

# Rainfall forecasting using PSpline and rice production with ocean-atmosphere interaction

**Rezzy Eko Caraka<sup>1†</sup>, Budi Darmawan Supatmanto<sup>2</sup>, Muhammad Tahmid<sup>3</sup>, Joko Soebagyo<sup>4</sup>, M Ali Mauludin<sup>5</sup>, Akbar Iskandar<sup>6</sup>, Bens Pardamean<sup>1,7</sup>**

<sup>1</sup> Bioinformatics and Data Science Research Center (BDSRC), Bina Nusantara University, Jakarta, Indonesia 11480

<sup>2</sup> Weather Modification Technology Center, Agency for the Assessment and Application of Technology (BPPT)

<sup>3</sup> Indonesia Agency for Meteorology Climatology and Geophysics, Indonesia

<sup>4</sup> College of Technology Wastukencana, Purwakarta

<sup>5</sup> Laboratory of Sociology and Extension, Padjadjaran University

<sup>6</sup> Informatics Engineering, STMIK AKBA, Makassar

<sup>7</sup> Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

Corresponding author: rezzyekocaraka@gmail.com<sup>†</sup>

**Abstract.** The role of climate can be affected by plants. The weather can accelerate and multiply the existence of various plant pests and diseases, accelerate the growth and development of grass among plants, and encourage the emergence of infection and significant damage to plants. The elements of climate that affect the growth of plants are one of them is rainfall. In this paper, we performed the simulation using the non-parametric penalized spline (PSPLINE) method and studied the effect on rice production in Lampung. It can be concluded that the increasing fluctuation, frequency, and intensity of climate anomalies in the last decade caused by the ENSO phenomenon have an impact on changes in distribution patterns, intensity, and period of the wet season so that the start of the rainy season and the dry season becomes too late. As a result, there is a seasonal shift from normal average conditions that can ultimately have severe implications for food crops. In a nutshell penalized spline gives high accuracy with  $R^2 = 96.227\%$  and MAPE= 1.62%.

**Keywords:** Rainfall, Ocean, Atmosphere, Penalized, Spline

## 1. Introduction

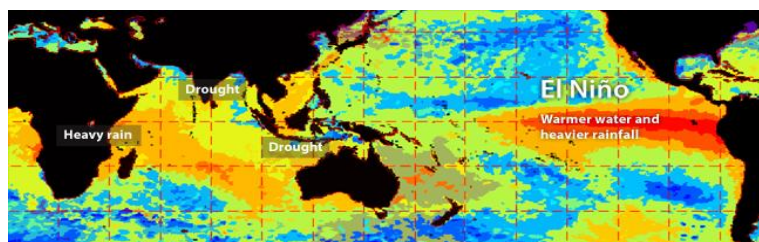
In realizing of the food security system, it is necessary to pay attention to the climate impacts [1]. Food demand, especially rice, will continue to increase in line with the increase in population and the increase in per capita consumption due to the rise in income. However, efforts to increase rice production are currently hampered by various constraints such as the ongoing conversion of fertile rice fields, climate anomaly, and technological fatigue phenomenon, the impact of soil to decreased productivity. The agricultural sector is inseparable from the atmospheric factor that crops will be



