

Background error in WRF model

Nato Kutaladze, George Mikuchadze
The Hydrometeorological department at The National Environmental Agency
150 ave., D. Agmashenebeli, 0112, Tbilisi
Georgia

Abstract: - WRF model have been tuned and tested over Georgia's territory for years. First time in Georgia the process of data assimilation in Numerical weather prediction is developing. 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. Results of modeling of background error covariance matrix for control variables using WRF model over Georgia with desired domain configuration are discussed and presented. The modeling was implemented in two different 3DVAR systems (WRFDA and GSI) and results were checked by pseudo observation benchmark cases using also default global and regional BE matrixes. The mathematical and physical properties of the covariances are also reviewed.

Key-Words: - Analyses increment, assimilation, matrix, numerical weather prediction

Received: February 27, 2020. Revised: August 18, 2020. Accepted: August 21, 2020. Published: August 25, 2020.

1 Introduction

In numerical weather prediction (NWP), it is impossible to know the full state of the system at each grid point. There are different noisy sources of information available. One is short-term prediction, valid at the analysis time (also known as first guess or background information), which may not be accurate. Another is observations, where neither every model variable is measured nor observations are available at every grid point, considering measurement errors.

Data assimilation (DA) methods are widely used for the objective of state-estimation, based on the weighted combination of different sources of information. During the process of data assimilation a correction to the forecast is applied based on a set of atmospheric observations and estimated errors that are present in both the observations and the forecast itself [1, 2].

Estimating background errors is essential to the success of data assimilation, as they are giving proper weight to the background and, therefore, implicitly also to the observations. They also are correlated in space, which provides a means to propagate the information of the observations in three dimensions. The background errors of various meteorological variables are also correlated, allowing multivariate adjustments to be made to the analysis, representing the dynamic and physical balance of the atmosphere [3,4].

Development of data assimilation process, began with simple horizontal interpolation methods, which

gradually became three-dimensional, including also multivariate relationships – 3D variational DA. Later, dynamical processes have been utilized into DA process and the 4D variational DA was developed and applied operationally. To consider the flow dependence of forecast errors in the DA process, various forms of ensemble Kalman filters and variational techniques have been tested. More recently, hybrids between variational and ensemble DA methods have been proposed and are mainstreamed into Global centers [5,6]. The background error statistics are measured, described and used by each of the centers in different ways [7,8].

Traditionally, two distinct techniques are used to specify forecast-error covariances: the NMC (National Meteorological Center) and ensemble methods which are often combined with algebraic operations such as matrix factorization or covariance localization [9, 10, 11]. Matrices with homogeneous and isotropic characteristics such as diagonal are often favored. Other approaches are based on numerical techniques involving convolution operations or resolution of diffusion equations [12,13].

Recent works have also investigated model error covariance modelling (often noted as matrix Q), which can be seen as the main contributor of the background matrix in a dynamical system [14].

Georgia has recently started the data assimilation. We use a three-dimensional variational data assimilation method at this stage. In this paper, we present

background covariance matrix' properties for WRF-ARW model configured and tuned over South Caucasus domain generated with GEN_BE code and tested results within the two assimilation systems GSI and WRFDA [15].

The first section of this document describes the role of the background error covariance matrix in the data assimilation framework; the second describes methods for formulation of the B matrix, with difficulties and opportunities of its estimation. The third one presents functional steps of the B matrix generation by GEN_BE v2.0 with some technical details in our application and provides results of pseudo observation case in two different systems of data assimilation (WRFDA and GSI) using different B matrix involving the same set of five control variables (CV5).

2 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 (or penalty) is a combination of forecast and observation deviations from the desired analysis, weighted by forecast and observation-error covariance matrices.

$$J(\delta x, x_g) = 1/2(\delta x_b - \delta x)^T B(\delta x_b - \delta x) + 1/2[y_o - H(x_g + \delta x)]^T R^{-1}[y_o - H(x_g + \delta x)] \quad (1)$$

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

δx is difference between the analysis x_a and reference state or the 'first guess' x_g , i.e.

$$x_a = x_g + \delta x \quad (2)$$

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 107×107 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 [16, 17].

