

Spatial Bayesian Model Averaging to Calibrate Short-Range Weather Forecast in Jakarta, Indonesia

Niswatul Qona'ah^{1a*}, Sutikno^{1b}, Purnadi^{1c}

¹ Department of Statistics, Faculty of Mathematics Computing and Data Science, Institut Teknologi Sepuluh Nopember, Kampus Sukolilo, Surabaya 60111, INDONESIA. E-mail: niswatulqonaah@gmail.com^a ; sutikno@statistika.its.ac.id^b ; purnadi@statistika.its.ac.id^c

* Corresponding Author: niswatulqonaah@gmail.com

Received: 21st April 2019

Revised : 6th August 2019

Published: 30th September 2019

DOI : <https://doi.org/10.22452/mjs.sp2019no2.6>

ABSTRACT Bayesian Model Averaging (BMA) is a statistical post-processing method to calibrate the ensemble forecasts and create more reliable predictive interval. However, BMA does not consider spatial correlation. Geostatistical Output Perturbation (GOP) considers spatial correlation among several locations altogether. It has spatial parameters that modifies the forecast output to capture spatial information. Spatial Bayesian Model Averaging (Spatial BMA) is a method which combines BMA and GOP. This method is applied to calibrate the temperature forecast at 8 stations in Indonesia that is previously predicted by Numerical Weather Prediction (NWP). Temperature forecasts of BMA are used to obtain simulated spatially correlated error that modify temperature forecasts. Spatial BMA is able to calibrate the temperature forecast better than raw ensemble whose coverage comes closer to the standard 50%. Based on Root Mean Square Error (RMSE) criteria, Spatial BMA is able to correct forecast bias NWP with RMSE value of 1.399° lower than NWP of 2.180°.

Keywords: BMA; Ensemble; GOP; NWP; Spatial BMA.

1. INTRODUCTION

In the last few years, BMKG (Meteorological, Climatological and Geophysical Agency) in Indonesia has already developed the numerical weather forecasting process using Numerical Weather Prediction (NWP) to support forecasters. But the NWP forecasting has high bias because it is measured on a global scale (homogeneous) and is unable to capture the dynamics fluctuating atmosphere (BMKG, 2011; Wilks, 2006). Statistical post-processing of NWP outcome using ensemble has been used for increasing the forecasting accuracy i.e. the NWP composite of some modeling methods, such as ARIMA, Neural Network, Principal Component Regression (PCR), etc.

However, the ensemble forecasting is often still underdispersive (the centered weather forecasting on a value with low variance). As a result, the forecast interval becomes narrow and the observations cannot be contained in the forecast interval, so the ensemble calibration is required (Schmeits and Kok, 2010). Several methods for forecast calibrating of ensembles are Bayesian Model Averaging (BMA), Geostatistical Output Perturbation (GOP) and Spatial BMA, which are a combination of BMA and GOP.

BMA is a method that combines the forecasts of all members of the ensemble based on weighted average, so it does not just consider the contribution of one model from any models as in most of statistical models (Raftery and Zheng, 2003). In the Meteorology

field, BMA is the most widely used method because its performance is quite satisfactory. However, this method has the disadvantage, i.e. only considering one location and ignoring spatial correlation.

One of the spatial-based weather forecasting methods is GOP. This method is able to generate large ensembles based on the identified spatial relationship of the model error. Then, the error is added to the forecasting result from a simple regression, to obtain a calibrated forecast which is capable in capturing the spatial phenomenon (Gel et al., 2004). However, GOP generally uses only one predictor from the NWP output, such as maximum temperature or minimum temperature.

Spatial BMA is a combination of BMA and GOP methods. This method is expected to overcome the weaknesses of BMA and GOP. Like in BMA, probability density function (pdf) predictive Spatial BMA is the weighted average of conditional pdf's centered on bias correction of ensemble member models, with weights that depend on the contribution of each member. In Spatial BMA model, conditional pdf is multivariate densities with covariance structures in order to consider the spatial structure from weather observation.

In addition, the Spatial BMA model parameters have to conform to the GOP model parameters. Spatial BMA method can be used to produce statistical ensembles from an entire area of simultaneous weather observation, of any size, and at minimum computing cost. In individual location, Spatial BMA can be reduced to BMA (Berrocal et al., 2007).

In this study, the forecast of air temperature in 8 meteorological stations in Jakarta, Indonesia was calibrated with Spatial BMA. Previously, each NWP parameter in the nine measurement grids was processed first with Principal Component Analysis (PCA) to

reduce the dimension. Members of the Spatial BMA ensemble are obtained from Partial Least Square (PLS), Principal Component Regression (PCR) and Ridge regression. This study aims to calibrate air temperature forecasts, in order to obtain a better method of optimizing the NWP output and is expected to be used for short-term forecasting.

2. LITERATURE REVIEW

2.1 *Ensemble Forecasting using Model Output Statistics (MOS)*

Prediction ensemble system is a system that consists of several combinations of models that process a single deterministic outcome (deterministic forecast) (Park, 2006). NWP is one of the single deterministic forecast models commonly used by many countries. However, in processing, NWP forecasts still have a high bias, so it needs to be optimized by combining multiple NWPs to obtain accurate, precise and calibrated forecasts.

