General Modelling for Kalman Filter Applying to Investigating Deep Pattern of Data and Motion Modelling

Shixiong Wang, Chongshou Li, and Andrew Lim

Abstract—Data is the carrier of information. Data mining techniques derived from information modelling is of great significance in data science. In this paper, the Time-variant Local Autocorrelated Polynomial (TVLAP) model with Kalman filter is proposed to estimate the instantaneous mean (trend) of the interested data series, specifically, time series. Theoretical analysis for reliability guarantee and divergence risk has demonstrated the sufficiency of our method. The advantages of our methods, beyond trend estimating, embody: (1) identifying and predicting the peak and valley values of a data series; (2) reporting and forecasting the current changing pattern (increasing or decreasing of the trend); and (3) being real-time and workable for sequential data, not just block data. We will show that our TVLAP model is actually the generalization of Local-Level model and Holt’s model in time series analysis community, and Constant Velocity model and Constant Acceleration model in moving-object tracking field. More interestingly and excitingly, we could also use the method we propose to explain the philosophy and nature of motion modelling in Physics, meaning the theoretical validity of existences of physical concepts found in experiments, beyond Velocity and Acceleration, like Jerk, Snap, Crackle, and Pop could be revealed.

Index Terms—Information Modelling, Time-Variant Local Autocorrelated Polynomial, Extrema Detecting, Trend Tracking, Kalman Filter.

I. INTRODUCTION

Data analysis plays important role in information extracting. As one key type of the data, time series draws scholars’ wide attention, as in [1]. Time series is a philosophical concept which makes a discrete time function evolving over time. When it is instantiated in different subject communities or research fields, the concepts like signal, traffic volume, real-time stock price, real-time sales volume, real-time position of a moving object, precipitation over time, temporal sequential data et al. come out. Time series analysis and processing, which are suggested to differentiate as two terms by [2], are critical and popular in many academic communities. They are widely used for solving forecasting problems in different areas such as population forecasting in social science and ecology [3], financial time series [4], sales prediction in retail [5], real-time target tracking of moving objects [6], traffic volume forecasting in transportation [7] and so on. In consideration of whether the dynamics of the system generating the interested time series is known or not, the time series could be casted into three types: (a) completely known dynamics, for example, \( x_{n+1} = f(x_n, u_n) \); (b) partially known dynamics, for example, \( x_{n+1} = f(x_n, u_n, \epsilon_n) \); and (c) completely unknown dynamics, meaning only the sequential observations are available, for example, \( \{x_0, x_1, x_2, ..., x_n\} \). The mentioned \( x_n \) denotes the observations of the interested time series at the time index \( n \); \( f(\cdot) \) denotes the possible dynamics (function mapping) of how to generate such a time series \( x_n \); \( u_n \) denotes the related exogenous variables of generating \( x_n \); \( \epsilon_n \) denotes the uncertain variables like random variable or chaotic variable. For the cases of completely known dynamics and partially known dynamics, they are popular branches in some specific research communities like estimation theory in signal processing [8]. However, unfortunately, many real and general time series analysis problems are without the explicit system dynamics (at least impossible to know the exact results). For example, the financial time series, the sales volume time series, the sun spot time series, the traffic volume time series, and so on. Thus in most cases when we mention the term time series analysis, we actually mean that only the observations are available. We use the term traditional time series analysis, which is shorted as time series analysis, to denote this branch (only available observations). The reason we make clear this point is because the increasing interdisciplinary studies give more and more different specific meanings to the term time series analysis. Note again that time series is a philosophical concept meaning any discrete time function evolving over time. This paper pays exclusive attention to traditional time series analysis problems.

Information modelling is one of the focused aspects of data analysis. As an example, in time series analysis, we aim to philosophically find out the internal mechanism of a system generating the interested time series, and subsequently build a proper mathematical model to refactor (reconstruct) the dynamics of this system so that we can analyze the past and predict the future with satisfying accuracy, just like the autoregressive moving average (ARMA) model does. Such the process of finding the internal mechanism/dynamics of the information system that we are interested in is termed as Information Modelling. Generally, a time series evolves in an uncertain way rather than a deterministic one which makes the problem more complex. Up to now, two widely accepted mathematical models of an uncertain time series are
stochastic process \cite{8} and chaotic process \cite{9}. Specifically, to the best knowledge of the authors, the existing literature and methodologies regarding time series analysis belong to one of the following six categories.

The first category refers to simple methods like average method, (seasonal) naive method, drift method, moving average method, exponential smoothing method, regressions on time as Holt’s and Holt-Winters method, and so on \cite{10,11}. Those methods are theoretically easy to understand and practically convenient to adapt to many real problems with acceptable performances. Thus they are still fashionable at present, especially in industry.

The second category admits the prestigious Box-Jenkins \cite{8} and autoregressive conditional heteroscedasticity (ARCH) \cite{12,13,14} methodological families. This Box-Jenkins methodology is also known as ARMA and ARIMA model. The Wold’s Decomposition theorem \cite{15,16} is the philosophy of the ARMA model. Based on this theorem, a regular (rational-spectra) wide-sense stationary stochastic process can be described by the ARMA model mathematically and sufficiently \cite{2,16}. With collected real samples, parameters of the model can be estimated by methods like maximum likelihood method \cite{15,8}, least square method \cite{15,8}, and spectral estimation method et al. \cite{16,17,15,18,19}. Consequently, the past information (collected samples) can be utilized to build the fundamental dynamics of the investigated stochastic process. Then prediction can be conducted. Notably, we here discuss three popular and classic models: ARIMA, SARIMA and ARMAX \cite{20}. The ARIMA/(SARIMA) uses difference (seasonal difference) operator with proper order to transform non-stationary (irrational-spectra wide-sense stationary) stochastic process to be stationary (rational). Therefore, the philosophy of the ARMA model is followed. The ARMAX model takes into account another exogenous variable \( x \) as predictor so that we can handle the co-variate characteristic and/or non-stationarity of the interested time series. The ARMA, (S)ARIMA and ARMAX models are used for solving problems in different areas such as computer science \cite{21}, power system \cite{22}, signal processing, automatic control \cite{19,23,24,18} as well as management science and operations research community \cite{25,26}. Without doubt, the ARMA, (S)ARIMA and ARMAX models are becoming canonical time series analysis methodologies \cite{15,10,11}. More interestingly, in 2019 \cite{2} point out the theoretical insufficiency of the (S)ARIMA model and introduce the ARMA-SIN methodology to time series analysis community which is the general method/methodology of many existing time series analysis methods like (S)ARIMA, exponential smoothing, moving average, decomposition methods like STL \cite{27}, and so on. According to Wang et al., the nature of the ARMA-SIN methodology is Spectral Analysis and Digital Filtering, which supports the third category. As an complement to Box-Jenkins families, the ARCH families take into account the problem of heteroscedasticity, and developed many member methods like GARCH \cite{28}, NGARCH \cite{29}, ZD-GARCH \cite{30} and Spatial GARCH \cite{31} and so on.

The third category shows interests in Spectral Analysis and Digital Filtering \cite{21,42,33,84,32} first conducted the Spectral Analysis, also known as Fourier Frequency Domain Analysis, to time series. It is followed by several monographs \cite{35,33,34}. Unfortunately, the spectral analysis is not widely studied in the current main streams of time series analysis literature \cite{15,10,8,11}. The only exception is Fourier Series Expansion which is also known as harmonic analysis. This analysis describes that a periodic time series can be formulated as the summation of combined trigonometric functions \cite{34}. Recently, \cite{2} pointed out that the spectral analysis for periodic series is exactly the harmonic analysis. Applying spectral analysis for aperiodic series is able to produce good results as well. The reason why there exist such a dilemma is two-fold: (1) the spectral analysis approach is not a classical and conventional methodology for time series area. Moreover, applying spectral analysis requires basic concepts of linear system analysis and knowledge of complex analysis \cite{2}; (2) the time series community does not pay attention on Digital Filtering technology sufficiently. Because the spectral analysis is only powerful for time domain problems if accompanied with digital filtering. A digital filter (DF) is designed to process a time series in frequency domain by preserving the desired frequency components and wiping away the unwanted frequency components of a time series. The spectral analysis and digital filtering are mutually complementary. Spectral analysis tells us what to do, meaning which frequency components should be preserved or wiped away, and Digital Filtering tells us how to do. This category also admits the Wavelet Transform \cite{36}, Hilbert-Huang Transform \cite{37} et al., which are extensions of frequency-domain analysis approaches. Notably, Hilbert-Huang transform is prestigious for its data-adaptivity presented in Empirical Mode Decomposition method, meaning the transformation basis is not fixed in advance, which is rather different from the Fourier transform and wavelet transform. Empirical Mode Decomposition method is just like the STL, X11 et al. methods, aiming at decomposing a time series into different components.

