Forecasting the Earth's Trapped Particle Distribution Using Hierarchical Bayesian Spatio Temporal Model

Wayan Suparta[1,*], **Gusrizal**[1]

[1]Institute of Space Sciences (ANGKASA), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia

E-mail: wayan@ukm.my

Abstract. We employed the Hierarchical Bayesian spatio temporal (HBST) Gaussian Process (GP) model for forecasting the distribution of the Earth's trapped particle. The model was applied in the South Atlantic Anomaly (SAA) region. Data from 1-30 January 2000 of >30 keV electron flux acquired by National Oceanic and Atmospheric (NOAA) 15 satellite was carried to model. The purpose was to forecast the flux value on 31 January 2000. Gridding process of 10x10 lot-lan was performed after cleaning and log transforming data. The HBST GP model was undertaken by implementing the Monte Carlo Markov Chain (MCMC) method. The forecasting result was interpolated by using Kriging technique to draw the distribution map of particle flux. Statistical validation represented by mean square error, root mean square error, mean absolute error, mean absolute percentage error, bias, relative bias, and mean relative separation shows good indicators. The visual validation also figured a quite similarity with NOAA's map that the model capable to forecast the particle flux.

1. Introduction

Trapped particle is one of major radiation in the space environment. It is produced by solar flares and CMEs and flows out to the Earth's atmosphere by the solar wind [1]. Due to geomagnetic field line the charged particle is trapped into two areas of Van Allen radiation belt: the inner radiation belt and the outer radiation belt. Trapped particle could cause harmful effects to the spacecraft that pass through their region especially for low earth orbit (LEO) satellite [2]. Therefore, it is important to model the distribution of Earth's trapped particle. There are several trapped particle models that have been developed, mostly are based on physics - magneto hydro dynamics (MHD) frame, such as SWMF/BATS-R-US with RCM [3], Fok Ring Current [4], Plasma sphere [5], CIMI [6]. The most used and the stable one, the AE-9/AP-9 [7] is run in a statistical modeling framework.

In this work, we choose to employ a statistical model, named hierarchical Bayesian spatio temporal model (HBST). The unique of this model because it works on a geographical coordinate rather than the (L, B) coordinate like the others. Although the approach of representing the trapped particle in magnetic coordinate is preferred in scientific application, the necessity of simple displayed model for the end user like satellite operator and designer is taking into account [8]. In addition, the model can be employed for both proton and electron in every condition of solar activity. The model can also perform a long time forecast like daily or weekly of solar trapped particle flux value. Our objective of this work is attempted to build the trapped particle distribution forecasting system for satellite operator and designer.

In this paper, the model developed is applying for the South Atlantic Anomaly (SAA) region, before applied in the whole area of Earth atmosphere. SAA is an atmosphere region centered over Brazil sky which lies from -90° to 40° of longitude and from 0° to -50° of latitude, at the altitude of ~500 km. The SAA is a main source of trapped particle in the inner radiation belt. We use the data acquired by National Oceanic and Atmospheric Administration (NOAA) 15 satellite, a polar orbit satellite with inclination 98.5° and altitude ~807 km. Based on this condition, our distribution forecasted is valid in the altitude of 300-800 km and in the shell of $L \leq 1$. We also perform the forecast on a daily basis.

2. Methodology

2.1. NOAA Data

The NOOA 15 data (http://satdat.ngdc.noaa.gov/sem/poes/data/avg/txt/) consists of 41 columns that contain time, locations, and flux values of observed particles, with the energy range from 30 keV to 200 MeV. For simplicity purpose, we chose the lowest energy level, >30 keV electron flux. The details explanation for the data can be access through NASA - National Geophysical Data Center (NGDC) (http://satdat.ngdc.noaa.gov/sem_poes/docs/readme_16s_ascii.txt). In this work, we use of electron data, based on the location of the SAA. Once data sorted, then the next phase are data cleaning, performing logarithmic transformation to stabilize the data distribution, and finally gridding them.

