Calibrating Weather Forecast using Bayesian Model Averaging and Geostatistical Output Perturbation

Muhammad Luthfi[1], Sutikno[1], Purhadi[1]

***Abstract*—Numerical Weather Prediction (NWP) has not yet been able to produce the weather forecast accurately. In order to overcome that, one approach could be taken is ensemble postprocessing. Ensemble is a combination of several methods to improve its accuracy and precision yet still possesses underdispersive nature. Bayesian Model Averaging (BMA) is intended to calibrate the ensemble prediction and create more reliable interval, though, does not consider spatial correlation. Unlike BMA, Geostatistical Output Perturbation (GOP) reckons spatial correlation among many locations altogether. Analysis applied to calibrate the temperature forecast at eight meteorological sites within Jakarta, Bogor, Tangerang and Bekasi (Jabotabek) are BMA and GOP. The ensemble members of BMA are the prediction of PLS, PCR, and Ridge. For training period over 30 days and based on some assessment indicators, BMA is better than GOP in terms of accuracy, precision, and calibration.**

***Keywords*—BMA, ensemble, GOP, NWP, Underdispersive.**

I. Introduction

In past few years, BMKG Indonesia began to apply numerical weather forecast, that is Numerical Weather Prediction (NWP), to aid the forecasters. However, its forecast bias was quite great since it has not been able to capture the dynamic atmosphere [1]. Hence, statistical post-processing needs to be applied to NWP output by using ensemble, such as combination of NWPs from several meteorological authorities. Though in many cases, ensemble forecast still possesses underdispersive nature, that is the forecast tends to concentrate at a point with low variance causing the observation outside the predictive interval, then as a consequence they need to be calibrated [2]. In order to handle such case, BMA and GOP could be applied to calibrate the ensemble forecast, among others.

As in [3], BMA combines the whole ensemble member forecast based on weighted mean, posterior probabilities, that depends to some statistical models instead of the single one. The BMA weights reflecting each member's relative skill then form the predictive PDFs. Despite its advantage, BMA merely considers weather forecast at single location and ignores the spatial correlation which frequently occurs [4]. Besides, two parameters, the weights and variances, could not be estimated by Maximum Likelihood and need the iterative approach like Expectation-Maximization (EM) algorithm.

GOP is a method of weather forecast being able to generate ensemble prediction of any size based on spatial association identified from the error correlation [5].

This method perturbs the outputs of NWP models spatially, such as error model, rather than their inputs. In other words, the simulated error has to be added to the regression linear forecast such that one would get spatially calibrated forecast. Like BMA, GOP also needs iterative approach, Limited-Memory BFGS (L-BFGS), to estimate its spatial parameters due to faster convergence when parameters being interest are large in size [6].

This research is intended to calibrate the ensemble temperature forecast at eight meteorological sites within Jabotabek, Indonesia using BMA and GOP to obtain better method being able to utilize NWP for short-range forecast, along with the brief derivation on how to obtain the parameter estimation of both methods. As this research is not equipped with enough NWP from various sources, the member of BMA consists of several statistical models, that is Partial Least Square Regression, Principal Component Regression, and Ridge Regression.

This paper is organized as follows. The materials section reviews the BMA, GOP and the indicator used to assess both method. The method section provides the data information and the way to do parameter estimation and calibrated forecasting. The results section presents the more detailed way of each method parameter estimation and forecasting. Finally, the last section gives the conclusion.

II. Materials

A. *Bayesian Model Averaging (BMA)*

BMA is a method to calibrate the underdispersive, overdispersive, and biased ensemble forecast where those might cause the forecast being less reliable, particularly the underdispersive one. The predictive PDF of BMA is linear combination of several competing model in which each of them has different weight to the PDF, relative to other

[1]Muhammad Luthfi, Sutikno, Purhadi are with Department of Statistics Faculty of Mathematics, Computing, and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: luthfi15@mhs.statistika.its.ac.id; sutikno@statistika.its.ac.id; purhadi@statistika.its.ac.id.

member [3]. They are called the posterior probabilities due to their changing value over sliding training period.

Suppose y is the observation of weather quantities with $f_1, f_2, \ldots, f_M$ are different M-forecast models where m = 1, 2, ..., M. Each member m forecast can be corrected using one of many possible methods that yields bias-corrected forecast f_m. The forecast f_m corresponds to a conditional PDF, $g_m(y|f_m)$, which could be defined as the conditional PDF of y on f_m if f_m is the best member [7].

As in [3], BMA predictive density is obtained using Eq. (1).

$$g(y \mid f_1, f_2, \ldots, f_M) = \sum_{m=1}^{M} w_m g_m(y \mid f_m), \quad (1)$$

where w_m is the m ensemble member weight or posterior probability recognized as the "best" one which is non-negative and sums up to 1, that is $w_1 + w_2 + \cdots + w_M = 1$. The weight w_m depends on forecast f_m's performance in training period.

In the case of normally distributed forecast f_m, then the interest y's posterior distribution given f_m is the best member is shown in Eq. (2).

$$y \mid f_m \sim N\left(\beta_{0,m} + \beta_{1,m} f_m, \sigma^2\right) \quad (2)$$

Based on [3], the BMA predictive mean, deterministic forecast, the weighted average of fitted member forecast, is given by Eq. (3).

$$E(y \mid f_1, f_2, \ldots, f_m) = \sum_{m=1}^{M} w_m \left(\hat{\beta}_{0,m} + \hat{\beta}_{1,m} f_m\right) \quad (3)$$

Following the BMA mean, the BMA variance is shown in Eq. (4)

$$\mathrm{var}(y \mid f_1, f_2, \ldots, f_m) = \sum_{m=1}^{M} w_m \left(\alpha_m - \sum_{m=1}^{M} w_m \alpha_m\right) + \sigma^2 , \quad (4)$$

With $\alpha_m = \hat{\beta}_{0,m} + \hat{\beta}_{1,m} f_m$ Based on Eq. (4), it could be said that the BMA variance is greater than both term on the right-hand side.

Unlike the bias-corrected coefficient β_0 and β_1, the weight w_m and variance σ^2 could not be estimated by MLE and replaced by iterative approach, that is Expectation-Maximization (EM) algorithm. It alternates between two steps, the Expectation or E step and the Maximization or M step [3]. This algorithm utilizes unobserved, latent variables z_{mt} , where z_{mt} = 1 if member m is the best forecast for time t, otherwise z_{mt} = 0.

