Choosing Initial Hyper-Parameter Based on Simple Feature Data for Gaussian Process Time Series State Space Models

S S Sholihat

Statistics Group, Institut Teknologi Bandung, Bandung, Indonesia

seli.siti.sholihat@gmail.com

Abstract. Choosing Initial hyper-parameters and covariance functions are important in Gaussian process time series (GPTS) state space model (SSM), because it should be done before doing predicting and forecasting inference. Some articles proposed optimization to choose the hyper-parameter but the optimization may not converge to the global maxima. Choosing covariance function and hyper-parameter randomly and subjectively is used to solve this problem. In this paper, we propose simple method for choosing initial hyper-parameter in GPTS SSM by feature of time series data that fit with the feature of prior GPTS. We do simulation and analysis feature of prior GPTS for some covariance functions using some hyper-parameters to know how covariance function and hyper-parameter influence prior GPTS. The simulation purposes are finding features of prior GPTS SSM for some hyper-parameter on covariance functions. The simulations show that different covariance functions and hyper-parameters influence prior GPTS. We applied this method in two real data time series to choose initial hyper-parameter based on features of data. We choose initial hyper-parameter based on simple features data. Simple features data here trend, statistics, periodicity of data. The results show that this method have smaller RMSE in forecasting and predicting than choosing covariance function and hyper-parameter randomly. This simple method can lead to better subjectivity.

INTRODUCTION

Nowadays, time Series Analysis has become one of the most important and widely used branches of Mathematical Statistics e.g. in econometrics [1], mathematical finance [2]:, meteorology [3], earthquake prediction [4] and many other application. A time series is a series of data points where the ordering matters. Mostly, time series data related with processes that data are obtained at sequence of time [5]. It means that data can be collected at regular time interval, such as daily, weekly, monthly, or annually. Change of ordering could change the meaning of data. The purposes of time series modelling are to find the features of time series pattern, to forecast in future and to analysis how past affects future.

The classic modelling time series illustration is shown in figure 1. Its model y_t dependson y_{t-1}, the linear model can be as $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma_n^2)$, called autoregressive order1(AR (1)). The autoregressive model is one of classic time series models. Furthermore, modelling time series can be generalized to prediction of $y_t$depends on $y_{1:t-1} = (y_1, y_2, \dots, y_{t-1})$.

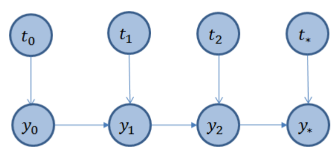


Figure 1. Illustration time series model: Autoregressive order 1(AR (1)).

Gaussian process time series is one of time series modelling which using Gaussian process. Gaussian Processes are one of the stochastics processes that have proved very successful to perform inference directly over the space of function [6]. Gaussian processes time series model generalizes classic time series model [7]. There are two kind Gaussian process time series (GPTS) model, a state space model (SSM) and an autoregressive Gaussian processes timeseries (ARGP) model [8]. Based "time domain" model, there are two type of time series models. First, the model depends on the series of past values as an input. Second, the regression model use time indices as an input. In this paper we discuss GPTS SSM that use time indices as an input, depend on state space model (SSM).

Gaussian process is defined by covariance matrix that is formed by covariance function. Covariance functions are kernel functions $k_\xi(t,t') \in \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$, ξ is collection of hyper-parameter, parameter on the covariance functions. Hyper-parameters terminology is defined to differ with parameter on distribution (mean, covariance matrix). Covariance function is discussed in subchapter, we discussed some covariance functions: exponential, squared exponential, gamma exponential, Matern class, rational quadratic, piecewise polynomial, periodic [6]. Once covariance function is selected then prior hyper-parameter is estimated by common approach maximum marginal likelihood, but it may not converge to the global maxima [9]. It is solved by choosing hyper-parameter randomly. Once covariance function is selected, choosing prior hyper-parameters randomly in GP regression have no significant impact on Gaussian process regression [10].

In this paper we do prior GPTS SSM simulation for some hyper-parameter in covariance functions. We want to see impact of hyper-parameter on covariance function to prior GPTS SSM. Then we will choose hyper-parameter based on feature of time series data that have similarity with prior GPTS SSM. There are four components of classic time series: trend, seasonal, cyclical, and random noise. Those similar components can be found in Gaussian processes e.g. stationary, fluidity, periodicity. Hence, there are some connectivity between time series models and Gaussian processes. Using those four features, we will analysis the real data time series for choosing the hyper-parameter and covariance function.

