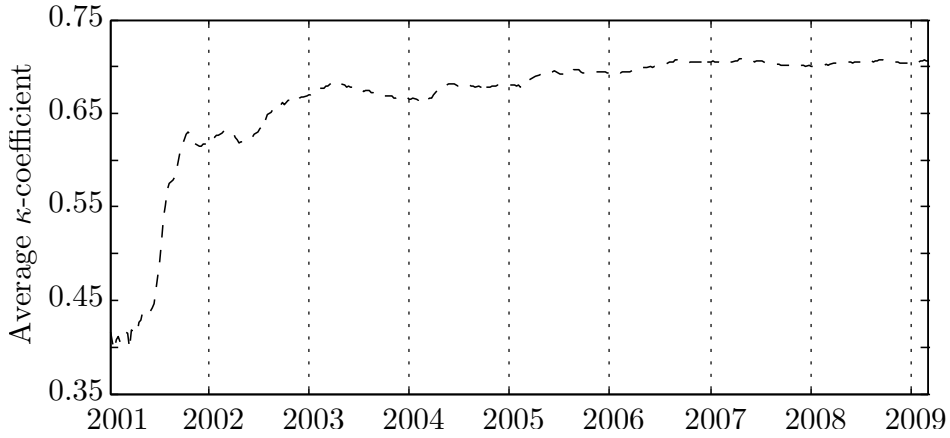


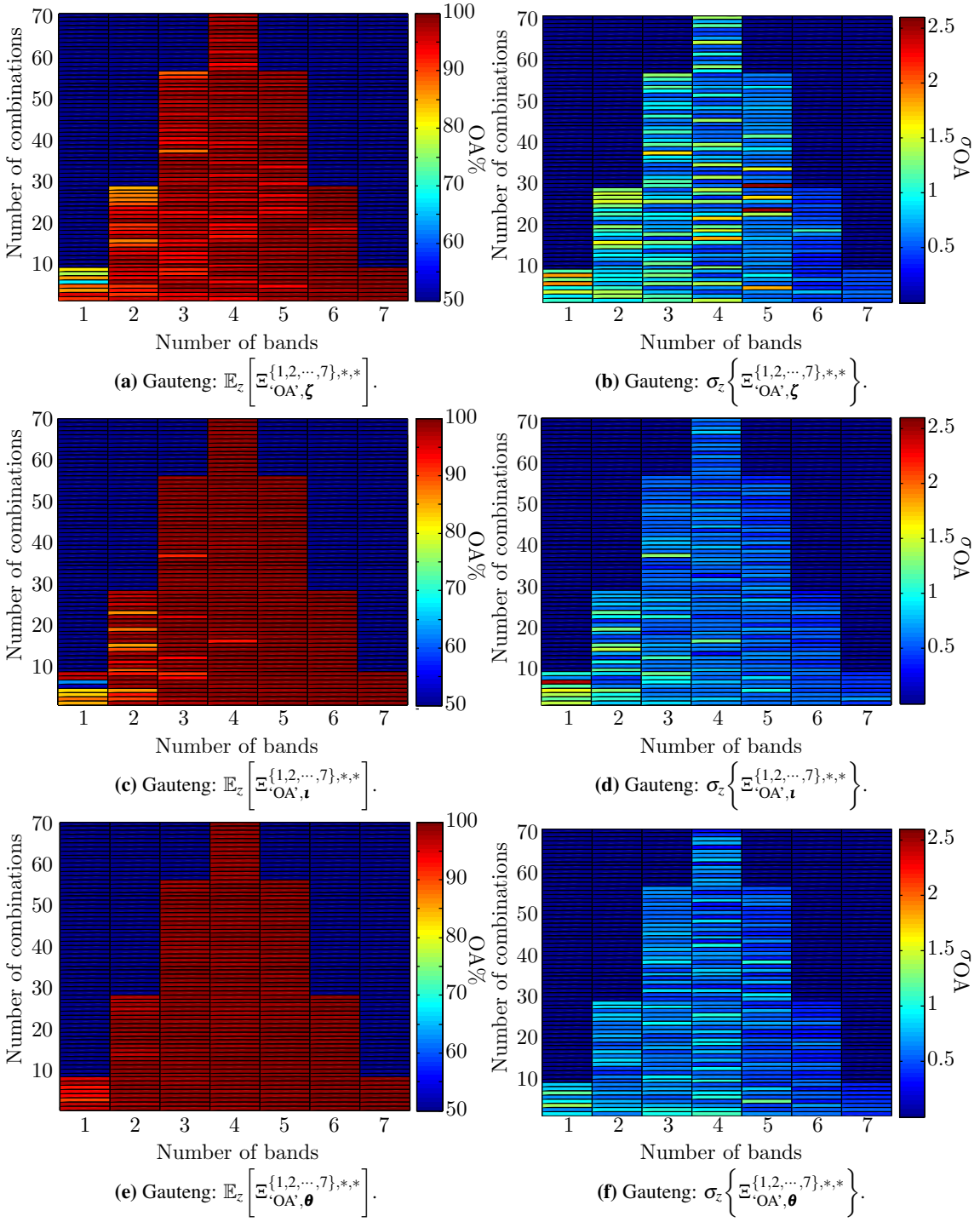
Figure 5.36: The average single-band classification performance of the time-varying maximum likelihood classifier for the Gauteng dataset as a function of time.

for the Gauteng dataset specifically the thresholds of the time-varying maximum likelihood classifier should be chosen in such a way that the classifier on average experiences (excluding outliers) at least a one-year delay (preferably two) before it can classify an observed sequence. The same result was obtained in another independent study, namely [78].

5.3.4 Preliminary SVM classification results: Gauteng

The SVM classifier and the proposed feature groups are discussed in Section 4.2.4 and Section 4.2.4.2 respectively. Recall from Section 4.2.4.2 that three feature groups are proposed, namely temporal features $\boldsymbol{\zeta}$, harmonic features $\boldsymbol{\iota}$ and noise-harmonic features $\boldsymbol{\theta}$. A linear SVM is used, since it produced sufficient classification results. The SVM is realised with the *SVM and Kernel Methods Matlab Toolbox* [203]. The SVM and Kernel Methods Matlab Toolbox requires two input parameters, which are determined via a standard grid search algorithm. The two input parameters that need to be set are (c, λ). The parameter c sets the bound on the Lagrangian multipliers, while λ is a conditioning parameter for the quadratic programming method used to determine the SVM hyperplane. Random split cross-validation (50% for training and 50% for validation) was employed by the grid search algorithm (50 independent experiments). The cross-validation OA results of the Gauteng dataset for all the possible band combinations of $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$ are presented in Figure 5.37. The standard deviation in Figure 5.37 for each feature group is small, indicating that the classification results are reliable and stable.

The Gauteng two-band classification results for $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$ are presented in Figure 5.38. The κ-



(a) Gauteng: $\mathbb{E}_z\left[\Xi_{\text{'OA'},\boldsymbol{\zeta}}^{\{1,2,\cdots,7\},*,*}\right]$.

(b) Gauteng: $\sigma_z\left\{\Xi_{\text{'OA'},\boldsymbol{\zeta}}^{\{1,2,\cdots,7\},*,*}\right\}$.

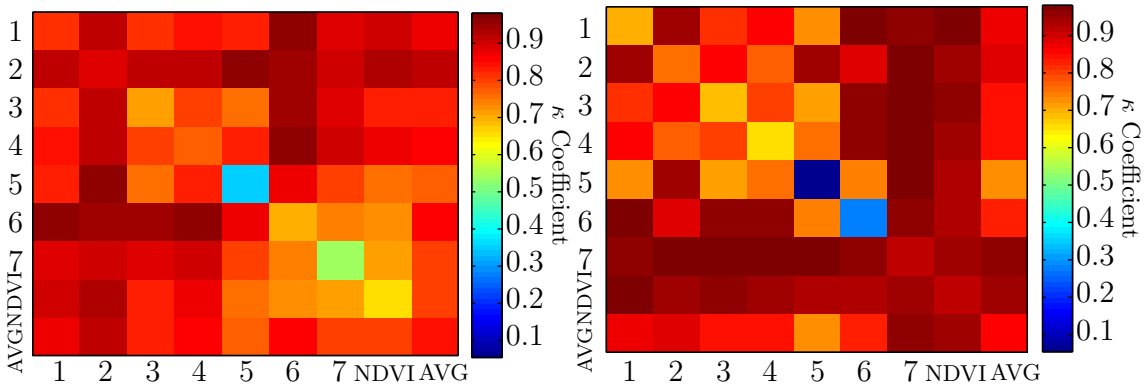
(c) Gauteng: $\mathbb{E}_z\left[\Xi_{\text{'OA'},\boldsymbol{\iota}}^{\{1,2,\cdots,7\},*,*}\right]$.