The important step in the data process is gridding the data. NOAA 15 completes its orbit in about 1 hour and 40 minutes. Due to that fact, a point of observation could not be observed in a long period of time. We have calculated that a point in a grid of 1x1 longitude and latitude will be observed again in about two months, and this make impossible to do a forecast due to the lack of data. Therefore, gridding data is proposed to solve this problem. We use the 10x10 gridding system (see Figure 1).

Each data in a grid was averaged with its counterparts and put the average value in the center of the grid. The new value and the new location then become a representative of a grid. We employed this procedure to 65 grids of 10x10 sizes. For validation purpose, we chose 15 points randomly as the validation points, and the rest as the model input as shown in Figure 1.
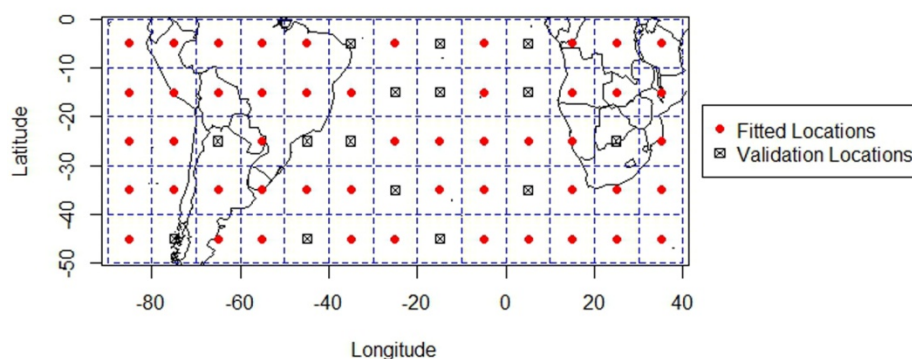
Figure 1. The validation and fit points in SAA with 10x10 grids

The validation process was conducted by comparing the values in validation points obtained by the forecasting process with the original values obtained by observation. Data from 1-30 January 2000 has chosen as the mode fitting. Our target is to forecast the flux value on 31 January 2000 with the model obtained.

2.2. Hierarchical Bayesian Spatio Temporal Model

HBST model is a statistical modeling technique that deals with space time modeling in Bayesian approach. Considering Gelfand [9], the HBST has a structure:

First stage	: Data model $[Z\|E, \theta]$
Second stage	: Process model $[E\|\theta]$
Third stage	: Parameter model $[\theta]$

where the data defined by *Z,* the (hidden) process is defined by E , and the unknown parameters are specified by θ.

First of all, we define the generic notations used throughout this paper, and the formula related to them [10]. Let $Z\,(s_i,\,t)$ is denoted the logarithmic value of SAA's > 30 keV electron flux at site s_i and time *t*, i = 1,…, n and t=1,…, T. $E(s_i,\,t)$ is as a true value corresponding to $Z(s_i,\,t)$. We put T = 30 at this work to accommodate selected data from 1-30 January 2000 as the model input. We put 31 January 2000 as our forecasting target, notified as T+1. We choose to define both *Z* and *E* in vector notation as $Z_t = (Z\,(s_i,\,t),…,\,Z\,(s_n,\,t))'$ and $E_t = (E\,(s_i,\,t),…,\,E\,(s_i,\,t))'$. Therefore the HBST GP model [11] is expressed by:

$$Z_t = E_t + \epsilon_t \tag{1}$$

$$E_t = X_t\beta + \eta_t \tag{2}$$

with $\epsilon_t = \left(\epsilon(s_1,t), …, \epsilon(s_n,t)\right)' \sim N(0, \sigma_\epsilon^2 I_n)$ is an error process. Included in this term, the σ_ϵ^2 is a nugget effect, and is homogeneous in space and time, whereas the I_n is the *n* x *n* identity matrix. The covariates influencing the Z value is denoted by X_t with *n x p* matrix size, and $\beta = (\beta_1,…, \beta_p)'$ is the *p x 1* vector of X_t, respectively. NOAA 15 data does not provide any covariate data for the trapped flux, so the intercept value is employed in Equation 2. Furthermore, the $\eta_t = \left(n(s_1,t), …, n(s_n,t)\right)' \sim N(0, \Sigma_n)$, is a spatially correlated error. Included in, the $\Sigma_n = \sigma_\eta^2 S_n = \sigma_\eta^2 \kappa(s_i, s_j; \phi, v)$ is a variance-covariance matrix, and have dimension *n* x *n*, *i, j* = 1,…, *n*. Afterward, the σ_η^2 is the site invariant common variance. The $\kappa(.; \phi, v)$ expresses the spatial correlation matrix with spatial decay ϕ, and smoothness parameter v. The error parameters ϵ_t, and η_t, are independent each other. For future reference, **z** is denoted all observed data, **x** is denoted all covariates data, and **E** represent all augmented values. Lastly, we use θ symbol to assert all parameters used $(\theta = (E, \beta, \sigma_\epsilon^2, \sigma_\eta^2, \emptyset))$.

For forecasting purpose, the HBST GP model at any observed point s_i on T+1 day is expressed in terms below:

$$Z(s_i, T + 1) = E(s_i, T + 1) + \epsilon(s_i, T + 1) \tag{3}$$

$$E(s_i, T + 1) = x'(s_i, T + 1)\beta + \eta(s_i, T + 1) \tag{4}$$

with posterior predictive distribution of $Z(s_i, T + 1)$ given **z** is denoted by:

$$\pi(Z(s_i, T + 1)|\mathbf{z}) = \int \pi(Z(s_i, T + 1)|\boldsymbol{\theta}, \mathbf{E}, E(s_i, T + 1), \mathbf{z})\,\pi(E(s_i, T + 1)|\boldsymbol{\theta}, \mathbf{z})$$

$$\pi(\boldsymbol{\theta}, \mathbf{E}|\mathbf{z})dE(s_i, T + 1)d\mathbf{E}d\boldsymbol{\theta} \tag{5}$$

Finally, to summarize this method, we perform the following algorithm to predict the $Z(s_i, T + 1)$, by using Monte Carlo Markov Chain (MCMC)-Gibbs sampling method with *j* iterations:

1. Draw a sample $\boldsymbol{\theta}^{(j)}$, and $\mathbf{E}^{(j)}, j \geq 1$ from Equation 5.
2. Draw $E^{(j)}(s_i, T + 1)$ from $N\left(x'_{T+1}\beta^{(j)}, \sigma_\eta^{2(j)}\right)$.
3. Finally draw $Z^{(j)}(s_i, T + 1)$ from $N\left(E^{(j)}(s_i, T + 1), \sigma_\epsilon^{2(j)}\right)$.

Once the value of Z at T+1 day obtained, then the result is displaying in a geographic map by using Kriging interpolation technique. Detail of the Kriging interpolation technique applied on the distribution of trapped particle over SAA region can be found at Suparta et al. [12]. Figure 2 is the flow chart of this work methodology.
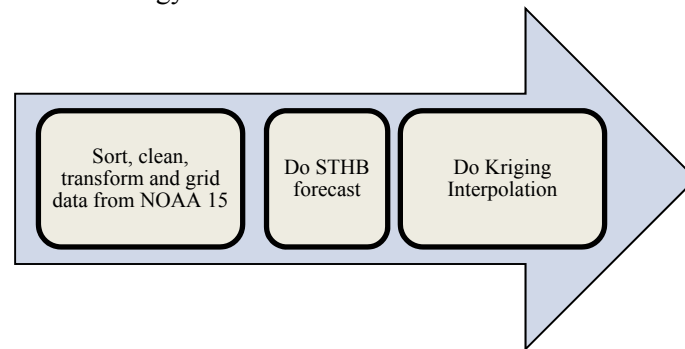
Figure 2. The flow chart of trapped particle flux distribution forecasting

3. Result and discussion

3.1. Statistical Validation

Statistical analysis was performed through seven validation parameters: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), bias (BIAS), relative bias (rBIAS) and the mean relative separation (rMSEP). Details formulas of the validation parameters can be found at Bakar and Sahu [13]. Table 1 shows the results of validation process.

