Mixture Factorized Ornstein-Uhlenbeck Processes for Time-Series Forecasting

Guo-Jun Qi
University of Central Florida
Orlando, Florida 32816
guojun.qi@ucf.edu

Jiliang Tang
Michigan State University
East Lansing, Michigan 48824
tangjili@msu.edu

Jingdong Wang
Microsoft Research Asia and Hefei University of Technology
Beijing, Beijing, China 100080
jingdw@microsoft.com

Jiebo Luo
University of Rochester
Rochester, New York 14627
jluo@cs.rochester.edu

ABSTRACT

Forecasting the future observations of time-series data can be performed by modeling the trend and fluctuations from the observed data. Many classical time-series analysis models like Autoregressive model (AR) and its variants have been developed to achieve such forecasting ability. While they are often based on the white noise assumption to model the data fluctuations, a more general Brownian motion has been adopted that results in Ornstein-Uhlenbeck (OU) process. The OU process has gained huge successes in predicting the future observations over many genres of time series, however, it is still limited in modeling simple diffusion dynamics driven by a single persistent factor that never evolves over time. However, in many real problems, a mixture of hidden factors are usually present, and when and how frequently they appear or disappear are unknown ahead of time. This imposes a challenge that inspires us to develop a Mixture Factorized OU process (MFOUP) to model evolving factors. The new model is able to capture the changing states of multiple mixed hidden factors, from which we can infer their roles in driving the movements of time series. We conduct experiments on three forecasting problems, covering sensor and market data streams. The results show its competitive performance on predicting future observations and capturing evolution patterns of hidden factors as compared with the other algorithms.

KEYWORDS

Mixture Factorized Ornstein-Uhlenbeck Process (MFOUP)

KDD'17, August 13–17, 2017, Halifax, NS, Canada.

ISBN 978-1-4503-4887-4/17/08. . . $15.00
DOI: http://dx.doi.org/10.1145/3097983.3098113

1 INTRODUCTION

The ability of forecasting the future values of time series [8, 10, 16] plays an important role in revealing the trend and fluctuation of dynamic data for many applications ranging from sensor networks [12], to financial and market analysis [5, 13] and beyond. It is usually performed by establishing a regression model that predict the future observations of a target data from its past values [35]. When building a forecasting model, our goal is not only to predict the future values of a target tim series, but also to reveal the its moving and fluctuation patterns. This can provide us deep insight into the internal schemes and driving factors that cause the dynamic change of time series data. For example, in analyzing the financial data, we wish to reveal the market volatility and predict stock prices from the past market data [20, 22, 32, 33]; for a sensor network system, we want to predict the future readings of a sensor by capturing its past trend and dynamic patterns of data fluctuations [7, 26, 34].

Among the time-series forecasting models [8], AutoRegression (AR) [15] is one of the most popular models in many applications. It explores the stationarity of a time series, and uses a linear model to forecast the future observations from the past inputs. The stationary assumption makes it possible to find time-invariant correlations between the observations at different time points, which enables the optimal predictions of future values. A non-stationary time series can be its stationary counterpart by differencing the consecutive observations in an input data stream, a standard technique to obtain a stationary time series [15]. There also exist variant AR variants that directly capture non-stationary properties by modeling time-varying prediction coefficients [14], evolving states [27] and switching autoregressive processes [21].

In spite of many applications, AR model is a discrete-time forecasting model, which uses IID (*independent identically distributed*) Gaussian distributions to simulate the white noises that drive the diffusion of target data [15]. In real world, a better choice of modeling the data diffusion is to use Brownian motion [17, 19] to characterize the uncertainty underlying the dynamic data. Unfortunately, a real Brownian motion has to be defined and analyzed in continuous-time domain rather than over a discrete set of time points. This leads to Ornstein-Uhlenbeck (OU) process [4, 24], which uses Stochastic Differential Equations (SDE) [25] to characterize the fluctuation of a time series. Formally, it uses the following SDE to model a multivariate time series $\{X_t\}$

$$dX_t = -\Psi (X_t - \mu) \, dt + \sigma dB_t \quad (1)$$

where $\{B_t\}$ is a Brownian motion process, σ is its strength, Ψ is a transition matrix modeling how the time series drift, and μ is its mean. The existence and uniqueness of the solution to this SDE have been established in the framework of the celebrated Ito calculus [19].

The conventional OU processes have been successfully applied to model many genres of time-series data, from stock prices to sensor streams. They are restricted to the use of a simple Brownian motion that is assumed to constantly drive the movement of dynamic data [18]. However, in many real scenarios, there often exist mixture of evolving factors over time, and there is no single one that would persistently drive the time series [1, 28, 31] through time. For example, in a sensor network used to monitor a mechanical machine, its readings might be affected by many unexpected factors such as external inferences and the operational dysfunction. These factors do not always persist through time. Instead, they would appear or disappear occasionally to affect the data movements and fluctuations. This inspires us to use a group of mixture Brownian motions to reveal the presence and the frequency of those hidden factors underlying time series.

For this purpose, we will develop a Mixture Factorized OU Process (MFOUP) in this paper. It can capture multiple unknown factors jointly to reveal their mixture effects. The evolution of these factors can be explicitly inferred to provide deep insight into how they affect the data fluctuations in the observed time series. Moreover, we will show that the MFOUP is able to capture the stationary time-series property that do not change over time, enabling it to generalize the time-invariant property as to predict future data values.

On the other hand, once we infer the hidden factors and fix them, the time series modeled by the MFOUP does not have to be stationary anymore. This gives it sufficient flexibility to handle the time-varying property driven by the evolving factors. We will reveal these statistic properties of the M-FOUP, and explore them to estimate the model parameters from the observed series. Through the experiments on three different time-series forecasting tasks, we will demonstrate the competitive performance of the MFOUP on modeling data fluctuations driven by evolving hidden factors.

