ON SELECTION OF PERIODIC KERNELS PARAMETERS IN TIME SERIES PREDICTION

Marcin Michalak

Institute of Informatics, Silesian University of Technology,
ul. Akademicka 16, 44-100 Gliwice, Poland
Marcin.Michalak@polsl.pl

ABSTRACT

In the paper the analysis of the periodic kernels parameters is described. Periodic kernels can be used for the prediction task, performed as the typical regression problem. On the basis of the Periodic Kernel Estimator (PerKE) the prediction of real time series is performed. As periodic kernels require the setting of their parameters it is necessary to analyse their influence on the prediction quality. This paper describes an easy methodology of finding values of parameters of periodic kernels. It is based on grid search. Two different error measures are taken into consideration as the prediction qualities but lead to comparable results. The methodology was tested on benchmark and real datasets and proved to give satisfactory results.

KEYWORDS

Kernel regression, time series prediction, nonparametric regression

1. INTRODUCTION

Estimation of a regression function is a way of describing a character of a phenomenon on the basis of the values of known variables that influence on the phenomenon. There are three main branches of the regression methods: parametric, nonparametric, and semiparametric. In the parametric regression the form of the dependence is assumed (the function with the finite number of parameters) and the regression task simplifies to the estimation of the model (function) parameters. The linear or polynomial regression are the most popular examples. In the nonparametric regression any analytical form of the regression function can be assumed and it is built straight from the data like in Support Vector Machines (*SVM*), kernel estimators, or neural networks. The third group is the combination of the two previously described. The regression task in this case is performed in two steps: firstly the parametric regression is applied followed by the nonparametric.

Time series are a specific kind of data: the observed phenomenon depends of some set of variables but also on the laps of time. The most popular and well known methods of time series analysis and prediction are presented in [1] which first edition was in 60's of the 20th century.

In this paper the semiparametric model of regression is applied for the purpose of time series prediction. In the previous works kernel estimators and *SVM* were used for this task [2][3] but

David C. Wyld et al. (Eds) : CST, ITCS, JSE, SIP, ARIA, DMS - 2014
pp. 01–10, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4101

these methods required mapping of the time series into a new space. Another approach was presented in [4] where the Periodic Kernel Estimator (*PerKE*) was defined. It is also the semiparametric algorithm. In the first step the regression model is built (linear or exponential) and for the rests the nonparametric model is applied. The final prediction is the compound of two models. The nonparametric step is the kernel regression with the specific kind of kernel function called periodic kernel function. In the mentioned paper two kernels were defined.

Because each periodic kernel requires some parameters in this paper the analysis of the influence of kernel parameters on prediction error becomes the point of interest. The paper is organized as follows: it starts from ashort description of prediction and regression methods, then the *PerKE* algorithm is presented. Afterwards, results of the experiments performed on time series are given. The paper ends with conclusions and the description of further works.

2. PREDICTION AND REGRESSION MODELS

2.1. ARIMA (SARIMA) Models

SARIMA(*Seasonal ARIMA*) model generalizes the Box and Jenkins *ARIMA* model (*AutoRegressive Integrated Moving Average*)[1] as the connection of three simple models: autoregression (*AR*), moving average (*MA*) and integration (*I*).

If B is defined as the lag operator for the time series x ($Bx_t = x_{t-1}$) then the autoregressive model of the order p(a_t is the white noise and will used also in other models) is given by the formula:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + a_t$$

and may be defined as:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p) x_t = a_t$$

In the *MA* models the value of time series depends on random component a_t and its q delays as follows:

$$x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

or as:

$$a_t\left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) = x_t$$

For the non-stationary time series the d operation of its differentiation is performed, described as the component $(1 - B)^d$ in the final equation. The full *ARIMA*(p, d, q) model takes the form:

$$\left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots \varphi_p B^p\right)(1 - B)^d x_t = a_t\left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$$

The *SARIMA* model is dedicated for time series that have strong periodic fluctuations. If s is the seasonal delay the model is described as *SARIMA*(p, d, q)(P,D,Q)^s where P is the order of seasonal autoregression $\left(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots \Phi_p B^{Ps}\right)$, Q is the order of seasonal moving average $\left(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots \Theta_p B^{Ps}\right)$and D is the order of seasonal integration $(1 - B^s - B^{2s} - \cdots B^{Ps})$.

