Hierarchical Bayesian Spatio Temporal Model Comparison on the Earth Trapped Particle Forecast

Wayan Suparta1, Gusrizal1

1 Space Science Centre (ANGKASA), Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia

E-mail: wayan@ukm.my

Abstract. We compared two hierarchical Bayesian spatio temporal (HBST) results, Gaussian process (GP) and autoregressive (AR) models, on the Earth trapped particle forecast. Two models were employed on the South Atlantic Anomaly (SAA) region. Electron of >30 keV (mepoel1) from National Oceanic and Atmospheric Administration (NOAA) 15-18 satellites data was chosen as the particle modeled. We used two weeks data to perform the model fitting on a 5°x5° grid of longitude and latitude, and 31 August 2007 was set as the date of forecast. Three statistical validations were performed on the data, i.e. the root mean square error (RMSE), mean absolute percentage error (MAPE) and bias (BIAS). The statistical analysis showed that GP model performed better than AR with the average of RMSE = 0.38 and 0.63, MAPE = 11.98 and 17.30, and BIAS = 0.32 and 0.24, for GP and AR, respectively. Visual validation on both models with the NOAA map’s also confirmed the superior of the GP than the AR. The variance of log flux minimum = 0.09 and 1.09, log flux maximum = 1.15 and 1.35, and in successively represents GP and AR.

1. Introduction
Trapped particle has a major role for low Earth orbit (LEO) satellite anomaly. The risk is increased for the LEO Equatorial inclination satellites due to their frequent pass over the South Atlantic Anomaly (SAA) region. The SAA is an area with dense distribution of trapped particle, especially proton and low energy electron [1]. The high energy proton causes the degradation of satellite solar array whereas the low energy electron causes the satellite surface charging [2]. Overcoming with this problem, we develop a LEO trapped particle distribution forecasting system based on a hierarchical Bayesian spatio temporal (HBST) model. The HBST is a statistical model that works for spatio temporal data in a Bayesian framework [3].

In our previous work [4], we have employed a simple HBST model, i.e. Gaussian Process (GP) [5]. In this work, we attempt to employ HBST-Autoregressive (AR) model [6], a modified version of the GP. The difference between the two models is the AR considering the previous day parameter in their true value calculation (Equation 11). Detail of both models is explained in Section 2.

As in our previous work [4], we still use the data taken by National Oceanic and Atmospheric Administration (NOAA) satellite. However, in this work we attempt to employ multiple satellites data instead of only using NOAA 15, i.e. NOAA 15-18. We also improved our gridding method by shrinking the size of our grid from 10°x10° of longitude and latitude in our previous work to become...
5°x5° of longitude and latitude as in Figure 1a. The application of 5°x5° grid size makes the SAA is divided in 12 layers at the latitude direction and randomly select four points of validation in each layer (as in Figure 1b). The detail about the gridding process is explained in our previous paper [4].

Since we start to employ multiple satellites data, then we reduce the data input, from 30 days data in our previous work to 14 days data in this work. This step is taken to optimize the time consumed while running computational process. We choose to employ the implementation of both models on the 30 keV of electron energy level (mep0e1). We choose mep0e1 due to the sensitivity of the medium energy proton and electron detector (MEPED) sensor installed in the NOAA satellites. The area of South Atlantic Anomaly (SAA) is still used in this work. We set 31 August 2007 as the target date of the forecasting. Kriging interpolation method still used to draw the distribution of forecasted particle flux value over the SAA region. Detail about the Kriging interpolation of our work can be found in Suparta et al. [7].

In the following section, we describe two HBST models mentioned, i.e. GP and AR models. Section 3 explains the forecasting results of the two models, performed in both statistical and visual analyses and discuss the results. The summary and future work is compiled in Section 4.

2. Hierarchical Bayesian Spatio Temporal Model

First, we define the notations used in this paper [8]. Let \( t \) defines time where \( t=1, \ldots, T \), and \( T \) is the total number of time units. In this work, we put \( T = 14 \). This is done by selecting data from 17-30 August 2007 as the model input. For the forecasting day (\( T+1 \)), we put 31 August 2007. The \( Z(s_i, t) \) is a logarithmic value of SAA’s 30 keV electron flux. We also denote \( E(s_i, t) \) as a true value corresponding to \( Z(s_i, t) \) at site \( s_i, i = 1, \ldots, n \) at time \( t \). Let \( Z_t = (Z(s_1, t), \ldots, Z(s_n, t))' \) and \( E_t = (E(s_1, t), \ldots, E(s_n, t))' \). We should denote all observed data by \( z \), and \( z^* \) will yield all missing data. As for \( E \), \( E \) defines all \( E_t \) for \( t=1, \ldots, T \). The total number of observations modeled is defined by \( N \), which \( N=nT \).

