

A Model for Non-Stationary Time Series and Its Applications in Filtering and Anomaly Detection

Shixiong Wang^{ID}, Graduate Student Member, IEEE, Chongshou Li^{ID}, and Andrew Lim^{ID}

Abstract—Time series measurements from sensing units (e.g., UWB ranging circuits) always suffer from uncertainties like noises, outliers, dropouts, and/or nonspecific anomalies. In order to extract the true information with high precision from the original corrupted measurements, the signal-model-based signal pre-processing units embedded in sensing circuits are usually employed. However, for a general signal to observe, its signal model cannot be obtained so that the signal-model-based signal processing methods are not applicable. In this article, the time-variant local autocorrelated polynomial (TVLAP) model in the state space is proposed to model the dynamics of a non-stationary stochastic process (i.e., a signal or a time series), through which the model-based signal processing methods could be utilized to denoise, to correct the outliers/dropouts, and/or to identify anomalies contained in the measurements. Besides, the presented method can also be used in change point detection for a time series.

Index Terms—Anomaly detection, change point detection, denoising, outlier correction, signal modeling, time-domain measurement.

I. INTRODUCTION

TIME series measurements from sensing systems, such as ultra-wideband (UWB) ranging circuit [1], seismic monitoring system [2], and power system [3], always suffer from uncertainties like noises [1], [2], [4], outliers [5], [6], dropouts [7], and specific anomalies [8] (i.e., we know what features it has or where it is from) or nonspecific anomalies [9], [10] (i.e., we do not know what it is and where it is from). Over the years, several powerful filtering [11]–[13], outlier/dropout correction [7], [14], [15], and anomaly detection [6], [8], [16] methods have been proposed to handle these problems. However, virtually all of those efficient methods require the use of the signal model, i.e., the signal process dynamics and the sensor measurement dynamics. For the model-free methods: (a) like IIR/FIR filters (e.g., exponential smoothing, moving average) for filtering [17], they innately have time

delay which denies the direct use in some strict real-time applications; (b) like the one-sided median method for outlier removal [18], it can neither handle the case where the outliers are dense nor simultaneously denoise in an online manner. Also, in model-free scenarios, it is challenging to describe and/or decide whether anomalies are contained in the measurements unless the offline methods like those in frequency domain [19], [20] are applied (note that block data, rather than streaming/sequential data, are required to do transformation and analysis in the frequency domain). This article is therefore concerned with modeling a non-stationary stochastic process (i.e., a random signal or a time series) in the state space so that we can use model-based signal processing methods like the Kalman filtering to denoise, to correct the outliers/dropouts, and/or to identify some types of anomalies contained in the corrupted measurements from inexact sensing units.

The contributions of this article are both theoretical and applicational.

- 1) *State-Space Modeling for a Non-Stationary Signal*: Signal model plays a very important role in signal processing. However, for a general measurement signal from a sensor, it is hard or even impossible to obtain the signal model, because sometimes we have no knowledge of the system (e.g., a sensor) generating the focused time series. Fortunately, our Time-Variant Local Autocorrelated Polynomial (TVLAP) model could help detour this issue.
- 2) *Denoising*: It is the use of derivatives of the mean function (DMF) of the signal that makes more satisfactory denoising/filtering performances for noised measurements from a sensor.
- 3) *Outlier/Dropout Correction*: The proposed method is powerful in identifying and correcting the outliers and dropouts in measurements. We treat dropouts as special outliers that are zero-valued.
- 4) *Fault Dignosis (i.e., Anomaly Detection)*: Sensors sometimes unavoidably suffer from anomalies (faults). We aim to use the DMF of the measurement time series to describe and identify those anomalies, and decide whether the sensing system is currently reliable or not.
- 5) *Change Point Detection*: The proposed method can be used to detect change points (e.g., turning points, extrema, etc.) of a time series.
- 6) *Theoretical Sufficiency*: The complete reliability guarantee of the proposed method is derived.

Manuscript received November 16, 2020; revised February 3, 2021; accepted February 7, 2021. Date of publication February 12, 2021; date of current version March 3, 2021. This work was supported by the National Research Foundation of Singapore under Grant NRF-RSS2016-004. The Associate Editor coordinating the review process was Gaigai Cai. (Corresponding author: Shixiong Wang.)

Shixiong Wang and Andrew Lim are with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576 (e-mail: s.wang@u.nus.edu; isealim@nus.edu.sg).

Chongshou Li was with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576. He is now with the School of Information Science and Technology, Institute of Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China (e-mail: cslichina2020@outlook.com).

Digital Object Identifier 10.1109/TIM.2021.3059321

II. PRELIMINARY ON STOCHASTIC PROCESS

Theorem 1 (Wold's Decomposition Theorem [21]): Any wide-sense stationary (WSS) discrete stochastic process $x(n)$ could be decomposed into two sub-processes: (a) regular process and (b) predictable process. Namely $x(n) = x_r(n) + x_p(n)$, where $x_r(n)$ is a regular process and $x_p(n)$ is a predictable process. Furthermore, the two processes are orthogonal: $\forall \tau, E\{x_r(n + \tau)x_p(n)\} = 0$.

The detailed concepts of regular process and the predictable process could be found in [21]. Intuitively, a regular WSS stochastic process is mathematically as $x = \text{ARMA}(p, q)$ (an autoregressive moving average process with autoregressive order of p and moving average order of q). Thus, this theorem reveals the validity of the ARMA model for a regular WSS signal.

III. PROBLEM FORMULATION

A. Notations

- 1) Let $\mathbf{v} = a : l : b$ define a vector \mathbf{v} being with the lower bound a , upper bound b and step length l . For example, $\mathbf{v} = 0 : 0.1 : 0.5$ means $a = 0$, $b = 0.5$, and $l = 0.1$. Thus $\mathbf{v} = [0, 0.1, 0.2, 0.3, 0.4, 0.5]^T$.
- 2) Let the function $\text{length}(\mathbf{x})$ return the length of the vector \mathbf{x} . For example, if $\mathbf{x} = [1, 2]^T$, we have $\text{length}(\mathbf{x}) = 2$.
- 3) Let t denote the continuous time variable, and n its corresponding discrete time variable. For example, if $t = 0 : 0.5 : 100$ (the time span is 100 s, and the sampling time is $T = 0.5$ s), we will have $n = t/T = 0 : 1 : [\text{length}(t) - 1] = 0 : 1 : 200$.
- 4) Let the function $\text{mean}(\mathbf{x})$ return the mean of a random variable \mathbf{x} , and $\text{var}(\mathbf{x})$ the variance of it. If \mathbf{x} is a stochastic process, then $\text{mean}(\mathbf{x})$ denotes the mean function and $\text{var}(\mathbf{x})$ the variance function.
- 5) Let \mathbf{G}' denote the transpose of the matrix \mathbf{G} .
- 6) Let the operator $\text{ARMA}(p, q|\boldsymbol{\varphi}, \boldsymbol{\theta})$ denote an ARMA process with autoregressive order of p and moving average order of q . Besides, the coefficient vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ are for autoregressive part and moving average part, respectively. $\text{ARMA}(p, q|\boldsymbol{\varphi}, \boldsymbol{\theta})$ is shorted as $\text{ARMA}(p, q)$.

B. A New Model for a Non-Stationary Stochastic Process

In this article, we consider a general model describing a non-stationary discrete stochastic process (signal) as the following form:

$$x(n) = f(n) + x_s(n) \quad (1)$$

where $x_s(n) := \text{ARMA}(p, q)$ is a regular WSS stochastic process [i.e., $x_s(n)$ is the measurement noise of a sensing unit]; $f(n)$ is a deterministic function denoting the mean function of the time series $x(n)$ [i.e., $f(n)$ is the true information contained in $x(n)$]. Since the expectation of the term $x_s(n)$ is zero, we have $\text{mean}(x) = f$, and $\text{var}(x) = \text{var}(x_s)$.

Note that by using the model (1) for a measurement time series, we are excluding the non-WSS measurement noises, such as the flicker noise sequence (a.k.a. $1/f$ noise) and

the random walk noise sequence (a.k.a. $1/f^2$ noise), etc., that mainly consist of low-frequency components [22], [23]. Namely, the basic assumption of the model (1) is that the measurement noise sequence is regularly WSS. Any low-frequency (e.g., non-zero-mean or zero-mean-but-unbounded-variance) component of a measurement time series would be regarded as a change of the measured quantity. See Fig. 9 for example.