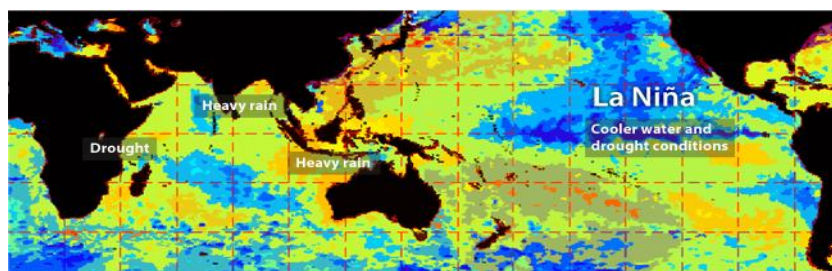
productive in certain climatic conditions. Earth's atmosphere is a system that is very diverse with the variability that occurs in a considerable range both in time and distance scales. The spread of vegetation of plant species will be limited by climatic [2] [3] [4] and soil conditions[3] and the adaptability of each species of the plant. In addition to the atmosphere can affect the growth and development of plants, the presence of vegetation can also change the surrounding environment[5]. Tropical climate conditions, especially rainfall is a very complex climate phenomenon that is influenced by global, regional and local factors. Indonesia is the largest tropical archipelagic country on earth with the longest coastline, and the mean sea-to-land ratio on the planet is 71.1% over 28.9%. With such a tremendous rate it is believed that Indonesia's climate is strongly influenced by oceans and areas around. The tropical continent of Indonesia in the tropics receives the most solar radiation, located between two continents (Asia and Australia) and two vast oceans (Indian and Pacific Oceans) causing this BMI region to be vulnerable to climate variability and climate change. Changes in climate variation in the Tropical continent region correlate with differences in sea surface temperatures both within Indonesia's waters and those in the Pacific Ocean and Indian Oceans. Also, there are two distinct understandings between climate variability and climate change. Climate variability further indicates the presence of climate differences in a relatively short period such as daily, monthly and yearly such as the El Niño and La Niña phenomena while climate change covers more than 30 years and with a wide coverage area. The impact of climate variability is very significant in the decrease in annual rainfall resulting in prolonged droughts and floods that are strictly related to food production. Indonesia as an agricultural country has a vast field of rice fields scattered in almost all parts of Indonesia. The problems of the agrarian sector are inseparable from climate variability. In the tropics, the length of the growing season depends on the falling rain except for areas with guaranteed irrigation[6]. Therefore, farmers have a tradition to match planting patterns to widespread rain distribution throughout the year.

## 2. Ocean-Atmosphere Interaction

The phenomenon of El Niño and La Niña[7][8] is due to the variability of the interaction conditions between the oceans and the atmosphere along the equatorial Pacific Ocean from its normal state. The El Niño incident in Indonesia is identified with the occurrence of a dry season that exceeds its normal condition. This is inversely proportional to the La Niña event that is capable of producing rainfall beyond its reasonable limits. El Niño will occur when the warmer waters of the central and eastern Pacific increase the temperature and humidity in the atmosphere above it. This incident led to the formation of clouds that will increase rainfall around the area[9]. The western part of the Pacific Ocean increased air pressure causing obstructions of cloud growth over the eastern seas of Indonesia, resulting in some regions of Indonesia a decline in rainfall far from normal. El Niño is often associated with Southern Oscillation with the abbreviation ENSO. South Oscillation is a 'pressure' system of air pressure between the eastern Pacific region and the Indonesian territory when the surface air pressure is low. This slow-moving, large-scale air pressure oscillation affects rainfall in some parts of Indonesia, although its influence varies across an archipelago and at different times. At the time of the El Niño event is generally accompanied by a negative southern oscillation index value.



**Figure 1.** El Nino Simulation[10]



**Figure 2.** La Nina Simulation [10]

### 3. Data and Analysis

In this research, we use data in Lampung area generally same as another area in Indonesia. Lampung located below the equator which has a wet tropical climate with wet sea wind blowing from the Indian Ocean has two seasons each year, and with an average air, humidity ranges from 80 to 88%. The rainy season occurs between November and May and the dry season lasts between May and September. Lampung area temperatures in terrain with elevations up to 60 m average between 26°C - 28°C. Maximum temperature is around 33°C, and minimum temperature is 22°C. We used the Nino 3.4 Data Index obtained from the National Center for Atmospheric Research (NCAR). Rainfall data used is rainfall data obtained from Lampung Climatology Station and data of annual agricultural productivity of each district (ton/ha) derived from Ministry of Agriculture Indonesia