The remainder of this paper is organized as follows. In Section 2, we briefly review the relevant work on time-series forecasting. We present the proposed Mixture Factorized Ornstein-Uhlenbeck Processes (MFOUPs) in Section 3, followed by an analysis of the property in Section 4. The computational method to estimate the model parameters is p-resented in Section 5. We report the experiment results on three time-series data sets in Section 6. Finally, we conclude the paper in Section 7.

2 RELATED WORK

There exist many time-series forecasting models in literature. As far as the time domain of the time series is concerned, a model belongs to one of two categories – either a discrete or continuous time-domain model.

The first category of models work on discrete-time domains [8]. In this case, a time series becomes a sequence sampled from a continuous-time data stream. For example, Autoregressive (AR) model [15] predicts the future values of a sequence by a linear regression of its past observations. Meanwhile, a sequence of IID Gaussian random variables are adopted in the AR model to simulate the fluctuations of the data values. The AR model can be used to generate stationary sequences, whose statical property does not change over time. It is the stationarity that guarantees the predicability of the model by generalizing time-variant property into the future values of a sequence. In addition, the Moving Averaging (MA) and the extended Autoregressive Moving Averaging (ARMA) models [9] can also generate stationary sequences which have extended the AR model by creating series of averages over a sliding time window.

Given a non-stationary sequence, a standard technique is to take a difference between consecutive observations which can convert it into a stationary counterpart [15]. Otherwise, time-dependent forecasting models have been developed to capture the non-stationary property. For example, nonstationary sequences can be modeled through a time-dependent autoregressive model by the use of a limited series expansion of the time-varying coefficients [14]; in the presence of regime shift in time series data, a general switching vector autoregressive processes are proposed to predict multiple time series subject to Markovian shifts in the regime [21]; recently, the State-Driven Dynamic Prediction (SDDP) model also extends the ordinary autoregressive model which introduces the prediction matrices as a function of evolving states over time [27].

In parallel with the discrete-time forecasting models, continuous-time models are often used to characterize time series on a continuous time domain. Discrete-time models can be seen as a special case of continuous-time counterparts by sampling an input time series on an evenly or unevenly spaced grid of time points. The continuous-time models are usually built on the celebrated Ito calculus [19], which integrates the movement of time series characterized by Brownian motion-s. Among the most representative continuous-time models are Ornstein-Uhlenbeck processes and its generalized forms based on Levy process [4, 24, 30]. They usually assume that the fluctuations of data values in a time series is driven by a persistent driving process, such as Brownian and Levy motions, whose presence does not change as time goes by. This is an unrealistic assumption since there often exist multiple hidden factors whose appearances or disappearances would occur through time, and their mixture effects dynamically characterize the movements of time-series values.

3 FORMULATION

In this paper, we wish to develop a novel paradigm of Ornstein-Uhlenbeck (OU) process driven by multiple hidden factors evolving over time. An ordinary OU process extends an Auto-Regressive (AR) model into continuous-time domain. The idea of OU model is to use Brownian motion to characterize the uncertainty underlying a time series. It leads to an Ito stochastic differential equation, whose solution integrates both system-deterministic transitions as well as stochastic fluctuations over time. It has gained huge successes in modeling sensor and financial data streams, since the Brownian motion provides a more precise description of dynamic uncertainty.

However, the conventional OU model does not take into consideration multiple mixed factors that jointly lead to the dynamics and fluctuations underlying time-series data. For example, for a sensor network, there are many unknown factors that would affect how a physical process evolves over time, which can be the unknown interference sources and unexpected system conditions. In a financial market, the volatility of stock prices reflect many economic and political factors which can occur simultaneously. What is more challenging is these hidden factors are not static; instead their appearance and disappearance are dynamically changing over time. In this paper, we attempt to reveal these hidden factors and characterize their dynamic evolution over time. In other words, we do not assume these factors are not always present, and as time goes by, they can dynamically affect the evolution of time series.

Formally, suppose we have a d-dimensional time series $X_t \in \mathbb{R}^d$ which is described as a stochastic process defined over probability space. Then we use the following Ito stochastic differential equation to describe the dynamics over this stochastic process.

$$dX_t = -\Psi\left(X_t - \mu\right)dt + \sum_{k=1}^{K} z_t^k \sigma^k dB_t^k \tag{2}$$

We call this stochastic differential equation the Mixture Factorized OU Process (MFOUP). It models K sequences of hidden factors that derive the fluctuations of a process. Specifically, we use K Factor Indication Processes (FIP) $\{z_t^k | z_t^k \in \{0,1\}\}$, $k = 1, \cdots, K$ to model if each factor k becomes active in driving the evolution of the time-series data.

Below we will explain different components of this basic equation in detail.

(1) Deterministic Evolution of Time-Series Data

First, $\Psi \in \mathbb{R}^{d\times d}$ and $\mu \in \mathbb{R}^d$ are two parameters characterizing the deterministic component of this process. In particular, the first term of the right hand side of the MFOUP equation models how the time series X_t would evolve around a mean vector μ with dynamic transition modeled by Ψ in the derivative form. Later on, we will show that Ψ drives the transition of X_t from time to time.

(2) Hidden Factors by Multiple Brownian Motions

We use K independent Brownian motions to model the K different hidden factors that drive the evolution of the observed process X_t, where σ^k is the strength for k-th Brownian motion.

In a sensor network, these factors represent multiple interference sources or uncertain statuses occurring in the operation of a system; in a financial market, they can correspond to several evolving factors driving the market volatile in asset prices. The Brownian motions are continuous over time, but almost non-differentiable everywhere. The differential "d" in the equation is not taken in common calculus sense. Instead, it represents continuous-time differential of the Brownian processes in the Ito calculus framework. This results in an ideal model to characterize a mixture of unknown stochastic factors in many applications.

(3) Continuous-time Factor Indication Processes

Suppose that for each factor k, $\mathcal{Z}^k = \{z_t^k\}$ is a continuous-time Markov process [3] defined on a binary set $z_t^k \in \{0,1\}$. Its value represents whether the state of kth hidden factor is active (when $z_t^k = 1$) or not (when $z_t^k = 0$) at time t.