C. Problem Statement

Mathematically, we aim to recover $f(n)$ and its high-order derivatives from noised $x(n)$, in an online manner. This is because the following conditions hold.

- 1) It is the incorporation of the DMF that allows us to describe the dynamics of a signal with higher accuracy, which is useful in denoising and outlier/dropout correction.
- 2) We can identify such the DMF as features to differentiate the faults/anomalies contained in the measurements.

Note that the difference-based method (e.g., $[x(n) - x(n-1)]/T$) is not reliable to estimate the derivatives of a noised time series because the noises would be amplified by the difference operator.

IV. MAIN RESULTS

A. TVLAP Model

In this section, we will introduce the TVLAP model with Kalman Filter to handle the online mean function estimation problem. As a demonstration, we in this section only take the special case of (1) as $x(n) = f(n) + G_w(n)$, where $G_w(n)$ denotes a white Gaussian series. The general case will be discussed later.

We can use a polynomial to regress the mean function in an online manner. We require that the regressed polynomial and the mean function of the raw time series are close enough. The theoretical validity and sufficiency of polynomial regression is from the prestigious Weierstrass approximation theorem [24]. However, the dilemma is the concern of real-time property, meaning we expect the algorithm to be able to work online with sequential data. Recall the Taylor's polynomial expansion

$$f(t) = f(t_0) + \frac{f^{(1)}(t_0)}{1!}(t - t_0) + \dots + \frac{f^{(k)}(t_0)}{k!}(t - t_0)^k + \dots \quad (2)$$

where $f^{(k)}(t_0)$ denotes the k^{th} -order derivative of $f(t)$ at t_0 . However, a function could be expanded as Taylor's series if and only if it is infinitely smooth, meaning infinitely differentiable. Thus we cannot directly apply the Taylor's series expansion over a general time series whose trend function $f(t)$ may be discontinuous in derivatives. To overcome this, we introduce an intermediate (temporary) function $p(t)$ as the Weierstrass approximation of $f(t)$. It means $p(t)$ is a polynomial with proper orders. Thus, we have $\forall \varepsilon > 0, \exists \bar{K} > 0$, such that

$$\sup_t |f(t) - p_{\bar{K}}(t)| < \varepsilon \quad (3)$$

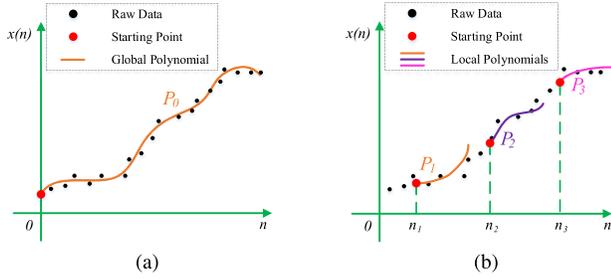


Fig. 1. Global and local polynomial. (a) Global polynomial. (b) Local polynomial.

in a closed and bounded interval, where $p_{\bar{K}}(t)$ denotes a polynomial with an order of \bar{K} . For simplicity, we ignore \bar{K} in notation. We have $p(t) = p_0 + p_1t + p_2t^2 + \dots + p_kt^k + \dots$.

Thus, when we have a time series $x(n)$, we could alternatively choose the polynomial in Taylor's form to regress $p(t)$ instead of $f(t)$ because only $p(t)$ is guaranteed to be infinitely differentiable. This will not lead to a disaster, according to (3). Suppose we have interests in the properties at the discrete time index n (i.e., $t_0 = nT$), (2) could then be rewritten as

$$p(t) = p(n) + \frac{p^{(1)}(n)}{1!}(t-n) + \dots + \frac{p^{(k)}(n)}{k!}(t-n)^k + \dots \quad (4)$$

As a result, the traditional polynomial regression $p(t) = p_0 + p_1t + p_2t^2 + \dots + p_kt^k + \dots$ could be regarded as a special case of (2) when we investigate the problem from the starting point of the time, namely, $t_0 = 0$. In other words, the polynomial (4) is a local polynomial, while $p(t) = p_0 + p_1t + p_2t^2 + \dots + p_kt^k + \dots$ is a global polynomial. For intuition see Fig. 1.

If we only pay attention to the case of $t = (n+1)T$ and truncate the polynomial on the order of K , we have (4) as

$$p(n+1) = \sum_{k=0}^K \frac{p^{(k)}(n)}{k!} T^k = \sum_{k=0}^K \frac{T^k}{k!} p^{(k)}(n) \quad (5)$$

where T denotes the time slot between the discrete time indices $n+1$ and n (i.e., the sampling time).

Interestingly, (5) holds the following powerful characteristics.

- 1) It is actually the state equation of a general time series. Note that the nature of the state equation is the recursive relationship of a time-related function from the former discrete time index n to the latter $n+1$.
- 2) It conveys the high-order derivatives up to the order of K^{th} of the function $p(t)$, which is attractive in signal processing.

In the state space, if we define our state vector as

$$\mathbf{X}(n) := \begin{bmatrix} X_0(n) \\ X_1(n) \\ X_2(n) \\ \dots \\ X_K(n) \end{bmatrix} := \begin{bmatrix} p^{(0)}(n) \\ p^{(1)}(n) \\ p^{(2)}(n) \\ \dots \\ p^{(K)}(n) \end{bmatrix} \quad (6)$$

meaning the first entry is the real-time value of $p(n)$ and the rest entries are the real-time values of the high-order derivatives of $p(n)$.

Consequently, we have the state space representation of (5) as

$$\mathbf{X}(n+1) = \begin{bmatrix} 1 & T & \frac{T^2}{2} & \dots & \frac{T^K}{K!} \\ 0 & 1 & T & \dots & \frac{T^{K-1}}{(K-1)!} \\ 0 & 0 & 1 & \dots & \frac{T^{K-2}}{(K-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \mathbf{X}(n). \quad (7)$$

Equation (7) implies that when we model the dynamics of $f(n)$, we actually admit the K^{th} -order derivative to remain constant over time.

As time series measurements from a sensor, only the sequential data $x(n)$ is obtainable, observable. Therefore, we in our state space adaptation should define the measure vector as

$$\mathbf{Y}(n) = x(n) = f(n) + x_s(n) = p(n) + x_s(n) \\ := f(n) + G_w(n). \quad (8)$$

By doing so, we have the measurement equation (also known as observation equation or output equation) as

$$\mathbf{Y}(n) := [1 \ 0 \ 0 \ \dots \ 0] \mathbf{X}(n) + \mathbf{V}(n) \quad (9)$$

where $\mathbf{V}(n)$ is used to model the measurement noise $G_w(n)$ [in general, the $x_s(n)$]. Note that, $\mathbf{Y}(n)$ and $\mathbf{V}(n)$ are all 1-D scalars. We write them in bold-face just to follow the notation convention for a state-space model.

Besides, we define

$$\Phi := \begin{bmatrix} 1 & T & \frac{T^2}{2} & \dots & \frac{T^K}{K!} \\ 0 & 1 & T & \dots & \frac{T^{K-1}}{(K-1)!} \\ 0 & 0 & 1 & \dots & \frac{T^{K-2}}{(K-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (10)$$

as our System Matrix, and

$$\mathbf{H} := [1 \ 0 \ 0 \ \dots \ 0] \quad (11)$$

as our Measurement Matrix. Therefore, Φ and \mathbf{H} are constant if given the order K . Suppose the state noise vector is $\mathbf{W}(n)$ with covariance $\mathbf{Q}(n)$ and its noise-driven matrix is \mathbf{G} ; the measurement noise state vector is $\mathbf{V}(n)$ with covariance $\mathbf{R}(n)$. We then have a state-space model for the time series $x(n)$ as

$$\begin{cases} \mathbf{X}(n+1) = \Phi \mathbf{X}(n) + \mathbf{G} \mathbf{W}(n) \\ \mathbf{Y}(n) = \mathbf{H} \mathbf{X}(n) + \mathbf{V}(n) \end{cases} \quad (12)$$

where $\mathbf{W}(n)$ denotes the modeling error. Since $\mathbf{V}(n)$ is a scalar, $\mathbf{R}(n)$ is also a scalar. Besides, due to the measurement noise process $x_s(n)$ is WSS, $\mathbf{R}(n)$ is constant over time. Let $\mathbf{R} = \mathbf{R} = \mathbf{R}(n)$. The first equality means \mathbf{R} is a scalar and the second means $\mathbf{R}(n)$ is constant.

