VOL. 14, NO. 15, AUGUST 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

STATISTICAL MODELLING FOR PREDICTION OF RICE PRODUCTION IN INDONESIA USING SEMIPARAMETRIC REGRESSION BASED ON THREE FORMS OF FOURIER SERIES ESTIMATOR

M. Fariz Fadillah Mardianto, Eko Tjahjono and Marisa Rifada
Department of Mathematics, Airlangga University, Surabaya, Indonesia
E-Mail: m.fariz.fadillah.m@fst.unair.ac.id

ABSTRACT

The primary food for Indonesian is rice. The rice production in Indonesia tends to decrease that influencing the increase of rice's price. The increase of rice's price has a big influence for inflation, like as noted by Indonesia Central Bureau of Statistics, in January 2018 inflation happened with rate 0,62% as the number 0,24% from increase of rice's price. Thus, the availability of rice must be controlled with make prediction for developing sustainable agriculture in Indonesia. In this study, we compared three forms of semiparametric regression method based on Fourier series estimator to make prediction of rice production for every province in Indonesia. The result is the best model for predict of rice production is modeled based on the best semiparametric regression based on Fourier cosine series estimator. The model met goodness criteria like the smallest MSE equals to 0,0000194 and the biggest determination coefficient near to 100%.

Keywords: rice production, prediction, semiparametric regression, fourier series estimator.

1. INTRODUCTION

Indonesia is an agrarian country that has rich resources of agriculture. One of the main agriculture resources is rice. Rice is a primary food for most of Indonesians, so it is important and the supply of rice must be controlled for national food sustainability. The problem is the rice plant cannot producing all of the time although advance agriculture's technology has developed. The period of great harvest for rice usually happened in February until June. There are some constrains that cause the great harvest for rice is not optimal, like climate, weather change, area of land change, land's productivity, human resource, pest factor and fertilizer availability (Chauhan *et al*., 2017).

Rice is the staple food of the Indonesian people which is difficult to replace. In the event of scarcity and increase in these commodities, this will result in national economic and political stability. For example, an increase in rice prices can make a high contribution to inflation. At the end of 2017 to January 2018 the increase in rice prices is one of the problems faced by the Indonesian people. In January 2018, Indonesia Central Bureau of Statistics recorded an inflation of 0,62%, with the largest contributor being an increase in rice prices. The increase in rice prices contributed 0,24% to national inflation (Central Bureau of Statistics, 2018). According to Indonesia Central Bureau of Statistics, the highest rice production cycle occurs in March to April which is the main rice harvest season in Indonesia. The rice harvest cycle in Indonesia is a periodic cycle. The condition of crop failure or the amount of national rice availability can result in little national food security. The problem of rice production in Indonesia is the comparison of rice production and national rice demands.

Although rice production in Indonesia has increased, this increase has not been able to compensate for the increase in national rice demand. For example, in 2010, national rice production amounted to 43.520 tons, in the same year the national rice demand reached 44,762 tons (Natawidjaja and Rum, 2012). Logically, it can lead for decreasing the number of Indonesia rice exports that presents, there is a trend towards the decline in Indonesia rice exports, supported by data from Indonesia Central Bureau of Statistics at 2017. In addition, the area of rice fields in Indonesia is decreasing, in 2013 the area of rice fields in Indonesia reached 8.128.499 hectares, in 2014 it reached 8.111.593 hectares, and in 2015 reached 8.087.393 hectares (Central Bureau of Statistics, 2017). In this discussion, information that be obtained, the problem of rice production in Indonesia is a reduction in agricultural land, and an increasing number of the rice demands. One factor increasing the need for rice is the increasing population. On the other hand, the Ministry of Agriculture is trying to increase the supply of rice to reach the target of 2045, Indonesia being the world's food barn. To achieve these targets requires planning, predicting, and meeting targets that are not easy. Thus, the availability of rice must be controlled with make prediction. Prediction result of rice production based on the productivity of agricultural land and the extent of agricultural land with observation every province in Indonesia can be used as an evaluation for government's policy about food sustainability, such as importing rice, giving information about the program of agriculture's intencification and extecification for achieving the peak of rice production target, specially in granary area.