2.2. Decomposition Method

This method tries to separate several components of the time series, each of them describing the series in the different way. Most important components are as follows:

- trend component (*T*): the long-time characteristic of the time series,
- seasonal component (*S*): the periodic changes of the time series,
- cyclical component (*C*): repeated but non-periodic changes of the time series,
- irregular (random) component (*e*).

Components are usually aggregated. It may be an additive aggregation when the final predicted value is a sum of all time series components or multiplicative aggregation when the final value is calculated as a multiplication of all time series components. First is called additive and the final predicted value is the sum of component time series values and the second is called multiplicative (aggregation is the multiplication of time series values).

2.3. Periodic Kernels

Periodic kernels belong to a wide group of kernel functions that are applied for the task of estimation of the regression function. They fulfil the typical conditions for the kernel function and some of the specific ones. As the most important typical features of the kernel function the following should be mentioned [5]:

- $\int_R K(u)du = 1$
- $\forall u \in R\ K(u) = K(-u)$
- $\int_R uK(u)du = 0$
- $\forall u \in R\ K(0) \geq K(u)$
- $\int_R u^2 K(u)du < \infty$

Furthermore, if we assume that the period of the analysed time series is *T*then there are the following specific conditions for the periodic kernel function:

- for each$k \in Z$ the value *K*(*kT*) is the strong local maximum,
- for each$x \in R \setminus \{0\}$ $K(0) > K(x)$,
- for each$n_1, n_2 \in N$ that $n_1 < n_2$ $K(n_1) > K(n_2)$.

In the paper [4] two periodic kernels were defined, named First Periodic Kernel (*FPK*) and Second Periodic Kernel (*SPK*). The formula of *FPK* is the multiplication of the exponential function and the cosine:

$$FPK(x) = \frac{1}{C} e^{-a|x|}(1 + \cos bx)$$

The constant *C* assures that *K* is integrable to one. This value depends on the values *a*and *b* as follows:

$$C = 2\int_0^\infty e^{ax}(1 + \cos bx)dx = \frac{4a^2 + 2b^2}{a(a^2 + b^2)}$$

In other to define the *FPK*it is to substitute *a* and *b* with the period *T* and parameterθthat is a function attenuation (the ratio of the two consecutive local maxima):

$$b = \frac{2\pi}{T}$$

$$\theta = \frac{K(t+T)}{K(t)} \Rightarrow -aT = \ln\theta \Rightarrow a = -\frac{\ln\theta}{T}$$

Based on this substitution the following formula is obtained:

$$K(x) = \frac{1}{C} e^{\frac{\ln\theta}{T}|x|}\left(1 + \cos\frac{2x\pi}{T}\right)$$
$$C = \frac{4T\ \ln^2\theta + 4T\pi^2}{-\ln^3\theta - 4\pi^2\ln^2\theta}$$

On the Figure 1. the sample *FPK* is presented.

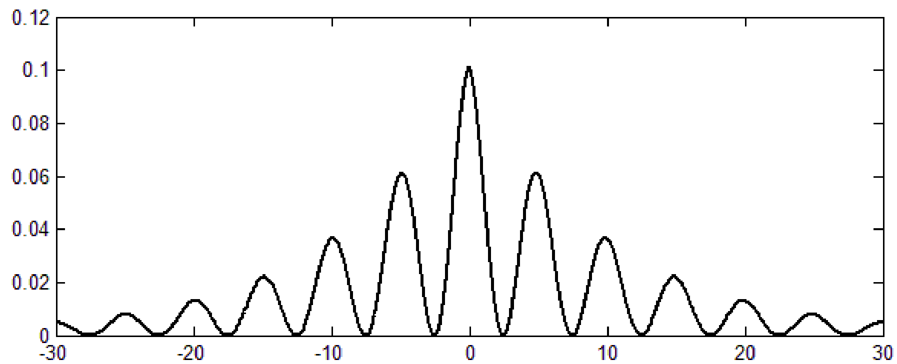

Figure 1. First Periodic Kernel generated with$T$=5, θ = 0.6

The second kernel (*SPK*) has a following formula:

$$SPK(x) = \frac{1}{C} e^{-a|x|} \cos^{\mathrm{n}}\ bx$$

where

$$C = 2\int_0^\infty e^{ax} \cos^{\mathrm{n}} bx\, dx = 2[I_n]_0^\infty$$

and I_n is an integral:

$$I_n = \int e^{ax} \cos^n bx$$

The final formula for the constant C calculated recurrently is following:

$$C = \left(-\frac{1}{a} - \sum_{i=1}^{n} \frac{a}{(a^2 + 4i^2)\prod_{k=0}^{i}\mu_i}\right)\prod_{i=1}^{n}\mu_i$$

with

$$\mu_0 = 1,\quad \mu_i = \frac{2i(2i-1)}{a^2 + 4i^2}$$

It is possible to calculate the value of the C in the analytical way when the software allows symbolic calculation. Experiments presented in this paper were performed in Matlab and the C was calculated in the symbolic way.

This kernel also may be defined with the period T and the attenuationθ:

$$K(x) = \frac{1}{C} e^{\frac{\ln\theta}{2}|x|} \cos^{\mathrm{n}} \frac{\pi x}{T}, \quad b(T) = \frac{\pi}{T}, \quad a(\theta) = -\frac{\ln\theta}{T}$$

The role of *n* parameter is to describe the „sharpness" of the function in the local maxima. On the Figure 2. the sample *SPK* is given.

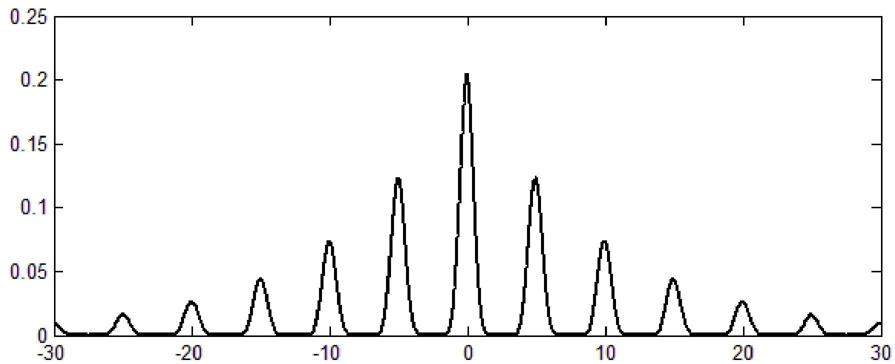

Figure 2. Second Periodic Kernel generated with *T*=5, θ = 0.6 and *n* = 10

3. PERKE ALGORITHM

*Periodic Kernel Estimator*is a member of the group of semiparametric (two step) methods [6][7]. Methods from this group consist of the initial parametric step and the final nonparametric one. After the parametric step the residuals are calculated and the nonparametric part of the model tries to explain the only variation of residuals. The final model can consists of addition or multiplication of the basic results. In this aspect the semiparametric method is similar to the decomposition method.

The *PerKE* models the residual part of the time series with the following formula:

$$x(t) = \frac{\sum_{i=1}^{k} x_{t-i} K(t-i)}{\sum_{i=1}^{k} K(t-i)}$$

where *k* is the number of previous observation in the train sample of the time series.

It may be noticed that this equationis derived from the Nadaraya-Watson kernel estimator [8][9]but the smoothing parameter *h* was removed. This may cause the situation of oversmoothing the data. It is observed in two variants: the predicted values are overestimated (bigger than real values) or overestimated (smaller than real values). In order to avoid this situation the parameter called underestimation α is introduced. It is the fraction of the predicted and original value:

$$\alpha_i = \frac{\tilde{x}_i}{x_i}$$

The underestimation is trained in the following way: if *p* is an interesting prediction horizon the last *p* observations from the train set are considered as the test set and predict them on the basis of the rest of the train set. Then the vector of underestimations is defined as the vector of fractions of predicted and real values. In the final prediction the values coming from the nonparametric step are divided by the corresponding α.