The fourth category includes the reputed Adaptive Filters, like Kalman Filter family \cite{38,15,10}, Information Filter \cite{10,39} and so on. Information Filter is mathematically equivalent to the Kalman filter, only showing different derivation and mathematical forms (notations) \cite{38}. The Kalman Filter family is mainly well studied and utilised in Signal Processing community especially like the target tracking field \cite{40}. The traditional Kalman filter aims to give the unbiased minimum variance estimates to the system states for a linear and white (uncorrelated) noise system, and it is proved infinite-many times to be amazing in practice. However, for the general time series analysis community like finance time series field, traffic volume prediction field et al., the dilemma of using the Kalman filter is that it is hard and/or even impossible to obtain the claimed State Equation, an equation analytically describing the dynamics (namely changing pattern) of the interested time series. It is this reason that Kalman filter failed to draw much enough attention in general time series analysis community. Even so, for some problems with special characteristics, the Kalman filter still shows its power \cite{7,10}. Worthy of mentioning is that Hyndman et al. and Durbin et al. used the state space method to bridge the gap between...
Kalman filter and some time series analysis methods like local level method, exponential smoothing method, Holt's method (known as linear trend model according to [41]) and so on [10], [41]. Note that the Kalman Filter(s) conceptually denotes a family of adaptive filters based on the traditional Kalman filter, like Extended Kalman Filter [42], Unscented Kalman Filter [43], Cubature Kalman Filter [42], [43], Tobit Kalman Filter [44], and many other Improved Kalman Filters [45], [43]. The filters but the traditional one aim to handle the Nonlinear, Non-Gaussian, Non-white (correlated/colored) and Unknown-disturbances accompanied scenarios [38].

The fifth category focuses on the newly boosting Machine Learning methods, in computer science. In this category, two sub-categories should be paid attention to. (a) Kernel based method. Note that all the aforementioned four categories model the uncertain time series as a stochastic process. However, there also exist some other theories that regard a time series as a chaotic process [9]. For the chaotic process model, the kernel based methods standout. Those methods are like kernel adaptive filters [46], kernel affine projection (KAP) algorithm [47], [9], kernel recursive least squares algorithm [48], [9], and kernel least mean kurtosis based method [49]. As reported in those publications (but the authors in this paper have not restudied and verified so much), the kernel based methods have amazing performances in predicting some complex time series. (b) Deep Learning method. Another powerful method from machine learning community to forecast a time series is Deep Learning [50], [51], [52]. As they stated [50], [51], [52], [53], there exists many advantages of deep learning over other methods like: (i) easy to extract features of a time series and further make satisfying prediction; (ii) recurrent Neural Networks (RNN) and Long Short Time Memory Network (LSTM) have inborn power to identify and express the underlying dynamics/patterns of a time series which evolves along the time; (iii) allowing the multiple inputs and multiple outputs scenarios.

The sixth category admits the combined methods like Vector auto-regression [11], TBATS [54], [11]. Bootstrapping [55], [11], Bagging [11], ARMA-SIN [2] and so on. Specifically, Vector auto-regression is a special case of ARMAX model with multiple exogenous variables; TBATS is a combination of Fourier Series Expansion, Exponential Smoothing, State Space Model, and Box-Cox transformation; Bootstrapping and bagging show the nature of multiple model methods which is capable of data fusion so that the integrated results is better than any of the ingredient method. For more on this point, please see [11], [56], [57].

However, in today’s literature, scholars only pay attention to understanding the values of a time series, giving no emphasis on the changing pattern (i.e., increasing, decreasing, and how large/fast the change is) and extrema values (peaks and/or valleys) from viewpoint of modelling. The changing pattern, as well as the extrema values and where they occurs, other than the value themselves are also very important to investigate, for example, the Stock Understanding and Forecasting problem. In other words, the information contained in a data series is more than what the values themselves of it are. Thus the Deep Patterns of the data should be further studied. To this end, in this paper, we will firstly model a time series as a non-stationary stochastic process presenting the properties of variant mean. Then the Time-Variant Local Autocorrelated Polynomial Model with Kalman Filter will be proposed to dynamically estimate the instantaneous mean (trend) of the interested time series. After that, we could further understand and forecast the focused time series.

The advantages (contributions) of our method embody:
(a) Extracting the instantaneous mean (trend) of a time series;
(b) Automatically identifying and predicting the peak and valley values of a time series;
(c) Automatically reporting and forecasting the current changing pattern (increasing or decreasing of the trend), termed as Trend Tracking here;
(d) Being real time, meaning also workable for the sequential data, not just block data, as exponential smoothing and moving average method do;
(e) Being theoretically sufficient, meaning the reliability guarantee is derived.

Remark 1: When we mention a time series processing method to be online (real-time), we actually mean it works for online data (sequential data). In this sense, exponential smoothing, moving average methods et al. are real time methods, while the ARMA method, STL decomposition, Fourier series expansion et al. are not real time methods. However, it does not mean we absolutely do not need any pre-collected samples. Because at least we need to properly select some related coefficients/parameters of algorithms we use based on some real samples. For example, the selection of smoothing coefficient α in exponential smoothing method.

For scholars interested in Kalman filter, our Time-Variant Local Autocorrelated Polynomial model provides the general State Equation for a given time series analysis problem, which allows the general choice and use of Kalman filter. Besides, and more excitedly, it is notable that the proposed Time-Variant Local Autocorrelated Polynomial model could also help us figure out the philosophy and nature of motion (kinematics) in physics.

II. PRELIMINARY ON STOCHASTIC PROCESS

Recall that in Introduction we have already point out the philosophy of ARMA and (S)ARIMA. They in fact aim to model the dynamics of a stochastic process so that we can use the past information from collected time series to satisfyingly predict the future. Mentioning this, we cannot ignore the reputed Wold’s Decomposition Theorem in stochastic process analysis.

Firstly we give the strict mathematical definition of the Wide-sense stationary (WSS) stochastic process. We use \( x(t) \) to denote a continuous time stochastic process and \( x(n) \) a discrete time one, meaning \( t = T \cdot n \), if the sampling time period is \( T \).

**Definition 1 (WSS Stochastic Process [16]):** A real-valued stochastic process \( x(t) \) is WSS if it satisfies:

- **Invariant Mean:** \( E\{x(t)\} = \eta \), where \( \eta \) is a constant;
- **Invariant Autocorrelation:** \( E\{x(t_1)x(t_2)\} = E\{x(t_1 + \tau)x(t_1)\} = R(\tau) \), meaning it only depends on \( \tau := t_2 - t_1 \), having nothing to do with \( t_1 \) and \( t_2 \).
Invariant autocorrelation immediately admits the Invariant Variance, since \( E\{x(t)\}^2 = R(0) \). Then we should turn to Wold’s Decomposition Theorem.

**Theorem 1** (Wold’s Decomposition Theorem [16]): Any wide-sense stationary (WSS) 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 (meaning uncorrelated): \( E\{x_r(n) + \tau x_p(n)\} = 0 \).

The detailed concepts of Regular process and Predictable process could be found in [16], [2]. Intuitively, the regular process is mathematically as \( x_r = ARMA(p,q) \), and predictable process \( x_p \) is the sum of trigonometric functions \( \cos(w_t n) \) and \( \sin(w_t n) \), for some \( w_t \). See also Eq. (2).

Alternatively, one can also understand a predictable process as a periodic discrete time function (periodic time series), since the Discrete Fourier Series/Transform theory asserts that any periodic discrete time function could be decomposed into the sum of trigonometric functions [2].

Thus if a WSS \( x \) is without the \( x_p \) part, we can use the ARMA method to model it and use the past information to train the coefficients so that we can refactor the dynamics of the interested \( x \). This is the philosophy of Box-Jenkins methodology [8]. If a WSS \( x \) is with the \( x_p \) part, sometimes we can use the ARIMA model to detrend with difference operator, and SARIMA model with seasonal difference operator to remove the seasonal components, and finally regard the transformed remainder as a ARMA process and then use the ARMA model to fit. These are the philosophies and natures of ARMA, ARIMA and SARIMA models. For more on this point, please refer to [2].