Generally, in Bayesian forecast modeling, there are \( p \) covariates, denote by \( X_t \) which are in the form of \( 1 \times p \) vector. These covariates could be some variable that influence \( Z \) value. The regression coefficients are denoted by \( \beta \), also in the form \( 1 \times p \) vector. The NOAA data does not provide the covariates values for this work. Therefore, we only employ the intercept value rather than \( \beta \). Finally, the \( \theta \) symbol is used to denote all parameters used in our model \((\theta = (E, \beta, \sigma^2, \sigma^2, \theta))\).

2.1. Forecasting using Gaussian Process (GP) Models.

The spatio-temporal Bayesian GP model [5] is expressed as

\[
Z_t = E_t + \epsilon_t
\]
\[ E_t = X_t \beta + \eta_t \]  

(2)

where \( \epsilon_t = (\epsilon(s_1, t), \ldots, \epsilon(s_n, t))^T \sim N(0, \sigma^2_\epsilon I_n) \) is a white noise error process. \( \sigma^2_\epsilon \) is the nugget effect, and is homogeneous in space and time. The term \( I_n \) is the \( n \times n \) identity matrix. The term \( \eta_t = (\eta(s_1, t), \ldots, \eta(s_n, t))^T \sim N(0, \Sigma) \) is a spatially correlated error. \( \Sigma = \sigma^2_\eta \Sigma_n = \sigma^2_\eta \kappa(s_i, s_j; \phi, \nu) \) is a variance-covariance matrix with have dimension \( n \times n \), \( i, j = 1, \ldots, n \). \( \sigma^2_\eta \) is the site invariant common variance, while \( \kappa(.; \phi, \nu) \) is the spatial correlation matrix that have spatial decay \( \phi \) and smoothness parameter \( \nu \). Both parameters, \( \epsilon_t \) and \( \eta_t \), are independent each other.

Hence, the GP model for 1-step ahead forecast for the distribution \( Z \) at any unobserved location \( S_0 \) is:

\[
\begin{align*}
Z(s_0, T + 1) &= E(s_0, T + 1) + \epsilon(s_0, T + 1) \\
E(s_0, T + 1) &= x'(s_0, T + 1) \beta + \eta(s_0, T + 1)
\end{align*}
\]

(3)

(4)

the posterior predictive distribution of \( Z(s_0, T + 1) \) for a given \( z \) act as the 1-step ahead forecast distribution, i.e.

\[
\pi(Z(s_0, T + 1)|z) = \int \pi(Z(s_0, T + 1)|\theta, E, E(s_0, T + 1), z) \pi(\theta, E|z) dE(s_0, T + 1) dE d\theta
\]

(5)

\( \pi(\theta, E|z) \) is a joint posterior distribution of \( E \) and \( \theta \). Because of the conditional independence of \( Z(s_0, T + 1) \) for \( Z \) given \( E \), so \( \pi(Z(s_0, T + 1)|\theta, E, E(s_0, T + 1), z) = \pi(Z(s_0, T + 1)|\theta, E(s_0, T + 1), z) \). As \( E(s_0, T + 1) \) does not depend on \( Z \) given \( E \), then we can replace \( \pi(E(s_0, T + 1)|\theta, z) \) by \( \pi(E(s_0, T + 1)|\theta) \).

A 1-step ahead forecast distribution (5) can be constructed through a marginal distribution as well as conditional distribution. For marginal distribution, at the \( j \)th Monte Carlo Markov Chain (MCMC) iterations, we have posterior samples, \( \theta^{(j)} \) and \( E^{(j)}(s_0, T + 1) \) is drawn from \( N\left(x^{(j)}_{T+1} \beta^{(j)}, \sigma^2_\eta^{(j)}\right) \). \( Z^{(j)}(s_0, T + 1) \) is obtained from \( N\left(E^{(j)}(s_0, T + 1), \sigma^2_\eta^{(j)}\right) \). To use conditional distribution, consider that the joint distribution of \( E_{T+1} = (E(s_1, T + 1), \ldots, E(s_n, T + 1))' \) is given by \( N\left(X_{T+1} \beta, \Sigma_\eta\right) \) and the conditional distribution of \( \pi(E(s_0, T + 1)|E_{T+1}) \) is normally distributed with mean

\[
x(s_0, T + 1) \beta + S_{n,12} S^{-1}_\eta (E_{T+1} - X_{T+1} \beta)
\]

