New Kernel Function in Gaussian Processes Model

SukonthipSuphachan[1], Poonpong Suksawang[1,*], Jatupat Mekparyup[2,]**

[1]College of Research Methodology and Cognitive Science, Burapha University,Chonburi, Thailand
[2]Department of Mathematics, Faculty of Science, Burapha University, Chonburi, Thailand
*Corresponding author: Asst. Prof. Poonpong Suksawang, PhD; College of Research Methodology and Cognitive Science, Burapha University, 169 Muang, Chonburi, 20131, Thailand; E-mail: poonpong@buu.ac.th
*Corresponding author: Asst. Prof. JatupatMekparyup, PhD; Department of Mathematics, Faculty of Science, Burapha University, 169 Muang, Chonburi 20131,Thailand; E-mail: Jatupat@buu.ac.th

Abstract- New Kernel Design Technique was presented in the form of the sum of Linear Kernel, the multiplication of 3 Kernel Functions, including Squared Exponential Kernel, Linear Kernel and Rational Quadratic Kernel, and the multiplication of 2 Kernel Functions, including Periodic Kernel and Linear Kernel, which was used as a component for finding answers in the Gaussian processes. The results showed that the mean absolute percentage error predicted by the New Kernel Function, when the sample size was 180 (8.50E-15), was lower than the Squared Exponential Kernel (SE), Periodic Kernel (PER), Rational Quadratic Kernel (RQ) and Linear Kernel (LIN), which gave the Average Absolute Error of 2.57E-07, 5.56E-02, 2.63E-08 and 4.35E-01, respectively.

[**Keywords**: Kernel Function, Gaussian Processes, Composing Kernels ,Forecasting]

1. Introduction

There were several techniques used for the most accurate forecasting results. In general, the data used for forecasting was quantitative prediction technique data, which was a technique using historical data to create mathematical models used in forecasting. The techniques used could be divided into 2 broad categories as follows. First, Time series method, it was the way that only uses historical data to predict data in the future or our interested patterns of data or variable variation. There were several methods of forecasting. Each method had different analysis procedures. Examples of forecasting by time series method included Composite Forecasting Method, Moving Average Forecasting Method, Exponential Smoothing Forecasting Method, Autoregressive integrated moving average model (ARIMA) Forecasting Method, and Holt-Winter's Seasonal Exponential Smoothing Forecasting Method. Second, Causal Forecasting Method, it was the method that studies causal relationships and effects of interested variables with other variables influencing the interested variables. Examples of Casual Forecasting Method included Multiple Linear Regression Analysis, both linear and non-linear Neural Network Series.

Forecasting via Artificial intelligence, this forecasting technique had been developed rapidly and progressively as the system was capable of learning and thinking like humans, as an artificial intelligence model, data mining, or non-focused variable database. By the past several decades, there were more researches that were conducted by applying artificial intelligence techniques for forecasting. Examples of Artificial Intelligence Forecasting Method included Artificial neural network (ANN) [1,2], Fuzzy systems [3], Knowledge based expert system (KBES) [4], Wavelet analysis method [5], Support vector machine (SVM) method [6], and the new techniques currently accepted called Gaussian Processes.

Gaussian Processes was used in data mining, data classification, regression and time series data forecasting [7-10]. Gaussian processes [11-17] were now recognized as an effective tool for solving problems of regression, classification, and decision in machine learning. It worked effectively even with less training data, had better convergence rate than SARIMA, ANN, and supported Regression Vector Machine [9,18]. Gaussian processes was advantages over other Machine Learning techniques because of its full capabilities in forecasted probability distribution, as well as prediction of uncertainty of forecasting. These features made Gaussian processes an ideal tool for forecasting purposes [19].

The key element of Gaussian processes was the value of Kernel Function, or sometimes called the Covariance Function. The forecasting accuracy of the Gaussian processes' algorithm depended on the proper selection of Kernel Functions to match the problems. The Kernel Function had stationary properties, which remained constant over time [18]. Kernel Functions was a highly efficient method. It could work well with various types of data, such as string, vector, messages, etc. Also, it could find various ways of relationships, such as ranking, classification, cluster, etc. [18,20]. Many Kernel Functions were used in Gaussian processes for forecasting, for example, Radial Basis Function (RBF) or Exponential Quadratic Kernel, Rational Quadratic Kernel, Matérn Kernel, etc. [20].

