



# Gaussian Process Kernels for Noisy Time Series: Application to Housing Price Prediction

Juntao Wang<sup>1(✉)</sup>, Wun Kwan Yam<sup>1</sup>, Kin Long Fong<sup>1</sup>, Siew Ann Cheong<sup>2</sup>,  
and K. Y. Michael Wong<sup>1(✉)</sup>

<sup>1</sup> The Hong Kong University of Science and Technology, Hong Kong, China  
{jwangcc,wkyam,klfong}@connect.ust.hk, phkywong@ust.hk

<sup>2</sup> Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore  
cheongsa@ntu.edu.sg

**Abstract.** We study the prediction of time series using Gaussian processes as applied to realistic time series such as housing prices in Hong Kong. Since the performance of Gaussian processes prediction is strongly dependent on the functional form of the adopted kernel, we propose to determine the kernel based on the useful information extracted from the training data. However, the essential features of the time series are concealed by the presence of noises in the training data. By applying rolling linear regression, smooth and denoised time series of change rates of the data are obtained. Surprisingly, a periodic pattern emerges, enabling us to formulate an empirical kernel. We show that the empirical kernel performs better in predictions on quasi-periodic time series, compared to popular kernels such as spectral mixture, squared exponential, and rational quadratic kernels. We further justify the potential of the empirical kernel by applying it to predicting the yearly mean total sunspot number.

**Keywords:** Gaussian processes · Time series prediction  
Housing price prediction · Empirical kernels

## 1 Introduction

In recent years, Gaussian processes (GPs) have become a useful technique for machine learning [1]. They are based on a principled approach to statistical inference as long as correlations among the data can be estimated *a priori*. Hence, it provides a Bayesian nonparametric approach to classification, regression and prediction, and excellent results were produced [2].

The prior of the inference embodies our natural expectation about the behaviors of the data, and many forms of kernels have been proposed [1,2]. For example, for quantities that have smooth variations in space or time, the Squared Exponential (SE) kernel is a suitable choice. For data with stronger correlation

range, one may choose the Rational Quadratic (RQ) kernel. To capture the characteristics of different kernels, mixtures of kernels such as Spectral Mixture (SM) kernel [2] have also been used.

The importance of the appropriate choice of kernels was illustrated in GP time series prediction [2,3]. For example, while the SE kernel is suitable to describe smoothly varying data, it performs poorly when processing time series with periodic variations. For time series data with both long and short term trends, kernels such as RQ can capture the long term trend but not short one, whereas kernels such as SE performs oppositely. To improve the performance, kernels composed of a spectral mixture [2] was proposed, and was able to overcome the above drawbacks.

In real applications of time series prediction in science and economics, the data are often very noisy. The quasi-periodic components of the data may be masked by the noise. The prediction of housing prices considered in this work is a typical example. And for the quasi-periodic time series, traditional kernels such as SE and RQ may not be able to make meaningful predictions and sophisticated kernels such as SM can suffer from over-parameterization and overfitting.

Financial time series have been investigated using different methods, such as fuzzy time series [4], Support Vector Machine [5,6], and Artificial Neural Network (ANN) [7], in recent decades for the interest of maintaining stable and profitable market environment. However, most of the methods fail to be applicable to real time series [4] and they are too complex and lack the power of interpretation to capture the essential properties of the real time series.

Among various markets, the housing market might be the most appealing one, because it is closely related to the livelihood of individuals. For example, the Japanese housing bubble burst in 1991 exerted serious impact not only on the housing market but also the overall Japanese economy [8], leading to the sudden decline in both investment and consumption. However, because of the stochastic nature of financial markets [9], it is extremely difficult for scientists to accurately predict the subsequent observations based on the current information. On the other hand, many researchers are seeking indicators which could reveal the hidden states of the financial markets. For example, to prevent great depression caused by housing bubble bursts, there were many efforts to identify the explosive behavior, which is regarded as bubble inflation, in housing price time series. Several statistical techniques were used to confirm the existence of explosive behavior, such as root test, augmented Dickey Fuller test, and Cointegration test [10]. One popular method was introduced by [11], called PSY method, and gained its success in identifying multiple bubbles in historical housing data [12]. However, it lacks evidence that it is practical in predicting future trends in housing markets.

