CHAPTER THREE

THE EXTENDED KALMAN FILTER

3.1 INTRODUCTION

As mentioned in chapter 1, one of the change detection methods proposed in this thesis is the Extended Kalman Filter (EKF) change detection method. As the name suggests, the EKF is a critical component of the change detection method and this chapter consequently aims to give some background on conceptual state-space filtering and in particular, the EKF. As will become more apparent in the chapters to follow, the underlying idea of the EKF change detection method lies in modeling an NDVI time-series as a triply modulated cosine function, and tracking the parameters of the model for each time-increment. The state-space filtering method is useful for this specific problem, as the parameters of the triply modulated cosine function could be characterized as a time-variant state-vector which relates to an observation model via the non-linear cosine function. As change detection is our primary objective, the near real-time nature of the state-space filtering method is also particularly useful, as the time from when the change occurred to when change is detected should ideally be minimal.

The EKF framework works on the basis that the posterior density of the state vector given the observed data is always assumed to be Gaussian, which makes for simple implementation and fast execution time [85]. Approximate grid based methods and Gaussian sum filters do not have the limitation of assuming Gaussian posteriors densities, but the computational complexity of these methods are very high which prevents their widespread use in practice [85]. The broader objective of this study is to implement the change detection methods operationally. This requires that the specific non-linear state space filter that is chosen be well understood and lend itself to be easily implemented. The EKF was consequently chosen as it is a well established method which is easily implementable [85]. The EKF was also compared to a sliding window FFT approach (section 4.5) and was found to be superior for the specific problem presented in this thesis (see chapter 6).

3.2 CONCEPTUAL STATE-SPACE FILTERING SOLUTION

In many applications, it is necessary to estimate the state of a dynamic system using only a time-series of noisy-measurements made on the system. In many cases, a discrete-time state-space approach is used to model the dynamic system. The underlying idea is that difference equations are used to model the evolution of the system over time and that measurements are available at discrete times. It is assumed that the state vector of the system contains a vector of state parameters that are able to accurately describe the behavior of a system. For example, in tracking systems, these parameters can be related to the kinematic characteristics of the target [85].

Using the state-space approach, at least two models are required to describe the noisy measurements obtained from the dynamic system. The first model (process model) describes the evolution of each state parameter from time-step $k-1$ to k. The second model (observation model) takes as input the state parameters at time-step k, obtained during the previous step, to produce an estimate of the output of the system at time-step k. For many applications, an estimate of the state parameters is required every time a measurement is received. The recursive nature of the state-space approach implies that a recursive filtering approach can be used where received data can be processed sequentially. Such a filter consists of two stages, namely predict and update. In the predict stage, the state PDF is predicted forward from one time-step to the next, which effectively broadens the state PDF. In the update stage, the latest available measurement is used to tighten the state PDF. [85]

The two models can be described formally as

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \tag{3.1}$$

and

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k. \tag{3.2}$$

Where $\mathbf{x}_k$ is the state vector at time-step k in the form

$$\mathbf{x}_k = [x_{k,1}\ x_{k,2}\ x_{k,3}\ ...\ x_{k,s}], \tag{3.3}$$

with s being the total number of state parameters. The relation between $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ is given by $\mathbf{f}$, a known but possibly non-linear function. The state vector $\mathbf{x}_k$ is related to the observation vector $\mathbf{y}_k$ via a known but possibly non-linear measurement function $\mathbf{h}$ It should be noted that $\mathbf{f}$ and $\mathbf{h}$ are allowed to be time-variant, but the time-invariant assumption was assumed in this thesis.. Both these models are possibly non-perfect, so the addition of process $\mathbf{w}_{k-1}$ and measurement $\mathbf{v}_k$ noise is included.

