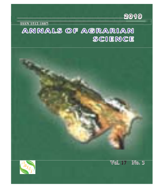
Annals of Agrarian Science

Journal homepage: http://journals.org.ge/index.php



# Background error covariance in numeral weather production

**Nato Kutaladze[a*], George Mikuchadze[a], David Loladze[a], Giorgi Sokhadze[b], David Mikava[b]**

[a]The Ministry of Environment Protection and Agriculture of Georgia, National Environmental Agency, 150, David Agmashenebeli, Ave.,Tbilisi, 0112, Georgia

[b]Geosciences and Technology Development Institute, 15a, Tamarashvili Str., Tbilisi, 0162, Georgia

Received: 29  March 2019; accepted: 15 May 2019

A B S T R A C T

Improving of weather forecast quality is a continuous work, as it is an invaluable for society and environment. WRF model have been tuned and tested over Georgia's territory for years. Nowadays as local meteorological network became denser and many remote observational sources are available data assimilation with variational methods is current challenge. First time in Georgia the process of data assimilation in Numerical weather prediction is developing, the way for forecast initial conditions' correction. Assessment of the forecast error is one of the first and most important steps in data assimilation. This work presents how forecast error statistics appear in the data assimilation problem through the background error covariance matrix – B, where the variances and correlations associated with model forecasts are estimated. Statistics of B are usually determined for a limited set of variables, called control variables that minimize the error covariance between variables. Results of generation and tuning of background error covariance matrix for five control variables using WRF model over Georgia with desired domain configuration are discussed and presented. The mathematical and physical properties of the covariances are also reviewed.

***Keywords:*** Weather forecast, Numerical Weather Prediction, Variational assimilation, Background errors statistics, Analysis increment, Distribution function.

*Corresponding author: Nato Kutaladze; e-mail address: cwlam08@gmail.com

## Introduction

Weather forecast accuracy very demanding on initial conditions, as small changes in initial conditions can lead to large changes in prediction. Variational data assimilation (VAR) is the method to estimate the initial state of the atmosphere for weather prediction to improve forecast quality [1, 2]. VAR usually combining measurements and models takes a forecast (also known as the first guess, or background information) and applies a correction to the forecast based on a set of observed data and estimated errors that are present in both the observations and the forecast itself. The forecast error is represented as background error covariance matrix (B). The specification of background error statistics is a key component of data assimilation since it affects the impact observations will have on the analyses. In the variational data assimilation approach, applied in geophysical sciences, the dimensions of the background error covariance matrix (B) are usually too large to be explicitly determined and B needs to be modeled. Statistics of the background error covariance matrix B are usually determined for a limited set of variables, called control variables that minimize the error covariance between variables. Then, several parameters need to be diagnosed to drive the series of operators that model B. There are now many leading centres around the world- European Centre for Medium-Range Weather Forecast (ECMWF), the National Centers for Environmental Prediction (NCEP), or the UK Met office etc. that use VAR for weather forecasting, and there are often differences in the way that forecast error statistics are measured, described and used by each [3-5]. In this paper, we present background covariance matrix' properties generated for WRF-ARW model with GEN_BE code over South Caucasus domain and testing results within the two assimilation systems GSI and WRFDA. Originally, the GEN_BE code was developed [6] as a component of a three dimensional variational data assimilation (3DVAR) method to estimate the background error of MM5 for a limited-area system. Since this initial version, various branches of code have been developed at NCAR and at the UK Met Office to address specific needs using different models such as (WRF) and the Unified Model (UM) on different data assimilation platforms such as the Weather Research Forecast Data Assimilation system (WRFDA) and the Grid point Statistical Interpolation system [7, 8]. The first section of this document describes the role of the background error covariance matrix B, difficulties and opportunities of it estimation. The second one presents general structure of GEN_BE code version 2.0. with some technical details in our application and provides results of pseudo observation case in two different systems of data assimilation (WRF-DA and GSI) using different B matrix involving the same set of five control variables (CV5). All the results presented in these papers have been obtained from WRF model output for 9.2 km resolution domain configured and tuned over South Caucasus domain (Fig.1).

## Methods and materials

2.1 Background error covariance matrix and initial state of atmosphere

The objective of VAR is a cost function $J(\delta x, x^g)$ minimization. This objective function is a combination of forecast and observation deviations from the desired analysis, weighted by forecast and observation-error covariance matrices.