Prediction of rice production in Indonesia based on the productivity of agricultural land and the extent of agricultural land in this research using regression approach. Regression analysis is a method in Statistics that used for modelling causality relation between response variable and predictor variables (Tjahjono *et al.,* 2018). In regression analysis, regression curve's shape can be substituted with mathematical's functions which corresponds to the data pattern. We use combination of linear and Fourier series function in additive form to

VOL. 14, NO. 15, AUGUST 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

estimate regression curve. This combination is called semiparametric regression with Fourier series estimator (Budiantara *et al.,* 2015). In this study we are not only use Fourier cosine series that be proposed by Bilodeau (1992), but also using Fourier series which include cosine and sine basis that be proposed by Biedermann *et al*., (2009) and Dette *et al*., (2016). The other estimator that be used is Fourier sine estimator, a new estimator in semiparametric regression based on Fourier series estimator. The purpose is presenting the performance of three form Fourier series estimator in semiparametric regression for modeling an application.

This study makes a model for prediction rice production in Indonesia with the best Fourier series estimator. The best Fourier series estimator is obtained by comparing the smallest Generalized Cross Validation (GCV) values for certain oscillation parameters obtained for three Fourier series estimators including cosines, sines, and estimators that contain both. This study can be done based on previous literature. Semiparametric regression with Fourier series estimator with a Fourier cosine series, has good performance to model rice production in Central Java Province, Indonesia (Asrini and Budiantara, 2014). In this study the relationship between the production of rice and the extent of agricultural land has linear pattern. The relationship between the production of rice and the productivity of agricultural land has fluctuate pattern, so it can be estimated with Fourier series estimator in regression.

2. MATERIALS AND METHODS

2.1 Data source

The data that be used in this study is secondary data about rice production as a response variable, the extent of agricultural land as a predictor variable that be estimated with linear function, and the productivity of agricultural land that be estimated with Fourier series function. The data is taken from Indonesia Central Bureau of Statistics. The data is divided as training and testing data. The training data is data that be published at 2016 which be observed at 2015 for making estimation in regression model. The testing data is data that published at 2017 which be observed at 2016 for making prediction based on estimation result in regression model. The observation area is all of provinces in Indonesia, so there are 34 provinces.

2.2 Semiparametric regression based on Fourier series estimator

Regression analysis is an analytical technique in statistics that can be used to explain the pattern of functional relationship between response variable and predictor variable. There are three types of estimator in regression analysis. There are parametric, nonparametric, and semiparametric regression. Semiparametric regression is a combination of a parametric and nonparametric regression. Parametric regression in this case can be estimated with linear function, and the nonparametric regression in this case can be estimated with Fourier series. The equation of a regression in semiparametric as follows:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{l=1}^{r} g_l(t_{il}) + \varepsilon_i \qquad (1)$$

In equation (1), the semiparametric regression approach used consists of parametric regression components which are approximated by linear functions with predictor variables x as much as p, and semiparametric regression functions with predictor variables t as much as r. Response variable denoted by y. Regression coefficient for parametric component denoted by β. The value of β can be gotten from estimation result. Regression curve for nonparametric component represented with $g_k(t_{ik})$, here i depends on the number of observations. An error random ε_i is independent and identically distributed with mean 0, and variance σ^2 denoted by ε (Pane *et al*., 2013).

Regression curve in (1) approached by Fourier series estimator. Fourier series is a function of trigonometric polynomials which has a high degree of flexibility. Fourier series is a curve that shows the sines and cosines functions (Suslov, 2003). If given $g(t)$ is a function that can be integrated and differentiable in intervals $[a, a + 2L]$, then the Fourier series representation of the interval associated with $g(t)$ containing the sines and cosines trigonometric components is as follows:

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos k^* t + b_n \sin k^* t) \qquad (2)$$

with $k^* \approx \frac{n\pi}{L}; n = 1, 2, 3, ...$

If $g(t)$ even functions, or if $g(-t) = g(t)$, the Fourier coefficient $b_n = 0$. Thus, the Fourier series is called the Fourier cosines series. So, if $g(t)$ can be integrated and differentiated in the interval $[0, L]$, then the Fourier cosines series is as follows:

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos k^* t \qquad (3)$$

If $g(t)$ odd functions, or if $g(-t) = -g(t)$, the Fourier coefficient $a_n = 0$. Thus, the Fourier series is called the Fourier sines series (Bilodeau, 2008). So, if $g(t)$ can be integrated and differentiated in the interval $[0, L]$, then the Fourier sines series is as follows:

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} b_n \sin k^* t \qquad (4)$$

By adjusting the Fourier series formulation, Bilodeau (1992) constructs the Fourier cosines series estimator in equation (3) by adding a trend function that can be used for nonparametric regression approaches with the Fourier series estimator for paired data (t_i, y_i) as follows:

$$g_l(t_{il}) = \frac{a_{0l}}{2} + \omega_l t_{il} + \sum_{k=1}^{K} a_{kl} \cos kt_{il} + \varepsilon_i \quad (5)$$

Equation (5) can become an analogue for the other Fourier series. So, the semiparametric regression equation based on Fourier series estimator is the substitution result equation (5) to equation (1).

The Fourier cosine series estimator for semiparametric regression is given as follows:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} + \sum_{l=1}^{r} \left(\frac{\hat{a}_{0l}}{2} + \hat{\omega}_l t_{il} + \sum_{k=1}^{K} \hat{a}_{kl} \cos kt_{il} \right) \quad (6)$$

The value of $a, \omega,$ and β can be determine based on the result of Ordinary Least Square (OLS) optimization (Mardianto and Budiantara, 2014). In this case k represents an oscillation parameter, that determine the number of oscillations for approaching the data pattern.

2.3 Selection of oscillation parameters

The selection of k values must be carried out optimally. Determination of optimal k can use the GCV (Generalized Cross Validation) method. The GCV method for oscillation parameters k is generally defined based on (Asrini and Budiantara, 2014). The GCV value depends on MSE (Mean Square Error) value because the numerator of GCV formula is MSE formula. The measure of model goodness is also determined by the value of the determination coefficient or R^2 which shows the percentage contribution of the predictor variable to the response variable. The best model that can be used for prediction met the goodness of criteria. The goodness of criteria is the smallest GCV value for an optimal oscillation parameter, the smallest MSE value, and the big of determination coefficient value (Takezawa, 2006).

2.4 Data analysis procedure

The procedure in analysis data that be related to prediction of rice production in Indonesia using semiparametric regression based on three forms of Fourier series estimator is given as follows:

a) Study literature related to rice production in Indonesia based on the distribution of provinces, and its relationship to the predictor variables used.
b) Perform descriptive statistics for each variables based on minimum, maximum, and average values.
c) Determine GCV value for each the number of oscillation parameter in three Fourier series estimators including cosines, sines, and estimators that contain both that be used based on training data.
d) Choose the smallest GCV value for every Fourier series that be used, and determine MSE and determination coefficient.
e) Comparing three Fourier series estimator that will be used to predict rice production based on the productivity of agricultural land and the extent of agricultural land.
f) Select the best model based on the smallest GCV value, MSE value, and the biggest determination coefficient value.
g) Using testing data, make prediction about rice production based on the productivity of agricultural land and the extent of agricultural land.

3. RESULTS AND DISCUSSIONS

3.1 Characteristics of rice agricultural in Indonesia

The following are described the characteristics of the variables that be used in this study, namely the production of rice (tons), harvested area (hectares), and land productivity (quintals/hectares) in 2016 - 2017. Each research variable in each year, taken the highest five provinces and the five provinces lowest.