GAUSSIAN PROCESS

Gaussian process is defined as a sequence of random variables, which any finite number of random variables have a multivariate joint Gaussian distribution. Mathematically, stochastic process $\{Y_t\}_{t\geq 0}$ is a Gaussian if for any $t_1, t_2, \dots, t_N \in T$, the distribution of the random vector $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}) \in \mathbb{R}^n$ is a Gaussian distribution in $\mathbb{R}^n$.

Let $(y_{t_1}, y_{t_2}, \dots, y_{t_n})$ are observation data with the input $t_1, t_2, \dots, t_N \in T$, respectively. The Gaussian process probability density function of the observations is in equation (1).

$$\boldsymbol{P}(\boldsymbol{y_{t_1}}, \boldsymbol{y_{t_2}}, \dots, \boldsymbol{y_{t_N}} | \boldsymbol{t_1}, \boldsymbol{t_2}, \dots, \boldsymbol{t_N}) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\pi}^{N/2}|\boldsymbol{\Sigma}|^{\mathbf{1/2}}} \boldsymbol{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}\,(\boldsymbol{y}-\boldsymbol{\mu})} \tag{1}$$

Where $\mu \in \mathbb{R}$ and $\Sigma \in \mathbb{R}^{NxN}$ are mean and covariance matrix. More clearly, the illustration of Gaussian processes is showed in Figure 3. Mathematically, for any set T, any mean function $\mu: T \to \mathbb{R}$ and any covariance function $k: T \times T \to \mathbb{R}$, there is a Gaussian process $f(t)$ on T, $\mathrm{E}[f(t)] = \mu(t)$, and $Cov(f(t), f(t')) = k_\xi(t,t')$, $\forall t,\ t' \in T$, it denotes $f \sim GP(\mu_t, k_\xi)$. Gaussian process is a distribution over function that can be fully determined by second-order statistics covariance matrix (Σ), which can be built by covariance function$k_\xi(\cdot,\cdot)$; ξ is collection of hyper-parameter ($\xi = (l, \alpha, \beta, \gamma)$). For simplicity, the Gaussian process can assume to have mean zero, defining covariance function defines behaviour of process. In this paper will discuss some of covariance functions in Table 1.

1.1. Covariance Function

The covariance function $k_\xi(t,t')$ is a function of the model inputs which yields variances and covariances values for the corresponding outputs [6]. Using the covariance function, we can get covariance of the outputs model from the input, $cov(y_t, y_{t'})$=$k_\xi(t,t')$. In time series observation indices $t_1, t_2, \dots, t_N \in T$as the input, we can get $cov\left(y_{t_i}, y_{t_j}\right)$=$k_\xi(t_i, t_j)$, then element of covariance matrix ($\Sigma_{ij} = cov\left(y_{t_i}, y_{t_j}\right) = k_\xi(t_i, t_j)$) are held, which is covariance matrix must be symmetric by definition which covariance functions are kernel, clearly. Defining the covariance function to construct covariance matrix of Gaussian processes can define stationary, isotropic, fluidity and periodicity. Some stationary and non-stationary covariance functions is below in Table 1. Simulation of variants hyper-parameter to some covariance functions are in figure 3.

Table 1. Covariance Function

		Name	Covariance Function $\boldsymbol{k(r); r = \|t - t'\|}$
Stationary	1	Exponential (E)	$k_E(r) = h^2 exp\left(-\frac{r}{l}\right), l > 0$
	2	Squared exponential (SE)	$k_{SE}(r) = h^2 exp\left(-\frac{r^2}{2l^2}\right), l > 0$
	3	γ-exponential (GE)	$k(r) = h^2 exp\left(-\left(\frac{r}{l}\right)^{\gamma}\right), untuk\ 0 < \gamma \leq 2$
	4	Matern Class (MC)	$k_{Matern}(r) = h^2 \frac{2^{1-v}}{\Gamma(v)}\left(\frac{\sqrt{2v}r}{l}\right)^{v} K_v\left(\frac{\sqrt{2v}r}{l}\right); v,$
	5	Rational Quadratic (RQ)	$k_{RQ}(r) = h^2\left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}\ ;\ \alpha, l > 0$
	6	Piecewise Polinomial (PP)	$k_{ppD,0}(r) = h^2(1-r)_+^j$, $j = \left\lfloor\frac{D}{2}\right\rfloor + q + 1$ $k_{ppD,1}(r) = h^2(1-r)_+^{j+1}((j+1)r + 1)$, $k_{ppD,2}(r) = h^2(1-r)_+^{j+2}((j^2 + 4j + 3)r^2 + (3j+6)r + 3)/3$ $D > 0, q > 0$
Non-Stationary	7	Periodic	$k(r) = h^2 exp\left(-\frac{2sin^2\left(\frac{r}{2}\right)}{l^2}\right), l > 0$

In the Table 1, there are h , l , α , γ, v, q as hyper-parameters: h is an output-scale amplitude, l is an input scale (length, or time), α is known as an index that is equivalent to scale mixture of squared exponential kernels, $\Gamma(v)$ is Gamma function, $K_v$is a modified Bessel function of second order, hyper-parameter v controls degree of differentiability of the resultant function, and q is polynomial degree [11].