III. Problem Formulation and Motivations

**A. Notations**

1) Let \( v = a : l : b \) define a vector \( v \) being with the lower bound \( a \), upper bound \( b \) and step length \( l \). For example, \( v = 0 : 0.1 : 0.5 \) means \( a = 0 \), \( b = 0.5 \), and \( l = 0.1 \). Thus \( v = [0, 0.1, 0.2, 0.3, 0.4, 0.5]^T \);
2) Let the function \( \text{length}(x) \) return the length of the vector \( x \). For example, if \( x = [1, 2, 3]^T \), we have \( \text{length}(x) = 3 \);
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 100s, and the sampling time is \( T_s = 0.5s \), we will have \( n = t/T_s = 0 : 1 : (\text{length}(t) - 1) = 0 : 1 : 200 \); Let \( N = \text{length}(n) \). For notation simplicity, we also use \( T \) as an alternative to \( T_s \);
4) Let \( x(n) \) or \( x_r(n) \) denote the interested time series, shorted as \( x \); Let \( y \) denote the transformed time series from \( x \);
5) Let \( x_0 \) denote the exact dynamics of \( x \), \( \hat{x} \) the estimated dynamics of \( x \). It means \( x \) is generated from its ground truth \( x_0 \). The closer between the transformed (estimated) series \( \hat{x} \) and \( x_0 \), the better the transform and the more proper the ARMA model could be used to refactor the dynamics of \( \hat{x} \);
6) Let the function \( \text{mean}(x) \) return the mean of a random variable \( x \), and \( \text{var}(x) \) the variance of it;
7) Let the operator \( ARMA(p,q|\varphi, \theta) \) denote a ARMA process with autoregressive order of \( p \) and moving average order of \( q \). Besides, the coefficient vectors \( \varphi \) and \( \theta \) are for autoregressive part and moving average part, respectively. \( ARMA(p,q|\varphi, \theta) \) is shorted as \( ARMA(p,q) \);
8) Let the operator \( SIN(K|A,B) \) denote a process that is the sum of trigonometric functions, namely
\[
x(n) = A_0 + 2 \sum_{k=1}^{K} \left[ A_k \cos\left(\frac{2\pi}{K}kn\right) + B_k \sin\left(\frac{2\pi}{K}kn\right) \right] \]
\[
= \sum_{k=0}^{K} X_k e^{j\frac{2\pi}{K}kn}
\]
where \( A_0, A_k, B_k \) and \( X_k \) could be found in [58], [2], and \( j \) is the complex unit. Note that \( X_k \) here actually means the Discrete Fourier Series/Transform (DFS/DFT) coefficients of \( x(n) \). \( SIN(K|A,B) \) is shorted as \( SIN(K) \). Note also that, by DFS/DFT theory, \( SIN(K) \) could be any periodic time series with the period of \( K \) [2].

B. General Model of Non-stationary Stochastic Process

In this paper, we consider a general model describing a non-stationary stochastic process with the following form
\[
x(n) = f(n) + [x_r(n) + x_p(n)],
\]
where \( x_r(n) := ARMA(p,q) \) is a regular process and \( x_p(n) := SIN(K) \) is a predictable process; \( f(n) \) is a deterministic function denoting the mean of the time series \( x(n) \). Note that the expectations of the terms \( x_r(n) \) and \( x_p(n) \) are all zero. Therefore, we have \( \text{mean}(x) = f \), and \( \text{var}(x) = [\text{var}(x_r) + \text{var}(x_p)] \).

**Remark 2**: Note that if \( [3] \) takes its special case as \( x(n) = x_r(n) + x_p(n) \), we have \( x(n) \) as a wide-sense stationary process. For a wide-sense stationary process, we can use the Box-Jenkins methodology (ARMA, ARIMA, SARIMA) to model, to some extent. For the more exact model, see ARMA-SIN methodology in [2].

Equation \( [3] \) admits
\[
x(n) = [f(n) + x_p(n)] + x_r(n).
\]

Since the term \( x_p(n) \) is not random, in some application scenario we can consider this part as the trend so that the phenomenon of Seasonal Trend (periodic trend) is possible, since \( x_p \) is periodic. Then \( [3] \) could be rewritten as
\[
x(n) = f(n) + x_r(n),
\]
meaning we can regard the non-stochastic part as a whole and termed as trend (general-sense trend) so that the seasonal trend is possible. The main reason of doing so is that in real-time application scenario, we cannot easily realize the pattern of periodicity, compared to block data analysis. As long as
a satisfying prediction could be made to some degree, we actually do not care whether the series is periodic or not. Another concern is that the periodicity only contributes more to the long-term prediction problem. Because in the long-run sense, the model taking into account the periodicity is more satisfying. That means, if in some cases we do not care much about the very long-term issue, it is reasonable to generalize the model as \( f(n) \) and \( f(n) \). Instead, when we need to differentiate the periodicity, we could use the time series decomposition methods, including the STL method, ARMA-SIN methodology and so on, to handle.

C. General Motivations

The real-time analysis and processing of a time series requires the algorithm to be able to: (a) handle the sequential data; (b) identify the peak values and the valley values. Many existing methods are workable for sequential data, like exponential smoothing, moving average, Hold’s method and so on. However, neither them nor other reported methods can deal with the minima or maxima detecting (Extrema Detecting) problem. The extrema detecting problem becomes harder for real-time processing scenario, especially the noised scenario. Thus we will in this paper investigate the real-time extrema detecting problem. Motivation will disclose our general idea to do so.

Motivation 1: Any continuous function could be approximated by a polynomial function with sufficient orders. Plus, the polynomial functions are high-order differentiable, meaning we can assert a point to be a: (a) minimum if the first-order derivative is zero-valued and second-order derivative is positive, (b) maximum if the first-order derivative is zero-valued and second-order derivative is negative, at this point. Thus we desire to use an order-sufficient polynomial to regress the interested time series in real time. Thus going here, the extrema detecting problem only concerns the real-time and no-time-delay polynomial regression.

Note that the forecasting to the real-time changing pattern of trend (namely the Trend Tracking problem) could also be handled by the first-order derivative, increasing if positive, or decreasing if negative. Note also that the extrema detecting problem and trend tracking problem concerns more on the low-frequency trend part of \( f(n) \), having little to do with high-frequency components of \( x_r(n) \) (because this part is generally regarded as noise in this sorts of problem settings). This point will be further discussed later.

However, the dilemmas are that existing polynomial regression methods: (a) like global traditional polynomial regression only works for block data, fails to work for real-time scenario, meaning it performs bad for relatively long-term forecasting; (b) like Holt’s method is order-deficiency, only holding the first-order (linear trend) polynomial which introduces potential time-delay problem when the changing rate of the interested time series is sharp. The issue of time-delay also exists in non-polynomial methods like exponential smoothing and moving average, reported by [2]. Thus we aim to in this paper introduce the Time-Variant Local Autocorrelated Polynomial Model with Kalman Filter (TVLAP-KF) to handle this issue.

As for the issue of the non-stationarity of the interested time series, we have the general idea to settle this in Motivation 2.

Motivation 2: Facing the non-stationarity problem, if we can estimate out the function \( f(n) \), we could then have an estimated \( \hat{x}_r(n) = [x(n) - \hat{f}(n)] \) which is considered to be a ARMA process, meaning we could use the Box-Jenkins methodology to handle. Here, \( \hat{f}(n) \) is the estimate to \( f(n) \). Then we can apply Box-Jenkins methodology to handle \( \hat{x}_r(n) \), and the non-stationarity problem is solved.

For terminological briefness, we in this paper refer to the estimation procedure of \( f(n) \) as Mean Estimating.

IV. MEAN ESTIMATING AND EXTREMA DETECTING

A. Time-Variant Local Autocorrelated Polynomial Model

In this section, we will introduce the Time-Variant Local Autocorrelated Polynomial Model with Kalman Filter to handle the online polynomial regression problem (mean estimation) and extrema detection problem. As a demonstration, we in this section only takes the special case of \( x(n) = f(n) + \text{white}(n) \), where \( \text{white}(n) \) denotes a white-noise (uncorrelated) series. Plus, readers are invited to refer to [38] for profoundly understanding about Kalman filter, including the estimation method and the prediction method. In consideration of paper length and necessity, we will not introduce more about the theory of Kalman filter.