Such a continuous-time Markov process is modeled by an infinitesimal generator between these two states, which is defined by the matrix $Q^k = \begin{bmatrix} -\alpha^k & \alpha^k \\ \beta^k & -\beta^k \end{bmatrix}$, each entry measuring how frequently the two states transit between each other in an infinitesimal time interval.

Specifically, we can obtain the transition matrix by solving the following ordinary differential equation

$$\frac{dP^k(\tau)}{d\tau} = Q^k P^k(\tau), P^k(0) = I_{2\times 2}$$

from which we can solve the transition matrix

$$P^k(\tau) = \exp(\tau Q) = \begin{bmatrix} \frac{\beta^k}{\alpha^k+\beta^k} + \frac{\alpha^k e^{-(\alpha^k+\beta^k)\tau}}{\alpha^k+\beta^k} & \frac{\alpha^k}{\alpha^k+\beta^k} - \frac{\alpha^k e^{-(\alpha^k+\beta^k)\tau}}{\alpha^k+\beta^k} \\ \frac{\beta^k}{\alpha^k+\beta^k} - \frac{\beta^k e^{-(\alpha^k+\beta^k)\tau}}{\alpha^k+\beta^k} & \frac{\alpha^k}{\alpha^k+\beta^k} + \frac{\beta^k e^{-(\alpha^k+\beta^k)\tau}}{\alpha^k+\beta^k} \end{bmatrix},$$

each entry of which defines the transition probability $P_{ij}^k(\tau) = P(z_{t+\tau}^k = j | z_t^k = i), i, j = 0, 1$ over a time interval τ between two states i and j for factor k.

Let $\tau \to +\infty$, the transition probability will converge, showing the existence of a stationary distribution for z_t. We will make use of this property to prove the existence of a stationary solution to the stochastic equation.

4 PROPERTY OF MFOUP

In this section, we will prove the existence of a stationary process defined by the MFOUP equation. This will shed light on how MFOUP models the observations with multiple factors.

To solve the MFOUP equation (2), we apply the Ito formula to $e^{\Psi t}(X_t - \mu)$, which results in

$$d\left(e^{\Psi t}\left(X_t - \mu\right)\right) = e^{\Psi t}\Psi\left(X_t - \mu\right)dt + e^{\Psi t}dX_t$$

Substituting Eq. (2) in the above equation, we have

$$d\left(e^{\Psi t}(X_t-\mu)\right)=\sum_{k=1}^{K} z_t^k e^{\Psi t}\sigma^k dB_t^k$$

Then taking Ito integral of both sides over an interval $[s,t]$, we obtain

$$e^{\Psi t}(X_t-\mu)=e^{\Psi s}(X_s-\mu)+\sum_{k=1}^{K}\int_s^t z_u^k e^{\Psi u}\sigma^k dB_u^k$$

Equivalently,

$$X_t=e^{-\Psi(t-s)}X_s+\left(I-e^{-\Psi(t-s)}\right)\mu+\sum_{k=1}^{K}\int_s^t z_u^k e^{\Psi(u-t)}\sigma^k dB_u^k \tag{3}$$

From this result, we can get the following theorem which establishes the existence of MFOUP solution as well as some basic properties.

Theorem 1 (Existence of MFOUP Solution). *There exists a covariance stationary solution X_t to MFOUP equation (2) if and only if all the eigenvalues of Ψ only have negative real parts, i.e.,* $\mathrm{Re}\lambda_i(\Psi)>0$ *for* $i=1,\cdots,d$. *Then the MFOUP solution becomes*

$$X_t=\mu+\sum_{k=1}^{K}\int_{-\infty}^{t} z_u^k e^{\Psi(u-t)}\sigma^k dB_u^k \tag{4}$$

which has the distribution of $\mu+\sum_{k=1}^{K}\int_0^{+\infty} z_u^k e^{-\Psi u}\sigma^k dB_u^k$.

Proof. Let $s\to-\infty$ in Equation (3). Then if all the eigenvalues of Ψ only have strictly positive real parts, the terms involving $e^{-\Psi(t-s)}$ will vanish for a fixed t. This leads to the result (4) since the sum of the covariance matrices of the terms on the right of Equation (3) is bounded. Thus, X_t given in Equation (4) has the distribution of $\mu+\sum_{k=1}^{K}\int_0^{+\infty} z_u^k e^{-\Psi u}\sigma^k dB_u^k$, if we replace the variable u in the integral with $u+t$, and consider the stationarity of the continuous time Markov processes $\{z_t^k\}$ for all k's and increment stationarity of Brownian motions [17]. □

The stationary solution is able to capture the time-invariant property that can be generalized as to predict the future values of a time series. In this sense, the existence of the stationary solution to the MFOUP justifies it as a valid model for the time-series forecasting problem with a mixture of evolving factors. Without loss of generality, for a non-stationary time series, we can adopt differencing of consecutive observations to get a stationary sequence [15].

It is also worth noting that the stationarity of the MFOUP is proved jointly on the target time series $\{X_t\}$ and the associated FIPs $\{z_t^k\}$. However, when $\{z_t^k\}$ are sampled or fixed, the conditional process $\{X_t\}$ might not be stationary anymore. In this sense, it also enables the MFOUP with ability of modeling time-varying property determined by the evolving factors. The basic lemma presented in the next section will explore such property and show how the model parameters can be estimated based on the lemma.

5 MODEL ESTIMATION

In this section, we discuss how to estimate the model parameters $\Theta=\{\Psi,\mu,\sigma^k,\alpha^k,\beta^k\}$ from the observations X_t. Usually, the observations are made at integer time clicks, i.e., we have observations $\{X_n=X_{\delta n}|n=1,\cdots N\}$ on evenly spaced sampling time points with an interval of δ. For example, in the stock market, we have the daily opening/close prices; in the sensor networks, the sensor readings are recorded when the clock clicks. In the rest of the paper, we will use X_n and z_n to denote $X_{\delta n}$ and $z_{\delta n}$ for brevity.