Note that the mathematical form of \mathbf{G} is not unique, meaning we can define it as any proper one. Some simple

examples are: 1) $\mathbf{G} = [\frac{T^K}{K!}, \dots, T, 1]'$ so that $\mathbf{W}(n)$ should be a 1-D scalar denoting the disturbance exerted to $X_K(n)$; 2) $\mathbf{G} = \text{diag}\{\frac{T^K}{K!}, \dots, T, 1\}$ so that $\mathbf{W}(n)$ should be a $(K + 1)$ -dimensional vector denoting the disturbance exerted to $\mathbf{X}(n)$; 3) \mathbf{G} as an identity matrix so that $\mathbf{W}(n)$ should be a $(K + 1)$ -dimensional vector denoting the disturbance exerted to $\mathbf{X}(n)$. The difference between 2) and 3) is reflected in their corresponding $\mathbf{Q}(n)$.

B. Estimate the R and $\mathbf{Q}(n)$

Actually, R is easy to estimate from the historical observations (measurements) of $x(\bar{n})$. Here \bar{n} is used to differentiate from n , meaning $x(\bar{n})$ could be any segment of $x(n)$ in the past, just as ground truth to estimate R . Suppose we use the traditional global polynomial $\bar{p}(t)$ to fit $x(\bar{n})$ (e.g., using a three-layer neural network [25]), we should have the fitting residual δ as $\delta(\bar{n}) := x(\bar{n}) - \bar{p}(\bar{n})$. According to our model assumption, $\delta(\bar{n})$ should be a WSS stochastic process, meaning the selected order of $\bar{p}(t)$ is proper if and only if $\delta(\bar{n})$ is WSS. Thus, we have $R := \text{var}(\delta)$.

As for the real-time estimation of $\mathbf{Q}(n)$ when given (12) and R , readers are invited to refer to [26], [27].

Remark 1: In practice, at many times there is no need to pursue the exactly true value of $\mathbf{Q}(n)$. Engineers could try different $\mathbf{Q}(n)$ to obtain different estimation performances. Note that the value of $\mathbf{Q}(n)$ actually adjusts our trust level toward the system model that we use [28]. The larger the value of $\mathbf{Q}(n)$, the less trust we have toward the dynamics model (i.e., the more trust toward the measurements). Therefore, for convenience, we suggest using $\mathbf{G} = [\frac{T^K}{K!}, \dots, T, 1]'$ so that $\mathbf{W}(n)$ would be a (constant) scalar which is easier to tune.

C. Applications of TVLAP-KF

Now, it is sufficient to use the Kalman filter to handle the linear system (12), during which we can estimate the real-time value \hat{X}_0 of $p(n)$, and real-time values of k^{th} -order derivative \hat{X}_k of $p(n)$, where $p(n)$ is the true signal (the mean function) of the focused time series $x(n)$. Note that the Kalman filter admits the regularized residual norm minimization so that the high-order derivatives of $p(n)$ exist as regularization terms [29], which helps avoid over-fitting.

We, in this article, term the presented method as TVLAP Model with Kalman Filter, shorted as TVLAP-KF. Time-Variant means the coefficients of the used polynomial model (5), namely $p^{(k)}(n)/k!$ and $X_k(n)$, change over time. The meaning of the word Local has been explained earlier in Fig. 1. Autocorrelated means the coefficients of the used polynomial are not independent, are instead highly related, because we have

$$p^{(k+1)}(n) = \frac{d[p^{(k)}(n)]}{dt}. \quad (13)$$

In the following, we show some possible applications of the proposed TVLAP-KF. The corresponding experiments will be conducted in experiment Section V. First, by using TVLAP-KF, the measurements from sensors could be denoised/filtered, i.e., the noise part $x_s(n)$ could be attenuated. Second, TVLAP-KF can predict future measurement based on

the historical measurements so that if the predicted measurement is far away from the collected one, we can identify the collected value as an outlier/dropout and replace it with the predicted value. This process corrects the outliers/dropouts contained in the measurements. Third, by using DMF (viz., DMF of a measurement signal) as features, we can easily describe and detect some types of anomalies contained in the measurements so that we can tell apart the associated malfunctioning sensor. Fourth, as a supplementary application scenario, we show the advantages of the proposed TVLAP-KF in Change Point Detection for a time series. See Algorithm 1.

Algorithm 1 Change Point Detection Method Based on TVLAP-KF

Definition: \mathbf{P} as state estimate covariance in Kalman filter; \mathbf{I} as identity matrix with proper dimension; ∞ as a big number; ϵ as a small number; $\text{abs}(x)$ as the absolute function which return the absolute value of a real number; \emptyset as an empty set

Reservation: Set \mathbb{E}_m to record minima, and Set \mathbb{E}^m to record maxima

Initialize: $\infty \leftarrow 10^5$, $\epsilon \leftarrow 10^{-6}$, $\mathbf{X} \leftarrow \mathbf{0}$, $\mathbf{P} \leftarrow \infty \times \mathbf{I}$, \mathbf{Q} , \mathbf{R} , $\mathbb{E}_m \leftarrow \emptyset$, $\mathbb{E}^m \leftarrow \emptyset$

Input: $x(n)$, $n = 0, 1, 2, 3, \dots$

```

1: while true do
2:    $n \leftarrow n + 1$ 
3:   // Estimate the DMF
4:    $\hat{\mathbf{X}}(n) = \text{Kalman\_Filter}[x(n)]$  // See [28] (Chapter 5.1)
5:   // Obtain the Estimated Mean Function
6:    $\hat{f}(n) \leftarrow \hat{X}_0(n)$ 
7:   // Turning Point Detection
8:   if  $\text{abs}(\hat{X}_1(n-1)) < \epsilon$  and  $\hat{X}_1(n) > 0$  then
9:     The time series starts to increase
10:  else if  $\text{abs}(\hat{X}_1(n-1)) < \epsilon$  and  $\hat{X}_1(n) < 0$  then
11:    The time series starts to decrease
12:  end if
13:  // Extrema Detection
14:  if  $\text{abs}(\hat{X}_1(n)) < \epsilon$  and  $\hat{X}_2(n) > 0$  then
15:     $\mathbb{E}_m \leftarrow \{n\} \cup \mathbb{E}_m$  // Minimum reached
16:  else if  $\text{abs}(\hat{X}_1(n)) < \epsilon$  and  $\hat{X}_2(n) < 0$  then
17:     $\mathbb{E}^m \leftarrow \{n\} \cup \mathbb{E}^m$  // Maximum reached
18:  end if
19: end while
Output: estimated mean  $\hat{f}(n)$ ; minima set  $\mathbb{E}_m$ ; maxima set  $\mathbb{E}^m$ 

```

D. Reliability Guarantee of TVLAP-KF

In this section, we analyze the performances of the proposed TVLAP-KF. That is, we need to investigate whether the TVLAP-KF could recursively approximate $p(t)$ and its derivatives defined in (4) with satisfying accuracy.

Definition 1: The linear time-invariant system defined as (12) is uniformly completely observable if the matrix \mathbf{O} defined by the matrices pair $[\Phi, \mathbf{H}]$:

$$\mathbf{O} = [\mathbf{H}', \Phi' \mathbf{H}', \dots, (\Phi^K)' \mathbf{H}'] \quad (14)$$

is of full rank.

Definition 2: The linear time-invariant system defined as (12) is uniformly completely controllable if the matrix C defined by the matrices pair $[\Phi, G]$:

$$C = [G, \Phi G, \dots, \Phi^K G] \quad (15)$$

is of full rank.

Lemma 1: $\Phi^K(T) = \Phi(KT)$.

