TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 11, November 2014, pp. 7912 ~ 7919
DOI: 10.11591/telkomnika.v12i11.6518

Optimization of Support Vector Regression using Genetic Algorithm and Particle Swarm Optimization for Rainfall Prediction in Dry Season

Gita Adhani*[1], Agus Buono[1], Akhmad Faqih[2]
[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, Bogor 16680, Indonesia
[2]Departmentof Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, Bogor 16680, Indonesia
*Corresponding author, e-mail: adhani.gita@gmail.com, pudesha@gmail.com, akhmadfaqih@gmail.com

Abstract

Support Vector Regression (SVR) is Support Vector Machine (SVM) is used for regression case. Regression method is one of prediction season method has been commonly used. SVR process requires kernel functions to transform the non-linear inputs into a high dimensional feature space. This research was conducted to predict rainfall in the dry season at 15 weather stations in Indramayu district. The basic method used in this study was Support Vector Regression (SVR) optimized by a hybrid algorithm GAPSO (Genetic Algorithm and Particle Swarm Optimization). SVR models created using Radial Basis Function (RBF) kernel. This hybrid technique incorporates concepts from GA and PSO and creates individuals new generation not only by crossover and mutation operation in GA, but also through the process of PSO. Predictors used were Indian Ocean Dipole (IOD) and NINO3.4 Sea Surface Temperature Anomaly (SSTA) data. This research obtained an SVR model with the highest correlation coefficient of 0.87 and NRMSE error value of 11.53 at Bulak station. Cikedung station has the lowest NMRSE error value of 0.78 and the correlation coefficient of 9.01.

Keywords*: rainfall in dry season, genetic algorithm, particle swarm optimization, support vector regression*

1. Introduction

Indonesia is country stated between 2 continents, Asia and Australia, and 2 oceans, Pacific and Hindia. Therefore Indonesia climate and weather are significantly affected by both oceans condition. Climate and weather factor has important role in many aspect of humankind. Besides, rainfall as variable which determining the climate condition is directly linked to agriculture and plantation success. As agrarian country, Indonesia depends on agriculture and plantation circumstance. High rate rainfall would cause flooding indicating great probability of failed crops. As bad as too long drought that would lead to not grown and dead plants.

Extreme weather can be related to climate deviation which defined as anomaly of weather and climate compared to normal environment in particular time range.One example of the deviations is occurrence of ENSO phenomenon namely El Nino and La Nina. El Nino case generally is connected to long time drought or dry season because of decrease in rainfall; otherwise La Nina is related to flooding. La Nina leads to overloaded accumulation of air mass that contains of a lot water vapor so that increase the potency of rain cloud formation.

Climatic phenomenon in Pacific Ocean can be seen on existence of Southern Oscillation Index (SOI) and Sea Surface Temperature Anomaly (SSTA) of NINO. Climate condition in Hindia on the other hand can be viewed on Indian Ocean Dipole (IOD). Beside of ENSO phenomenon in Pacific Ocean, IOD also affect significantly sea surface and atmosphere status. IOD and SSTA NINO3.4 play role as indicators to monitor the ENSO phenomenon.

Indramayu is one of Indonesian district which is center of agricultural products such as rice [1]. This place is very vulnerable to drought and flooding, especially when ENSO happens in Indonesia. Based on data collected in Annual Report of Indramayu Department of Agriculture, it is known that in previous years when El Nino and La Nina ensued, Indramayu encountered plenty food plants (rice) damages [2]. According to Estiningtyas [3], the main factor of the crop

Received July 25, 2014; Revised September 11, 2014; Accepted September 26, 2014

failure is drought (79.8%), pest organisms (15.6%) and flooding (5.6%) which are strongly influenced by climate deviation.

The research was focused on rainfall forecasting in dry season in Indramayu. Predictors used were variables related to dry season rainfall. Those were Indian Ocean Dipole (IOD) and Sea Surface Temperature Anomaly (SSTA) in NINO3.4 area. Method applied was Support Vector Regression (SVR) optimized by Genetic Algorithm and Particle Swarm Optimization.