First, in the next subsection, we prove a lemma which connects the observations with the model parameters. This lemma plays a key role in estimating the model parameters.

Then, the model estimation proceeds in two alternating steps in an Expectation-Maximization (EM) fashion. In Maximization step Section 5.2, maximum likelihood estimation will be derived based on the discretized samples of FIP $\{z_n\}$; with the fixed model parameters, the sampling of $\{z_n\}$ is performed to estimate the time-evolving hidden factors. Finally, we summarize the model estimation algorithm in Section 5.4.

5.1 Basic Lemma

Based on Eq. (3), consider the observations X_n, and X_{n-1} at two successive time steps n and $n-1$, we can obtain the following model

$$X_n=e^{-\Psi\delta}X_{n-1}+\left(I-e^{-\Psi\delta}\right)\mu+\varepsilon_n \tag{5}$$

where the last term,

$$\varepsilon_n=\sum_{k=1}^{K}\int_{(n-1)\delta}^{n\delta} z_u^k e^{\Psi(u-n\delta)}\sigma^k dB_u^k, \tag{6}$$

characterizes the stochastic evolution of X_t driven by the K hidden factors, in addition to the first two terms decided by the deterministic component (Ψ,μ).

The property of ε_n is stated in the following lemma.

Lemma 1. *Conditioned on a sequence of* $\mathcal{Z}=\{z_n^k|n=1,\cdots,N,k=1,\cdots,K\}$, $\{\varepsilon_n\}$ *is an indepndent sequence with zero mean and covariance matrix* $\mathrm{cov}(\varepsilon_n|\mathcal{Z})$

$$\begin{aligned}\Omega_n^{\mathcal{Z}} &\triangleq \mathrm{cov}(\varepsilon_n|\mathcal{Z})\\ &=\sum_{k=1}^{K}\frac{\alpha^k}{\alpha^k+\beta^k}\left(\sigma^k\right)^2(\Psi+\Psi^T)^{-1}\left(I-e^{-(\Psi+\Psi^T)\delta}\right)\\ &+\sum_{k=1}^{K}\frac{(1-z_{n-1}^k)\alpha^k-\beta^k z_{n-1}^k}{\alpha^k+\beta^k}\left(\sigma^k\right)^2\\ &\times(\Psi+\Psi^T-(\alpha^k+\beta^k)I)^{-1}\left(e^{-(\Psi+\Psi^T)\delta}-e^{-(\alpha^k+\beta^k)\delta}I\right)\end{aligned}$$

The proof of this lemma applies the increments of Brownian motion and the Isometry property of the Ito integral. We give its proof in Appendix A.

This lemma characterizes the statics of the remainder term ε_n, which will play an important role in the model parameter estimation of the following section.

5.2 Maximum Likelihood Estimation with Laplace Approximation

According to the basic lemma and Eq. (5), we can conclude that given X_{n-1} and the sampled FIP $\mathcal{Z}$, the conditional mean M_n of X_n can be expressed as

$$M_n = \mathbb{E}(X_n|X_{n-1}, \mathcal{Z}) = e^{-\Psi\delta}X_{n-1} + \left(I - e^{-\Psi\delta}\right)\mu$$

and its covariance is exactly $\Omega_n^{\mathcal{Z}}$ as stated in the basic lemma.

Unfortunately, in general, X_n does not follow a multivariate normal distribution, if we marginalize the FIP z_u^k over the interval $[(n-1)\delta, n\delta]$ in Eq. (6). However, we can use a normal distribution with the same mean M_n and covariance $\Omega_n^{\mathcal{Z}}$ to approximate the true conditional distribution of X_{n-1}. This is known as Laplace approximation in literature [6, 29], which uses the first two orders of moments to approximate an intractable distribution.

Therefore, this Laplace approximate leads to the following problem of log-likelihood maximization over a sequence of observations

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{n=1}^{N}\log P(X_n|X_{n-1}, \mathcal{Z}) + \sum_{n,k=1}^{N,K}\log P(z_n^k|z_{n-1}^k)\\ &= \sum_{n=1}^{N}\log\mathcal{N}(M_n, \Omega_n^{\mathcal{Z}}) + \sum_{n,k=1}^{N,K}\log P(z_n^k|z_{n-1}^k)\end{aligned} \tag{7}$$

where $P(z_n^k|z_{n-1}^k)$ is the transition probability for FIP. This objective is differentiable w.r.t. model parameters. Any standard optimization algorithm can be adopted to solve these parameters. In this paper, we adopt the stochastic gradient method, which updates the model parameters with their gradients computed over each pair of observations and sampled factors (X_n, z_n) and (X_{n-1}, z_{n-1}). The advantage of stochastic gradient method is its update can be performed efficiently in an online fashion by scanning through the observations and factors over the sequence. Even more, the model parameters can be initialized with the result from the last likelihood maximization step since they are gradually updated with a new set of sampled IFP $\mathcal{Z}$ in each step.

Next, we present how to sample the FIP sequences $\mathcal{Z}$ in the following section.

5.3 Sampling of Hidden Factors

As illustrated in Figure 3, the sampling of FIP $\mathcal{Z}$ can be performed by embedding them into K chains of random processes. For each chain, the transition matrix for a factor k is given by the Continuous-Time Markov Process over an interval of δ as $P\left(z_n^k|z_{n-1}^k\right) = \exp\left(\delta Q^k\right)$, and the corresponding emission distribution is given by the conditional distribution $P\left(X_n|X_{n-1}, z_n^{1:K}, z_{n-1}^{1:K}\right) = \mathcal{N}(M_n, \Omega_n)$ derived in the last section.