For the normally distributed BMA model, the E step is written on Eq. (5).

$$\hat{z}_{mt}^{(i)} = \frac{w_m g_m\left(y_t \mid f_{mt}, \sigma^{(i-1)}\right)}{\sum_{l=1}^{M} w_l g_l\left(y_t \mid f_{lt}, \sigma^{(i-1)}\right)} \quad (5)$$

The superscript i in Eq. (5) refers to the i-th iteration with the density $g_m\left(y_t \mid f_{mt}, \sigma^{(i-1)}\right)$ is normal with mean $\hat{\beta}_{0,m} + \hat{\beta}_{1,m} f_{mt}$ and standard deviation $\sigma^{(i-1)}$ evaluated on observation y_t [3]. In the next M step, iterative estimation of w_m and σ^2 is calculated based on current estimate of latent z in Eq. (5). Hence, exist Eq. (6)

$$w_m^{(i)} = \frac{1}{T}\sum_{t=1}^{T} \hat{z}_{mt}^{(i)} \; ; \; \sigma^{2(i)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M} \hat{z}_{mt}^{(i)} \left(y_t - f_{mt}\right)^2 , \quad (6)$$

where T is the number of observation in the training period. Both E and M step are iterated to convergence, which is no change greater than tolerance limit in terms of parameter value and zmt in one iteration [8].

B. Geostatistical Output Perturbation (GOP)

GOP is a spatial method modifying and perturbing NWP deterministic outputs by taking the errors correlation among locations of interest into account [5]. Such errors were obtained through geostatistical simulation to get calibrated weather field forecast which is reliable and sharp as well. Considering multivariate among sites, let $\mathbf{y}_t = [y_{1t}, y_{2t}, \mathrm{K}, y_{st}]'$ and $\mathbf{x}_t = [x_{1t}, x_{2t}, \mathrm{K}, x_{st}]'$ denote the $s \times 1$ vector of observed weather quantity and vector of interest NWP output, respectively, where t = 1,2,K,T. Then, GOP model is represented in Eq. (7)

$$\mathbf{y}_t = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_t + \boldsymbol{\varepsilon}_t , \quad (7)$$

where $\mathbf{1}$ is an s x 1 unity vector and $\boldsymbol{\varepsilon}_t$ is residual vector. As in [4], error $\boldsymbol{\varepsilon}_t$ in Eq. (7) follows the normal distribution which has covariance $\boldsymbol{\Sigma}$ relying on covariance structure of error spatially. From here on, error is referred to the difference between observation and bias-corrected forecast.

Given C ($\boldsymbol{s}_i$,$\boldsymbol{s}_j$) is stationer and isotropy exponential correlation function, the (ij)th element of $\boldsymbol{\Sigma}$ is obtained based on Eq. (8).

$$\frac{1}{2}\mathrm{var}\left(\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j)\right) = \rho^2 + \sigma^2\left(1 - C(\mathbf{s}_i, \mathbf{s}_j)\right) , \quad (8)$$

where ρ^2 is nugget effect, that is the variance of measurement error as well as small-scale variability [4]. Then, the marginal variance of error is known as sill, obtained from $\rho^2 + \sigma^2$. As in Eq. (8), the error correlation might be identified through common exponential semivariogram, denoted in Eq. (9)

$$\gamma(\mathbf{d}) = \rho^2 + \sigma^2\left(1 - \exp\left(-\mathbf{d}/r\right)\right) , \quad (9)$$

where $\boldsymbol{d}$ is obtained from $\|s_i - s_j\|$ denotes the Euclidean distance between set of pair of location $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. Range r (in km) indicates the distance from which the spatial error correlation began to diminishing exponentially [4].

In order to estimate ρ^2, σ^2, and r, the applied approach is the iterative one using Limited-Memory BFGS (L-BFGS). It minimize the objective function as shown in Eq. (10), derived from weighted least square

$$g\left(\rho^2, \sigma^2, r\right) = \sum_{l=1}^{N} n_l \left(\frac{\hat{\gamma}(d_l) - \left[\rho^2 + \sigma^2\left(1 - \exp\left(-d_l / r\right)\right)\right]}{\rho^2 + \sigma^2\left(1 - \exp\left(-d_l / r\right)\right)}\right)^2 , \quad (10)$$

where n_l is number of location pair in bin B_l and N is number of obtained bin [4].

C. Verification Method

This subsection presents some methods used to assess the predictive quality obtained from calibrated ensemble forecast. Calibration is the consistency between the

ensemble forecasts and observations. For the research attempting to calibrate the weather forecast, RMSE or MAE is not sufficient to conclude the best model in terms of accuracy. It also needs other tools, such as CRPS and coverage, to verify the bias correction level and sharpness. RMSE and CRPS is expected as little as possible, while coverage is expected much closer to 50% and 90 %, respectively for BMA and GOP in this research.

1) Root Mean Square Error (RMSE)

As an accuracy indicator, RMSE in Eq. (11) is calculated from squared root of MSE, the average of sum of square of difference between forecast and verifying observation,

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (11)$$

where *n* is the number of observation [9].

2) Coverage

The sharpness of ensemble forecast could be verified from coverage by comparing the standard coverage and empirical coverage. If an observation lies within the ensemble range, then it could be said that the observation is inside the coverage [9]. The standard coverage is given by Eq. (12).

$$\frac{M-1}{M+1} x100\% \quad (12)$$

The notation *M* denotes the number of ensemble member. The ensemble forecasts is calibrated if the empirical coverage is much closer to the standard coverage.

3) Continuous Rank Probability Score (CRPS)

CRPS is used to verify how reliable or precise the predictive interval obtained from BMA or other probabilistic forecast. The less the CRPS, the more reliable the prediction interval [7]. CRPS is written on Eq. (13)

$$\mathrm{CRPS} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{crps}\left(F_i, y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[F_i^{forecast}(y) - F_i^{obs}(y)\right]^2 dy \quad (13)$$

where *n* is the number of observation, *i* is the time basis, $F_i^{forecast}(y)$and $F_i^{obs}(y)$ are the predictive CDF and empirical CDF at *i*-th time, respectively. The CDF $F_i^{obs}(y) = 1$ for observation ≥ forecast, otherwise 0.

III. METHOD

The data in this research are obtained from Meteorology, Climatology, and Geophysics Agency (BMKG) Indonesia, that is NWP Conformal Cubic Atmospheric Model (CCAM) output within period January 1st, 2009 until December 31st, 2010 or about 708 observation days. The meteorological stations of interest within Jabotabek, Indonesia are of Kemayoran, Priok, Cengkareng, Pondok Betung, Curug, Dermaga, Tangerang, and Citeko, shown on figure 1 with red dot.