In this paper, we investigate the useful statistical information extracted from a real time series, namely, the Hong Kong housing price, and propose a comprehensive kernel using the prior knowledge obtained from the time series. By using this kernel, we can improve the predictive power of the GP models. In Sect. 2 we first review the theoretical framework of GPs. In Sect. 3, we describe

how the Fourier spectrum of the auto-covariance of the time series can be used to construct the High Order Periodic (HOPE) kernel for Gaussian processes. In Sect. 4, we describe the construction of the kernel from the Hong Kong housing price data. In Sects. 5 and 6, we will present the predictions on housing price data and yearly mean total sunspot number given by the GP models using the HOPE kernel and some other alternative popular kernels. The paper is concluded in Sect. 7.

## 2 Gaussian Processes

Given a set of points  $X = \{x_1, x_2, \dots, x_t\}$ , a Gaussian process [1] assumes that  $f(x_i)$ , the true value generated at point  $x_i$ , satisfies the joint Gaussian distribution

$$\mathbf{f} = [f(x_1) \dots f(x_t)]^T \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}(X, X)), \quad (1)$$

with mean vector  $\boldsymbol{\mu} = [\mu(x_1), \mu(x_2), \dots, \mu(x_t)]^T$  and covariance matrix  $\mathbf{K}$ , where  $K_{ij} = k(\boldsymbol{\theta}, x_i, x_j)$  is the kernel parameterized by hyperparameter  $\boldsymbol{\theta}$ . And  $\mathcal{N}$  is the Gaussian distribution. Generally, the true value  $f(x_i)$  cannot be known and only their noisy versions  $y_i = f(x_i) + \epsilon$  can be observed, here  $\epsilon$  is the independent identically distributed Gaussian noise with variance  $\sigma_n^2$ . Based on the given observations and prior distribution, the expected value and variance of the true value generated at a set of points  $X'$  can be inferred. Specifically, to estimate the true values  $\mathbf{f}'$  at  $X'$  based on the known observations and prior, we have the joint distribution

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}' \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}' \end{bmatrix}, \begin{bmatrix} \mathbf{K}(X, X) + \sigma_n^2 I & \mathbf{K}(X, X') \\ \mathbf{K}(X, X') & \mathbf{K}(X', X') \end{bmatrix} \right), \quad (2)$$

where  $I$  is the identity matrix, and the posterior predictive distribution

$$p(\mathbf{f}' | \mathbf{y}, X, X') = \mathcal{N}(\mathbf{m}, \mathbf{C}), \quad (3)$$

where  $\mathbf{m}$  is the estimated mean and  $\mathbf{C}$  is the estimated variance at points  $X'$ , which can be analytically derived as:

$$\mathbf{m} = \boldsymbol{\mu}' + \mathbf{K}(X, X')(\mathbf{K}(X, X) + \sigma_n^2 I)^{-1}(\mathbf{y} - \boldsymbol{\mu}), \quad (4)$$

$$\mathbf{C} = \mathbf{K}(X', X') - \mathbf{K}(X, X')(\mathbf{K}(X, X) + \sigma_n^2 I)^{-1}\mathbf{K}(X', X). \quad (5)$$

Therefore, specifying the kernel  $k$ , the variance  $\sigma_n^2$ , and the mean function  $\mu$  determines the GP, and hence, the estimated mean value can be calculated by the model. To tune the hyperparameters of GP, we need to optimize the marginal log-likelihood  $\mathcal{L}$  with respect to the unknown hyperparameters in the covariance matrix and the mean function:

$$\begin{aligned} \mathcal{L} = \log p(\mathbf{y} | \boldsymbol{\theta}, X) = & -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{K} + \sigma_n^2 I)^{-1}(\mathbf{y} - \boldsymbol{\mu}) \\ & - \frac{1}{2} \log \det (\mathbf{K} + \sigma_n^2 I) - \frac{|X|}{2} \log 2\pi, \quad (6) \end{aligned}$$

where  $X$  is the set of observation points,  $\mathbf{y}$  is the corresponding vector of observed values at  $X$ , and  $|X|$  is the number of observations. Thus, given the estimated hyperparameters, variance  $\sigma_n^2$ , and mean function that maximize the log-likelihood  $\mathcal{L}$ , we are able to predict the expected value and standard deviation of the unknown true value generated at points  $X'$  by the predictive distribution.