2.2. Background errors covariance matrix modeling

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 .

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

$$\delta x = B^{1/2}u \quad (3)$$

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

$$J(u) = 1/2u^T u + 1/2(d - HB^{1/2}u)^T R^{-1}(d - HB^{1/2}u) \quad (4)$$

Where d is the innovation vector defined as $d = (y_o - H(x_b))$ 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 Equation (3) is decomposed to a series of sub-matrices, each corresponding to an elemental transform that can be individually modeled:

$$U = SUpUvUh \quad (5)$$

Where, S is composed of the standard deviations of the background errors and is a diagonal matrix. Up matrix (Physical Transformation) defines the cross-correlations between different analysis variables via statistical balance (linear).

Uh - Horizontal Transform - defines the horizontal auto-correlations for the control variables. It is modeled through successive applications of recursive filters [18].

The matrix Uv defines the vertical auto-correlations for each of the control variables [19].

The modeled matrix B can be understood in terms of its horizontal and vertical structure functions. These structural functions determine the propagation of information from observations by background corrections for nearby horizontal and vertical grid points in accordance with the correlation (between gridpoints and variables) captured by the "offdiagonals" of B matrix.

3 Calculation and results

3.1. Variational Data assimilation

The theoretical framework for variational data assimilation is supported by Bayes' theorem: the best estimate of the state can be found from the conditional probability of the state given the

observations (posterior). This is mathematically a function of the probability of the state (prior) and conditional probability of the observations occurring given the state (likelihood). The best estimate of the state is known as the analysis. The variational approach involves finding the analysis that minimises a cost function J (see eq.1). Let J write as the sum of two members - the penalty contribution by the model (J_b) and observations (J_o)

$$J(\mathbf{x}) = J_b(\mathbf{x}, \mathbf{x}_b, \mathbf{B}) + J_o(\mathbf{x}, \mathbf{y}, \mathbf{R}) \quad (6)$$

Assuming that the prior and likelihood follow a Gaussian probability distribution, we apply Bayes' theorem to determine the posterior. J_b is larger if \mathbf{x} deviates more from \mathbf{x}_b with smaller background errors. J_o is larger if \mathbf{x} deviates more from its corresponding observations with smaller observation errors.

The analysis ($\mathbf{x}=\mathbf{x}_a$) that minimizes this cost function has the maximum a posteriori probability; it is the mode of the posterior.

The solution \mathbf{x}_a

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{B}, \mathbf{R})\mathbf{d}(\mathbf{x}_b, \mathbf{y}) \quad (7)$$

Where, \mathbf{K} is commonly referred to as the gain matrix, as a function of \mathbf{B} and \mathbf{R} and \mathbf{d} is the innovation - difference between observations and the equivalent observations constructed from the background. (\mathbf{Kd}) quantitatively estimates increment - the corrections to add to the background. This is the mathematically optimal approach for finding the best estimate of the state in variational data assimilation.

3.2. Calculation of the B matrix

For this study, we calculate Background error covariance matrix \mathbf{B} using GEN_BE code version 2.0 (compiled with Intel compilers) in WRFDA for WRF-ARW model over the 9.2 km domain (Fig.1) with 151 x 100 x 36 grid cells.

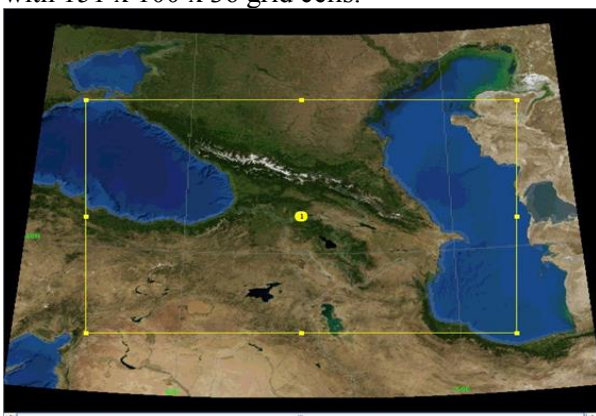


Fig.1, Extension of the WRF-ARW computational domain

The run comprises from 5 stages, having separate input output infrastructure and managed via namelist file, where control variables and all parameters to model \mathbf{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 [20] 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 (e.g. one month).