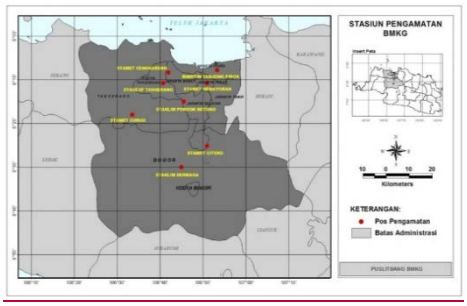

Figure 1. Meteorological stations of interest over Jabotabek [1].

The predictand (Y) of BMA and GOP is observed temperature, consisting of the maximum one and the minimum one. To produce deterministic BMA forecast, it should define the ensemble members, that is Partial Least Square Regression (PLSR), Principal Component Regression (PCR) and Ridge fitted value of temperature. Such value is obtained first by reducing dimension of each 32 NWP parameter using Principal Component (PC). Each NWP parameter has 9 grids or 3 x 3 in square that corresponds to each station, roughly illustrated on figure 2.

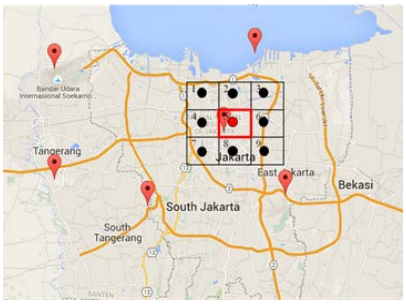

Figure 2. Implementation of 3 x 3 grid of NWP parameter.

As could be seen in figure 2, the middle red dot represents the grid nearest to the station and the rests (black dots) surround the station itself. Preprocessing using PC might be the appropriate way to summary the correlation among 9 grids solely in each NW P parameter. The short description about NWP parameters used to generate calibrated weather forecast is given on table 1.

TABLE 1.
NWP CCAM PARAMETERS SHORT DESCRIPTION [10].

NWP Parameter (code)	Level	Unit
Surface Pressure Tendency (dpsdt)	Surface	hPa
Water Mixing Ratio (mixr)	1, 2, 4	g/kg
Vertical Velocity (omega)	1, 2, 4	knot
PBL depth (pblh)	Surface	Meter
Surface Pressure (ps)	Surface	hPa
Mean Sea Level Pressure (psl)	Surface	hPa
Screen Mixing Ratio (qgscm)	Surface	g/kg
Relative Humidity (rh)	1, 2, 4	%
Precipitation (rnd)	Surface	mm
Temperature	1, 2, 4	Celcius
Maximum Screen Temperature (tmaxcr)	Surface	Celcius
Minimum Screen Temperature (tmincr)	Surface	Celcius
Pan Temperature (tpan)	Surface	Celcius
Screen Temperature (tscrn)	Surface	Celcius

Zonal Wind (u)	1, 2, 4	knot
Friction Velocity (ustar)	Surface	m/sec
Meridional Wind (v)	1, 2, 4	knot
Geopotential Height (zg)	1, 2, 4	meter

Besides seven parameters measured at three kind of levels, there are eleven parameters measured at surface level, 2 meters above sea surface. Thus, one obtain 32 parameters in total. Since this research seems to involve many meteorological variables, which are highly correlated, then it should be noted that the ensemble which consists of three statistical models is in favor to overcome multicolinearity despite the BMKG's lack of NWP source.

Thus, the predictor (X) of BMA is temperature forecast produced by PLSR, PCR, and Ridge. Meanwhile, the single predictor (X) of GOP is tmaxcr and tmincr for maximum and minimum temperature forecast, respectively, whose grid is very close to the station. This research considers 24-hour ahead forecast of NWP parameters, hence the BMA and GOP have to provide the temperature forecast for the same time ahead over 30-day training period. Based on [3], such training length gives more stability and ability to adapt with dynamical properties of weather.

IV. Results and Discussion

A. Derivation of BMA parameters estimation

The BMA parameters of each member m are able to be classified to two parts, bias-corrected coefficient $\boldsymbol{\beta}_m$, that is $\beta_{0,m}$ and $\beta_{1,m}$, alongside the weight and variance (w_m and σ^2). These parameters are used to yield the calibrated weather forecast and predictive interval. Since the weather quantity of interest is temperature which is normally distributed, then it is easily shown based on [10], either by Maximum Likelihood (ML) or Ordinary Least Square (OLS), that the estimation of $\beta_{0,m}$ and $\beta_{1,m}$ are expressed on Eq. (14)

$$\hat{\beta}_{1,m} = \frac{T\sum_{t=1}^{T} f_{mt}y_t - \left(\sum_{t=1}^{T} f_{mt}\right)\left(\sum_{t=1}^{T} y_t\right)}{T\sum_{t=1}^{T} f_{mt}^2 - \left(\sum_{t=1}^{T} f_{mt}\right)^2} = \frac{\sum_{t=1}^{T}\left(f_{mt}-\bar{f}_m\right)\left(y_t-\bar{y}\right)}{\sum_{t=1}^{T}\left(f_{mt}-\bar{f}_m\right)^2}$$

$$\hat{\beta}_{0,m} = \bar{y} - \hat{\beta}_{1,m}\bar{f}_m \tag{14}$$

Given y_t and f_{mt} are temperature and ensemble member forecast verified at time t where t = 1, 2, ..., T. Since BMA utilizes the concept of sliding training period, the entire parameters, including $\beta_{0,m}$ and $\beta_{1,m}$, changes relative to the trend value of observation and ensemble forecast [7].

Unlike $\beta_{0,m}$ and $\beta_{1,m}$ which easily could be estimated, the weight w_m and variance σ^2 is not able to be estimated by ML. In order to overcome that, BMA has to extend the derivation using iterative method, for instance using EM algorithm which considers the complete-data likelihood $L(\boldsymbol{\theta};\mathbf{y},\mathbf{z})$ based on incomplete-data likelihood $L(\boldsymbol{\theta};\mathbf{y})$ [8]. The steps needed to yield the estimation of w_m and σ^2 are briefly given below.