### 3 Construction of Kernels

In this section, we construct a new kernel for noisy realistic time series. Although empirical correlation can be questionable in reflecting the true theoretical underlying processes [2], we may show that it is still possible to improve predictions on noisy quasi-periodic time series by using the information extracted from statistical analysis, such as auto-covariance analysis and Fourier Transform.

The cross-covariance  $\gamma_{FG}(l)$  for two finite discrete real time series  $F = \{F_t : t \in T\}$  and  $G = \{G_t : t \in T\}$  is defined as

$$\gamma_{FG}(l) = \sum_t F_{t+l} G_t, \quad (7)$$

where  $l$  is the time that  $G$  lags  $F$ . For the case that  $F = G$ ,  $\gamma_{FG} = \gamma_{FF}$  is the auto-covariance of the time series  $F$ . For the housing price data we adopt the definition that is without the subtraction of expected value of the time series.

Given a time series  $F$  with  $N$  observations, the auto-covariance  $\gamma_{FF}(l)$  can characterize the similarity in the data up to lag  $N - 1$ . To construct the kernel function describing the similarities between arbitrary data, we transform the auto-covariance  $\gamma_{FF}(l)$  into its frequency-domain:

$$X(f) = \sum_{l=0}^{N-1} \gamma_{FF}(l) e^{-i2\pi fl/N}, \quad (8)$$

and obtain the auto-covariance spectrum  $\frac{|X(f)|}{N}$ .

Therefore, we select the significant periodic components having large enough amplitudes to construct the kernel and ignore the other weaker components, since the components with small amplitudes may be just noises or redundancies and have little contribution to the true connections between the data. Thus we propose the one-dimensional High Order Periodic (HOPE) kernel with order  $n$ :

$$k_n(x_i, x_j) = \alpha e^{-(x_i - x_j)^2 / (2\ell^2)} \sum_{m=1}^n A_m \cos\left(2\pi f_m |x_i - x_j|\right), \quad (9)$$

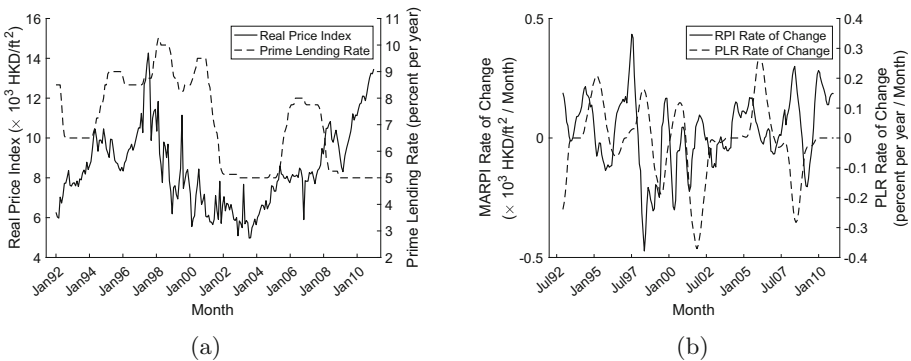
where  $n \in \mathbb{Z}^+$ ,  $\alpha$ ,  $\ell$ , and  $f_m, m = 1, \dots, n$ , are hyperparameters, and the weights  $A_m, m = 1, \dots, n$  is the  $m$ -th largest amplitudes (peaks) of the components in the auto-covariance spectrum. The introduction of the exponential factor  $e^{-(x_i - x_j)^2 / (2\ell^2)}$  is to ensure that the kernel vanishes when the space (time) difference between two observations points approaches infinity. Intuitively, two data

points well separated in space (time) should have weak correlations. The HOPE kernel can be considered as an extension of the Spectral Mixture (SM) kernel [2] to deal with noisy quasi-periodic data, whereby we determine a number of hyperparameters directly from the statistical features of the data to alleviate the problem of over-parameterization and overfitting, which will be shown in Sects. 5 and 6. Then by optimizing the log-likelihood (Eq. 6) with respect to these hyperparameters and the mean function  $\mu$ , we obtain a trained GP regression model.