The task at hand is to produce an estimate of $\mathbf{x}_k$ given all available measurements $\mathbf{Y}_k = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, ..., \mathbf{y}_k\}$ up to time k. It is thus required to construct the posterior PDF $p(\mathbf{x}_k|\mathbf{Y}_k)$. The initial estimate of $p(\mathbf{x}_k)$ should be provided where-after $p(\mathbf{x}_k|\mathbf{Y}_k)$ is estimated recursively using the predict and update stages mentioned previously. The predicted PDF $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ is obtained by means of the Chapman-Kolmogoroff equation [85]:

$$p(\mathbf{x}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{Y}_{k-1})d\mathbf{x}_{k-1}. \tag{3.4}$$

Where $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is obtained using (3.1) and known statistics of $\mathbf{w}_{k-1}$. When the observation at time k ($\mathbf{y}_k$) becomes available, the state PDF is updated via Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})}, \tag{3.5}$$

where the normalization constant $p(\mathbf{y}_k|\mathbf{Y}_{k-1})$ can be written as

$$p(\mathbf{y}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_{k-1})d\mathbf{x}_k. \tag{3.6}$$

The likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$ can be obtained by using (3.2) and known statistics of $\mathbf{v}_k$. Knowledge of the posterior density $p(\mathbf{x}_k|\mathbf{Y}_k)$ enables one to not only compute the optimal state estimate with respect to any criterion, but also to determine the measure of accuracy of the state estimate [85]. For example, if $p(\mathbf{x}_k|\mathbf{Y}_k)$ is a multivariate Gaussian distribution, the covariance matrix can be used to determine the state estimate accuracy.

In the case that $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ in (3.1) and (3.2) are Gaussian distributed and both $\mathbf{f}(\mathbf{x}_{k-1})$ and $\mathbf{h}(\mathbf{x}_k)$ are linear functions, the functional recursion of (3.4) and (3.5) is the Kalman filter [86]. The Kalman filter will be discussed in further detail in the following section.

3.3 KALMAN FILTER

The Kalman filter is named after Rudolf E. Kalman and is a well-established method that was published in 1960 [86]. Since the algorithm was proposed, it has been widely used, especially in military and space applications.

As stated previously, the Kalman filter assumes that the process and observation noise is Gaussian distributed and that both the $\mathbf{f}$ and $\mathbf{h}$ functions in (3.1) and (3.2) are linear. It follows that (3.1) and (3.2) can be written as:

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \tag{3.7}$$

and

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \tag{3.8}$$

where $\mathbf{F}$ and $\mathbf{H}$ are known matrices defining the linear functions [85]. It is assumed that $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are zero-mean Gaussian distributed with covariances $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$ respectively. The distributions $p(\mathbf{x}_{k-1}|\mathbf{Y}_{k-1})$, $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$ and $p(\mathbf{x}_k|\mathbf{Y}_k)$ given in equations (3.4) and (3.5) can then be expressed as:

$$p(\mathbf{x}_{k-1}|\mathbf{Y}_{k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \widehat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \tag{3.9}$$

$$p(\mathbf{x}_k|\mathbf{Y}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \widehat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{3.10}$$

$$p(\mathbf{x}_k|\mathbf{Y}_k) = \mathcal{N}(\mathbf{x}_k; \widehat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), \tag{3.11}$$

where $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P})$ is a Gaussian distribution with argument $\mathbf{x}$, mean ($\mathbf{m}$) and covariance ($\mathbf{P}$) given as:

$$\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}) = \sqrt{|2\pi\mathbf{P}|}\; e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})}. \tag{3.12}$$

The mean and covariance parameters in (3.9)-(3.11) is given as [86]:

$$\widehat{\mathbf{x}}_{k|k-1} = \mathbf{F}\,\widehat{\mathbf{x}}_{k-1|k-1} \tag{3.13}$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \mathbf{F}\,\mathbf{P}_{k-1|k-1}\mathbf{F}^T \tag{3.14}$$

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\widehat{\mathbf{x}}_{k|k-1}) \tag{3.15}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T, \tag{3.16}$$

where

$$\mathbf{S}_k = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_{\mathbf{k}}, \tag{3.17}$$

is the innovation term and

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}, \tag{3.18}$$

is the Kalman gain. The Kalman filter thus effectively computes the mean and covariance of the Gaussian posterior $p(\mathbf{x}_k|\mathbf{Y}_k)$ and is optimal in the linear Gaussian environment. Unfortunately, many real world applications are non-linear and quite often non-stationary. In this case, approximation methods have to be used. A popular extension of the Kalman filter to the non-linear case is the Extended Kalman Filter, which will be discussed in more detail in the following section.