Because of Kernel Function was the primary function in Gaussian processes, so the accuracy of forecasted data depended on the selection of Kernel Function and proper parameters adjustment [18,20,21]. Therefore, the problem was what Kernel Functions should be used. These problems needed to be tested and solved in order to obtain the accurate forecasting model. This research was aimed to create an effective new Kernel Function, compatible with many problematic conditions by mainly considering the time series data. Because the time series data of one variable might consist of only one, two, three, or all four types of data, so selecting one Kernel Function might not enough to deal with the problems of the prediction. Thus, the researchers combined Kernel Function models together as a new Kernel, which had Superposition properties [20] separating the variables that control the features of each functions freely.

2. Material and Method

2.1 Gaussian processes [22,23] was a random process or stochastic process, which could be defined as the distribution on the time function $f(x)$with the mean (m: x), Covariance or known as Kernel Function $k(x,x^{'})$ or $k(\tau)$ where $\tau = x - x^{'}$, which could be generated from the time function $f(x)$, evaluating the match of the knowledge from Observation Set $y = [y_1, y_2, y_3, \ldots, y_N]^T$, as a vector size where $N \times 1$, with Observation Set Input $X = [x_1, x_2, x_3, \ldots, x_N]^T$ with the same size of $N \times 1$ [21]. This was defined as a Gaussian process [24]

$$y_i = f(x_i) + \varepsilon_i \tag{1}$$

$$f(x_i) \sim \mathcal{GP}\left(m(x_i), k(x_i, x_j)\right) \tag{2}$$

Where i was the index of measure, and $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ was a Gaussian Distributed Error Model with Zero Mean and σ_ε^2 variance.

The design model of observation was $y_i = f(x_i) + \varepsilon_i$ according to Equation 1, where Covariance was equal to$cov(y_i, y_j) = \boldsymbol{K}(x_i, x_j) + \sigma_\varepsilon^2 \delta_{ij}$; δ_{ij} represented Kronecker Delta and $\delta_{ij} = 1$; $i = j$; and others were equal to 0. The correlation between observation data and test target)Target: f_j) was based on Equation 3.

$$\begin{bmatrix} y \\ f_j \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \boldsymbol{K}(X,X) + \sigma_\varepsilon^2 I_N & \boldsymbol{k}(X, x_j) \\ \boldsymbol{k}(X, x_j)^T & \boldsymbol{k}(x_j, x_j) \end{bmatrix}\right) \tag{3}$$

According to Equation 3, it was found that the conditional probability of $P(f_j | x_i, \boldsymbol{y})$ was distributed on the function $f(x_j)$ with the mean of $m(x_j)$ and Covariance of $cov(f_j)$ [25].

$$m(x_j) = \boldsymbol{k}(X, x_j)^T (\boldsymbol{K}(X,X) + \sigma_\varepsilon^2 I_N)^{-1} \boldsymbol{y} \tag{4}$$

$$cov(f_j) = \boldsymbol{k}(x_j, x_j) - \boldsymbol{k}(X, x_j)^T (\boldsymbol{K}(X,X) + \sigma_\varepsilon^2 I_N)^{-1} \boldsymbol{k}(X, x_j) \tag{5}$$

From the definition of the Gaussian process in (4), the Minimum Mean Square Error was used to forecast $f(x_j)$ where:

$$f(x_j) \sim m(x_j) = E\left[f(x_j) \middle| x_j, \boldsymbol{y}\right] \tag{6}$$

Therefore, it was possible to forecast the tested data by the mean of Gaussian processes, which was [26].

$$f(x_j) \sim m(x_j) = \boldsymbol{k}(X, x_j)^T (\boldsymbol{K}(X, X) + \sigma^2 I_N)^{-1} \boldsymbol{y} \tag{7}$$

The forecasted result was given by $f(x_j) \sim m(x_j)$. From the answer of Equation (7), the accuracy of the Gaussian processes' algorithm depended on the selection of appropriate functions for the Kernel Function. In addition, the relationship in Equation (7) could also be given in the form of a linear combination of n Kernel Functions [26].

$$f(x_j) = \sum_{i=1}^{N} \alpha_i \, \boldsymbol{k}(x_i, x_j) \tag{8}$$

Where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T = (\boldsymbol{K}(X, X) + \sigma^2 I_N)^{-1} \boldsymbol{y}$**.**

2.2 Kernel Functions

Kernel Function was a part of the Machine Learning algorithms that are closely related to Gaussian Processes [18].