### 4. Application

Data Index Nino 3.4 is a calculation of sea surface temperature anomalies. Next, determine the years of climate variability with the sea surface temperature and anomaly criteria in the Pacific Ocean region. The criteria for anomalous sea level temperatures:

*El Niño* :  $> 0.5^{\circ}\text{C}$

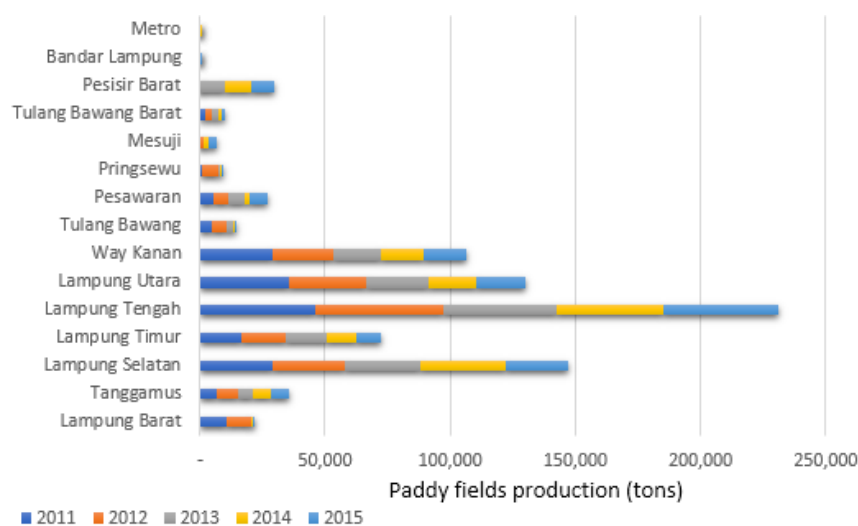
Normal:  $\pm 0.5^{\circ}\text{C}$

*La Niña* :  $< -0.5^{\circ}\text{C}$

To determine whether or not the effect of El Niño and La Niña on season variability and rice productivity is through comparative analysis or comparison to the above data processing result that is the initial variability and the length of the seasons based on the anomaly scale of sea level temperature period. The analysis phase is as follows (1) Describe the anomalous criteria of sea surface temperature of the Pacific Ocean region, in this case, the El Niño and La Niña events. (2) Analysis of the initial normal distribution and the length of the rainy and dry seasons throughout Lampung zone season based on the calculation of rainfall and season criteria according to Meteorological, Climatological, and Geophysical Agency Indonesia (BMKG). (3) Analysis of the influence of the Pacific Ocean SML anomaly on the early variability of the dry season and the beginning of the rainy season on each zone season against the norm. (4) Analysis of the influence of the Pacific Ocean anomaly on the large variability of the dry season and the length of the rainy season on each zone season against its normal. (5) Analysis of the influence of season variability on rice productivity in each district of Lampung Province



**Figure 3.** Zone Season in Lampung, Indonesia



**Figure 4.** Paddy Fields Production (tons)

The early variability and the length of the seasons caused by the anomalous climate of the Pacific Ocean affect the productivity of paddy in Lampung regency. At the time, El Niño that resulted in early wet season retreats resulted in farmers will have difficulty in determining the beginning of the planting period and the longer pre-dry season will also affect the growth phase of rice crops. In some parts of Indonesia generally, El Niño is closely linked to crop failure and declining productivity. The productivity of paddy in the Lampung region at the time of El Niño generally decreased substantially from the average. Almost all districts have decreased from  $\pm 0.1$  to  $\pm 0.8$  Tons/ha. While for Metro, Tanggamus and South Lampung districts do not show the same effect, these three areas have increased productivity in the presence of El Niño. In La Niña years the productivity of paddy in the Lampung region generally decreased as well as when the El Niño occurred even though the decline in La Niña was not as significant as the El Niño year. However, in 2010 when La Niña with moderate rice productivity category in all Lampung Regency increased productivity by  $\pm 0.1$  to  $\pm 0.8$  ton/ha. La Niña which resulted in early wet season retreat in most of Lampung region may indicate a productivity disruption because it is difficult to determine the beginning of the planting period. Longer pre-dry season with too high rainfall intensity at the time of La Niña can lead to flooding and disrupt the