The ensemble forecast is performed by an integrated approach of several statistical modeling methods that process the same NWP output and is known as MOS. The study used 3 statistical modeling methods as ensemble members, i.e. Partial Least Square (PLS), Principal Component Regression (PCR) and Ridge Regression.

i. Partial Least Square (PLS)

PLS models the relationship between the response \mathbf{Y} and the predictor \mathbf{X} based on the latent variables simultaneously (Wold et al., 2001). PLS corresponds to PCR, i.e. forming a matrix of latent component \mathbf{T} in (1) of size $n \times c$ as a linear transformation of the predictor matrix \mathbf{X} :

$$\mathbf{T} = \mathbf{XW} \tag{1}$$

with \mathbf{W} is being the weighted matrix $p \times c$. Index n denotes the number of observations, p denotes the number of predictors, and c denotes the number of latent components. The latent component

\mathbf{T} is as a random variable \mathbf{X} and is used to predict \mathbf{Y} . The response \mathbf{Y} of size $q \times c$ with q is the number of responses. When \mathbf{T} is formed, then \mathbf{Q}' is obtained by the least squares method based on (2):

$$\mathbf{Q}' = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \tag{2}$$

Based on Wold et al. (2001), one way to model \mathbf{Y} is:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \tag{3}$$

By applying substitution involving (3), we obtain a weighted \mathbf{B} in (4):

$$\begin{aligned} \mathbf{XB} + \mathbf{F} &= \mathbf{TQ}' + \mathbf{F} \\ \mathbf{XB} &= \mathbf{XWQ}' \\ \mathbf{B} &= \mathbf{WQ}' = \mathbf{W}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \end{aligned} \tag{4}$$

Thus, it is obtained (5) to guess the response:

$$\hat{\mathbf{Y}} = \mathbf{XB} = [\mathbf{TW}^{-1}\mathbf{W}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y}] = [\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y}] \tag{5}$$

ii. **Principal Component Regression (PCR)**

Suppose that $\mathbf{x}' = [x_1 \ x_2 \ \dots \ x_p]$ is a predictor vector of a matrix \mathbf{X} of size $n \times p$. If the matrix \mathbf{A} is an orthogonal

matrix $p \times p$ with the m -th column containing the m -th eigenvector from $\mathbf{X}'\mathbf{X}$ and assumes $m \leq p$, then the PC score for each observation is as in (6):

$$\mathbf{Z} = \mathbf{XA} \tag{6}$$

where element (i) from \mathbf{Z} represents the score of the m -th PC for the i -th observation where $i = 1, 2, \dots, n$. Based on the orthogonal properties of \mathbf{A} where

$\mathbf{AA}' = \mathbf{I}$, multiple linear regression with correlated predictors can be converted into PCR as in (7):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{7}$$

with $\mathbf{XA} = \mathbf{Z}$ and $\mathbf{A}'\boldsymbol{\beta} = \boldsymbol{\gamma}$.

iii. Ridge Regression

Based on Draper and Smith (1992), ridge regression uses a non-negative constant λ to calculate the more

efficient regression coefficients, while reducing the singularity due to multicollinearity. (8) is used to calculate the Ridge regression coefficient:

$$\hat{\beta}_{ridge} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{y}. \tag{8}$$

So that obtained the estimated model based on (9):

$$\hat{y} = \mathbf{X}\hat{\beta}_{ridge}. \tag{9}$$

The λ constant can be selected intuitively (by default) or by the cross-validation technique based on the lowest RMSE. The higher λ causes the $\hat{\beta}_{ridge}$ coefficient closer to 0 or causes the parameter to have less effect on the response variable.

2.2 Spatial Bayesian Model Averaging (Spatial BMA)

Spatial Bayesian Model Averaging (Spatial BMA) is a method for post-processing ensembles statistically which is a combination of BMA (Raftery and Zheng, 2003) and GOP (Gel et al., 2004). Like BMA, pdf predictive Spatial BMA is the weighted average of

conditional pdf that centered on bias correction of ensemble member models, with the weights which are associated with the contribution of each member.

The Spatial BMA method considers the weather location $\mathbf{Y} = \{Y(s) : s \in S\}$, where S is a quite large set of locations but is finite and conditional on an ensemble,

$$\mathbf{F}_1 = \{f_1(s) : s \in S\}, \dots, \mathbf{F}_M = \{f_M(s) : s \in S\}$$

of M weather forecasts simultaneously, rather than just a single deterministic weather forecast. The Spatial BMA predictive pdf for weather forecasts is

$$g(\mathbf{Y}|\mathbf{F}_1, \dots, \mathbf{F}_M) = \sum_{m=1}^M w_m g_m(\mathbf{Y}|\mathbf{F}_m) \tag{10}$$

where w_m is the BMA weight, equal to the probability that member of m is the best among the members of the forecast ensemble, and $g_m(\mathbf{Y}|\mathbf{F}_m)$ is a conditional pdf of \mathbf{Y} if member of m is known to be the best. In practice,

conditional pdf is multivariate densities centered on the forecast member's bias correction, $\beta_{0,m}\mathbf{1} + \beta_{1,m}\mathbf{F}_m$, and has covariance structures spatially Σ_m . This condition can be denoted by:

$$(\mathbf{Y} | \mathbf{F}_m) \sim MVN(\beta_{0,m} \mathbf{1} + \beta_{1,m} \mathbf{F}_m, \Sigma_m) \tag{11}$$

In Equation (11),

$$\Sigma_m = \frac{\sigma^2}{\rho_m^2 + \tau_m^2} \Sigma \tag{12}$$

where σ^2 is BMA variance, Σ is spatial structure of GOP covariance matrix, each of ρ_m^2 and τ_m^2 is a GOP covariance parameter.

Like GOP, Spatial BMA has a multivariate predictive pdf for weather forecasts. The requirement of ensemble member m can be the best as described in equation (13):

$$\mathbf{Y} | \mathbf{F}_m = \beta_{0,m} \mathbf{1} + \beta_{1,m} \mathbf{F}_m + \mathbf{E}_{1m} + \mathbf{E}_{2m} \tag{13}$$

where each of \mathbf{E}_{1m} and \mathbf{E}_{2m} denotes a part of continuous and discrete of the conditional error.

Here is the algorithm for getting members of Spatial BMA ensemble:

- 1) Take as many $m \in \{1, \dots, M\}$ samples, with probabilities given by BMA weights w_1, w_2, \dots, w_M . This is to get members of the dynamic ensemble.
- 2) Simulate the realization of continuous and discrete parts, \mathbf{E}_{1m} and \mathbf{E}_{2m} , on the conditional error of each conditional pdf.
- 3) Use the right-hand side in equation (13) to get a bias-corrected of weather forecast $\beta_{0,m} \mathbf{1} + \beta_{1,m} \mathbf{F}_m$, with conditional error simulation, \mathbf{E}_{1m} and \mathbf{E}_{2m} .

