CHAPTER SEVEN

EXTENDED KALMAN FILTER FEATURES

7.1 OVERVIEW

In this chapter, the Extended Kalman filter (EKF) is used as a feature extraction method, and is studied in-depth. The chapter discusses how the state space variables are used within the EKF, followed by how these are used to separate a set of time series into several classes. The importance of the initial parameters used to set the EKF is discussed in section 7.2.3, illustrating how the behaviour is dependent on these initial parameters.

A novel criterion called the Bias-Variance Equilibrium Point (BVEP), is proposed in section 7.2.4, which defines a desired set of initial parameters that will provide optimal performance. The BVEP criterion is derived using both the temporal and spatial information to design a system with desirable behaviour. A specifically designed search algorithm called the Bias-Variance Search Algorithm (BVSA) is proposed that will adjust the Bias-Variance Score (BVS) to best satisfy the BVEP criterion that will provide good initial parameters for the EKF. The chapter concludes by briefly overviewing the Autocovariance Least Squares (ALS) method, which will be used as benchmark when evaluating the method proposed in section 7.2.4.

7.2 CHANGE DETECTION METHOD: EXTENDED KALMAN FILTER

7.2.1 Introduction

An EKF is discussed as a feature extraction method in this section, which is based on the assumption that the parameters of the underlying model can be used to separate a set of time series into different classes. The model is based on the seasonal behaviour of a specific land cover class. It should be noted that a certain model would better describe a particular land cover class than another and that proper model selection must be done for each different land cover class. It follows that more separable

parameters derived by the EKF make it easier to detect changes in the assigned classes.

Lhermitte *et al.* proposed a method that separates different land cover classes using a Fourier analysis of NDVI time series [116]. It was concluded that good separation is achievable when evaluating the magnitude of the coefficients of the Fourier transform associated with the NDVI signal's mean and amplitude components. Kleynhans *et al.* proposed a method which jointly estimates the mean and seasonal component of the Fourier transform using a triply modulated cosine function [30]. The EKF uses the triply modulated cosine function to model NDVI time series by updating the mean (μ), amplitude (α), and phase (θ) parameters for each time increment.

The method proposed in this section expands on the method of Kleynhans [30] *et al.* by modelling the spectral bands separately and addresses the second constraint of the manual estimation of the initial parameters for the EKF to ensure proper tracking of the observation vectors. The initial parameters include the initial state-space vector, process noise covariance matrix and observation noise covariance matrix. An operator typically uses a training set to supervise the adjustment of the initial parameters until acceptable performance is obtained for a set of time series.

7.2.2 The method

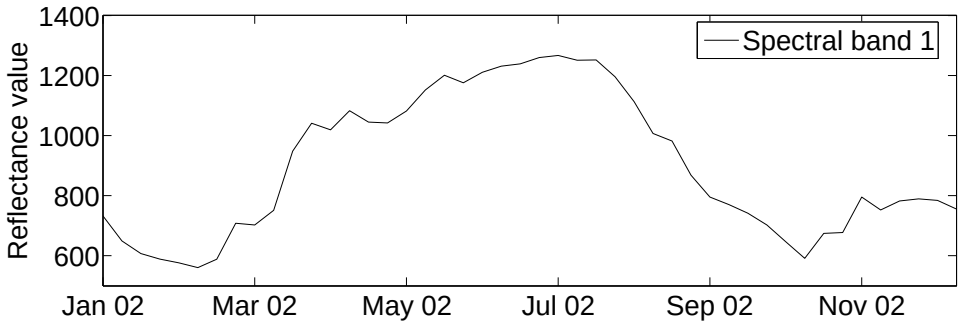
The EKF is a non-linear estimation method, which estimates the unobserved parameters using noisy observation vectors of a related observation model. The EKF has been used in the remote sensing community for parameter estimation of values related to physical, biogeochemical processes or vegetation dynamics models [204, 205].

In figure 7.1, a Fourier transform is used to observe that the majority of the signal energy is contained in the mean and seasonal component of the first spectral band. This implies that the time series in spectral band 1 are well represented in the time domain as a single cosine function with a mean offset, amplitude and phase, as shown in figure 7.2.

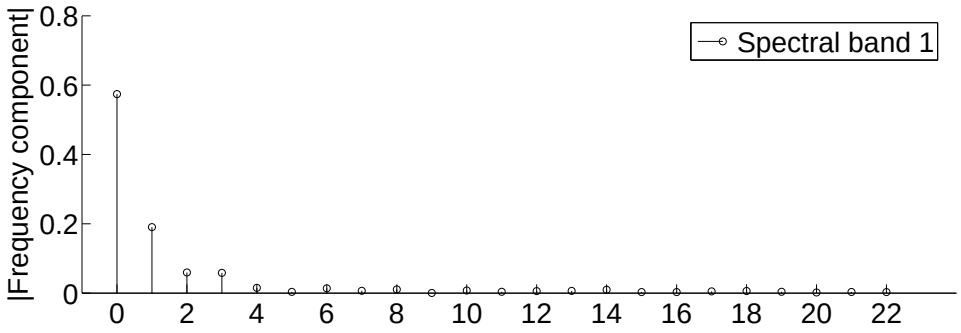
This single cosine model is, however, not a good representation if the time series is non-stationary, which is often the case; for example, inter-annual variability or land cover change. The triply modulated cosine function proposed in [30] is extended here to model a spectral band as

$$x_{i,k,b} = \mu_{i,k,b} + \alpha_{i,k,b} \cos(2\pi f_{\text{samp}} i + \theta_{i,k,b}) + v_{i,k,b}. \tag{7.1}$$

The variable $x_{i,k,b}$ denotes the observed value of the b^{th} spectral band's time series, $b \in \{1, 7\}$, of the k^{th} pixel, $k \in [1, N]$, at time index i, $i \in [1, \mathcal{I}]$. The noise sample of the k^{th} pixel at time i for each spectral band is denoted by $v_{i,k,b}$. The noise is additive with an unknown distribution on all the spectral bands. The cosine function model is separately fitted to each of the spectral bands and is based on several different parameters; the frequency f_{samp} can be explicitly calculated based on the annual



(a) Time series of reflectance values recorded by the MODIS spectral band 1.