$J(\delta x,x^g)=1/2(\delta x^b-\delta x)^T B(\delta x^b-\delta x)+1/2[y_o-H(x^g+\delta x)]^T R^{-1}[y_o-H(x^g+\delta x)]$ (Eq.A.1)

Where $x$ is the state vector composed of the model variables (e.g. winds, pressure, temperature, humidity, etc.) to analyses, at every grid pointDist of the 3-dimensional (3-D) model computational grid [6].

$\delta x$ is difference between the analysis $x^a$ and reference state or the 'first guess' $x^g$, i.e.

$x^a= x^g+ \delta x$ (Eq.A.2)

$y_o$ is the vector of observations and $H$ called the observation operator, is a mapper from the gridded model variables to the irregularly distributed observation locations. R is the observational error covariance matrix. B is the background error covariance matrix. The background error covariance matrix describes the probability distribution function (PDF) of forecast errors. Theoretically exact knowledge of R and B would require the knowledge of the true state of the atmosphere at all times and everywhere on the model computational grid, what is not possible. Therefore, both matrices have to be estimated in practice. Dimension of the B matrix is the square of the 3-D model grid multiplied by the number of analyzed variables. For typical geophysical applications as in meteorology, the size of the B matrix, comprised of nearly $10^7 \times 10^7$ entries, is too large to be calculate explicitly nor be stored in present computer memories. As a result, the B matrix needs to be parameterized [9, 10].

2.2 Background errors covariance matrix modeling.

The cost function as defined in Eq. (A.1) is usually minimized after applying the change of a variable:

δx= B1/2u (Eq.A.3)

$B^{1/2}$ is the square root of the background error covariance matrix. The variable $u$ is called the control variable and the cost function becomes:

J(u)=1/2uTu+1/2(d−HB1/2u)TR−1(d−HB1/2u) (Eq.A.4)

Where $d$ is the innovation vector defined as $d = (yo-H(xb))$ and it represents the difference between observations and their modeled values using a non-linear observation operator.

The square root of the B matrix as defined in Eq. (A3) is decomposed to a series of sub-matrices, each corresponding to an elemental transform that can be individually modeled:

U= SUpUvUh (Eq.A.5)

Where, S diagonal matrix and composed of the standard deviations of the background errors.

$U_p$ matrix - Physical Transform - defines the cross-correlations between different analysis Variables via statistical balance (linear).

$U_h$ - Horizontal Transform - defines the horizontal auto-correlations for the control variables. It is modeled through successive applications of recursive filters [11], The matrix $U_v$ defines the vertical auto-correlations for each of the control variables [12].

If the EOF (Empirical Orthogonal Function) decomposition is used, the eigenvectors model the vertical transform ($U_v$) and the associated eigenvalues represent the variance. The length scale is estimated in the EOF space and represents the horizontal transform ($U_h$). In the data assimilation process, the eigenvalues weight the analysis increment and the recursive filter first spreads out the information in the EOF space according to length scale value. Then, the transformation from EOF mode to physical space spreads out the information vertically.

## Calculations and results

For this study WRF-ARW model over the 9.2 km domain (Fig. B.1) with 151 x 100 x 36 grid cells have been used.

Background error covariance matrix **B** was generated using GEN_BE code version 2.0 in WRFDA. The code comprises from 5 stages, having separate input output infrastructure and managed via name list file, where control variables and all parameters to model B are defined by user.

Since the background error covariance matrix is a statistical entity, samples of model forecasts are required to estimate the associated variances and correlations of desired variables. The input data for gen_be are WRF forecasts, which are used to generate model perturbations, used as a proxy for estimates of forecast error.

NMC (named for the National Meteorological Center) method [13] was used to represent a sample of model background errors, where differences between two forecasts valid at the same time but initiated at different dates (time lagged forecast, e.g. 24-minus 12 h forecasts) was taken. This is done for many different dates to build up a large sample size for calculating statistics. Climatological estimates of background error may then be obtained by averaging these forecast differences over a period of time (e.g. one month).

For this run, February 2018 12 and 24-hour WRF-ARW forecasts, initialized both at 00 and at 12 UTC was used. Thus in all 60 pairs of perturbations are utilized to generate WRF-ARW Background Error.