Support Vector Machine (SVM) chosen in regression case is Support Vector Regression (SVR). The research adopted previous SVR method applied by Adhani [4] about rainfall prediction in dry season using SOI data and NINO 3.4 sea surface temperature. SVR process needs kernel to transform non-linier input to high dimension feature room. The research only applied the RBF kernel because in former study by Adhani[4] that has shown higher correlation value and smaller NRMSE error of RBF kernel compared to Linier or Polynomial kernels. Moreover, RBF kernel is the simple one by its parameter C and γ. Kernel has parameter value that have to be determined at first. The research implemented merger of two optimation method in order to define the optimal kernel function parameter which is Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) with abbreviation of GAPSO [5].

2. Research Method

Flowchart of research methods can be seen in Figure 1.

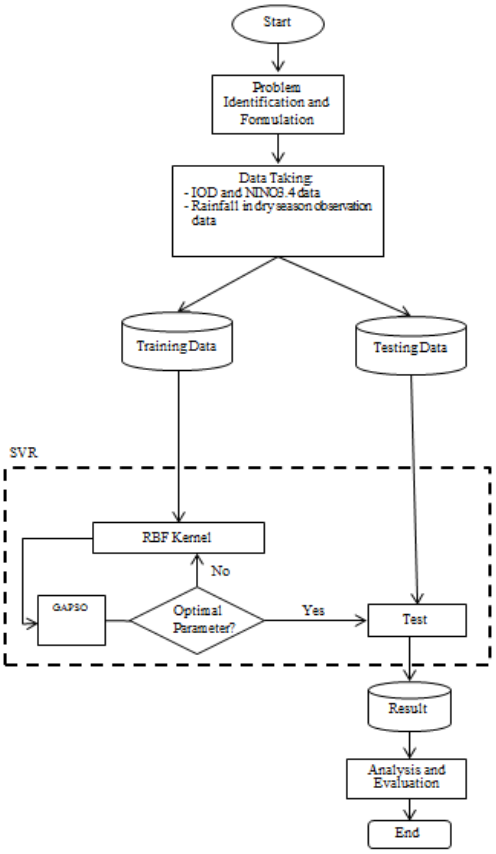


Figure 1. Research Method Flowchart

2.1. Data and Predictors Selection

Indian Ocean Dipole (IOD) and Sea Surface Temperature Anomaly (SSTA) NINO are indicators to monitor the ENSO phenomenon. ENSO has great role in extreme rain variability condition. Fluctuation of ENSO in Pacific ocean is highly related to rainfall in Indonesia [6]. IOD

is sea phenomenon followed by atmosphere phenomenon in Hindia ocean equator that influences the climate of Australia and other country surrounding Hindia ocean cavity [7]. IOD is identified as deviation of physical condition in atmosphere–sea interaction in tropical Hindia ocean that is assumed can lead to drought in Indonesia [8]. NINO Sea Surface Temperature Anomaly (NINO SSTA) is index of sea surface temperature in some regions. There are 4 NINO areas according to IRI [9], such as NINO1+2, NINO3, NINO3.4, and NINO4. NINO 3.4 is stated between equator latitude of 5°S–5°N d 170°–120°E and has high variability in El Nino time scale. NINO 3.4 is commonly used in global climate variability that has broad impact. Sea surface temperature variability in this area has strongest impact on rainfall friction on West Pacific [9].

IOD data from 1979 to 2008 was obtained by calculating difference of Sea Surface Temperature (SST) between west and east end of Hindiaocean. The data was collected from IRI site by opening ERSST data link on IRI Data Library (IRIDL). On that link IOD data (in part of data selection) was chosen based on desired time range and area. NINO3.4 data has same year range as IOD. This data can also be gained from IRI sites by applying the same way. Observation data was rainfall data ranged from 1979–2008 in15 weather stations in Indramayu. IOD and NINO3.4 SSTA were used as predictors otherwise rainfall data in dry season on May, June, July and August were ones predicted. Those rainfall data were divided into 15 weather stations namely: Bangkir, Bulak, Bondan, Cidempet, Cikedung, Juntinyuat, KedokanBunder, Krangkeng, Losarang, Lohbener, Sukadana, Sumurwatu, Sudimampir, Tugu and Ujungaris.

Data collecting was objected to gain training and testing data. Training data was used to build SVR model, whereas testing data to count accuration of finished SVR model. Testing data used were only in period of 1 year. Research method flowchart can be seen below in Figure 1.