1) Step 1: Obtaining the incomplete-data likelihood

As usual, the incomplete-data likelihood given on Eq. (15) is carried out to estimate the weight of ensemble member m w_m.

$$L(\boldsymbol{\theta};\mathbf{y}) = \prod_{t=1}^{T} g(y_t;\boldsymbol{\theta}) = \prod_{t=1}^{T}\left(\sum_{m=1}^{M} w_m g_m(y_t \mid f_{mt})\right) \tag{15}$$

Eq. (15) might be transformed to log-likelihood to make the estimation easier and the result is expressed on Eq. (16)

$$\sum_{t=1}^{T}\left[\frac{g_j(y_t \mid f_{jt}) - g_M(y_t \mid f_{Mt})}{\sum_{m=1}^{M} w_m g_m(y_t \mid f_{mt})}\right] = 0 \tag{16}$$

It might be said that the solution for w_m based on Eq. (16) is not exist since it is definitely complicated to be derived. Therefore, the EM algorithm considering latent variable $\mathbf{Z}$ where $\mathbf{Z} = \mathbf{Z}_1^{\mathrm{T}}, \mathbf{Z}_2^{\mathrm{T}}, \mathrm{K}, \mathbf{Z}_{\mathrm{T}}^{\mathrm{T}}$ has to be applied to estimate w_m as well as σ^2 [8]. The subscript T denotes the T-th sliding training period where t = 1, 2, ..., T. It is also given that $\mathbf{Z}_{\mathrm{t}} = (\mathrm{Z}_{1t}, \mathrm{Z}_{2t}, \mathrm{K}, \mathrm{Z}_{Mt})$. For any t-th time, only one element of $\mathbf{Z}_t$ is 1, otherwise 0. Hence, for k = 1, 2, ..., M, there exists Eq. (17)

$$\mathbf{I}_{(z_{mt}=k)} = \begin{cases} 1, & z_{mt} = k \\ 0, & z_{mt} \neq k \end{cases}, \tag{17}$$

where k denotes the best ensemble member.

2) Step 2: Obtaining the complete-data likelihood

As the subsequent step after performing incomplete-data likelihood, the complete-data one is obtained by taking latent $\mathbf{Z}$ into account [8]. By applying indicator function $\mathbf{I}_{(Z_{mt}=k)}$to represent the latent, so that one obtain the likelihood on Eq. (18).

$$\begin{aligned} L(\boldsymbol{\theta};\mathbf{y},\mathbf{z}) &= \prod_{t=1}^{T}\sum_{m=1}^{M}\mathbf{I}_{(z_{mt}=k)} w_m g_m(y_t \mid f_{mt}) \\ &= \exp\left(\sum_{t=1}^{T}\sum_{m=1}^{M}\mathbf{I}_{(z_{mt}=k)}\left[\ln w_m - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2\sigma^2}\left(y_t - f_{mt}\right)^2\right]\right) \end{aligned} \tag{18}$$

3) Step 3: Performing Expectation (E) step

This E step is iteratively conducted to obtain the expectation from likelihood or log-likelihood function of complete-data. Given i as the i-th iteration, then exists Eq. (19).

$$\begin{aligned} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(i)}\right) &= E_{\boldsymbol{\theta}^{(i)}}\left[\ln L(\boldsymbol{\theta};\mathbf{y},\mathbf{z})\right] \\ &= \left[\sum_{t=1}^{T}\sum_{m}^{M} E_{\boldsymbol{\theta}^{(i)}}\mathbf{I}_{(z_{mt}=k)}\left(\ln w_m - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2\sigma^2}\left(y_t - f_{mt}\right)^2\right)\right] \end{aligned} \tag{19}$$

Based on further derivation to the expectation on Eq. (19) as in [8], it would be easily shown

$E_{\boldsymbol{\theta}^{(i)}}\mathbf{I}_{(z_{mt}=k)} = E_{\boldsymbol{\theta}^{(i)}}(z_{mt} \mid y) = \Pr_{\boldsymbol{\theta}^{(i)}}(z_{mt} = 1 \mid y) = z_{mt}^{(i)}$,

so that Eq. (20) exists

$$z_{mt}^{(i)} = \frac{w_m g_m\left(y_t \mid f_{mt}, \sigma^{(i-1)}\right)}{\sum_{l=1}^{M} w_l g_l\left(y_t \mid f_{lt}, \sigma^{(i-1)}\right)} . \tag{20}$$

The latent guess $Z_{mt}^{(i)}$ given by Eq. (20) is the posterior probability of the observation at t-th time of ensemble member m [8]. Thus, one would obtain Eq. (21) to proceed to the next M step.

$$Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(i)}\right)=\left[\sum_{t=1}^{T}\sum_{m}^{M}z_{mt}^{(i)}\left(\ln w_m-\frac{1}{2}\ln\sigma^2-\frac{1}{2}\ln 2\pi-\frac{1}{2\sigma^2}\left(y_t-f_{mt}\right)^2\right)\right] \quad (21)$$

4) Step 4: Performing Maximization (M) step

This M step selects $\boldsymbol{\theta}^{(i+1)}$ which maximizes $Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(i)}\right)$ of Eq. (21) such that $Q\left(\boldsymbol{\theta}^{(i+1)}\middle|\boldsymbol{\theta}^{(i)}\right)\geq Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(i)}\right)$ given $w_1+w_2+\mathrm{K}+w_M$ and $\boldsymbol{\theta}=(\mathbf{w}',\sigma^2)'$ [8]. The estimation of $\mathbf{w}$ might be carried out through first-order derivation of $Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(i)}\right)$ with respect to , $\frac{\partial Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(i)}\right)}{\partial\mathbf{w}}=0$, with

$$Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(i)}\right)=\left[\sum_{t=1}^{T}z_{1t}^{(i)}\right]\ln w_1+\left[\sum_{t=1}^{T}z_{2t}^{(i)}\right]\ln w_2+\ldots+\left[\sum_{t=1}^{T}z_{Mt}^{(i)}\right]\ln w_M+C$$

and

$$C=\sum_{t=1}^{T}\sum_{m}^{M}z_{mt}^{(i)}\left(-\frac{1}{2}\ln\sigma^2-\frac{1}{2}\ln 2\pi-\frac{1}{2\sigma^2}\left(y_t-f_{mt}\right)^2\right).$$

For instance, there are 2 ensemble members so that M = 2, then w_2 = 1- w_1. It would yield

$$\begin{aligned}\frac{\partial Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(i)}\right)}{\partial w_1}&=\frac{\partial}{\partial w_1}\left(\left[\sum_{t=1}^{T}z_{1t}^{(i)}\right]\ln w_1+\left[\sum_{t=1}^{T}z_{2t}^{(i)}\right]\ln w_2+C\right)\\&=\frac{\partial}{\partial w_1}\left(\left[\sum_{t=1}^{T}z_{1t}^{(i)}\right]\ln w_1+\left[\sum_{t=1}^{T}z_{2t}^{(i)}\right]\ln(1-w_1)+C\right)\\&=\frac{\sum_{t=1}^{T}z_{1t}^{(i)}}{w_1}-\frac{\sum_{t=1}^{T}z_{2t}^{(i)}}{1-w_1}=0\rightarrow w_1^{(i)}=\frac{\sum_{t=1}^{T}z_{1t}^{(i)}}{\sum_{t=1}^{T}z_{1t}^{(i)}+\sum_{t=1}^{T}z_{2t}^{(i)}}=\frac{1}{T}\sum_{t=1}^{T}z_{1t}^{(i)}\end{aligned}$$