The basic aspects that can be defined through the covariance function are processes of stationarity, isotropic, fluidity and periodicity [6]. Stationary in this case refers to the process of 'behavior concerning the separation of two points tand t'. A process is said to be stationary when its covariance function depends on $t - t'$, whereas if it is non-stationary, it depends on the actual position of t and t'. If the process depends only on $r = \|\mathbf{t} - \mathbf{t}'\|$, Euclidean distance (no direction) between t and t', then the process is considered isotropic.

GAUSSIAN PROCESS TIME SERIES STATE SPACE MODELS

Discrete time series data are categorized in two model approaches, state space model and Autoregressive model. The difference between these two approaches are the state space model focused on estimation "state", while autoregressive depend on past observation. The idea of Gaussian processes time series state space models is using Gaussian process as transition mapping [12]. It means that Gaussian process is used on estimation "state". Gaussian process time series is continuous state space model, it can do estimation of state space in continue indices parameter. This is the advantage of Gaussian process in comparison with the discrete time series data. Illustration of Gaussian time series state space model is on figure 2.

Let $(y_{t_1}, y_{t_2}, \dots, y_{t_n})$ are data observation from input $t_1, t_2, \dots, t_N \in \mathbb{R}$, respectively. In Gaussian processes time series, time indices $t_i (i = 1,2,..,N)$ are used as inputs while $y_{t_i}$are outputs.

$$y_t = f(t) + \epsilon_t, \quad f \sim GP(0, k_\xi),\ \xi = (l, \alpha, \gamma, h, v) \quad ;\ \epsilon_t \sim N(0, \sigma_n^2) \tag{2}$$

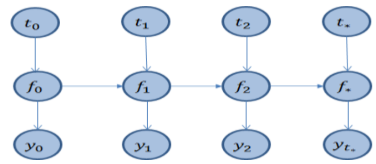


Figure 2. Illustration of Gaussian time series state space model. Estimation function $f_0, f_2, \dots, f_t$ are estimation "state" using Gaussian process.

Where ξ is GP hyper-parameter and σ_n^2 is noise variance. Thus, the mean function of Gaussian Process is set to zero. Gaussian prediction time series model from data $(y_{t_1}, y_{t_2}, \dots, y_{t_{n-1}})$ is below in equation (3),

$$p(y_{t_n} | y_{t_1}, y_{t_2}, \dots, y_{t_{n-1}}, \theta_m) = N(m_{t_n}, v_{t_n}) \tag{3}$$

where:

$m_{t_n} = k_*^T (K + \sigma_n^2 I)^{-1} \boldsymbol{y}; \boldsymbol{y} = [y_{t_1}, y_{t_2}, \dots, y_{t_{n-1}}]; k_* = k(y_{t_1:t_{n-1}}, y_t); K = k(y_{t_1:t_{n-1}}, y_{t_1:t_{n-1}})$

$v_{t_n} = k(y_{t_n}, y_{t_n}) - k_*^T (K + \sigma_n^2 I)^{-1} k_*$

$\theta_m \in \boldsymbol{\lambda} := (\boldsymbol{\xi}; \sigma_n^2)$

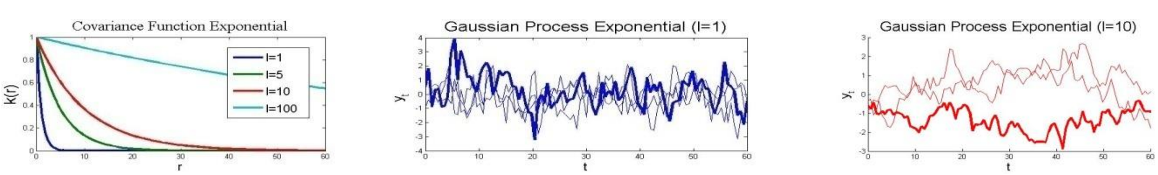
K is covariance matrix of Gaussian process which is constructed from input $t_1, t_2, \dots, t_N$. Distribution in equation (3) is called posterior distribution after considered to output $y_{t_1}, y_{t_2}, \dots, y_{t_{n-1}}$. This distribution can be used in prediction or forecasting of y_{t_n}value using m_{t_n}.