(b) Discrete Fourier transform of the time series shown in (a).

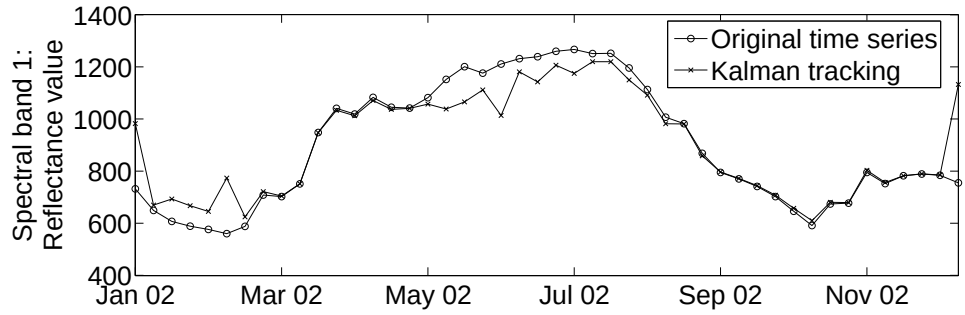
FIGURE 7.1: The time series recorded by the first spectral band for a geographical area is shown in (a) with the corresponding discrete Fourier transform shown in (b).

vegetation growth cycle, and the sampling rate of the MODIS sensor. Given the 8 daily composite MCD43A4 MODIS data set, f_{samp} is set to $\frac{8}{365}$. The non-zero mean of the b^{th} spectral band of the k^{th} pixel at time index i is denoted by $\mu_{i,k,b}$, the amplitude by $\alpha_{i,k,b}$ and the phase by $\theta_{i,k,b}$. The values of $\mu_{i,k,b}$, $\alpha_{i,k,b}$ and $\theta_{i,k,b}$ are dependent on time and must be estimated for each pixel k, $\forall k, k \in [1, N]$, given the spectral band observation vectors $x_{i,k,b}$ for i, $\forall i, i \in [1, \mathcal{I}]$, and b, $b \in \{1, 7\}$.

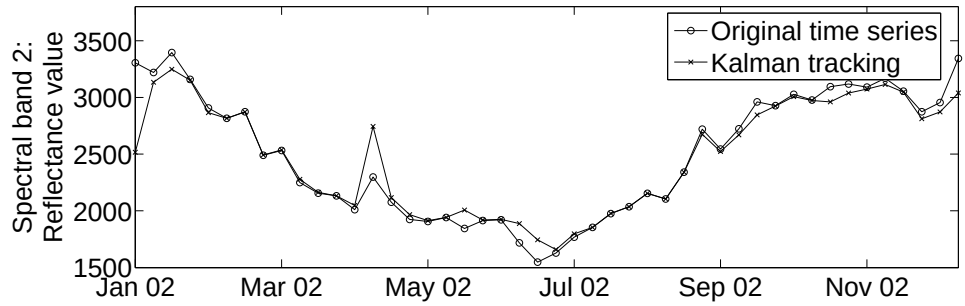
The MODIS spectral bands however are assumed to be uncorrelated and are treated independently in this method. The index b is omitted for convenience, with no loss in generality in the description of this method. A state-space vector is estimated by the EKF at each time increment i for each spectral band and contains all the parameters. This is expressed as

$$\vec{W}_{i,k} = [W_{i,k,1} \;\; W_{i,k,2} \;\; W_{i,k,3}] = [W_{i,k,\mu} \;\; W_{i,k,\alpha} \;\; W_{i,k,\theta}]. \tag{7.2}$$

For the present example of land cover classification, it is assumed that the state-space vector $\vec{W}_{i,k}$ does not change significantly through time; hence, the process model is linear. The measurement model, however, contains the cosine function and, as such, is evaluated via the standard Jacobian formulation, through linear approximation of the non-linear measurement function around the current state-space vector. The state-space vector $\vec{W}_{i,k}$ is related to the observation vector $x_{i,k}$ via a non-linear measurement function. Both the transition function and measurement function are assumed to be



(a) Extended Kalman filter tracking the observation vectors extracted from spectral band 1.



(b) Extended Kalman filter tracking the observation vectors extracted from spectral band 2.

FIGURE 7.2: The tracking of the first two spectral bands using the triply modulated cosine function.

non-perfect, so the addition of process and observation noise is required.

Converting state-space vectors to land cover classes

A machine learning algorithm is used to process the estimated state-space vectors to assign class labels. A class label is assigned to each state-space vector for each pixel at each time increment. This is expressed as

$$\mathcal{C}_{i,k} = \mathcal{F}_{\mathcal{C}}(W_{i,k,1}, \ldots, W_{i,k,S}) = \mathcal{F}_{\mathcal{C}}(\vec{W}_{i,k}), \tag{7.3}$$

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

$$\mathcal{C}_{i,k} \neq \mathcal{C}_{j,k}, \quad 0 \leq i \leq j, \forall i, j. \tag{7.4}$$

The importance of the initial parameters will be discussed in the next section.

7.2.3 Importance of the initial parameters

The EKF recursively solves the state-space form of a linear dynamic model [185, Ch. 1]. In this section the importance of the initial estimates of the system's variables is shown.