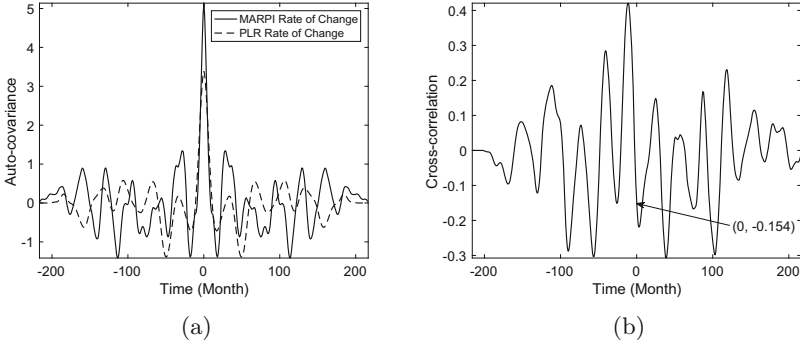
## 4 Hong Kong Housing Price

The data of the Hong Kong property transaction prices were obtained through negotiations from EPRC Limited, a wholly-owned subsidiary of the Hong Kong Economic Times that specializes in providing property information to market-related industries. The dataset contains records of transacted properties for the period 1992 to 2010 and amounts to 2,492,842 transaction records. Figure 1(a) illustrates the monthly average of real price index (MARPI) (in HKD per square foot normalized by the consumer price index) of Hong Kong housing transactions and the Prime Lending Rates (PLR) in Hong Kong from Jan 1992 to Dec 2010 (totally 228 data points). Since the data of MARPI are very noisy, to pre-process and denoise the data, we investigate the time series of MARPI rate of change over 12 months (a year), defined as the slope of the linear regression line for the past 12 months MARPI data. Similarly, the time series of PLR rate of change over 12 months can also be calculated. As a comparison, these two time series are illustrated in Fig. 1(b).

As shown in Fig. 1(b), quasi-periodic features exist in both the MARPI rate of change and the PLR rate of change. The quasi-periodicity is exhibited even more clearly in the auto-covariance (Eq. 7) of the rate of change as shown in Fig. 2(a).

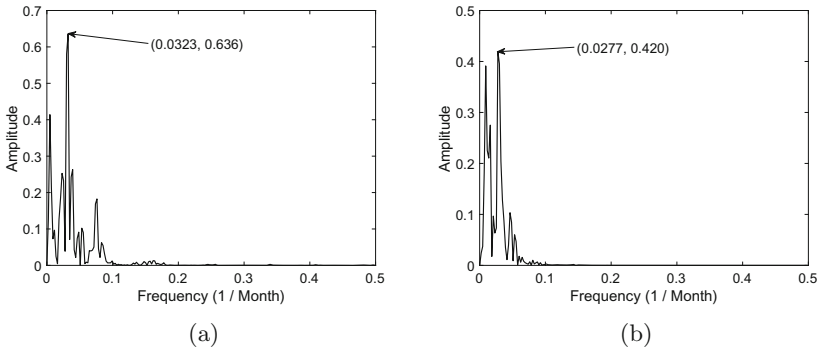


**Fig. 1.** (a) The monthly average real price index (MARPI) per square foot of Hong Kong housing and the prime lending rate (PLR). (b) The MARPI rate of change and the PLR rate of change.



**Fig. 2.** (a) Auto-covariance of MARPI rate of change and PLR rate of change. (b) Cross-correlation (normalized version of cross-covariance) of MARPI rate of change and PLR rate of change.

The periodicity of the rate of change observed from the auto-covariance is around 30 months, as confirmed by the result of the auto-covariance spectrum in Fig. 3.