growth of rice crops. However, in Metro, Tanggamus and South Lampung districts, their productivity has increased. It is suspected that in these three regions rainfall did not significantly affect the productivity of rice so that the Pacific anomaly resulting in season variability did not affect productivity. These three areas are high production when El Niño occurs, and La Niña productivity tends to increase.

Next, we use a non-parametric penalized spline to predict rainfall on Lampung. Smoothing is one of the methods used in nonparametric data analysis[11][12]. The purpose of smoothing is to minimize the error and estimate the behavior of data that tends to be different and has no effect so that the characteristics of the data will appear more clearly and carefully. One of the regression models with a nonparametric approach that can be used to estimate the regression curve is spline regression. Spline regression is an approach to matching data while still taking into account the curve. The spline approach has its virtues because spline is a piecewise polynomial of order  $m$  that has a continuous segmented property that adequately describes the local characteristics of data functions [13] The spline regression model with the order of  $m$  and knot point  $\tau_1, \tau_2, \dots, \tau_K$  can be write as bellows :

$$y_i = \beta_0 + \beta_1 x + \dots + \beta_{m-1} x^{m-1} + \sum_{k=1}^K \beta_{m-1+k} (x - \tau_k)_+^{m-1} + \varepsilon_i$$

With truncated function:

$$(x - \tau_k)_+^{m-1} = \begin{cases} (x - \tau_k)^{m-1}, & x \geq \tau_k \\ 0, & x < \tau_k \end{cases}$$

Where  $m$  is a polynomial orde,  $\tau_k$  is the  $k$ -th knot point with  $k=1, 2, \dots, K$  and  $\varepsilon_i$  is an error with mean zero and variance  $\sigma^2$  [14]. Spline has an advantage in overcoming the pattern of data showing sharp ups and downs with the help of knots, and the resulting curve is relatively smooth. Knots are a common combination that shows the behavioural changes of the spline function at different intervals[5][15]. One form of spline regression is the penalized spline obtained by minimizing Penalized Least Square (PLS), which is an estimation function that combines the least square function and smoothness of the curve. The smoothing parameter is the balance controller between the curve conformity to data and curves. On the other hand, it is desirable to have an estimator form in addition to having a degree of graduation, also by the data. Therefore, it is essential to choose an optimal finishing parameter. Selecting a fining parameter in principle is the same as selecting many optimal knots that produce the optimum knot value which results in the minimum GCV value [16]

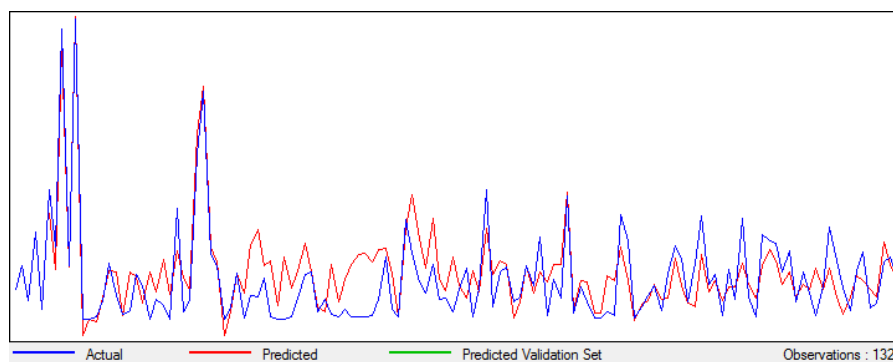
$$GCV = n^{-1} \frac{RSS(\lambda)}{(1 - n^{-1} df)^2}$$