Let $\mathbf{x}_k = \{\vec{x}_{i,k}\}_{i=1}^{i=\mathcal{I}}$, $k \in [1, N]$, denote the k^{th} time series in the set of time series consisting of observation vectors, with each observation vector denoted by $\vec{x}_{i,k} = x_{i,k}$ as the spectral bands are treated independently. Let $\vec{W}_{i,k} = \{W_{i,k,s}\}_{s=1}^{s=S}$ denote the corresponding state-space vector for $x_{i,k}$. Then it is said that the EKF solves the state-space form recursively using the transition equation given as

$$\vec{W}_{i,k} = \mathbf{f}(\vec{W}_{(i-1),k}) + z_{(i-1),k}, \tag{7.5}$$

and the measurement equation given as

$$\vec{x}_{i,k} = \mathbf{h}\left(\vec{W}_{i,k}\right) + v_{i,k}. \tag{7.6}$$

The transition function is denoted by $\mathbf{f}$ and the measurement function is denoted by $\mathbf{h}$. A brief overview of the operations of the EKF which is shown in section 5.5 is revisited for convenience. It is well known from estimation theory that many prediction results simplify when Gaussian distributions are assumed. The process noise vector and observation noise vector are thus assumed to be Gaussian distributed. The process noise vector is thus denoted by $z_{(i-1),k}, z_{(i-1),k} \sim \mathcal{N}(0, \mathcal{Q}_{(i-1),k})$, and the observation noise vector is denoted by $v_{i,k}, v_{i,k} \sim \mathcal{N}(0, \mathcal{R}_{i,k})$.

The EKF recursively adapts the state-space vector for each incoming observation vector by predicting and updating the vector. In the prediction step the state-space vector $\vec{\hat{W}}_{(i|i-1),k}$ and covariance matrix $\mathfrak{B}_{(i|i-1),k}$ are predicted. The predicted state-space vector's estimate $\vec{\hat{W}}_{(i|i-1),k}$ is computed as

$$\vec{\hat{W}}_{(i|i-1),k} = \mathbf{f}\left(\vec{\hat{W}}_{(i-1|i-1),k}\right), \tag{7.7}$$

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

$$\mathfrak{B}_{(i|i-1),k} = \mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-1|i-1),k}\mathbf{F}_{\text{est}}^{\text{T}}. \tag{7.8}$$

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

$$\vec{\hat{W}}_{(i|i),k} = \vec{\hat{W}}_{(i|i-1),k} + \mathfrak{K}_{i,k}\left(\vec{x}_{i,k} - \mathbf{h}\left(\vec{W}_{i,k}\right)\right), \tag{7.9}$$

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

$$\mathfrak{K}_{i,k} = \mathfrak{B}_{(i|i-1),k}\mathbf{H}_{\text{est}}^{\text{T}}\mathcal{S}_{i,k}^{-1}. \tag{7.10}$$

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

$$\mathcal{S}_{i,k} = \mathbf{H}_{\text{est}}\mathfrak{B}_{(i|i-1),k}\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{i,k}. \tag{7.11}$$

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

$$\mathfrak{B}_{(i|i),k} = \mathfrak{B}_{(i|i-1),k} - \mathfrak{K}_{i,k}\mathcal{S}_{i,k}\mathfrak{K}_{i,k}^{\text{T}}. \tag{7.12}$$

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

$$\mathcal{E}_{\vec{x},i,k} = |\vec{x}_{i,k} - \vec{\hat{x}}_{i,k}| = \left|\vec{x}_{i,k} - \mathbf{h}\left(\vec{W}_{(i|i),k}\right)\right|. \tag{7.13}$$

In equation (7.13), it is observed that the state-space vector $\vec{\hat{W}}_{(i|i),k}$ determines the observation error $\mathcal{E}_{\vec{x},i,k}$. Thus the state-space vector $\vec{\hat{W}}_{(i|i),k}$ can be selected to minimise the observation error. The MODIS spectral bands are assumed to be uncorrelated and only produce a single reflectance value for each pixel. This simplifies equation (7.13) to

$$\mathcal{E}_{\vec{x},i,k} = |x_{i,k} - \hat{x}_{i,k}| = \left|x_{i,k} - \mathbf{h}\left(\vec{W}_{(i|i),k}\right)\right|. \tag{7.14}$$

The observation error is easily minimised by significantly varying $\vec{\hat{W}}_{(i|i),k}$ to accommodate the fluctuation in observation vectors. This does not bode well if the underlying structure of the system is also being analysed. A significantly varying state-space vector $\vec{\hat{W}}_{(i|i),k}$ is indicative of an unstable model. The conclusion is that the state-space model must be kept stable, while also attempting to minimise equation (7.14).

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

$$\hat{\vec{W}}_{(0|0),k} = \underset{\vec{W}\in\mathcal{W}}{\operatorname{argmin}} \left\{ \left| \vec{x}_{0,k} - \mathbf{h}\left(\vec{W}\right) \right| \right\}, \tag{7.15}$$

then

$$\mathcal{E}_{\vec{x},0,k} = \left| \vec{x}_{0,k} - \mathbf{h}\left(\hat{\vec{W}}_{(0|0),k,b}\right) \right|. \tag{7.16}$$

The recursive adaptation of the state-space vector's estimate $\hat{\vec{W}}_{(i|i),k}$ is then calculated using the predicted step given in equation (7.7) and the updating step in equation (7.9). Equation (7.7) is substituted into equation (7.9) to yield

$$\hat{\vec{W}}_{(i|i),k} = \mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right) + \mathfrak{K}_{i,k}\left(\vec{x}_{i,k} - \mathbf{h}\left(\mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right)\right)\right). \tag{7.17}$$