**Fig. 3.** Single-sided auto-covariance spectrum of (a) MARPI rate of change and (b) PLR rate of change. The components with the largest amplitude in (a) and (b) have periodicity  $1/0.0323 = 31.0$  Month and  $1/0.0277 = 36.1$  Month respectively.

It is interesting to consider why housing price time series is quasi-periodic. We found evidence that the housing price is correlated with PLR, which happened to have a quasi-periodic variation during this period of study. While the underlying reason is beyond the scope of this study, it is natural to expect the occurrence of cycles in the world economy. When consumption heats up, PLR will be raised to prevent overheating, and when the economy is weak, PLR will be reduced to stimulate consumption. As shown in Figs. 1, 2, 3, there is a close correlation between the periods of two time series. Furthermore, the cross-correlation (normalized version of cross-covariance) is negative near 0 months showing that housing prices drop (rise) when the mortgage rate rises (drops).

## 5 Experiments

In this section, we show that the HOPE kernel can be applied to the real time series to improve the predictive power of the GP. At the same time, to contrast the prediction results using the new kernel, other popular kernels, such as Spectral Mixture (SM) with number of components  $Q = 12$ , Squared Exponential (SE), and Rational Quadratic (RQ) kernel (as shown in Table 1 [1, 2]) are also applied to predict the data. Since the size of our data is small (around 200 data points), Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [13] with cubic interpolation is applied to optimize the log-likelihood function. The `fitrgp.m` and `fminunc.m` functions from MATLAB [14] are used to implement the GP and optimization. The mean function  $\mu$  of the GP model is set to be a constant function  $\mu(x) = c$ . For the optimization of hyperparameters, we randomly sample 40 initializations of hyperparameters for each model and optimizations are repeated. The final optimal hyperparameters are the ones which yield the highest log-likelihood  $\mathcal{L}$  for the model. During the sampling, the order  $n$  in HOPE is sampled between 1 and 12.

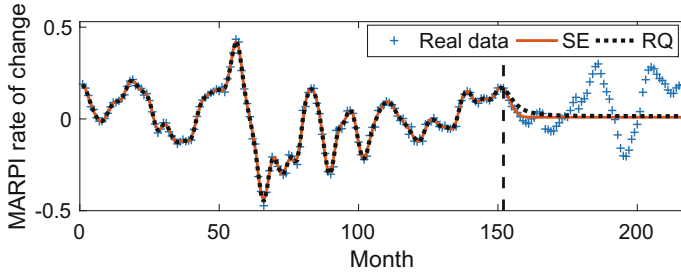
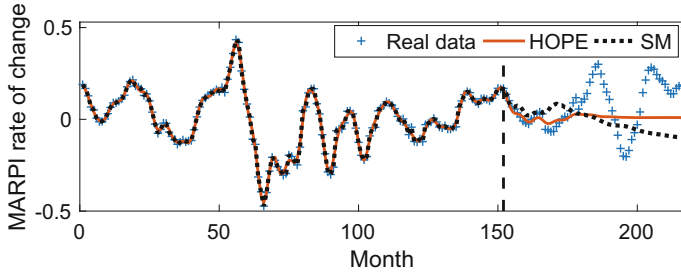
**Table 1.** Expressions of one-dimensional popular kernels, where  $\tau = |x_i - x_j|$ .

Kernel	Expression	Hyperparameters
SM	$k_{\text{SM}}(x_i, x_j) = \sum_{q=1}^Q \omega_q \exp\left(-2\pi^2 \tau^2 \nu_q\right) \cos(2\pi \tau \mu_q)$	$\omega_q, \nu_q, \mu_q$
SE	$k_{\text{SE}}(x_i, x_j) = \sigma^2 \exp\left(-\frac{\tau^2}{2\ell^2}\right)$	$\sigma, \ell$
RQ	$k_{\text{RQ}}(x_i, x_j) = \sigma^2 \left(1 + \frac{\tau^2}{2\alpha\ell^2}\right)^{-\alpha}$	$\sigma, \alpha, \ell$