3.4 EXTENDED KALMAN FILTER

The extended Kalman filter (EKF) is the nonlinear version of the popular Kalman filter. Similar to the standard Kalman filter, for every increment of k (the discrete time) a state vector $\mathbf{x}_k$ is defined containing the parameters to be estimated. If one were, for example, to estimate the mean (μ) amplitude (α) and phase (ϕ) of a cosine function, the state vector could be in the form $\mathbf{x}_k = [\mu_k \ \alpha_k \ \phi_k]^T$. The state vector can be estimated over time k by recursive iteration based on the observation data $\mathbf{Y}_k$ up to time k. For the EKF, equations (3.7) and (3.8) can be reformulated as:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \tag{3.19}$$

and

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k. \tag{3.20}$$

In this formulation, either or both functions $\mathbf{f}$ and $\mathbf{h}$ are non-linear functions. The basic idea is that these non-linear functions can be sufficiently described using local linearization. The posterior PDF $p(\mathbf{x}_k|\mathbf{Y}_k)$ is approximated by a Gaussian distribution which implies that (3.9)-(3.11) are assumed to hold. Equations (3.13)-(3.18) can then be rewritten as:

$$\begin{aligned}
\widehat{\mathbf{x}}_{k|k-1} &= \mathbf{f}(\widehat{\mathbf{x}}_{k-1|k-1}) && (3.21)\\
\mathbf{P}_{k|k-1} &= \mathbf{Q}_{k-1} + \mathbf{F}\,\mathbf{P}_{k-1|k-1}\mathbf{F}^T && (3.22)\\
\widehat{\mathbf{x}}_{k|k} &= \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\widehat{\mathbf{x}}_{k|k-1})) && (3.23)\\
\mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T, && (3.24)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{S}_k &= \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R_k} && (3.25)\\
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}, && (3.26)
\end{aligned}$$

and $\mathbf{F}$ and $\mathbf{H}$ are the local linearization of the non-linear function $\mathbf{f}$ and $\mathbf{h}$ respectively. They are defined as Jacobians evaluated at $\widehat{\mathbf{x}}_{k-1|k-1}$ and $\widehat{\mathbf{x}}_{k|k-1}$ respectively [85]:

$$\widehat{\mathbf{F}} = \| \nabla_{\mathbf{x_{k-1}}} \mathbf{f}^T(\mathbf{x_{k-1}})^T \|_{\mathbf{x_{k-1}}=\widehat{\mathbf{x}}_{\mathbf{k-1|k-1}}} \quad (3.27)$$

$$\widehat{\mathbf{H}} = \| \nabla_{\mathbf{x_k}} \mathbf{h}^T(\mathbf{x_k})^T \|_{\mathbf{x_k}=\widehat{\mathbf{x}}_{\mathbf{k|k-1}}}, \quad (3.28)$$

where

$$\nabla_{\mathbf{x_k}} = \left[\frac{\partial}{\partial x_{k,1}} \cdots \frac{\partial}{\partial x_{k,s}} \right]. \quad (3.29)$$

The EKF is referred to as an analytical approximation because the Jacobians $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{H}}$ are computed analytically.

3.5 EXAMPLE OF AN EKF TRACKING APPLICATION

In this section, an example of tracking a cosine function with varying amplitude and phase using an EKF is shown. Assume that noisy observations are made of a process that is governed by the following function:

$$y_k = \alpha \cos(\omega k + \phi) + n_k, \quad (3.30)$$

where $\omega = 2\pi f$. It is assumed that the fundamental frequency f is known and that the task at hand is to estimate the value of α and ϕ for each time-step k given observations up to time-step k. It is proposed that this non-linear problem be solved using an EKF. The state vector can be defined as:

$$\mathbf{x_k} = [\alpha_k \ \phi_k]^T. \quad (3.31)$$

The process and observation models can be formulated as:

$$\mathbf{x_k} = \mathbf{x_{k-1}} + \mathbf{w_{k-1}}, \quad (3.32)$$

and

$$y_k = h(\mathbf{x}_k) + v_k, \quad (3.33)$$

where $h(\mathbf{x}_k) = x_{k,1} \cos(\omega k + x_{k,2})$. In the process model (3.32), it is assumed that the state vector remains constant from one time-step to the next with an additive process noise component. This implies that:

$$\mathbf{f}(\mathbf{x}) = \mathbf{F}\mathbf{x} = \mathbf{I}\mathbf{x}, \quad (3.34)$$

where $\mathbf{I}$ is a 2×2 Identity matrix. It is assumed that the noise component $\mathbf{w_{k-1}}$ is Gaussian distributed having zero mean and covariance $\mathbf{Q_{k-1}}$, i.e. $p(\mathbf{w_{k-1}}) = \mathcal{N}([0\ 0]^T, \mathbf{Q_{k-1}})$

The observation model (3.33) is based on (3.30) with the amplitude and phase parameter being replaced by the state parameters representing each of these variables and the noise component v_k also being assumed to be Gaussian distributed with zero mean and variance $\mathbf{R_k} = \sigma_v^2$, i.e. $p(\mathbf{v_k}) = \mathcal{N}(0, \sigma_v^2)$.