The Kalman gain $\mathfrak{K}_{i,k}$ determines the rate of change in the error between the predicted and estimated state-space vector. If the observation error is large and the Kalman gain is large, then large changes will be made to the current state-space vector. If the observation error is large and the Kalman gain is small, then the state-space's estimate $\hat{\vec{W}}_{(i|i),k}$ will adapt slowly, which typically leads to a large observation error $\mathcal{E}_{\vec{x},i,k}$ (equation (7.13)) until it eventually converges. If the observation error is small and the Kalman gain is large, then the state-space vector will struggle to converge, as it will continually overshoot the desired state-space vector that will minimise equation (7.13). Substituting the optimal Kalman gain given in equation (7.10) into equation (7.17) expands it to

$$\hat{\vec{W}}_{(i|i),k} = \mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right) + \mathfrak{B}_{(i|i-1),k}\mathbf{H}_{\text{est}}^{\text{T}}\mathcal{S}_{i,k}^{-1}\left(\vec{x}_{i,k} - \mathbf{h}\left(\mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right)\right)\right). \tag{7.18}$$

The Kalman gain is dependent on the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ and innovation term $\mathcal{S}_{i,k}$. The innovation term controls the trust region within the state-space vector's space. This is dependent on the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ and observation covariance noise $\mathcal{R}_{i,k}$. Substituting the innovation term given in equation (7.11) into equation (7.18) results in

$$\begin{aligned} \hat{\vec{W}}_{(i|i),k} \quad = \quad & \mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right) + \mathfrak{B}_{(i|i-1),k}\mathbf{H}_{\text{est}}^{\text{T}}(\mathbf{H}_{\text{est}}\mathfrak{B}_{(i|i-1),k}\mathbf{H}_{\text{est}}^{\text{T}} + \\ & \mathcal{R}_{i,k})^{-1}\left(\vec{x}_{i,k} - \mathbf{h}\left(\mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right)\right)\right). \end{aligned} \tag{7.19}$$

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

$$\begin{aligned}\hat{\vec{W}}_{(i|i),k} &= \mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right) + (\mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-1|i-1),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} \\ &\quad (\mathbf{H}_{\text{est}}(\mathcal{Q}_{(i-1),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-1|i-1),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{i,k})^{-1} \\ &\quad \left(\vec{x}_{i,k} - \mathbf{h}\left(\mathbf{f}\left(\hat{\vec{W}}_{(i-1|i-1),k}\right)\right)\right). \end{aligned} \tag{7.20}$$

The transition function $\mathbf{f}$ and measurement function $\mathbf{h}$ are assumed to be known. The observation vector $\vec{x}_{i,k}$ is supplied by the real system. The only variables left within equation (7.20) are: (1) previous state-space vector's estimate $\hat{\vec{W}}_{(i-1|i-1),k}$, (2) process noise's covariance matrix $\mathcal{Q}_{(i-1),k}$, (3) previous estimate of covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$, and (4) observation noise's covariance matrix $\mathcal{R}_{i,k}$.

The previous estimation of the covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$ will be briefly explored, as it is part of equation (7.20). The covariance matrix $\mathfrak{B}_{(i-1|i-1),k}$ is updated with

$$\mathfrak{B}_{(i-1|i-1),k} = \mathfrak{B}_{(i-1|i-2),k} - \mathfrak{K}_{(i-1),k}\mathcal{S}_{(i-1),k}\mathfrak{K}_{(i-1),k}^{\text{T}}. \tag{7.21}$$

Substituting the Kalman gain of equation (7.10) into equation (7.21) yields

$$\mathfrak{B}_{(i-1|i-1),k} = \mathfrak{B}_{(i-1|i-2),k} - (\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}}\mathcal{S}_{(i-1),k}^{-1})\mathcal{S}_{(i-1),k}(\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}}\mathcal{S}_{(i-1),k,b}^{-1})^{\text{T}}. \tag{7.22}$$

Substituting the innovation term of equation (7.11) into equation (7.22) yields

$$\begin{aligned}\mathfrak{B}_{(i-1|i-1),k} &= \mathfrak{B}_{(i-1|i-2),k} - (\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}}(\mathbf{H}_{\text{est}}\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})^{-1}) \\ &\quad (\mathbf{H}_{\text{est}}\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})(\mathfrak{B}_{(i-1|i-2),k}\mathbf{H}_{\text{est}}^{\text{T}}(\mathbf{H}_{\text{est}}\mathfrak{B}_{(i-1|i-2),k} \\ &\quad \mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})^{-1})^{\text{T}}. \end{aligned} \tag{7.23}$$

The predicted covariance matrix $\mathfrak{B}_{(i-1|i-2),k}$ given in equation (7.8) is substituted into equation (7.23), which yields

$$\begin{aligned}\mathfrak{B}_{(i-1|i-1),k} &= (\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}}) - ((\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} \\ &\quad (\mathbf{H}_{\text{est}}(\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})^{-1})(\mathbf{H}_{\text{est}}(\mathcal{Q}_{(i-2),k} + \\ &\quad \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})((\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} \\ &\quad (\mathbf{H}_{\text{est}}(\mathcal{Q}_{(i-2),k} + \mathbf{F}_{\text{est}}\mathfrak{B}_{(i-2|i-2),k}\mathbf{F}_{\text{est}}^{\text{T}})\mathbf{H}_{\text{est}}^{\text{T}} + \mathcal{R}_{(i-1),k})^{-1})^{\text{T}}. \end{aligned} \tag{7.24}$$

Equation (7.20) is computed for every newly obtained observation vector. The state-space vector's estimate $\hat{\vec{W}}_{(i|i),k}$ requires the results from equation (7.24) to compute the current estimates. The transition function $\mathbf{F}_{\text{est}}$ and measurement function $\mathbf{H}_{\text{est}}$ are known, then the only variables left to compute in equation (7.24) are: (1) initial covariance matrix $\mathfrak{B}_{(0|0),k}$, (2) process covariance matrix $\mathcal{Q}_{(i-1),k}$, and (3) observation noise's covariance matrix $\mathcal{R}_{i,k}$. The conclusion from equation (7.20) and equation (7.24) is that the initial parameters of importance are:

1. the initial state-space vector's estimate $\hat{\vec{W}}_{(0|0),k}$,
2. the initial covariance matrix estimate $\mathfrak{B}_{(0|0),k}$,
3. the process covariance matrix $\mathcal{Q}_{(i-1),k}$, and
4. the observation covariance matrix $\mathcal{R}_{i,k}$.

