International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531

(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Maximum Temperature Forecast Using NWP Output and Station Data in Equatorial Region: Preliminary Result for West Java, Indonesia

Urip Haryoko[a*], Hidayat Pawitan[b], Edvin Aldrian[c], Aji Hamim Wigena[d]

[a]Indonesian Agency for Meteorology Climatology and Geophysics, Jakarta, Indonesia

[b] Department of Geophysics and Meteorology, Bogor Agricultural University, Dramaga Campus, Bogor, Indonesia

[c] Center for Research and Development, Indonesian Agency for Meteorology Climatology and Geophysics, Jakarta, Indonesia

[d] Department of Statistics, Bogor Agricultural University, Dramaga Campus, Bogor, Indonesia

[a]Email: urip.haryoko@bmkg.go.id

[b]Email: hpawitan@gmail.com

[c]Email: e_aldrian@yahoo.com

[d]Email: ajiwigena@ymail.com

Abstract

Model Output Statistics (MOS) is one of the statistical downscaling methods in post-processing of Numerical Weather Prediction (NWP) output to get weather forecasts at a point of observation stations. The problem in MOS is how to determine the spatial domain of NWP to be used as predictor in the development stage. This paper uses methods for determining the optimal NWP spatial domain and to predict maximum temperature in the Jabodetabek area using NWP output from Global Forecast System (GFS) generated by the National Oceanic and Atmospheric Administration (NOAA).

----

* Corresponding author.

E-mail address: urip.haryoko@bmkg.go.id

Maximum temperature observation data was taken from eight stations in the Jabodetabek area. In the first stage, spatial domain of NWP was defined as 8x8 grids, then attempts are made to reduce this to smaller domains, i.e. 2x2, 3x3, 3x4, 4x4 and 5x5grids. The three methods for determining suitable spatial domains were spatial correlation analysis, singular value decomposition (SVD) and partial least square regression (PLSR). These three analysis methods generally showed similar results, however spatial domain of size 3x3grids is the most suitable. By using this domain, maximum temperatures were predicted with PLSR and Principal Component Regression (PCR) methods. Besides that, these methods were also implemented in two models. Model I used a time base predictand and Model II used a time base predictor. Both methods and respected models showed comparable accuracies to forecast maximum temperature at three days ahead. The accuracy of day-i^{th} prediction was better then day-$(i+1)^{th}$ prediction. This model could increase the accuracy of GFS prediction by reducing the average error from 1,9^{o}C to 1,1^{o}C.

Keywords: statistical downscaling; numerical weather prediction; single value decomposition; partial least square regression; principle component regression.

1. Introduction

Nowadays, users need to get weather information more quickly and more precisely. The Indonesian Agency of Meteorology, Climatology and Geophysics (BKMG) has released daily weather forecasts up to one day ahead (http://www.bmkg.go.id/BMKG_Pusat/Meteorologi/ Prakiraan_Cuaca_Indonesia.bmkg) for major cities in Indonesia, including the parameters : maximum temperature (T-max), minimum temperature (T-min), maximum humidity (RH-max), the minimum humidity (RH-min), and the weather especially for rainfall occurrence (RR). Weather forecast for one week ahead are released by BMKG as weekly weather prospects as well (http://meteo.bmkg.go.id/prakiraan/mingguan). Based on the recent daily verification, percent correct of daily weather forecast (T-max, T-min, max-RH, RH-min and the weather) is about 75%.

Short-term weather forecasts were subjectively created because they rely on the skill of the forecaster and not based on the objective standard methods. The existing subjective short-term weather forecasts at BMKG needed to be replaced by an objective method in creating short-term weather forecasts. The development of a short-term weather forecast requires a long time to prepare the facilities and infrastructure, database, etc. This makes it necessary to find a weather forecast method which is fast, accurate, inexpensive and easily operated. Statistical forecasting method can be an alternative to overcome the problems [6].

Model Output Statistics (MOS) is an objective weather forecast that described the statistical relationships between predictand variables and numerical model outputs at several lead time projections. This technique was applied to predict surface wind, probability of precipitation, maximum temperature, cloud amount and conditional probability of snowfall. MOS is a useful technique for objective weather forecasting [1,2]. MOS is a process in which the statistical relationship between the output of numerical weather prediction (NWP) and observations to improve the weather forecasts. This process is often used to answer the problem of weather forecasting when certain variables are not predicted by NWP, or for downscaling if the spatial resolution of the NWP is very rough. Rough terrain, insufficient observation network and lack of knowledge about the physical processes were other issues which could reduce the performance of NWP output and it required additional processes, such as MOS [15].

Statistical interpretation of numerical model outputs was required in surface weather forecast. It was based on : 1) there is a significant difference between the real world and NWP model. The NWP uses a simplified and made homogeneous surface condition in each grid box, 2) tropical circulation and weather system were triggered by physical processes, while the physical processes of the subgrid-scale is represented in the form of a parameterization, so that NWP output may not represent the location and variables required, 3) NWP model has not perfect result, and the results still have an error, such as systematic errors that is caused by a deficiency in physical modeling, 4) NWP model are deterministic and can not fully explain the stochastic weather processes, however NWP information which is used in conjunction with statistical methods allowed us to quantify and explain the uncertainty by connecting with different conditions or probabilistic forecasts. [11,12].

Maini and his colleagues [4] developed maximum and minimum temperature forecasts using statistical interpretation of numerical weather models output. In this research, it was stated that the limitations of general circulation model (GCM) in predicting surface weather parameters required statistical interpretation of the GCM products. Forecast model of maximum and minimum temperatures at 12 locations in India has been developed with the approach of perfect prognosis method (PPM). Several models of MOS verification conducted by Maini and his colleagues [4] showed that the value of the root mean square error (RMSE) for maximum temperature forecast at the day 1 to day 4 were around 1.69°C - 3.1°C. Federico, 2011 [3] developed maximum temperature forecasts at day 1 until day 4 for the Carribean and RMSE were approximately 2.4^{o}C to 2.9^{o}C.

MOS is a post-processing method of NWP output that can be implemented by utilizing the data, facilities and infrastructure that existed at BMKG, without purchasing new equipment and data. This model does not require high performance computer. MOS utilize global forecast of NWP. NWP is a short- term weather forecasts up to three to seven days ahead. The data which could be obtained from NWP were weather forecasts with a spatial resolution to 0.5^{o}x0.5^{o} or approximately (50x50) km2 and at some levels. The objectives of this study are : 1) to specify the spatial domain of the NWP output for post-processing process, 2) to develop maximum temperature forecast model at three days ahead with one day timestamp in several places around Jakarta and West Java .