As we state in Introduction, the Kalman filter is powerful only when the required state equation (also known as system equation, system dynamics equation, or transfer function in state space et al. in control theory and signal processing community) is known. This limits the wide utilization of Kalman filter in time series analysis and signal processing, because a general time series is without the explicit evolution pattern (state equation), unlike many problems in control theory and signal processing. However, the kalman filter is extremely attractive for us due to that: (a) it is an online algorithm; (b) it is an optimal estimation method in linear-system and white-noise sense. Since we are concerned with the mean estimating (trend estimating) problem, and the mean of a time series is generally low-frequency, why not to use an order-sufficient polynomial to refactor the changing pattern (dynamics) of the mean of a time series, namely, the \( f(n) \) part in \( x(n) \)? That is to say, why do not we modelling the information dynamics contained in the data?

The theoretical validity and sufficiency of polynomial regression is from the prestigious Weierstrass approximation theorem [59]. However, the dilemma is the real-time and extrema detecting issues, meaning we expect the algorithm to be able to not only work online but also simultaneously return the current first-order and second-order derivatives of such a well-approximated polynomial. Well-approximation here means the regressed polynomial and the raw time series are close enough.

Fortunately, the dilemma is possible to detour when we ask for help from the Taylor’s expansion, asserting that a real-valued function \( f(t) \) that is infinitely differentiable at a real number \( t_0 \) could be the power series with the form of

\[
f(t) = f(t_0) + \frac{f^{(1)}(t_0)}{1!}(t-t_0) + \ldots + \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k + \ldots,
\]
where \( f^{(k)}(t_0) \) denotes the \( k^{th} \)-order derivative of \( f(t) \) at \( t_0 \).

Note that (6) is also a polynomial with a special mathematical form rather than its general form below

\[
    f(t) = p_0 + p_1 t + p_2 t^2 + ... + p_k t^k + ..., \tag{7}
\]

where \( p_k \) is a constant.

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 (high-order) 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 \( \forall \varepsilon > 0, \exists K > 0, \) such that

\[
    \sup_t |f(t) - p_K(t)| < \varepsilon, \tag{8}
\]

over a compact (i.e., closed in real space) interval, where \( p_K(t) \) denotes the polynomial with order of \( K \). For simplicity, we ignore \( K \) in notation. We have

\[
    p(t) = p_0 + p_1 t + p_2 t^2 + ... + p_k t^k + ..., \tag{9}
\]

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 disaster, according to (8). Suppose we have interests in the properties at the discrete time index \( n \), Eq. (6) could then be rewritten as (10).

\[
    p(t) = p(n) + \frac{p(1)(n)}{1!} (t-n) + ... + \frac{p(k)(n)}{k!} (t-n)^k + ... \tag{10}
\]

Thus, the traditional polynomial regression (7) could be regarded as the special case of (6) when we investigate the problem from the starting point of the time, namely, \( t_0 = 0 \), meaning the polynomial (10) is a local polynomial, while (7) is a global polynomial. For intuitive understanding, see Fig.

Fig. 1. Global and local polynomial

If we only pay attention to the case of \( t = n+1 \) and truncate the polynomial on the order of \( K \), we have (10) as

\[
    p(n+1) = \sum_{k=0}^{K} \frac{p(k)(n)}{k!} T^k = \sum_{k=0}^{K} \frac{T^k}{k!} p(k)(n), \tag{11}
\]

where \( T \) denotes the time slot between the discrete time index \( n+1 \) and \( n \), also known as Sampling time which bridge the gap between the continuous time \( t \) and the discrete time index \( n \) as \( t = nT \). For the detailed concepts of the Sampling, the continuous time variable \( t \), and the discrete time variable \( n \), please see [2].

Interestingly, Eq. (11) holds the following powerful characteristics:

1) It is actually the required State Equation in Kalman filter. 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 \( f(t) \), which is attractive in extrema detection and other analysis.

Remark 3: The state equation expressed by Eq. (11) is our model to refactor the dynamics (changing pattern) of the trend of \( x(n) \), namely \( f(n) \) of \( x(n) \).

Note that the online extrema detecting problem only cares about the trend \( f(n) \) of \( x(n) \), have little to do with \( g(n)x_r(n) := \text{white}(n) \). Thus the trick here is that we treat \( x(n) \) as a series noised by \( x_r(n) \) from \( f(n) \). From the viewpoint of Estimation theory and Kalman filter [38], \( f(n) \) is true value of the state, \( x(n) \) is measurement, and \( x_r(n) \) is measurement noise. Thus in this sense, \( X_0(n) \) in Eq. (12) is the estimate to \( p(n) \), further speaking, also to \( f(n) \), while \( x(n) \) is the measure to \( f(n) \). This perspective is the basis of using Kalman filter.

Now it is possible for us to apply the Kalman filter as long as we could have the state space representation of (11). In consideration of the fact that the terms \( p^{(k)}(n), \ k = 0, 1, 2, ..., K \) actually change through the time and convey the explicit physical meanings of \( p(n) \) (the complete changing patterns), we could choose them as our state variables. Note that only a variable instead of a constant could be considered as a state variable. Thus we define our state vector as

\[
    X(n) = 
    \begin{bmatrix}
    X_0(n) \\
    X_1(n) \\
    X_2(n) \\
    \vdots \\
    X_K(n) \\
    \end{bmatrix} =
    \begin{bmatrix}
    p^{(0)}(n) \\
    p^{(1)}(n) \\
    p^{(2)}(n) \\
    \vdots \\
    p^{(K)}(n) \\
    \end{bmatrix}, \tag{12}
\]

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 (11) as

\[
    X(n+1) =
    \begin{bmatrix}
    1 & T & T^2 & \cdots & T^K \\
    0 & 1 & T & \cdots & T^{K-1} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & 1 & \cdots & 1 \\
    \end{bmatrix}
    \begin{bmatrix}
    X_0(n) \\
    X_1(n) \\
    X_2(n) \\
    \vdots \\
    X_K(n) \\
    \end{bmatrix}, \tag{13}
\]

Eq. (13) implies that when we model the dynamics of \( f(n) \), we actually admit the \( K^{th} \)-order derivative to remain constant over time. This is therefore the cost of truncation, meaning we just use the Level Model (0-order holder, 0-order local polynomial) to smooth the \( K^{th} \)-order derivative. To be more specific, Level model models the slow-changing pattern of
the $K^{th}$-order derivative rather than fast-changing pattern. For more on this point, see the philosophy, implementation and performances of Level model introduced in [41].

In time series analysis scenario, only the sequential data $x(n)$ is obtainable, measurable. Thus we in our state space adaptation should define the measure vector as

$$Y(n) = x(n) = f(n) + x_r(n) = p(n) + x_r(n) := f(n) + \text{white}(n). \quad (14)$$

By doing so, we have the Measurement Equation (also known as Observation Equation, or Output Equation) as

$$Y(n) = [1 \ 0 \ 0 \ \cdots \ 0] X(n) + V(n), \quad (15)$$

where $V(n)$ is used to model the white$(n)$ part.

Besides, let’s define

$$\Phi = \begin{bmatrix} 1 & T & T^2 & \cdots & T^K \\ 0 & 1 & T & \cdots & \frac{T}{(K-1)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \quad (16)$$

as our System Matrix in Kalman filter, and

$$H = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (17)$$

as our Measurement Matrix. Note that the $\Phi$ and $H$ are constant if given the order $K$. Suppose the state noise vector is $W(n)$ with covariance $Q(n)$ and its noise-driven matrix is $G$; the measurement noise state vector is $V(n)$ with covariance $R(n)$, we then have a state-space model as a stochastic process to a time series $x(n) = p(n) + \text{white}(n) = f(n) + \text{white}(n)$ in (18). $G'$ denotes the transpose of $G$.

$$\left\{ \begin{array}{l}
X(n + 1) = \Phi X(n) + GW(n) \\
y(n) = HX(n) + V(n),
\end{array} \right. \quad (18)$$

where $W(n)$ denotes the modelling error. Specifically, it models the difference between (13) and (9), that is, between (11) and (9).