(d) Gauteng: $\sigma_z\left\{\Xi_{\text{'OA'},\boldsymbol{\iota}}^{\{1,2,\cdots,7\},*,*}\right\}$.

(e) Gauteng: $\mathbb{E}_z\left[\Xi_{\text{'OA'},\boldsymbol{\theta}}^{\{1,2,\cdots,7\},*,*}\right]$.

(f) Gauteng: $\sigma_z\left\{\Xi_{\text{'OA'},\boldsymbol{\theta}}^{\{1,2,\cdots,7\},*,*}\right\}$.

Figure 5.37: The cross-validation results [with $(\infty, 0.05)$ and $e = 50$] of the Gauteng dataset for all the possible band combinations of $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$.



(a) Gauteng: $\mathbb{E}_z\left[\Xi^{\{1,2\},*,*}_{\kappa,\boldsymbol{\zeta}}\right]$ packed into a symmetrical matrix.

(b) Gauteng: $\mathbb{E}_z\left[\Xi^{\{1,2\},*,*}_{\kappa,\boldsymbol{\iota}}\right]$ packed into a symmetrical matrix.

(c) Gauteng: $\mathbb{E}_z\left[\Xi^{\{1,2\},*,*}_{\kappa,\boldsymbol{\theta}}\right]$ packed into a symmetrical matrix.

(d) Average between $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$ results for Gauteng.

Figure 5.38: The two-band classification results (average κ) of $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$ for the Gauteng dataset [with $(\infty, 0.05)$ and $e = 50$].

coefficients reported in Figure 5.38 are the average of 50 random split cross-validation experiments. The following observations and conclusions can be made from Figure 5.38:

1. When the SVM classifier using $\boldsymbol{\zeta}$ uses data from bands $\{1,2,4,6\}$ it classifies better than when the SVM classifier using $\boldsymbol{\zeta}$ uses data from bands $\{3,5,7,\text{NDVI}\}$. The SVM classifier using $\boldsymbol{\zeta}$ obtains its best classification results when data from band 2 are used in combination, while the SVM classifier using $\boldsymbol{\zeta}$ performs at its worst when data from band 5 are used in combination. The two-band combination $\{2,5\}$ performs best, while the two-band combination $\{7,\text{NDVI}\}$ performs worst of all the two-band combinations. The best and worst single-bands that the SVM classifier using $\boldsymbol{\zeta}$ can use are bands 2 and 5 respectively.

2. When the SVM using the harmonic features $\boldsymbol{\iota}$ uses data from bands $\{1,2,7,\text{NDVI}\}$ it achieves

better classification results than when it uses data from bands $\{3,4,5,6\}$. When the SVM using the harmonic features $\boldsymbol{\iota}$ use data from band 7 in combination it produces the best classification results, while using band 5 in combination leads to the worst classification results. Band combination $\{3,5\}$ performs best, while band combination $\{2,7\}$ performs worst of all the possible two-band combination. The best and worst single-bands that the SVM classifier using $\boldsymbol{\iota}$ can use are NDVI and band 5 respectively.

3. The SVM using the noise-harmonic features $\boldsymbol{\theta}$ perform better when it uses data from bands $\{1,2,4,7\}$ than when it uses data from bands $\{3,5,6,\text{NDVI}\}$. The highest average κ values are achieved when band 2 is used in combination, while the lowest average κ values are achieved when using band 3 in combination. The two-band combination that perform best is $\{1,2\}$, while $\{3,4\}$ perform the worst among all the possible two-band combinations. The best and worst single-bands that the SVM classifier using $\boldsymbol{\theta}$ can use are bands 2 and 3 respectively.

4. Generally for the Gauteng dataset, the SVM using $\boldsymbol{\theta}$ outperforms the SVMs using $\boldsymbol{\zeta}$ and $\boldsymbol{\iota}$. The SVM using $\boldsymbol{\iota}$ outperforms the SVM using $\boldsymbol{\zeta}$. The fact that the feature group $\boldsymbol{\theta}$ outperforms $\boldsymbol{\iota}$ is as expected, since the parameter Hellinger distance indicated in Section 5.1.3.3 that the noise and phase parameters of the CSHO are extra discerning features, which implies that they can be used to extend the classification potential of $\boldsymbol{\iota}$.

5. The average κ-coefficients of the SVMs using $\boldsymbol{\zeta}, \boldsymbol{\iota}$ and $\boldsymbol{\theta}$ are presented in Figure 5.38d, from which it is clear that on average combining $\{1,2,3,4\}$ with $\{6,7,\text{NDVI}\}$ produces high κ values. Furthermore, generally the following band combinations, $\{1,2\},\{5,2\}$ and $\{3,2\}$, also perform well.

6. Recall from Section 5.1.2.3 that for the Gauteng dataset the temporal Hellinger distance metric predicted that a classifier using data from bands $\{1,2,3,4\}$ would provide better class differentiability than a classifier using data from bands $\{5,6,7,\text{NDVI}\}$ when the classifier in question relies on temporal features. Similarly, recall from Section 5.1.3.3 that the parameter Hellinger distance metric predicted that a classifier using data from bands $\{1,2,7,\text{NDVI}\}$ would provide better class differentiability than a classifier using bands $\{3,4,5,6\}$ when the classifier in question employs the parameters of the CSHO as classification features. As seen in Figure 5.38, these two predictions are close to the actual observed behaviour of the three SVM classifiers investigated in this section. The correlation between the predictions and reality implies that the

separability metrics introduced in Section 5.1 can be used to select classification features. The small discrepancy between the observed behaviour and the predicted behaviour can be ascribed to the fact that the relation between actual performance and predicted separability based on a single metric is not a perfect one-to-one relation, as the single metric does not incorporate all the factors (for instance spectral dependence is not measured by the Hellinger metric) that influence the performance of a specific classifier.

5.3.5 Classification results: Gauteng

Finally, the bar graph representing the $\mathbb{E}_{y(x)}\left[\mathbb{E}_z\left[\Xi_{`\kappa',\mu}\right]\right]$ values for each classification approach mentioned (for all μ) is presented in Figure 5.39. Note that there is no cross-validation experiments for the benchmarking approaches. In the case of the benchmarking approaches only one sample mean is taken over all the band combinations. The values $\mathbb{E}_{y(x)}\left[\sigma_z\left\{\Xi_{`\kappa',\mu}\right\}\right]$ for the SVM approaches are displayed in Figure 5.39b. Recall that the index $x \in \{1,2,\cdots,8\}$ is used to indicate the band restriction value. So if $x = 5$, the band restriction value is equal to 5. The following observations can be made from Figure 5.39a and Figure 5.39b:

1. The SVM classifier using the noise-harmonic features outperform all the other classifiers when $x \leq 3$. When $x > 3$ however the time-varying maximum likelihood classifier performs better than the SVM classifier using the noise-harmonic features. Furthermore, when $x > 3$ the SVM classifier using the harmonic features produces similar classification results as the SVM classifier using the noise-harmonic features.

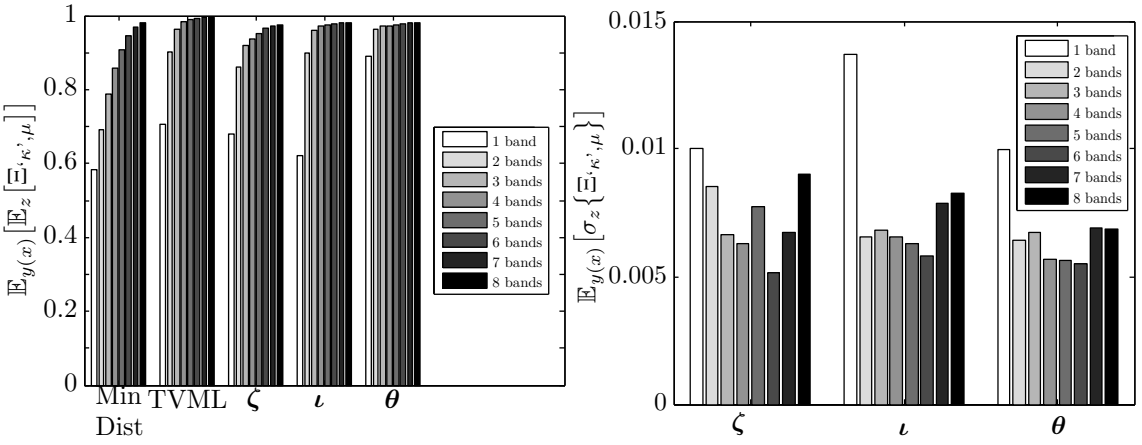
2. For all values of x, the time-varying maximum likelihood classifier on average achieves higher classification accuracies compared to the remaining classifiers (when the SVM using $\boldsymbol{\theta}$ is excluded).