2. Methodology

2.1 Data

This study used the outputs of Global Forecasting System (GFS) with a spatial resolution (gridsize) 0.5^{o} for the period September 2010 to December 2012. NWP output is produced by mathematical model of the atmosphere and oceans to predict the weather based on current weather conditions. The NWP output used for this study has two dimensions in space. The time dimensions was forecasts time up to seven days ahead with three hour timestamp, while the spatial dimensions consist of grid boxes at the earth's surface and model level. The grids used in this study were the (8 x 8) grids located around West Java , Banten and Jakarta Province; each grid are numbered from 1 (bottom left) to 64 (top right). The sample of grid was based on daily to weekly weather phenomena defined by the WMO, the horizontal scale of daily to weekly weather variation were between 10 km to 100 km [13]. The surface weather data used in this study was maximum temperature observed by the Meteorological Station of Kemayoran (KMO), Tanjung Priok (PRI), Curug (CRG), Cengkareng (CKG), Pondok Betung (PBT), Tangerang (TNG), Bogor (BGR) and Citeko (CTK). Station locations and grid number are shown in figure 1, spatial domain number 1 to 6 are shown in Table 1 and list of station name are in Table 2.

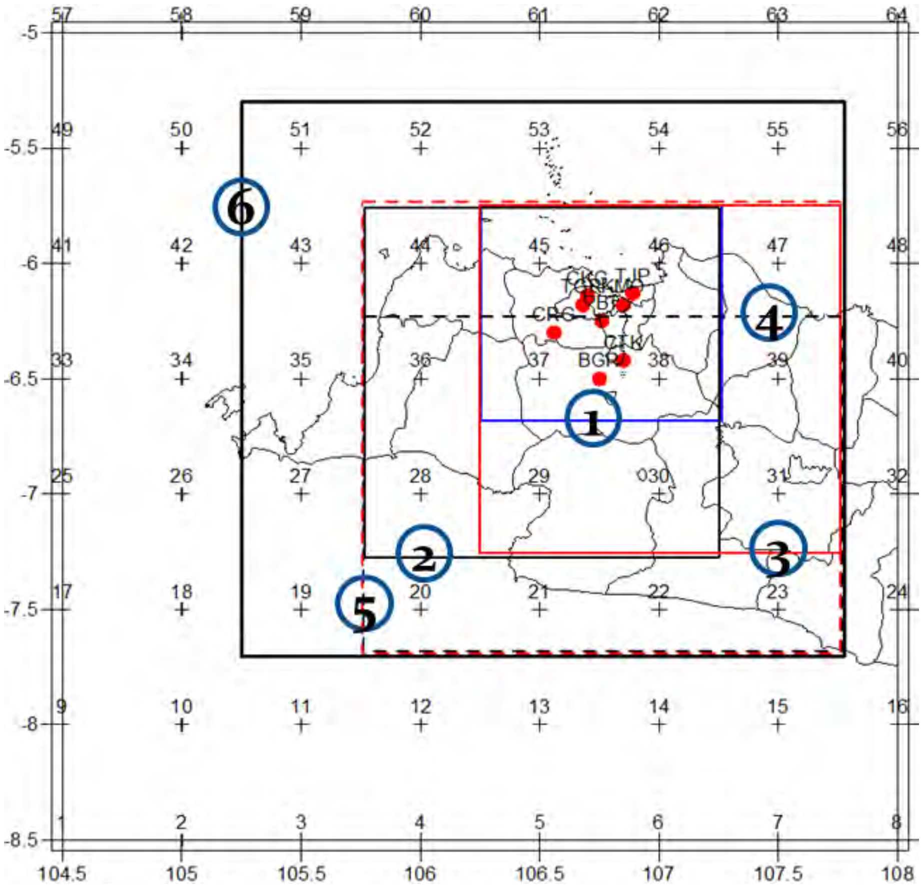

Figure 1: Station locations and grid numbers

Table 1: Spatial domain of NWP

Domain	Grid size	Grid number
domain_1	2 x 2	37 38 45 46
domain_2	3 x 3	28 29 30 36 37 38 44 45 46
domain_3	3 x 3	29 30 31 37 38 39 45 46 47
domain_4	3 x 4	20 21 22 23 28 29 30 31 36 37 38 39
domain_5	4 x 4	20 21 22 23 28 29 30 31 36 37 38 39 44 45 46 47
domain_6	5 x 5	19 20 21 22 23 27 28 29 30 31 35 36 37 38 39 43 44 45 46 47 51 52 53 54 55

2.2 Method

In this study, the processing stages are arranged as in figure 2. NWP data was obtained from the National Oceanic and Atmospheric Administration (NOAA) site on address ftp://nomads.ncdc.noaa.gov /GFS/Grid4/.The data is available in a GRIB2 format (grid binary version 2). This is a standard format recommended by WMO to compress the data into smaller sized files. The NWP forecast are initiated at 00, 06, 12 and 18 UTC, for each initiation data is produced up to 180 hours with interval of 3 hours. Furthermore, the data is converted into a text file with comma separated value format (.csv). The next process is the selecting of data by spatial and level cropping. Spatially, it only selects NWP data on a 8x8 grid, whereas the only five levels selected are the surface layer, 1000 mb, 925, 850 mb and700 mb. Local weather data was checked using a range test and if there was some suspicious and erroneous data it was checked against the original data.

Table 2: List of Station Name

No.	Station name	Code	Latitude	Longitude	Elevation (m)
1	Stasiun Meteorologi Tanjung Priok	PRI	-6.13	106.89	2
2	Stasiun Meteorologi Kemayoran	KMO	-6.18	106.85	4
3	Stasiun Meteorologi Cengkareng	CKG	-6.14	106.70	8
4	Stasiun Geofisika Tangerang	TNG	-6.18	106.68	14
5	Stasiun Klimatologi Pondokbetung	PBT	-6.25	106.76	26
6	Stasiun Meteorologi Curug	CRG	-6.30	106.56	46
7	Stasiun Meteorologi Dermaga Bogor	BGR	-6.50	106.75	207
8	Stasiun Meteorologi Citeko	CTK	-6.42	106.85	920

The statistical process was started with the spatial dimension reduction of NWP data. From the study area of 8x8 grid, the number of grids will be determined which statistically affects temperature at each point of observation. The method used in dimension reduction is isocorrelation maps, the single value decomposition (SVD) and partial least square regression (PLSR). Isocorrelation map shows the closeness between observed maximum temperature at each station with a maximum temperature from the NWP.