4. SELECTION OF KERNEL PARAMETERS

4.1. Discretisation of Periodic Kernels

In the experiments a simplified –a discretized – form of periodic kernels was used. Let assume that only the values of the kernel for the period multiple are interesting: $K(x)$ where $x = kT, k \in Z$. Then the formula for *FPK* simplifies to the following one:

$$K(kT) = \frac{2}{C} e^{|k| \ln \theta}$$

Discretisation of the *SPK* leads to the same formula. The only difference between two discretized kernels is the value of the C constant which can be tabularised before the experiments. It speeds up calculation because each constant C (for each demanded form of periodic kernel) was calculated once and was read in a constant time.

On the basis of the discretized form of periodic kernels and the kernel regression formula of residual part of the series, it might be claimed, that both types of periodic kernels give the same results.

4.2. The Error Evaluation

The error of prediction was measured with two different quality functions:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{|y_i - \tilde{y}_i|}{|y_i|} \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}$$

Each of them describes a different kind of an error. The first one points the averaged absolute error and is more resistant when the test samples have values from very wide range. The second one measures the error in the unit of the analysed data so it can be more interpretable in some cases.

4.3. Setting the Parameters for *SPK*

Let`s consider the very popular time series describing the number of passengers in America (*G* series from Box and Jenkins [1]). It contains 144 monthly values of number of passengers (in millions) between 01.1949 and 12.1960. Its natural period is 12. This time series is presented on the Figure 3.

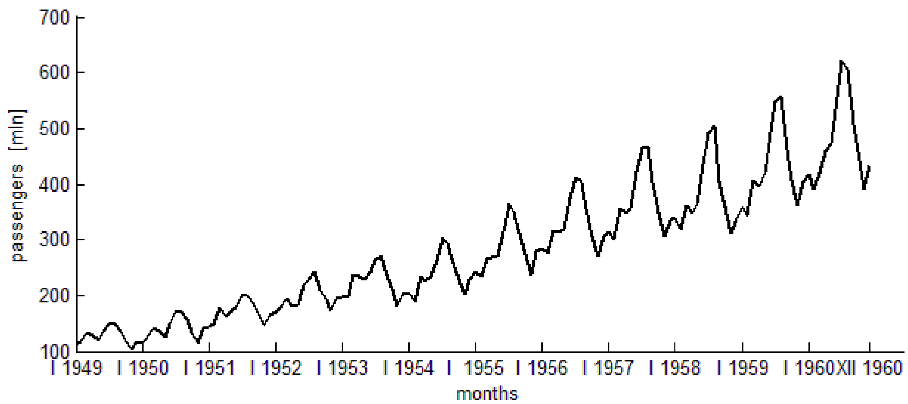

Figure 3. *G* time series

For the purpose of the analysis of an influence of the *SPK* parameters on the prediction accuracy the following optimization step was performed. Instead of calculation of the *C* value for each prediction task, the array of *C* values for the predefined periodic kernel parameters was created. The attenuation was changing from $\theta = 0.1$ to $\theta = 0.9$ with the step 0.1. The sharpness was changing from $n = 2$ to $n = 60$ with the step 2.

The error of the prediction depending on the kernel parameters is shown on the Figure 4.

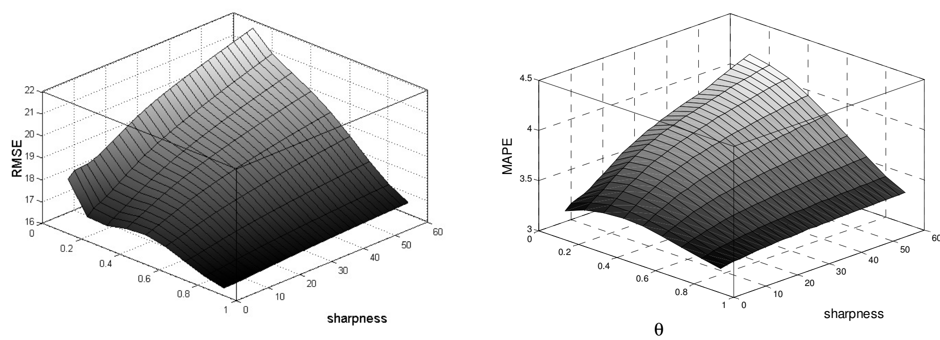

Figure 4. *G* time series prediction error (*MAPE* on the left and *RMSE* on the right) as the function of θ and sharpness