Then $\mathcal{Z}$ is sampled over these K chains according to Gibbs Sampling approach [11] from the following distribution

$$\prod_{n=0}^{N-1}\left[P\left(X_{n+1}|X_n, z_n^{1:K}\right)\prod_{k=1}^{K}P\left(z_{n+1}^k|z_n^k\right)\right]$$

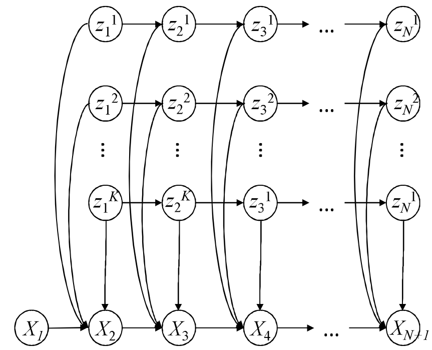


Figure 1: Dependency between the K FIP sequences and the observed time series $\{X_n\}$. The hidden factors are sampled by embedding FIP sequences into K chains of random processes and the Gibbs sampling approach is applied to sample them sequentially. Note that as shown in the $\Omega_n^{\mathcal{Z}}$ of the basic lemma, an observation X_n depends on the previous factors $\{z_{n-1}^k|k = 1, \cdots, K\}$, since we only access the past knowledge of the hidden factors to infer X_n for each step.

Algorithm 1 Estimating Model Parameter

---

input an input sequence $\{X_n|n = 1, 2, \cdots, N\}$
Initialize the model parameter
// Iterate between M-Step, and E-Step.
E-Step Sample the sequence of the FIP $\mathcal{Z}$ according to Eq. (8);
M-Step Update the current model parameters by solving (7) with the sampled FIP $\mathcal{Z}$.

Precisely, for each z_n^k, we have

$$z_n^k \sim P\left(X_{n+1}|X_n, z_n^{-k}, z_n^k\right)P\left(z_n^k|z_{n-1}^k\right) \tag{8}$$

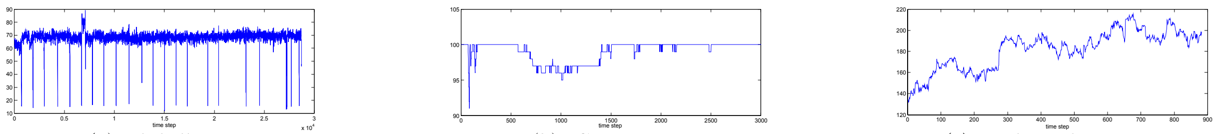
where z_n^{-k} denotes $z_n^{1:K}$ excluding z_n^k, $P\left(X_{n+1}|X_n, z_n^{-k}, z_n^k\right)$ is a normal distribution $\mathcal{N}(M_{n+1}, \Omega_{n+1}^{\mathcal{Z}})$ where the FIP samples at time n are set to $\mathcal{Z}_n = z_n^{-k} \cup \{z_n^k\}$, and $P\left(z_n^k|z_{n-1}^k\right)$ is the transition probability for the FIP of factor k.

5.4 Summary

In summary, the model parameter estimation proceeds in two alternating steps as depicted in the last two subsections. We summarize it in Algorithm 1. Given the sampled states $\mathcal{Z} = \{z_n^k|n = 0, 1, \cdots, N, k = 1, \cdots, K\}$, we maximize the likelihood objective (7) to update the current estimate of model parameters. Given the current model parameters, we sequentially sample $\mathcal{Z}$ according to Eq. (8). We iterate between these two subroutines of algorithms to find the optimal model parameters.

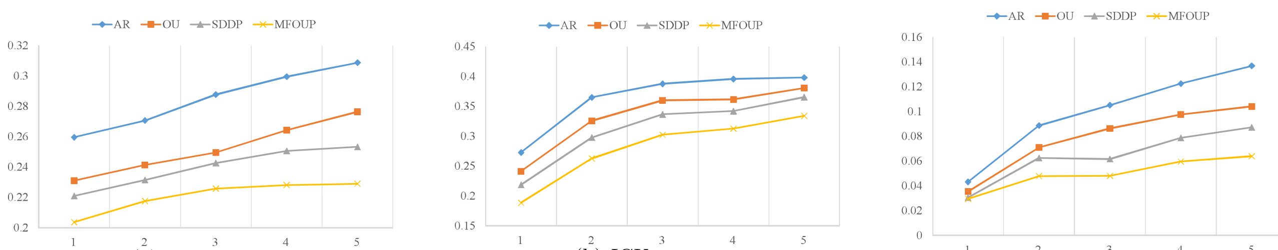
6 EXPERIMENTS

In this section, we perform experiments to test the proposed MFOUP model for three time-series prediction tasks. First,



(a) oil drilling (b) ICU sensor (c) stock market price

Figure 2: Examples of time series from the three data sets for the experiments.



(a) oil drilling (b) ICU sensor (c) stock market price

Figure 3: Comparison of different algorithms: the prediction errors (vertical axis) versus the number of prediction steps (horizontal axis) into future.

we will use two sensor network data sets – an oil drilling sensor network and a medical sensor network – to track and predict future sensor readings. Then, we will test the model on a stock market dataset to predict the trend of stock prices. We will explain how they are collected and some statistics about them.

6.1 Data Sets

We test the proposed time series forecasting approach on the following three data sets. Figure 2 illustrates some examples of these time series.

Oil drilling sensor network This sensor network is deployed to monitor the status of the oil drilling system, diagnosing and predicting the potential risk of dysfunction. The sensor network consists of sensors deployed at the surface, as well as sensors along the drill string, that operate within the wellbore in the subsurface of the earth. As the drill moves for vertical and sometimes horizontal drilling, these sensors encounter many time-varying factors, such as different geologic formations and relative positions of these sensors to the earth’s surface. They can affect the observed measurements made by the sensors. We consider a set of 33 such surface and wellbore sensors and demonstrate how our state based sensor selection approach can enable significant reductions in the number of sensors needed, while minimizing prediction error.