SVD can be seen from three viewpoints. First, SVD can be seen as a tool for transforming the correlated variables into a set that is not correlated each other so it will be better to explain the variation of the original data. Second, at the same time, the SVD is a method identifying and ordering the greatest dimension of variation to the smallest. Third, by knowing the greatest variation of the data, then it can be determined the most appoximation of data smaller dimensions. Thus SVD can be viewed as a method for reducing data dimension [2].

PLSR is a method for constructing prediction models when the independent variables are highly correlated [8,9]. The emphasis of this method is to predict the response variable, not to search for relationships between variables. PLSR method was used to build the regression model between eight observation of maximum temperature as predictand and maximum temperature of NWP as predictors at the spatial domains of different grid sizes [8] i.e. 2x2, 3x3, 3x4, 4x4, 5x5 (see figure 1). Correlation values and RMSE between forecasts and observations were used to select the spatial domain which produced the best model [14].

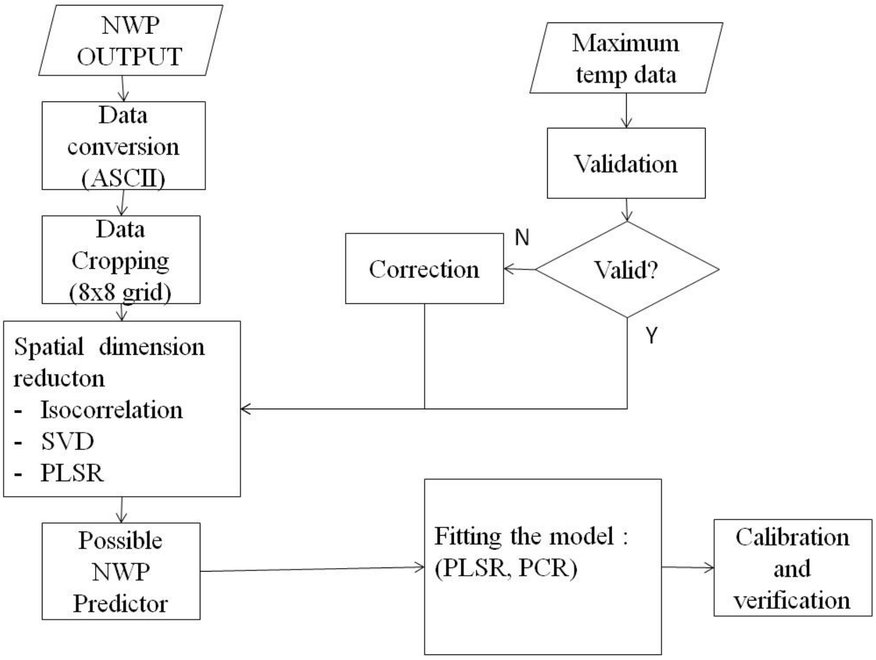

Figure 2: Processing step

The next step is to build the maximum temperature forecast model. In the development step of the model, the data is divided into two parts. The first part (September 2010 to September 2012) was used to build the model and the second part (October 2012 - December 2012) was used for verification. PLSR and Principal Component Regression (PCR) were used for developing maximum temperature forecasts model.

The statistical models are built using predictor at time t and predictand at time T. MOS regression equations of forecasts y_f were written as follows [5] :

Model I: $y_o(T) = f[x_m(t)] + \text{error}$

$y_f(t, T) = f[x_m(t)]$

Model II: $y_o(t, T) = f[x_m(t)] + \text{error}$

$y_f(t, T) = f[x_m(t)]$

where

y : predictand

x : predictors

subsctript o : observation data

subsctript f : forecasts

subsctript m : model

t : time for predictor

T : time for predictand

f : linear function where the parameters estimated from the data

The time reference used to determine predictor and prdictand are shown in Figure 3. Maximum temperature usually occures at around 14 Local Time (LT) or 07 UTC, so the NWP which was selected as a predictor of maximum temperature is NWP of 06 UTC (for day-1 forecast), 30 UTC (for day-2 forecast), 54 UTC and 78 UTC.

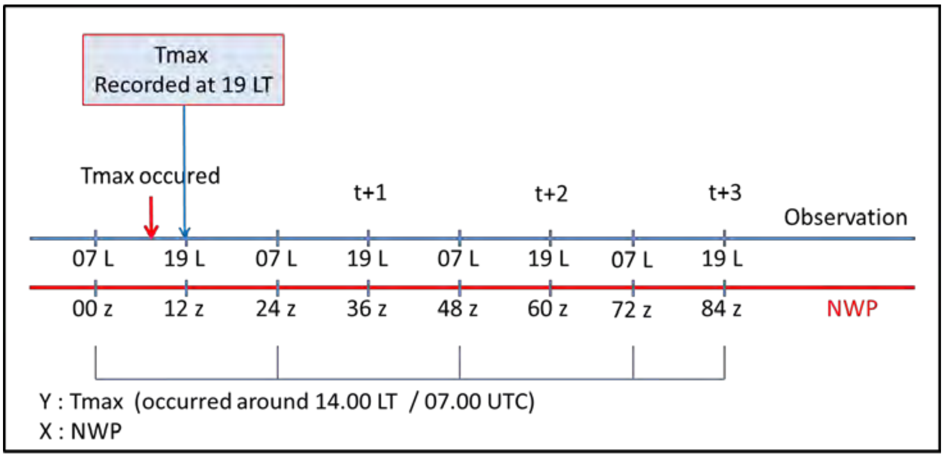

Figure 3: The time reference used to determine predictors (X) and preditand (Y)

Tabel 3: Time reference (UTC) of predictor and predictand for Model I and II

Day	Model I			Model II		
	Development		Implementation	Development		Implementation
	Y	X	X	Y	X	X
1	07 *)	6	6	07 *)	6	6
2	-	-	30	07 *)	30	30
3	-	-	54	07 *)	54	54
4	-	-	78	07 *)	78	78

Note : Y time reference based on observation date

X time reference based on NWP initiation date

*) maximum temperature occurance

Selection of predictand and predictor time in the development stage and model implementation can be seen in

Table 3. In Model I, at the development stage, using time reference 07 UTC for predictand and 06 UTC for predictors, while at the implementation stage using time reference 30 UTC for predictor (or 06 UTC at the next day) to forecast the second day, and using the predictors at 54 hours UTC and 78 UTC for day-3 and day-4. Model II, at this stage of development, using the predictor of 06 UTC, 30 UTC (forecasts 06 UTC at the next day), 54 UTC and 78 UTC, and also in the implementation stage.