For this run, spring 2020 12 and 24-hour WRF-ARW forecasts, initialized both at 00 and at 12 UTC was used. Thus in all 180 pairs of perturbations are utilized to generate WRF-ARW Background Error.

On the initial stage analyses control variables stream function (ψ) and unbalanced velocity potential (χ_u) are calculated from u and v wind, then differences for following 5 control variables: stream function (ψ), unbalanced velocity potential (χ_u), Temperature (T), Relative Humidity (q), Surface Pressure (ps) have been created. 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 [21].

On the stage, 3 regression coefficients and balanced part of χ , T and ps variables are computed. The simplest way to model correlation between them is to use linear regression. Firstly, the regression coefficients between variables are calculated, and then linear regressions are performed to derive uncorrelated control variables and then remove the balanced part for each other variable. This part is achieved by the physical transformation (U_p). It models correlations between variables and allows transforming to the diagonal matrix in the control variables (uncorrelated) space and computes unbalanced parts for the same variables:

$\chi_u' = \chi' - \chi_b$; $Tu' = T' - T_b$; $ps_u' = ps' - ps_b$ is the preliminary step before estimating the vertical and horizontal auto-correlation parameters for each control variable.

Stage 4 Removes mean for χ_u' , Tu' & ps_u' and computes eigenvectors and eigen values for vertical error covariance matrix of ψ , Tu' , χ_u' and q fields, variance of ps_u' and eigen decomposition of ψ' , χ_u' , Tu' and q fields.

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

Bellow on the fig. 2 some properties of B matrix displayed.

Namely Fig.2 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.3, horizontal length scales are shown for the same 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.

3.3. A single pseudo observation test

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 the analysis equation:

$$x_a = x_b + BH^T(HBH^T + R)^{-1}[y_o - H(x_b)] \quad (8)$$

It's assumed that for any control variable $[y_o - H(x_b)] = 1.0$; $R = I$. Thus, $x_a - x_b = B^* \text{ constant delta vector}$ 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 also estimated.

We design our single observation experiment in this way: wind U component was increased with 1 m/s in the center of the domain on the 500-hpa height.

The benchmark case was carried out into two variational data assimilation systems WRFDA and GSI, with similar background forecast files and namelist settings and 3 different B matrices: our domain specific B, Bnam and default Global B.dat_cv3. We performed several runs with Bnam and B.dat_cv3 matrices, with and without tuning (lengthscale and variance options).

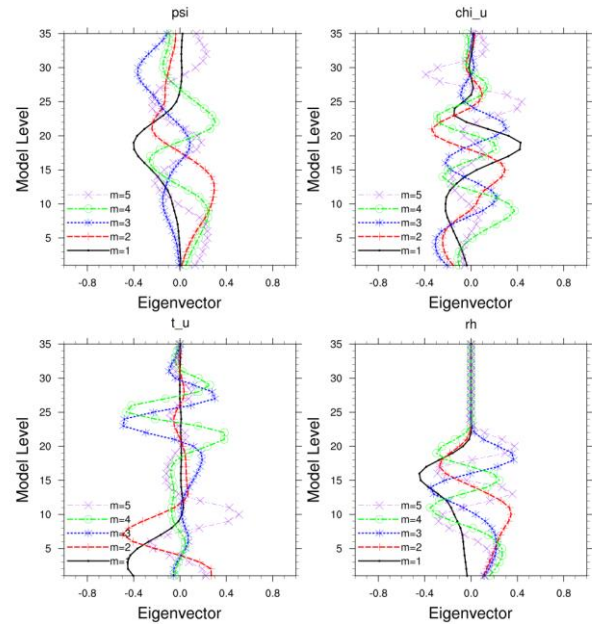


Fig.2, Five eigenvectors of psi – Stream function, chi_u, - unbalanced part of velocity potential, t_u, - unbalanced part temperature and rh-relative humidity variables