Hence, it would be easily shown that

$$w_2^{(i)}=1-w_1^{(i)}=1-\frac{1}{T}\sum_{t=1}^{T}z_{1t}^{(i)}=\frac{1}{T}\sum_{t=1}^{T}z_{2t}^{(i)}$$

Generally, as in [3], for an ensemble with M in size where m = 1, 2, ..., M, there exists Eq. (22)

$$w_m^{(i)}=\frac{1}{T}\sum_{t=1}^{T}z_{mt}^{(i)} \quad (22)$$

Similar with w_m, the M step of which variance σ^2 should carry out yields the below derivative

$$\begin{aligned}\frac{\partial Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(i)}\right)}{\partial\sigma^2}&=\frac{\partial}{\partial\sigma^2}\left(-\frac{1}{2}\sum_{t=1}^{T}\sum_{m}^{M}z_{mt}^{(i)}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2+D\right)\\&=-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}+\frac{1}{2\left(\sigma^2\right)^2}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2+0\end{aligned}$$

with $D=\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(\ln w_m-\frac{1}{2}\ln 2\pi\right)$.

Therefore, the current estimation of σ^2 is given on Eq. (23).

$$\frac{1}{2\sigma^{2(i)}}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}=\frac{1}{2\left(\sigma^{2(i)}\right)^2}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2$$

$$\frac{1}{2\sigma^{2(i)}}\left(\sum_{t=1}^{T}z_{1t}^{(i)}+\ldots+\sum_{t=1}^{T}z_{Mt}^{(i)}\right)=\frac{1}{2\left(\sigma^{2(i)}\right)^2}\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2 \quad (23)$$

$$\sigma^{2(i)}T=\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2\rightarrow\sigma^{2(i)}=\frac{\sum_{t=1}^{T}\sum_{m=1}^{M}z_{mt}^{(i)}\left(y_t-f_{mt}\right)^2}{T}$$

As could be seen, the estimation of weight w_m and variance σ^2 on Eq. (22) and Eq. (23) is the same with Eq. (6) in materials section. For i = 0, 1, 2, ..., convergence happened with a sequence of incomplete-data likelihood values that are bounded above or in other words, $L\left(\boldsymbol{\theta}^{(i+1)};\mathbf{y}\right)\geq L\left(\boldsymbol{\theta}^{(i)};\mathbf{y}\right)$.

B. Derivation of GOP spatial parameters estimation

Before estimating the spatial parameters ρ^2, σ^2 and r, the first step one should to do is calculating empirical semivariogram $\hat{\gamma}(d_1)$ based on Eq. (24)

$$\hat{\gamma}\left(d_l\right)=\frac{1}{2k}\sum_{l=1}^{k}\left(\varepsilon\left(\mathbf{s}_{i+d_l}\right)-\varepsilon\left(\mathbf{s}_i\right)\right)^2 \quad (24)$$

where k is the number of pair of distance between two locations and $d_1=d_1(1),d_1(2),K$ is the distance which represents the whole pair of two locations being involved [4]. The next step is carrying out the estimation of those three parameters based on objective function given on Eq. (25).

$$g\left(\rho^2,\sigma^2,r\right)=\sum_{l=1}^{k}n_l\left(\frac{\hat{\gamma}\left(d_l\right)}{\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)}-1\right)^2,\rho^2,\sigma^2,r\geq 0 \quad (25)$$

Due to complicated derivative, then the step of those parameters estimation is only represented by nugget ρ^2. By applying the first-order derivation of Eq. (25) with respect to ρ^2, then equals to 0, one would obtain Eq. (26).

$$\begin{aligned}\frac{\partial g\left(\rho^2,\sigma^2,r\right)}{\partial\rho^2}&=\frac{\partial}{\partial\rho^2}\left(\sum_{l=1}^{k}n_l\left(\frac{\hat{\gamma}\left(d_l\right)}{\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)}-1\right)^2\right)\\&=\sum_{l=1}^{k}n_l\frac{\partial}{\partial\rho^2}\left(\hat{\gamma}\left(d_l\right)\left[\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)\right]^{-1}-1\right)^2\\&=\sum_{l=1}^{k}n_l2\left(\hat{\gamma}\left(d_l\right)\left[\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)\right]^{-1}-1\right)\times\\&\quad\left(-\hat{\gamma}\left(d_l\right)\left[\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)\right]^{-2}\right)\\&=-2\sum_{l=1}^{k}n_l\frac{\hat{\gamma}\left(d_l\right)}{[\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)]^2}\left(\frac{\hat{\gamma}\left(d_l\right)}{\rho^2+\sigma^2\left(1-\mathrm{e}^{-d_l/r}\right)}-1\right)\end{aligned} \quad (26)$$

Eq. (26) shows unclosed-form of ρ^2 estimation as it still contains partial sill σ^2 and range r which are going to be estimated as well. Thus, one need L-BFGS approach to obtain ρ2, σ2 and r estimation simultaneously. If $\mathbf{z}=\left[\rho^2,\sigma^2,r^2\right]$, then exists the gradient vector

$$\nabla g(\mathbf{z}) = \left[\frac{\partial g\left(\rho^2, \sigma^2, r\right)}{\partial \rho^2} \quad \frac{\partial g\left(\rho^2, \sigma^2, r\right)}{\partial \sigma^2} \quad \frac{\partial g\left(\rho^2, \sigma^2, r\right)}{\partial r} \right]'$$

with the following steps to carry out L-BFGS, where k = 0, 1, 2, ... denotes the iteration [6].