GFS NWP parameters which used in this study were listed at Table 4. All parameters will be used as predictors so that there are no variation of parameters was excluded in the model. The reason is consistent with the use of PLSR and PCR methods, the both methods could eliminate multicolinierity factors. PCR method will only take the main components of 90% variation, while PLSR method will take all variation of predictor.

Tabel 4: List of predictor for maximum temperature forecast model

Parameter	Description	Unit	Surface	1000	925	850	700
Height	Geoptential height	m		x	x	x	x
Tmax	Maximum temperature	K	x				
Tmin	Minimum temperature	K	x				
T	Temperature	K	x	x	x	x	x
RH	Relative humidity	%		x	x	x	x
U	U-component	m/sec		x	x	x	x
V	V-component	m/sec		x	x	x	x
Vvel	Vertical Velocity	m/sec		x	x	x	x
Psurf	Surface pressure	Pa	x				
PRMSL	Pressure at mean sea level	Pa	x				
Pwat	Precipitable water	mm	x				

3. Result and *Discussion*

3.1 Spatial dimension reduction

3.1.1 Spatial isocorrelation

Isocorrelation map between observed maximum temperatures at each station and NWP maximum temperature at all grid point (Figure 4a and 4b) show that correlation value which is above 0.4 (significant at α = 5 %) occurred at grid number 30 and 31. While the correlation between the observed maximum temperature of each station and NWP surface temperatures at 06 UTC which greater than 0.4 occurred at grid 29, 30, 31, 32 and 36, 37, 38, 39, 40. The grids location which has a higher correlation than 0.4 is located at the grid box which has the station in it.

Grid area of NWP which have correlation greater than 0.4, almost have the same area for each station, i.e. an around one to three grids. However, for correlation with NWP surface temperature, indicating that the area is more wider for each station, i.e. about six to nine grids. Comparing with the correlation of NWP maximum temperature, the correlation with surface temperatures have more widely. By looking this correlation value and area, it can be concluded that the spatial domain which has an influence on the maximum temperature is the grid around station sites with an area of approximately one to 9 grid.

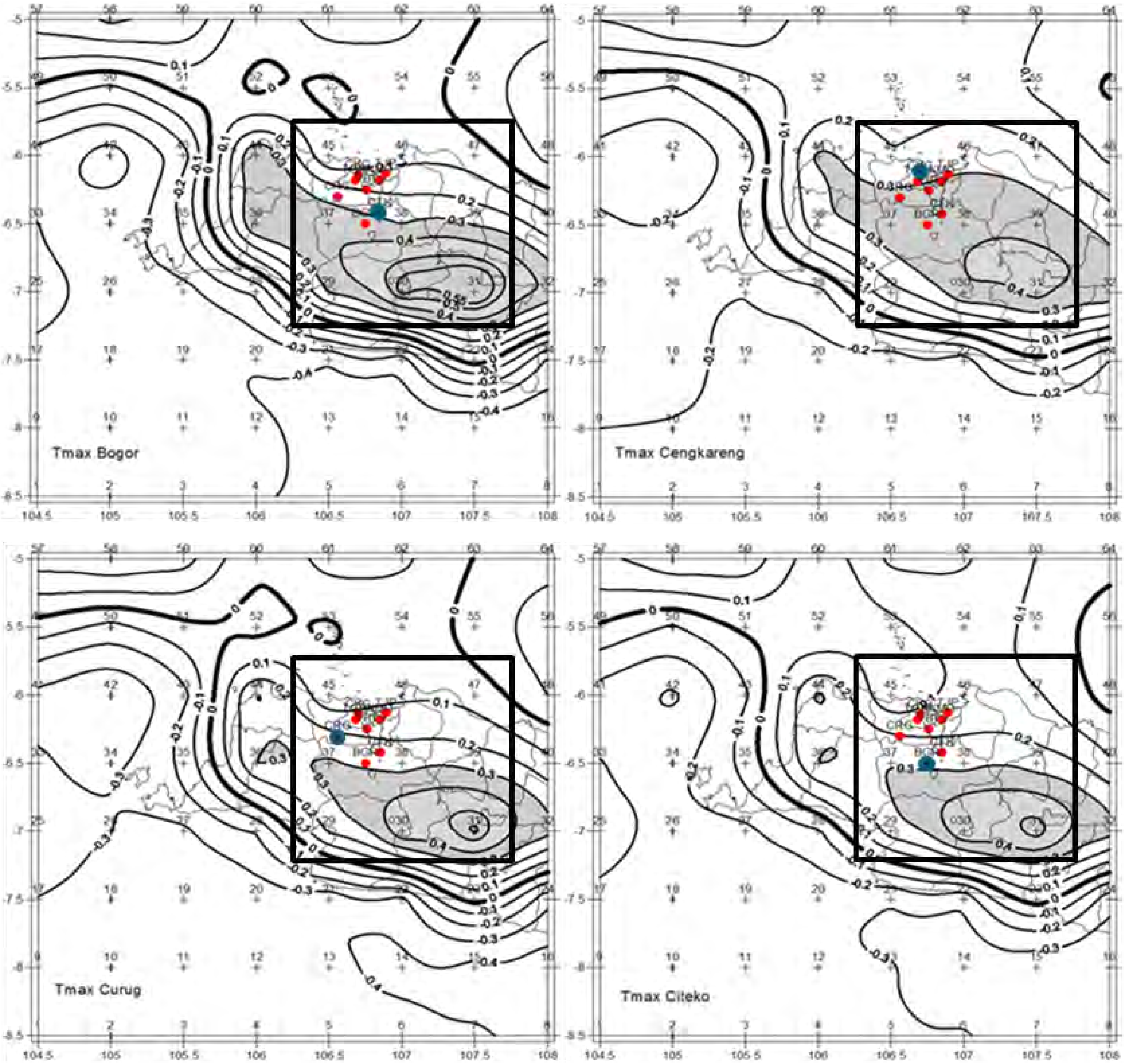

Figure 4.a: Isocorrelation map between observed maximum temperature (Bogor, Cengkareng, Curug dan Citeko) and maximum temperature of NWP for 06 UTC.