The initial state-space vector's estimate $\hat{\vec{W}}_{(0|0),k}$ is initialised using equation (7.15). Even if an incorrect estimate is used, the state-space vector $\hat{\vec{W}}_{(i|i),k}$ should converge to the correct vector as $i \to \infty$. The same is true about the initial covariance matrix $\mathfrak{B}_{(0|0),k}$. As $i \to \infty$, the covariance matrix $\mathfrak{B}_{(i|i),k}$ should tend to converge to the correct matrix. The usual operation of the EKF sets the initial covariance matrix equal to an identity matrix.

The initial covariance matrix $\mathfrak{B}_{(0|0),k}$ will stabilise, as equation (7.8) is known as a discrete Riccati equation, and under certain circumstances will converge, which results in equation (7.24) converging to a stable state [206]. The conditions for convergences of the discrete Riccati equation are:

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

$$\operatorname{rank}\left[z_{(i-1),k} | \mathbf{F}_{\text{est}} z_{(i-1),k} | \mathbf{F}_{\text{est}}^2 z_{(i-1),k} | \dots | \mathbf{F}_{\text{est}}^{N-1} z_{(i-1),k}\right] = N, \tag{7.25}$$

4. and the pair $(\mathbf{F}_{\text{est}}, \mathbf{H}_{\text{est}})$ is observable, *i.e.*,

$$\operatorname{rank}\left[\mathbf{H}_{\text{est}}^{\text{T}} | \mathbf{F}_{\text{est}}^{\text{T}} \mathbf{H}_{\text{est}}^{\text{T}} | (\mathbf{F}_{\text{est}}^{\text{T}})^2 \mathbf{H}_{\text{est}}^{\text{T}} | \dots | (\mathbf{F}_{\text{est}}^{\text{T}})^{N-1} \mathbf{H}_{\text{est}}^{\text{T}}\right] = N, \tag{7.26}$$

with $N \in \mathbb{N}$. Under the above conditions, the predicted covariance matrix $\mathfrak{B}_{(i|i-1),k}$ converges to a constant matrix

$$\lim_{i\to\infty} \mathfrak{B}_{(i|i-1),k} = \mathfrak{B}_{\text{const}}, \tag{7.27}$$

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

The system can also estimate $\hat{\vec{W}}_{(0|0),k}$ and $\mathfrak{B}_{(0|0),k}$ using an offline training phase. Offline refers to observation vectors that are stored and are used recursively for estimation. The process covariance matrix $\mathcal{Q}_{(i-1),k}$ and observation covariance matrix $\mathcal{R}_{i,k}$ are assumed to be constant throughout the recursive estimation of the observation vector. This is usually manually set by a system analyst in an offline training phase through successive adjustments. In this thesis the initial EKF is defined as:

1. The initial state-space vector $\hat{\vec{W}}_{(0|0),k}$ is estimated offline.
2. The initial covariance matrix $\mathfrak{B}_{(0|0),k}$ is estimated offline.
3. The process covariance matrix $\mathcal{Q}_{(i-1),k}$ is set to a fixed matrix.
4. The observation covariance matrix $\mathcal{R}_{i,k}$ is set to a fixed matrix.

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

7.2.4 Bias-Variance Equilibrium Point

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

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

$$\mathcal{R}_{i,k} = E[(x_{i,k} - E[x_{i,k}])^2]. \tag{7.28}$$

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

$$\mathcal{Q}_{i,k} = \begin{pmatrix} E[(W_{i,k,1}-E[W_{i,k,1}])(W_{i,k,1}-E[W_{i,k,1}])] & \dots & E[(W_{i,k,1}-E[W_{i,k,1}])(W_{i,k,S}-E[W_{i,k,S}])] \\ \vdots & \ddots & \vdots \\ E[(W_{i,k,S}-E[W_{i,k,S}])(W_{i,k,1}-E[W_{i,k,1}])] & \dots & E[(W_{i,k,S}-E[W_{i,k,S}])(W_{i,k,S}-E[W_{i,k,S}])] \end{pmatrix}. \quad (7.29)$$

The state-space variables within the state-space vector are assumed to be uncorrelated; the process covariance matrix simplifies to

$$\mathcal{Q}_{i,k} = \text{diag}\{E[(W_{i,k,s}-E[W_{i,k,s}])^2]\}, \ \ \forall s. \quad (7.30)$$

The setting of the initial parameters has a major effect on the overall system performance. The initial state-space vector $\vec{W}_{(0|0),k}$ for the first observation vector $\vec{x}_{0,k}$ is optimised using equation (7.15). The initial estimated covariance matrix $\mathfrak{B}_{(0|0),k}$ is usually set to the identity matrix. This only leaves the estimation of the observation covariance matrix $\mathcal{R}_{i,k}$ and process covariance matrix $\mathcal{Q}_{i,k}$. Let the uncorrelated observation covariance matrix's diagonals be placed into a vector called the observation candidate vector $\Upsilon_{\mathcal{R},i,k}$, were $\Upsilon_{\mathcal{R},k}$ is selected from the space $\upsilon_{\mathcal{R}}$, and it is expressed as

$$\Upsilon_{\mathcal{R},i,k} = 10^{\zeta_{i,k}/10}, \quad (7.31)$$

with

$$\zeta_{i,k} = 10\log_{10}\left(E[(\vec{x}_{i,k}-E[\vec{x}_{i,k}])^2]\right). \quad (7.32)$$