1. Determining the initial value of $\mathbf{z}$ and $\mathbf{H}$, that is $\mathbf{z}_0$ and $\mathbf{H}_0$. On k-th iteration, $\mathbf{z}_0$ should be non-negative with $\mathbf{H}_0$ is symmetric positive definite matrix, such as identity matrix. This first step determines a positive integer m as well to seek how long $\mathbf{H}_0$ information used to renew iteration. Then, determining β and γ where
$$0 < \gamma < 0.5$$
$$\gamma < \beta < 1$$
In general, m should be less than 10 in order that the iteration is running shortly as well as effectively.
2. Calculating $\Delta \mathbf{z}_k = -\mathbf{H}_\mathbf{k} \nabla_g(\mathbf{z}_k)$ and $\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha_k \Delta \mathbf{z}_k$ where $\nabla_g(\mathbf{z}_k)$ is the gradient at $\mathbf{z}_k$ with constant α_k satisfying Wolfe condition.
3. If $\left\| \nabla_g(\mathbf{z}_{k+1}) - \nabla_g(\mathbf{z}_k) \right\| < \varepsilon$, with ε arguably small value, then the current iteration should be terminated. Otherwise, the iteration should proceed to next step.
4. Updating matrix $\mathbf{H}_k$, shown on Eq. (27), based on the information from Hessian $\mathbf{H}_0$ m times so that one obtain $\mathbf{H}_{k+1}$ with $\hat{m} = \min(k+1, m)$
$$\mathbf{H}_{k+1} = \mathbf{V}_k^T \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T \quad (27)$$
where
$$\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}; \mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$$
$$\{\mathbf{s}_i, \mathbf{y}_i\} = \{\mathbf{z}_{i+1} - \mathbf{z}_i, \nabla g(\mathbf{z}_{i+1}) - \nabla g(\mathbf{z}_i)\}, i = k - \hat{m} + 1, \ldots, k$$
5. Returning to the second step to obtain the new $\Delta \mathbf{z}_k$ and updating $\mathbf{z}_{k+1}$ to check the convergence where $k = k + 1$.

From here on, the discussion of Dermaga station in Bogor, West Java is explained more specific rather than other stations since Bogor is not only the location at which the weather changes more uncertain, but also the steps taken to analyze the PCA pre-processing until produce the calibrated forecast using BMA and GOP are similar.

C. PCA Pre-Processing and Statistical Model Forecast

As have been said, the NWP parameters should be dimensionally reduced by PCA. PCA pre-processing might reduce the model complexity as well, so that it enables the further analysis and interpretation easier to handle. For Dermaga station, table 2 provides the number of PC representing each NWP parameter (variable) with its cumulative variance.

TABLE 2.
PC REPRESENTATION OF EACH NWP PARAMETER AT DERMAGA.

Variable	Num. of PC	Cum. Variance	Variable	Num. of PC	Cum. Variance
dpsdt	1	99.98%	temp2	1	93.13%
mixr1	1	84.80%	temp4	1	97.26%
mixr2	1	92.96%	tmaxscr	1	96%
mixr4	1	95.76%	tminscr	1	83.80%
omega1	2	88.56%	tpan	1	88.61%
omega2	2	87.37%	tscrn	1	88.72%
omega4	2	84.13%	u1	1	83.41%
pblh	1	83.15%	u2	1	89.02%
ps	1	94.77%	u4	1	98.63%
psl	1	99.95%	ustar	2	82.29%
qgscrn	2	86.17%	v1	2	87.25%
rh1	2	92.98%	v2	2	88.74%
rh2	1	91.70%	v4	1	95.64%
rh4	1	95.45%	zg1	1	97.55%
rnd	1	81.76%	zg2	1	87.89%
temp1	1	88.25%	zg4	2	98.57%

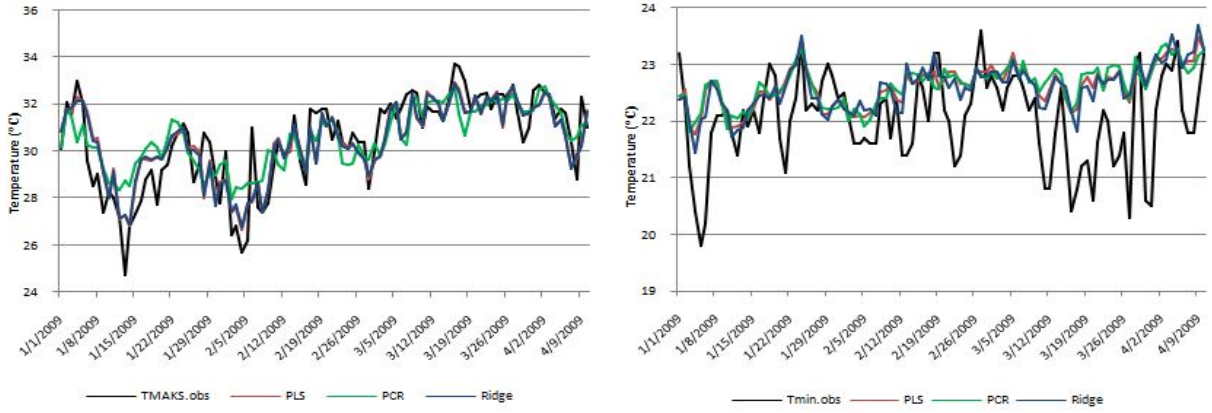

Figure 3. Trend of ensemble member forecasts and observations for (a) maximum temperature and (b) minimum temperature.

Table 2 shows the PCA pre-processing yields 41 PCs in total from 32 NWP parameters being involved. The explained cumulative variances range from 81.76% to almost 100%. Therefore, the association of weather quantities among 9 grids of each NWP parameter itself is extremely high. Those PCs thus would be involved as the predictor in three aforementioned statistical models, that is PLSR, PCR, and Ridge, which yield the forecasts shown only for the first 100 days on figure 3. Figure 3 implied that the forecast of each ensemble member initially could have been follow the general trend of maximum temperature and minimum temperature, they were going up if the observed

temperature acted the same and were doing so when it went down. But, the lingering problem was under-fitting or over-fitting which happened on the same day. Such problem even was more significantly seen on minimum temperature.

Based on figure 3, it could be said the forecasts produced by those 3 probabilistic models are quite far beyond the verifying temperatures even if they could capture the pattern of temperature. Therefore, it might be necessary to calibrate those models as the ensemble member to produce more accurate and precise weather --temperature-- forecast as well as being calibrated. The weighting adopted by BMA supposedly minimizes the impact of under-fitting or over-fitting, even of seasonal pattern.

C. BMA Calibrated Temperature Forecast

The calibration is performed in order that the variance would adapt to inevitably under-dispersive nature experienced by ensemble member before. After calibrating, one would obtain more reliable forecast with more proportional variance and narrower predictive interval. To assess whether the ensemble forecast is under dispersive or not, the Verification Rank Histogram (VRH) on figure 4 was necessarily carried out. The under dispersive ensemble is recognized from shaped-U histogram, while the over dispersive one is recognized from bell-shaped histogram [9]. As shown on figure 4, both ensemble forecast, either for maximum or minimum temperature, still possess under dispersive properties since each VRH resembles U-shape. They implied that many verifying temperatures are beyond ensemble range, the difference between maximum and minimum value of an ensemble forecast.