RESULTS

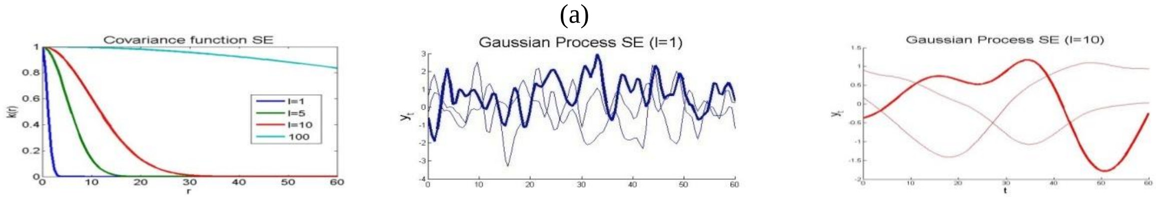
Firstly, we do simulation of prior GPTS to learn how features of prior GPTS. Prior GPTS is got by simulation of some hyper-parameter of covariance functions. Then we forecast with GPTS SSM.

1.2. Simulation of prior GPTS

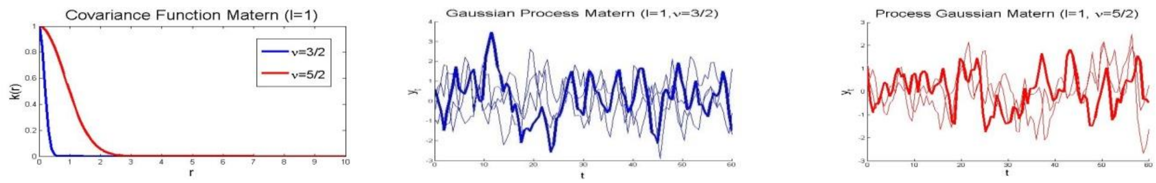
Simulation GPTS using some of variants hyper-parameter on covariance functions show below in figure 3. Here, we use output scale h=1.



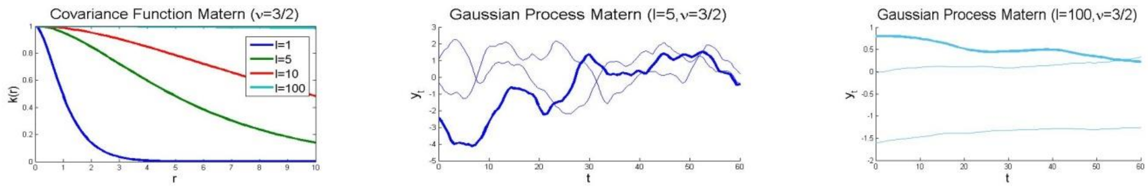
(a)



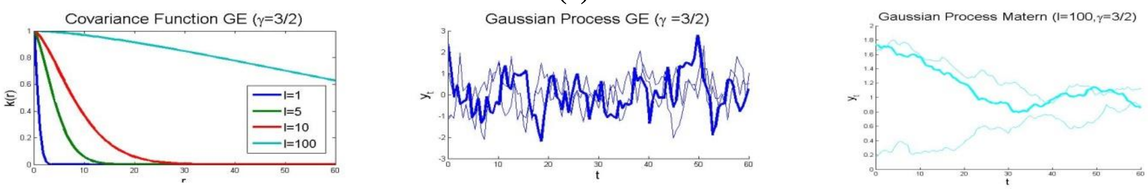
(b)


Covariance Function Matern (l=1)
ν=3/2
ν=5/2
Gaussian Process Matern (l=1, ν=3/2)
Process Gaussian Matern (l=1, ν=5/2)

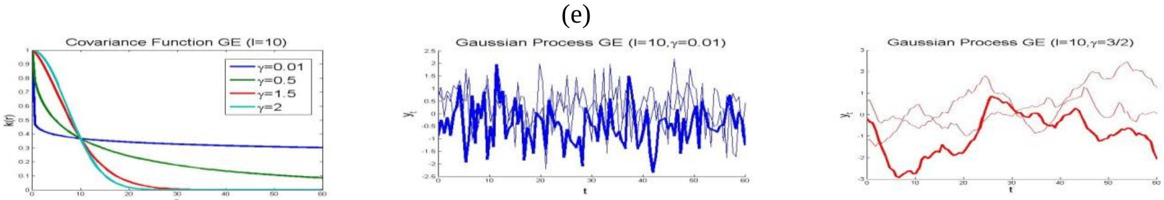
(c)


Covariance Function Matern (ν=3/2)
l=1
l=5
l=10
l=100
Gaussian Process Matern (l=5,ν=3/2)
Gaussian Process Matern (l=100,ν=3/2)