Proof: Actually, there exists a matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (16)$$

such that $\Phi(T) = e^{AT}$. Thus, $\Phi^K(T) = e^{KAT} = \Phi(KT)$. That is,

$$\Phi^K(T) = \begin{bmatrix} 1 & KT & \frac{(KT)^2}{2} & \dots & \frac{(KT)^K}{K!} \\ 0 & 1 & KT & \dots & \frac{(KT)^{K-1}}{(K-1)!} \\ 0 & 0 & 1 & \dots & \frac{(KT)^{K-2}}{(K-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (17)$$

Lemma 2: The Vandermonde matrix defined as

$$V = \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{n-1} \\ 1 & a_3 & a_3^2 & \dots & a_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_m & a_m^2 & \dots & a_m^{n-1} \end{bmatrix} \quad (18)$$

is of full rank if $\forall i \neq j$, we have $a_j \neq a_i$.

Proof: Since $\det(V) = \prod_{1 \leq i < j \leq n} (a_j - a_i)$ (see [30], Chapter 6.1), the lemma stands. \square

Lemma 3: The linear time-invariant system defined in (12) is uniformly completely observable, if K is finite.

Proof:

$$O = \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^K \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & T & \frac{(T)^2}{2} & \dots & \frac{(T)^K}{K!} \\ 1 & 2T & \frac{(2T)^2}{2} & \dots & \frac{(2T)^K}{K!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & KT & \frac{(KT)^2}{2} & \dots & \frac{(KT)^K}{K!} \end{bmatrix}. \quad (19)$$

Note that, if K tends to infinity, many entries of O would tend to zeroes. Thus, if K is finite, by Lemma 2, we have

$$\begin{aligned} \text{rank}(O) &= \text{rank} \left(\begin{bmatrix} 0^0 & 0^1 & 0^2 & \dots & 0^K \\ 1^0 & 1^1 & 1^2 & \dots & 1^K \\ 2^0 & 2^1 & 2^2 & \dots & 2^K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K^0 & K^1 & K^2 & \dots & K^K \end{bmatrix} \right) \\ &= K + 1 \end{aligned} \quad (20)$$

meaning O is of full rank. According to the definition of observability, this lemma stands. \square

Lemma 4: The linear time-invariant system defined in (12) is uniformly completely controllable, if K is finite and G is given as one of the following cases.

1) $G_1 = [\frac{T^K}{K!}, \dots, T, 1]^T$.

2) $G_2 = \text{diag}\{\frac{T^K}{K!}, \dots, T, 1\}$.

\square 3) G_3 as an identity matrix I with proper dimensions.

Proof: Let $C_{\Phi, G}$ denotes the controllability matrix defined by the pair $[\Phi, G]$. Since $C = [G, \Phi G, \dots, \Phi^K G]$, it is easy to check that $\text{rank}(C_{\Phi, G_3}) = K + 1$ (full rank). Due to $\text{rank}(C_{\Phi, G_2}) = \text{rank}(C_{\Phi, G_3})$, $\text{rank}(C_{\Phi, G_2}) = K + 1$ also holds. As for C_{Φ, G_1} , we have (21), as shown at the bottom of the page.

By the binomial theorem, the entry of C_{Φ, G_1} at $(I+1, J+1)$ is therefore

$$\begin{aligned} C_{\Phi, G_1}(I+1, J+1) &= \sum_{i=0}^{K-I} \frac{(JT)^i T^{K-I-i}}{i!(K-I-i)!} \\ &= \frac{1}{(K-I)!} (JT+T)^{K-I} \end{aligned} \quad (22)$$

$$\begin{aligned} C_{\Phi, G_1} &= [G_1, \Phi G_1, \dots, \Phi^K G_1] \\ &= \begin{bmatrix} \frac{T^K}{K!} & \sum_{i=0}^K \frac{(1T)^i (T)^{K-i}}{i! (K-i)!} & \dots & \sum_{i=0}^K \frac{(KT)^i (T)^{K-i}}{i! (K-i)!} \\ \frac{T^{K-1}}{K-1!} & \sum_{i=0}^{K-1} \frac{(1T)^i (T)^{K-1-i}}{i! (K-1-i)!} & \dots & \sum_{i=0}^{K-1} \frac{(KT)^i (T)^{K-1-i}}{i! (K-1-i)!} \\ \vdots & \vdots & \ddots & \vdots \\ T & \sum_{i=0}^1 \frac{(1T)^i (T)^{1-i}}{i! (1-i)!} & \dots & \sum_{i=0}^1 \frac{(KT)^i (T)^{1-i}}{i! (1-i)!} \\ 1 & 1 & \dots & 1 \end{bmatrix}. \end{aligned} \quad (21)$$

where $I, J = 0, 1, 2, \dots, K$, giving \mathbf{C}_{Φ, G_1} further as

$$\mathbf{C}_{\Phi, G_1} = \begin{bmatrix} \frac{T^K}{K!} & \frac{(2T)^K}{K!} & \frac{(3T)^K}{K!} & \dots & \frac{[(K+1)T]^K}{K!} \\ \frac{T^{K-1}}{K-1!} & \frac{(2T)^{K-1}}{K-1!} & \frac{(3T)^{K-1}}{K-1!} & \dots & \frac{[(K+1)T]^{K-1}}{K-1!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T & 2T & 3T & \dots & (K+1)T \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}. \quad (23)$$

Note that, if K tends to infinity, many entries of \mathbf{C}_{Φ, G_1} would tend to zeroes. Thus, if K is finite, by Lemma 2, we have

$$\begin{aligned} \text{rank}(\mathbf{C}_{\Phi, G_1}) &= \text{rank} \left(\begin{bmatrix} 1^K & 2^K & \dots & (K+1)^K \\ 1^{K-1} & 2^{K-1} & \dots & (K+1)^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1^1 & 2^1 & \dots & (K+1)^1 \\ 1^0 & 2^0 & \dots & (K+1)^0 \end{bmatrix} \right) \\ &= K + 1. \end{aligned} \quad (24)$$

Since \mathbf{C}_{Φ, G_1} defined in (21) is rank-sufficiency, this lemma stands. \square

Theorem 2: For any given norm-finite $\hat{\mathbf{X}}_{0|0}$, if $\Phi, \mathbf{G}, \mathbf{Q}$ and \mathbf{R} are bounded, $[\Phi, \mathbf{H}]$ is uniformly completely observable, and $[\Phi, \mathbf{G}]$ is uniformly completely controllable, then

$$\hat{\mathbf{X}}_{n|n} \rightarrow_d \mathbf{X}_n, \quad \text{as } n \rightarrow \infty \quad (25)$$

meaning

$$\hat{p}^{(k)}(n) \rightarrow_d p^{(k)}(n), \quad \text{as } n \rightarrow \infty \quad \forall k = 0, 1, 2, \dots, K. \quad (26)$$

Besides, the convergence rate is exponential (i.e., very fast).

Remark 2: Note that in Theorem 2, the notation \rightarrow_d means convergence in distribution, for example, $\hat{\mathbf{X}}_{n|n} \rightarrow_d \mathbf{X}_n$ admits $[\hat{\mathbf{X}}_{n|n} - \mathbf{X}_n] \rightarrow_d \mathbf{N}(\mathbf{0}, \mathbf{P}_{n|n})$ where $\hat{\mathbf{X}}_{n|n}$ means the *a posteriori* estimation of \mathbf{X}_n given by the Kalman filter; $\mathbf{N}(\cdot, \cdot)$ means a multivariate normal distribution; and $\mathbf{P}_{n|n}$ is the *a posteriori* estimation error covariance returned by the Kalman filter.