On the initial stage analyses control variables stream function ($\psi$) and unbalanced velocity potential ($\chi_u$) are calculated from u and v wind, then differences for following 5 control variables: stream function ($\psi$), unbalanced velocity potential ($\chi_u$), Temperature (T), Relative Humidity (q), Surface Pressure (ps) have been crated. On the next stage statistics are calculated, such as mean from differences, created on the initial stage, then performs perturbation for each control variable and computes covariance of the respective fields [14].

On the stage 3 regression coefficient & balanced part of $\chi$, T and $p_s$ variables computed. The estimation error for one analysis variable may affect the value of another if they are correlated. The simplest way to model them is to use linear regression. Firstly, the regression coefficients between variables calculated, then, linear regressions are performed to derive uncorrelated control variables and then remove the balanced part for each other variable. This part achieves the **U**p transform: it models correlations between variables and allows transforming the matrix as a diagonal bloc in the control (uncorrelated) space. Computes unbalanced parts for the same variables:

$\chi_u' = \chi' - \chi_b$; $T_u' = T' - T_b$; $p_{s\_u}' = p_s' - p_{s\_b}$ is the preliminary step before estimating the vertical and horizontal auto-correlation parameters for each control variable.

Stage 4 Removes mean for $\chi_u'$, $T_u'$ & $p_{s\_u}'$ and computes eigenvectors and eigen values for vertical error covariance matrix of $\psi$ , $T_u'$, $\chi_u'$ and q fields, variance of $p_{s\_u}'$ and eigen decomposition of $\psi'$, $\chi_u'$, $T_u'$ and q fields.

On the last stage "lengthscale (s)" calculated for each variable and each eigen mode.

Bellow on the fig. B. 2 some properties of B matrix displayed. Namely Fig.B.2. a) (left panel) represents the first five eigenvectors of psi –Stream function, chi_u, -unbalanced part of velocity potential**,** *t*_u, - unbalanced part temperature and rh-relative humidity variables. The eigenvectors are the results of EOF decomposition of the vertical auto covariance matrix and define vertical transform. On the Fig.B.2. b) horizontal length scales are shown for the same 4 variables.

The stream function and the potential velocity have the largest length scale value reaching 160 km and 120 km correspondingly. While, the unbalanced temperature length scale has a strong variation for the three first EOF passing approximately from 5 to 15 vertical modes and from there decreases from 40km to reach 10 km for the last EOF mode.

As the domain specific forecast error statistics computed, for diagnose and visualize B matrix properties is a good chose to run a single observation test, where only one (pseudo) observation is assimilated from a specific time and location within the analysis domain. In this case in analysis equation:

$$x^a = x^b + BH^T(HBH^T + R)^{-1}[y^o - H(x^b)] \quad \text{(Eq.A.6)}$$

it's assumed that for any control variable $[y^o - H(x^b)] = 1.0$ ; $R = I$. Thus, $x^a - x^b = B^*$ constant delta vector and only B matrix is corresponding on spread of increments in the point across the domain horizontally and vertically. In addition, how it affects the other variables.

We design our single observation experiment in this way: temperature was increased with 1 Kelvin in the center of the domain on the 500-hpa height. Two variational data assimilation systems WRFDA with WRF-ARW domain specific background errors and GSI with NAM regional background errors have been used. For GSI we performed two runs with $B_{nam}$ matrix. One of them was without tuning (lengthscale and variance options were set to 1) and with tuning (hzscl= 0.373, 0.746, 1.50). Background Forecast files have been similarly defined in all cases.

Fig. 3 shows analyses innovation for T, U and V variables for above-mentioned three runs. Fig.B.3. a) – corresponds to the results from GSI with regional $B_{nam}$ (without tuning), Fig.B.3. b) – GSI with regional $B_{nam}$ (with tuning), Fig.B.3. c) - WRFDA with our B. Each part of Fig.B.3 (a,b,c) shows two panels together left side - horizontal (XY at 11th sigma level) cross-sections of above mentioned three variables and right part - vertical cross-sections (XZ).

The first row on all figures show how temperature increment in the domain center spreaded horizontally and vertically. From figure 3a to 3b area where increment affects surrounded area reduces due to tuning length scale and variance parameters. On the figure 3c the affected area more concentrated in the center and more reduced.