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

$$\Upsilon_{\mathcal{Q},i,k} = 10^{[\varsigma_{i,k,1} \ \dots \ \varsigma_{i,k,S}]/10} = 10^{\vec{\varsigma}_{i,k}/10}, \quad (7.33)$$

with

$$\varsigma_{i,k,s} = 10\log_{10}\left(E[(W_{i,k,s}-E[W_{i,k,s}])^2]\right). \quad (7.34)$$

It should be noted that the EKF only updates recursively the state-space vector $\vec{W}_{(i|i),k}$, and covariance matrix $\mathfrak{B}_{(i|i),k}$. The time index of the observation covariance matrix $\mathcal{Q}_{i,k}$ has been left

inserted to emphasise the time effect in a dynamic linear system. The EKF, however, does not alter the observation covariance matrix at each time increment and is thus constant for all time indices. This is formally stated as $\mathcal{Q}$=$\mathcal{Q}_i$, $\forall i$. The process covariance matrix is also retained as a constant for all time indices and this is stated as $\mathcal{R}$=$\mathcal{R}_i$, $\forall i$. This concludes that the observation covariance matrix and process covariance matrix are independent of time. This property allows the observation candidate vector to be rewritten as

$$\Upsilon_{\mathcal{R},k} = 10^{\zeta_k/10} \quad \forall k, \tag{7.35}$$

and the process candidate vector rewritten as

$$\Upsilon_{\mathcal{Q},k} = 10^{[\varsigma_{k,1} \ \cdots \ \varsigma_{k,S}]/10} = 10^{\vec{\varsigma}_k/10} \quad \forall k. \tag{7.36}$$

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

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

$$\sigma_{\mathcal{E}} = \min_{\Upsilon_{\mathcal{R},k}\in\upsilon_{\mathcal{R}},\Upsilon_{\mathcal{Q},k}\in\upsilon_{\mathcal{Q}}} \left\{ \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| \hat{x}_{i,k} - x_{i,k} \right\| \right\}, \tag{7.37}$$

then

$$[\mathcal{R}_{\sigma_{\mathcal{E}}}, \mathcal{Q}_{\sigma_{\mathcal{E}}}] = \underset{\Upsilon_{\mathcal{R},k}\in\upsilon_{\mathcal{R}},\Upsilon_{\mathcal{Q},k}\in\upsilon_{\mathcal{Q}}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| \hat{x}_{i,k} - x_{i,k} \right\| \right\}. \tag{7.38}$$

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

$$\sigma_{\mathcal{E}} = \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| \hat{x}_{i,k} - x_{i,k} \right\| \Bigg|_{\mathcal{R}=\mathcal{R}_{\sigma_{\mathcal{E}}},\mathcal{Q}=\mathcal{Q}_{\sigma_{\mathcal{E}}}}. \tag{7.39}$$

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

$$\sigma_s = \min_{\Upsilon_{\mathcal{R},k}\in\upsilon_{\mathcal{R}},\Upsilon_{\mathcal{Q},k}\in\upsilon_{\mathcal{Q}}} \left\{ \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| W_{i,k,s} - E[W_{i,k,s}] \right\| \right\}, \quad \forall s, \tag{7.40}$$

then

$$[\mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}] = \underset{\Upsilon_{\mathcal{R},k}\in\upsilon_{\mathcal{R}},\Upsilon_{\mathcal{Q},k}\in\upsilon_{\mathcal{Q}}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| W_{i,k,s} - E[W_{i,k,s}] \right\| \right\}, \quad \forall s. \tag{7.41}$$

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

$$\sigma_s = \sum_{k=1}^{N}\sum_{i=1}^{\mathcal{I}} \left\| W_{i,k,s} - E[W_{i,k,s}] \right\| \Bigg|_{\mathcal{R}=\mathcal{R}_{\sigma_s},\mathcal{Q}=\mathcal{Q}_{\sigma_s}}. \tag{7.42}$$

The spatial information is included through the use of a set of time series all located in a specific geographical area. The set of N time series for a geographical area is denoted by $\{\vec{x}_{i,k}\}$. Let $q_{i,\mathcal{E}}$ denote the probability density function derived at time index i from the residuals given over the set of observations $\{x_{i,k}\}_{k=1}^{k=N}$ such that $P[a \leq \mathcal{E} \leq b] = \int_a^b f(e)\mathrm{d}e = \int_a^b f(e,\mathcal{R},\mathcal{Q})\mathrm{d}e$ i.e.,

$$P[a \leq \mathcal{E} \leq b] = \int_a^b q(e,\mathcal{R},\mathcal{Q})\mathrm{d}e = \int_a^b q_{i,\mathcal{E}}\mathrm{d}e. \tag{7.43}$$

Let $q_{i,s}$ denote the probability density function for the state-space variable s derived at time index i from the deviations given over the set of state-space vectors $\{W_{i,k,s}\}_{k=1}^{k=N}$ such that $P[a \leq s \leq b] = \int_a^b f(s')\mathrm{d}s' = \int_a^b f(s',\mathcal{R},\mathcal{Q})\mathrm{d}s'$ i.e.,

$$P[a \leq s \leq b] = \int_a^b q(s',\mathcal{R},\mathcal{Q})\mathrm{d}s' = \int_a^b q_{i,s}\mathrm{d}s'. \tag{7.44}$$