Where$\tau_{i-j} = x_i - x_j$. Thus, the Cross Covariance of $\boldsymbol{k}(X, x_j), \boldsymbol{K}(X, X)$ would be as follows:

$$\mathbf{K}(X, X) = \boldsymbol{K}(\tau) \sim \begin{bmatrix} k(\tau_0) & k(\tau_{-1}) & \ldots & k(\tau -_{(N-1)}) \\ k(\tau_1) & k(\tau_0) & \ldots & k(\tau_{-(N-2)}) \\ \vdots & \vdots & \vdots & \vdots \\ k(\tau_{(N-1)}) & k(\tau_{(N-2)}) & \ldots & k(\tau_0) \end{bmatrix} \tag{9}$$

$$\boldsymbol{k}(X, x_j) \sim \left[k(x_i, x_j) \ldots k(x_N, x_j)\right]^T \tag{10}$$

$$\sim [k(\tau_0) \ldots k(\tau_{N-1})]^T$$

The key element of Gaussian Process was the Kernel Function or sometimes called Covariance Function [20], where $k(\tau_j)$ was set to Covariance data between the pairs of Functions $f(x_i)$ and $f(x_j)$ at the right position matching the input x_i and x_j, respectively, which was $(\tau_{i-j}) = k(x_i, x_j) = cov\left(f(x_i), f(x_j)\right)$. Covariance Function would be selected to match the features of the function, which depended on the pattern of problems, whether time signal, flat signal, linear signal or polynomial signal. That is, there was no need to modify the form of the algorithm in Equation (7). When the function of interested problems changed, what to do is just adjust the Covariance Function or Kernel Function to suit the problem. Moreover, finding the distribution on the kernel's hyper-parameters could explain various properties of the data as well, such as rate of variation, periodicity, and smoothness [20]. In general, the Kernel Function was a pair-mapping of the inputs $x_i \in X$and $x_j \in X$ into the $\mathbb{R}$ domain and the covalence. Thus, for function $f(x_i) \in \mathbb{R}$ with Zoro Mean, the Kernel Function would be defined as [20].

$$k(\tau_{i-j}) = k(x_i, x_j) = cov\left(f(x_i), f(x_j)\right) = \mathbb{E}\left[f(x_i) f^*(x_j)\right] \tag{11}$$

Equation (11) was used as a Kernel Function of Gaussian processes according to Equation (5) and Equation (7), so any matrix$\boldsymbol{K}(X,X) = \boldsymbol{K}(\tau)$ with elements in $K_{ij} = k(x_i, x_j) = k(\tau_{i-j})$ must be a positive semi- definite matrix [21], with the condition stated that $z^T K z \geq 0$ for all $z \in \mathbb{R}^N$.

Squared exponential kernel

Squared exponential kernel known as radial basis function(RBF) or exponential quadratic kernel is a kernel that has been widely used with features as a function of time$f(x)$ with the smoothness and slow change by $\sigma$and$l$is responsible for determining the size and amplitude sensitivity of the time changing [18] with a form of function being:

$$k_{SE}(x,x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right) \quad (12)$$

Rational Quadratic Kernel

For exponential kernel with the sensitivity of the change of time function $f(x)$ will depend on the scale value αorl, the scope of time function $f(x)$ created by rational quadratic kernel is an infinite sum of the squared exponential kernel with the $l$length that is different. The quadratic kernel was created to design complex data by the αvalue of the rational quadratic kernel if $\alpha \to \infty$, it can be converted back into a squared exponential [[18with the form of the function is as below.

$$k_{RQ}(x,x') = \sigma^2 \left(\frac{(1+(x-x')^2/p}{2\alpha\ell^2}\right)^{-\alpha} \quad (13)$$

Periodic kernel

Kernel function designed for use with functions generated from Gaussian processes that appear to be redundant as proposed by MacKay in 1998 with the hyper parameters being$\theta = \{p, \ell, \sigma\}$. $p$is the length of the function signal in time.ℓandσ is the length of the repeating period and the size of the signal. If the hyper- parameter is correct, periodic signals from$f(x_j)$ functions of the Gaussian processor may compare to find the hyper- parameter that can be generate of the Gaussian process or may be compared to find the hyper-parameter with the function being:

$$k_{PER}(x,x') = \sigma^2 \exp\left(\frac{2\sin^2(\pi(x-x')/p}{\ell^2}\right) \quad (14)$$