Furthermore, we get weather forecast ensemble of Spatial BMA, with various

ensemble sizes which we want, and with minimum computing costs (Berrocal et al., 2007).

2.3 Goodness of Fit Model Evaluation

Goodness of fit models that aim to calibrate weather forecasts are not adequately measured if by using only RMSE. The other measures are also needed to check the level of bias correction and sharpness forecasts ensemble, i.e. CRPS and coverage (Feldmann, 2012).

2.3.1 Root Mean Square Error (RMSE)

The RMSE as an indicator of accuracy in (14) is obtained from the square root of MSE, which is the sum of the squares of the difference between the forecast and observation values.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

where n is the number of observation.

2.3.2 Continuous Rank Probability Score (CRPS)

CRPS, calculated based on (15), is used to check how precise the predictive

intervals produced by the calibration methods, such as Spatial BMA. The lower the CRPS value, the more reliable the predictive interval (Feldmann, 2012).

$$CRPS = \frac{1}{n} \sum_{i=1}^n crps(F_i, y_i) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} [F_i^{forecast}(y) - F_i^{obs}(y)]^2 dy \quad (15)$$

where n is number of observation, i is the time period (e.g. daily), $F_i^{forecast}(y)$ is the predictive cumulative distribution function (cdf) at time i^{th} , and $F_i^{obs}(y)$ is the empirical cdf at time i^{th} (Anggraeni, 2013). If the threshold forecast < observation, then $F_i^{obs}(y) = 0$, and 1 if the threshold forecast \geq observation.

2.3.3 Coverage

The sharpness of the ensemble forecasts can be identified through coverage in (16). If the observation is in the ensemble range, then the observation is identified to be in the coverage, where the coverage standard is derived from

$$\frac{M - 1}{M + 1} \times 100\% \quad (16)$$

The ensemble forecast is called to be calibrated if the value of coverage closes to the standard of calculation in (16).

2.4 Temperature

The atmospheric temperature is the measure of the temperature at various levels in the Earth's atmosphere that can be affected by solar radiation, humidity and altitude, thus the effect is a complex relationship among the biosphere, the lithosphere and the atmosphere (Tanudidjaja, 1993). Energy is constantly moving from the surface to the air above it which causes heat transfer.

3. METHODOLOGY

3.1 Data Source

The data used in this study are the secondary data from Meteorology, Climatology and Geophysics Agency (BMKG), i.e. the data of CCAM (conformal cubic atmospheric model) NWP data from 1st of January 2009 to 31st of December 2010 or 708 days. The location of research focus is meteorological station, i.e. Kemayoran, Priok, Cengkareng, Pondok Betung, Curug, Dermaga, Tangerang and Citeko.

3.2 Research Variables

Response variable in this study is air temperature of the observation data, i.e. maximum temperature (Celsius). Predictors

are air temperature forecasts which are modeled by PLS, PCR, and Ridge regression. Predictors are obtained from the outcome of the CCAM NWP parameter, shown in Table 1, which were initially reduced by PCA.

Table 1: NWP parameters (BMKG, 2011).

| NWP Parameter (code) | Level | Unit |
|--|--------------|-------------|
| <i>Surface Pressure Tendency</i> (dpsdt) | surface | hPa |
| <i>Water Mixing Ratio</i> (mixr) | 1, 2, 4 | g/kg |
| <i>Vertical Velocity</i> (omega) | 1, 2, 4 | knot |
| <i>PBL depth</i> (pblh) | surface | meter |
| <i>Surface Pressure</i> (ps) | surface | hPa |
| <i>Mean Sea Level Pressure</i> (psl) | surface | hPa |
| <i>Screen Mixing Ratio</i> (qgscm) | surface | g/kg |
| <i>Relative Humidity</i> (rh) | 1, 2, 4 | % |
| <i>Precipitation</i> (rnd) | surface | mm |
| <i>Temperature</i> | 1, 2, 4 | Celcius |
| <i>Maximum Screen Temperature</i> (tmaxcr) | surface | Celcius |
| <i>Minimum Screen Temperature</i> (tmincr) | surface | Celcius |
| <i>Pan Temperature</i> (tpan) | surface | Celcius |
| <i>Screen Temperature</i> (tscrn) | surface | Celcius |
| <i>Zonal Wind</i> (u) | 1, 2, 4 | knot |
| <i>Friction Velocity</i> (ustar) | surface | m/sec |
| <i>Meridional Wind</i> (v) | 1, 2, 4 | knot |
| <i>Geopotential Height</i> (zg) | 1, 2, 4 | meter |

Besides the 7 parameters that are measured in the different pressure levels, 11 other parameters were measured only in the surface level with a height of ± 2 meters above sea level. Thus, the number of NWP parameters is 32 parameters. Then, each of the 32 parameters is measured on the nine grid (3 x 3) measurements, so there are 288 parameters in total.

3.3 Steps of Analysis

The steps to apply Spatial BMA to obtain a calibrated temperature forecast are as follows:

- a. Standardize data, both for observation data and NWP parameter data.
- b. Reduce the dimensions of each NWP parameter using PCA based on the covariance matrix, then we get PC scores.
- c. Predict the air temperature with temperature observation as response and PC score as predictor.

Forecasts are generated from PLS, PCR, and Ridge regressions.

- d. Calibrate the ensemble forecasts for air temperature using BMA based on 30-day training window, starting from estimation of regression coefficient of $\beta_{0,m}$ and $\beta_{1,m}$ and calculating BMA calibrated forecast on certain day.
- e. Model the air temperature forecast using spatial modeling as in GOP, with BMA forecast results as a predictor based on 30-day training, starting from analyzing empirical semivariogram, obtaining forecasting temperature that has been added with spatial errors effect until goodness examination of Spatial BMA model.