With  $RSS = Residual Sum Square$ ;  $RSS(\lambda) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ . The degree of freedom is equivalent to the trace value of the matrix of hat  $S$   $S(\lambda) = (X(X^T X + \lambda^{2m-1} R)^{-1} X^T)$ ;  $df = tr(S(\lambda)) = tr(X(X^T X + \lambda^{2m-1} R)^{-1} X^T)$ . So, the GCV function can be expressed as follows:

$$GCV = n^{-1} \frac{RSS(\lambda)}{(1 - n^{-1} tr(S(\lambda)))^2}$$

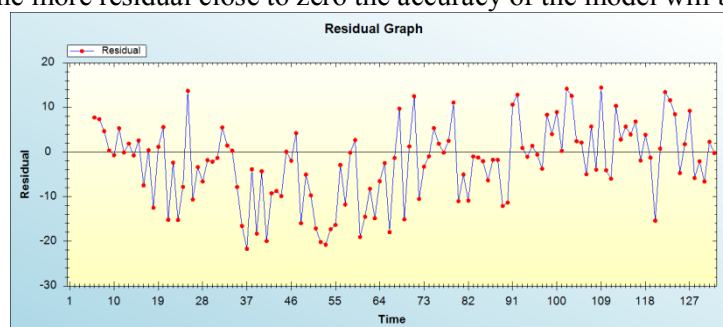
In the selection of many optimal knots using a full-search algorithm that is by selecting many knots and parameters optimally tuned that produce the minimum GCV value. Based on a simulation the minimum GCV value of 577.84 which is at  $K=5$  with an optimal finishing parameter of 305.3856. Thus  $K = 5$  is chosen as many optimum knots on the order of 2 with an optimal finishing parameter of 15.117. The P-Spline model is formed:

$$f(x_i) = \sum_{r=0}^1 \beta_r x_i^r + \sum_{k=1}^5 \beta_{2k} |x_i - \tau_k|^{3+} + \varepsilon_i$$



**Figure 5. Actual vs Prediction Rainfall**

Based on figure 5 can be seen that the actual data fit predicted data so it can be concluded that the estimation of spline penalize has good performance. Figure 6 residual plot to see if the error value is close to zero or not. the more residual close to zero the accuracy of the model will be higher



**Figure 6. Residual Graph**

Model performance used in forecasting can be seen based on the  $R^2$  value of the data in the sample and Mean Absolute Percentage Error (MAPE) from the out sample data.  $R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2} = 96.227\%$ . This is indicate that the effect of rain on the past amount on the future amount of rain at 96.227% and the remaining 3,773% is influenced by other variables. MAPE is a quantity error that compares actual data rates with forecasting data from the estimation of the penalized spline model. With the value of rain data ( $y_i$ ) and prediction results ( $\hat{y}_i$ ).  $MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \%$ . So that obtained MAPE = 1.62%. Based on the calculation, the MAPE value is less than 10%, which is 1.62%, so it can be concluded that the estimation of the penalized spline model has a high accurate forecasting.

## Conclusion

Season variability also affects early indications of El Niño and duration of the El Niño itself. If the occurrence of El Niño in the middle of the dry season this can lead to an increasingly long dry season and the duration of the more widespread El Niño phenomenon that can lead to the beginning of the rainy season is getting backward. The local influence is also likely to affect whether or not the El Niño effect on season variability, such as in ZOM 39 areas on the plateau at the time of El Niño occurs. At the time of La Niña, generally, there was an increase in rainfall that affected the beginning of the season and the length of the season. Used when the La Niña region in Indonesia experienced the slower dry season or retreat and the beginning of the rainy season is faster or advanced while the Long Rainy Season is longer or longer than normal. At the time of El Niño, rice productivity in the



Lampung region tended to decrease from  $\pm 0.1$  to  $\pm 0.8$  Ton/ha. While for Metro Municipality, Tanggamus Regency and Lampung Selatan El Niño do not significantly affect rice productivity in which these three areas have increased productivity. At the time of La Niña, rice productivity in the Lampung region also tended to decrease. As for Metro, Tanggamus Regency and South Lampung productivity increased. The productivity of paddy in East Lampung is not disrupted when La Niña occurs or is equal to the average.