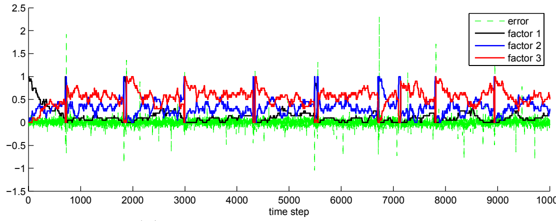
ICU sensor network The second data set records the readings of ICU (Intensive Care Unit) sensors. They are used to monitor the health status of patients. It contains the readings of seven types of ICU sensors, including *heart rate, body temperature, saturation of peripheral oxygen, diastolic blood pressure, mean blood pressure* and *respiration*, collected from 357 patients. Different patients have been accommodated in the ICUs for varying lengths of periods. Some patients are treated only for a few hours, while some others can be cared in hospital for several months. The health conditions of these patients keep changing over time, and we would like to study how a model can be adapted to these changing conditions.

Stock Market Data Set We gather the history prices of eight stocks selected from the technology sector exchanged on US market, including *Apple, Cisco, Facebook, Intel, IBM, MSFT, Oracle* and *Taiwan Semiconductor*. Their daily opening prices are sampled from 18 May 2012 to 18 November 2015. We select 18 May 2012 as the starting date as it was the first day for *Facebook* being traded on the public market. This results in a data set with 882×8 price points. The market volatility is constantly affected by many political, economic, social and technological factors, and we wish to test the effect of these hidden factors on the trend of stock prices.

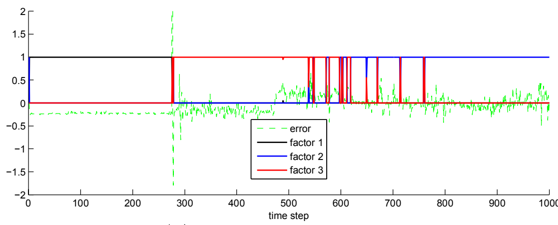
6.2 Compared Algorithms

We compare with the following algorithms to evaluate the proposed model.

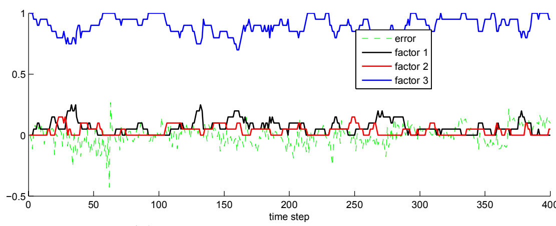
- AR (AutoRegression): the baseline benchmark. This model uses a conventional auto-regressive model [2, 23] to predict the future observations in a data stream, and a multi-variate AR model is created to jointly model multiple sequence.
- OU (Ornstein-Uhlenbeck process): this is the conventional OU model [24] that continuously predicts over a data stream. This is the baseline on which the proposed MFOUP is built.



(a) oil drilling sensor data set



(b) ICU sensor data set



(c) stock market price data set

Figure 4: Plots of prediction errors and the samples of FIPs over time. We sample the FIP for each factor multiple times so that we plot the beliefs over these factors at each time step. This figure shows how different factors reflect the evolution of time series through different stages. For the illustration purpose, the prediction errors are shifted so that their time-average mean centers at zero.

- SDDP (State-driven dynamic prediction): this model enhances the conventional AR model with the time-evolving states [27]. These dynamic states model the changing correlations between multiple sequences, thus yielding more precise predictions on the future values in sequential data streams. However, this model is unable to reveal mixture factors that affect the fluctuations underlying input dynamic data.
- MFOUP (Mixture Factorized Ornstein-Uhlenbeck Process): this is the proposed model. As compared with the other models, it explores mixture factors jointly and reveals their effects on the observations of time-series data. We will plot the sampled paths of these mixture factors below in the experiment details.

In the experiments, we normalize each time series to have zero mean and unit variance. All the data sets are split into three partitions: a training set for estimating the model parameters, a validation set for tuning hyper-parameters and a test set for evaluating the prediction errors. Specifically, for the proposed MFOUP model, we need to decide the number of hidden factors. This number is empirically chosen from 1 to 10 based on the performance on the validation set. Once it is fixed, the root mean square deviation is reported as prediction error on the independent test set for performance comparison.

6.3 Results

Figure 3 compares the experiment results obtained by different algorithms on the three data sets. We vary the number of prediction steps into future observations from 1 to 5. From the results, the proposed MFOUP model outperforms the other compared algorithms on these data sets. As we expected, the prediction errors increase as we attempt to predict in a larger number of steps forward into the future. The OU process performs slightly better than the AR model, and the MFOUP outperforms the SDDP as it captures multiple mixture factors. This becomes more evident as shown in Figure 4.

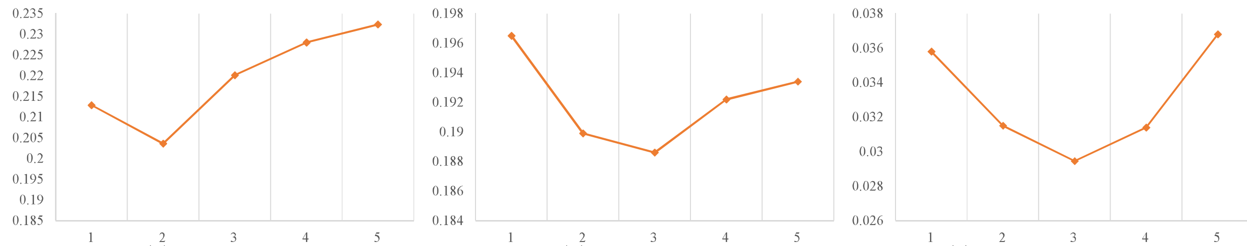
In Figure 4, we evaluate one-step prediction errors by the MFOUP model and plot the hidden factors inferred over time. The paths of FIPs are sampled twenty times and the beliefs (i.e., the frequency of a factor being sampled) are plotted together with the prediction errors. For the comparison, we fix the number of hidden factors to three on all three data sets. From the results, we can observe different patterns of hidden factors captured from these data sets.