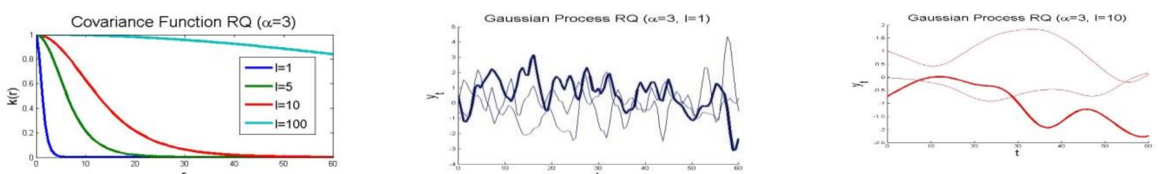
(d)


Covariance Function GE (γ=3/2)
l=1
l=5
l=10
l=100
Gaussian Process GE (γ =3/2)
Gaussian Process Matern (l=100,γ=3/2)

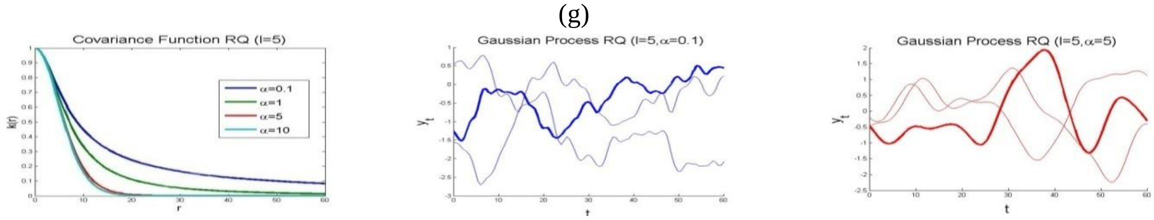
(e)


Covariance Function GE (l=10)
γ=0.01
γ=0.5
γ=1.5
γ=2
Gaussian Process GE (l=10,γ=0.01)
Gaussian Process GE (l=10,γ=3/2)

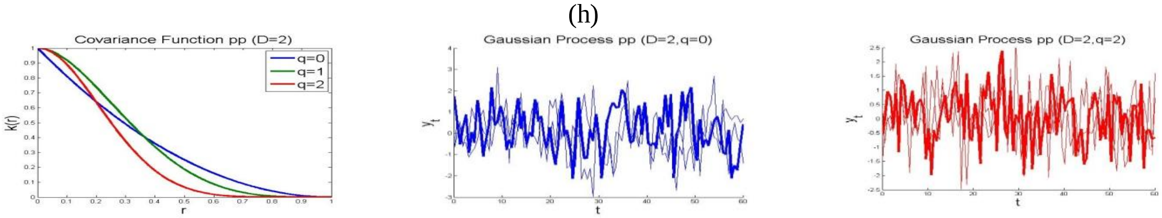
(f)


Covariance Function RQ (α=3)
l=1
l=5
l=10
l=100
Gaussian Process RQ (α=3, l=1)
Gaussian Process RQ (α=3, l=10)

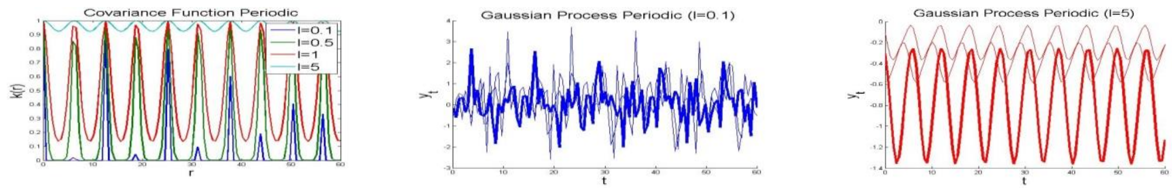
(g)


Covariance Function RQ (l=5)
α=0.1
α=1
α=5
α=10
Gaussian Process RQ (l=5,α=0.1)
Gaussian Process RQ (l=5,α=5)

(h)


Covariance Function pp (D=2)
q=0
q=1
q=2
Gaussian Process pp (D=2,q=0)
Gaussian Process pp (D=2,q=2)

(i)



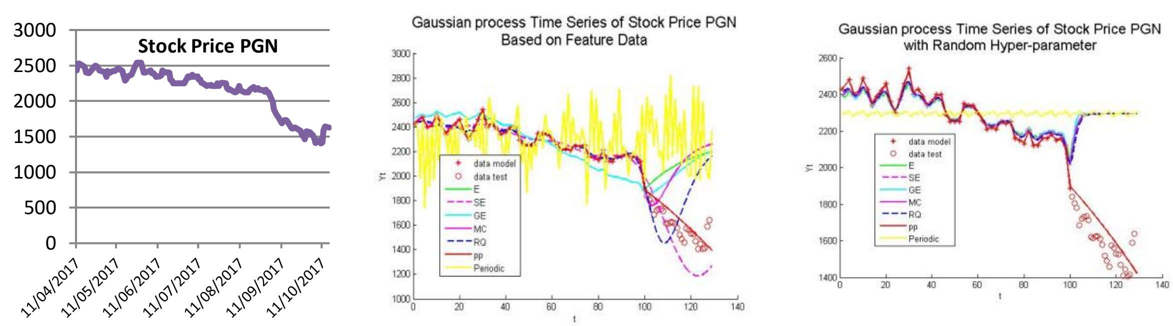
(j)