4. RESULTS AND DISCUSSION

4.1 Standardizing Data and Dimension Reduction using PCA

The NWP parameters to be processed as predictors have various measurement units. Therefore, it is necessary to standardize using scaled and center method. It aims to minimize the difference of measurement scale between NWP parameters so that the model formed will be balanced.

Furthermore, NWP parameters need to be reduced using PCA as there is an indication of spatial relationship between grids in a NWP parameter. PCA also aims to simplify modeling and is expected to shorten the computation process without reducing

precision and accuracy. For the Dermaga meteorological station, each NWP parameter produces 1 to 2 components. Resulting in a total of 41 components of 32 NWP parameters. Similarly with other meteorological stations, it also produces 1 to 2 components on each NWP parameter. The variability of NWP parameters explained by the PC varies from 80% to 100%.

Thus, the dependency of weather conditions between grids in a NWP parameter is relatively high. The ensemble member models will be discussed in detail for Dermaga station. Meanwhile, the other stations have the same steps. Furthermore, 41 components of NWP parameter will be involved in MOS modeling in which PC scores from each of the 41 selected components are used as predictors to obtain forecasts of ensemble members.

4.2 PLS Regression for Maximum Temperature

Predicted residual error sum of square (PRESS) value are obtained from validation data that are generally randomly selected. This is to see if a regression model has accurate forecasting capabilities. For Dermaga station, the lowest PRESS value for maximum temperature is in the PLS model with 16 components, which is 0.641. After the optimal number of components is obtained and used as a latent component contributing to the modeling, the next step is to apply the PLS regression for maximum temperature. Table 2 shows the regression coefficient of PLS (in the standardized form).

Table 2: The regression coefficient of PLS for maximum temperature.

| Predictors | T _{MAX} | Predictors | T _{MAX} |
|------------|------------------|------------|------------------|
| PC.dpsdt | -0.112 | PC.tmaxscr | -0.538 |
| PC.mixr1 | 0.130 | PC.tminscr | -0.128 |
| PC.mixr2 | 0.139 | PC.tpan | 0.092 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| PC.temp1 | -0.115 | PC.zg2 | 0.020 |
| PC.temp2 | 0.318 | PC1.zg4 | -0.0002 |
| PC.temp4 | 0.017 | PC2.zg4 | 0.067 |

The next step is to form the PLS model based on the regression coefficient in Table 2. However, the model is still in the form of the PC that can be returned to the variable form in nine grids. To return to the nine grids variable, multiply the PC with each eigenvector. Here is

a summary of the eigenvector for each 41 components in Dermaga station with the first column showing the eigenvector of dpsdt (surface pressure tendency) and the last column is the eigenvector of zg4 (geopotential height).

$$\mathbf{E} = \begin{bmatrix}
 \text{dpsdt} & \text{mixr1} & \text{mixr2} & \text{mixr4} & \text{omega1} & \text{omega1} & \dots & \text{zg4} & \text{zg4} \\
 -0,3333 & -0,3411 & -0,3353 & -0,3337 & -0,2908 & -0,4402 & \dots & -0,2007 & 0,4516 \\
 -0,3333 & -0,3220 & -0,3339 & -0,3346 & -0,2546 & -0,5264 & \dots & -0,3958 & -0,2278 \\
 -0,3333 & -0,3131 & -0,3239 & -0,3301 & -0,2129 & -0,5401 & \dots & -0,4396 & 0,0187 \\
 -0,3334 & -0,3494 & -0,3396 & -0,3370 & -0,3826 & 0,0026 & \dots & -0,0704 & 0,5045 \\
 -0,3334 & -0,3565 & -0,3432 & -0,3395 & -0,3931 & 0,1631 & \dots & -0,4394 & -0,0633 \\
 -0,3334 & -0,3478 & -0,3393 & -0,3355 & -0,3614 & 0,1186 & \dots & -0,4080 & 0,1937 \\
 -0,3333 & -0,3270 & -0,3260 & -0,3233 & -0,3439 & 0,2151 & \dots & 0,1027 & 0,4975 \\
 -0,3333 & -0,3226 & -0,3281 & -0,3337 & -0,3556 & 0,3005 & \dots & -0,2504 & -0,4224 \\
 -0,3333 & -0,3174 & -0,3301 & -0,3323 & -0,3593 & 0,2454 & \dots & -0,4149 & 0,1481
 \end{bmatrix}$$

4.3 PCR for Maximum Temperature

Similar to PLS regression, the step before modeling the weather with PCR regression is to determine the optimal number of components from 41 components of the Dermaga station. Although there are similarities in the initial procedure, but PCR does not select the optimal component based on the lowest RMSE such as PLS regression. For PCR, the selected component is the cumulative number of components which is capable to represent a predictor variance of at least 80%, it means that the PC should be able

to explain data variability of at least 80% (Johnson and Wichern, 2007). It was found that 8 of the 41 components were able to represent more than 80% predictor variance, so it was decided that the optimum components number for the maximum temperature PCR model were 8.

After knowing the optimal number of components to be used in the modeling, the next step is obtaining the PCR regression for maximum temperature. The PCR regression coefficients (in the standardized form) are shown in Table 3

Table 3: The regression coefficient of PCR for maximum temperature.

| Predictor | T _{MAKS} | Predictor | T _{MAX} |
|-----------|-------------------|------------|------------------|
| PC.dpsdt | 0.011 | PC.tmaxscr | -0.072 |
| PC.mixr1 | -0.028 | PC.tminscr | -0.054 |
| PC.mixr2 | 0.015 | PC.tpan | -0.073 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| PC.temp1 | -0.058 | PC.zg2 | -0.011 |
| PC.temp2 | -0.049 | PC1.zg4 | -0.038 |
| PC.temp4 | -0.028 | PC2.zg4 | 0.016 |

Furthermore, to obtain the maximum temperature forecasts is to form the PCR model based on the regression coefficient in **4.4 Ridge Regression for Maximum Temperature**

After obtaining ensemble members from PLS and PCR models, then weather modeling with Ridge regression is able to reduce the multicollinearity effect. Unlike the previous

Table 3, then do the same steps as PLS to obtain the model in the form of variables in nine grids.

two methods, this method does not require the selection of many optimal components. This method uses a constant λ to minimize the impact of the singularity of $\mathbf{X}^T\mathbf{X}$. Figure 1 assists the visual determination of the constant λ .