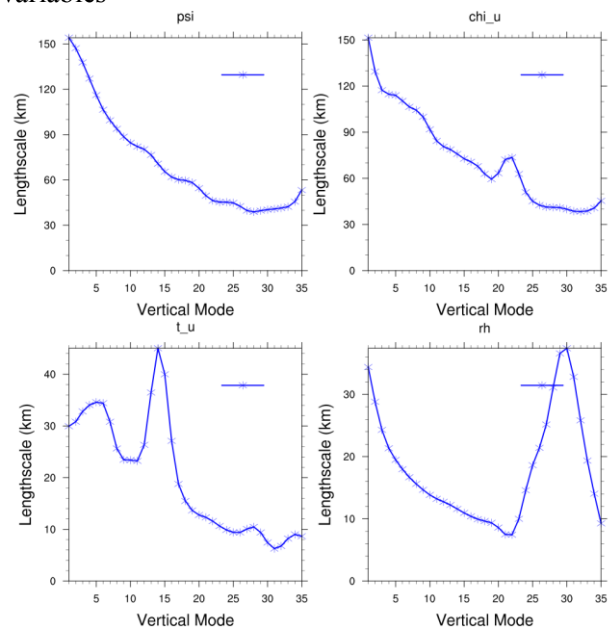


Fig. 3, length scale factor for the Stream function, unbalanced part of velocity potential, unbalanced part temperature and relative humidity variables

Information spreading areas horizontally and vertically, for U-wind were compared from all cases; also, its impact on V wind and Temperature increment propagation was studied.

Results of how the increment of U_wind in WRFDA with our B in the domain center, affects temperature and V_wind horizontally are presented on the fig. 4. From all simulation is clear that Bnam and B.dat_cv3

matrices are producing very large spread area, with distorted shapes. It's undesirable to use them without tuning, which was expected.

The Perturbation area produced from GSI recursive filter, with Bnam is larger than from WRFDA produced one with EOF mode, but due to tuning length scale and variance parameters affected area became more concentrated in the center and reduced. Even after tuning, the advantage of our B matrix in both systems is obvious. Only the U_{wind}'s innovation has a larger impact on the vertical cross-sections XZ, using our B than Bnam. Direct comparison of these statistics is difficult, as they are performed with different models, configurations, and physical options, but it's also worth to mention, that properties of B depends on data sample and its size, for which it was modeled. B_{ma} modeled from NMC method applied to 60 perturbations taken over a year, while our B was constructed from 2-months data set. Our domain specific B matrix was validated within both assimilation systems via the single observation tests' successfully. The results have sound physical meaning and are well expected.

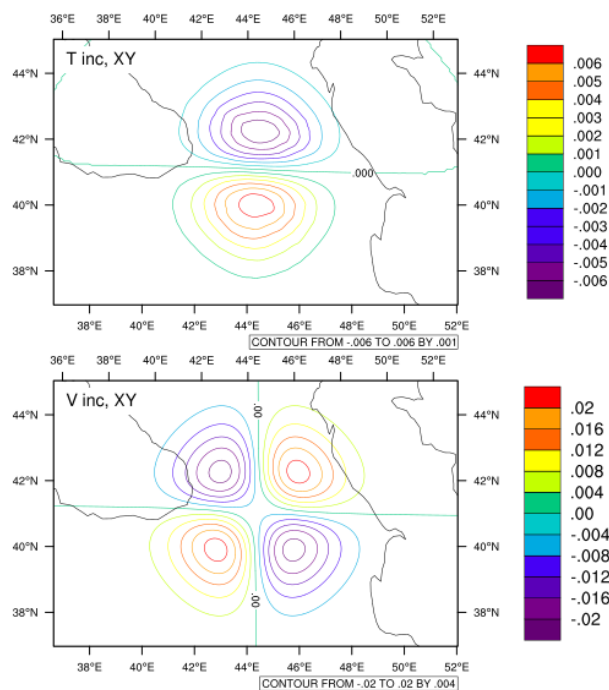


Fig.4, Analyses innovation for T and V variables in WRFDA with our B matrix.