A conditioned observation probability density function $q^*_{i,\mathcal{E}}$ is defined as the probability density function $q_{i,\mathcal{E}}$ in equation (7.43), which uses the set $[\mathcal{R}_{\sigma_\mathcal{E}}, \mathcal{Q}_{\sigma_\mathcal{E}}]$ to satisfy the condition given in equation (7.39) as

$$P[a \leq \mathcal{E} \leq b] = \int_a^b q(e,\mathcal{R}_{\sigma_\mathcal{E}},\mathcal{Q}_{\sigma_\mathcal{E}})\mathrm{d}e = \int_a^b q^*_{i,\mathcal{E}}\mathrm{d}e. \tag{7.45}$$

A conditioned process probability density function $q^*_{i,s}$ is defined as the probability density function $q_{i,s}$ in equation (7.44), which uses the set $[\mathcal{R}_{\sigma_s}, \mathcal{Q}_{\sigma_s}]$ to satisfy the condition given in equation (7.42) as

$$P[a \leq\ s \leq b] = \int_a^b q(s',\mathcal{R}_{\sigma_s},\mathcal{Q}_{\sigma_s})\mathrm{d}s' = \int_a^b q^*_{i,s}\mathrm{d}s'. \tag{7.46}$$

The performance of the current estimate $\Upsilon_{\mathcal{R},k}$ and $\Upsilon_{\mathcal{Q},k}$ is defined by a criterion that evaluates how

well the conditions stated in equation (7.37) and equation (7.40) are satisfied. The current estimates are recursively updated and are denoted by $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$, where ι denotes the iteration number. The current estimates $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$ are used to derive the set of probability density functions $\{\hat{q}^{\iota}_{i,\mathcal{E}}\}$, $\forall i$, and $\{\hat{q}^{\iota}_{i,s}\}$, $\forall i$.

A f-divergent distance known as the Hellinger distance [207, 208] is used to measure the similarity between the probability density functions $\hat{q}^{\iota}_{i,\mathcal{E}}$ and $q^{*}_{i,\mathcal{E}}$. The modified Hellinger distance $\mathcal{H}_{i,\mathcal{E}}$, $\mathcal{H}_{i,\mathcal{E}} \in [0, 1]$, is computed as

$$\mathcal{H}_{i,\mathcal{E}} = 1 - \sqrt{1 - \sqrt{\int_{-\infty}^{\infty} \hat{q}^{\iota}_{i,\mathcal{E}}\, q^{*}_{i,\mathcal{E}} \mathrm{d}e}}, \tag{7.47}$$

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

$$\mathcal{H}_{i,s} = 1 - \sqrt{1 - \sqrt{\int_{-\infty}^{\infty} \hat{q}^{\iota}_{i,s}\, q^{*}_{i,s} \mathrm{d}s'}}, \tag{7.48}$$

where a value of $\mathcal{H}_{i,s} \rightarrow 1$ means high similarity between $\hat{q}^{\iota}_{i,b,s}$ and $q^{*}_{i,b,s}$, while $\mathcal{H}_{i,s} \rightarrow 0$ means low similarity.

The BVS is defined to encapsulates all similarity metrics as

$$\Gamma_i = \min\Big(\{\mathcal{H}_{i,s}\}_{s=1}^{s=S} \cup \{\mathcal{H}_{i,\mathcal{E}}\}\Big). \tag{7.49}$$

Finding optimal estimates for $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$ requires a stable covariance matrix $\mathfrak{B}_{(i|i),k}$. Equation (7.27) states that the predicted covariance matrix $\mathfrak{B}_{(i|i),k}$ should converge to a constant matrix under certain prerequisite conditions. Let $\mathcal{I}_T$, $\mathcal{I}_T \ll \mathcal{I}$, denote the number of time steps required to ensure that the predicted covariance matrix $\mathfrak{B}_{(\mathcal{I}_T|\mathcal{I}_T-1),k}$ converges to ensure a stable covariance matrix $\mathfrak{B}_{(\mathcal{I}_T|\mathcal{I}_T),k}$. The BVS is deemed accurate at $\mathcal{I}_T$, which is defined as

$$\Gamma_{\mathcal{I}_T} = \min\Big(\{\mathcal{H}_{\mathcal{I}_T,s}\}_{s=1}^{s=S} \cup \{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\}\Big). \tag{7.50}$$

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

$$\Gamma^{*}_{\mathcal{I}_T} = \max_{\Upsilon^{\iota}_{\mathcal{R},k} \in \upsilon_{\mathcal{R}}, \Upsilon^{\iota}_{\mathcal{Q},k} \in \upsilon_{\mathcal{Q}}} \{\Gamma_{\mathcal{I}_T}\}. \tag{7.51}$$

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

for this as

$$\Gamma_{\mathcal{I}_T} = \min\left\{\left\{\{\mathcal{H}_{\mathcal{I}_T,b,s}\}_{s=1}^{s=S}\right\}_{b=1}^{b=B}\{\mathcal{H}_{\mathcal{I}_T,b,\mathcal{E}}\}_{b=1}^{b=B}\right\}. \tag{7.52}$$

In this thesis however it was assumed that the spectral bands were uncorrelated.

7.2.5 Bias-Variance Search algorithm

The BVSA is proposed in this section, which will attempt to estimate $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$ to satisfy the BVEP criterion using the BVS given in equation (7.50). The BVSA starts by creating ideal operating conditions for each parameter in the EKF, followed by using a hill-climbing algorithm to search for a set of $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$ that will satisfy at best the ideal operating conditions for all the parameters within the EKF.

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

$$q^*_{i,\mathcal{E}} = \left\{q_{i,\mathcal{E}} : \{\zeta_k\} \to -\infty; \{\varsigma_{k,s}\} \to \infty, \forall\, s\right\}. \tag{7.53}$$

Under perfect conditions the probability density function $q^*_{i,\mathcal{E}}$ should tend to be an impulse of unity power situated around the zero position, meaning zero error residual is measured.