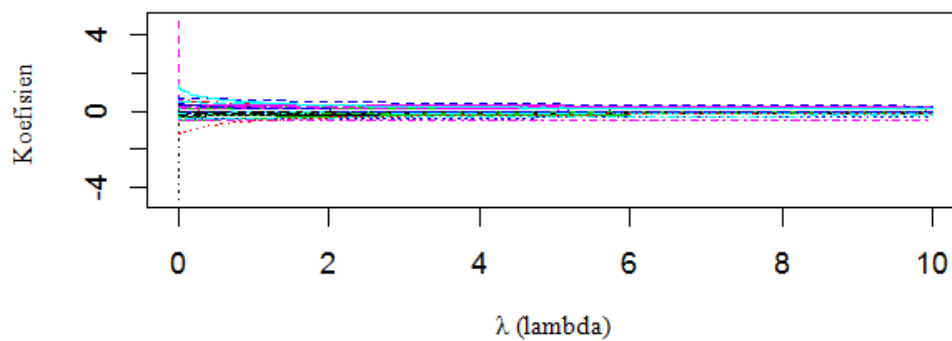


Figure 1: Convergence of ridge regression coefficients.

Determination of convergence based on Figure 1 should be avoided because it is subjective. However, in this case, the constant λ remains non convergent even when λ approaches 100. So it is decided to use visual assistance only in determining λ . Based on Draper and Smith (1992), the determination of the lower and upper limits for λ is not strictly

determined. However, the upper limit should not be greater than 10 or 20 to rid of the regression coefficients that are insignificant because they are close to 0. Based on Figure 1, it is indicated that the regression coefficients for maximum temperature to converge is when λ is 9 or greater. Table 4 is the Ridge regression coefficient for Dermaga station with $\lambda = 9$.

Table 4: The ridge regression coefficient for maximum temperature.

| Predictor | T _{MAKS} | Predictor | T _{MAX} |
|-----------|-------------------|------------|------------------|
| PC.dpsdt | -0.075 | PC.tmaxscr | -0.513 |
| PC.mixr1 | 0.123 | PC.tminscr | -0.108 |
| PC.mixr2 | 0.102 | PC.tpan | 0.053 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| PC.temp1 | -0.108 | PC.zg2 | 0.022 |
| PC.temp2 | 0.256 | PC1.zg4 | 0.001 |
| PC.temp4 | 0.008 | PC2.zg4 | 0.079 |

After the regression coefficient is obtained in the previous stage, then the next step is to form a model based on the regression coefficient in Table 4.

4.5 Description of Ensemble Member Weather Forecast

After obtaining the model to be used to predict the air temperature, the next step is to compare the forecast results of each ensemble member, i.e. PLS, PCR, and Ridge, and temperature observation values. Descriptive

analysis is performed to see how well the ensemble member predictions before being calibrated by Spatial BMA. Figure 2 below shows the first 100-day trend of 2009 from the ensemble member's predictions and maximum temperature observations for the Dermaga station.

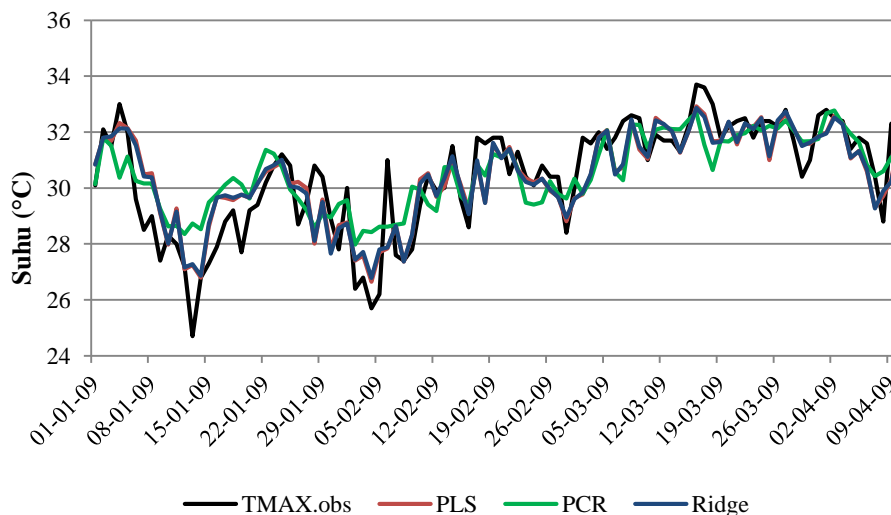


Figure 2 : Forecasts trend of ensemble members and observation of maximum temperatures.

Figure 2 indicates that each ensemble member, either PLS, PCR or Ridge, has been able to follow the general pattern of maximum temperature which is, if the maximum temperature trend increases then forecasts follow the rise, and the same if the temperature

trend decreases, the forecasts will also decrease. However, the problems occur are under-fitting (predictions below the value of observation) or over-fitting (predictions are above the value of observation) that consistently occur on the same day. Although

forecasts ensemble members can capture the temperature patterns that occur, but the forecast generated is far from the observation. Therefore, it is necessary to calibrate the model to produce more accurate and precise weather forecasts.

4.6 Calibrating Weather Forecast using Spatial BMA

Based on the previous discussion, it is indicated that the results of ensemble forecasting still have a fairly low accuracy. Therefore, statistical processing methods are needed to calibrate the forecasting results to make the forecast bias lower. Calibration is done to make adjustments to the variance, to obtain a more reliable forecast with a proportional variance and has a narrower predictor interval.