The state parameter $(\widehat{\mathbf{x}}_{k|k-1})$ prediction given in (3.21), can be re-written for this specific problem as:

$$\widehat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\widehat{\mathbf{x}}_{k-1|k-1}) \tag{3.35}$$
$$= \mathbf{I}\mathbf{x}_{k-1|k-1} \tag{3.36}$$
$$= \mathbf{x}_{k-1|k-1}. \tag{3.37}$$

The predicted covariance term ($\mathbf{P}_{k|k-1}$), as calculated in (3.22) can also be re-written as:

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \mathbf{F}\,\mathbf{P}_{k-1|k-1}\mathbf{F}^T \tag{3.38}$$
$$= \mathbf{Q}_{k-1} + \mathbf{I}\,\mathbf{P}_{k-1|k-1}\mathbf{I}^T \tag{3.39}$$
$$= \mathbf{Q}_{k-1} + \mathbf{P}_{k-1|k-1}. \tag{3.40}$$

In the state vector update phase of the EKF (3.23), the observation $(\mathbf{y}_k)$ is used to update the current state vector $\widehat{\mathbf{x}}_{k|k-1}$ and can be written as:

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\widehat{\mathbf{x}}_{k|k-1})) \tag{3.41}$$
$$= \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \widehat{\mathbf{H}}\widehat{\mathbf{x}}_{k|k-1}), \tag{3.42}$$

where $\widehat{\mathbf{H}}$ is the local linearization of function $\mathbf{h}(\mathbf{x})$ given as:

$$\widehat{\mathbf{H}} = \| \bigtriangledown_{\mathbf{x}_k} h(\mathbf{x}_k)\|_{\mathbf{x}_k=\widehat{\mathbf{x}}_{k|k-1}} \tag{3.43}$$
$$= \left[\frac{\partial}{\partial x_{k,1}}h(\mathbf{x}_k) \quad \frac{\partial}{\partial x_{k,2}}h(\mathbf{x}_k)\right]_{\mathbf{x}_k=\widehat{\mathbf{x}}_{k|k-1}}, \tag{3.44}$$

where

$$\frac{\partial}{\partial x_{k,1}}h(\mathbf{x}_k) = \frac{\partial}{\partial x_{k,1}}\ x_{k,1}\cos(\omega k + x_{k,2}) \tag{3.45}$$
$$= \cos(\omega k + x_{k,2}), \tag{3.46}$$

and

$$\frac{\partial}{\partial x_{k,2}} h(\mathbf{x}_k) = \frac{\partial}{\partial x_{k,2}} x_{k,1} \cos(\omega k + x_{k,2}) \tag{3.47}$$

$$= \frac{\partial}{\partial x_{k,2}} x_{k,1} \left[\cos(\omega k) \cos(x_{k,2}) - \sin(\omega k) \sin(x_{k,2}) \right] \tag{3.48}$$

$$= -x_{k,1} \left[\sin(\omega k) \cos(x_{k,2}) + \cos(\omega k) \sin(x_{k,2}) \right]. \tag{3.49}$$

The term $\mathbf{S}_k$ in the parameter covariance update equation (3.24) can also be rewritten for the present case as:

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \tag{3.50}$$

$$= \widehat{\mathbf{H}}_k \mathbf{P}_{k|k-1} \widehat{\mathbf{H}}_k^T + \sigma_v^2. \tag{3.51}$$

A time-series was generated with a Signal-to-noise Ratio (SNR) of 5 dB (Figure 3.1A). Figures 3.1B and 3.1C show the corresponding amplitude and phase being tracked using the EKF framework. In Figure 3.1A, the noisy observations as well as the actual signal is shown together with the filtered EKF estimate. It can be seen that the filter requires an initial number of observations before the EKF state parameters start to stabilize. The stabilized state vector corresponds to the accurate tracking of the underlying signal by the EKF, i.e. as soon as the state parameters start to stabilize, the value of the underlying signal is very accurately tracked by the EKF.