Proof: According to [31] (see Theorem 4) and [32] (see Chapter 4.4), with support of our Lemma 3 and Lemma 4, this theorem holds. Note that uniformly complete controllability (respective observability) implies the uniformly complete stabilizability (respective detectability). Note also that $\text{rank}(\mathbf{O}_{\Phi, \mathbf{H}}) = \text{rank}(\mathbf{O}_{\Phi, \mathbf{H}\mathbf{R}^{1/2}})$, and $\text{rank}(\mathbf{C}_{\Phi, \mathbf{G}}) = \text{rank}(\mathbf{C}_{\Phi, \mathbf{G}\mathbf{Q}^{1/2}})$, where $\mathbf{R}^{1/2}(\mathbf{R}^{1/2})' = \mathbf{R}$ and $\mathbf{Q}^{1/2}(\mathbf{Q}^{1/2})' = \mathbf{Q}$. Since \mathbf{R} and \mathbf{Q} are positive definite, the decomposition can be made. $\mathbf{O}_{\Phi, \mathbf{H}}$ denotes the observability matrix defined by the pair $[\Phi, \mathbf{H}]$. The notation conventions keep the same to $\mathbf{C}_{\Phi, \mathbf{G}}$, $\mathbf{O}_{\Phi, \mathbf{H}\mathbf{R}^{1/2}}$, and $\mathbf{C}_{\Phi, \mathbf{G}\mathbf{Q}^{1/2}}$. \square

As we can see, the observability and controllability of the system (12) are the sufficient conditions for the convergence results in Theorem 2. Intuitively, the observability guarantees that the system state $\mathbf{X}(n)$ [i.e., $p^{(k)}(n)$] is able to be

observed/estimated from the system output $\mathbf{Y}(n)$. Otherwise, if (12) is not observable, the output $\mathbf{Y}(n)$ will give no enough information to estimate $\mathbf{X}(n)$. On the other hand, the controllability (which implies the stabilizability) guarantees that the state estimation error covariance (i.e., $\mathbf{P}_{n|n}$) is bounded (therefore reliable) [32].

E. Select the Model Order K

It is easy to see that the core of the TVLAP model is the matrix Φ defined in (10). It relates to the parameter K . In theory, for K , the larger, the better. However, in practice, due to the existence of noise and the Runge phenomenon in polynomial fitting, K should not be extremely large. According to the authors' experiences from experiments, the suggested value of K should be $2 \sim 5$ (0 and 1 are also useful for some cases). $K = 3, 4$ are typical options.

F. General Methodology for Non-White Noise

We, in this section, derive the TVLAP-KF model for a time series with colored (non-white) noise. That is, we no longer assume $x_s(n)$ to be white. Instead, we investigate the general colored case of it.

Suppose the noise part $x_s(n)$ could be modeled by ARMA($p, q|\varphi, \theta$) with the transfer function as

$$H(z) = \frac{\theta_0 + \theta_1 z^{-1} + \dots + \theta_q z^{-q}}{1 + \varphi_1 z^{-1} + \dots + \varphi_p z^{-p}}. \quad (27)$$

It means that the input of this ARMA system is a 1-D Gaussian white sequence $\boldsymbol{\varepsilon}(n)$ and the output is 1-D $x_s(n) =: \mathbf{V}(n)$. Since $x_s(n)$ is WSS, $\boldsymbol{\varepsilon}(n)$ is also WSS. Note that the first coefficient of the denominator polynomial is normalized to 1. Note also that the white noise case is the special case of ARMA(p, q) with $H(z) = \theta_0$, that is, ARMA(0, 0).

Let $r := \max\{p, q\}$, $\varphi_j := 0, \forall j > p$, and $\theta_j := 0, \forall j > q$. Then we have an alternative representation of (27) as

$$\begin{aligned} H(z) &= \frac{\theta_0 + \theta_1 z^{-1} + \dots + \theta_r z^{-r}}{1 + \varphi_1 z^{-1} + \dots + \varphi_r z^{-r}} \\ &= \frac{\theta_0 z^r + \theta_1 z^{r-1} + \dots + \theta_r}{z^r + \varphi_1 z^{r-1} + \dots + \varphi_r} \\ &= \theta_0 + \frac{\beta_1 z^{r-1} + \dots + \beta_r}{z^r + \varphi_1 z^{r-1} + \dots + \varphi_r} \end{aligned} \quad (28)$$

where $\beta_i := \theta_i - \theta_0 \varphi_i$, $i = 1, 2, \dots, r$.

Therefore, the state-space counterpart of (27) is

$$\begin{cases} \boldsymbol{\xi}(n+1) = \boldsymbol{\Xi} \boldsymbol{\xi}(n) + \boldsymbol{\Upsilon} \boldsymbol{\varepsilon}(n) \\ \mathbf{V}(n) = \boldsymbol{\Pi} \boldsymbol{\xi}(n) + \boldsymbol{\Lambda} \boldsymbol{\varepsilon}(n) \end{cases} \quad (29)$$

where

$$\boldsymbol{\Xi} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\varphi_r & -\varphi_{r-1} & -\varphi_{r-2} & \dots & -\varphi_1 \end{bmatrix} \quad (30)$$

$$\boldsymbol{\Upsilon} = [0 \ 0 \ \dots \ 0 \ 1]' \quad (31)$$

$$\boldsymbol{\Pi} = [\beta_r \ \beta_{r-1} \ \dots \ \beta_2 \ \beta_1] \quad (32)$$

and

$$\Lambda = \theta_0. \quad (33)$$

Note that this $V(n)$ is conceptually similar to the one in (9).

Thus, the entire state space model for our general model $x(n) = f(n) + x_s(n)$, namely, (1) should be

$$\begin{cases} X(n+1) = \Phi X(n) + G W(n) \\ Y(n) = H X(n) + V(n) \\ \xi(n+1) = \Xi \xi(n) + \Upsilon \varepsilon(n) \\ V(n) = \Pi \xi(n) + \Lambda \varepsilon(n) \end{cases} \quad (34)$$

which, by augmenting the state vector, is equivalent to

$$\begin{cases} \bar{X}(n+1) = \bar{\Phi} \bar{X}(n) + \bar{w}(n) \\ Y(n) = \bar{H}(n) \bar{X}(n) + \bar{v}(n) \end{cases} \quad (35)$$

where

$$\bar{X}(n) := \begin{bmatrix} X(n) \\ \xi(n) \end{bmatrix} \quad (36)$$

$$\bar{\Phi} := \begin{bmatrix} \Phi & \mathbf{0} \\ \mathbf{0} & \Xi \end{bmatrix} \quad (37)$$

$$\bar{w}(n) := \begin{bmatrix} G & \mathbf{0} \\ \mathbf{0} & \Upsilon \end{bmatrix} \begin{bmatrix} W(n) \\ \varepsilon(n) \end{bmatrix} \quad (38)$$

$$\bar{H} := \begin{bmatrix} H & \Pi \end{bmatrix} \quad (39)$$

and

$$\bar{v}(n) := \Lambda \varepsilon(n). \quad (40)$$

The system (35) could be handled by the Colored Kalman filter (see [28], Chapter 7.1). Note that the covariance matrix between the process noise $\bar{w}(n)$ and the measurement noise $\bar{v}(n)$ is

$$\begin{aligned} E[\bar{w}(n)\bar{v}^T(j)] &:= M(n)\delta_{k-j} = \begin{bmatrix} \mathbf{0} \\ \Upsilon \bar{R} \Lambda' \end{bmatrix} \delta_{n-j} \\ &= \begin{bmatrix} \mathbf{0} \\ \Upsilon \bar{R} \Lambda \end{bmatrix} \delta_{n-j} \end{aligned} \quad (41)$$

where δ_{n-j} is the Kronecker delta function; $\bar{R} = \bar{R} = \bar{R}(n)$ denote the variance of $\varepsilon(n)$. Note that $\bar{R}(n)$ is 1-D and constant over time. Now, the last thing to do is to estimate the value of \bar{R} .

Equation (27) reveals how $\varepsilon(n)$ generates $x_s(n) = V(n)$. Since $x_s(n)$ is a WSS process with constant variance R , we can have the constant variance \bar{R} of $\varepsilon(n)$, according to [17] (see Chapter 2.11.1), implicitly defined as

$$\begin{aligned} R &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 \cdot \bar{R} d\omega \\ &= \bar{R} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega \end{aligned} \quad (42)$$

where $H(e^{j\omega}) = H(z)|_{z=e^{j\omega}}$ is the Fourier frequency response of $H(z)$; the term $|H(e^{j\omega})|^2 \cdot \bar{R}$ denotes the power spectra of the output sequence $x_s(n)$. Suppose the impulse response of the system $H(z)$ is $h(n)$. According to the Parseval's theorem (see [17], Chapter 2.9.11), we further have

$$R = \bar{R} \cdot \sum_{n=-\infty}^{\infty} h^2(n) \quad (43)$$



Fig. 2. UWB ranging module and its supporting pole. (a) UWB module. (b) Supporting pole.

namely,

$$\bar{R}(n) = \bar{R}(n) = \bar{R} = \frac{R}{\sum_{n=0}^{\infty} h^2(n)}. \quad (44)$$

Note that in (44), R has already been estimated from the residual series $\delta(\bar{n})$ (see section IV-B). Note also that when $H(z)$ is stable, $\sum_{n=0}^{\infty} |h(n)|$ is convergent, which means that $\sum_{n=0}^{\infty} h^2(n)$ is also convergent. Besides, when $H(z)$ is causal, $h(n) = 0, \forall n < 0$. For a real system $H(z)$, the stability and causality are guaranteed.