Figure 3. Covariance function curve with variants hyper-parameter are in figure (a), (b), (c), …, (j), left column. The middle and right column are Gaussian process simulation that covariance functions on the left side and use $\mu = 0$.

Covariance function defines the similarity and closeness between data point, these can be seen in Figure 3. In figure 3 show hyper-parameter manner in covariance function. Those influence the Gaussian process. Generally, Gaussian processes simulation using hyper-parameter l (input scale) increasingly makes two outputs in distance correspondingly, Gaussian process become smooth and existence of trend in time series. Increasing hyper-parameter ν on covariance function Matern in figure 3 (c) have no significant impact on Gaussian process simulation. Increasing hyper-parameter γ on covariance function GE in figure 3 (f) has impact to smoothness on Gaussian process and existence of trend in time series. Increasing hyper-parameter α on covariance function RQ in figure 3 (h) have impact on smoothness Gaussian process and small distance between time series value. Periodic covariance function is showed in figure 3 (j), increasing hyper-parameter l made decreasing of length periodicity. These mean that if we have data observations with periodicity in high time interval then we can use small hyper-parameter l, on contrary if data observations have periodicity in small time interval then we can use high hyper-parameter l.

1.3. Analysis GPTS for Stock Price of Perusahaan Gas Negara (PGN)

Figure 4 (a) are stock price data of Perusahaan Gas Negara (PGN) since 11 April 2017-16 October 2017, source: finance.yahoo.com. These time series data will be used to forecast stock price using GPTS SSM.



(a) (b) (c)

Figure 4. (a) Stock price of PGN in period 11 April-16 Oct 2017, (b) GPTS PGN based on feature data, (c) GPTS PGN using random hyper-parameter.

First, we learn feature of PGN data time series. Even though the stock prices are not smooth, it has a descending trend generally. First, choosing hyper-parameter for every covariance functions based on Gaussian processes simulation that have trend and fluctuation : $\boldsymbol{h = \sigma_{data}}$; $\boldsymbol{l_E}$ = 30; $\boldsymbol{l_{SE}}$ = 30; $\boldsymbol{l_{GE} = 30, \gamma = 1.5}$; $\boldsymbol{l_{MC} = 30}$; $\boldsymbol{l_{RQ} = 30, \alpha = 1}$; $\boldsymbol{l_{periodic} = 30}$. Hyper-parameter $\boldsymbol{h}$ is output scale with trend which variance output can influence more To GPTS than mean output, hence choose $\boldsymbol{h = \sigma_{data}}$. Thus, choosing hyper-parameter $\boldsymbol{l}$ based on length input forecasting are tested for length 30. Hyper-parameter $\boldsymbol{\gamma = 1.5}$ is choosed based on feature data that have trend. Hyper-parameter $\boldsymbol{\alpha = 1}$ is chosen based on data feature that has generally descending trend, it could be close to linear. And we choose covariance function pp that close to linear, $\boldsymbol{k_{ppD,0}}$. Prior Gaussian process time series model is built from 50 data models. Then we calculate prediction of 50 data model and do forecasting for the 30 data. Calculations of root mean square error (RMSE) for every GPTS based on random hyper-parameter (figure 4. (c)) and GPTS based on feature data (figure 4. (b)) are showed in table 2.

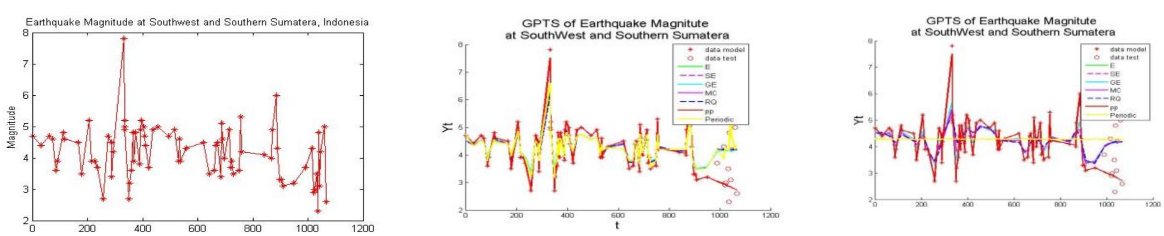
Table 2 show that RMSE of prediction GPTS SSM based on feature data are lower than GPTS SSM using random hyper-parameter. GPTS SSM using covariance function piecewise polynomial has lowest RMSE for both, hyperparameter based on feature data and random.