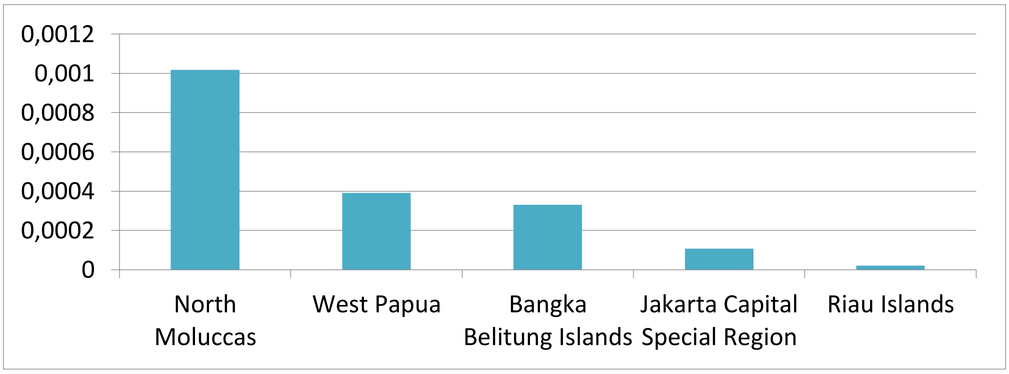

Figure-1. Five Provinces with The Lowest Rice Production (tons) in 2016.

VOL. 14, NO. 15, AUGUST 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

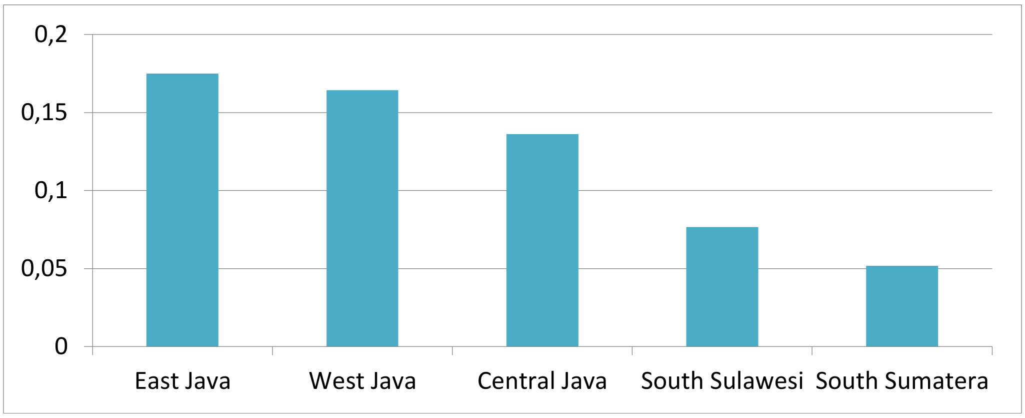

Figure-2. Five Provinces with The Highest Rice Production (tons) in 2016.

Figure-1 shows the five provinces with the lowest rice production (tons) in 2016. Based on Figure-1, it is seen that these provinces are North Maluku at 0,0010173; West Papua of 0,0003905; Bangka Belitung Islands is 0,0003314; Jakarta Capital Special Region of 0,0001064; and Riau Islands amounted to 0,000020. Figure-2 shows the five provinces with the highest rice production (tons) in 2016. Based on Figure-2, it is known that the five provinces are East Java at 0,1749847; West Java at 0,1643681; Central Java of 0,1361833; South Sulawesi of 0,0765895; and South Sumatra of 0,0518083.