Calibration of weather forecast on 8 sites individually using BMA was done by

Luthfi (2017). Based on Raftery et al. (2005), it is said that calibration with BMA will result in better forecasts if the ensemble range has a significant correlation with the degree of forecasting error. Ensemble range is the difference between the maximum and minimum of an ensemble. However, the significant correlation does not necessarily guarantee the calibrated ensemble forecasts which are no longer under-dispersive or over-dispersive and can be identified from the Verification Range Histogram (VRH).

Figure 3 is used to identify whether the raw ensemble forecasts are under-dispersive or over-dispersive. A U-shaped histogram indicates an under-dispersive ensemble, while a histogram resembling a normal distribution curve indicates an over-dispersive ensemble. In this case, the raw ensemble consists of PLS, PCR, and Ridge which are the result of simultaneous ensemble forecasts.

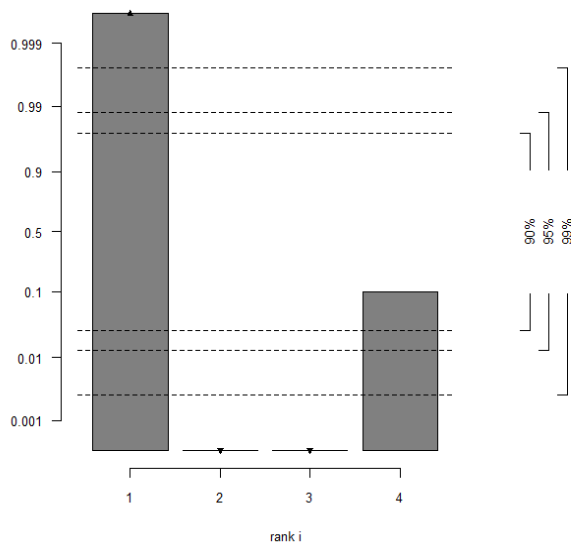


Figure 3: VRH raw ensemble of Dermaga site for maximum temperature 1st Jan '09 - 30th Jan'09.

Figure 3 shows that the raw ensemble forecasts for maximum temperatures are still under-dispersive because the shape of the histograms. This indicates that there are still many maximum temperature observations that are outside the ensemble range, i.e. the

difference between the maximum and minimum of an ensemble forecast. Based on Figure 3, the coverage of ensemble is 19.23%. This value is still far below the standard, which is 50%. It means that MOS ensemble forecasts are still under-dispersive. This can make the

predictive intervals less precise, so it is needed to calibrate using Spatial BMA.

Based on Luthfi (2017), it is indicated that BMA as a non-spatial approach that can calibrate the ensemble forecasts quite well. By applying the Spatial BMA method, it is expected to capture the spatial phenomena that occur and also utilize information from ensemble forecasts, so it can result in accurate and reliable weather forecasts. The first step in

Spatial BMA modeling is to simultaneously regress all stations of each ensemble member to the observation so as to obtain regression bias coefficient, that is, $\beta_{0,m}$ and $\beta_{1,m}$ for each member m . The regression coefficient and weight for each member m and time t are the same for 8 sites. For example, Table 5 presents the regression coefficients, weights, forecasts that will be predictors in the Spatial BMA model.

Table 5: Parameter estimates for ensemble member model of BMA, maximum temperature, Kemayoran, 8th of February 2009.

| Model | β_0 | β_1 | w | Ensemble Member Forecasts (°C) | Obs. (°C) | Simultaneous BMA (°C) |
|-------|-----------|-----------|-------|-----------------------------------|--------------|--------------------------|
| PLS | 0.27 | 0.97 | 0.771 | 29.26 | | |
| PCR | -0.77 | 0.99 | 0.095 | 28.20 | 29.30 | 28.60 |
| Ridge | 0.31 | 0.97 | 0.134 | 2.23 | | |

Table 5 shows that PLS has the highest contribution to forecasts for maximum temperatures due to w weights of 0.771, higher than PCR and Ridge whose respective weights are 0.095 and 0.134. Based on Table 5 it can be said that the simultaneous BMA forecast accuracy on February 8, 2009 did not differ significantly with the ensemble member

forecasts. Furthermore, the forecast result of the simultaneous BMA used as a predictor for the formation of Spatial BMA model with response is the observation of maximum temperature. Before that, we calculated the value of Moran's I and p -value to see the significance of spatial dependencies. Table 6 presents the values of Moran's I and p -value.

Table 6: Parameter estimates for ensemble member model of BMA, Maximum temperature, Kemayoran, 8th of February 2009.

| Weather Element | Moran's I | st.dev | p -value |
|------------------|-----------|--------|------------|
| T _{MAX} | 0.135 | 0.141 | 0.048 |

The Moran's I and p -value in Table 6 indicate that there are spatial dependencies for maximum temperatures at the significance level of $\alpha = 0.05$. Positive Moran's I values indicate that air temperatures at adjacent sites tend to have a higher relationship than distant locations. Having proven spatial dependencies between 8 sites, the first step in Spatial BMA

modeling is to obtain the estimates of parameter β_0 and β_1 , and also residual estimation to form empirical semivariogram. Estimation of parameters and presented in Table 7.

Table 7: Parameter estimation of spatial BMA model for maximum temperature.

| Weather Element | $\hat{\beta}_0$ | $\hat{\beta}_1$ | SE($\hat{\beta}_0$) | SE($\hat{\beta}_1$) |
|------------------|-----------------|-----------------|-----------------------|-----------------------|
| T _{MAX} | 0.542 | 0.990 | 0.741 | 0.025 |

Based on Table 7, it can be seen that the β_0 parameters in the T_{MAX} weather element are not significant, while the β_1 parameters are significant. Although there are insignificant parameters, this does not affect the coverage and predictive interval width of the Spatial BMA model. Furthermore, semivariogram

formed empirical exponential model of spatial parameters ρ^2 , τ^2 and r estimated based on the iterative method of L-BFGS. The semivariogram is formed from residual model Spatial BMA based on the estimation of β_0 and β_1 from Table 7. Figure 4 is a semivariogram formed.