3.1.2 Single Value Decomposition

SVD analysis was applied to the six domains data, at each domain the value of the squared covariance fraction (SCF) was calculated as well as the correlation between the coefficients A and B in each expansion. Spatial linkages between the observed maximum temperature with NWP maximum temperature can be identified by looking at the variance proportion or the value of SCF. The i^{th} SCF expansion can be interpreted as the variance proportion of the observed maximum temperature and of maximum temperature from the NWP which is described by pair of i^{th} spatial pattern. The value of SCF or variance which is generated in the first expansion of all domains has reached 99 % and the highest value occurred in domain 4, namely the grid size of the 3x4 grid. This means that the 99% variance of observed maximum temperature at eight stations have been explained by the first expansion.

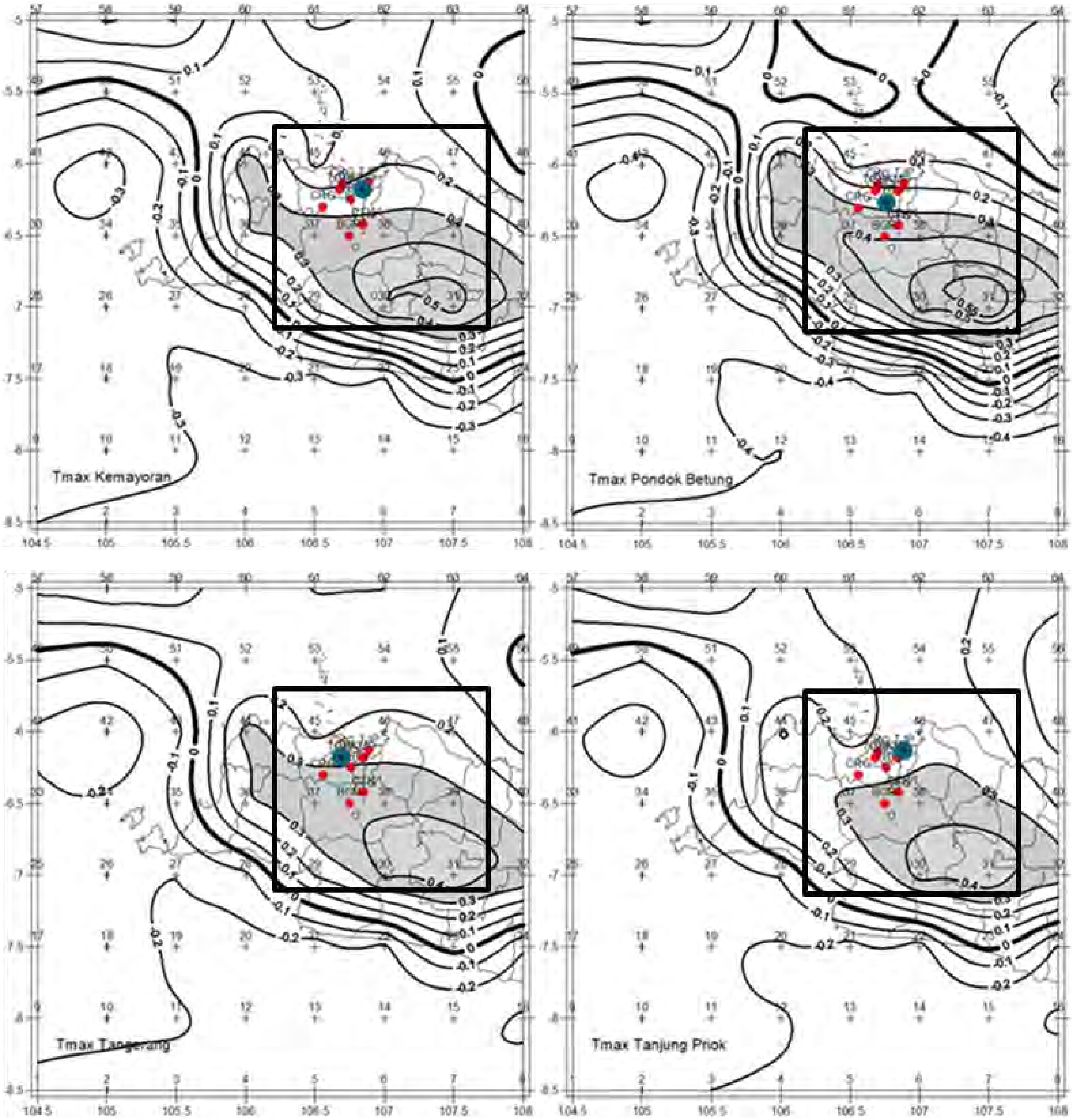

Figure 4.b: Isocorrelation map between observed maximum temperature (Kemayoran, Pondok Betung, Tangerang dan Tanjung Priok) and maximum temperature of NWP for 06 UTC.

SCF values of all domains do not show a significant difference, and all of them have reached 99%. This suggests that the SVD analysis does not distinguish between their respective domains. Furthermore, based on the correlation coefficient of the first to eight-th expansion for all domains, by the first expansion it was seen that the correlation value increases proportionally with the increase of grid domain. For domain 1 (4 grids) correlation only reached 0.41, domains 2 and 3 (9 grids) reached an average of 0.5 and domain 4 (12 grids) reached above 0.66. Based on the SVD analysis above, it can be determined that the best domain is domain 4 with grid size 3x4.

3.1.3 Partial Least Square Regression

PLSR method is used to develop forecast model between observed maximum temperature and NWP maximum temperature. The verification results using correlation value of observational data and forecasts indicated that there is a significant difference in correlations between each domain. Generally, the highest correlations occur in domain 3, namely domain with size of 3 x 3 grid (grid number : 29 30 31 37 38 39 45 46 47, further it is called as grid 1, 2 ….9 respectively). Similarly, RMSE value indicates that the smallest value recorded at domain 3.

3.2 Maximum temperature forecast model

PLSR and PCR methods are used to develop maximum temperature forecast model, each method is implemented for two models. Model I use at predictor which have the same time with predictand, while Model II use predictor at the time of NWP forecasts, so for model I it will be obtained 8 equations and Model II 32 equations for each PLSR and PCR methods. Procedure of predictors determination is an important part in the model development stage of MOS [7]. Candidate of predictors were listed in Table 4. Total number of predictors are [30 (parameter x level) x 9 grid] or 270.