Thus, the temperature Perturbation area produced from GSI recursive filter is larger than from WRFDA produced one with EOF mode. On the vertical cross-sections *XZ*, the temperature innovation has a larger impact on the vertical using our B than $B_{nam}$. These differences come from the dataset used to model these B matrices, the statistics in $B_{nam}$ are more climatological as they are averaged over time and they are interpolated on the mesh grid of our domain during the data assimilation process. While our B constructed from 2-month data set.

The second and third rows of figure 3 show how wind U and V components response on temperature perturbation. Horizontal and vertical cross-sections of this parameters show similar features.

To validate B matrix within both assimilation systems the single observation tests' result are realistic and very close to each other with expected differences.

## Conclusions

WRF model is the main tool for weather forecast in Georgia. The model have been tuned and tested over Georgia's territory for years. Nowadays as local meteorological network became denser and many remote observational sources are available to assimilate with variational methods is current challenge. We are working with two variational assimilation platforms suitable for this model namely WRFDA and GSI.

To estimate model forecast error in variational assimilation system, background error covariance matrix B, was successfully modeled and validated for Georgia's territory. To model B matrix GEN_BE v2.0 code has been used where model univariate or multivariate covariance errors from five control variables was taken as an input. This code gathers some methods and options that can be easily applied to different model inputs and used on different data assimilation platforms.

Different stages and transforms that lead to the modeling of the background error covariance matrix B and testing results by performing single observation tests was described and shown in this paper. B matrix modeled for our domain was tested on WRFDA platform using the EOF decomposition and was compared with the similarly designed test results on GSI platform using the recursive filters to model the vertical transform. The test shows similar results with comprehensive differences for the set of five control variables.