3. The SVM classifier using the harmonic features classifies better than the SVM using $\boldsymbol{\zeta}$ and the minimum distance classifier except when $x = 1$, then $\boldsymbol{\zeta}$ produces better classification results than $\boldsymbol{\iota}$ or the minimum distance classifier.

4. The SVM classifier using the temporal features outperforms the minimum distance classifier when $x < 8$. When $x = 8$ the minimum distance classifier outperforms the SVM using $\boldsymbol{\zeta}$.

5. For all classifiers, increasing the spectral dimension increases classification accuracy.

6. The standard deviation for each SVM classifier (for all x) is small, implying that the classification results are stable and reliable.



(a) $\mathbb{E}_{y(x)}\left[\mathbb{E}_z\left[\Xi_{\text{`}\kappa\text{'},\mu}\right]\right]$ for the Gauteng dataset.

(b) $\mathbb{E}_{y(x)}\left[\sigma_z\left\{\Xi_{\text{`}\kappa\text{'},\mu}\right\}\right]$ for the Gauteng dataset.

Figure 5.39: The bar graph representing the $\mathbb{E}_{y(x)}\left[\mathbb{E}_z\left[\Xi_{\text{`}\kappa\text{'},\mu}\right]\right]$ values for the benchmarking and SVM classifiers [with $(\infty, 0.05)$ and $e = 50$] (Gauteng).

5.3.6 Classification results: Limpopo

To avoid repetition, only the most important classification results for the Limpopo data set will be mentioned. As mentioned in Section 5.3.3, the time-varying maximum likelihood classifier is an important facet of the thesis and the equivalent of Figure 5.36 is therefore presented in Figure 5.40 for the Limpopo dataset.

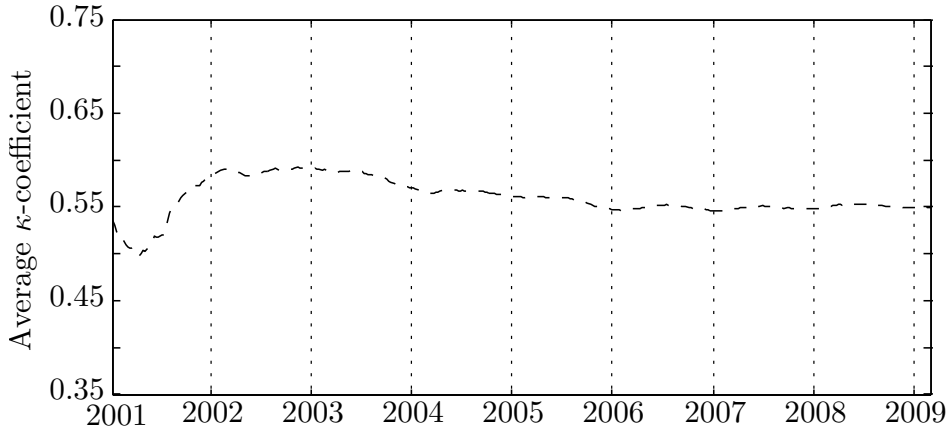
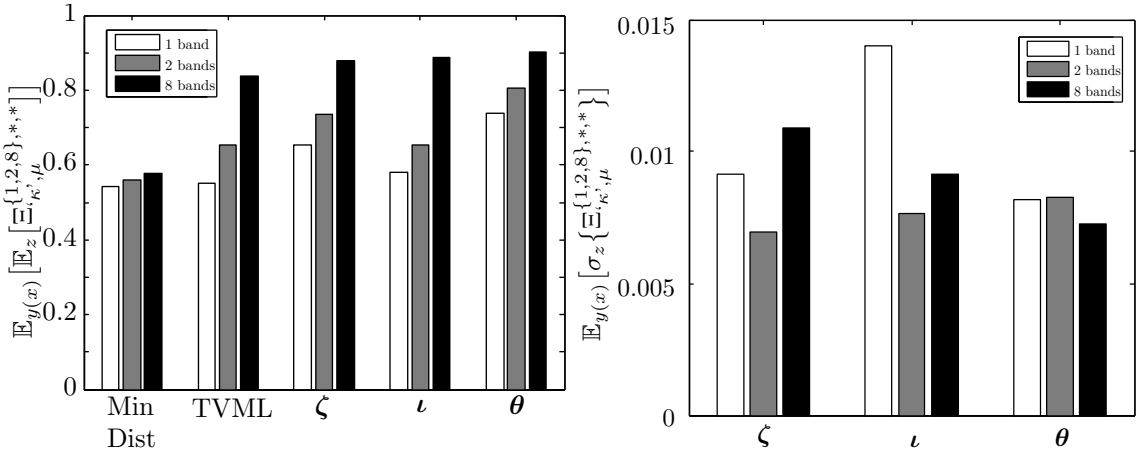


Figure 5.40: The average single-band classification performance of the time-varying maximum likelihood classifier for the Limpopo dataset as a function of time.

Generally the average κ-coefficient in Figure 5.40 increases during the first two years. A steep increase is observable in the first year and a smaller increase during the second year. After the second year however the average κ-coefficient starts to decrease steadily. The temporal dynamics over the first two years of the average κ-coefficient in Figure 5.40 strengthens the one-year (preferably two-year) sequential detection delay rule formed in Section 5.3.3, while inspecting Figure 5.36. The behaviour in Figure 5.36 and Figure 5.40 however differs from 2002 onwards. The decrease in κ observable in Figure 5.40 implies that the vegetation and settlement classes in the case of the Limpopo dataset become less separable over time. The opposite behaviour is seen in Figure 5.36, implying that the separability does not decrease in the case of the Gauteng dataset. The loss in separability in the case of the Limpopo dataset is not directly observable when inspecting Table 5.8, which implies that the separability loss happens gradually.

The equivalent of Figure 5.39 is presented in Figure 5.41 for the Limpopo dataset, as it sums up the performance of each classifier on the Limpopo dataset. As mentioned in Section 5.3.4, a grid search is required to obtain the SVM software parameters (c, λ). Random split cross-validation (50% for training and 50% for validation) was also employed by the grid search algorithm for the Limpopo dataset (10 independent experiments). The content details of Figure 5.41 are discussed in Section 5.3.5. Also recall from Section 5.3.5 that $x \in \{1, 2, 8\}$ indicates the band restriction value.



(a) $\mathbb{E}_{y(x)}\left[\mathbb{E}_z\left[\Xi_{\kappa,\mu}^{\{1,2,8\},*,*}\right]\right]$ for the Limpopo dataset. **(b)** $\mathbb{E}_{y(x)}\left[\sigma_z\left\{\Xi_{\kappa,\mu}^{\{1,2,8\},*,*}\right\}\right]$ for the Limpopo dataset.

Figure 5.41: The bar graph representing the $\mathbb{E}_{y(x)}\left[\mathbb{E}_z\left[\Xi_{\kappa,\mu}\right]\right]$ values for the benchmarking and SVM classifiers [with $(\infty, 1)$ and $e = 10$] (Limpopo).

The following observations can be made from Figure 5.41:

1. The SVM classifier using the noise-harmonic features on average achieves higher classification accuracies than any of the other classifiers for all x.

2. The SVM classifier using the temporal features on average produces better classification results if compared to all the classification results produced by the remaining classifiers (if the SVM classifier using $\boldsymbol{\theta}$ is excluded) except when $x = 8$. When $x = 8$ the SVM classifier using $\boldsymbol{\iota}$ outperforms the SVM using $\boldsymbol{\zeta}$.

3. The time-varying maximum likelihood classifier and the SVM classifier using the harmonic features on average achieve better classification accuracies if compared with the classification results produced by the minimum distance classifier.

4. When $x = 1$ or $x = 8$ the SVM using the harmonic features outperforms the time-varying maximum likelihood classifier. When $x = 2$ exactly the opposite behaviour is observed.

5. For all classifiers, increasing the spectral dimension improves the accuracy of the classifier.

6. The standard deviation for each SVM classifier (for all x) is small, implying that the classification results are reliable and stable.