In general, it may be seen that the error of the prediction decreaseswhen the θincreases. Additionally, it is observed that the influence of the sharpness is opposite. In other wordsthe decrease of the sharpness implies the decrease of the error.

Because the period of this series is 12 (the number of months) periodic kernel parameters were established on the basis of prediction on 144 – 12 *G* series values (all data without the last 12 values). Both error measures were considered. The smaller time series were called train series.

Table 1 compares the errors on the train series and on the whole series. The best results (typed with bold font) for the train series were for $\theta = 0.9$ and sharpness = 2. Performing the grid experiment for the whole series the best results were for 0.9 and 2 (with *MAPE*) and for 0.9 and 4 (with *RMSE*) respectively. It can be seen, that on the basis of the *MAPE* results for train data the best values of parameters (with the assumed grid steps) were found and with the *RMSE* results – almost the best.

Table 1. Comparison of best results and kernel parameters for train and whole time series.

Train series			
θ	**sharpness**	***MAPE***	***RMSE***
0.9	2	**3.6877**	**17.8085**
0.9	4	3.6938	17.8752

Whole series			
θ	**Sharpness**	***MAPE***	***RMSE***
0.9	2	**3.1989**	16.1038
0.9	4	3.2084	**16.0972**

5. REAL DATA APPLICATION

Selection of periodic kernel parameters was applied for the real time series, describing the monthly production of heat in one of the heating plant in Poland. This series (denoted as *E*) contained 97 values. The series is presented on the Figure 5.

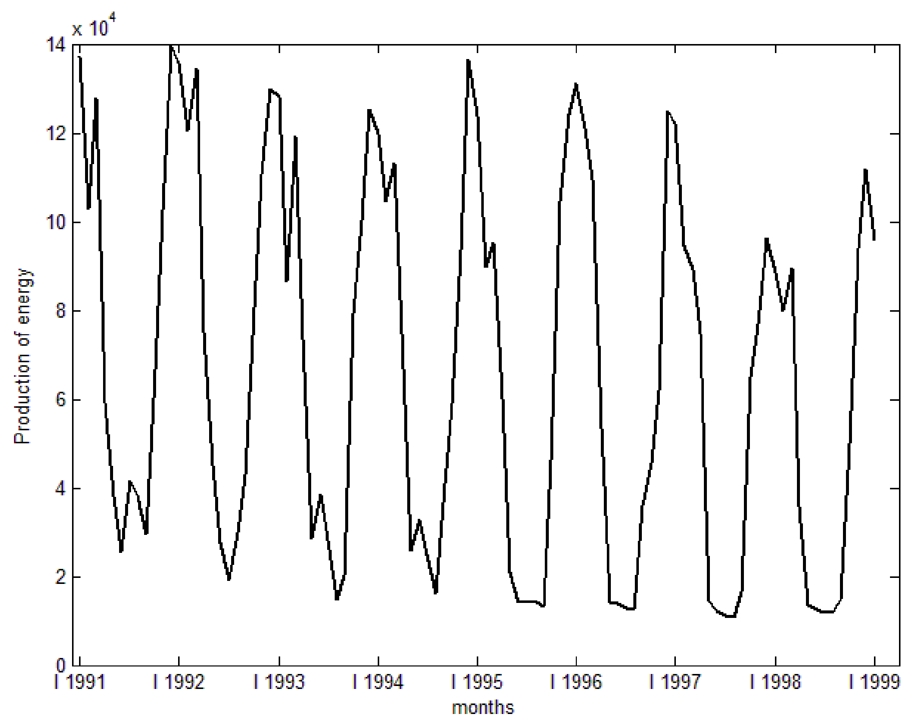

Figure 5. *E* time series prediction – monthly production of heat.