Linear kernel

Linear functionality changed for a long-term from Gauss linear. The format of the function being:

$$k_{Lin}(x,x') = \sigma(x-\ell)(x'-\ell) \quad (15)$$

2.3 Composing Kernels

Kernel function in each category under the Gaussian process. Influence on different models of time function simulation. For this research, the researcher will newly create a kernel function based on time series data, which characteristics of the time series of a variable consists of four parts: a cycle, trend, a seasonal variation, and fluctuations from unusual events These 4 aspects are consistent with Each kernel function i.e. the trend looks consistent with the pattern of learning the long-term trends and the constant variance. It is a feature of the Linux X squared and linear kernel, cyclical character corresponds to the function model for Information sessions are repeated regularly, but not the type of time period kernel Seasonal variation is consistent with Patterns that are repetitive in time. It's corresponds to the time-type kernel function, which is a function for constant learning, even with variations and fluctuations from unusual events, and corresponds to Kernel quadratic , which is a function of complex change but also changes slowly and because the time series data of one variable may consist of one, two, three, or all four. One of the kernel functions may not cover the problem of the forecast so taking the form of the kernel, the 4 functions are combined to be a new kernel with superposition [21] with the variable control features function independently in each category separately.

The research is focused on how to integrate the kernel function is Sum and Product.

$$k_a + k_b = k_a(x, x') + k_b(x, x') \tag{16}$$

$$k_a * k_b = k_a(x, x') * k_b(x, x') \tag{17}$$

Applying the form of Squared Exponential Kernel (SE), Periodic Kernel (PER), Rational Quadratic Kernel (RQ), and Linear Kernel (LIN) combined used in the simulation with Matlab program by usingMonte Carlo technique.

The determination of the kernel function error to be used comparing the value of each kernel form by using:

Mean absolute deviation (MAD)

$$MAD = \frac{1}{t}\sum_{i=1}^{t}\left|e_i(1)\right| \tag{18}$$

Mean Square Error (MSE)

$$MSE = \frac{1}{t}\sum_{i=1}^{t}[e_i(1)]^2 \tag{19}$$

Mean absolute percentage error (MAPE)

$$MAPE = \frac{100}{t}\sum_{i=1}^{t}\left|\frac{e_i(1)}{Z_i}\right| \tag{20}$$

The kernel functions using low MAD, MSE or MAPE is better than the kernel functions using high MAD, MSE or MAPE.

.3 Results and discussions

When combining the kernel function, Squared Exponential Kernel (SE), the Periodic Kernel (PER), Rational Quadratic Kernel (RQ), and Linear Kernel (LIN) together by applying the Sum and Product structures, the new forms of Kernel function totally 3,639 profiles derived, as shown in Table 3.1.

Table 3.1 shows the new kernel function model.

Rank	Kernel function mode	Number of mode
1	(SE*PER)+(SE*RQ)+(SE*LIN)	20
2	(SE*PER*RQ*LIN)+SE	4
3	(SE*PER*RQ*LIN)+SE+PER	6
4	(SE*PER*RQ*LIN)+SE+PER+RQ	4
5	(SE*PER*RQ*LIN)+SE+PER+RQ+LIN	1
6	(SE*PER)+(SE*RQ)+(SE*LIN)+SE	80
7	(SE*PER)+(SE*RQ)+(SE*LIN)+SE+PER	120
8	(SE*PER)+(SE*RQ)+(SE*LIN)+SE+PER+RQ	80
9	(SE*PER)+(SE*RQ)+(SE*LIN)+SE+PER+RQ+LIN	20
10	(SE*PER*RQ)	4
11	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ)	80
12	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ)+SE	320
13	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ)+SE+PER	480
14	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ)+SE+PER+RQ	320

Rank	Kernel function mode	Number of mode
15	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ)+SE+PER+RQ+LIN	80
16	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ*LIN)	20
17	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ*LIN)+SE	80
18	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ*LIN)+SE+PER	120
19	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ*LIN)+SE+PER+RQ	80
20	(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*RQ*LIN)+SE+PER+RQ+LIN	20
21	(SE*PER*RQ)+(SE*PER)	24
22	(SE*PER*RQ)+(SE*PER)+SE	96
23	(SE*PER*RQ)+(SE*PER)+SE+PER	144
24	(SE*PER*RQ)+(SE*PER)+SE+PER+RQ	96
25	(SE*PER*RQ)+(SE*PER)+SE+PER+RQ+LIN	24
26	(SE*PER*RQ)+(SE*PER)+(SE*PER)	60
27	(SE*PER*RQ)+(SE*PER)+(SE*PER)+SE	240
28	(SE*PER*RQ)+(SE*PER)+(SE*PER)+SE+PER	360
29	(SE*PER*RQ)+(SE*PER)+(SE*PER)+SE+PER+RQ	240
30	(SE*PER*RQ)+(SE*PER)+(SE*PER)+SE+PER+RQ+LIN	60
31	(SE*PER*RQ*LIN)+(SE*PER)	6
32	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)	15
33	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)+(SE*PER)	20
34	(SE*PER*RQ*LIN)+(SE*PER)+SE	24
35	(SE*PER*RQ*LIN)+(SE*PER)+SE+PER	36
36	(SE*PER*RQ*LIN)+(SE*PER)+SE+PER+RQ	24
37	(SE*PER*RQ*LIN)+(SE*PER)+SE+PER+RQ+LIN	6
38	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)+SE	60
39	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)+SE+PER	90
40	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)+SE+PER+RQ	60
41	(SE*PER*RQ*LIN)+(SE*PER)+(SE*PER)+SE+PER+RQ+LIN	15
	Total	3639