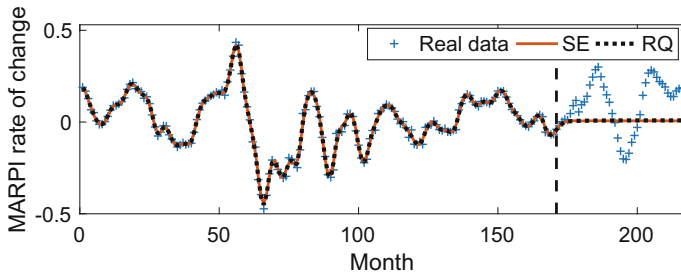
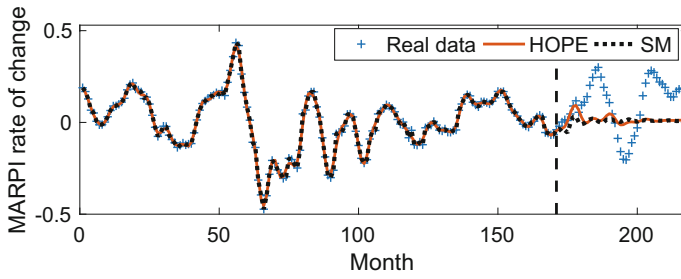
### 5.1 Predictions on MARPI Rate of Change

To illustrate the predictive power of the GP model using the HOPE, we choose two subsets from the time series of the MARPI rate of change as the training set, one containing the first 152 observations in the time series and the other constituting the first 171 observations, and the remaining data as testing sets. The predictions using different kernels can be found in Fig. 4.

GP models using all kernels perform well in reconstructing the training inputs, since the predictions on training data given by the models almost overlap with the real data in the training set. However, in future predictions, models using SE and RQ have little predictive power, since the estimated mean observations in the future (outside the training set) diverges largely and quickly from the testing data, and their values tend to fall back to the mean function  $c$ . The predictions using SM are unsatisfactory, since there are many huge gaps between the estimated mean observations and real data. On the other hand, the predictions given by the model using HOPE do not deviate from the testing data dramatically and can capture the trend of the future data up to a considerable number of steps. At the same time, Fig. 4 indicates that generally, the errors may increase when we predict the values to a more distant future.



(a)



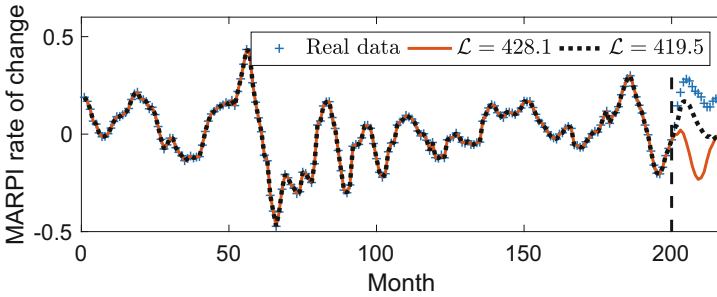
(b)

**Fig. 4.** Predictions on MARPI rate of change using different kernels. The training data are on the left hand side of the vertical dashed line. (a) The first 152 observations as the training inputs. (b) The first 171 observations as the training inputs.



Another observation is that for the training set with 152 (171) observations, the optimized log-likelihood  $\mathcal{L}$  of HOPE and SM are 270.9 (311.7) and 300.4 (329.9) respectively. Although the SM has higher log-likelihood, Fig. 4 shows that it has poorer performance in predicting the unobserved data. One suggested reason for this is that using SM, possibly overparameterized, leads to the overfitting so that the predictive power of the GP is suppressed. Meanwhile, this observation can be considered as the evidence that compared to SM, HOPE is able to effectively prevent the GP from overfitting and give better predictions on the future data.

Moreover, we use another example to illustrate the problem of overfitting induced by the SM. For one set of training inputs (first 200 observations of MARPI rate of change), the predictions given by two GP models using SM ( $Q = 12$ ) are shown in Fig. 5. Compared to the model with log-likelihood  $\mathcal{L} = 419.5$ , the model with the higher log-likelihood  $\mathcal{L} = 428.1$  has rather poorer performance on predicting the future data.



**Fig. 5.** Predictions on MARPI rate of change with first 200 observations as training inputs, which are on the left hand side of the vertical dashed line.  $\mathcal{L}$  is the log-likelihood of the model. Both models are using SM with  $Q = 12$ .