2.2. Support Vector Regression (SVR) Process

Training data was processed using SVR training to obtain model which using rainfall data in dry season as input for the training. Kernel function applied in SVR processing was Radial Basis Function (RBF). This function has parameter value that must be determined at first, such as parameter C and γ. Those values affect significantly the resulted SVR model. More optimal the parameter leads to better built model. Search of the kernel function optimum parameter was assisted by mergering optimization algorithms of Genetic Algorithm and Particle Swarm optimization (GAPSO). SVR is application of Support Vector Machine (SVM) in term of regression. In regression case, output is real or continous number. SVR method is able to settle the over-fitting (condition when model turning too complex then causing bad prediction results) so can generate great performance [10].

SVR uses kernel function to transform non-linear input into feature room with higher dimesion because generally real world problem is rarely linear separable. Kernel function can solve non-linear separable cases like this. Afterthat, SVR will do linearcalculation to find optimal hyperplane in the feature room. Kernel projects data into high dimension feature room to increase computing ability of linear studying machine. Equation (1) of Radial Basis Function (RBF) kernel function can be seen below:

$$k(x,y) = \exp(-\gamma\|x - y\|^2) \quad (1)$$

2.3. Optimization of Support Vector Regression (SVR) using Genetic Algorithm and Particle Swarm Optimization (GAPSO)

The research implemented mergering of two optimization methods to generate optimal kernel function parameter, such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) with abbreviation of GAPSO. Previous study related to method optimized by GA and PSO (GAPSO) was conducted by Kao and Zahara [5] and Ririd [11]. Kao and Zahara [5] adjusted GAPSO optimization to multimodal function. This hybrid technique combined concepts of GA and PSO and generated new generation individual, not only by GA crossover operation and mutation but also PSO processing. The result showed advantage of GAPSO solution quality and hybrid approach convergence compared to 4 other approaches which applied 17 multimodal functions obtained from literature. Figure 2 describes concept of the two algorithms merger.

Juang[12] observed optimization of recurrent neural and fuzzy networks design. This study compared performances of algorithms optimized by GAPSO with GA and PSO. Result indicated better GAPSO performance among others which used GA or PSO.

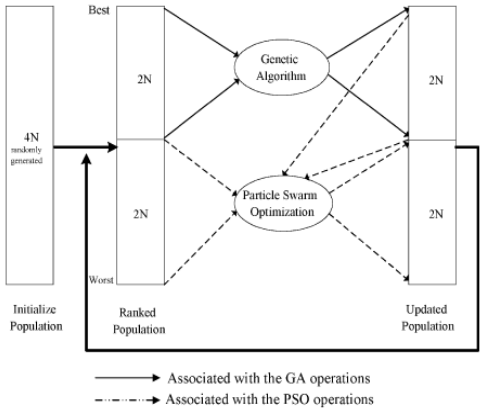


Figure 2. GA and PSO Method Classification Diagram[5]

2.4. Analysis and Evaluation

Analysis and evaluation step to prediction result was conducted after the testing. Testing data was used as input in SVR model to achieve output in form of prediction value. Accuracy and error measurement of prediction result which was obtained from SVR model to testing data used Normalized Root Mean Square Error (NRMSE) (2) and correlation coefficient (3). Error application was objected to determine deviation of predicted value compared to actual value. Error calculation used NRMSE. Correlation coefficient (R) defines connection strength between two variables. Model suitability can be achieved if R value comes near 1 and NRMSE comes near 0. Besides, analysis and evaluation also can be performed using Taylor diagram [13]. This diagram is able to evaluate several aspects from a complex model or assess reliability of some models at once. Taylor diagram was built from Root Mean Square Error (RMSE), standard deviation andcorrelation between prediction and observation.

$$NRMSE=\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{N_p}\left(y_{oi}-y_{pi}\right)^2}{\sigma_y}} \quad (2)$$

$$R=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{\left[n\sum_{i=1}^{n}x_i^2-\left[\sum_{i=1}^{n}x_i\right]^2\right]\left[n\sum_{i=1}^{n}y_i^2-\left[\sum_{i=1}^{n}y_i\right]^2\right]}} \quad (3)$$

y_{oi} = observation data in period*i*to *n*
y_{pi} = prediction result in period*i*to *n*
n = number of data
σ_y = standard deviationof prediction