Based on figure 4 as well, the empirical coverages obtained were 20% and 6.7% for maximum and minimum temperature, respectively. It is identified from the percentage of observations lie within the second rank and the third rank. It could be said that each coverage is far less than standard coverage 50% and convinces us that the ensemble forecasts are uncalibrated due to under dispersive. It would influence the predictive interval to be unreliable, so that it needs to be calibrated by BMA.

The first step of BMA calibration is performing regression for each ensemble member (PLSR, PCR, and Ridge) with respect to the verifying observation. For instance, table 3 presents the regression coefficient and the weight of each member along with BMA mean as deterministic forecast on November 14th, 2009, particularly for Dermaga station based on 30-day training period.

It should be noted from table 3 that PLSR has the biggest contribution to BMA maximum temperature forecast as its weight is 0.724, greater than PCR's and Ridge's which have weight 0.276 and 0, respectively. However, the latter has the biggest contribution to BMA minimum temperature forecast rather than PLSR and Ridge. It might be said then the BMA's accuracy for both temperatures on that day were not significantly distinguished compared to the each member forecast.

The next optional analysis is graphing the BMA predictive density to find out how fit the calibration done by BMA for maximum temperature at the same station on the same day, as shown on figure 5. It could be seen that figure 5 shows the observation (vertical solid line) lies inside 95% BMA predictive interval (dashed line).

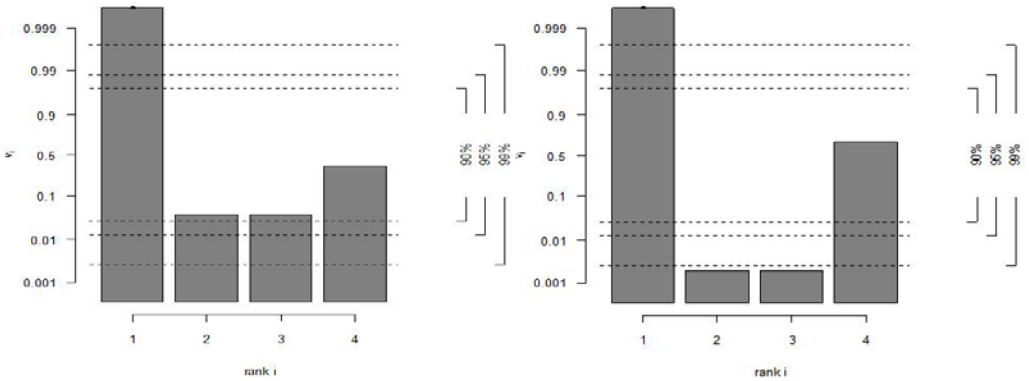

Figure 4. VRH for three ensemble members of (a) maximum temperature and (b) minimum temperature, January 31st 2009 – December 31st 2009.

TABLE 3.
BMA ESTIMATION AND FORECAST ON NOVEMBER 14TH, 2009.

Maximum Temperature						
Model	**β_0**	**β_1**	***w***	**Ensemble Forecast (°C)**	**Obs.(°C)**	**BMA(°C)**
PLS	1.12	0.95	0.72	32.81		
PCR	-3.31	1.09	0.28	32.24	32.6	32.9
Ridge	0.89	0.96	0.00	32.75		
Minimum Temperature						
Model	**β_0**	**β_1**	***w***	**Ensemble Forecast (°C)**	**Obs.(°C)**	**BMA(°C)**
PLS	12.35	0.47	0.02	22.93		
PCR	13.73	0.41	0.01	22.89	22.2	23.16
Ridge	10.92	0.53	0.97	22.83		

As indicated by figure 5, BMA yields more reliable interval, particularly for November 14th, 2009. The maximum temperature observation lies inside the ensemble range as well. It means that BMA manages to increase each member's precision, shown from PDF BMA which shifts closer to the middle. It likely behaves the same for the minimum temperature. One thing should be noted, though, is the variance of BMA was somehow always larger than each member's itself due to its role to broaden the variance.

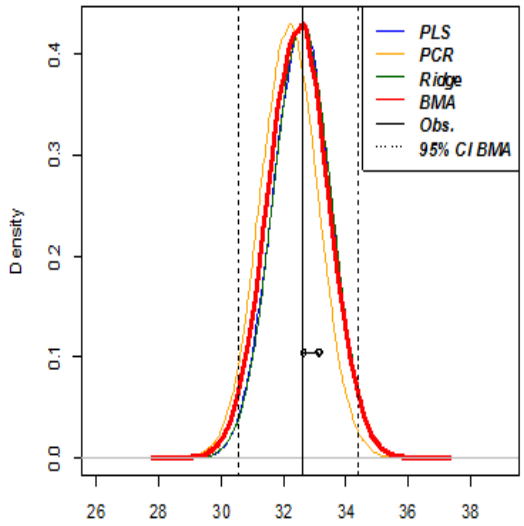

Figure 5. BMA predictive PDF and its members on November 14th, 2009.

Then, one should compare the forecast produced by BMA and deterministic NWP merely for Dermaga station with 30-day sliding training period, as shown on table 4. Based on Table 4, it implied that BMA manages to improve the forecast accuracy over 50% as it has less RMSE. In addition to the improved accuracy, BMA is able to yield the narrower predictive interval rather than raw ensemble since it has less CRPS.

TABLE 4.
ASSESSMENT FOR THE FORECASTS OVER 30-DAY TRAINING PERIOD.

	RMSE (°C)		CRPS		Coverage (%)	
	NWP	BMA	*Raw Ensemble*	BMA	*Raw Ensemble*	BMA
$T_{MAX.}$	2.18	0.950	0.653	0.517	20.65	49.41
$T_{MIN.}$	1.66	0.777	0.566	0.431	6.7	49.26

As the main goal, BMA manages to calibrate the ensemble forecast for temperature in particular. It could be seen from the coverage which advances significantly ever than before, 20.65% to 49.41% and 6.7% to 49.26%, respectively for maximum temperature and minimum temperature. It means that BMA is able to produce the weather, in this case temperature, forecast which is more consistent.

D. GOP Calibrated Temperature Forecast

The first step of GOP is to estimate β_0 and β_1 bias-corrected coefficient in order to obtain the residual for constructing empirical semivariogram. With 30-day training period, the GOP models for both temperature are given on Eq. (28).

$$\begin{aligned} \max.temp_{s,t} &= 1.785 + 0.959 tmaxscr_{s,t} \\ \min.temp_{s,t} &= 25.963 - 0.131 tminscr_{s,t} \end{aligned} \quad (28)$$

In general, for GOP model, the bias-corrected coefficient on Eq. (28) is retained. The next step then is to construct exponential semivariogram, particularly for the maximum temperature, shown on Figure 7.