## References

- [1] R. E. Caraka, Supari, and M. Tahmid, "Copula-Based Model for Rainfall and El- Niño in Banyuwangi Indonesia," *J. Phys. Conf. Ser.*, vol. 1008, no. 1, 2018.
- [2] I. D. Kurniawan, H. R. . Soesilhadi, C. Rahmadi, R. E. Caraka, and B. Pardamean, "The difference on Arthropod communities' structure within show caves and wild caves in Gunungsewu karst area, Indonesia," *Int. journal-Ecology, Environ. Conserv.*, vol. 24, no. 1, pp. 81–90, 2018.
- [3] I. D. Kurniawan, C. Rahmadi, R. E. Caraka, and T. A. Ardi, "Short Communication: Cave-dwelling Arthropod community of Semedi Show Cave in Gunungsewu Karst Area, Pacitan, East Java, Indonesia," *Biodiversitas*, vol. 19, no. 3, pp. 857–866, 2018.
- [4] R. E. Caraka *et al.*, "Ecological Show Cave and Wild Cave : Negative Binomial Gllvm ' s Arthropod Community Modelling," *Procedia Comput. Sci.*, vol. 135, pp. 377–384, 2018.
- [5] R. E. Caraka and A. R. Devi, "Application of Non Parametric Basis Spline (B-Spline) in Temperature," *Statistika*, vol. 4, no. 2, pp. 68–74, 2016.
- [6] R. E. Caraka and H. Yasin, *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis*, 1st ed. MOBIUS GRAHA ILMU, 2017.
- [7] Supari, F. Tangang, E. Salimun, E. Aldrian, A. Sopaheluwakan, and L. Juneng, "ENSO modulation of seasonal rainfall and extremes in Indonesia," *Climate Dynamics*, pp. 1–22, 2017.
- [8] Supari, F. Tangang, L. Juneng, and E. Aldrian, "Observed changes in extreme temperature and precipitation over Indonesia," *Int. J. Climatol.*, vol. 37, no. 4, pp. 1979–1997, 2017.
- [9] R. E. Caraka and S. A. Bakar, "Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations," *J. Eng. Appl. Sci.*, vol. 13, no. 15, 2018.
- [10] Tokyo Climate Center, "WMO Regional Climate Center in RA II (Asia)," Japan, 2018.
- [11] B. Pardamean, A. Budiarto, and R. E. Caraka, *Bioinformatika dengan R Tingkat Lanjut*, 1st ed. Yogyakarta: Teknosains, 2018.
- [12] Jamilatuzzahro, R. . Caraka, and R. Herliansyah, *Aplikasi Generalized Linear Model dengan R*, 1st ed. Yogyakarta: INNOSAIN Graha Ilmu Yogyakarta, 2018.
- [13] R. L. Eubank, "A simple smoothing spline, III," *Comput. Stat.*, vol. 19, no. 2, pp. 227–241, 2004.
- [14] D. Ruppert, "Nonparametric Regression and Spline Smoothing," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1522–1523, 2001.
- [15] A. Prahutama, T. W. Utami, R. E. Caraka, and D. Zumrohtuliyosi, "Pemodelan Inflasi Berdasarkan Harga-Harga Pangan Menggunakan Spline Multivariabel," *Media Stat.*, vol. 7, no. 2, pp. 89–94, 2014.
- [16] M. A. Lukas, F. R. de Hoog, and R. S. Anderssen, "Practical use of robust GCV and modified GCV for spline smoothing," *Comput. Stat.*, vol. 31, no. 1, pp. 269–289, 2016.