The error of each kernel function was calculated by simulating with Matlab. The errors are shown in Table 3.2.

Table 3.2 shows the error of each kernel function.

kernel function	Error
(SE*PER)+(SE*RQ)+(SE*LIN)	MSE = 5.0606e-08 , MAPE = 1.5773e-08 , MAD = 0.00016173
(SE*RQ)+(SE*LIN)+(PER*RQ)	MSE = 8.9383e-08 , MAPE = 1.9872e-08 , MAD = 0.00023003
(SE*LIN)+(PER*RQ)+(RQ*LIN)	MSE = 1.1278e-07 , MAPE = 2.2979e-08 , MAD = 0.00026739
⋮	⋮
(SE*PER)+(SE*RQ)+(SE*LIN)+(SE*PER*LIN)+PER+RQ	MSE = 7.3625e-07 , MAPE = 2.095e-08 , MAD = 0.00029638
(SE*RQ)+(SE*LIN)+(PER*RQ)+(SE*PER*LIN)+PER+RQ	MSE = 1.077e-07 , MAPE = 2.0197e-08 , MAD = 0.00021762
(SE*LIN)+(PER*RQ)+(RQ*LIN)+(SE*PER*LIN)+PER+RQ	MSE = 1.3068e-06 , MAPE = 3.1908e-08 , MAD = 0.00045056
⋮	⋮
(SE*PER*LIN)+(RQ*LIN)+LIN	MSE = 1.2045e-06 , MAPE = 3.1356e-08 , MAD = 0.00044047
(SE*RQ*LIN)+(LIN*PER)+LIN	MSE = 2.7497e-12 , MAPE = 1.1273e-11 , MAD = 2.9533e-07
(SE*RQ*LIN)+(RQ*LIN)+LIN	MSE = 3.9171e-07 , MAPE = 9.5851e-09 , MAD = 0.00014228
⋮	⋮
(SE*PER*RQ*LIN)+(SE*LIN)+(PER*RQ)+SE+PER+RQ+LIN	MSE = 1.2207e-07 , MAPE = 2.4202e-08 , MAD = 0.0002773
(SE*PER*RQ*LIN)+(SE*LIN)+(PER*LIN)+SE+PER+RQ+LIN	MSE = 1.2234e-06 , MAPE = 2.9656e-08 , MAD = 0.00042943
(SE*PER*RQ*LIN)+(PER*LIN)+(RQ*LIN)+SE+PER+RQ+LIN	MSE = 1.1191e-06 , MAPE = 2.4754e-08 , MAD = 0.00036224

The results showed that the most efficient new kernel function was as follows;

$$k_{(SE*RQ*LIN)+(LIN*PER)+LIN}$$

The form of the new kernel function derived according to the equation (21).

$$k(x,x') = \left(\sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right)\left(\frac{(1+\frac{(x-x')^2}{p}}{2\alpha\ell^2}\right)^{-\alpha} (x-\ell)(x'-\ell)\right) + \left(\sigma^2 exp\left(\frac{2\sin^2(\pi(x-x')/p}{\ell^2}\right)(x-\ell)(x'-\ell)\right) + (\sigma(x-\ell)(x'-\ell)) \quad (21)$$

When comparing the performance among the new kernel functions, Squared Exponential Kernel, Periodic Kernel, Rational Quadratic kernel, and Linear kernel, when the sample size were different, the results were show in Table 3.3.

Table 3.3 shows the performance comparison among the new kernel functions, Squared Exponential Kernel, Periodic Kernel. Rational Quadratic kernel, and Linear kernel, when the sample size are 36, 60, and 180 .