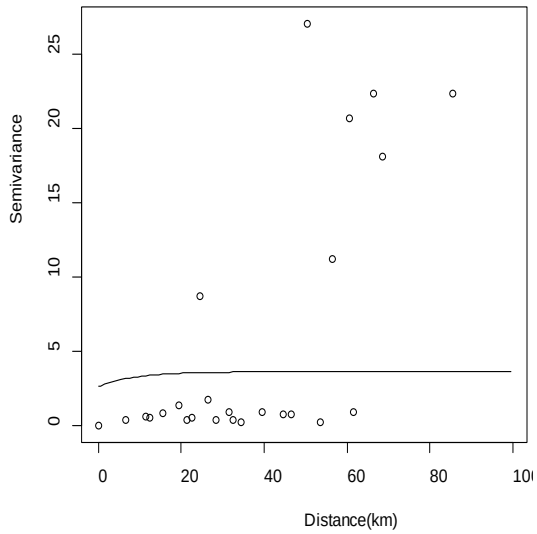

Figure 7. Empirical semivariogram of maximum temperature over 30-day training period.

The semivariance on figure 7 is roughly constant after reaching the distance about 8.69 kilometres. It implied that the maximum temperature between two stations is no longer dependent after 8.69 km, with sill recorded at 3.65. The high amount of sill might cause the greater variance of estimation or influence the forecast.

Figure 7 also indicates that spatial inconsistency is exist upon maximum temperature in particular. On some distance pairs, there are few bins which have semivariance far greater than others. Since these patterns are seen on bins representing distance 50 km or more, it might be triggered by Citeko station situated on Puncak plateau. It has been the biggest effect on the unexpected inconsistency. Regardless of above fact, Table 6 shows the forecast along with GOP predictive interval, represented by 5th percentile and 95th percentile on January, 31st 2009.

TABLE 6.
GOP FORECAST AND ITS CONFIDENCE LIMIT OF MAXIMUM TEMPERATURE ON JANUARY 31ST, 2009.

Station	Obs. (°C)	NWP(°C)	GOP(°C)	P_5 (°C)	P_{95} (°C)
Kemayoran	28.8	26.43	26.82	24.28	30.14
Priok	28.7	26.55	26.48	23.73	30.52
Cengkareng	28.6	26.42	28.75	24.41	29.75
Pd. Betung	29.0	26.49	27.80	24.44	30.25
Curug	28.3	26.26	26.64	23.63	29.66
Tangerang	29.2	26.34	26.22	23.93	30.79
Citeko	25.0	26.77	28.73	24.60	31.34
Dermaga	27.8	26.73	27.60	24.41	31.00

As shown on Table 6, GOP clearly manages to improve the bias correction rate, reflected on the RMSE of GOP (**2.12° C**) which is higher than of NWP (**2.18° C**), although the rate is somewhat less than 5%. It might have been the impact of spatial inconsistency mentioned before. It means GOP is severely vulnerable and risky in case of inadequate location of interest, inappropriate data properties, data mishandle, etc.

E. The assesment of the best calibration method

After carrying out both calibration method, BMA and GOP, one should verify and compare which method gives the better forecast based on few indicator presented on Table 7, using the same 30-day training period. Based on Table 7, the maximum temperature forecast is more accurate using BMA, while the minimum temperature using mean ensemble (the average of PLSR, PCR, and Ridge). Furthermore, BMA manages to better calibrate both temperature than GOP, indicated by less CRPS. Hence, BMA would yield the weather forecasts which are more accurate and reliable than GOP, particularly on those eight stations at Jabotabek.

TABLE 7.
THE COMPARATION OF FORECAST METHODS OVER 30-DAY TRAINING PERIOD.

Assessment Method	Type of Forecast	$T_{MAX.}$	$T_{MIN.}$
RMSE (°C)	NWP	2.745	2.01
	Mean ensemble	1.058	0.805
	BMA	1.053	0.819
	GOP	3.07	2.67
CRPS	BMA	0.576	0.451
	GOP	1.54	1.43

V. CONCLUSION

Some parameters of BMA and GOP, such as the weight and variance of BMA and the spatial parameters of GOP, should be estimated by iterative approach, for instance EM and L-BFGS respectively. For 30-day training period, the accuracy of BMA is not different than of the three members, while the former was more reliable, indicated by less CRPS. Furthermore, BMA manages to calibrate the forecast, indicated by the coverage closer to 50%. Lack of fit, though, is still owned by GOP since it has higher RMSE. From both method, BMA forecasts apparently have greater accuracy and precision.

ACKNOWLEDGEMENT

The entire data used in this work were supported by the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia. The authors also appreciate the Ministry of Research, Technology, and Higher Education of Indonesia on their invaluable support for this work as part of the strategic national research.

REFERENCES

[1] BMKG, *Kajian dan Aplikasi Model CCAM (Conformal Cubic Atmospheric Model) untuk Prakiraan Cuaca Jangka Pendek Menggunakan MOS (Model Output Statistics)*. Jakarta: Puslitbang BMKG, 2011.

[2] M. J. Schmeits, K. J. Kok, M. J. Schmeits, and K. J. Kok, "A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression Using ECMWF Ensemble Precipitation Reforecasts," *Mon. Weather Rev.*, vol. 138, no. 11, pp. 4199–4211, Nov. 2010.

[3] A. E. Raftery *et al.*, "Using Bayesian Model Averaging to Calibrate Forecast Ensembles," *Mon. Weather Rev.*, vol. 133, no. 5, pp. 1155–1174, May 2005.

[4] V. J. Berrocal, A. E. Raftery, T. Gneiting, V. J. Berrocal, A. E. Raftery, and T. Gneiting, "Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts," *Mon. Weather Rev.*, vol. 135, no. 4, pp. 1386–1402, Apr. 2007.

[5] Y. Gel, A. E. Raftery, and T. Gneiting, "Calibrated Probabilistic Mesoscale Weather Field Forecasting," *J. Am. Stat. Assoc.*, vol. 99, no. 467, pp. 575–583, 2004.

[6] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1, pp. 503–528, Aug. 1989.

[7] K. Feldmann, "Statistical Postprocessing of Ensemble Forecasts for Temperature: The Importance of Spatial Modeling," Ruprecht-Karls-Universitat Heidelberg, 2012.

[8] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.

[9] A. Möller, "Multivariate and Spatial Ensemble Postprocessing Methods," Ruperto-Carola University of Heidelberg, 2014.

[10] A. C. Rencher and G. B. Schaalje, *Linear Models in Statistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007.