Eq. (18) is the linear system model of $x(n)$ in state space. For linear system theory, please see [2]. Note that the mathematical form of $G$ is not unique, meaning we can define it as any proper one. Some simple examples are: (a) $G = [T^k, \ldots, T, 1]$’ so that $W(n)$ should be a 1-dimensional vector denoting the disturbance exerted to $X_k(n)$; (b) $G = \text{diag}\{T^k, \ldots, T, 1\}$ so that $W(n)$ should be a $(K+1)$-dimensional vector denoting the disturbance exerted to $X(n)$; (c) $G$ as an identity matrix so that $W(n)$ should be a $(K+1)$-dimensional vector denoting the disturbance exerted to $X(n)$. The difference between (b) and (c) is reflected in their corresponding $Q(n)$.

For our model (5) studied in this paper, it is stationary in noise, meaning $R(n)$ is constant over time. Let $R := R(n)$.

Remark 4: If we allow the heteroscedasticity in noise, we could re-model (5) as

$$x(n) = f(n) + g(n)x_r(n), \quad (19)$$

where $g(n)$ is another deterministic function.

B. Estimate the $R$ and $Q(n)$

Actually, $R$ is easy to estimate from the historical observations (measures) 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 $x(\bar{n})$, we should have the fitting residual $\delta$ as

$$\delta(\bar{n}) := x(\bar{n}) - \bar{p}(\bar{n}). \quad (20)$$

According to our model assumption, $\delta(\bar{n})$ should be a wide stationary stochastic (WSS) process, specifically, a rational-spectra WSS process, meaning the selected order of $\bar{p}(t)$ is proper if and only if $\delta(\bar{n})$ is rational-spectra wide stationary. For the details on testing the rational-spectra wide stationarity, see [2].

Thus we have

$$R = \text{var(}\delta). \quad (21)$$

Note that in (18), $V$ is 1-dimensional, meaning $R$ is also a scalar rather than a vector, denoted as $\bar{R}$.

As for the real-time, adaptive estimate to $Q(n)$ when given (18) and $\bar{R}$, it is a bit more complicated. Readers are invited to refer to [60], [61], [62] and other similar literature concerning the process covariance estimation problem.

C. TVLAP Model with Kalman Filter

Now, it is sufficient to use the Kalman filter to handle the linear system (18), during which we could also estimate the real-time value $\bar{X}_0$ of $p(n)$, and real-time values of $k^{th}$-order derivative $\bar{X}_k$ of $p(n)$, where $p(n)$ is the mean function of the interested time series $x(n)$. The estimates to derivatives admit the feasibility of extrema detecting and the changing-pattern prediction (to predict the increasing pattern or decreasing pattern). The Kalman filter is given in Algorithm IV-C.

We in this paper term the presented method in this section as Time-Variant Local Autocorrelated Polynomial Model (TVLAP) with Kalman Filter, shorted as TVLAP-KF. Time-Variant means the coefficients of the used polynomial model (11), namely $p^{(k)}(n)/k!$ and $\bar{X}_k(n)$, change over time. The meaning of the word Local has been explained before in Fig. [4].

Autocorrelated means the coefficients of the used polynomial are not independent, are instead highly related. Because we have

$$p^{(k)}(n) = \frac{d}{dt} \left[ \frac{p^{(k+1)}(n)}{(k+1)!} \right]. \quad (22)$$

Finally we present the entire algorithm of Online Mean Estimating and Extrema Detecting in Algorithm IV-C.

Remark 5: Note that the RISK BOUND of the Algorithm IV-C is revealed by the covariance matrix $P$. This is easy to understand from the theory of Kalman filter [38].

D. Reliability Guarantee of TVLAP-KF

In this section, we are concerned to analyze the performances of the proposed TVLAP-KF. That is, we need to investigate whether the TVLAP-KF could recursively approximate $p(t)$ defined in (10) with satisfying accuracy.
Before we start, we first give two definitions regrading the observability and contractility of a linear time-invariant system.

Definition 2: The linear time-invariant system defined as (18) is uniformly completely observable if the matrix $O$ defined by the matrices pair $[\Phi, H]$ is of full rank.

\[
O = [H', \Phi' H', \ldots, (\Phi')^K H'],
\]

is of full rank.

Definition 3: The linear time-invariant system defined as (18) is uniformly completely controllable if the matrix $C$ defined by the matrices pair $[\Phi, G]$ is of full rank.

As we can see, in order to calculate the rank of the matrices $O$ and $C$, we must cope with the calculation of the power of the matrix $\Phi$, specifically, $\Phi^K(T)$. According to the speciality of our defined $\Phi$, we have Lemma 1.

**Lemma 1:**

\[
\Phi^K(T) = \Phi(\alpha T).
\]

**Proof:** Actually, there exists a matrix

\[
A = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix},
\]

such that

\[
\Phi(T) = e^{AT}.
\]

Thus, $\Phi^K(T) = e^{KAT} = \Phi(\alpha T)$. That is,

\[
\Phi^K = \begin{bmatrix}
1 & KT & \frac{(KT)^2}{2} & \cdots & (\frac{KT)^K}{K!} \\
0 & 1 & KT & \cdots & (\frac{KT)^K}{(K-1)!} \\
0 & 0 & 1 & \cdots & (\frac{KT)^K}{(K-2)!} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}.
\]

Note that not all the transition matrix $\Phi$ of a general linear time-invariant system can be represented as $[27]$.

**Lemma 2:** The Vandermonde matrix defined as

\[
V = \begin{bmatrix}
1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\
1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\
1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \alpha_m & \alpha_m^2 & \cdots & \alpha_m^{n-1}
\end{bmatrix},
\]

is of full rank if $\forall i \neq j, \alpha_j \neq \alpha_i$.

**Proof:** Since $\det(V) = \prod_{1 \leq i < j \leq n}(\alpha_j - \alpha_i) [63]$ (see chapter 6.1), the lemma stands.

**Lemma 3:** The linear time-invariant system defined in (18) is uniformly completely observable, if $K$ is not very large.
The linear time-invariant system defined in (18) is uniformly completely observable, then

\[
\begin{bmatrix}
    T^K \\
    K! \\
\end{bmatrix} \sum_{i=0}^{K-1} (T)^{i} \frac{1}{i!} (K-i)! \\
= \begin{bmatrix}
    \frac{T^K}{K!} \\
    \vdots \\
    \frac{T^K}{K!} \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\begin{bmatrix}
    T \\
    1 \\
    \vdots \\
    K \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\begin{bmatrix}
    0^0 0^1 0^2 \ldots 0^K \\
    1^0 1^1 1^2 \ldots 1^K \\
    2^0 2^1 2^2 \ldots 2^K \\
    \vdots \\
    K^0 K^1 K^2 \ldots K^K \\
\end{bmatrix} = K + 1, \\
\end{align}
\]

meaning \( O \) is of full rank. According to Definition 2 this lemma stands.

Lemma 4: The linear time-invariant system defined in (18) is uniformly completely controllable, if \( K \) is not very large and \( G \) is given as one of the following cases:

(a) \( G_1 = \left[ \frac{T^K}{K!}, \ldots, T, 1 \right] \);
(b) \( G_2 = \text{diag} \left( \frac{T^K}{K!}, \ldots, T, 1 \right) \);
(c) \( G_3 \) as an identity matrix \( I \). \( I \) denotes the identity matrix with proper dimensions.

Proof: Let \( C_{\Phi,G} \) denotes the controllability matrix defined by the pair \( \Phi, G \).

Since

\[
C = [G, \Phi G, \ldots, \Phi^K G],
\]

\[
\begin{bmatrix}
    \frac{T^K}{K!} \\
    \vdots \\
    \frac{T^K}{K!} \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\begin{bmatrix}
    T \\
    1 \\
    \vdots \\
    K \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} \frac{1}{i!} (K-i)! \\
\begin{bmatrix}
    0^0 0^1 0^2 \ldots 0^K \\
    1^0 1^1 1^2 \ldots 1^K \\
    2^0 2^1 2^2 \ldots 2^K \\
    \vdots \\
    K^0 K^1 K^2 \ldots K^K \\
\end{bmatrix} = K + 1, \\
\end{align}
\]

\[
\text{By binomial theorem, the entry of } C \text{ at } (I, J) \text{ is therefore}
\]

\[
C(I, J) = \sum_{i=0}^{K-I} \frac{(I-J)^i}{i!} (K-I-i)! \\
= \frac{1}{(K-I)!} (J+T)^{K-I},
\]

where \( I, J = 0, 1, 2, \ldots, K \), giving \( C \) further as

\[
C = \begin{bmatrix}
    \frac{T^K}{K!} \\
    \vdots \\
    \frac{T^K}{K!} \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} (K-i)! \\
\begin{bmatrix}
    T \\
    1 \\
    \vdots \\
    K \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} (K-i)! \\
\end{bmatrix} \sum_{i=0}^{K-1} \frac{T^i}{i!} (K-i)! \\
\begin{bmatrix}
    0^0 0^1 0^2 \ldots 0^K \\
    1^0 1^1 1^2 \ldots 1^K \\
    2^0 2^1 2^2 \ldots 2^K \\
    \vdots \\
    K^0 K^1 K^2 \ldots K^K \\
\end{bmatrix} = K + 1.
\]

Since \( C_{\Phi,G} \) defined in (33) is rank-sufficiency, this lemma stands.