The second ideal condition is a system that employs a stable state-space variable. This ideal condition is used to create the probability density function $q^*_{i,s}$. This is obtained by

$$q^*_{i,s} = \left\{q_{i,s} : \{\zeta_k\} \to \infty; \{\varsigma_{k,\{s\}\setminus s}\} \to \infty; \{\varsigma_{k,s}\} \to -\infty\right\}. \tag{7.54}$$

This condition creates an environment which attempts to track the state-space variable s with the smallest variation.

After the ideal observation conditions' probability density functions $q^*_{i,\mathcal{E}}$ and $q^*_{i,s}$ have been computed, a hill-climbing search algorithm is applied to find a set of initial parameters that will best satisfy all these ideal conditions. The BVSA iteratively searches the parameter space and is described briefly below.

Step 1: The BVSA starts with the initial parameters set as $\zeta^0_k = 0$dB, $\forall\, k$, and $\varsigma^0_{k,s} = 0$dB, $\forall\, k, s$.

Step 2: Compute the state-space vector $\vec{W}_{(\mathcal{I}_T|\mathcal{I}_T),k}$ at time $\mathcal{I}_T$ using the same $\hat{\Upsilon}^{\iota}_{\mathcal{R},k} = \zeta^{\iota}_k$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k} = \{\zeta^{\iota}_k\}_{s=1}^{s=S}$ for every time series in the set $\{\mathbf{x}_k\}_{k=1}^{k=N}$.

Step 3: Obtain the probability density function of the residual errors $q^{\iota}_{i,\mathcal{E}}$ over the N time series at time index $\mathcal{I}_T$.

Step 4: Obtain the probability density function of the residual errors $q^{\iota}_{i,s}$ of the state-space variable s over the N time series at time index $\mathcal{I}_T$.

Step 5: Compute the modified Hellinger distance $\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}$ as shown in equation (7.47).

Step 6: Compute the modified Hellinger distance $\mathcal{H}_{\mathcal{I}_T,s}$ as shown in equation (7.48).

Step 7: Determine the best performing condition $\mathcal{H}_{\text{best}}$ as

$$\mathcal{H}_{\text{best}} = \max\left\{\{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\}\ \{\mathcal{H}_{\mathcal{I}_T,s}\}\right\}. \tag{7.55}$$

Step 8: Determine the worst performing condition $\mathcal{H}_{\text{worst}}$ as

$$\mathcal{H}_{\text{worst}} = \min\left\{\{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}}\}\ \{\mathcal{H}_{\mathcal{I}_T,s}\}\right\}. \tag{7.56}$$

Step 9: Adjust the new ζ^{ι}_k according to its relative position to the best and worst performing parameters using a threshold $\rho_{\mathcal{H}}$, $\rho_{\mathcal{H}} \in [0,1]$, $\rho_{\mathcal{H}} \in \mathbb{R}$. The adjustment is made as

$$\zeta^{\iota+1}_k = \begin{cases} \zeta^{\iota}_k + \gamma^{\iota} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} > \rho_{\mathcal{H}}\right) \\ \zeta^{\iota}_k - \gamma^{\iota} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,\mathcal{E}} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} \leq \rho_{\mathcal{H}}\right) \end{cases}. \tag{7.57}$$

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

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

$$\varsigma^{\iota+1}_{k,s} = \begin{cases} \varsigma^{\iota}_{k,s} + \gamma^{\iota} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,s} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} > \rho_{\mathcal{H}}\right) \\ \varsigma^{\iota}_{k,s} - \gamma^{\iota} & \text{if } \left(\frac{\mathcal{H}_{\mathcal{I}_T,s} - \mathcal{H}_{\text{worst}}}{\mathcal{H}_{\text{best}} - \mathcal{H}_{\text{worst}}} \leq \rho_{\mathcal{H}}\right) \end{cases}. \tag{7.58}$$

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

Repeat steps 2–10 until one of the parameters ζ_k or $\varsigma_{k,s}$ stabilises. After the search algorithm converges, the estimates $\hat{\Upsilon}^{\iota}_{\mathcal{R},k}$ and $\hat{\Upsilon}^{\iota}_{\mathcal{Q},k}$ are used to initialise the EKF.

7.3 AUTOCOVARIANCE LEAST SQUARES METHOD

In this section a method known as the ALS is investigated as an alternative for setting the initial parameters of the EKF. If complete system knowledge about the measurement function $\mathbf{h}$ and transition

function $\mathbf{f}$ were known, then the EKF only requires knowledge of the observation covariance matrix $\mathcal{R}$ and process covariance matrix $\mathcal{Q}$. Several different approaches have been formulated to solve the estimation of these covariance matrices [209–211]. All these methods assumed that the noise-shaping matrix in the transition equation is known. In the absence of information on the noise-shaping matrix the linear dynamic model is modelled as a Gaussian noise vector. The method that is investigated is the ALS method, which operates in the absence of information on the noise shaping matrix [212]. The ALS method assumes that:

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

The method estimates the observation covariance matrix $\mathcal{R}$ and process covariance matrix $\mathcal{Q}$ by minimising an objective function [212]. The objective function is a function of the measurement function $\mathbf{h}$, transition function $\mathbf{f}$ and the noise-shaping matrix (if present). The motivation for using this method is that it avoids a complicated non-linear estimation approach used by methods that employ a maximum likelihood estimation approach [213].

7.4 SUMMARY

In this chapter a novel BVEP criterion was proposed, which computes the process covariance matrix and observation covariance matrix using spatial and temporal information. This criterion could easily be extended, as shown in equation (7.52), to include spectral information if the spectral bands are correlated.

The derived matrices in the BVS were then used to initialise the EKF, which is used as a feature extraction method. The BVSA provides covariance matrices that could be used for a variety of different applications. A variety of different search algorithms can be used with the BVEP criterion, such as interior point, active set, simulated annealing, etc. These methods will be explored in chapter 8.