V. EXPERIMENTS

The experiments are based on UWB ranging signals for range-based positioning problem. The source codes and data are available online at GitHub: <https://github.com/Spratm-Asleaf/TVLAP-KF>. In order to improve the positioning performances, we deploy many (more than required three for 2-D positioning) UWB anchors (i.e., UWB bases). Since this article mainly contributes to provide a model for a non-stationary signal through which the signal-model-based signal processing methods could be used, we do not conduct intensive comparison experiments regarding the performances of different signal-model-based signal processing methods. Only the representative exponential smoothing (ETS) method which is signal-model-free is adopted to compare with TVLAP-KF.

A. UWB Ranging Device

The UWB wireless transceiver chip for ranging that we adopt is the DW1000, produced by Decawave (<https://www.decawave.com/product/dw1000-radio-ic/>). The UWB chip, UWB antenna, power, power amplifier circuit, clock, communication buses [i.e., Serial Peripheral Interface (SPI)], micro-controller (i.e., STM32F103C8T6), etc., are integrated together to build the UWB ranging module. See Fig. 2. The ranging protocol is the symmetric-double-sided two-way time-of-arrival (SDS-TW-TOA) based on the IEEE Standard 802.15.4a [33].

B. Outlier/Dropout Correction

The first experiment is conducted outdoors. The field configuration of the UWB positioning environment with eight UWB anchors is shown in Fig. 3. This field is 100 m by 10 m. The UWB tag starts its trajectory near the A_0 and follows the orange dotted rectangle counterclockwise.

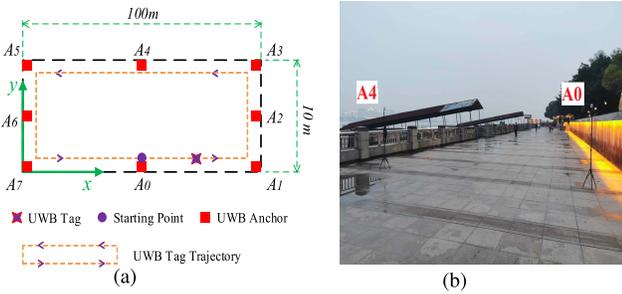


Fig. 3. Field configuration of the UWB positioning environment. We use eight UWB anchors indexed $A_0 \sim A_7$. In (b), only A_0 and A_4 are illustrated. (a) Topology. (b) Real field.

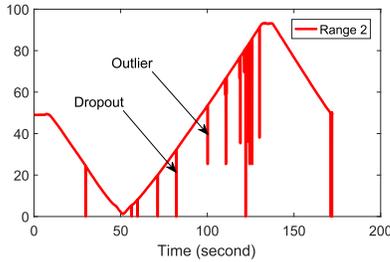


Fig. 4. Illustration of outliers and dropouts contained in the UWB ranging measurements. The data are from A_2 . Dimension in y axis: meter.

Taking ranging measurements from A_2 during one test as an example, it contains many outliers and dropouts. See Fig. 4. The sampling time is 0.1 s.

We aim to use the proposed TVLAP-KF ($K = 3$, $R = (0.05/3)^2$, $Q = 0.01^2$, $T = 0.1$) to identify and correct such outliers and dropouts. We use $(0.05/3)^2$ for R because the error range of UWB ranging sensors are ± 5 cm. Therefore, by using the popular “ 3σ ” rule, we value the ranging variance of UWB sensors as $(0.05/3)^2$. The results are shown in Fig. 5. From Fig. 5, we can see that no matter the outliers/dropouts are sparse [see like (a), (b), (d)] or dense [see like (c)], TVLAP-KF can always handle them.

After using an efficient multiple-bases TDOA method [34], the real-time position in y-axis is shown in Fig. 6. The comparison experiment is conducted with the ETS method which is free of the signal model. As we can see, the traditional ETS method (filtering parameter: 0.11) may suffer from severe time delay and not robust enough for dense outliers/dropouts [see, for example, Fig. 6(b) around 48th s].

C. Anomaly Detection (i.e., Fault Diagnosis) of Sensors

The second experiment is conducted indoors. In order to focus on the problem of anomaly detection, we choose range measurements that do not contain outliers/dropouts. Due to signal sheltering and complex electromagnetic environment, ranging signals from different anchors may have different ranging performances at different areas. Thus, we aim to select sensors without large errors from all the available anchors in one area to localize the moving target. The essence of the above issue is actually to diagnose the sensor fault (or detect the anomalies in ranging signals), in an online manner. Ranging signals provided by three of all available anchors are shown in Fig. 7(a).

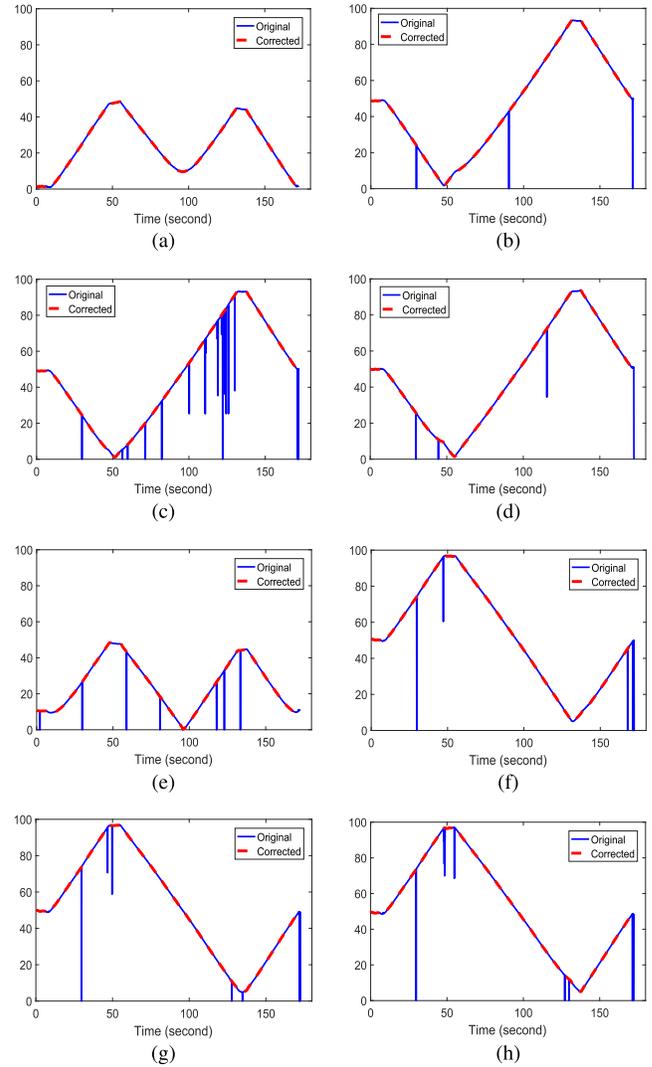


Fig. 5. Outliers/dropouts correction results using TVLAP-KF. (a) Range 0. (b) Range 1. (c) Range 2. (d) Range 3. (e) Range 4. (f) Range 5. (g) Range 6. (h) Range 7.