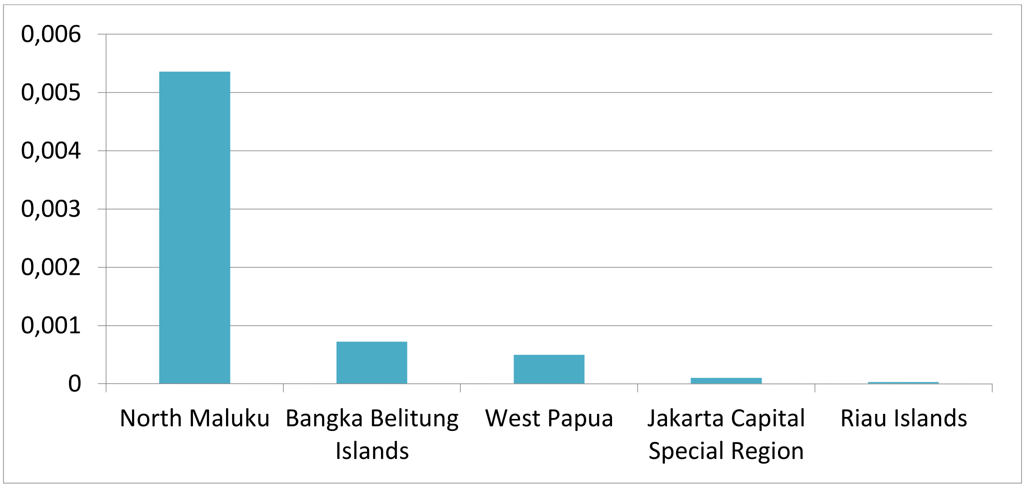

Figure-3. Five Provinces with The Lowest Harvest Area (hectares) in 2016.

Figure-3 shows the five provinces with the lowest harvested area (hectares) in 2016. Based on Figure-3, it can be seen that these provinces are North Maluku at 0,00536; Bangka Belitung Islands is 0,000721; West Papua of 0,000499; Jakarta Capital Special Region amounting to 0,000101; and Riau Islands amounted to 0,000028.

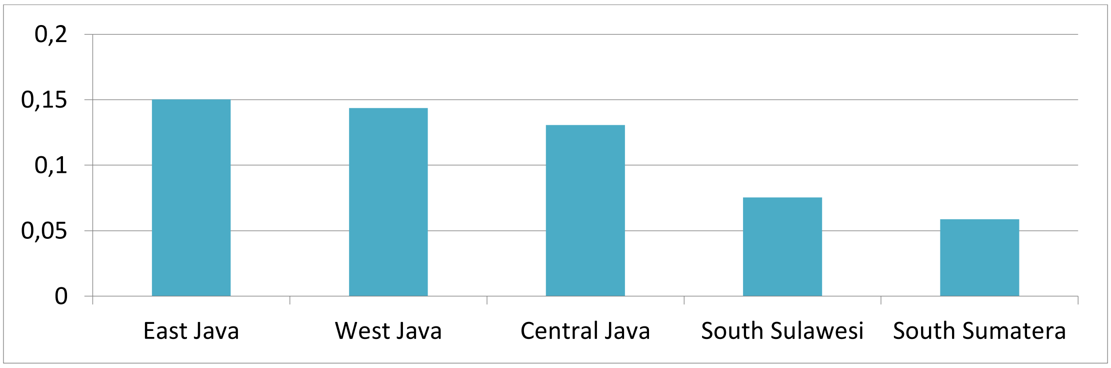

Figure-4. Five provinces with the highest harvest area (hectares) in 2016.

VOL. 14, NO. 15, AUGUST 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Figure-4 shows the five provinces with the highest harvested area (hectare) in 2016. Based on Figure-4, it is known that the five provinces include East Java of 0,150218; West Java of 0,143492; Central Java at 0,130526; South Sulawesi of 0,075379; and South Sumatra of 0,058772.

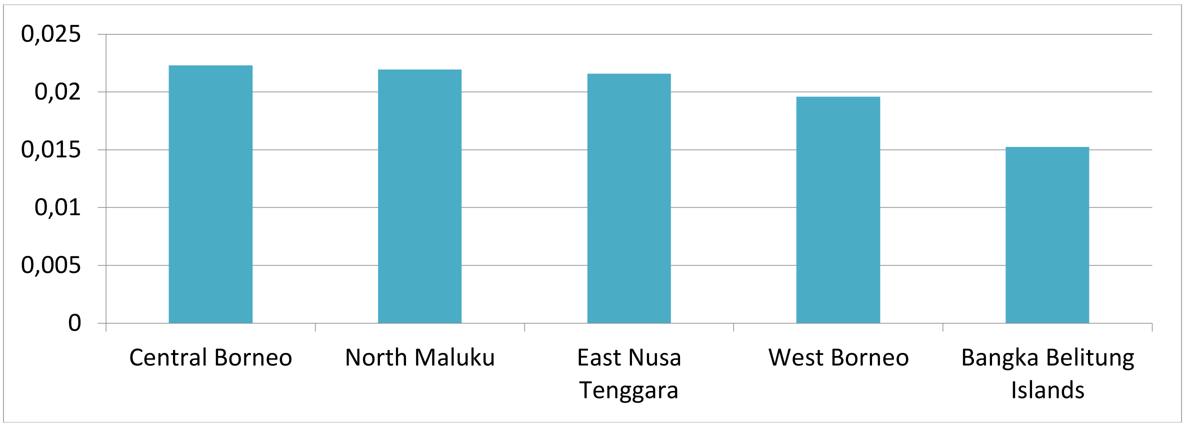

Figure-5. Five provinces with the lowest productivity of land (quintals / hectares) in 2016.