## 5.2 Performance and Stability

Since GPs can be really sensitive to the initialization of the hyperparameters as well as the input training data, we further compare the prediction stability of different kernels by measuring their mean squared error (MSE) and following the procedure of time series prediction performance assessment introduced by Hyndman [15]. Specifically, experiments are repeated using different training sets, each one having additional one observation than the previous one, and during each experiment and for each training set the MSE in  $h$ -step horizon (that is, from 1-step prediction up to  $h$ -step prediction) is computed to investigate the performance of prediction. Initially, the training set has 151 observations. And the average MSE of experiments of each kernel can be found in Table 2. The smaller the average MSE, the more stable the model, because small average MSE indicates that the model can maintain relatively good performance when the training inputs are changed. As a comparison, the variance of the whole data

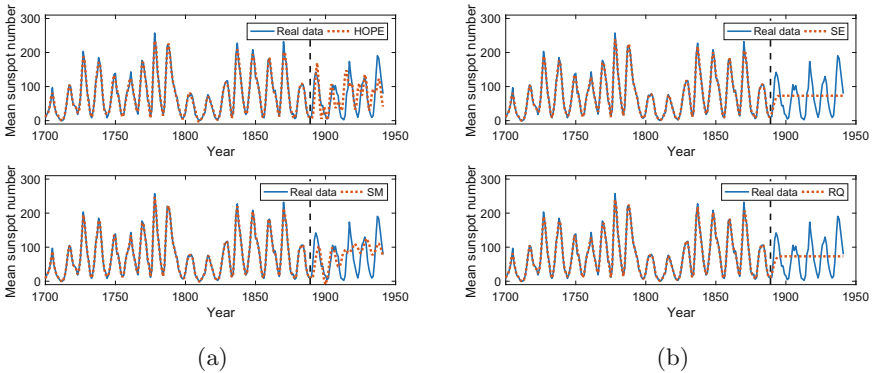
set is 0.0232. Table 2 shows that the stability of HOPE outperforms other kernels in several-step horizon prediction, which also justifies that HOPE has better performance in capturing the future trends of the time series. It also indicates that the predictive power fades away when the GP model is trying to estimate the data in distant future, since the average MSE increases dramatically when the prediction horizon increases. And we can also see that the performances of SM are relatively poor, which might be the consequences of overfitting.

**Table 2.** Average MSE using different kernels

Horizon	HOPE	SM	SE	RQ
6-step	<b>0.0078</b>	0.0088	0.0092	0.0086
9-step	<b>0.0118</b>	0.0149	0.0130	0.0130
12-step	<b>0.0148</b>	0.0171	0.0153	0.0159

## 6 Application to Sunspot Time Series

To further illustrate the power of HOPE, we apply it to another well known time series, yearly mean total sunspot number, which can be obtain from <http://www.sidc.be/silso/datafiles>. To make comparisons, the popular kernels in Table 1 (SM with  $Q = 20$ ) are also applied. The data from year 1700 to year 1899 (totally 190 data points) are used as training inputs, and the data from year 1900 to year 1941 (totally 42 data points) are considered as testing set. The mean function of the GP is set to be constant function  $\mu(x) = c$ . BFGS algorithm and 100 random initializations of hyperparameters are applied to search for optimal estimations of hyperparameters for each model, and the order  $n$  in HOPE is sampled between 1 and 20.



**Fig. 6.** Predictions on yearly mean total sunspot number with the first 190 observations as training inputs, which are on the left hand side of the vertical dashed line. (a) Predictions given by HOPE and SM with  $Q = 20$ . (b) Predictions given by SE and RQ.

Predictions on the yearly mean total sunspot number given by different kernels are illustrated in Fig. 6. Similar to the cases of predictions on MARPI rate of change, for sunspot time series, SM ( $\mathcal{L} = -857.5$ ) has a larger log-likelihood than HOPE ( $\mathcal{L} = -885.0$ ). However, for the performance on predicting the testing data, compared to HOPE (MSE = 1947.5), SM (MSE = 2272.3) gives poorer predictions, which indicates the problem of overfitting caused by SM. Also, predictions on future data given by both SE (MSE = 2588.4) and RQ (MSE = 2599.7) are quite poor, which can be seen in Fig. 6(b), though their log-likelihood values ( $\mathcal{L}_{SE} = -887.7$  and  $\mathcal{L}_{RQ} = -887.5$ ) are comparable to that of HOPE.