PerKE algorithm was performed in three ways: periodic kernels with arbitrarily set kernel parameters (two types of periodic kernels) and the *SPK* with the presented methodology of parameters setting. Additionally, two popular time series prediction methods were used as the reference points for kernel prediction results: SARIMA and decomposition method.

The results of all experiments are shown in the Table 2. (*G* series) and Table 3. (*E* series). In the first case periodic kernel parameters did not depend on the chosen measure. The final prediction quality is still better than the quality of other popular prediction methods.

Table 2. Comparison of the *G* time series prediction results.

method	*MAPE*	*RMSE*	annotations
SARIMA	4.80%	26.95	$(1,0,0)(2,0,0)^{12}$
decomp.	4.51%	26.60	exponential+multiplicative
FPK	3.20%	16.10	
SPK	3.72%	21.00	T=12, θ=0.4, n =60
SPK(*MAPE/RMSE*)	3.20%	16.10	T=12, θ=0.9, n =2

Table 3. Comparison of the *E* time series prediction results.

method	*MAPE*	*RMSE*	annotations
SARIMA	20.95%	10 115.91	$(1,0,0)(2,0,0)^{12}$
decomp.	22.10%	9 010.87	linear+additive
FPK	69.13%	19 855.28	
SPK	20.08%	8 638.12	T=12, θ=0.9, n =80
SPK(*MAPE*)	19.13%	14 735.66	T=12, θ=0.9, n =2
SPK(*RMSE*)	18.26%	15 861.22	T=12, θ=0.1, n =2

In the second case (*E* series) the selected set of periodic kernel parameters depended on the quality measure. But for each of them the decrease of relative error is observed.

6. CONCLUSIONS AND FURTHER WORKS

In the paper the analysis of the periodic kernel parameters influence on the prediction error was analysed. Two types of periodic kernels were taken into consideration and the error of the prediction was measured with two different methods. On the basis of the analysis of the *G* time series and the *E* time series it may be said that the methodology of finding the periodic kernel parameters gives satisfying results.

Further works will focus on the application of *PerKE* and periodic kernels to time series with the different time interval between observations. It is expected that more differences between the two kernels will occur. It is also possible that the sharpness will have the bigger influence on the prediction error.

ACKNOWLEDGEMENTS

This work was supported by the European Union from the European Social Fund (grant agreement number: UDA-POKL.04.01.01-106/09).

REFERENCES

[1] Box, George & Jenkins, Gwilym (1970)Time series analysis. Holden-Day, San Francisco.
[2] Michalak, Marcin (2011)“Adaptive kernel approach to the time series prediction”,Pattern Analysis and Application, Vol. 14, pp. 283-293.
[3] Michalak, Marcin (2009)“Time series prediction using new adaptive kernel estimators”,Advances in Intelligent and Soft Computing, Vol. 57, pp. 229-236.
[4] Michalak, Marcin (2011)“Time series prediction with periodic kernels”,Advances in Intelligent and Soft Computing, Vol. 95, pp. 137-146.
[5] Scott, David (1992)Multivariate Density Estimation. Theory, Practice and Visualization,Wiley & Sons.
[6] Abramson, Ian (1982) “Arbitrariness of the pilot estimator in adaptive kernel methods”,Journal of Multivariate Analysis, Vol. 12, pp. 562-567.

[7] Hjort, Nils & Glad, Ingrid (1995)"Nonparametric density estimation with a parametric start", Annals of Statistics, Vol. 23, pp. 882-904.
[8] Nadaraya, Elizbar (1964)"On estimating regression",Theory of Probability and Its Applications,Vol. 9, pp.141-142.
[9] Watson, Geoffrey (1964)"Smooth regression analysis",Sankhya - The Indian Journal of Statistics,Vol. 26, pp. 359-372.

AUTHOR

Marcin Michalak was born in Poland in 1981. He received his M.Sc. Eng. in computer science from the Silesian University of Technology in 2005 and Ph.D. degree in 2009 from the same university. His scientific interests is in machine learning, data mining, rough sets and biclustering. He is an author and coauthor of over 40 scientific papers.