7. According to the classification results the Limpopo dataset is less separable than the Gauteng dataset.

5.3.7 Important classification conclusions

The following important conclusions can be made from the results presented in Section 5.3:

1. As the conclusions in Section 5.3.3 and Section 5.3.4 show, the metrics proposed in Section 5.1 can be used to choose efficient spectral bands and classification features.

2. As mentioned in Section 5.3.5 and Section 5.3.6, increasing the spectral dimension of a classifier improves its classification accuracy.

3. The feature group $\boldsymbol{\theta}$ is an extension of $\boldsymbol{\iota}$, not only in size but also in terms of its classification capability. This is clearly seen in Figure 5.11, Figure 5.14 and Figure 5.38. Moreover, Figure 5.39 and Figure 5.41 show that the SVM classifier that uses $\boldsymbol{\theta}$ outperforms all the other

classifiers investigated (except when $x > 2$, in the case of the Gauteng datset, which is when the time-varying maximum likelihood classifier starts achieving a slightly better classification accuracy). It is also significant to mention that the SVM classifier using $\boldsymbol{\theta}$ performs much better than the other classifiers when the spectral dimension of the classifier is low. This unique characteristic of $\boldsymbol{\theta}$ is significant, since it is obviously more advantageous to classify more accurately without having to increase the spectral view.

4. The Limpopo dataset is less separable than the Gauteng dataset, according to the classification results in Section 5.3.5 and Section 5.3.6. This is not surprising for the temporal or the CSHO feature classifiers. Even though the temporal Hellinger distance metrics in Table 5.3, are similar for the Gauteng and Limpopo datasets it is clear from Figure 5.40 that the separability of the Limpopo dataset deteriorates over time, explaining the weaker performance (in the case of the Limpopo dataset) of the minimum distance classifier, the time-varying maximum likelihood classifier and the SVM using $\boldsymbol{\zeta}$ in spite of good separability in the time-varying models of the Limpopo dataset. For $\boldsymbol{\iota}$ and $\boldsymbol{\theta}$ it is clear from Figure 5.11 and Figure 5.14 that the parameters of the CSHO provide better classification capability in the case of the Gauteng dataset than for the Limpopo dataset.

5. The most important sequential result found in [23] was reproduced, which is stated next. The temporal dynamics of the average κ-coefficient in Figure 5.36 and Figure 5.40 suggest that in general the thresholds of the time-varying maximum likelihood classifier should be chosen in such a way that the classifier on average experiences (excluding outliers) at least a one-year delay (preferably two) before it can classify an observed sequence.

5.4 CHANGE DETECTION RESULTS: GAUTENG AND LIMPOPO

This section focuses on the performance of the sequential change detection algorithm presented in Section 4.3.3, namely CUSUM. The CUSUM algorithm is benchmarked against the popular band differencing algorithm discussed in Section 4.3.2. This section starts by introducing the different change detection metrics employed, followed by the presentation of the change detection results of band differencing and CUSUM.

5.4.1 Change detection accuracy metrics

As change detection metrics P_{D}, P_{FA}, $\mathbb{E}\{(T-\tau)^{+368}\}$ and A_e are employed as metrics instead of $d_l(T)$ and $f(T)$, which were introduced in Section 3.6, since metrics that can be fairly compared to non-sequential change detection algorithms are required. Here P_{D} is the probability of correctly detecting a change within the eight-year observation period, P_{FA} is the probability of detecting a change when there was no change in the eight-year period, $\mathbb{E}\{(T-\tau)^{+368}\} = \mathbb{E}\{\min\{\max\{T-\tau,0\},368-\tau\}\}$ is the positive expected delay truncated to 368 observations and $A_e = \frac{1}{2}[(1-P_{\text{D}})+P_{\text{FA}}]$.

5.4.2 Results of Lunetta et al.'s scheme: Gauteng and Limpopo

The scheme developed by Lunetta et al. is discussed in detail in Section 4.3.2 and is also known as the band differencing algorithm [7]. The band differencing algorithm is a popular time-series change detection benchmarking method, which requires two parameters as input, the amount of frequency components ν to preserve and the decision threshold h_l^b [10,28]. A grid search was performed to find suitable values for ν and h_l^b. In this section the algorithm is applied to the Gauteng and Limpopo datasets. The values for ν, h_l^b and the change detection accuracies can be found in Table 5.10 and Table 5.11 for the Gauteng and Limpopo datasets, respectively. As the band differencing approach is used for benchmarking, the entire Gauteng and Limpopo datasets are used for training and validation. In a previous study [30], the band differencing algorithm was applied to the Gauteng and Limpopo datasets. In [30], the focus was on preserving the structure of the original signal and for that reason ν was set equal to 10. Figure 5.42 displays the effect of ν, which should make the reason for setting ν equal to 10 in the previous study clear. In this section, however the structure of the original signal is not one of the design criteria, and ν is thus found via a grid search algorithm. The number of frequency components ν to keep (as found via the grid search algorithm) is respectively set to two and three for the Gauteng and Limpopo datasets.

The following observations and conclusions can be made from Table 5.10 and Table 5.11:

1. In Table 5.10 and Table 5.11 the lower value of ν increases the accuracy of the band differencing change detector. The remaining observations assume that the lower value of ν is used.

2. For the Gauteng dataset it is clear from Table 5.10 that when the band differencing algorithm employs bands $\{1,3,4,\text{NDVI}\}$ the accuracy of the band differencing change detector is higher

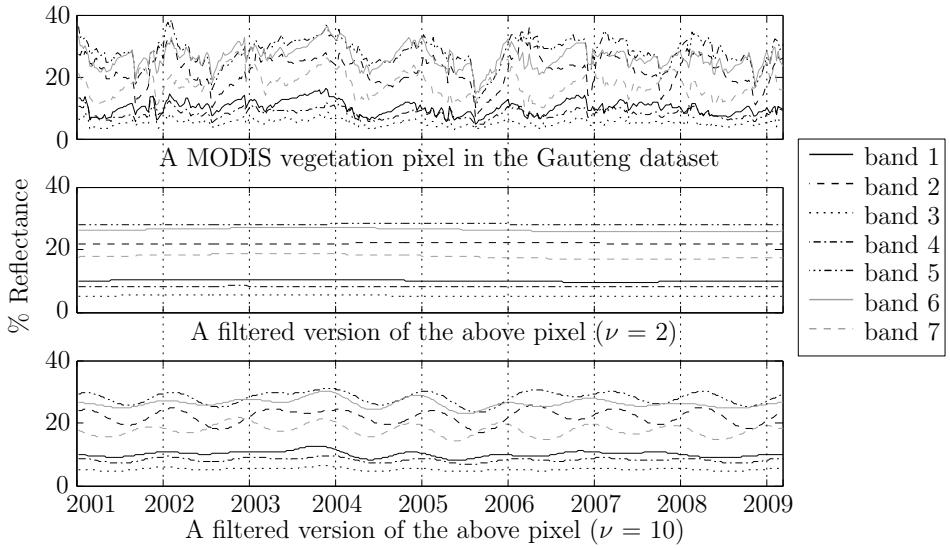


Figure 5.42: A true vegetation pixel belonging to the Gauteng dataset and two different filtered versions of the vegetation pixel.

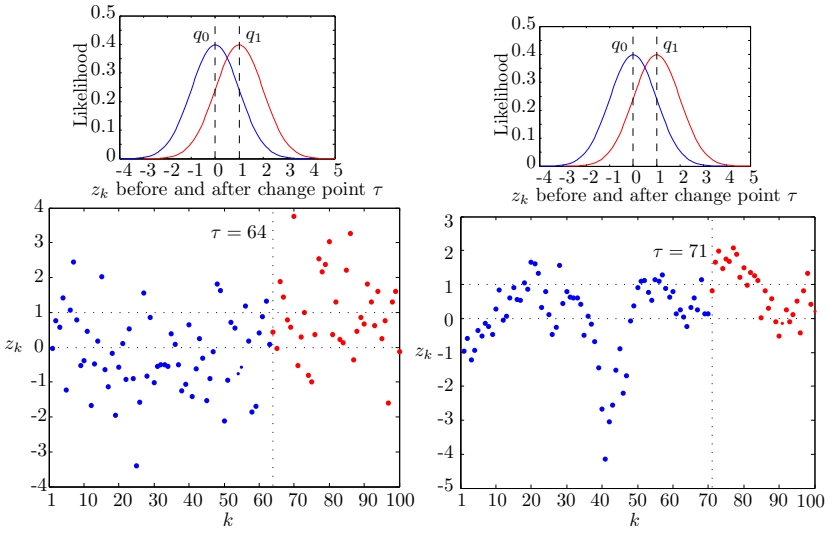
than if it uses bands $\{2,5,6,7\}$. Using band 1 leads to the lowest average error A_e, while using band 5 leads to the highest average error.