KERNEL	MSE			MAPE			MAD		
	N=36	N=60	N=180	N=36	N=60	N=180	N=36	N=60	N=180
New	1.94E-19	7.41E-13	6.70E-11	8.50E-15	1.23E-11	6.62E-11	1.79E-10	2.90E-07	1.43E-06
SE	1.46E-05	2.97E-06	4.09E+04	2.57E-07	1.19E-07	1.31E-02	2.17E-04	9.96E-05	1.60E+02
PER	9.52E+05	9.51E+05	7.81E+05	5.56E-02	5.58E-02	6.11E-02	8.06E+02	7.99E+02	7.26E+02
RQ	2.54E-07	4.86E-08	3.03E-07	2.63E-08	1.20E-08	3.44E-08	3.86E-04	1.72E-04	4.20E-04
LIN	5.96E+07	5.87E+07	6.09E+07	4.35E-01	4.23E-01	5.09E-01	5.84E+03	5.77E+07	5.84E+03

The results of the performance comparison among the new kernel functions, Squared Exponential Kernel, Periodic Kernel, Rational Quadratic Kernel, and Linear Kernel are shown in Figure 3.1.

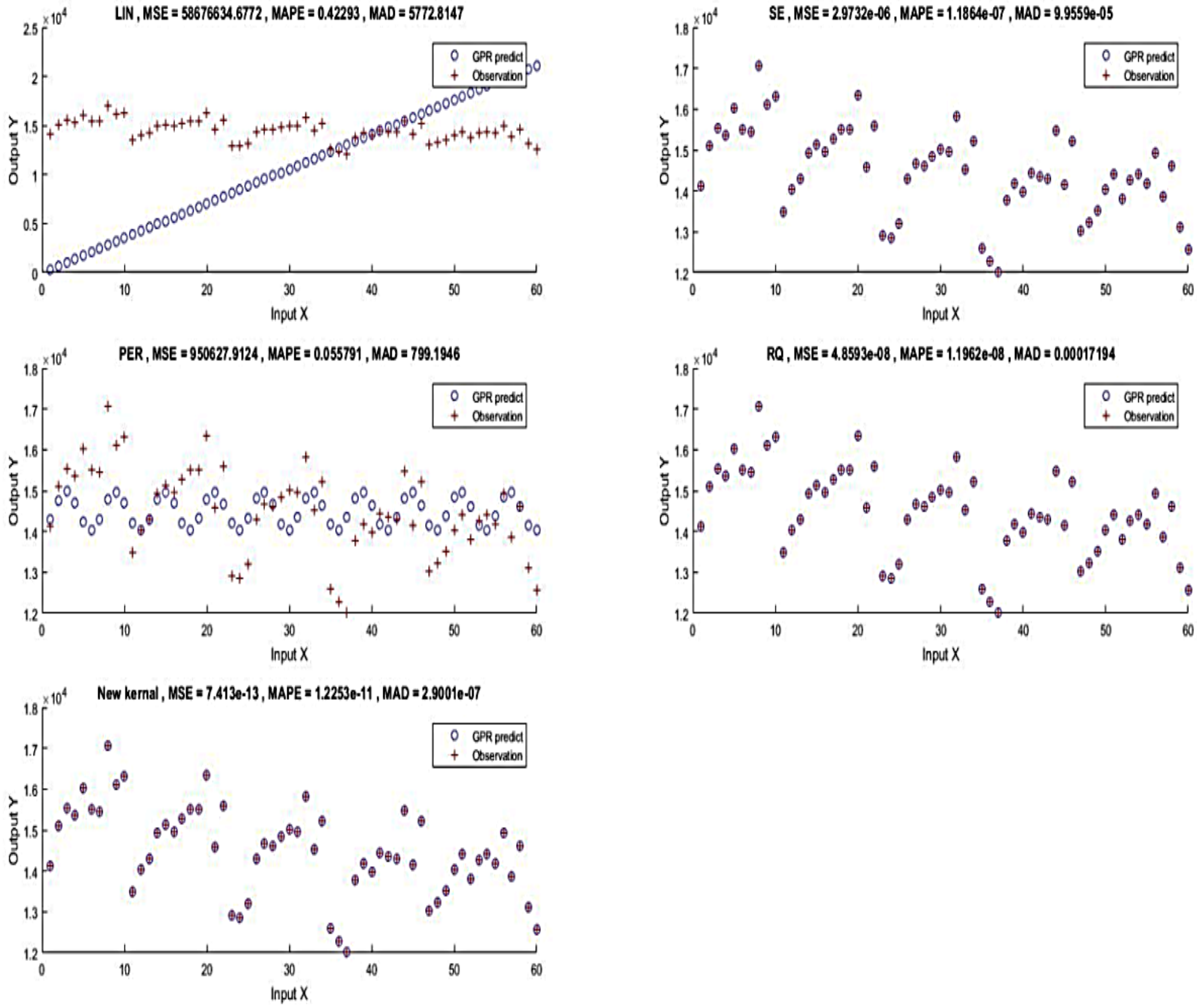

Figure 3.1 shows the performance comparison among the new kernel function, Squared Exponential kernel, Periodic kernel, Rational Quadratic kernel, and Linear kernel when the sample size is 60