Now, it is sufficient to give the theorem below to guarantee the reliability of our TVLAP-KF.

Theorem 2: For any given norm-finite \( \hat{X}_{0,0} \), if \( \Phi, G \) and \( R \) are bounded, \( \Phi, H \) is uniformly completely observable, and \( \Phi, G \) is uniformly completely controllable, then

\[
\hat{X}_{n,n} \rightarrow X_n, \text{as } n \rightarrow \infty,
\]

meaning

\[
p^{(k)}(n) \rightarrow p^{(k)}(n), \text{as } n \rightarrow \infty, \forall k = 0, 1, 2, \ldots, K.
\]

Proof: According to [64], [65], [66] (see chapter 4.4), with support of our Lemma 3 and Lemma 4 this theorem holds. Note that

\[
\text{rank}(O_{\Phi,H}) = \text{rank}(O_{\Phi,H} R^{-1/2}),
\]
and
\[ \text{rank}(C_{\Phi,G}) = \text{rank}(C_{\Phi,GQ^{-1/2}}), \]
(40)
where \( R^{-1/2}(R^{-1/2}) = R \) and \( Q^{-1/2}(Q^{-1/2}) = Q \). Since \( R \) and \( Q \) are positive definite, the decomposition is possible. In above, \( O_{\Phi,H} \) denotes the observability matrix defined by the pair \( [\Phi,H] \). The notation conventions keep same to \( C_{\Phi,G}, O_{\Phi,H} R^{-1/2} \) and \( C_{\Phi,GQ^{-1/2}} \).

E. Divergence Risk of TVLAP-KF and The Solutions

Although Theorem 2 asserts the sufficiency of the proposed TVLAP-KF, we should mention here that our TVLAP-KF model may still suffer from the divergence risk. This is because the real trend of \( x(n) \) is \( f(n) \), not \( p(n) \). From the viewpoint of Kalman filter, the true system dynamics is \( f(n) \) while we use \( p(n) \) to approximate it and filter. Thus there must be a underlying divergence risk.\(^{7,8,9}\) Facing this problem of modelling error and/or unknown disturbances, the adaptive Kalman filter\(^{10}\), Multiple model methods\(^{7}\), Kalman filter with unknown disturbances\(^{9,7}\) and/or unknown inputs\(^{7,9}\) are developed to bridge and/or eliminate the gap due to modelling error. Specifically, instead of studying\(^{11}\), we should focus on
\[
\begin{align*}
X(n+1) &= \Phi X(n) + MA(n) + GW(n) \\
Y(n) &= HX(n) + V(n),
\end{align*}
\]
(41)
where \( A(n) \) is unknown input driven by known matrix \( M \). The effort here should be on estimating \( X(n) \) in the presence of unknown \( A(n) \). Obviously, in \( A(n) \) is used to compensate or eliminate the modelling bias (modelling error).

Thus in this paper, the TVLAP-KF does not merely refer to the canonical Kalman filter, also includes any proper members in Kalman families.

F. Further Investigation on TVLAP model

As closing remarks, in this subsection, we should investigate more on our TVLAP model, especially the relationship between TVLAP and existing canonical methods. We have Remark 6 and Remark 7.

Remark 6: The Level model and Holt’s method (also known as Linear Trend model) mentioned in \(^{41}\) are special cases of TVLAP. When \( K = 0 \), TVLAP becomes the recursive-form Level model. If \( K = 1 \), TVLAP degenerates into the recursive-form Holt’s method.

Remark 7: In target tracking community\(^{72}\), one branch of signal processing problem, the canonical Static model, Constant Velocity (CV) model, and Constant Acceleration (CA) model are special cases of TVLAP. When \( K = 0 \), TVLAP gives the Static model. If \( K = 1 \), we have the CV model. If \( K = 2 \), TVLAP degenerates into the CA model.

G. Simulation Study

We in this subsection illustrate the practical validity of the presented TVLAP model using a simple scenario. Suppose \( t = 0 : 0.1 : 120 \) (thus \( T = 0.1 \)), \( x(n) = 5 \sin(0.1t) + WG(\text{length}(t)) \), and let \( K = 4 \), \( Q = \text{diag}(0,0,0,0.01^2) \), \( R = 1^2 \), where \( WG \) means the White Gaussian process with mean of zero and variance of 1, we then apply the Algorithm\(^{15,16,17}\) to Mean Estimating and Extrema Detecting, and have the simulation results in Fig. 2. Fig. 2 as well show a result of the 200-step ahead prediction. The prediction includes the mean forecasting and extrema forecasting. Note that we desire the low-frequency component of the raw series, since the measure is noised, we should trust more the system equation than the measurements. It is this reason that we have process variance be less than measurement variance. Note also that the choose of \( Q \) and \( R \) does not necessarily depend on the true variance of raw series \( x(n) \).\(^{23}\) For simplicity and with loss of mathematical strictness (see Remark 5), we do not provide the plot of the risk bound of the forecasting in Figure 2.

The satisfying prediction performances in Fig. 2 comes from the fact that the model we designed could preserve the high-order information of the changing pattern.

This experiment supports the validity of\(^{5}\). Although the part \( 5 \sin(0.1t) \) is a periodic component, which should be modeled as \( x_p(n) \) in model\(^{5}\). However, as discussed in Subsection II-B, it is as well proper to admit the existence of periodic trend, as long as we pay none attention to the very-long-term prediction.

H. Issue of Selecting the Model Order \( K \) and the Time Gap \( T \)

It is easy to see that the core of the TVLAP model is the matrix \( \Phi \) defined in \(^{16} \). It relates to the parameters \( K \) and \( T \), and the model performances depend much on the proper values of them.

- **Choosing** \( T \). For a typical time series, \( T = 1 \) in theory. However, in practice, the suggested value of \( T \) should be \( T \leq 1 \). Because the original series \( x(n) \) contains the noise (high-frequency) component, meaning the estimation error to derivatives would never be zero even though the true values are zero. Therefore, if we have \( T < 1 \), the impact introduced by the estimation errors to high-order derivatives could be weaken or eliminated. This is because the term \( T^k/k! \) will rapidly converges to zero if \( T < 1 \) as \( k \) increases.

- **Choosing** \( K \). In theory, for \( K \), the larger, the better. However, in practice, due to the exists of noise and Runge phenomenon in polynomial fitting, \( K \) should not be extremely large. The suggested value of \( K \) should be \( 2 \sim 8 \). \( K = 4 \) is a typical option. This is the experience of authors obtained in simulation studies. Note that if \( K = 0,1 \), the TVLAP model degenerated to Level model and Holt’s model, respectively. On the other hand, as stated in Lemma\(^{5} \) and Lemma\(^{4} \) the selected \( K \) should not be very large so that the observability and controllability matrices are non-singular.

It is possible that \( T \) may has explicit meaning in practice for a general time series, meaning we cannot assign value to it arbitrarily. For example, if the time series is the quantity of sale of apple in a store per day. The meaning of \( T \) should be 1 with unit of \( \text{Day} \). It seems inconsistent to our suggestion of \( T \leq 1 \). However, fortunately this is not an issue because \( f^{(n)} \) is a variable, meaning the difference could be compensated by
the estimate to \( f(n) \). Plus, the estimates to \( f(n) \) may become larger than their real values, advantaging that they become easier to observe.