3. Exactly the same performance is observable for the Limpopo dataset as for the Gauteng dataset, except that for the Limpopo dataset, using band 4 leads to the lowest average error, while using band 2 leads to the highest average error.

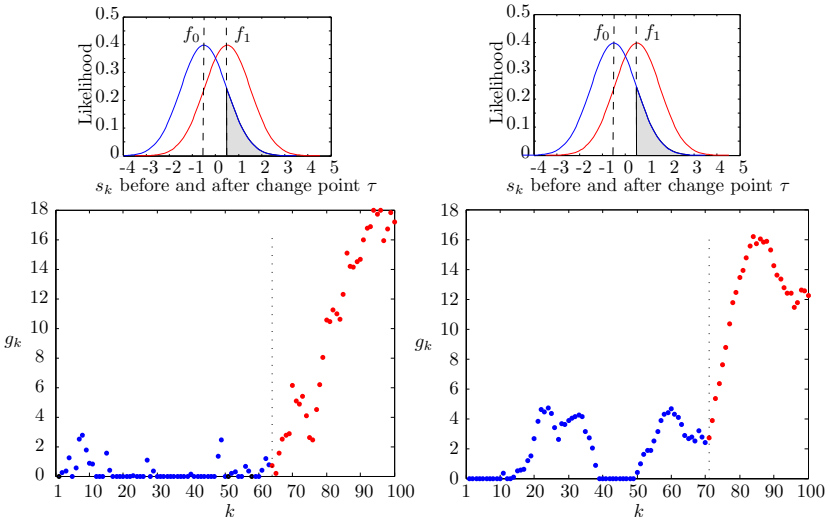
4. Generally the Gauteng dataset produces lower average errors when it is compared with the Limpopo dataset.

5.4.3 Temporal dependence and the CUSUM threshold

The influence of temporal dependence on the CUSUM threshold needs to be investigated before CUSUM can be applied to MODIS data. The influence of temporal dependence on the CUSUM threshold is explained in this section with the aid of an example. Assume therefore that there is a sequence z_k which is an i.i.d. sequence. The sequence is drawn independently from q_0 before change point τ and from τ onwards drawn from q_1. The density $q_0 \sim \mathcal{N}(0,1)$ and the density $q_1 \sim \mathcal{N}(1,1)$. An example of such a sequence can be found in Figure 5.43a with $\tau = 64$. The CUSUM statistic g_k generated from z_k with Equation 3.55 when $y = 0$ in Equation 3.55 is displayed in Figure 5.43c. Note that Equation 3.55 employs the log likelihood ratio s_k defined in Equation 3.56. While g_k stays



(a) z_k is independent. **(b)** z_k is not independent.



(c) g_k for independent case. **(d)** g_k for dependent case.

Figure 5.43: g_k for an independent and dependent example.

Table 5.10: Band differencing applied to the Gauteng dataset.

MODIS band	h_l^b	P_{D}	P_{FA}	A_e
$\frac{\nu=2}{\nu=10}$				
1	$\frac{1.2598}{1.8874}$	$\frac{0.7901}{0.6133}$	$\frac{0.0693}{0.2247}$	$\frac{0.1396}{0.3057}$
2	$\frac{1.3699}{0.7477}$	$\frac{0.6077}{0.9613}$	$\frac{0.1774}{0.8986}$	$\frac{0.2848}{0.4687}$
3	$\frac{1.1827}{1.5350}$	$\frac{0.7735}{0.7790}$	$\frac{0.1233}{0.3919}$	$\frac{0.1749}{0.3064}$
4	$\frac{1.2598}{1.5681}$	$\frac{0.7735}{0.7182}$	$\frac{0.0591}{0.3311}$	$\frac{0.1428}{0.3064}$
5	$\frac{1.4635}{0.7918}$	$\frac{0.5359}{0.9392}$	$\frac{0.2010}{0.8733}$	$\frac{0.3326}{0.4670}$
6	$\frac{1.5460}{2.7407}$	$\frac{0.5746}{0.2210}$	$\frac{0.1233}{0.0591}$	$\frac{0.2744}{0.4191}$
7	$\frac{1.2763}{1.7222}$	$\frac{0.6961}{0.6796}$	$\frac{0.1470}{0.2787}$	$\frac{0.2254}{0.2996}$
NDVI	$\frac{1.2047}{1.5185}$	$\frac{0.8011}{0.8122}$	$\frac{0.1622}{0.4645}$	$\frac{0.1805}{0.3262}$

below a threshold h no change is declared. After crossing the threshold h a change is declared. If z_k is associated with q_0 and q_1 then s_k will be associated with $f_0 \sim \mathcal{N}(-\frac{1}{2},1)$ and $f_1 \sim \mathcal{N}(\frac{1}{2},1)$ because

$$\begin{aligned} s_k &= \ln \frac{q_1(z_k)}{q_0(z_k)} \\ &= ln \frac{e^{-\frac{1}{2}(z_k-1)^2}}{e^{-\frac{1}{2}z_k^2}} \\ &= z_k - \frac{1}{2}. \end{aligned}$$

The densities f_0 and f_1 are displayed in Figure 5.43c.

The focus now shifts to the discretised Ornstein-Uhlenbeck process (see Section 4.1.2.1), which is a dependent sequence with generating equation

$$z_k = e^{-\lambda} z_{k-1} + (1-e^{-\lambda})\mu + \sigma \sqrt{\frac{1-e^{-2\lambda}}{2\lambda}} \eta_k,$$

where $\lambda > 0$ determines the degree of dependence (as well as the mean reversion rate), μ is the long-term mean, $\sigma > 0$ is the volatility of the random fluctuations and η_k is i.i.d. and has density $\mathcal{N}(0,1)$. Recall from Section 4.1.2.1 that the dependent sequence z_k is distributed according to density $\mathcal{N}\left(\mu, \frac{\sigma^2}{2\lambda}\right)$ (the equilibrium density) as long as z_0 is also distributed according to the equilibrium density. The *closer* λ *gets to zero* the *higher the temporal dependence* in the sequence z_k. To mi-

Table 5.11: Band differencing applied to the Limpopo dataset.

MODIS band	h_l^b	P_{D}	P_{FA}	A_e
$\frac{\nu=3}{\nu=10}$				
1	$\frac{2.2287}{2.4379}$	$\frac{0.6496}{0.4957}$	$\frac{0.0822}{0.0788}$	$\frac{0.2163}{0.2915}$
2	$\frac{2.4600}{1.7553}$	$\frac{0.3077}{0.4615}$	$\frac{0.0741}{0.3634}$	$\frac{0.3832}{0.4509}$
3	$\frac{1.8323}{2.2673}$	$\frac{0.7436}{0.5641}$	$\frac{0.1757}{0.0969}$	$\frac{0.2160}{0.2664}$
4	$\frac{2.2452}{2.0085}$	$\frac{0.6325}{0.6410}$	$\frac{0.0601}{0.1844}$	$\frac{0.2138}{0.2717}$
5	$\frac{1.9314}{2.1682}$	$\frac{0.5812}{0.4188}$	$\frac{0.2224}{0.1777}$	$\frac{0.3206}{0.3794}$
6	$\frac{1.9204}{2.4049}$	$\frac{0.6752}{0.3761}$	$\frac{0.2151}{0.0888}$	$\frac{0.2699}{0.3564}$
7	$\frac{1.7497}{1.6011}$	$\frac{0.7436}{0.8120}$	$\frac{0.2445}{0.4502}$	$\frac{0.2504}{0.3191}$
NDVI	$\frac{2.0305}{1.7828}$	$\frac{0.6923}{0.7778}$	$\frac{0.1383}{0.3180}$	$\frac{0.2230}{0.2701}$

mic the independent case, the variance of the equilibrium density should be equal to 1. By choosing $\sigma = \frac{1}{\sqrt{2}}$ and $\lambda = \frac{1}{4}$ a variance of 1 is obtained. Furthermore, assume that before change point τ, $\mu = 0$ and from τ onwards $\mu = 1$. The dependent sequence z_k is shown in Figure 5.43b ($\tau = 71$) and z_k's CUSUM sequence g_k is shown in Figure 5.43d. It is clear from Figure 5.43 that the change detection threshold h will usually be higher for the dependent case than the independent case (even though both cases are equally separable), since the *higher temporal dependence* found in the dependent case causes a *larger noise floor*. A larger noise floor is observable for the dependent case, as the probability of s_k to be positive increases due to the dependence (see the gray area in Figure 5.43d).