Discussions

In this paper ,we have presented a brief outline of the conceptual and mathematical basis of GP modelling of time series. We found that the structures are often capable of accurate extrapolation in complex time-series datasets, and used kernel combination methods on a variety of prediction tasks. The goal of automating the choice of kernel function , we introduced a space of composite kernels defined compositionally as sums and products . The learned kernels often yield decompositions of a signal into diverse and interpretable components. We believe that a data-driven approach to used new kernel structures can help make forecasting and classification methods accessible to non-experts.

4 Conclusions

The current forecast that is needed to find out about the future has been highly focus because forecast plays an extremely important role in the plot and decision making about the operation of every individual profession and various organizations. The forecasted value will be used to plan and make decision. In the past, there have been researches conducted on the forecast with the techniques used were ARMA, SVR, artificial neural networks, and Gaussian processes.

This study presents a method to predict the Gaussian kernel to forecast by designing new functions in line with the characteristics of time series data by blending squared exponential kernel periodic,rational, quadratic,and lineartechnique in which the application presentation of the kernel function to fit a long-term trend. It is the hyper parameters of 16 kernel functions, which is multiplied by the squared exponential kernel is the introduction of long-term relationships of the model or change the relationship slowly and multiplied with a linear kernel. This is equivalent to the multiplication of functions that have been modeled by linear functions resulting in standard deviations without affecting the relationship functional value. When multiplied by the kernel function quadratic, it would likely cause a change in the long run complex has a smooth slow contained hyper- 8parameters: θ_1 Actively controls the amplitude of the dataθ_2 , the sensitivity of the information. It also serves to weight the variance of the data $\theta_3\theta_4$, and $\theta_6$generated by the change. $\theta_5$adjusts the frequency of each period. $\theta_7$controls the size of the difference of information and θ_8 controls the variance of the data. The result of multiplication of the three kernels cause a change in the trend of complex changes in the long run, repeated pattern. It also contains discrepancies in information including seasonality. To be effective for volatile data generated by other external variables, multiplication was conducted with a linear kernel by θ_9 is responsible for determining sensitivity. Change of information to correspond to θ_{10} with the duty to adjust the frequency in the repetition period and θ_{11} is responsible for determining the data period. This will make the output repeatable and varied withθ_{12}andθ_{13} being responsible for controlling the size of the difference of data, determining the size of the linear kernel data multiplication time-type kernel function resulted in less variance with a wider range of amplitude and information in the form of long-term linear trend with a positive linear kernel making the trend in the long run smooth with more constant variability containing hyper-parameters being $\theta_{14}$to control the amplitude of the θ_{15} andθ_{16}data to control the difference of information to a constant variance.

From the test results, it was found that this new kernel function using 16hyper-parameters had the ability to forecast the precision and efficiency, and water when making comparisons to the kernel, the squared exponential kernel Periodic kernel and rational quadratic and linear kernel yielded more effective precision.

5 References

[1] Yetis, Yunus, and Mo Jamshidi. "Forecasting of Turkey's electricity consumption using Artificial Neural Network." *World Automation Congress (WAC), 2014*.IEEE, 2014.

[2]Tamizharasi, G., S. Kathiresan, and K. S. Sreenivasan. "Energy forecasting using artificial neural networks." *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* 3.3 (2014): 7568-7576.

[3] Tao, Zhou, Tang Zhong, and Ren Shuyan. "Medium and long term load forecasting based on fuzzy times series." *Advanced Mechatronic Systems (ICAMechS), 2013 International Conference on*.IEEE, 2013.

[4] Torres, G. Lambert, B. Valiquette, and D. Mukhedkar. "Short-term feeder load forecasting: An expert system using fuzzy logic." *IFAC Power Systems and Power Control*. 2014.

[5] Chen, Yanhua, et al. "A hybrid application algorithm based on the support vector machine and artificial intelligence: an example of electric load forecasting." *Applied Mathematical Modelling* 39.9 (2015): 2617-2632.