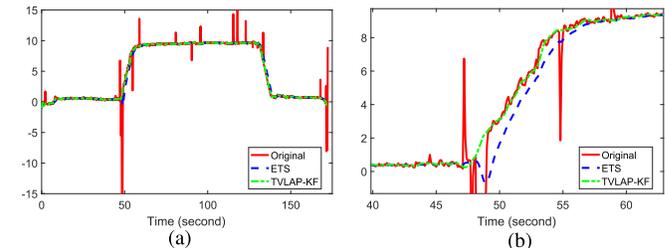


Fig. 6. Real-time position in y-axis. Unit in y-axis: meter. (a) Real-time position in y-axis. (b) Closeup of (a).

Intuitively, the Sensor 3 is with large error, since its ranging signal jumps at many places (for instance, when $t = 25 \sim 30$, around $t = 48$, and $t = 70 \sim 80$, t is time). Those jumps are in fact errors because a real moving target cannot maneuver in such a sharp way. On the other hand, if they are indeed generated from sharp maneuvers, Sensor 1 and Sensor 2 should have the same jumps in their ranging signals as well.

We aim to differentiate Sensor 3 from Sensor 1 and Sensor 2 so that Sensor 3 would be excluded to participate in positioning in this area. If we use TVLAP-KF ($K = 4$, $R = (0.05/3)^2$, $Q = 2.5^2$, $T = 0.1$; R is estimated from

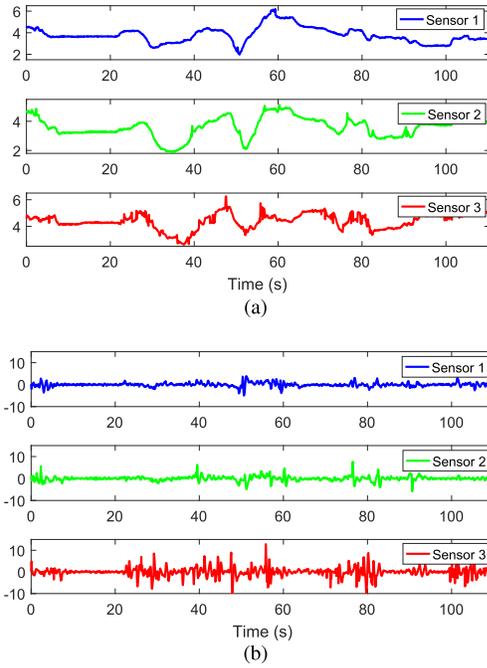


Fig. 7. UWB ranging signals and their first-order DMF. The variances of the three time series in (b) are 0.6424, 0.8293, and 3.3216, respectively. (a) UWB ranging signals from three of anchors. (b) DMF (first-order) of three UWB ranging signals.

the real data; Q is set to be relatively larger because we in this scenario emphasize more on observations than the system model) to estimate the changing pattern (first-order derivative) of ranging signals, we have Fig. 7(b).

From Fig. 7, it is easy to tell apart Sensor 3 from Sensor 1 and Sensor 2 because Sensor 3 has significantly large variance (or more outliers) in the first-order DMF of ranging signals. Note that the variance estimation method of a zero-mean sequence $x(n)$ is given as $\sum_{i=1}^n [x(i)]^2 / (n-1)$ (its online version, namely recursive version, is easy to derive).

D. Denoising and Prediction

In this part, we demonstrate the denoising (i.e., filtering) and prediction performance of TVLAP-KF with different K .

Since we do not know the true real-time positions (i.e., true signal) of a moving UWB tag, we cannot compare the denoised signal with the true signal. This is because we cannot control the true trajectory of the UWB tag to ideally follow the given trajectory with a constant moving speed. Therefore, we conduct the experiments over simulated data. Suppose we have $t = 0 : 0.1 : 120$, $x(n) = 5 \sin(0.1t) + \exp(0.03t) + G_w(\text{length}(t))$, the filtering results and the 200-step ahead predictions given by TVLAP with $K = 4$, $K = 1$, and $K = 0$, respectively, are displayed in Fig. 8, in which $Q = 300^2$, $R = 1^2$, and $T = 0.1$.

All the results given in Fig. 8 are the respectively best ones among ten simulations. The corresponding denoising (i.e., time from 0 to 100) and prediction (i.e., time from 100 to 120) mean square error (MSE) are given in Table I. As we can see, it is the use of DMF of signal that allows us to make more satisfactory filtering and prediction. Moreover, the order-insufficient models cannot promptly track the relatively sharp

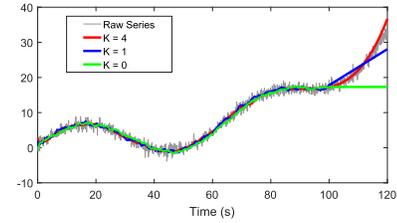


Fig. 8. Denoising and prediction performances of TVLAP with different K .

TABLE I

DENOISING AND PREDICTION MSE OF TVLAP

	Denoising MSE	Prediction MSE
K=4	0.0689	2.5979
K=1	0.0790	17.5410
K=0	0.1709	62.3891

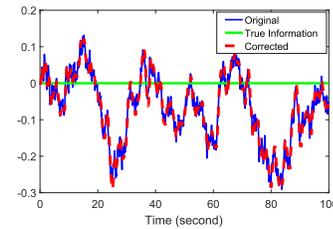


Fig. 9. Proposed method cannot handle the random walk noises contained in the measurement time series. The low-frequency component of the noise sequence is identified as non-zero readings of the measured quantity.

changing pattern of a time series. Therefore, high-order models (with relatively large K) are expected.

Note again that the outliers/dropouts correction is based on signal prediction. Therefore, the good performances in forecasting guarantee the performances in outliers/dropouts correction.

VI. DISCUSSIONS AND CONCLUSION

This article provides a state-space model for a non-stationary signal so that the model-based signal processing methods could be utilized. Its possible applications in engineering are discussed. Simulation suggests that the incorporation of DMF of a signal helps improve the filtering/denoising and forecasting performances, which further accounts for the high-accuracy and precision outliers/dropouts correction, and anomalies detection for measurement signals from sensors. However, the proposed model has the following shortcomings that give rise to alerts to users.

- 1) As mentioned in section III-B, non-WSS noises, such as the random walk noise, etc., cannot be handled. For example see Fig. 9.
- 2) If the measured quantity changes suddenly at a time instant, for example, there is a step contained in the true information $f(n)$ of (1), the proposed method would be likely to identify this sharp change as anomalies so that some true information would be, on the contrary, negatively influenced. See, for example, Fig. 10(a). However, this problem could be mitigated by an extra strategy: if successively many (e.g., 3 or 5) “outliers” occur and those “outliers” are not very far away from

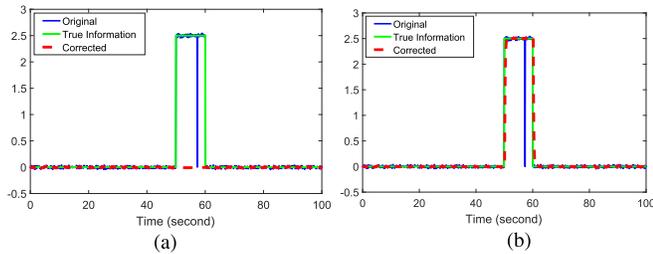


Fig. 10. Step change contained in the true information of the measurement time series. (a) Sudden change is wrongly treated as anomalies. (b) Performance after using the proposed remedy strategy. (a) Treat a step change as outliers. (b) Using the remedy strategy.