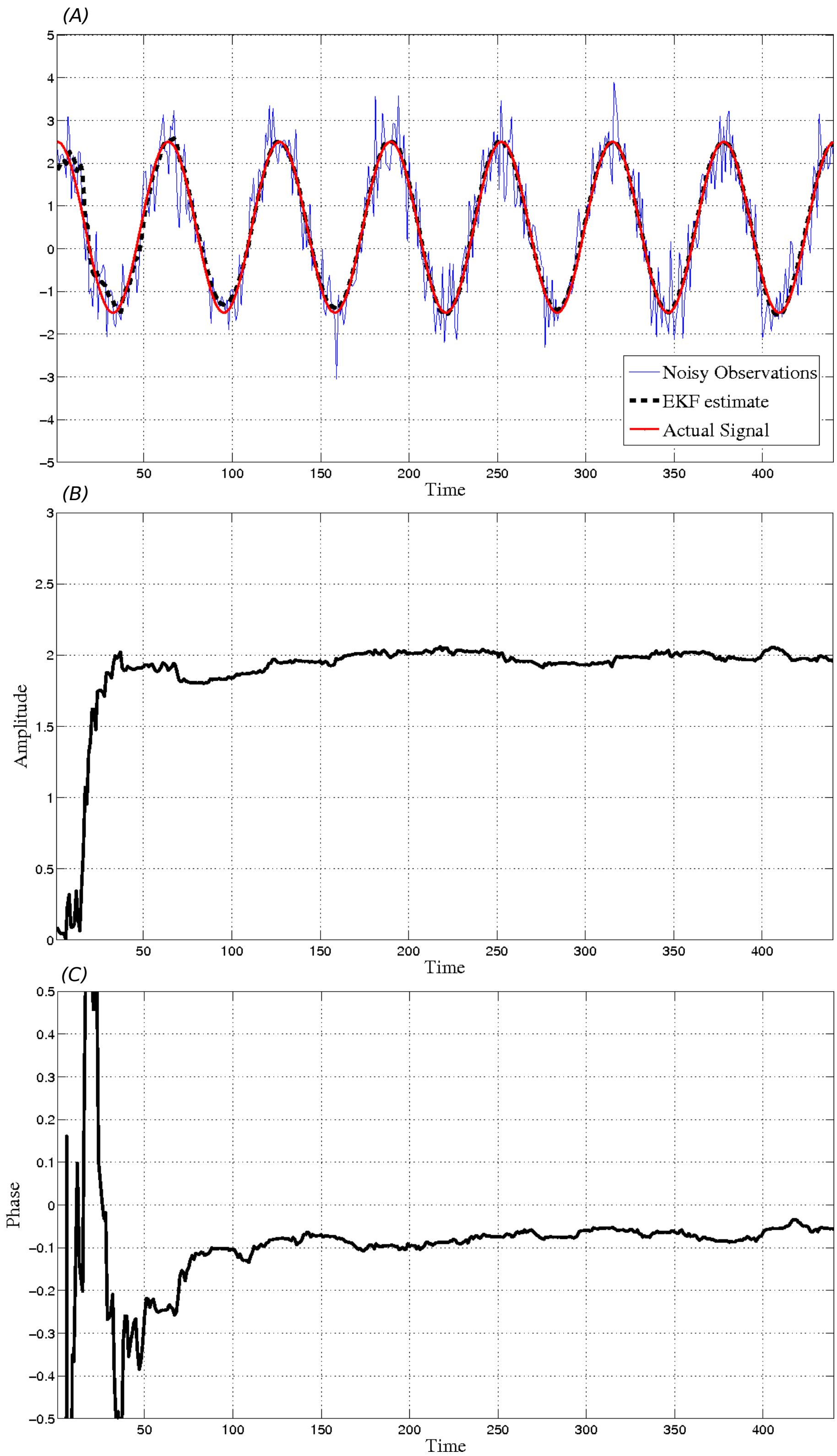

FIGURE 3.1: Figure *A* shows the noisy observation, actual signal and EKF output. Figures *B* and *C* show the estimated amplitude and phase state parameter for each time-step.