[6] Selakov, A., et al. "Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank." *Applied Soft Computing* 16 (2014): 80-88.

[7] Rasmussen, Carl Edward, and Hannes Nickisch. "Gaussian processes for machine learning (GPML) toolbox." *Journal of Machine Learning Research* 11.Nov (2010): 3011-3015.

[8] Earls, Cecilia, and Giles Hooker. "Bayesian covariance estimation and inference in latent Gaussian process models." *Statistical Methodology* 18 (2014): 79-100.

[9] Ploysuwan, Tuchsanai. "Spectral mixture kernel for pattern discovery and time series forecasting of electricity peak load." *TENCON 2014-2014 IEEE Region 10 Conference*.IEEE, 2014.

[10] Lourenco, J. M., and P. J. Santos. "Short-term load forecasting using a Gaussian process model: The influence of a derivative term in the input regressor." *Intelligent Decision Technologies* 6.4 (2012): 273-281.

[11]Wu, Qi, Rob Law, and Xin Xu. "A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong." *Expert Systems with Applications* 39.5 (2012): 4769-4774.
[12] Salcedo-Sanz, Sancho, et al. "Prediction of daily global solar irradiation using temporal gaussian processes." *IEEE Geoscience and Remote Sensing Letters* 11.11 (2014): 1936-1940.
[13] Sun, Alexander Y., Dingbao Wang, and Xianli Xu. "Monthly streamflow forecasting using Gaussian process regression." *Journal of Hydrology* 511 (2014): 72-81.
[14] Hachino, T., et al. "Improvement of Gaussian Process Predictor of Electric Power Damage Caused by Typhoons Considering Time-Varying Characteristics." (2015).
[15] Lei, Yu, et al. "Prediction of Length-of-day Using Gaussian Process Regression." *Journal of Navigation* 68.03 (2015): 563-575.
[16] Senanayake, Ransalu, Simon O'Callaghan, and Fabio Ramos. "Predicting Spatio-Temporal Propagation of Seasonal Influenza Using Variational Gaussian Process Regression." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
[17] Ludkovski, Michael, James Risk, and Howard Zail. "Gaussian Process Models for Mortality Rates and Improvement Factors." (2016).
[18] Rasmussen, Carl Edward. "Gaussian processes for machine learning." (2006).
[19] Clavería González, Óscar, Enric Monte Moreno, and Salvador Torra Porras. "Modelling cross-dependencies between Spain's regional tourism markets with an extension of the Gaussian process regression model." *Series-Journal Of The Spanish Economic Association, 2016, vol. 7, num. 3, p. 341-357* (2016).
[20] Duvenaud, David. *Automatic model construction with Gaussian processes*. Diss. University of Cambridge, 2014.
[21] Simionovici, Ana Maria. *LOAD PREDICTION AND BALANCING FOR CLOUD-BASED VOICE-OVER-IP SOLUTIONS*. Diss. University of Luxembourg, Luxembourg, Luxembourg, 2016.
[22]Wilson, Andrew, and Ryan Adams. "Gaussian process kernels for pattern discovery and extrapolation." *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*. 2013.
[23] Barkan, Oren, Jonathan Weill, and Amir Averbuch. "Gaussian process regression for out-of-sample extension." *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 2016.
[24]Kowal, Daniel R., David S. Matteson, and David Ruppert. "Gaussian Processes for Functional Autoregression." *arXiv preprint arXiv:1603.02982* (2016).
[25] Ghoshal, Sid, and Stephen Roberts. "Extracting predictive information from heterogeneous data streams using Gaussian Processes." *Algorithmic Finance* 5.1-2 (2016): 21-30.
[26]Van Vaerenbergh, Steven, Jesus Fernandez-Bes, and Victor Elvira. "On the relationship between online Gaussian process regression and kernel least mean squares algorithms." *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 2016.

AUTHORS

First Author – SukonthipSuphachan, College of Research Methodology and Cognitive Science, Burapha University, 169 Muang , Chonburi, 20131, Thailand; E-mail: Suphachan@yahoo.com
Second Author –DR. Poonpong Suksawang, Asst. Prof.; College of Research Methodology and Cognitive Science, Burapha University, 169 Muang , Chonburi, 20131, Thailand; E-mail: poonpong@buu.ac.th
Third Author –DR. JatupatMekparyup, Asst. Prof.; Department of Mathematics, Faculty of Science, Burapha University, 169 Muang, Chonburi 20131,Thailand; E-mail: Jatupat@buu.ac.th