5.4.4 Results of the CUSUM test: Gauteng and Limpopo

In this section CUSUM is applied to the Gauteng and Limpopo datasets in order to detect when vegetation pixels in the study areas change into settlement pixels. The CUSUM algorithm is discussed in detail in Section 3.6 and Section 4.3.3. The CUSUM result section is divided into two parts or phases. In the first part of the section an off-line optimisation algorithm is used to determine the best threshold h by performing a sweep of h from 1 to 100 on simulated data to establish an intuitive base of the performance of CUSUM on MODIS data in the study areas. The simulated data that are used during this first phase are generated by the CSHO simulator discussed in Section 4.1.2.7 and Section 5.2.1. In the second part of the section the performance of the off-line determined h (using random split cross-validation) is evaluated on real world MODIS change data and is compared to the

band differencing method (on the same data). The following algorithm, with input vector (j,k,l,m,n), is proposed to determine the best off-line threshold h:

1. Use j pixels of the no-change vegetation data (real world no-change data) to learn the parameters needed by the simulator (training set).

2. Use k no-change settlement pixels (real world no-change data) to estimate the parameters needed by the simulator (training set).

3. Now using the trained simulator, simulate l pixels of each class, and use them to create the 45 probability density functions that span a year.

4. Simulate m pixels of each class, and use those to create simulated change data, where the change point τ has density U$[1,300]$. The change is simulated by using linear blending over a six-month period [10].

5. Simulate n pixels of no-change vegetation pixels.

6. For each threshold h perform the CUSUM algorithm on each band and determine P_{D}, P_{FA}, $\mathbb{E}\{(T-\tau)^{+368}\}$ and A_e using the simulated change and no-change data.

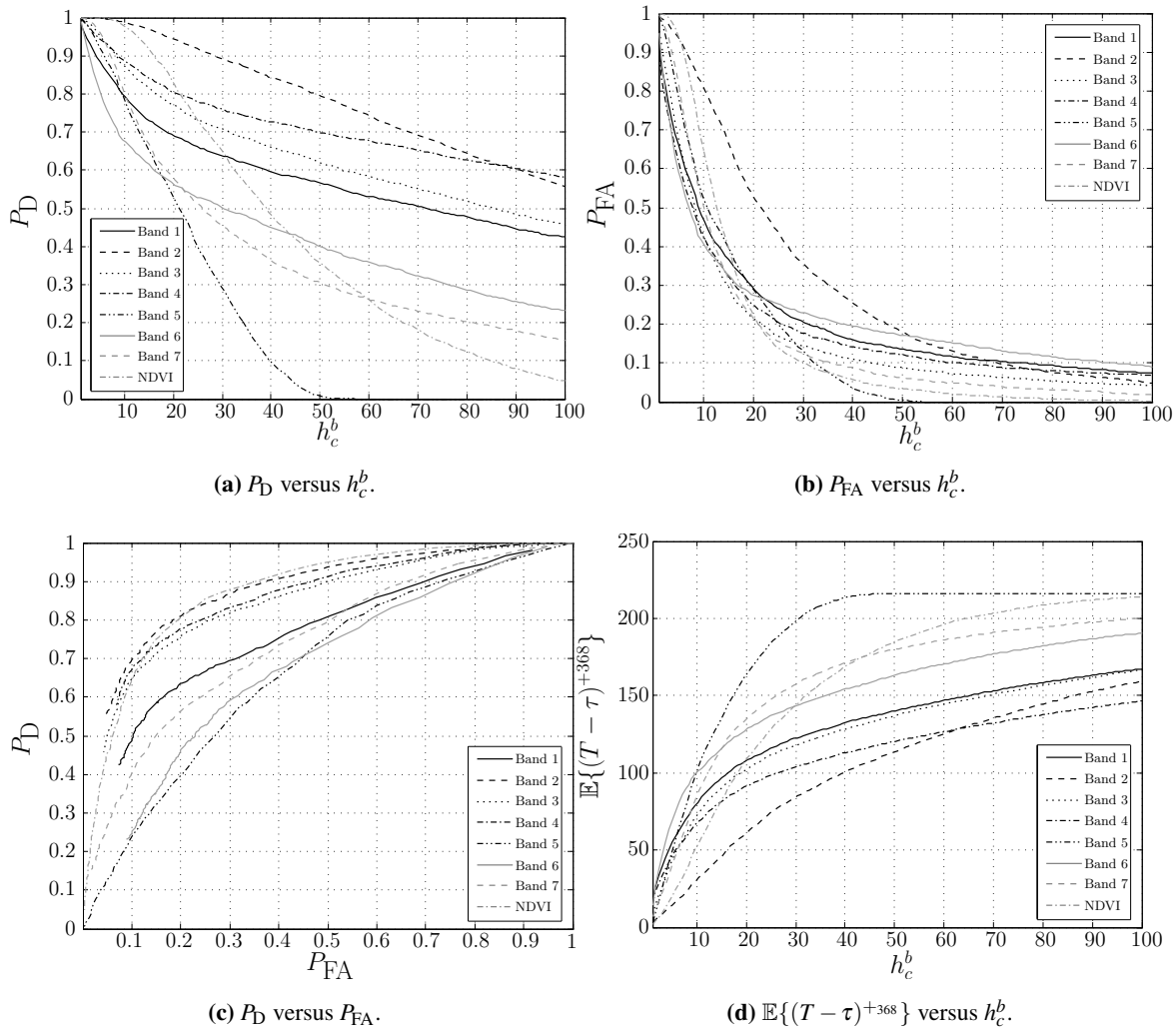
7. To determine the best h for each band, calculate the κ-coefficient Equation 5.7 (based on the number of correctly detected changes and the number of incorrectly detected changes) at each h in the sweeping interval and then select the h value that produces the largest κ-coefficient.

As an example of the output generated by the off-line optimisation algorithm, the resulting P_{D}, P_{FA} and $\mathbb{E}\{(T-\tau)^{+368}\}$ metrics, determined for the Gauteng data set with input vector $(592,333,3000,3000,3000)$, are shown in Figure 5.44.

The resulting P_{D}, P_{FA} and $\mathbb{E}\{(T-\tau)^{+368}\}$ generated by the off-line optimisation algorithm for the Limpopo dataset with input vector $(1497,1735,3000,3000,3000)$ is displayed in Figure 5.45.

The following observations can be made from Figure 5.44 and Figure 5.45:

1. The effect of h_c^b on the metrics P_{D}, P_{FA} and $\mathbb{E}\{(T-\tau)^{+368}\}$ is different for each value of b and each dataset.



(a) P_D versus h_c^b.

(b) P_{FA} versus h_c^b.