x_i = observation datainperiod *i*to*n*
y_i = predictionresultinperiod *i*to*n*

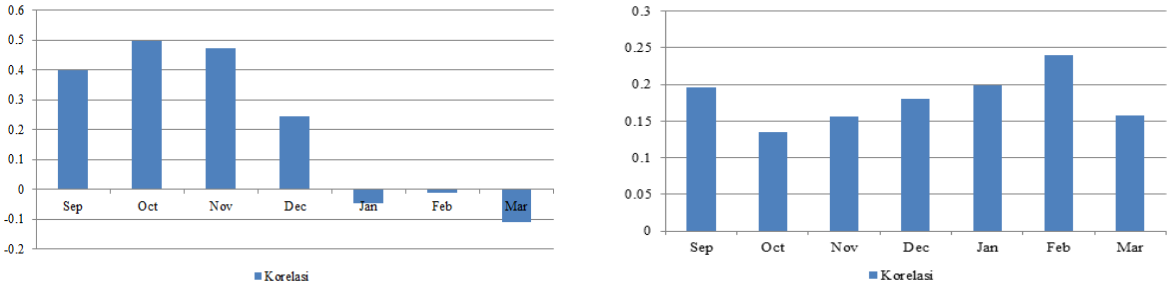
3. Results and Analysis

3.1. Data and Predictor Selection

Predictor selection in order to predict dry season rainfall on May, June, July and August (MJJA) was conducted by correlating IOD and SSTA NINO3.4 each month to dry season rainfall MJJA data from observed weather stations. Output resulted by correlating those two values showed months with IOD and SSTA NINO3.4 significantly relating each other to dry season rainfall. Not all the IOD and NINO3.4 predictor months were used, there were only ones which had highest correlation using Pearson method with used dry season rainfall selected as predictor.

Figure 3(a) shows highest IOD correlation occurred on October with value of 0.50, second highest was on November with value of 0.47 and third was on September with value of 0.40. Figure 3(b) describes NINO3.4 predictors' correlation values. Highest correlation value in NINO3.4 was obtained on February which is 0.24, January of 0.20 and September of 0.20.

Correlation value of IOD and NINO3.4 to dry season rainfall by Pearson method has negative and positive value. Negative value in IOD data correlation means inversely proportional relationship. Higher the IOD value, lower dry season rainfall. Positive value in NINO3.4 correlation means directlyproportional relationship. Higher the NINO3.4 value, higher dry season rainfall. Based on correlation result, the research is held on September, October and November as IOD predictors and September, January and February as NINO3.4 predictors.



(a) (b)

Figure 3. Correlation Value of (a) IOD and (b) NINO3.4 with MJJA dry season rainfall

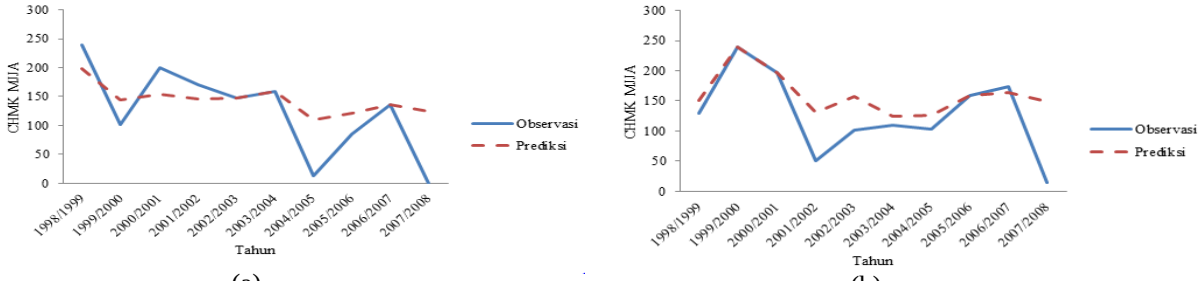
3.2. Model Performance Based on Optimization Algorithms

The research was conducted to training data of 20 years. Performance of SVR kernel function can be seen on correlation level and prediction error value compared to observation data. Model performance is assessed well if the correlation level is high and prediction error value is low.

Training using SVR needs parameter fits with its kernel. In order to obtain optimal kernel, when training occured, optimization was conducted using GAPSO (Genetic Algorithm and Particle Swarm Optimization) hybrid algorithms. Parameter optimized in RBF kernel was parameter C and γ (gamma).