Figure-5 shows the five provinces with the lowest land productivity (quintal / hectare) in 2016. Based on Figure-5, these provinces are known as Central Borneo of 0,022309; North Maluku of 0,021947; East Nusa Tenggara is 0,021592; West Borneo at 0,019585; and Bangka Belitung Islands is 0,015242. Figure 6 shows the five provinces with the highest land productivity (quintals / hectares) in 2016. Based on Figure 8, it is known that these provinces include Bali at 0,038797; East Java of 0,038597; West Java at 0,037958; Yogyakarta Special Region is 0,037345; and Jakarta Capital Special Region amounting to 0,034757.

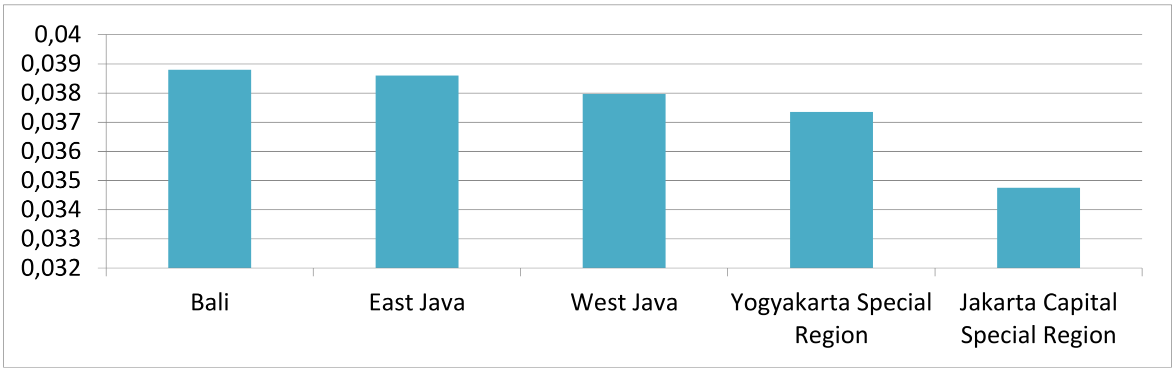

Figure-6. Five provinces with the highest productivity of land (quintals / hectares) in 2016.

From this section can be summarized that there are gap in each provinces. Provinces that probably become rice granary area in Indonesia is provinces that located in Java, islands with the largest population and the most crowd in Indonesia. Provinces that have small rice production, generally provinces with small harvest area, and mostly locatted in East Indonesia and Islands Province. But, some province like Bali, South Sumatera and South Sulawesi have probability to become rice granary in Indonesia.

3.2. The estimation result

The results of the optimal GCV value that be calculated to R software using training data are presented in the following table:

Table-1. GCV value based on Fourier cosines series estimator.

k	**GCV Value**
3	144,263330
4	0,574006
5	1.214,853000

Table-2. GCV value based on Fourier sines series estimator.

k	**GCV Value**
20	1.436,965
21	1.105,573
22	1.316,003

ARPN Journal of Engineering and Applied Sciences

Table-3. GCV value based on Fourier series that include cosines and sines estimator.

k	GCV Value
24	1.360,189
25	1.272,236
26	4.682,301

Based on Table-1, the minimum GCV value is 0,574006 with k equal to 4 is chosen. Based on optimal osilation parameter value (k) as equal as 4, obtained Fourier series estimator model in semiparametric regression as follows:

$$\hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i + \frac{\hat{\alpha}_0}{2} + \hat{\omega} t_i + \hat{\alpha}_1 co + \hat{\alpha}_2 \cos 2t_i + \hat{\alpha}_3 \cos 3t_i + \hat{\alpha}_4 \cos 4t_i$$