## 7 Conclusion

In this paper, we have discussed the Gaussian process (GP), which is a powerful regression model to predict the time series. Using a new kernel (HOPE) based on the information extracted from the auto-covariance of the training time series, we improved the predictive power of the GP model and it compares favourably with models using other popular kernels. Our work shows that the choice of the kernel is a very essential factor to the performance of GP. Choosing a kernel from some popular choices may not work for noisy quasi-periodic data, in which case extracting the kernel from the covariance functions provides an alternative.

The Gaussian process can be a powerful predictor in financial time series analysis. Since we have only used Gaussian processes to analyze the single time series, we believe that by incorporating multiple time series, such as Prime Lending Rate and trading volume, Gaussian processes may capture more information underlying the series, and give more accurate predictions. However, the model cannot predict the value at arbitrary time steps ahead, so we may further investigate the applicable range that the predicted values are still close to the real data. On the other hand, we may also focus on predicting a few steps ahead so that we can determine its accuracy and stability in short-term prediction.

**Acknowledgements.** We thank Prof. Chihiro Shimizu for forwarding the data to us. This work is supported by the Research Grants Council of Hong Kong (grant numbers 16322616 and 16306817).

## References

1. Rasmussen, C.E., Williams, C.K.: Gaussian Process for Machine Learning. MIT Press, Cambridge (2006)
2. Wilson, A.G., Adams, R.P.: Gaussian process kernels for pattern discovery and extrapolation. In: Dasgupta, S., McAllester, D. (eds.) Proceedings of the 30th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 28, pp. 1067–1075. PMLR, Atlanta (2013)
3. Brahim-Belhouari, S., Bermak, A.: Gaussian process for nonstationary time series prediction. *Comput. Stat. Data Anal.* **47**(4), 705–712 (2004)

4. Lee, C.H.L., Liu, A., Chen, W.S.: Pattern discovery of fuzzy time series for financial prediction. *IEEE Trans. Knowl. Data Eng.* **18**(5), 613–625 (2006)
5. Kim, K.J.: Financial time series forecasting using support vector machines. *Neurocomputing* **55**(1–2), 307–319 (2003)
6. Cao, L.J., Tay, F.E.H.: Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. Neural Netw.* **14**(6), 1506–1518 (2003)
7. Kim, T.Y., Oh, K.J., Kim, C., Do, J.D.: Artificial neural networks for non-stationary time series. *Neurocomputing* **61**, 439–447 (2004)
8. Hamada, K., Kashyap, A.K., Weinstein, D.E.: *Japan’s Bubble, Deflation, and Long-Term Stagnation*. MIT Press, Cambridge (2011)
9. Tsay, R.S.: *Analysis of Financial Time Series*, vol. 543. Wiley, Hoboken (2005)
10. Hui, E.C., Yue, S.: Housing price bubbles in Hong Kong, Beijing and Shanghai: a comparative study. *J. Real Estate Finance Econ.* **33**(4), 299–327 (2006)
11. Phillips, P.C., Shi, S., Yu, J.: Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. *Int. Econ. Rev.* **56**(4), 1043–1078 (2015)
12. Yiu, M.S., Yu, J., Jin, L.: Detecting bubbles in Hong Kong residential property market. *J. Asian Econ.* **28**, 115–124 (2013)
13. Fletcher, R.: *Practical Methods of Optimization*. Wiley (2013)
14. MATLAB: *Statistics and Machine Learning Toolbox Release 2017b*. The MathWorks Inc., Natick (2017)
15. Hyndman, R.J.: Measuring forecast accuracy. In: Gilliland, M., Sglavo, U., Tashman, L. (eds.) *Business Forecasting: Practical Problems and Solutions*, pp. 177–184. Wiley (2015)