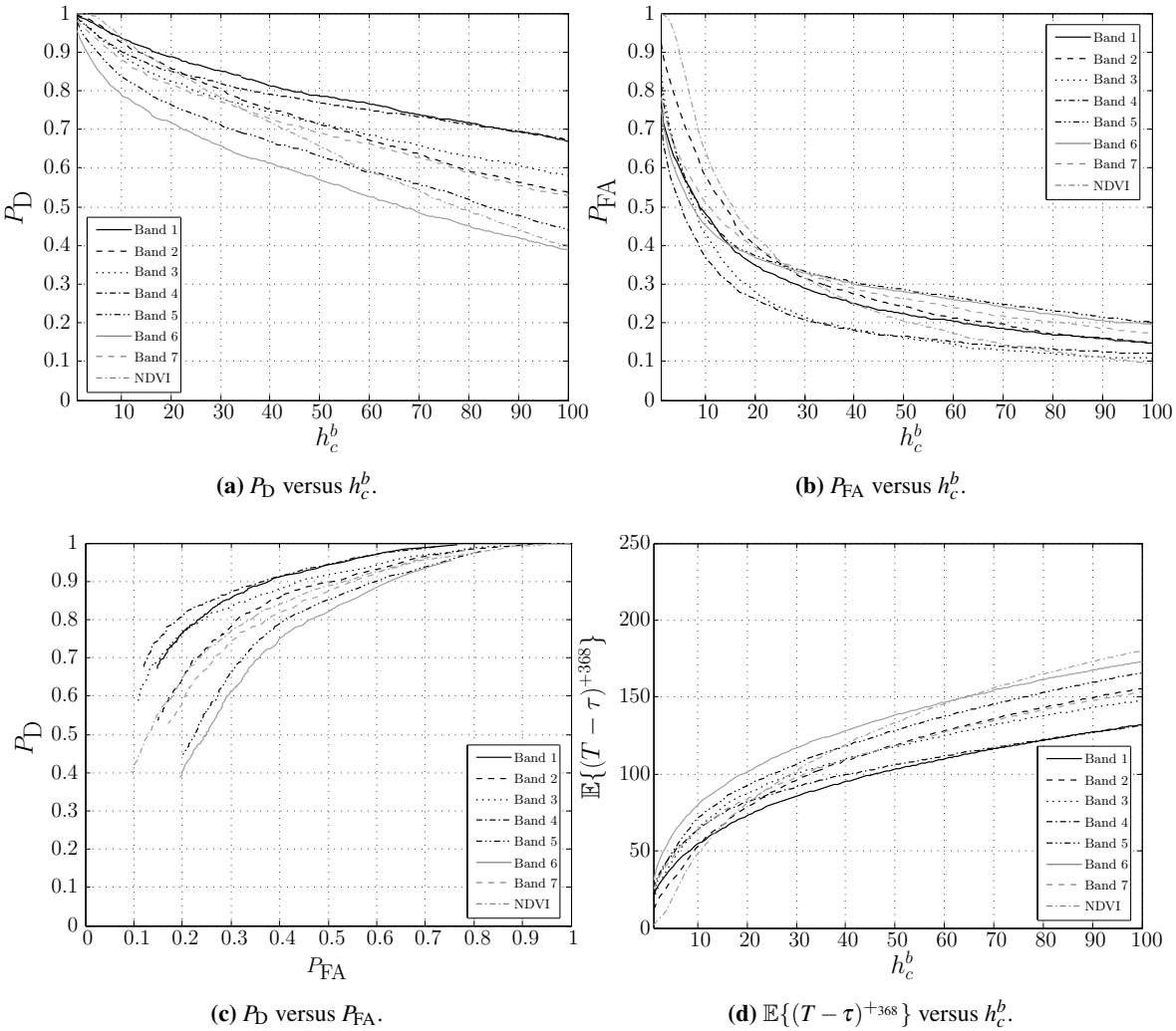
(c) P_D versus P_{FA}.

(d) $\mathbb{E}\{(T-\tau)^{+368}\}$ versus h_c^b.

Figure 5.44: Measured P_D, P_{FA} and $\mathbb{E}\{(T-\tau)^{+368}\}$ values for the simulated data in Gauteng [30] © IEEE 2012.

2. The most important graphs in Figure 5.44 and Figure 5.45 are the Receiver Operating Curves (ROCs) in Figure 5.44c and Figure 5.45c, since they display the probability of correctly detecting a change in the eight-year observation period against declaring a change during the observation period if none occurred. It is clear from Figure 5.44c that the CUSUM change detector produces higher change detection accuracies when using simulated data from bands $\{2,3,4,\text{NDVI}\}$ than if it uses simulated data from bands $\{1,5,6,7\}$. In Figure 5.45c it is clear that the CUSUM change detector achieves better change detection results if it uses simulated data from bands $\{1,2,3,4\}$ than if it uses simulated data from bands $\{5,6,7,\text{NDVI}\}$.

3. It is important to realise that the delay metric $\mathbb{E}\{(T-\tau)^{+368}\}$ presented in Figure 5.44d and



(a) P_D versus h_c^b.

(b) P_{FA} versus h_c^b.

(c) P_D versus P_{FA}.

(d) $\mathbb{E}\{(T-\tau)^{+368}\}$ versus h_c^b.

Figure 5.45: Measured P_D, P_{FA} and $\mathbb{E}\{(T-\tau)^{+368}\}$ values for the simulated data in Limpopo.

Figure 5.45d is an inefficient delay metric, as it is a truncated metric (and is easily affected by outliers). The metric is only reported here to be compatible with [30].

4. Better conclusions about the value of the threshold h_c^b, to optimise the delay of the CUSUM algorithm, can be made from the delay metric discussion in Section 5.3.6. The CUSUM algorithm and the time-varying maximum likelihood algorithm are based on the same underlying principles (see Chapter 3, Section 4.2.3 and Section 4.3.3). Since both approaches use the same underlying principles (CUSUM is merely a repeated time-varying SPRT–Section 4.3.3) both approaches are limited by the same expected decision delay imposed on them by the separability of the dataset to which they are applied. The one approach cannot suddenly be ten times faster than the other approach, as the underlying engines are based on the same design. As

seen in Figure 5.36 and Figure 5.40, the sequential threshold of the time-varying maximum likelihood classifier should be chosen in such a way that the classifier experiences a delay of at least a year (preferably two) before making a decision. A good design criterion for the CUSUM threshold h_c^b would therefore be to choose h_c^b in such a way that the expected decision delay (excluding outliers) of the CUSUM algorithm is at least one year (preferably two). Of course this delay is based on the fact that the change was instantaneous. The threshold should obviously be adapted to compensate for gradual change if required (meaning the expected delay of the detector should be increased).

To evaluate the performance of CUSUM on real world data, the metrics P_{D}, P_{FA} and A_e are used. No delay could be measured, as the true change points of the Gauteng and Limpopo datasets are unknown. The following methodology is proposed to determine the effectiveness of CUSUM on the Gauteng and Limpopo datasets:

1. Use the off-line optimisation algorithm with input vectors $(296, 333, 1000, 1000, 1000)$ and $(749, 1735, 1000, 1000,\ 1000)$ for the Gauteng and Limpopo datasets respectively, to determine the threshold h for each study region. Note that only 50% (random 50%) of the no change vegetation data was used to learn the parameters needed by the simulator and 50% of the real data was left for validation.

2. Apply the best h value to the no-change real vegetation data (validation dataset) and the real change data to determine P_{D} and P_{FA} (for each study region).

Random split cross-validation is performed by repeating the above experiment 50 times. The results of the random split cross validation experiments are displayed in Table 5.12. A training dataset and a validation dataset (equal in size), which were least correlated with each other (from all possible training and validation data sets) are also investigated. Since spatial independence is assumed by CUSUM, the worst case experiment is required to investigate whether spatial independence could be assumed (without detrimental effects) for the datasets under consideration. The results of the worst case experiment are found in Table 5.13.

The following observations and conclusion can be made from Table 5.12 and Table 5.13:

1. In the case of the real change Gauteng dataset the CUSUM algorithm achieves better change detection accuracies when applied to bands $\{1, 2, 3, 4\}$ than when the CUSUM algorithm is

Table 5.12: Random split cross-validation of CUSUM (50 experiments, 50% for training and 50% for validation) applied to the Gauteng and Limpopo datasets.

MODIS band	h_c^b	$\sigma_{h_c^b}$	P_{D}	$\sigma_{P_{\mathrm{D}}}$	P_{FA}	$\sigma_{P_{\mathrm{FA}}}$	A_e
Gauteng **Limpopo**							
1	37.33 49.77	6.34 8.03	0.9835 0.7695	0.0047 0.0167	0.1932 0.2483	0.0381 0.0281	0.1048 0.2394
2	55.88 39.19	5.87 6.46	0.8718 0.6736	0.0030 0.0144	0.1225 0.3070	0.0196 0.0281	0.1254 0.3167
3	28.67 36.90	5.55 8.42	0.9846 0.6638	0.0023 0.0376	0.1737 0.1920	0.0348 0.0307	0.0946 0.2641
4	38.02 38.53	7.29 10.77	0.9835 0.6866	0.0008 0.0308	0.1593 0.1930	0.0271 0.0311	0.0879 0.2532
5	16.30 22.78	2.78 5.90	0.7884 0.6860	0.0652 0.0360	0.5363 0.4517	0.0552 0.0352	0.3740 0.3829
6	18.05 19.06	5.21 5.30	0.2136 0.6986	0.0809 0.0416	0.3471 0.4700	0.0558 0.0456	0.5668 0.3857
7	19.27 40.72	2.88 7.45	0.7114 0.7236	0.0730 0.0281	0.2495 0.3180	0.0853 0.0408	0.2690 0.2972
NDVI	21.21 35.53	1.65 5.17	0.8170 0.7499	0.0502 0.0488	0.1484 0.4316	0.0561 0.0519	0.1657 0.3409

applied to bands $\{5,6,7,\mathrm{NDVI}\}$. Exactly the same behaviour is observed in the case of the real change Limpopo dataset. The CUSUM change detection algorithm performs much better on the Gauteng dataset than on the Limpopo dataset. The top performing band sets obtained for the off-line optimization algorithm (discussed earlier in this section) are practically the same as the top performing bands sets obtained when CUSUM is applied to the real change datasets. The discrepancy can be explained by the fact that the simulated change dataset is certainly not a carbon copy of the real change dataset and as such a guarantee can therefore not be given that CUSUM will produce exactly same results when applied to the simulated change dataset and the real change dataset.