Based on the results of calculations using R software, the parameter values in the model can be written as follows:

$$\hat{y_i} = 13.487{,}088 + 1{,}066x_i + 3{,}261t_i - 17.155{,}348\cos t_i + 14.809{,}671\cos 2t_i - 19.484{,}707\cos 3t_i + 8.343{,}262\cos 4t_i$$

The estimator based on Fourier cosines series in semiparametric regression is the best estimator because when compared with Fourier sines series and Fourier series that have basis sines and cosines the estimator is the most parsimony or has simple estimator form. For estimator based on Fourier cosines an optimal oscillation parameter can be reached when k equals to 4. For Fourier sines series estimator, based on Table 2, an optimal oscillation parameter can be reached when k equals to 21. For Fourier series estimator that include cosines and sines, based on Table 3, an optimal oscillation parameter can be reached when k equals to 25.

The estimator based on Fourier cosines series in semiparametric regression has the smallest GCV value equals to 0,574. For estimator based on Fourier sines series the GCV value equals to 1.105,573. For estimator based on Fourier series the GCV value equals to 1.272,236. Consequently, the estimator based on Fourier cosines series in semiparametric regression has the smallest MSE value equals to 0,0000194. For estimator based on Fourier sines series the MSE value equals to 0,004854140. For estimator based on Fourier series the MSE value equals to 0,601119. And the last goodness criteria that be compared is determination coefficient. The estimator based on Fourier cosines series in semiparametric regression has the bigest determination coefficient value approaches to 100%. For estimator based on Fourier sines series the determination coefficient value equals to 96%. For estimator based on Fourier series the determination coefficient value equals to 92%. So, we use Fourier cosine series in semiparametric regression to make prediction.

3.3 The prediction result

Rice production the following year, which is 2017 is predicted based on the best model with the cosines Fourier series approach using testing data that be observed at 2017, the prediction results are presented in Table 4. Based on the ratings analysis of rice production and the rice production prediction results, there are 6 provinces that experienced a significant shift in rankings. The six provinces experienced a shift of more than 5 ranking units, while there are 16 other provinces that experienced a rating shift but are not significant. The provinces of East Java, West Java, Central Java, South Sulawesi, South Sumatra, North Sumatra, Lampung and West Sumatra do not experience changes in rank and are sequentially ranked 1 to 9. This indicates that the model is made using the Fourier series estimator in semiparametric regression have good performance to predict rice production in Indonesia theoretically. Statistically the model has the best criteria with k values equal to 4, GCV equals 0.574, MSE equals 0.0000194, and R^2 approaches 100%. So, the model has met the criteria for the goodness of the model.

ARPN Journal of Engineering and Applied Sciences

Table-4. Comparison between Prediction of Rice Production in 2017 and Rice Production in 2017 as Testing Data.

Code	Province	Rice Production	Prediction	Code	Province	Rice Production	Prediction
1	Aceh	0,030917	0,024135	18	West Nusa Tenggara	0,032062	0,028736
2	North Sumatra	0,053646	0,051473	29	East Nusa Tenggara	0,012574	0,013215
3	West Sumatra	0,033829	0,034645	20	West Kalimantan	0,016920	0,029087
4	Riau	0,005225	0,002231	21	Central Kalimantan	0,011847	0,012851
5	Jambi	0,007182	0,005596	22	South Kalimantan	0,028386	0,032448
6	South Sumatra	0,056340	0,056941	23	East Kalimantan	0,005422	0,001720
7	Bengkulu	0,007675	0,005323	24	North Kalimantan	0,001487	0,003472
8	Lampung	0,048302	0,046477	25	North Sulawesi	0,008941	0,005336
9	Bangka Belitung Islands	0,000359	0,006159	26	Central Sulawesi	0,013467	0,011496
10	Riau Islands	0,000013	0,005939	27	South Sulawesi	0,072572	0,077329
11	Jakarta Capital Special Region	0,000084	0,001695	28	South East Sulawesi	0,008763	0,005447
12	West Java	0,150842	0,156789	29	Gorontalo	0,004393	0,000664
13	Central Java	0,149891	0,137158	30	West Sulawesi	0,006125	0,002112
14	Yogyakarta Special Region	0,012535	0,014736	31	Maluku	0,001562	0,003547
15	East Java	0,174474	0,165466	32	North Maluku	0,000998	0,004234
16	Banten	0,029033	0,027352	33	West Papua	0,000401	0,005552
17	Bali	0,011323	0,016799	34	Papua	0,002411	0,002478