**Table 1**. Validation result of GP model

MSE	RMSE	MAE	MAPE	BIAS	rBIAS	rMSEP
0.5577	0.7468	0.6691	19.3589	0.1796	0.0457	0.4969

From Table 1, we can conclude that the forecast values and the observed data have a high degree of similarity since the validation parameters produced the small numbers. This allows us to draw the conclusion that the HBST GP model can work well and give an accurate result in the term of forecasting.

3.2. Visual Analysis

The visual analysis was performed by comparing the Kriging results, both estimates and variance with several patterns of map distribution, i.e. daily data plot, NOAA's distribution map, Kriging interpolation on the raw data and Kriging interpolation on the grid data. Figures 3a and 3b present the electron >30 keV daily plot on 31 January 2000 and its distribution map drawn by NOAA, respectively. Figures 4, 5, and 6 present the distribution maps of electron estimation and variance obtained by ordinary Kriging method on STHB-GP model forecast result, empirical, and grid data, respectively. We could understand from Figure 3a that the data collected is so lack, and is a challenging work to deal with it. As for Figure 3b, nine days data are collected by NOAA to get the distribution map. On the other hand, we could produce a 1-day forecast value of SAA electron flux and display in a distribution map as in Figure 4a, while Figure 4b refers to the variance of Kriging interpolation on the forecasting result. When we compare the estimation of Kriging from Figures 4a, 5a, and 6a, we could see that the pattern of forecasting estimate (Figure 4a) is slightly different from the others and the pattern is more widened rather than concentrated. It is also reflected a decreasing value of flux maximum log in a significant rate, that is, from > 6 in the observed data to <5 in the

forecast estimate. This visual analysis showed an opposite result with the statistical validation which showed very good results. Gridding process is probably as major factor to this problem as a consequence of averaging and centering.
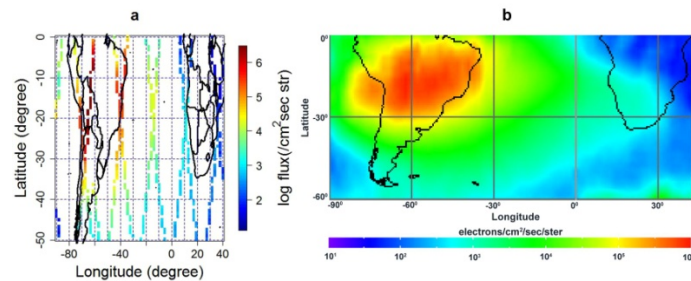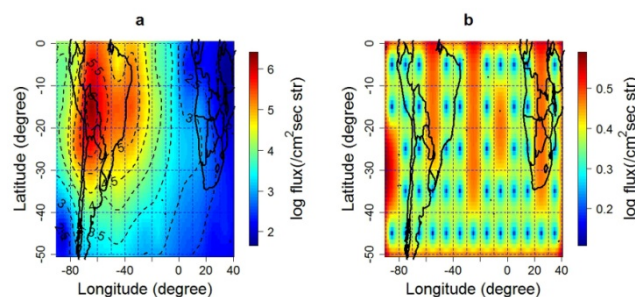

Figure 3. Mep0e1 data on 31 January 2000 for (a) raw data and (b) NOAA's SAA on 23-31 January 2000
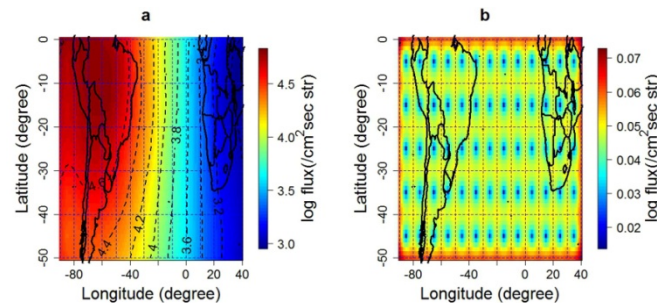
Figure 4. Flux Distribution map of >30 keV electron over SAA region on 31 January 2000, for (a) forecast estimation and (b) forecast variance
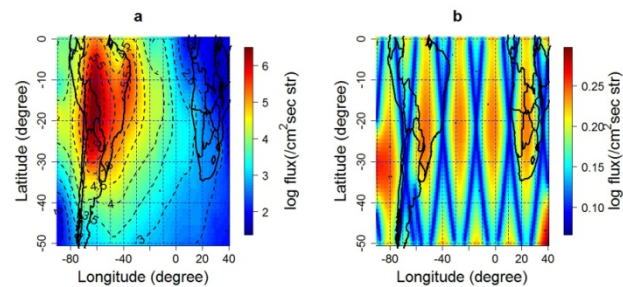
Figure 5. Flux Distribution map of >30 keV electron over SAA region on 31 January 2000, for (a) raw data estimation and (b) raw data variance