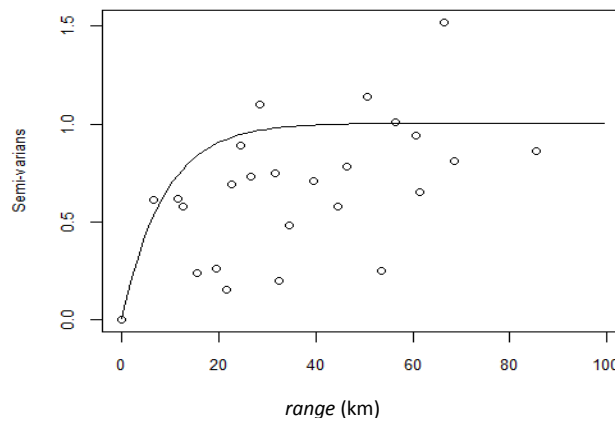


Figure 4: Empirical semivariogram for maximum temperature.

In Figure 4, the semivariogram value is constant after the range of 8.69 km or more, with a sill value (nugget + partial sill) of 1.005. A larger sill value may cause the estimation variance to become larger. This means that there is a possibility that the precision of the maximum temperature forecast is high so that it impacts the interval of Spatial BMA model that can be evaluated with CRPS. The Spatial BMA model runs a simulation process by

modifying residuals to get calibrated weather forecasts. The process is run to get 99 member realization of the ensemble. The reason for the use of 99 realizations according to Gel et al. (2004) is to accommodate the observed value when Spatial BMA is said to be calibrated when the t^{th} temperature observation falls between the two percentiles. Table 8 presents the RMSE and predictive interval of the Spatial BMA model on 2nd of March, 2009.

Table 8 : RMSE of maximum temperature forecasts using Spatial BMA and NWP, 2nd of March, 2009.

| Stamet | Obs. (°C) | NWP (°C) | Spatial BMA (°C) | RMSE Spatial BMA | RMSE NWP |
|------------|-----------|----------|------------------|------------------|----------|
| Kemayoran | 33 | 29.87 | 31.38 | | |
| Priok | 32.4 | 29.58 | 31.18 | | |
| Cengkareng | 31.9 | 29.77 | 31.04 | | |
| Pd. Betung | 34 | 29.59 | 32.57 | | |
| Curug | 33.2 | 29.23 | 31.41 | 1.399° | 2.180° |
| Tangerang | 33 | 29.67 | 32.13 | | |
| Citeko | 26 | 29.65 | 24.84 | | |
| Dermaga | 31.8 | 29.69 | 29.76 | | |

Based on Table 8, it can be said that Spatial BMA is able to correct forecast bias NWP. This is indicated by the lower RMSE forecast of Spatial BMA than the NWP forecast at the maximum temperature parameter. In addition, the coverage value is used as the calibrated indicator of forecasts. Forecast results are said to be calibrated if the value of coverage approaches the standard coverage value. The coverage standard is the

percentage comparison between the number of ensembles minus 1 by the number of ensembles plus 1. In this case, according to the number of simulated ensemble realizations, the number of ensembles is 99. So the standard coverage used is 98%. This coverage indicator can be visualized with Verification Rank Histogram (VRH). Figure 5 represents the VRH for the forecast of maximum temperature using Spatial BM.

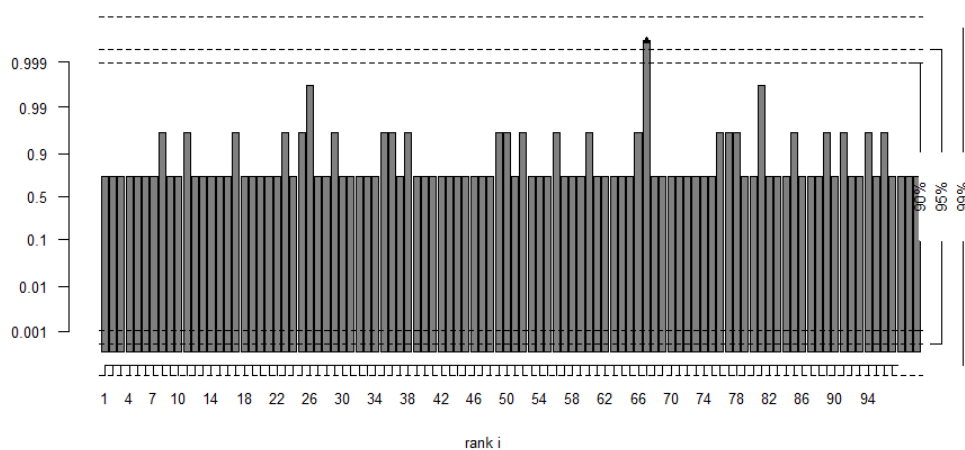


Figure 5: VRH Spatial BMA forecasts, number of ensembles = 99.

Based on Figure 5, it is known that the coverage forecasts for Spatial BMA T_{MAKS} is 87.94% which is the percentage of the number of observations that are in 2nd rank to 98th rank. The coverage value of T_{MAKS} parameters is close to 98%. This indicates that Spatial BMA is sufficiently able to calibrate the maximum

temperature forecast parameters. Furthermore, in order to be able to compare directly with the forecast results before calibration, the number of simulations of the ensemble realization is 3, in accordance with the number of ensemble members before calibration. Thus, the default coverage value used is 50%. Figure 6 is a

verification Rank Histogram (VRH) of the T_{MAX} forecasts result using Spatial BMA with the number of ensembles of 3.