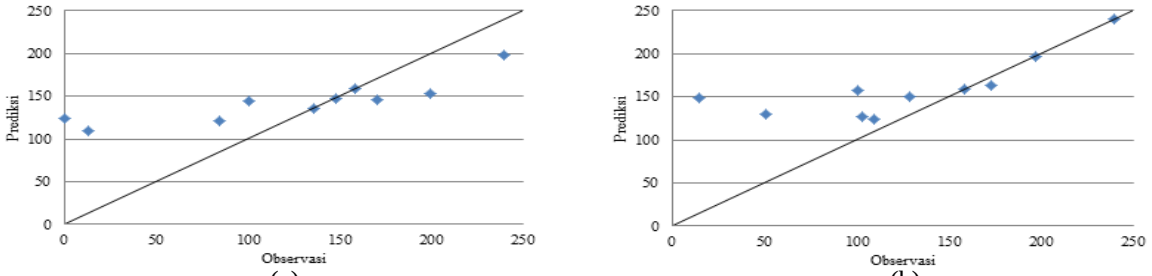
Table 1. Correlation and NRMSE Value in Rain Station of Indramayu

Station	Correlation	NRMSE
Bangkir	0.72	13.93
Bulak	0.87	11.53
Bondan	0.72	9.47
Cidempet	0.67	16.02
Cikedung	0.78	9.01
Juntinyuat	0.72	16.56
KedokanBunder	0.82	15.16
Krangkeng	0.13	32.55
Losarang	0.32	15.10
Lohbener	0.78	12.22
Sukadana	0.57	20.85
Sumurwatu	0.73	15.60
Sudimampir	0.49	18.49
Tugu	0.87	10.43
Ujungaris	0.49	17.98



(a) (b)

Figure 4. Comparison Chart of Observation and Prediction CHMK MJJA in (a) Bulak station which has highest correlation coefficient value (b) Cikedung station which has lowest NRMSE error value



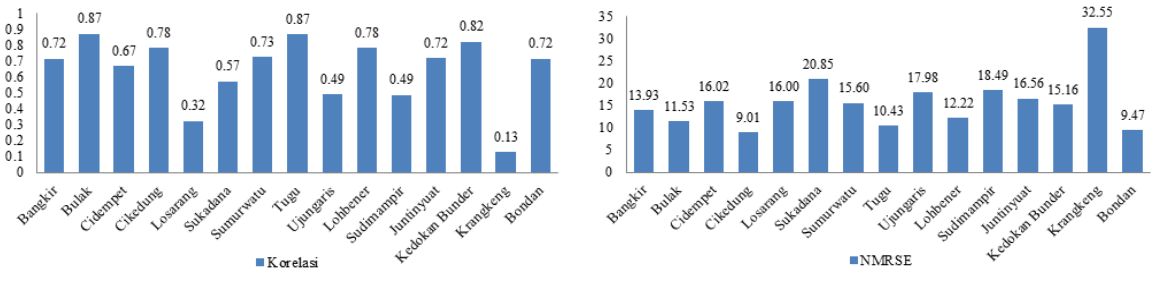
(a) (b)

Figure 5. Scatter Plot of Observation with Prediction (a) Bulak and (b) Cikedung station

Based on optimization calculation result of RBF kernel parameter using 24 populations, 100 times of iterations each process in GA and PSO, 10 iterations in GAPSO processing, value of min_c= 0.1, max_c=50, min_gamma= 0 and max_gamma=10, correlation value and NRMSE of dry season rainfall prediction MJJA each stations were obtained. Table 1 described correlation and NRMSE value each weather stations in Indramayu. More detail description of RBF kernel function performance in SVR model can be seen in comparison chart in Figure 4. The chart showed connection between observation value and dry season rainfall prediction result on May, June, July, August (CHMK MJJA). Bold connection between observation and prediction indicates stronger correlation and lower error between observed and predicted values. Figure 4 defines observation value and CHMK MJJA prediction result of Bulak station has highest correlation coefficient value, which is 0.87, and Cikedung station has lowest NRMSE error value, which is 9.01. Scatter plot in Figure 5 showed connection pattern between observation value and prediction result. Linear connection that forms straight line indicates there is firm connection between observation and prediction result.

There were extreme rainfall value in some points in observation data of year 2001/2002, 2004/2005 and 2007/2008 and other extreme points which are minimum from observation rainfall. Assumption using IOD and NINO3.4 data in those extreme points had not resulted optimal prediction value yet because its insensitiveness in responding the extreme pattern.