At the first stage RMSE between the observed maximum temperature at eight stations and the NWP maximum temperature in grid 1 to 9 was calculated for data in October 2012 to December 2012. This stage is to test the accuracy of NWP output to forecast maximum temperature without using a statistical downscaling model.

Tabel 5: RMSE of observed and NWP maximum temperature (^{o}C) at each grid

		Station							
No	grid	PRI	CKG	TNG	KMO	PBT	CRG	BGR	CTK
1	29	5.4	4.7	5.2	5.8	5.6	5.0	4.4	2.1
2	30	5.9	5.3	5.8	6.4	6.2	5.6	5.0	2.1
3	31	8.2	7.6	8.1	8.7	8.5	7.8	7.3	2.2
4	37	3.6	2.9	3.3	3.9	3.8	3.2	2.6	3.8
5	38	2.5	1.9	2.3	2.8	2.6	2.1	1.6	5.0
6	39	*1.4*	*1.5*	*1.5*	*1.4*	*1.3*	*1.3*	*1.3*	***6.8***
7	45	2.9	2.1	2.5	3.2	3.1	2.4	2.0	4.6
8	46	3.5	2.7	3.2	3.8	3.7	3.0	2.5	3.9
9	47	3.7	2.9	3.4	4.0	3.8	3.2	2.6	3.7

Table 5 is the RMSE value between the observed maximum temperature at eight stations with NWP maximum temperature for grid 1 to 9. The smallest RMSE recorded in the Grid 6 9grid number 39) for each station, except for Citeko. Citeko is located in the Puncak Area, Bogor, which has a higher altitude (920 m) compared to the other stations (DRAMAGA Bogor = 207 m). Based on this fact, it indicates that the NWP can not forecast accurately the weather parameter at a place which has higher elevation than surrounding properties, in this case for Citeko Station.

Tables 6 and 7 are the results of model verification for three months (October 2012 to December 2012). Correlation in the table are between the maximum temperature forecasts with observed maximum temperature, while the RMSE is the average error value of the maximum temperature forecasts against observations. Day 1 states the time set as the initial forecasts, while the second day is the next day, and so on.

In general, PLSR both in Model I and Model II has almost the same level of accuracy, there was no significant difference in either of the value of the correlation and RMSE. PLSR method, both Model I and Model II, shows that the further the time predicted the level of accuracy decreases. It can be seen from the correlation value on day 1 is greater than the day 2, day 2 correlation is greater than day 3 and so on. Likewise, the value of RMSE , day 1 is smaller than day 2, day 2 smaller that day 3, and so on. The best model is for Bogor Station and the worst in Tanjung Priok Station and Tangerang.

PCR method results did not differ with PLSR. PCR method, both for Models I and II, also have the same level of accuracy. The best model is for Bogor and the worst is in Tangerang Station. Similarly, the accuracy of the forecast day 1 is more accurate than day 2, day 2 is more accurate than day 3, and so on .

The difference of RMSE between forecasts and observations without and with modeling is quite significant. Average RMSE without modeling (not including Citeko Station) is about 1.9 ° C, while the average RMSE with modeling (PCR and PLSR method for Model I or II) is 1.1 ^{o}C. Thus by using a statistical model, it can improve the accuracy of forecasting with an average of about 0.8 ^{o}C.

Tabel 6. Verification of PLSR method for Model I and Model II

Model	Statistic		Station							
		Forecast	PRI	KMO	CKG	TNG	PBT	CRG	BGR	CTK
I	Correlation	day 1	0.56	0.59	0.47	0.43	0.7	0.62	***0.74***	0.6
		day 2	0.4	0.46	0.28	0.29	0.6	0.53	0.65	0.5
		day 3	0.4	0.46	0.28	0.29	0.49	0.53	0.53	0.52
		day 4	0.25	0.4	0.25	***0.24***	0.34	0.38	0.5	0.51
	RMSE (^{o}C)	day 1	1.64	1.26	1.06	1.06	0.99	1.03	***0.8***	0.89
		day 2	1.69	1.26	1.09	1.12	1.12	1.13	0.9	1
		day 3	1.69	1.33	1.18	1.17	1.26	1.12	1.05	1
		day 4	***1.73***	1.31	1.17	1.15	1.39	1.23	1.05	1.01
	Correlation	day 1	0.57	0.58	0.46	0.38	0.69	0.61	***0.74***	0.6

		day 2	0.48	0.52	0.43	0.35	0.63	0.52	0.72	0.49
		day 3	0.48	0.49	0.32	0.34	0.54	0.56	0.62	0.47
		day 4	0.32	0.49	0.29	***0.26***	0.47	0.49	0.62	0.46
II	RMSE (^oC)	day 1	1.57	1.23	1.05	1.09	1.01	1.04	***0.79***	0.89
		day 2	1.59	1.21	1.05	1.08	1.12	1.13	0.82	1.05
		day 3	1.61	1.29	1.12	1.1	1.23	1.11	0.95	1.04
		day 4	***1.68***	1.23	1.11	1.05	1.27	1.11	0.9	1.03

Tabel 7. Verification of PCR method for Model I and Model II

Model	Statistic	Station								
		Forecast	PRI	KMO	CKG	TNG	PBT	CRG	BGR	CTK
I	Correlation	day 1	0.58	0.51	0.43	0.5	0.63	0.63	0.71	0.42
		day 2	0.56	0.49	0.45	0.43	0.59	0.56	0.69	0.38
		day 3	0.48	0.45	0.3	0.41	0.55	0.59	0.61	0.38
		day 4	0.41	0.38	0.17	0.3	0.41	0.47	0.54	0.41
	RMSE (^oC)	day 1	1.84	1.51	1.16	1.09	1.12	1.12	0.97	1.06
		day 2	1.8	1.44	1.12	1.08	1.11	1.1	0.91	1.05
		day 3	1.82	1.49	1.2	1.1	1.19	1.12	1	1.05
		day 4	1.84	1.5	1.25	1.1	1.25	1.18	1	1.03
II	Correlation	day 1	0.58	0.51	0.43	0.5	0.63	0.63	0.71	0.42
		day 2	0.59	0.5	0.45	0.42	0.61	0.57	0.7	0.38
		day 3	0.56	0.49	0.34	0.41	0.6	0.58	0.65	0.4
		day 4	0.48	0.42	0.22	0.33	0.49	0.46	0.58	0.41
	RMSE (^oC)	day 1	1.84	1.51	1.16	1.09	1.12	1.12	0.97	1.06
		day 2	1.84	1.44	1.11	1.07	1.1	1.13	0.94	1.05
		day 3	1.89	1.56	1.21	1.13	1.19	1.21	1.05	1.06
		day 4	1.93	1.6	1.25	1.12	1.24	1.26	1.06	1.07

Plot of observed and predicted maximum temperature using PCR Model I shown in Figure 5a and 5b. It apppears that both of them have the same pattern, however the model can not predicted accurately for the extreme value. Another interesting thing is the plot of Tanjung Priok Station, the difference between observations and forecasting is quite large (> 2 ° C), occurred in October 2012.