Table 2. Root Mean Square Error (RMSE) of Prediction GPTS PGN Stock Price

	GPTS based on feature data	GPTS random-1	GPTS random-2	GPTS random-3	GPTS random-4
E	$(l = 30)$	(h=2, l =2)	(h=3, l =4)	(h=4, l =6)	(h=40, l =0.01)
	227,1485	331.0524	321.5085	311.2283	344.3989
SE	$(l = 30)$	(h =2, l =2)	(h =3, l =4)	(h =4, l = 6)	(h =40, l = 0.01)
	100,2646	329.7020	316.7206	297.6668	344.3989
GE	$(l = 30, \gamma = 1.5)$	(h=2, l=2, $\gamma = 2$)	(h=3, l=4, $\gamma = 2$)	(h=4, l=6, $\gamma = 2$)	(h=40, l=0.01 $\gamma = 2$)
	153,1967	332.3003	329.0423	312.2935	334.5973
MC	$(l = 30, \nu = 1)$	(h=2,l=2, $\nu = 1$)	(h=3,l=4, $\nu = 1$)	(h=4,l=6, $\nu = 1$)	(h=40,l=0.01, $\nu = 1$)
	72,1625	330.2274	319.2657	304.7188	344.3989
RQ	$(l = 30, \alpha = 1)$	(h=2, l=2, $\alpha = 3$)	(h=3, l=4, $\alpha = 3$)	(h=4, l=6, $\alpha = 3$)	(h=40, l=0.01, $\alpha =$ 9)
	62.2795	328.1890	312.9686	291.3520	344.3989
PP	q=1, D=2	(h =2 ,q=1)	(h =3 ,q=1)	(h =4 ,q=1)	(h =40, l =0.01, q=1)
	52,9262	54.6277	53.6292	53.3036	52.9108
Periodic	$(l = 0.001)$	(h=2, l =2)	(h=3, l =4)	(h=4, l =6)	(h=40, l =0.01)
	339,5159	353.5434	353.5437	353.5692	339.5000

1.4. GPTS for Earthquake of Southwest and Southern Sumatera

We use the earthquake magnitude data at Southwest Sumatera and Southern Sumatera in Indonesia. The data are in figure 5 (a), from 4 January 2015 until 3 January 2018 (source: www.bmkg.go.id). Sumatera is one of the island in Indonesia that has high tendency to earthquake. That is caused of Sumatra is an active tectonic region in which there is a subduction zone in southwest Sumatra [13]. BMKG has wrote that there are 81 earthquakes in 4 January 2015-3 January 2018. We apply some covariance functions in Table 1 to prior Gaussian process time series from 70 test data. We calculate some prediction by equation (3). We choose some random hyper-parameters for every covariance function. It is done to compare between the influence hyper-parameter randomly and w hyper-parameter based on feature data. We calculate the root mean square error (RMSE) to see performance, it is showed in Table 3.



(a) (b) (c)

Figure 5. (a) Earthquake Magnitude at Southwest and Southern Sumatera in Indonesia, (b) GPTS of Earthquake magnitude at Southwest and Southern Sumatera in Indonesia using hyper-parameter based on feature data. (c) GPTS of Earthquake magnitude at Southwest and Southern Sumatera in Indonesia using random hyper-parameter.

Feature data time series of earthquake magnitude is shown in figure 5 (a). It shows no trend and rough-fluctuate. First, choosing hyper-parameter for every covariance functions based on Gaussian processes simulation that have no trend and fluctuation: $h = 5.02$; l_E=1; l_{SE}=1; $l_{GE} = 1, \gamma = 0.1$; $l_{MC} = 1$; $l_{RQ} = 1, \alpha = 1$; $l_{periodic} = 1$. Hyper-parameter h as output scale of time series with no trend, hence we choose $h = \sigma_{magnitude} + \mu_{magnitude} = 5.02$. Those hyper-parameter l are chosen based on fluctuate with no trend, then smaller l can do good prediction as on simulation prior GPTS in figure 3. Hyper-parameter $\gamma = 0.1$ is choosed based on feature data that have rough feature, smaller γ can give rough simulation prior GPTS. Hyper-parameter is chosen $\alpha = 0.5$ for rough data time series. For simplicity, we choose covariance function pp, $k_{ppD,0}$. Prior Gaussian process time series model is built from 50 data models. Figure 5 (b) shows the GPTS SSM use hyper-parameter based on feature data and figure 5 (c) shows the GPTS SSM by random hyper-parameter.