3.3. Analysis and Evaluation



(a) (b)

Figure 6.Value Chart of (a) correlation coefficient and (b) NRMSE error of prediction and observation result in dry season rainfall prediction in each stations

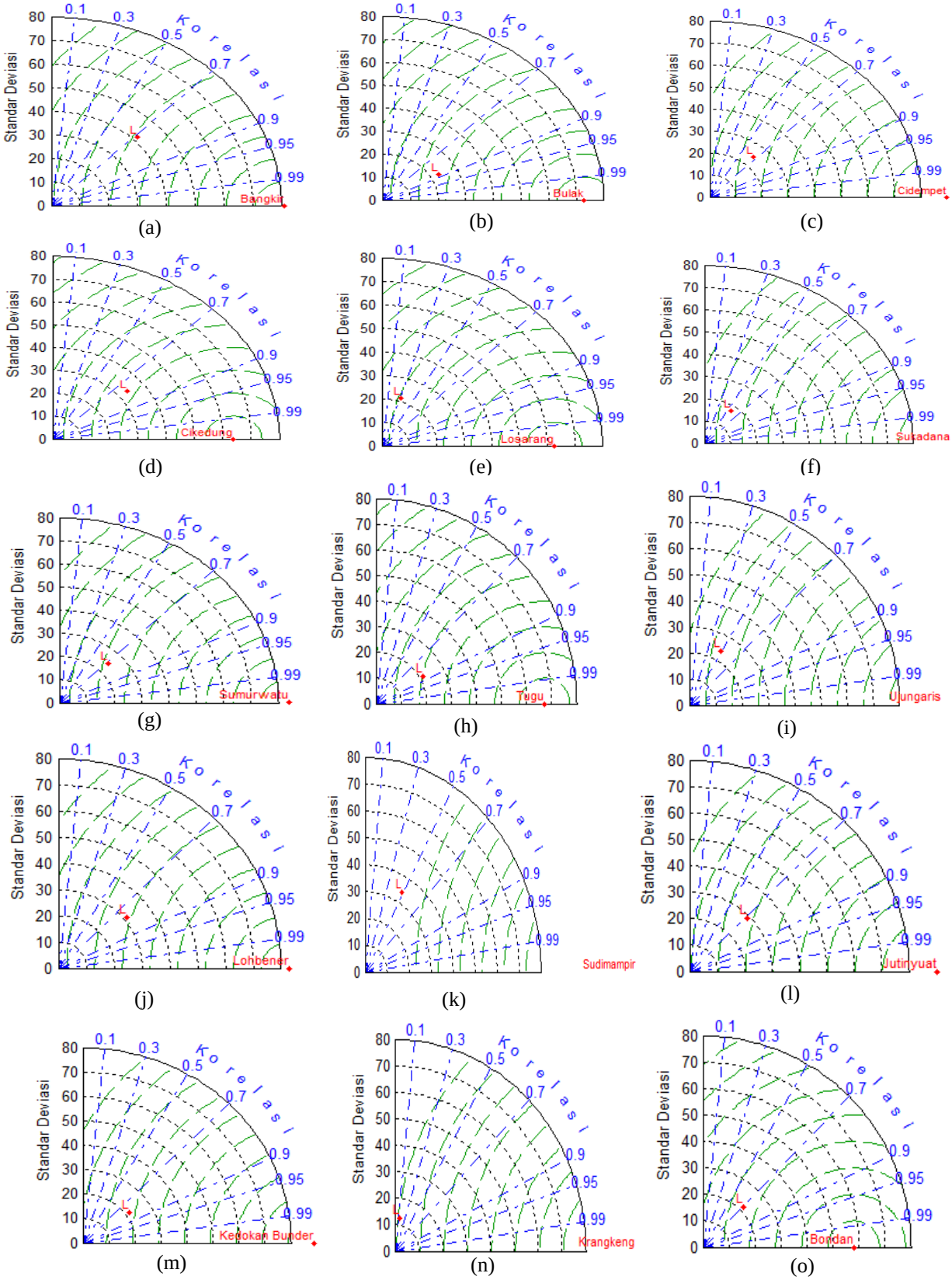


Figure 7. Taylor Diagram of each weather station in Indramayu (a) Bangkir, (b) Bulak, (c) Cidempet, (d) Cikedung, (e) Losarang, (f) Sukadana, (g)Sumurwatu, (h) Tugu, (i) Ujungaris, (j) Lohbener, (k) Sudimampir,(l) Juntinyuat, (m) Kedokan Bunder, (n)Krengkeng, (o) Bondan

Prediction of rainfall in dry season using SVR resulted varied correlation coefficient and NRMSE errorvalue. Based on RBF kernel function, Bulak station has highest correlation value whereas Cikedung station has lowest NRMSE error value. Correlation coefficient and NRMSE error value between prediction result and observation data of dry season rainfall in Indramayu are completely described in Figure 6.