## Appendix B: Figures

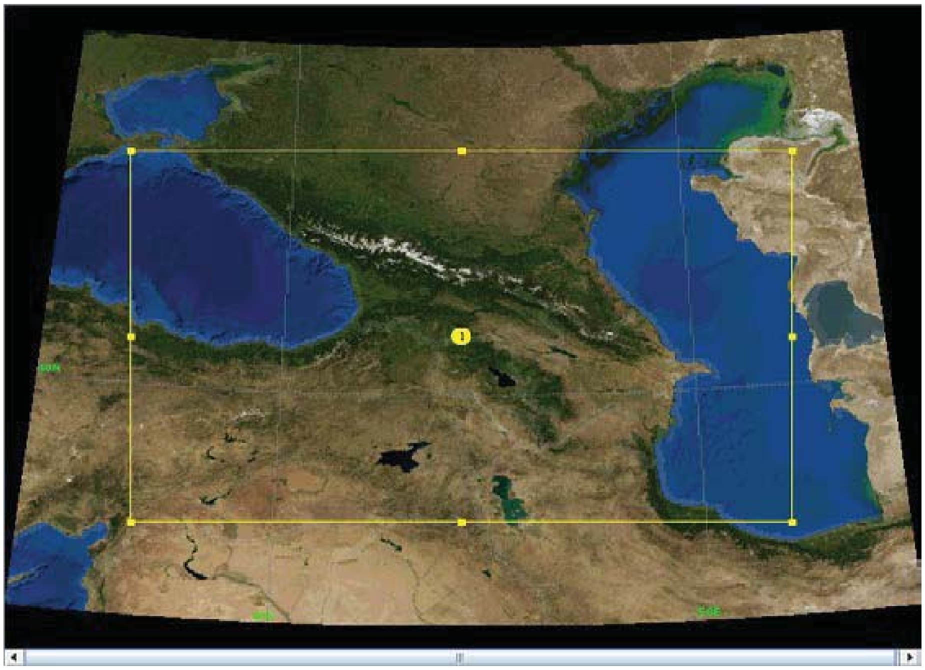


**Fig. B.1.** *Extension of the WRF-ARW computational domain*

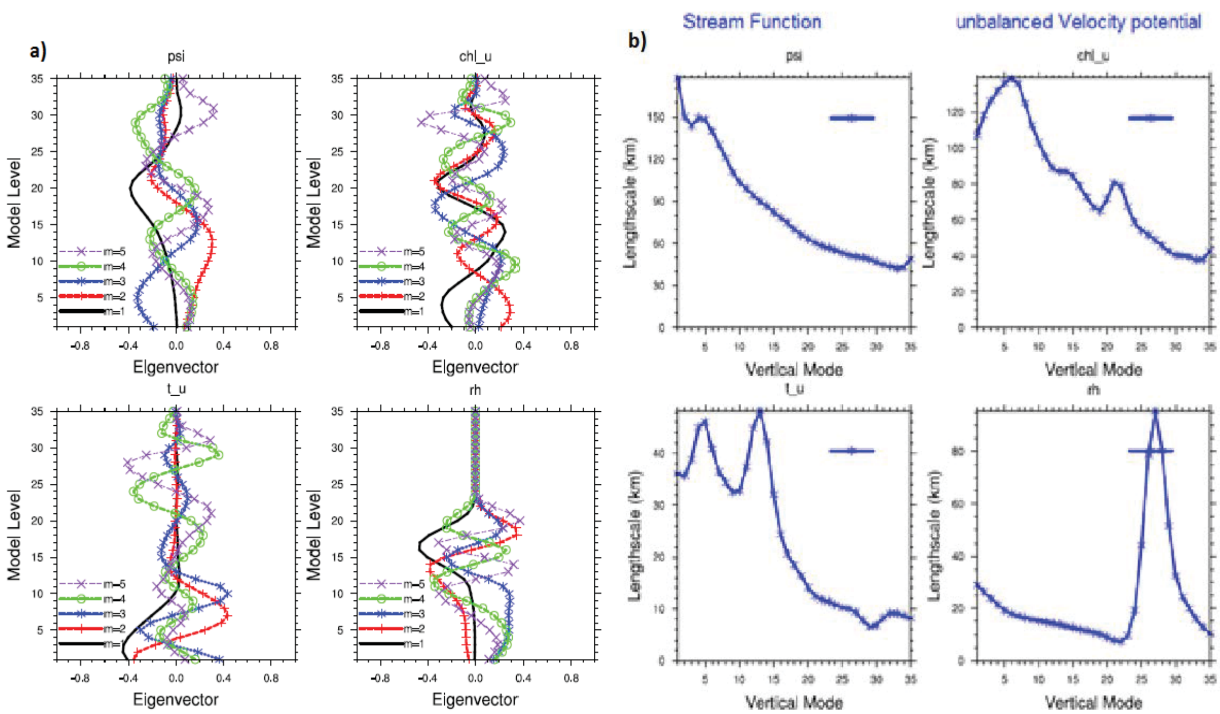



**Fig. 2. a).** *Five eigenvectors of psi –Stream function, chi_u, -unbalanced part of velocity potential, t_u, - unbalanced part temperature and rh-relative humidity variables; b) length scale factor for the same variables.*