Figure 6. Flux Distribution map of >30 keV electron over SAA region on 31 January 2000 for (a) grid data estimation and (b) grid data variance

4. Summary and Future Works

HBST-GP model was successful to perform the forecasting of Earth trapped particle over the SAA region. We also achieved to deliver the convenient way to figure the >30 keV electron flux distribution map by implementing the Kriging interpolation method on a geographical map. Statistical analysis showed good results whereas the visual gave a slightly different pattern. It was also seen a significant decreasing of the flux maximum log value. To accomplish these problems, two ways are suggested to apply at the future work, i.e. shrinking the grid size and using more data from NOAA satellite series.

Acknowledgments

References

[1]   Vainio R , Desorgher L, Heynderickx D, Storini M, Fluckiger E, Horne R B , Kovaltsov G A , Kudela K, Laurenza M, Lawlor S M, Rothkaehl H, and Usoskin I G 2009 Dynamics of the earth's particle radiation environment *Space Sci. Rev.* **147** 187

[2]   Horne R B , Glaurent S A, Meredith N P, Boscher D, Maget V, Heynderickx D, and Pitchford D 2013 Space weather impact on satellites and forecasting the earth's electron radiation belts with SPACECAST *Space Weather* **11** 169

[3]   Tóth G et al. 2005 Space Weather Modeling Framework: A new tool for the space science community *J. Geohys. Res.* **110** A12226

[4]   Fok M C, Wolf R A, Spiro R W, and Moore T E 2001 Comprehensive computational model of the earth's ring current *J. Geohys. Res.* **106** 8417

[5]   Pierrard V, Goldstein J, Andre N, Jordanova V K, Kotova G A, Lemaire J F, Liemohn M W, Matsui H 2009 Recent progress in physics-based models of the plasmasphere *Space Sci. Rev.* **145** 193

[6]   Fok M C, Glocer A, Zheng Q, Horne R B, Meredith N P, Albert J M, and Nagai T 2011 Recent developments in the radiation belt environment model *J. Atmos.Sol-Terr Phys.* **73** 1435

[7]   Ginet G P et al. 2013 AE9, AP9 and SPM: New model for specifying the trapped energetic particle and space plasma environment *Space Sci. Rev.* **179** 579

[8]   Hastings D and Garreth H 2004 *Spacecraft-Environment Interactions* 2004 (Cambridge: Cambridge University Press) chapter 3 p 66

[9]   Gelfand A E 2012 Hierarchical modeling for spatial data Problems *Spatial Statistics* **1** 30

[10]   Sahu S K, Bakar K S, and Awang N 2013 Bayesian forecasting using hierarchical spatio-temporal models with applications to ozone levels in the eastern united states *Technical Report* School of Mathematics University of Southampton Southampton

[11]   Banerjee S, Carlin B P, and Gelfand A E 2004 *Hierarchical modeling and analysis for spatial data* (Boca Raton: Chapman & Hall/CRC)

[12]   Suparta W, Gusrizal, Mohd. Alauddin M A, Ahmad N 2013 Distribution of energetic particle fluxes in the vicinity of saa according to kriging interpolation *Proc. 2013 IEEE Intern. Conf. Space Sci. Comm. (ICONSPACE)*, p 17

[13]   Bakar K S and Sahu S K 2013 **spTimer**: Spatio-temporal bayesian modeling using R *Technical Report* School of Mathematics University of Southampton Southampton