4 Conclusion

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 were 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. Successful modeling of the B matrix is a prerequisite for further development of NWP system in the region, where the orography is quite complex and peculiarities of locally developed weather phenomena at any time a year are often characterized with diversity and extremity. It should also be taken into account that the quantity and quality of observational data for assimilation process really is not sufficient in the region, which underlines the demand for the reliability of this matrix.

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 pp.1951–1970.
- [3] N. Gustafsson, T. Janjić, C. Schraff et al. "Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres." *Q J R Meteorol Soc.* 2018 144 pp. 1218–1256. <https://doi.org/10.1002/qj.3179>.
- [4] JCK. Lee, X-Y. Huang, "Background error statistics in the Tropics: Structures and impact in a convective-scale numerical weather prediction system." *QJR Meteorol Soc.* 2020 pp. 1–20. <https://doi.org/10.1002/qj.3785>
- [5] J. Bédard, J. Caron, M. Buehner, S. Baek, and L. Fillion. "Hybrid Background Error Covariances for a Limited-Area Deterministic Weather Prediction

System.” *Wea. Forecasting*, 35, 2020 pp. 1051–1066, <https://doi.org/10.1175/WAF-D-19-0069.1>.

[6] T. Ishibashi, “Improvement of Accuracy of Global Numerical Weather Prediction Using Refined Error Covariance Matrices.” *Mon. Wea. Rev.*, 148, 2020 pp. 2623–2643, <https://doi.org/10.1175/MWR-D-19-0269.1>.

[7] V. Valler, J. Franke and S. Brönnimann “Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation,” *Clim. Past*, 15, 2019, pp. 1427–1441, <https://doi.org/10.5194/cp-15-1427-2019>

[8] S. Cheng, J-P. Argaud, B. Iooss, D. Lucor, A. Ponçot. “Background Error Covariance Iterative Updating with Invariant Observation Measures for Data Assimilation”. 2019. hal-02307657

[9] T. Ishibashi. Tensor formulation of ensemble-based background error covariance matrix factorization. *Monthly Weather Review*, 143 (12) 2015. pp. 4963–4973

[10] C. Liu and M. Xue. Relationships among four-dimensional hybrid ensemble–variational data assimilation algorithms with full and approximate ensemble covariance localization. *Monthly Weather Review*, 144(2) 2016 pp. 591–606.

[11] L. Berre, “Estimation of synoptic and mesoscale forecast error covariances in a limited area model.” *Mon. Weather Rev.* 128, 2000, pp. 644–667.

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

[13] O. Pannekoucke, L. Berre and G. Desroziers, “Background error correlation length scale estimates and their sampling statistics”. *QJRMS*. 134, 2008, pp. [14] 497-511

[15] D. Dreano, P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit. Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Quarterly Journal of the Royal Meteorological Society*, 143(705) 2017 pp. 1877 – 1885.

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

[17] TW. Schlatter, Variational assimilation of meteorological observations in the lower atmosphere: a tutorial on how it works. *J. Atmos. Solar–Terr. Phys.* 62, 2000, pp. 1057–1070.

[18] J. F. Caron, and L. Fillion, “An examination of background error correlations and rotational wind

over precipitation regions”, *Mon. Weather Rev.*, 138, 2010, pp. 563–578, doi:10.1175/2009MWR2998.1

[19] 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

[20] T. Montmerle, and L. Berre, “Diagnosis and formulation of heterogeneous background error covariances at mesoscale”, *Q. J. Roy. Meteor. Soc.*, 136, 2010, pp.1408–1420, doi:10.1002/qj.655

[21] D.F. Parrish and J.C. Derber, “The National Meteorological Center’s Statistical Spectral Interpolation Analysis System” *Monthly Weather Review*, 109, 1992 pp. 1747-1763. <http://dx.doi.org/10.1175/1520-0493>

[22] E. Andersson, and H. Järvinen, “Variational quality control”. *Q. J. R. Meteorol. Soc.*, 125, 1999, pp. 697–722.

Sources of funding for research presented in a scientific article or scientific article itself

The research was funded by the Shota Rustaveli National Science Foundation (SRNSF) (grant FR17_529).

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US