(6)

and variance

\[
\sigma^2_\eta (1 - S_{n,12} S^{-1}_\eta)
\]

(7)

where \( S_{n,21}' = S_{n,12} = e^{-\phi d_{12}} \) and \( d_{12} = (||s_1 - s_0||, \ldots, ||s_n - s_0||)' \).
Similar with any unobserved point, a 1-step ahead forecast at any observed point is modeled by:

\[ Z(s_i, T + 1) = E(s_i, T + 1) + \epsilon(s_i, T + 1) \]  
\[ E(s_i, T + 1) = x'(s_i, T + 1)\beta + \eta(s_i, T + 1) \]

the same step is applied for getting Z as in the unobserved point.

2.2. Forecasting using Auto-Regressive (AR) Models.

The AR model [6] is given by

\[ Z_t = E_t + \epsilon_t \]  
\[ E_t = \rho E_{t-1} + X_t \beta + \eta_t \]

where \( \epsilon_t \) and \( \eta_t \) form is defined as in the previous GP model. The \( \rho \) denotes the autoregressive process parameter, with \( |\rho| < 1 \). For \( \rho = 0 \), then our AR model turn to GP model as in (2). For predicting the distribution of \( Z(s_0, T + 1) \) at any unobserved point, it is determined by the value of \( O(s_0, T + 1) \). According to (11), \( E(s_0, T + 1) \) follows the normal distribution and has site variance \( \sigma^2_\eta \) and mean \( \rho E(s_0, T) + x'(s_0, T + 1) \). Because dependent to \( E(s_0, T) \), we have to specify all the random variables \( E(s_0, T) \), for \( k=0,\ldots,T \). For the first step of simulation, we first simulate the conditional distribution of \( E(s_0, 0) \) given \( E_0 \). Thus, at \( j \)th MCMC iterations, we consecutively simulate \( E^{(j)}(s_0, k) \) given \( E^{(j)}(s_0, k - 1) \) for \( k = 1,\ldots,T + 1 \) from a normal distribution with mean \( \rho^{(j)}E^{(j)}(s_0, k - 1) + x'(s_0, k)\beta^{(j)} \) and variance \( \sigma^2_\eta^{(j)} \).

Similarly, for forecasting at any observed point \( s_i \), we draw \( Z^{(j)}(s_i, T + 1) \) from the normal distribution with mean \( \rho^{(j)}E^{(j)}(s_i, T) + x'(s, T + 1)\beta^{(j)} \) and variance \( \sigma^2_\epsilon^{(j)} \).

3. Results and Discussion

We use R \textit{spTimer} package [9] to compute MCMC Gibbs sampling procedure as defined in previous section. We set 13,000 for \( j \) value and discharge the first 3,000 iterations to avoid the starting effect. The forecasting accuracy of both models was then be evaluated by three statistical validations, i.e. the root mean square error (RMSE), mean absolute percentage error (MAPE) and bias (BIAS) on the validation points (Figure 1b).

The results of statistical validation of GP and AR models are displayed in Figure 2. Figures 2a and 2c, mentioned for RMSE and BIAS with units in log flux (particles/cm\(^2\) sec str), whereas for MAPE in Figure 2b, the unit is in percentage. The accuracy of forecasting result is pointed by MAPE, which defines how close the forecasting result to the real data. Our models, GP and AR, give 11.98 and 17.30 for MAPE values which indicates their high degree of forecasting accuracy. On the other hand, RMSE result indicates variance between forecasting value and real data. We find that the average values of RMSE for GP and AR is 0.38 and 0.63, respectively. We could also evaluate the nature of flux distribution variability from the RMSE. The higher of RMSE value in a layer means the higher of variability among the flux in that layer. In this work, we could find that the L3 and L4 layers have the highest values of RMSE. It means that the high variability of flux particle distribution occurred at the SAA core. The trend of flux particle distribution variability shows that the distribution of flux is quite uniform outside the SAA area (L9 to L10). We also find that at SAA boundaries area such as L1, L2,
L5-L8, the variability of flux starts to rise. In addition, we can accept the variability at L11 and L12 since there are boundary layers between the two dense areas, the SAA and the polar. From BIAS result, we find that our model mostly produced overestimate values in each layer except for L3, L7 and L8, where the AR gives the underestimate forecasts. The average values of BIAS for GP and AR are 0.32 and 0.24, respectively.