V. UNDERSTANDING THE MOTION MODELLING

In the basis of Remark [7] let’s study the natural philosophy, specially the motion description (modelling) of an macroscopic object in physics (in particular, in Theoretical Mechanics). Suppose:

- \( s(n) := f(n) \) and \( f^{(k)}(n) := 0 \), \( \forall k \geq 1 \), we have \([1]\) as
  \[
  s(n + 1) = s(n);
  \]
- \( v(n) := f^{(1)}(n) \) and \( f^{(k)}(n) := 0 \), \( \forall k \geq 2 \), we have \([1]\) as
  \[
  s(n + 1) = s(n) + v(n)T;
  \]
- \( a(n) := f^{(2)}(n) \) and \( f^{(k)}(n) := 0 \), \( \forall k \geq 3 \), we have \([1]\) as
  \[
  s(n + 1) = s(n) + v(n)T + \frac{1}{2}a(n)T^2.
  \]

When we use \( s \) to denote the displacement, \( v \) the velocity, and \( a \) the acceleration of a moving object, we have the well-known kinematics equations \([42] \sim [44]\) in physics. Note that \( f^{(2)}(n) := 0 \), \( \forall k \geq 3 \) admits that \( a(n) := f^{(2)}(n) \) keeps constant over time. Thus we have Philosophy \([1]\) to disclose the nature of motion modelling. Before we have Philosophy \([1]\) let’s first have an axiom.

Axiom 1: The motion of an physical macroscopic object is continuous, meaning it is impossible to see a position change of an object at a time moment. Thus none time elapse denotes none position change.

Although the Axiom \([1]\) seems right intuitively and at least we cannot disprove it using any existing theoretical frames or experimental observations, we cannot assert it is absolutely right. Just as an axiom, let’s suppose it indeed holds. Because this is the premise of the discussion of our Philosophy \([1]\).

Philosophy 1: Before the emergence of humans, the phenomena of motion already exist. When we try to understand the natural law of the motion, most of us lost or started believing the existence of God (maybe the theism is correct, we do not hold standpoint here), until the great Newton was born. He found his Second Law and disclosed the relationship among the force, the mass and the acceleration. Further, he creatively used his concepts of integration and mathematically expressed the displacement with the velocity and the acceleration, which is thought to be the start of the Scientific Analysis and Natural Philosophy. Thus now, we should consider why Newton was right. What is underlying possibility of the correctness of this scientific frame? The viewpoints of this paper are as follows. According to Axiom \([1]\) the motion law of an object could be represented by \([1]\) with sufficient order \( K \) in any accuracy. The magic and interesting coincidence here is that the term \( f^{(2)}(n) \) is proportional to the externally exerted force. Thus we assign a Philosophical concept to \( f^{(2)}(n) \) and term as Acceleration. Subsequently, the other philosophical concepts like velocity and displacement were come up with and the relationships were bridged as \([42] \sim [44]\), if the higher-order terms of \([1]\) are omitted.

As complements to Philosophy \([1]\) we show Philosophy \([2]\) and Remark \([8]\) below.

Philosophy 2: Some other issues should be raised. Is the time-variant local autocorrelated polynomial model the unique one to explain the motion law of a moving object? Does it exist other approximation methods that are also plausible or interesting? Note that we do not know does not mean it does not exist. If exist, maybe we could have other philosophical concepts which are the conceptual counterparts of Acceleration and/or Velocity. For example, when we use Fourier series/transform to approximate the motion (changing pattern) of a signal \([58]\), or of a time series \([2]\), we have the concepts of Frequency rather than acceleration. The dominant difference here is whether and which parameters of the model we use could come across some physically existed concepts. For TVLAP, it is the pair of \( f^{(2)}(n) \) and acceleration. For Fourier’s, it is the pair of \( \frac{k}{\kappa T} \) and frequency. For more on \( \frac{k}{\kappa T} \), see \([2]\) and \([2]\). Note also that the mathematics exists before we can realize them.

Remark 8: The natural concern when we compare \([44]\) and \([1]\) is that does the truncation of the high-order parts, namely the terms with \( k \geq 3 \), matter? Should this be the insufficiency of the current theoretical frame of classical mechanics? Is this the underlying reason that Newton’s Second Law failed to explain some changing patterns of other entities, for instance, electron? Maybe this is an interesting problem to investigate. However, it is beyond the topic that this paper mainly concerns. Inspired readers could think more about this issue.
Interestingly and excitingly, the validity of above derivation could be supported by [24], [25]. In the mentioned literature of physics, beyond the concepts of Velocity and Acceleration, the concepts of Jerk, Snap, Crackle and Pop are assigned to the high derivatives of 3-order \( f^{(3)}(n) \), 4-order \( f^{(4)}(n) \), 5-order \( f^{(5)}(n) \) and 6-order \( f^{(6)}(n) \), respectively. The much higher order derivatives (orders larger than 6) have not been studied in physics, and we expect the advances in this direction.

VI. GENERAL METHODOLOGY FOR NON-GAUSSIAN NON-WHITE NOISE

Unfortunately, the Kalman filter is optimal only for white noise (we regard the non-trend part as noise for trend tracking problem). However, the Kalman filter is still to some degree effective to return a feasible solution. However, the good news is that there exists the exact method for colored noise Kalman filter. For briefless, we ignore the details here and invite the interested readers to refer to [38], [76], [77], [78]. Note that the philosophy/trick here is that we treat the interested time series as the real-time position of a moving object so that the time series analysis problem in this sense is a target tracking problem.

VII. CONCLUSION

This paper discusses the Time-Variant Local Autocorrelated Polynomial model with Kalman filter to handle the online non-stationary time series analysis and processing including the extrema detecting and forecasting problem, and trend tracking and forecasting problem. All the methods or methodologies presented in this paper are just complements to existing solutions like ARMA-SIN methodology, TBATS methodology, Box-Jenkins methodology, Box-Cox transformation, regression methods, machine learning methods, exponential smoothing method, moving average methods and so on, asserting no dominant position over other methods. Because although powerful our methods are in some application scenarios, we still admit they are confined within the realm of the no-free-lunch theorem that there exists no the unique-best solution to all of the problems raised in practice. For example, it seems no method could outperform the logarithmic transformation for Air-Passenger forecasting problem.

Additionally, the proposed model can also explain motion modelling in physics, which provides a new angle for understanding for this area. We anticipate our model could help scholars in physics find out more natures of motion.