4. CONCLUSIONS

Descriptions of rice production in Indonesia based on the productivity of agricultural land and the extent of agricultural land generally have gaps. For example provinces with large and densely populated areas are rice granary areas in Indonesia, especially provinces in Java such as West Java, Central Java and East Java. For make a plan and target, the rice production in Indonesia is modeled based on the best semiparametric regression based on Fourier cosine series estimator. The estimator satisfies parsimony model, the smallest GCV and MSE value, and the biggest determination coefficient value. Using testing data, obtained the results of prediction of rice production in Indonesia which is not much different from training data. The ranking of the largest to smallest provinces in rice producing has not changed much. Recommendations for the government in manage the needs of rice in Indonesia, the government can mantain the stability of rice stock in provinces that has rank top in the rice production without reducing agricultural land for national food sustainability.

ACKNOWLEDGEMENTS

The authors give high appreciation for Indonesia Central Bureau of Statistics who have provided data, Perum Bulog that manage rice stock in Indonesia, which have received the results of the study with good response, and Airlangga University has funded this publication.

REFERENCES

Asrini L. J. and Budiantara I. N. 2014. Fourier series semiparametric regression models (case study: the production of lowland rice irrigation in Central Java). ARPN Journal of Engineering and Applied Sciences. 9(9) : 1501-1506.

Biedermann S., Dette H. and Hoffmann P. 2009. Constrained Optimal Discrimination Designs for Fourier Regression Models. Ann Inst Stat Math Journal. 61(2): 143-157.

Bilodeau M. 1992. Fourier Smoother and Additive Models. Canadian Journal of Statistics. 20(3): 257-269.

Budiantara I. N., Ratnasari V., Zain I., Ratna M. and Mardianto M. F. F. 2015. Modeling of HDI and PQLI in East Java (Indonesia) using Biresponse Semiparametric Regression with Fourier Series Approach. ATABS Journal. 5(4): 21-28.

Central Bureau of Statistics. 2017. Indonesia Statistics *2017* Indonesia Central Bureau of Statistics Jakarta.

Central Bureau of Statistics. 2018. Official News about Statistics, No. 10/02/Th.XX1, 1st February 2018 Indonesia Central Bureau of Statistics Jakarta.

Chauhan B. S., Jabran K. and Mahajan G. 2017. Rice Production Worldwide Springer Switzerland.

Dette H., Melas V. B. and Shpilev P. 2016. T -optimal Discriminating Designs for Fourier Regression Models. Journal of Computational Statistics and Data Analysis. 26(3): 1-11.

Hardle W. 1990. Applied Nonparametric Regression Cambridge University Press New York.

Mardianto M. F. F and Budiantara I. N. 2014. Estimation of Biresponse Semiparametric Regression Model with Fourier series Approach Proceeding National Seminar Udayana University Denpasar.

Natawidjaja R. S. and Rum I. A. 2012. Food Security Situation and Policy in Indonesia Center for Agrifood Policy and Agribusiness Studies Padjajaran University Bandung.

Pane R., Budiantara I. N., Zain I. and Otok B. W. 2013. Parametric and Nonparametric Estimators in Semiparametric Regression and Their Characteristics Journal of Applied Mathematical Science. 8(102): 5053-5064.

Suslov S. K. 2003. An Introduction to Basuc Fourier Series Springer-Science Arizona.

Takezawa K. 2006. Introduction to Nonparametric Regression John Wiley and Sons Inc. New Jersey.

Tjahjono E., Mardianto M. F. F and Chamidah N. 2018. Prediction of Electricity Consumption using Fourier Series Estimator in Biresponse Semiparametric Regression Model. Far East Journal of Mathematical Sciences (FJMS). 103(8): 1251-1263.