Correlation coefficient value showed connection pattern between observation and prediction. Bulak station has highest correlation value of 0.87 which means 87% of observation value total diversity can be defined by its linear connection with prediction value. Figure 7 is Taylor Diagram that shows model resulted in this research generated varied outputs. The best SVR model in each stations is model with Taylor diagram position stated closest to observation point, by looking at standard deviation, RMSE and correlation. Observation point is standard deviation point in an observed location [13].

4. Conclusion

The research has successfully built model of Support Vector Regression (SVR) optimized by GAPSO hybrid algorithm in predicting rainfall in dry season with highest correlation coefficient value and lowest NMRSE value using IOD and SSTA NINO3.4 data. That SVR model was obtained using Radial Basis Function (RBF) kerenel with 24 populations, 100 times iteration each GA and PSO, 10 iterations of GAPSO, min_c = 0.1, max_c=50, min_gamma=0 and max_gamma=10. Station weather of Bulak has highest correlation coefficient value among others, which is 0.87, and NRMSE error value of 11.53. Cikedung station has lowest NMRSE error value, which is 9.01, and correlation coefficient value of 0.78. It was caused by function form that was not matched with data, or wrong parameter range collected when optimization occurred.

References

[1] Zein. PemodelanBackpropagation Neural Networks dan Probabilistic Neural Network untuk Pendugaan Awal Musim Hujan Berdasarkan Indeks Iklim Global. PhD Thesis. Bogor: Postgraduate IPB; 2014.
[2] Suciantini, Boer R, Hidayat R. Evaluasi Prakiraan Curah Hujan BMG: Studi Kasus Kabupaten Indramayu. *J. Agromet*. 2006; 20(1): 34–43.
[3] Estiningtyas W. Pengembangan Model Asuransi Indeks Iklim untuk Meningkatkan Ketahanan Petani Padi dalam Menghadapi Perubahan Iklim. Phd Dissertation. Bogor: Postgraduate IPB; 2012.
[4] Adhani G, Buono A, Faqih A. *Support Vector Regression Modelling For Rainfall Prediction in Dry Season Based on Southern Oscillation Index and Nino 3.4*. In Advanced Computer Science and Information Systems (ICACSIS). Bali. 2013: 315–320.
[5] Kao YT, Zahara E. A Hybrid Genetic Algorithm and Particle Swarm Optimization For Multimodal Functions. *Applied Soft Computing.* 2008; 8: 849–857.
[6] Aldrian E, LD Gates, FH Widodo. Seasonal Variability of Indonesian Rainfall in ECHAM4 Simulations and in The Reanalyses: The role of ENSO. *Theoretical and Applied Climatology*. 2007; 87: 41–59.
[7] Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. A Dipole Mode in the Tropical Indian Ocean. *Nature.* 1999; 401: 360-363.
[8] Ashok K, Guan Z, Yamagata T. A Look at the Relationship Between The ENSO and The Indian Ocean Dipole. *J Meteorological Society*. 2003; 18(1): 41-56.
[9] [IRI] The International Research Institute for Climate and Society(US). 2007. Monitoring ENSO. Available: http://iri.columbia.edu/climate/ENSO/background/monitoring.html. [downloaded 2012 Nov 25].
[10] Smola AJ, Schölkopf B. 2004. A Tutorial on Support Vector Regression. *Statistics and Computing*. 2004; 14: 199-222.
[11] Ririd ARTH, Arifin AZ, Yuniarti A. 2010. OptimasiMetode Discriminatively Regularized Least Square Classification DenganAlgoritmaGenetika. *JITI*. 2010; 5(3): 166-174.
[12] Juang CF. A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design. *IEEE Trans. Syst., Man. Cybern., B*. 2004; 34(2): 997-1006.
[13] Taylor KE. 2001. Summarizing Multiple Aspect of Model Performance in a Single Diagram. *J Geophysical Research: Atmospheres*. 2001; 106(D7): 7183-7192.