We perform a visual validation for the results as depicted in Figure 3. Figure 3a defines the plot of raw data acquired by NOAA 15-18 on 31 August 2007, whereas Figure 3b defines the NOAA map of mep0e1 on 31 August 2007. Figures 3c and 3d define the estimation of our forecast values after being interpolated by a Kriging technique, for GP and AR, respectively. The Kriging interpolation was computed by using R fields package [10]. The quality of Kriging interpolation is defined by its variance values and represented by Figures 3e and 3f, which successively represent GP and AR. The difference between our results in Figures 3c-3d to the NOAA map in Figure 3b is the period of time. NOAA’s map in Figure 3b is generated for 4 days distribution (28-31 August 2013) on a nowcast basis. On the other hand, our results in Figure 3c-3d showed forecast maps and designed for one day distribution, 31 August 2007.

As illustrated in Figures 3c-3d, the AR produced a significant overestimate number of a maximum and underestimate number of a minimum. The log maximum number of real data is 7.65 and 1.09 for a minimum. Thus, the GP gives 1.09 for a maximum variance and 0.09 for a minimum, whereas the AR gives 1.35 and 1.15 for a maximum and a minimum, respectively. From the comparison between Figures 3e and 3f, we could see that the GP result Kriging interpolation is better than the AR since it produced less variance than AR. From this result, we can conclude that the autoregressive parameter (in Equation 11) is not suitable for our work.
Figure 2. Statistical validation for (a) RMSE, (b) MAPE (in percentage), and (c) BIAS. Both units for RMSE and BIAS are in yy.
4. Summary and Future work
This work has been successfully compared two HBST models to forecast trapped particle flux distribution on 31 August 2007 from 14-day input of NOAA’s data. The GP obtained 88.02 % for forecasting accuracy weather the AR obtained 82.7 %. We assume that the less accurate of AR due to the AR temporal correlation parameter used in this work, that needs to be modified regarding to our data. Compared to our previous work, the employment of multiple satellites data in this work omits the missing values occurrence at validation points. In the future work, we propose to shrink the grid into $2^\circ \times 2^\circ$ size of longitude and latitude to reduce the average bias effect in the gridding process.

Figure 3. (a) Raw data plot of NOAA 15-18 on 31 August 2007 for mep0e1, (b) NOAA’s flux distribution map of mep0e1 on 28-31 August 2007, (c) Kriging estimation on GP forecast result, (d) Kriging Estimation on AR forecast result, (e) Kriging variance on GP forecast result and (f) Kriging variance on AR forecast result.
Acknowledgments

This work is supported by the Ministry of Science, Technology and Innovation (MOSTI) Malaysia under Science Fund 06-01-02-SF0808 grant. The authors also thank to NOAA for data used. Special thanks to Mr. Muhammad Marizal for his advice on R programming.

References

[1] Benton E R and Benton E V 2001 Nucl. Instrum. Methods Phys. Res., Sect. B 184 255
[2] Gubby R and Evans J 2002 J Atmos Sol-Terr Phy 64 1723
[3] Cressie N A and Wikle C K 2011 Statistics for Spatio-Temporal Data (New York: John Wiley & Sons)
[4] Suparta W, Gusrizal 2013 paper presented at the 2014 International Conference on Science & Engineering in Mathematics, Chemistry and Physics (ScieTech 2014) Jakarta 13-14 January 2014
[5] Banerjee S, Carlin B P and Gelfand A E 2004 Hierarchical Modeling and Analysis for Spatial Data (Boca Raton : Chapman & Hall/CRC)
[6] Sahu S K, Gelfand A E, and Holland D M 2007 J. Am. Statist. Assoc. 102 1221
[7] Suparta W, Gusrizal, Mohd Alauddin M A and Ahmad N 2013 Proceeding of the 2013 IEEE International Conference on Space Science and Communication (ICON SPACE) Malacca 1-3 July 2013
[8] Sahu S K, Bakar K S, and Awang N 2013 Technical Report (Southampton: School of Mathematics University of Southampton)
[9] Bakar K S and Sahu S K 2013 Technical Report (Southampton: School of Mathematics University of Southampton)
[10] Furrer R, Nychka D, and Sain S 2013 fields: Tools for Spatial Data (http://CRAN.R-project.org/package=fields)