2. Recall from Section 5.1.2.3 that for the Gauteng dataset the temporal Hellinger distance metric predicted that a change detector using data from bands $\{1,2,3,4\}$ would provide better change detection accuracies than a change detector that uses data from bands $\{5,6,7,\mathrm{NDVI}\}$ when the change detector in question relies on temporal features. From Table 5.12 it is clear that this prediction is accurate. This is no surprise, as a similar result is found in Section 5.3.3 and Section 5.3.4. A similar conclusion can be made when inspecting the real change Limpopo dataset.

3. As mentioned in Section 5.4.3, the CUSUM threshold is affected by the amount of temporal

Table 5.13: CUSUM applied to the worst possible correlated training set (for the Gauteng and Limpopo datasets).

MODIS band	h_c^b	P_{D}	P_{FA}	A_e
Gauteng / Limpopo				
1	23.59 / 60.80	0.9890 / 0.8034	0.2500 / 0.3178	0.1305 / 0.2572
2	65.45 / 44.19	0.8729 / 0.6752	0.1554 / 0.3284	0.1412 / 0.3266
3	29.57 / 61.46	0.9834 / 0.6410	0.1250 / 0.2510	0.0708 / 0.3050
4	39.54 / 32.23	0.9834 / 0.7778	0.1216 / 0.3418	0.0691 / 0.2820
5	11.63 / 26.91	0.9558 / 0.5983	0.7770 / 0.3578	0.4106 / 0.3798
6	12.96 / 34.22	0.2099 / 0.5299	0.3243 / 0.3845	0.5572 / 0.4273
7	14.29 / 43.52	0.8287 / 0.7009	0.5845 / 0.3204	0.3779 / 0.3098
NDVI	22.26 / 37.54	0.8122 / 0.8120	0.3311 / 0.5901	0.2595 / 0.3891

dependence in the data. The higher the dependence, the higher the threshold should be. This threshold phenomenon is observable in Table 5.12 when inspecting the top performing bands $\{1,2,3,4\}$ of the real change Gauteng dataset. The bands in $\{1,2,3,4\}$ that have a higher amount of dependence between their observations (which can be estimated from Table 5.6) have higher thresholds than the bands that have a lower amount of dependence between their observations. A similar conclusion can be drawn from the real change Limpopo dataset. It is however important to realise that the λ value is only an estimate of the amount of dependence between the observations of a band and can therefore not always be trusted to predict the CUSUM threshold as is seen in the case of the NDVI threshold for the real change Gauteng dataset.

4. As the results in Table 5.12 and Table 5.13 are similar, it shows that a spatial independent assumption is an allowable assumption for the current datasets.

5.4.5 Important change detection conclusions

The following important conclusions can be made from the results presented in Section 5.4:

1. From Table 5.10 and Table 5.11 it is clear that the band differencing algorithm performs better

if structure preservation is ignored when choosing the value of ν.

2. From Table 5.12 and Table 5.13 it is clear that the sequential change detection algorithm CUSUM can be effectively applied to MODIS. According to Figure 5.36 and Figure 5.40, a good design criterion for the CUSUM threshold h_c^b would be to choose h_c^b in such a way that the expected decision delay (excluding outliers) of the CUSUM algorithm is at least one year (preferably two).

3. According to Table 5.10 and Table 5.12 the CUSUM algorithm outperforms band differencing in the case of the Gauteng dataset. The exact opposite happens in the case of the Limpopo dataset. This is not surprising, since Figure 5.40 shows that the separability of the Limpopo dataset deteriorates over time. Since the CUSUM approach is a sequential algorithm, it is severely affected by this deterioration (more than band differencing that relies on detecting pixel outliers).

5.5 CONCLUSION

The chapter presented the classification and change detection accuracies and rankings of the different sequential and non-sequential hypertemporal classification and change detection algorithms investigated in this thesis. The most important conclusions from this chapter are summarised in Section 6.1.

CHAPTER 6

CONCLUSION

The results and conclusion of Chapter 5 are summarized in this chapter.

6.1 MAIN CONCLUSIONS

The following table highlights the most important sections contained in the thesis:

Table 6.1: The most important sections of the thesis.

Algorithm	Description	Results	Main conclusions
Simulation	Section 4.1.2	Section 5.2	Section 5.2.6
Minimum distance classifier	Section 4.2.2	Section 5.3	Section 5.3.7
Time-varying maximum likelihood classifier	Section 4.2.3		
$\boldsymbol{\theta}, \boldsymbol{\iota}$ and $\boldsymbol{\zeta}$	Section 4.2.4.2		
Band differencing	Section 4.3.2	Section 5.4.2	Section 5.4.5
CUSUM	Section 4.3.3	Section 5.4.4	

The sections listed in the right most column of Table 6.1 contain the most important conclusions made in the thesis.

6.2 SUMMARY OF WORK

In this thesis new hypertemporal techniques were proposed for the settlement detection problem in South Africa. The hypertemporal techniques were applied to study areas in the Gauteng and Limpopo

provinces of South Africa. To be more precise, new sequential (windowless) and non-sequential hypertemporal techniques were investigated. The time-series employed by the new hypertemporal techniques were obtained from the MODIS sensor, which is on board the earth observation satellites Aqua and Terra. One MODIS dataset was constructed for each province.

An SVM that uses a novel noise-harmonic feature set was implemented to detect existing human settlements. The noise-harmonic feature set is a non-sequential hypertemporal feature set and was constructed by using the CSHO. The CSHO consists of an SHO which is superimposed on the Ornstein-Uhlenbeck process. The noise-harmonic feature set is an extension of the classic harmonic feature set. The classic harmonic feature set consists of a mean and a seasonal component. For the case studies in this thesis, it is observed that the noise-harmonic feature set not only extends the harmonic feature set, but also improves on its classification capability.

The noise-harmonic feature SVM was also compared with the minimum distance classifier, the time-varying classifier (which is based on sequential analysis) and a temporal feature SVM. In general the noise-harmonic feature SVM outperformed the minimum distance classifier, the time-varying classifier and the temporal feature SVM. It is also worth mentioning that the noise-harmonic feature SVM performs much better than the other classifiers when the spectral dimension of the classifiers are low.

The CUSUM algorithm was developed by E.S. Page in 1954. In its original form it is a sequential (windowless) hypertemporal change detection technique. Windowed versions of the algorithm have been applied in a remote sensing context. In this thesis CUSUM was used in its original form to detect settlement expansion in South Africa and is benchmarked against the classic band differencing change detection approach of R.S. Lunetta et al. In the case of the Gauteng study area, the CUSUM algorithm outperformed the band differencing technique.

Sequential hypertemporal techniques are data-intensive and an inductive MODIS simulator was consequently also developed (to augment datasets). The proposed simulator is also based on the CSHO. Two case studies showed that the proposed inductive simulator accurately replicates the temporal dynamics and spectral dependencies found in MODIS data.

The main result of this thesis is that the noise-harmonic feature set and the CUSUM algorithm are promising hypertemporal techniques that may achieve good results (competitive) when applied in a remote sensing context. The main academic contribution of this thesis however was the successful

application of sequential hypertemporal techniques in the remote sensing field.

6.3 FUTURE WORK

The following is a list of possible work that can be done in the future to extend the results in this thesis:

1. The Shiryaev-Roberts stopping time, which is a sequential change detection algorithm, is defined in Equation 3.65. It would be interesting to compare the performance of the Shiryaev-Roberts stopping time when it is applied to the datasets in Section 2.8 with the performance results of the CUSUM (Section 4.3.3) algorithm presented in Section 5.4.4.

2. Figure 5.43 seems to indicate that if there is dependence between observations then the CUSUM statistic derived from the dependent observations exhibit sinusoidal behaviour. It would be worth investigating whether applying different filtering techniques to the CUSUM statistic could improve the performance of the CUSUM algorithm whenever it is applied to a dataset with dependent observations.

3. The AR(p) model is defined as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t,$$

where $\{\varphi_1, \ldots, \varphi_p\}$ are the parameters of the model, c is a constant, and ε_t is white noise. Since the Ornstein-Uhlenbeck proces is actually the continues-time analogue of the AR(1) process, it would make sense to also try and model the residual $\eta_c^b(t)$ with a higher order AR process. It would be interesting to determine whether the set $\{\varphi_1, \ldots, \varphi_p\}$ can provide good class discernibility.