Table 2. Root Mean Square Error (RMSE) of Prediction GPTS PGN Stock Price

	GPTS based on feature data	**GPTS random-1**	**GPTS random-2**	**GPTS random-3**	**GPTS random-4**
E	$(h = 5; l = 1)$ 0.6310	(h=2, l =2) 0.6595	(h=2, l =30) 0.7147	$(h = 1000; l = 0.001)$ 0.6246	(h=10, l =0.01) 0.6272
SE	$(h = 5;\ l = 1)$ 0.6310	(h =2, l =2) 0.6609	(h =2, l = 30) 0.7962	$(h = 1000; l = 0.001)$ 0.6246	(h =10, l = 0.01) 0.6272
GE	$(h = 5; l = 1; \gamma = 0.1)$ 0.6310	(h=2, l=2, $\gamma = 2$) 0.6590	(h=2, l=30, $\gamma = 0.5$) 0.7147	$(h = 10^3; l = 10^{-3}, \gamma = 0.1)$ 0.6246	(h=10, l=0.01, $\gamma = 2$) 0.6272
MC	$(h = 5; l = 1, \nu = 1)$ 0.6310	(h=2,l=2, $\nu = 1$) 0.6600	(h=2,l=30, $\nu = 1$) 0.7639	$(h = 10^3; l = 10^{-3}, \nu = 1)$ 0.6246	(h=10,l=0.01, $\nu = 1$) 0.6272
RQ	$(h = 5; l = 1, \alpha = 0.1)$ 0.6284	(h=2, l=2, $\alpha = 3$) 0.6717	(h=2, l=30, $\alpha = 1$) 0.7725	$(h = 10^3; l = 10^{-3}, = 0.1)$ 0.6246	(h=10, l=0.01, $\alpha = 9$) 0.6272
PP	$(h = 5$; q=1,D=2) 0.6698	(h =2, q=1) 0.6703	(h =2, l =30,q=1) 0.6704	$(h = 1000$; q=1;D=2) 0.6700	(h =10,l =0.01,q=1) 0.6698
Periodic	$(h = 5;\ l = 1)$ 0.8471	(h=2, l =2) 0.8841	(h=2, l =30) 0.9138	$(h = 1000; l = 0.001$ 1.4521	(h=10, l =0.01) 1.1039

Conclusion

Choosing hyper-parameter for GPTS SSM based on feature data have smaller RMSE or close to small RSME by choosing hyper-parameter randomly. Simulation of prior GPTS for every covariance function can figure prior GPTS and show the feature of prior GPTS. Simulation can be the illustration before choosing hyper-parameter based on feature data.

References

[1] Kleppea T S and Oglend A 2017 Estimating the competitive storage model: A simulated likelihood approach *Econometrics and Statistics.* **4** pp 39-56

[2] Peia O and Roszbachc K 2015 Finance and growth: time series evidence on causality *J. of Financial Stability*. **19** pp 105-118

[3] Reikard G, Haupt S E, and Jensen T 2017 Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models *Renewable Energy*. **112** pp 474-485

[4] Vogel E E, Saravia G, Pastén D, and Muñoz V 2017 Time-series analysis of earthquake sequences by means of information recognizer *Techtonophysics*. pp 723-728

[5] Frigola R and Alcalde 2015 *Bayesian Time Series with Gaussian Processes.* (Cambridge: University of Cambridge)

[6] Rasmussen C E and Williams K I 2006 *Gaussian Process for Machine Learning*. (London: MIT Press) pp 79-102

[7] Murray R, Smith and Girard A 2001 Gaussian process priors with ARMA noise models *Proceeding of Irish Signals and Systems Conference.* (Ireland) pp 147–152

[8] Turner R 2011 *Gaussian Processes for State Space Models and Change Point Detection*. (Cambridge: University of Cambridge)

[9] Wilson A C 2014 *Covariance Kernels for Fast Automatics Pattern Discovery and Extrapolation with Gaussian Processes*, (Cambridge: University of Cambridge)

[10] Chen Z, and Wang B 2018 How priors of initial hyperparameters affect Gaussian process regression models *Neurocomputing*. **275** pp 1702-1710

[11] Roberts S, Osborn M, Ebden M, Reece S, Gibson N et al 2013 Gaussian processes for time-series modelling *R. Society*. A p 371

[12] Eleftheriadis S, Nicholson T F W, Deisenroth M P, and Hensman J 2017 Identification of Gaussian process state space models *arXiv: 1705.10888v2*

[13] PusGen,PusatLitbang Perumahan dan pemukiman 2017, *Peta Sumber dan Bahaya Gempa Indonesia Tahun 2017*.p 182