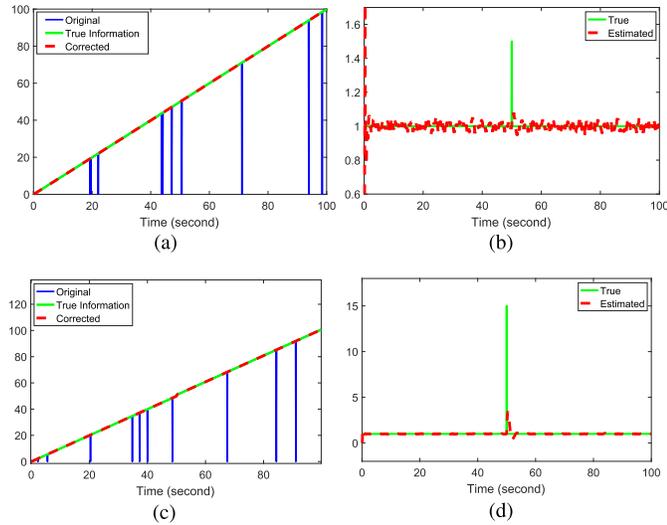


Fig. 11. Outlier is contained in the first-order derivative of the true information of the measurement time series. The outlier in (a) and (b) has a smaller amplitude as 1.5 so it cannot be identified, while in (c) and (d) has a larger amplitude as 15 so it can be identified. However, we can only identify that the outlier happens at this time instant. We cannot estimate the true amplitude of this outlier (i.e., the estimated amplitude is about 2.5 instead of 15). The measurement noise level (with a standard deviation of 0.05/3) keeps the same in the sub-figures (a) and (c). (a) Measurement time series. (b) First-order derivative of (a). (c) Another measurement time series. (d) First-order derivative of (c).

each other, we no longer treat those as outliers and use the predicted values to replace them. Instead, we accept them as truly believable measurements and directly feed to the proposed method. According to Theorem 2, the proposed method would let the estimated quantities (e.g., $p(n)$ and its derivatives) quickly converge to their true values again [see Fig. 10(b)].

- 3) An error caused by a change in one of the derivatives of the measurement time series might not be detected because the contribution of this change does not produce a significant dispersion in the measurement data. For example [see Fig. 11(a) and (b)] in which an outlier is contained in the first-order derivative. However, this depends a lot on the relative amplitude of the outlier compared to the measurement noise level. If the relative amplitude of the outlier is larger, it is still possible to identify the outlier contained in the first-order derivative [see Fig. 11(c) and (d)].

For more practical issues about implementing the TVLAP-KF, see <https://github.com/Spratm-Asleaf/Range-Correction>.

REFERENCES

- [1] T. Risset, C. Goursaud, X. Brun, K. Marquet, and F. Meyer, "UWB ranging for rapid movements," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2018, pp. 1–8.
- [2] S. Bose, A. De, and I. Chakrabarti, "Framework for automated earthquake event detection based on denoising by adaptive filter," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 9, pp. 3070–3083, Sep. 2020.
- [3] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, Jul. 2017.
- [4] Y. Zhu, B. Zhu, H. H. T. Liu, and K. Qin, "A model-based approach for measurement noise estimation and compensation in feedback control systems," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8112–8127, Oct. 2020.
- [5] M. Compagnoni *et al.*, "A geometrical–statistical approach to outlier removal for TDOA measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3960–3975, Aug. 2017.
- [6] Y. He, Y. Peng, S. Wang, and D. Liu, "ADMOST: UAV flight data anomaly detection and mitigation via online subspace tracking," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 4, pp. 1035–1044, Apr. 2019.
- [7] S. Sun, T. Tian, and H. Lin, "Optimal linear estimators for systems with finite-step correlated noises and packet dropout compensations," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5672–5681, Nov. 2016.
- [8] H. Ozkan, F. Ozkan, and S. S. Kozat, "Online anomaly detection under Markov statistics with controllable Type-I error," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1435–1445, Mar. 2016.
- [9] Q. Zhai and Z.-S. Ye, "Robust degradation analysis with non-Gaussian measurement errors," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 11, pp. 2803–2812, Nov. 2017.
- [10] S. Y. Cho, "Two-step calibration for UWB-based indoor positioning system and positioning filter considering channel common bias," *Meas. Sci. Technol.*, vol. 30, no. 2, Feb. 2019, Art. no. 025003.
- [11] H. H. Afshari, S. A. Gadsden, and S. Habibi, "Gaussian filters for parameter and state estimation: A general review of theory and recent trends," *Signal Process.*, vol. 135, pp. 218–238, Jun. 2017.
- [12] Y. S. Shmaliy, F. Lehmann, S. Zhao, and C. K. Ahn, "Comparing robustness of the Kalman, H_{∞} , and UFIR filters," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3447–3458, Jul. 2018.
- [13] M. Bai, Y. Huang, Y. Zhang, and G. Jia, "A novel progressive Gaussian approximate filter for tightly coupled GNSS/INS integration," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3493–3505, Jun. 2020.
- [14] H. Ferdowsi, S. Jagannathan, and M. Zawodniok, "An online outlier identification and removal scheme for improving fault detection performance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 908–919, May 2014.
- [15] Z. Li, Y. Yao, J. Wang, and J. Gao, "Application of improved robust Kalman filter in data fusion for PPP/INS tightly coupled positioning system," *Metrol. Meas. Syst.*, vol. 24, no. 2, pp. 289–301, Jun. 2017.
- [16] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, 2015.
- [17] P. S. Diniz, E. A. Da Silva, and S. L. Netto, *Digital Signal Processing: System Analysis and Design*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [18] S. Basu and M. Meckesheimer, "Automatic outlier detection for time series: An application to sensor data," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 137–154, Feb. 2007.
- [19] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013.
- [20] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 760–771, May 2013.
- [21] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.
- [22] W. H. Press, "Flicker noises in astronomy and elsewhere," *Comments Astrophys.*, vol. 7, no. 4, pp. 103–119, 1978.
- [23] R. F. Voss, "1/f (flicker) noise: A brief review," in *Proc. 33rd Annu. Symp. Freq. Control*, May 1979, pp. 40–46.
- [24] E. Kreyszig, *Introductory Functional Analysis with Applications*, vol. 1. New York, NY, USA: Wiley, 1978.

- [25] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 2018. San Francisco, CA, USA: Determination Press, 2015.
- [26] J. Duník, O. Straka, O. Kost, and J. Havlík, “Noise covariance matrices in state-space models: A survey and comparison of estimation methods—Part I,” *Int. J. Adapt. Control Signal Process.*, vol. 31, no. 11, pp. 1505–1543, 2017.
- [27] A. H. Mohamed and K. P. Schwarz, “Adaptive Kalman filtering for INS/GPS,” *J. Geodesy*, vol. 73, no. 4, pp. 193–203, May 1999.
- [28] D. Simon, *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley, 2006.
- [29] A. H. Sayed, “A framework for state-space estimation with uncertain models,” *IEEE Trans. Autom. Control*, vol. 46, no. 7, pp. 998–1013, Jul. 2001.
- [30] H. Roger and R. J. Charles, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
- [31] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” *J. Basic Eng.*, vol. 83, no. 1, pp. 95–108, 1961.
- [32] B. D. Anderson and J. B. Moore, *Optimal Filtering*. North Chelmsford, MA, USA: Courier Corporation, 1979.
- [33] Z. Sahinoglu and S. Gezici, “Ranging in the IEEE 802.15.4a standard,” in *Proc. IEEE Annu. Wireless Microw. Technol. Conf.*, Dec. 2006, pp. 1–5.
- [34] S. Cao, X. Chen, X. Zhang, and X. Chen, “Combined weighted method for TDOA-based localization,” *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 1962–1971, May 2020.



Shixiong Wang (Graduate Student Member, IEEE) received the B.Eng. degree in detection, guidance, and control technology and the M.Eng. degree in systems and control engineering from the School of Electronics and Information, Northwestern Polytechnical University, China, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore.

His research interest includes statistics and optimization theories with applications in signal processing (especially optimal estimation theory) and control technology.



Chongshou Li received the B.Eng. degree from Northwestern Polytechnical University, Xi’an, China, in 2010, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2015.

He was a Research Scientist with Nanyang Technological University, Singapore, from 2015 to 2016, and a Research Assistant Professor with the National University of Singapore, Singapore, from 2017 to 2021. He is currently an Associate Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. His research areas include hierarchical learning, forecasting, data mining, and machine learning.



Andrew Lim received the B.Sc. and Ph.D. degrees in computer and information sciences from the University of Minnesota, Minneapolis, MN, USA, in 1987 and 1992, respectively.

He is currently a Professor with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore. Many of his works have been published in leading journals, such as *Operations Research*, *Management Science*, and *Transportation Science*. His research interests include big data analytics, demand generation, and supply management problems in the domains of logistics, transportation, and healthcare.

Prof. Lim is one of the five Returning Singaporean Scientists and was a recipient of the China 1000 Talents Scheme.

Prof. Lim is one of the five Returning Singaporean Scientists and was a recipient of the China 1000 Talents Scheme.