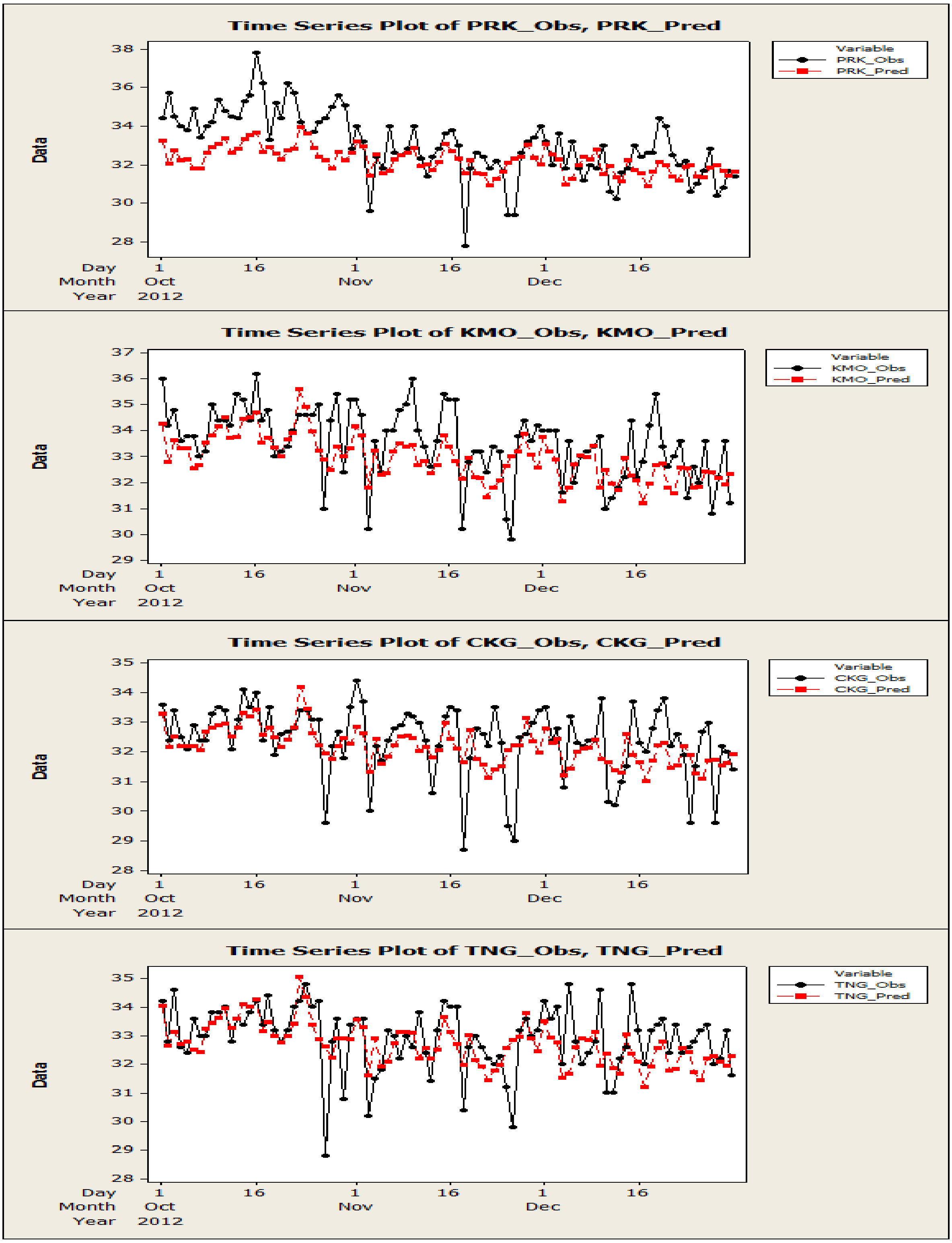

Figure 5a: Plot of observed and forecast maximum temperature of PCR Model I for Tanjung Priok, Kemayoran, Cengkareng and Tangerang.