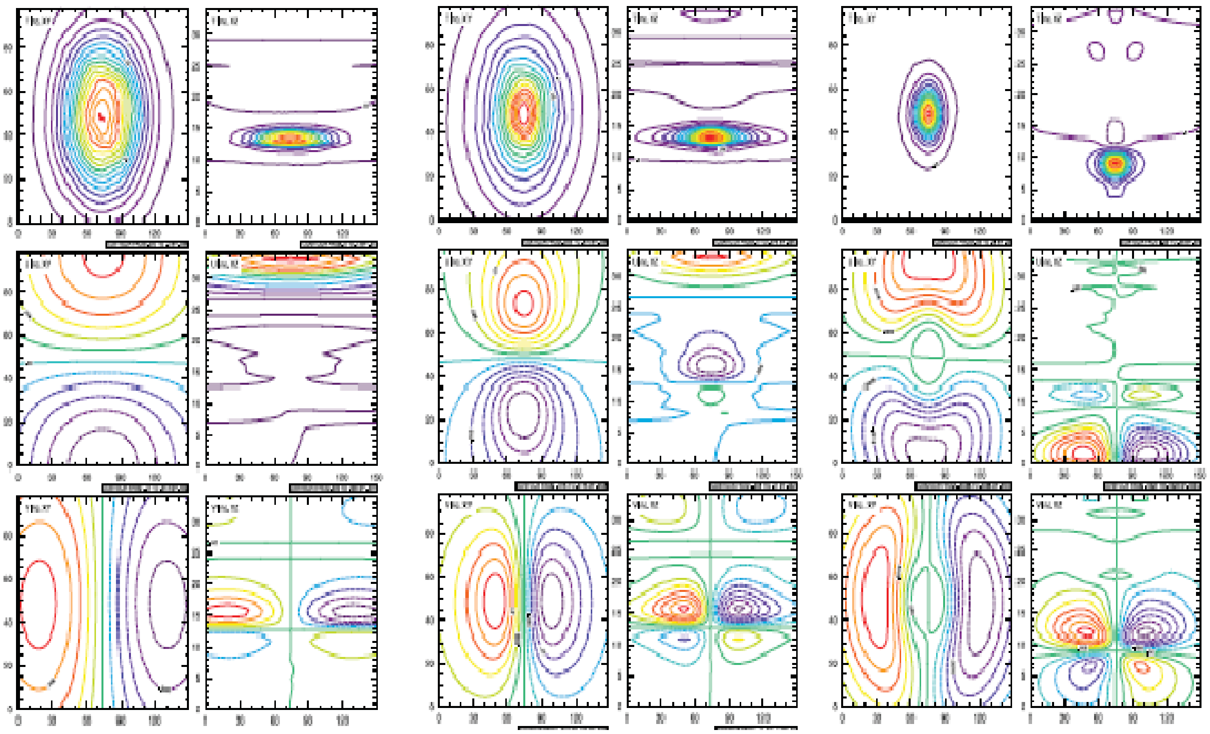



**Fig. 2. a).** *Analyses innovation for T, U and V variables. 3 a –results from GSI with regional Bnam (without tuning), 3b – GSI with regional Bnam (with tuning), 3c - WRFDA with our B. Each part of figure 3 (a,b,c) shows two panels together left side - horizontal (XY at 11th sigma level) cross-sections of above mentioned three variables and right part - vertical cross-sections (XZ).*

## Acknowledgment

The authors would like to thank to the Shota Rustaveli National Science Foundation (SRNSF) for the financial support (grant FR17_529).

## References

[1] R. Daley, Atmospheric Data Analysis. Cambridge University Press, 1991.

[2] RN. Bannister, A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Q. J. R. Meteorol. Soc., 134 (2008) 1951–1970.

[3] N.B. Ingleby, The statistical structure of forecast errors and its representation in the Met Office global 3-dimensional variational data assimilation system, Quart. J. Roy. Met. Soc., 127 (2001) 209-231.

[4] M. Fisher, Background error covariance modelling. In: ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 September 2003, ECMWF, Reading UK (2003) 45– 64.

[5] L. Berre, Estimation of synoptic and mesoscale forecast error covariances in a limited area model. Mon. Weather Rev. 128 (2000) 644–667.

[6] DM. Barker, W. Huang, YR. Guo, AJ. Bourgeois, QN. Xiao, A three-dimensional variational assimilation system for MM5: Implementation and initial results. Mon. Weather Rev. 132 (2004) 897– 914.

[7] O. Pannekoucke, L. Berre and G. Desroziers, Background error correlation length scale estimates and their sampling statistics. QJRMS. 134 (2008) 497-511.

[8] G. Descombes, T. Aulígné, F. Vandenberghe, and D. M. Barker, Generalized Background Error covariance matrix model (GEN_BE v2.0) Geosci. Model Dev. Discuss., 7 (2014) 4291–4352.

[9] TW. Schlatter, Variational assimilation of meteorological observations in the lower atmo-

sphere: a tutorial on how it works. J. Atmos. Solar–Terr. Phys. 62 (2000) 1057–1070.
[10] J. F. Caron, and L. Fillion, An examination of background error correlations and rotational wind over precipitation regions, Mon. Weather Rev., 138 (2010) 563–578,doi:10.1175/2009MWR2998.1.
[11] RN. Bannister, A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. Q. J. R. Meteorol. Soc. 134 (2008) 1971–1996.
[12] T. Montmerle, and L.Berre, Diagnosis and formulation of heterogeneous background error covariances at mesoscale, Q. J. Roy. Meteor. Soc., 136, 1408–1420 (2010) doi:10.1002/qj.655.
[13] D.F. Parrish and J.C. Derber, The National Meteorological Center's Statistical Spectral Interpolation Analysis System. Monthly Weather Review, 109 (1992) 1747-1763. http://dx.doi.org/10.1175/1520-0493 (1992 ) 120<1747:TNMCSS>2.0.CO;2
[14] E. Andersson, and H. Järvinen, Variational quality control. Q. J. R. Meteorol. Soc., 125 (1999) 697–722.