First, In Figure 4(a), we can find that on the oil drilling data set there exhibit periodic patterns in its result, which might be caused by the periodically repeated drilling operations. On the other hand, these periodic patterns are witnessed by the cycles present in the paths of the inferred FIPs.

Second, Figure 4(b) illustrates the result on the ICU data set. At the beginning, the fluctuation in prediction error is relatively small, and the FIP path corresponding to the factor 1 dominates which reflects this stable condition. After that, the stable condition corresponding to the factor 1 disappears, and the patient's condition becomes unstable, being witnessed by the two emerging factors 2 and 3 that switch between each other. Eventually the factor 2 dominates over the other factors, accounting for the fluctuation of sensor readings starting from around the step 750. These dynamic factors can be presented to the physicians to understand their medical implications on the patient's conditions.

Finally, Figure 4(c) plots result on stock market price data set. The result shows that, on one hand, Factor 3 is the dominating factor that accounts for the most of market volatility. On the other hand, the other two factors tend to interleave with one another, which could model the periodic factors that drive the short-term move in stock prices.

These sampled FIP paths show how they characterize the dynamics underlying different genres of time-series data, which partly explains the competitive performance achieved by the MFOUP.



(a) oil drilling (b) ICU sensor (c) stock market price

Figure 5: Prediction errors (vertical axis) versus varying numbers of hidden factors (horizontal axis) on the three validation sets.

6.4 Effect of Hidden Factor Numbers

While the number of hidden factors are set based on an independent validation set, it is intriguing to see its effect on the model performance. Figure 5 plots the prediction errors versus varying numbers of hidden factors on the validation set of three data sets. For different applications, the best number of hidden factors varies. On the oil drilling sensor data set, two hidden factors are chosen to generate the best result on validation set, which turn out to be complementary in modeling the periodic drilling operations. However, on both the ICU sensor and stock market price data sets, the best number is chosen to be 3. As plot in Figure 4, one factor on ICU sensor time series accounts for the stable state with low prediction error, while the other two factors model the unstable states. On the stock market price data set, one factor captures the market volatility in stock prices in the background, while the other two captures periodic volatility in short terms.

7 CONCLUSION

This paper presents a novel paradigm of Ornstein-Uhlenbeck (OU) process capable of modeling multiple mixture factors evolving over time. Compared with the conventional time-series forecasting methods, the proposed MFOUP (Mixture Factorized OU Process) can model the evolutions of the mixture factors, which appear and disappear through time. Modeling the evolving mixture factors provides a more accurate characterization of data dynamics, thus yielding a better performance on the time-series forecasting problem. We prove some useful statistics to characterize the property of the process, which are later used to develop the algorithm to estimate its parameters. The experimental results demonstrate its competitive performance, along with the paths of hidden factors underlying the patterns of data fluctuations over time.

ACKNOWLEDGEMENT

G.-J. Qi was partly supported by NSF under grant 1560302, and acknowledged the donations from NVIDIA and Adobe. J. Wang was partly supported by the 973 program of China under grant 2014CB347600.

A PROOF OF LEMMA 1

In this section, we prove the first conclusion in Lemma 1.

Proof. The independence of $\{\varepsilon_n\}$ can be seen by applying the independent increments of the Brownian motion to the definition of this sequence. The zero mean of ε_n can be obtained by the zero mean and the independent increments of Brownian processes.

The covariance of ε_n can be proved by applying the following sequence of equalities:

$$\begin{aligned}
\Omega_n &\triangleq \mathrm{cov}(\varepsilon_n) \overset{1}{=} \sum_{k=1}^{K} \mathbb{E}\left(\int_{(n-1)\delta}^{n\delta} z_u^k e^{\Psi(u-n\delta)}\sigma^k dB_u^k\right)^2 \\
&\overset{2}{=} \sum_{k=1}^{K} \mathbb{E}\left(\int_{(n-1)\delta}^{n\delta} z_u^k \left(\sigma^k\right)^2 e^{\Psi(u-n\delta)} e^{\Psi^T(u-n\delta)} du\right) \\
&\overset{3}{=} \sum_{k=1}^{K} \left(\int_{(n-1)\delta}^{n\delta} \mathbb{E}(z_u^k | z_{n-1}^k) \left(\sigma^k\right)^2 e^{\Psi(u-n\delta)} e^{\Psi^T(u-n\delta)} du\right)
\end{aligned}$$

where all the expectations $\mathbb{E}$ are taken with the condition on $\{z_n\}$. The first equality applies the independence between the Brownian processes of different factors, the second equality uses the Ito's isometry property, and the third equality applies the dominated convergence theorem to change the order of the expectation and the integral.

By the transition matrix of continuous time Markov process, we have the following conditional expectation of z_u on z_n, for $\delta n < u < \delta(n+1)$.

$$\begin{aligned}
\mathbb{E}(z_u^k | z_{n-1}^k) &= P(z_u^k = 1 | z_{n-1}^k) \\
&= \frac{\alpha^k}{\alpha^k + \beta^k} + \frac{-(1 - z_{n-1}^k)\alpha^k + \beta^k z_{n-1}^k}{\alpha^k + \beta^k} e^{-(\alpha^k + \beta^k)(u-(n-1)\delta)}
\end{aligned}$$