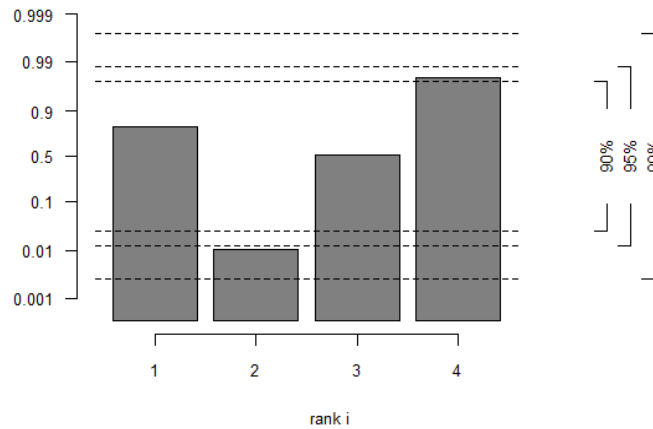


Figure 6: VRH Spatial BMA forecasts, number of ensembles = 3.

Figure 6 shows that the Spatial BMA forecast coverage for the T_{MAX} parameter is 48.63%, which is the percentage of the number of observations entered in 2nd rank and 3rd rank. The value is close to the 50% coverage standard. This indicates that the Spatial BMA with the number of ensembles of 3 has been sufficiently able to calibrate the results of the maximum temperature forecast. This indication is supported by the value of forecast coverage before being calibrated which is still far from the standard that is 19.23%.

Based on RMSE and coverage standards, it can be said that Spatial BMA method is able to improve the sharpness of the ensemble forecasts by making more observation values in the ensemble range. But in terms of accuracy, Spatial BMA has not been able to correct the forecast bias. This is supported by the CRPS values of Spatial BMA that is higher than the Simultaneous BMA. Table 9 shows the comparison of CRPS values between the Spatial BMA model and the Simultaneous BMA model (before spatially calibrated).

Table 9: CRPS value of Spatial BMA and Simultaneous BMA.

| Response | CRPS BMA Simultaneous | CRPS Spatial BMA | |
|-----------|-----------------------|------------------|------------|
| | | 99 Ensemble | 3 Ensemble |
| T_{MAX} | 0.562 | 0.580 | 0.763 |

Table 9 shows that the CRPS value of Simultaneous BMA for maximum temperature is smaller than the CRPS of Spatial BMA either with 3 ensembles or 99 ensembles. This indicates that the Spatial BMA model has not been able to correct forecast bias and improve the predictive pdf of the Simultaneous BMA

model. Table 9 also shows that the CRPS model value of Spatial BMA with 99 ensembles is smaller than the Spatial BMA model CRPS with only 3 ensembles. So that, for the Spatial BMA model it is better to use a large number of simulated realization

ensembles in order to calibrate the forecast optimally.

5. CONCLUSION

Based on coverage standards, it can be said that Spatial BMA method is able to improve the sharpness of the ensemble forecasts by making more observation values in the ensemble range. Coverage of Spatial BMA forecasts for maximum temperature increased to closer coverage standard (50%), i.e. 48.63% from previous 19.23%. In addition, based on RMSE value, Spatial BMA is able to correct forecast bias of NWP prediction with RMSE value of 1.399° lower than NWP of 2.180°.

6. ACKNOWLEDGEMENT

The entire data used in this study were supported by the Meteorology, Climatology and Geophysics Agency (BMKG) of Indonesia. The fund for this study is supported by the Ministry of Research, Technology and Higher Education of Indonesia for the grant of the National Strategic Research 2018.

7. REFERENCES

- Anggraeni, D. (2013). Kalibrasi Peramalan *Ensemble* Data Curah Hujan Dengan Metode *Ensemble Model Output Statistics* (EMOS) dan *Bayesian Model Averaging* (BMA). *Tesis*. Insitut Teknologi Sepuluh Nopember, Surabaya.
- Berrocal, V.J., Raftery, A.E., and Gneiting, T. (2007). Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecast. *Monthly Weather Review AMS*, 135: 1386-1402.
- BMKG. (2011). *Kajian dan Aplikasi Model CCAM (Conformal Cubic Atmospheric Model) untuk Prakiraan Cuaca Jangka Pendek Menggunakan MOS (Model Output Statistics)*. Jakarta: Pusat Penelitian dan Pengembangan BMKG.
- Draper, N.R. and Smith, H. (1992). *Applied Regression Analysis Second Edition*. New York: John Wiley and Sons, Inc.
- Feldmann, K. (2012). *Statistical Postprocessing of Ensemble Forecasts for Temperature: The Importance of Spatial Modeling. Diplomarbeit*. Ruperto-Carola University of Heidelberg, Germany.
- Gel, Y., Raftery, A.E., and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, 99 (467): 575–583.
- Johnson, R.A. dan Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis 5th Edition*. New Jersey: Prentice Hall.
- Luthfi, M. (2017). Bayesian Model Averaging dan Geostatistical Output Perturbation untuk Prakiraan Cuaca Jangka Pendek Terkalibrasi. *Thesis*, Insitut Teknologi Sepuluh Nopember, Surabaya.
- Park, Y.Y. (2006). Recent development of ensemble forecast system. *ASEAN-ROK Cooperation Training Workshop for the Use of Numerical Weather Prediction Products*, KMA, Seoul, South Korea, 93-177.
- Raftery, A.E. and Zheng, Y. (2003). Discussion: Performance of Bayesian Model Averaging. *Journal of the*

- American Statistical Association*, 98: 931-938.
- Raftery, A.E., Gneiting, T., Balabdoui, F. and Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review AMS*, 133: 1155-1174.
- Schmeits, M.J. and Kok, K.J. (2010). A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging and Extended Logistic Regression Using ECMWF Ensemble Precipitation Forecast. *Monthly Weather Review AMS*, 138: 4199-4211.
- Tanudidjaja. (1993). *Ilmu Pengetahuan Bumi dan Antariksa*. Jakarta: Penerbit Departemen Pendidikan dan Kebudayaan.
- Wilks, D.S. (2006). *Statistical Methods in the Atmospheric Sciences 2nd Edition*. Boston: Elsevier.
- Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109-130