REFERENCES

[1] S. K. Jensen, T. B. Pedersen, and C. Thomsen, “Time series management systems: A survey,” IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 11, pp. 2581–2600, 2017.
[2] S. Wang, C. Li, and A. Lim, “Why are the arima and sarima not sufficient,” arXiv preprint arXiv:1904.07632, 2019.
[3] A. R. Ives, K. C. Abbott, and N. L. Ziebarth, “Analysis of ecological time series with arima (p, q) models,” Ecology, vol. 91, no. 3, pp. 858–871, 2010.
[4] A. C. Harvey and P. Todd, “Forecasting economic time series with structural and box-jenkins models: A case study,” Journal of Business & Economic Statistics, vol. 1, no. 4, pp. 299–307, 1983.
[5] C. Li and A. Lim, “A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing,” European Journal of Operational Research, vol. 269, no. 3, pp. 860–869, 2018.
[6] S. Wang, C. Li, and A. Lim, “Optimal joint estimation and identification theorem to linear gaussian system with unknown inputs,” Signal Processing, vol. 161, pp. 268–288, 2019.
[7] I. Okatanli and Y. I. Shepely, “Dynamic prediction of traffic volume through kalman filtering theory,” Transportation Research Part B: Methodological, vol. 18, no. 1, pp. 1–11, 1984.
[8] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control. John Wiley & Sons, 2015.
[9] W. Liu, J. C. Principe, and S. Haykin, Kernel adaptive filtering: a comprehensive introduction. John Wiley & Sons, 2011, vol. 57.
[10] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media, 2008.
[11] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018.
[12] “Autoregressive conditional heteroskedasticity,” https://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedasticity, accessed on Sep 4, 2019.
[13] R. F. Engle, “Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation,” Econometrica: Journal of the Econometric Society, pp. 987–1007, 1982.
[14] C. Brooks, Introductory econometrics for finance. Cambridge university press, 2019.
[15] J. D. Hamilton, “Time series analysis,” Economic Theory, II, Princeton University Press, USA, pp. 625–630, 1995.
[16] A. Papoulis and S. U. Pillai, Probability, random variables, and stochastic processes. Tata McGraw-Hill Education, 2002.
[17] S. L. Marple Jr and W. M. Carey, “Digital spectral analysis with applications,” 1989.
[18] J. Cadzow, “High performance spectral estimation—a new arma method,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 5, pp. 524–529, 1980.
[19] B. Friedlander and B. Porat, “The modified yule-walker method of arma spectral estimation,” IEEE Transactions on Aerospace and Electronic Systems, no. 2, pp. 158–173, 1984.
[20] H. Liu, L. Zhu, Z. Pan, F. Bai, Y. Liu, Y. Liu, M. Patel, E. Farrantos, and N. Bhatt, “Arms-based transfer function model identification using wide-area measurement for adaptive and coordinated damping control,” IEEE Transactions on Smart Grid, vol. 8, no. 3, pp. 1105–1115, 2017.
[21] R. N. Calheiros, E. Masoumi, R. Ranjan, and R. Buyya, “Workload prediction using arima model and its impact on cloud applications qos,” IEEE Transactions on Cloud Computing, vol. 3, no. 4, pp. 449–458, 2015.
[22] K. Yunus, T. Thiringer, and P. Chen, “Arima-based frequency-decomposed modeling of wind speed time series,” IEEE Transactions on Power Systems, vol. 31, no. 4, pp. 2546–2556, 2016.
[23] C. Yu, K. You, and L. Xie, “Quantized identification of arma systems with colored measurement noise,” Automatica, vol. 66, pp. 101–108, 2016.
[24] F. Ding, L. Xu, F. E. Alsaadi, and T. Hayat, “Iterative parameter identification for pseudo-linear systems with arma noise using the filtering technique,” IET Control Theory & Applications, vol. 12, no. 7, pp. 892–899, 2018.
[25] K. Gilbert, “An arma supply chain model,” Management Science, vol. 51, no. 2, pp. 305–310, 2005.
[26] M. M. Ali, M. Z. Babai, J. E. Boylan, and A. A. Syntetos, “Supply chain forecasting when information is not shared,” European Journal of Operational Research, vol. 260, no. 3, pp. 984–994, 2017.
[27] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpening, “Stil: A seasonal-trend decomposition,” Journal of official statistics, vol. 6, no. 1, pp. 3–73, 1990.
[28] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” Journal of econometrics, vol. 31, no. 3, pp. 307–327, 1986.
[29] P. Posedel, “Analysis of the exchange rate and pricing foreign currency options on the croatian market: the nargarch model as an alternative to the black-scholes model,” Financial theory and practice, vol. 30, no. 4, pp. 347–368, 2006.
[30] D. Li, X. Zhang, K. Zhu, and S. Ling, “The zd-garch model: A new way to study heteroscedasticity,” Journal of Econometrics, vol. 202, no. 1, pp. 1–17, 2018.
[31] P. Otto, W. Schmid, and R. Garthoff, “Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity,” Spatial Statistics, vol. 26, pp. 125–145, 2018.
[49] X. Wang, Y. Liang, Q. Pan, C. Zhao, and F. Yang, “Nonlinear gaussian
[43] S. Sarkka and J. Hartikainen, “On gaussian optimal smoothing of non-
[39] K. P. B. Chandra, D.-W. Gu, and I. Postlethwaite, “Square root cubature
[38] D. Simon, Optimal state estimation: Kalman, H infinity, and nonlinear
[34] P. Bloomfield, Fourier analysis of time series: an introduction
[32] C. W. J. Granger and M. Hatanaka, Spectral Analysis of Economic Time
[31] W. Liu and M. Farooq, “An arma model based scheme for maneuvering
[30] D. Eager, A.-M. Pendrill, and N. Reistad, “Beyond velocity and acceler-
[29] T.-c. Li, J.-y. Su, W. Liu, and J. M. Corchado, “Approximate gaussian
covariance: parametric recursive filtering under nonlinearity, multimodal-
[28] S. Siami-Namini and A. S. Namin, “Forecasting economics and financial
[27] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time
[26] S. R. Li and V. P. Jilkov, “Survey of maneuvering target tracking, part v.
multiple-model methods,” IEEE Transactions on Aerospace and
Electronic Systems, vol. 41, no. 4, pp. 1255–1321, 2005.
[25] J. Durbin and S. J. Koopman, Time series analysis by state space
methods,” Oxford university press, 2012.
[24] X. Wang, Y. Liang, Q. Pan, C. Zhao, and F. Yang, “Nonlinear gaussian
smoothers with colored measurement noise,” IEEE Transactions on
Automatic Control, vol. 60, no. 3, pp. 870–876, 2015.
[23] S. Sarkka and J. Hartikainen, “On gaussian optimal smoothing of non-
linear state space models,” IEEE Transactions on Automatic Control,
vol. 55, no. 8, pp. 1935–1941, 2010.
[22] H. Geng, Z. Wang, Y. Cheng, F. E. Alsaadi, and A. M. Doabaie, “State
estimation under non-gaussian levy and time-correlated additive
detector noise: A modified tobit kalman filtering approach,” Signal
Processing, vol. 154, pp. 120–129, 2018.
[21] Y. Liang, D. H. Zhou, Q. Pan et al., “A finite-horizon adaptive kalman
filter for linear systems with unknown disturbances,” Signal Processing,
vol. 84, no. 11, pp. 2175–2194, 2004.
[20] W. Ma, J. Duan, W. Man, H. Zhao, and B. Chen, “Robust kernel
adaptive filters based on mean p-power error for noisy chaotic time
series prediction,” Engineering Applications of Artificial Intelligence,
v. 58, pp. 101–110, 2017.
[19] C. Richard, J. C. M. Bermudez, and P.HONEINE, “Online prediction of
time series data with kernels,” IEEE Transactions on Signal Processing,
v. 57, no. 3, pp. 1058–1067, 2009.
[18] M. Han, S. Zhang, M. Xu, T. Qiu, and N. Wang, “Multivariate chaotic
time series online prediction based on improved kernel recursive least
squares algorithm,” IEEE transactions on cybernetics, no. 99, pp. 1–13,
2018.
[17] Q. Hua, M. Wen-Tao, Z. Ji-Hong, and C. Ba-Dong, “Kernel least mean
kurtosis based online chaotic time series prediction,” Chinese Physics
Letters, vol. 30, no. 11, p. 110505, 2013.
[16] Y. Zhu, R. Dai, G. Liu, Z. Wang, and S. Lu, “Power market price
forecasting via deep learning,” in IECON 2018-44th Annual Conference
of the IEEE Industrial Electronics Society. IEEE, 2018, pp. 4935–4939.
[15] S. Liao, J. Chen, J. Hou, Q. Xiong, and J. Wen, “Deep convolutional
neural networks with random subspace learning for short-term traffic
flow prediction with incomplete data,” in 2018 International Joint
Conference on Neural Networks (IJCNN). IEEE, 2018, pp. 1–6.
[14] S. Siami-Namini and A. S. Namin, “Forecasting economics and financial
time series: Arima vs. Istm,” arXiv preprint arXiv:1803.06386, 2018.
[13] D. W. Lu, “Agent inspired trading using recurrent reinforcement learning
and lstm neural networks,” arXiv preprint arXiv:1707.07338, 2017.
[12] M. De Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time
series with complex seasonal patterns using exponential smoothing,”
Journal of the American Statistical Association, vol. 106, no. 496, pp.
1513–1527, 2011.
[11] M. Mudelsee, Climate time series analysis. Springer, 2013.
[10] K. Doya, K. Samejima, K.-i. Katagiri, and M. Kawato, “Multiple model-
based reinforcement learning,” Neural computation, vol. 14, no. 6, pp.
1347–1369, 2002.
[9] R. Murray-Smith and T. Johansen, Multiple model approaches to non-
linear modelling and control, CRC press, 1997.
[8] L. Chaparro and A. Akan, Signals and Systems using MATLAB. Aca-
demic Press, 2018.
[7] E. Kreyszig, Introductory functional analysis with applications. wiley
New York, 1978, vol. 1.
[6] K. Myers and B. Tlapley, “Adaptive sequential estimation with unknown
noise statistics,” IEEE Transactions on Automatic Control, vol. 21, no. 4,
pp. 520–523, 1976.
[5] M. Mudelsee, Climate time series analysis. Springer, 2013.
[4] D. W. Lu, “Agent inspired trading using recurrent reinforcement learning
and Istm neural networks,” arXiv preprint arXiv:1707.07338, 2017.
[3] M. De Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time
series with complex seasonal patterns using exponential smoothing,”
Journal of the American Statistical Association, vol. 106, no. 496, pp.
1513–1527, 2011.
[2] S. Siami-Namini and A. S. Namin, “Forecasting economics and financial
time series: Arima vs. Istm,” arXiv preprint arXiv:1803.06386, 2018.