Substituting it in the last equation, we have the following result

$$\begin{aligned}
&\sum_{k=1}^{K} \left(\int_{(n-1)\delta}^{n\delta} \frac{\alpha^k}{\alpha^k + \beta^k} \left(\sigma^k\right)^2 e^{\Psi(u-n\delta)} e^{\Psi^T(u-n\delta)} du\right) \\
&+ \sum_{k=1}^{K} \int_{(n-1)\delta}^{n\delta} \frac{-(1 - z_{n-1}^k)\alpha^k + \beta^k z_{n-1}^k}{\alpha^k + \beta^k} e^{-(\alpha^k + \beta^k)(u-(n-1)\delta)} \\
&\times \left(\sigma^k\right)^2 e^{\Psi(u-n\delta)} e^{\Psi^T(u-n\delta)} du \\
&\overset{4}{=} \sum_{k=1}^{K} \left(\frac{\alpha^k}{\alpha^k + \beta^k} \int_0^{\delta} \left(\sigma^k\right)^2 e^{-\Psi u} e^{-\Psi^T u} du\right) \\
&+ \sum_{k=1}^{K} \int_0^{\delta} \frac{-(1 - z_{n-1}^k)\alpha^k + \beta^k z_{n-1}^k}{\alpha^k + \beta^k} e^{(\alpha^k + \beta^k)(u-\delta)} \\
&\times \left(\sigma^k\right)^2 e^{-\Psi u} e^{-\Psi^T u} du
\end{aligned}$$

where the forth equality applies the change of integral variable.

The following result follows by applying the integral of exponential function of matrices to the above equation.

$$\sum_{k=1}^{K} \frac{\alpha^k}{\alpha^k+\beta^k} (\sigma^k)^2 (\Psi+\Psi^T)^{-1} \left(I - e^{-(\Psi+\Psi^T)\delta}\right)$$
$$+ \sum_{k=1}^{K} \frac{(1-z_{n-1}^k)\alpha^k - \beta^k z_{n-1}^k}{\alpha^k+\beta^k} (\sigma^k)^2$$
$$\times (\Psi+\Psi^T - (\alpha^k+\beta^k)I)^{-1} \left(e^{-(\Psi+\Psi^T)\delta} - e^{-(\alpha^k+\beta^k)\delta} I\right)$$

This proves the lemma. □

REFERENCES

[1] I. Adelman and C. T. Morris. A factor analysis of the interrelationship between social and political variables and per capita gross national product. *The Quarterly Journal of Economics*, pages 555–578, 1965.

[2] C. Aggarwal, Y. Xie, and P. Yu. On dynamic data-driven selection of sensor streams. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, August 2011.

[3] W. J. Anderson. *Continuous-time Markov chains: an applications-oriented approach*. Springer Science & Business Media, 2012.

[4] O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.

[5] D. A. Bessler. Relative prices and money: a vector autoregression on brazilian data. *American Journal of Agricultural Economics*, 66(1):25–30, 1984.

[6] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[7] P. Bonnet, J. Gehrke, and P. Seshadri. Towards sensor database systems. In *Mobile Data Management*, pages 3–14. Springer, 2001.

[8] G. E. Box, G. M. Jenkins, and G. C. Reinsel. *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons, 2011.

[9] G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.

[10] R. G. Brown. *Smoothing, forecasting and prediction of discrete time series*. Courier Corporation, 2004.

[11] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

[12] D.-I. Curiac, O. Banias, F. Dragan, C. Volosencu, and O. Dranga. *Malicious node detection in wireless sensor networks using an autoregression technique*. June 2007.

[13] J. Friedman and Y. Shachmurove. Co-movements of major european community stock markets: A vector autoregression analysis. *Global Finance Journal*, 8(2):257–277, 1998.

[14] Y. Grenier. Time-dependent arma modeling of nonstationary signals. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 31(4):899–911, 1983.

[15] J. D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[16] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[17] T. Hida. *Brownian motion*. Springer, 1980.

[18] R. A. Holley and D. W. Stroock. Generalized ornstein-uhlenbeck processes and infinite particle branching brownian motions. *Publications of the Research Institute for Mathematical Sciences*, 14(3):741–788, 1978.

[19] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[20] K.-j. Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1):307–319, 2003.

[21] H.-M. Krolzig. Predicting markov-switching vector autoregressive processes. Technical report, University of Oxford, Department of Economics, 2000.

[22] T. C. Mills and R. N. Markellos. *The econometric modelling of financial time series*. Cambridge University Press, 2008.

[23] S. Nassar, K.-P. SCHWARZ, N. EL-SHEIMY, and A. Noureldin. Modeling inertial sensor errors using autoregressive (ar) models. *Navigation*, 51(4):259–268, 2004.

[24] E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of the ornstein-uhlenbeck type. *Mathematical finance*, 13(4):445–466, 2003.

[25] B. Øksendal. *Stochastic differential equations*. Springer, 2003.

[26] S. Papadimitriou, J. Sun, and C. Faloutsos. Streaming pattern discovery in multiple time-series. In *Proceedings of the 31st international conference on Very large data bases*, pages 697–708. VLDB Endowment, 2005.

[27] G.-J. Qi, C. Aggarwal, D. Turaga, D. Sow, and P. Anno. State-driven dynamic sensor selection and prediction with state-stacked sparseness. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 945–954. ACM, 2015.

[28] K. H. Riitters, R. O'neill, C. Hunsaker, J. D. Wickham, D. Yankee, S. Timmins, K. Jones, and B. Jackson. A factor analysis of landscape pattern and structure metrics. *Landscape ecology*, 10(1):23–39, 1995.

[29] H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

[30] T. H. Rydberg. The normal inverse gaussian lévy process: simulation and approximation. *Communications in statistics. Stochastic models*, 13(4):887–910, 1997.

[31] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis. Parallel factor analysis in sensor array processing. *Signal Processing, IEEE Transactions on*, 48(8):2377–2388, 2000.

[32] S. J. Taylor. Modelling financial time series. 2007.

[33] R. S. Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.

[34] D. Tulone and S. Madden. Paq: Time series forecasting for approximate query answering in sensor networks. In *Wireless Sensor Networks*, pages 21–37. Springer, 2006.

[35] A. S. Weigend. Time series prediction: forecasting the future and understanding the past. *Santa Fe Institute Studies in the Sciences of Complexity*, 1994.