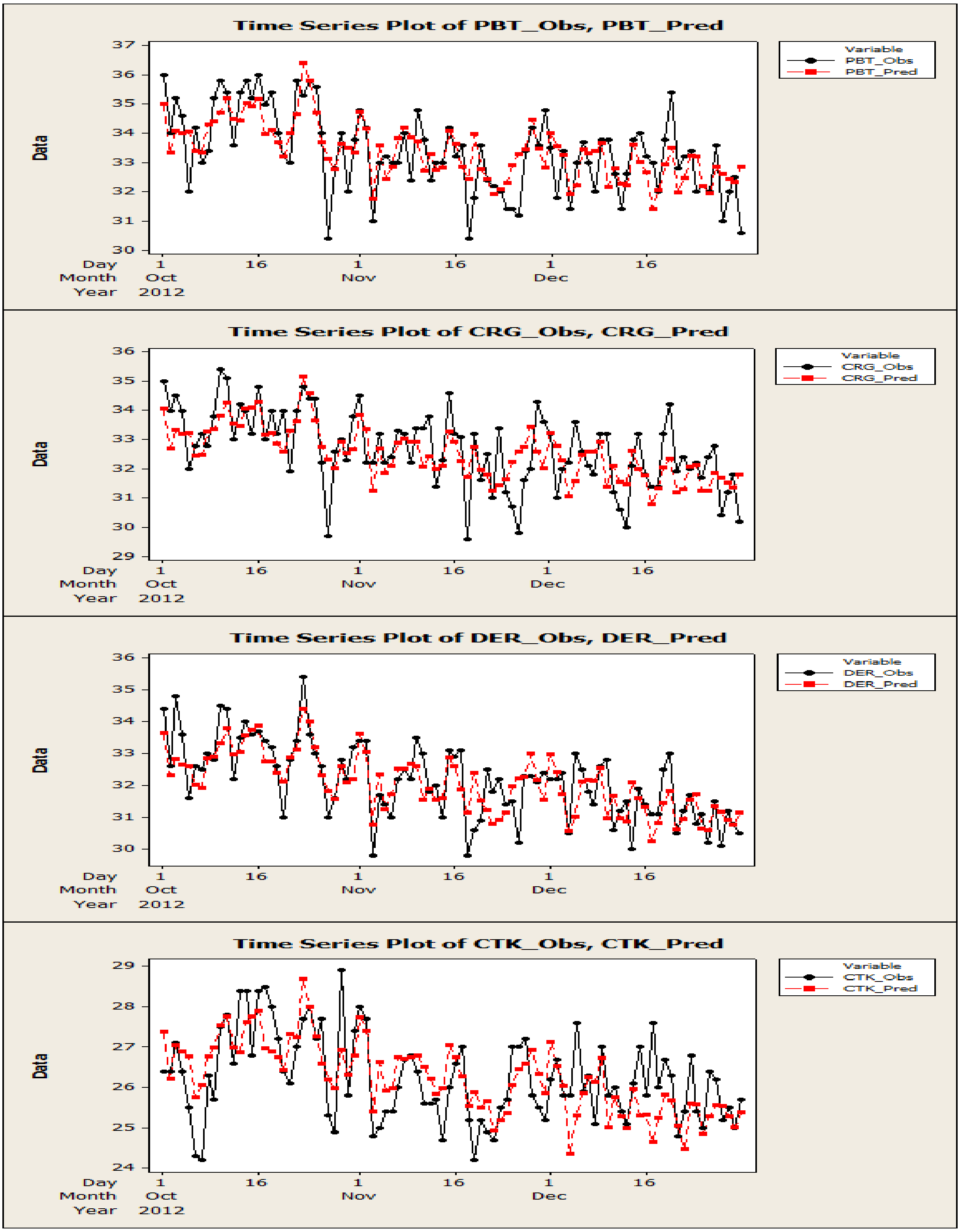

Figure 5b: Plot of observed and forecast maximum temperature of PCR Model I for Pondok Betung, Curug, Bogor and Citeko.

4. Summary and Comment

GFS NWP output has not been able to predict accurately the maximum temperature for a point location which has different characteristics to the surrounding area. This can be seen in inaccuracies of the NWP maximum temperature forecasts in Citeko Meteorological Station. Citeko Station has a significantly different height comparing with the surrounding area.

Grid Domain of 3x3 on position 29, 30, 31, 37, 38, 39, 45, 46 and 47 is potentially used as a spatial domain in the development of MOS models around West Java, Banten and Jakarta using GFS NWP data. Analysis of spatial correlation, SVD and PLSR used for determining the spatial domain showed almost the same results. However, the most optimal model for the determination of its domain is the PLSR method, because it can easily to select by verifying the value of correlation and RMSE.

MOS modeling using PLSR and PCR methods are not significantly different, both of them show the same accuracy with an average RMSE of 1.1^{o}C. These models are able to increase the accuracy of NWP forecasts which have an average RMSE of 1.9^{o}C. Similarly, the results of Model I and Model II, both showed the same accuracy of forecast, so it is sufficient to NWP outputs at the same time with the observation as predictor. The significant difference is on the accuracy of the model to predict the next day. The day-1 forecast is more accurate than the day-2, day-2 forecasts is more accurate than day-3, and so on.

It is necessary to find another method that can improve the accuracy of forecasts, especially which are able to predict the next seven days with the same accuracy level.

References

[1] Glahn HR, Lowry DA. 1972. The use of model output statistics (MOS) in Objectives Weather Forecasting. *Journal of Applied Meteorology* 11:1203-1211.

[2] Baker K. 2005. Singular Value Decomposition Tutorial [Internet] [downloaded 4 September 2012] Site: http://amsglossary. allenpress.com /glossary/.

[3] Federico S. 2011. Verification of minimum temperature, mean, and maximum temperature forecast in Caribia for summer 2008. *Natural Hazard and Earth System Sciences* Copernicus Publications 11:487-500.

[4] Maini P, Kumar A, Rathore LS, Singh SV. 2003. Forecasting maximum and minimum temperatures by statistical interpretation of numerical weather prediction model output. *Journal of Weather and Forecasting* 18:938-952.

[5] Marzban C, Sandgathe S, and Kalnay E. 2006. MOS, Perfect Prog, and Reanalysis. *Monthly Weather Review* 134:657-663.

[6] Sutikno, 2008. Statistical downscaling luaran GCM dan pemanfaatannya untuk peramalam produksi padi. Disertasi S-3 IPB.

[7] Termonia P and Deckmyn A. 2007. Model-inspired predictors for Model Output Statistics (MOS). *Monthly Weather Review* 135:3496-3505.

[8] Wigena AH. 2006. *Pemodelan statistical downscaling dengan regresi projection pursuit untuk*

peramalan curah hujan bulanan. Disertasi, Sekolah Pasca Sarjana. Institut Pertanian Bogor.

[9] Wigena AH and Djuaridah A. 2009. Pendekatan Regresi Kuadrat Terkecil Parsial Robust dalam Model Kalibrasi. *Forum Statistika dan Komputasi* 14(1):34-41.

[10] Wigena, A. H., 2011. Regresi Kuadrat Terkecil Parsial multirespon untuk statistical Downscaling (Multi Response Partial Least Square for Statistical Downscaling). Forum Statistika dan Komputasi .16(2):

[11] Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences, An Introduction.* Academic Press Inc. New York.

[12] Wilks, D. S., Hamill, T. M., 2007. Comparison of ensemble-MOS methods using GFS reforecasts. Monthly Weather Review. 135(6): 2379-2390

[13] [WMO] Word Meteorological Organization. 1999. Method of Interpreting Numerical Weather Prediction Output for Aeronautical Meteorology. Technical Note No. 195, WMO-No 770, Secretariat WMO, Geneva, Switzerland.

[14] [WMO] Word Meteorological Organization. 2000. Guideline on Performance Assessment of Public Weather Service. WMO/TD 1-23, WMO, Geneva, Switzerland.

[15] Yuval, Hsieh WW. 2003. An adaptive nonlinear MOS scheme for precipitation forecast using neural network. *Weather and Forecasting* 18:303-310.