

Formal estimation of the random component in global maps of total electron content

V. Ya. Choliy*

Taras Shevchenko National University of Kyiv, Glushkova ave., 4, 03127, Kyiv, Ukraine

Random component of the total electron content (TEC) maps, produced by global navigation satellite system processing centres, was analysed. Helmert transform (HT) and two-dimension singular spectrum analysis (2dSSA) were used. Optimal parameters (in the sense calculation speed versus quality) of the 2dSSA windows were determined along with precision estimations.

Key words: ionosphere modelling and forecasting

INTRODUCTION

Helmert transform (HT) and singular spectrum analysis (SSA) in one-dimension variant were used by author and his colleagues in their works [1, 2, 3, 4, 5]. This article closes the series. Beside the TEC map analysis the article contains the comparison of Helmert transform (HT) and two-dimensional SSA (2dSSA). It should be noted that latter of the methods needs much more calculations than the first one. That is why determination of the optimal values of 2dSSA window parameters as of independent significance.

All files containing TEC maps in IONEX format have been taken from NASA server¹ for six successive solstices and equinoxes: 22/09/2014, 22/12/2014, 21/03/2015, 20/06/2015, 22/09/2015, 22/12/2015. Very short information about the data sources is given in Table 1. Different files contain different amount of maps calculated for different time moments inside the day. The only maps equally distributed during the day with two hour interval were used. It means 13 maps per day per file. All the maps have identical dimensions: 71 rows (latitude step is 2.5°) and 72 columns (longitude step is 5°).

The second column of Table 1 contains six symbols. Each of them is to be used for the single date from the just specified date list. Their meanings are: (+) the maps of the file were used for comparison, (−) the maps were rejected during steps of comparison, (0) there was no file for the date. Each data processing centre uses different methods to build TEC map: different amount of satellites, stations, smoothing, combination, forecasting or just present raw data. None of those differences were taken into account during the analysis. We just used the maps as they were provided. That is why in Table 1 we have presented the data from file headers only with

minimum comments. It should be taken in mind that combined solutions *Igrg* and *Igsg* for different dates were created with different composing source maps.

Table 1: Main properties of the TEC maps.

Name	Key	Office	Comment
<i>C1pg</i>	++++++	AIUB	1 day forecast
<i>C2pg</i>	++++++		2 days forecast
<i>Codg</i>	++++++		
<i>Corg</i>	++++++		raw
<i>E1pg</i>	-----0	ESA/ESOC	1 day forecast
<i>E2pg</i>	0-----0		1 day forecast (?)
<i>Ehrg</i>	++++++		
<i>Emrg</i>	000--+	NRCan/CGS	
<i>Esag</i>	++++++	ESA/ESOC	
<i>Esrg</i>	++++++		
<i>Igrg</i>	++++++	GRL/UWM	combined
<i>Igsg</i>	++++++		combined
<i>Jplg</i>	++++++	JPL-GNISD	
<i>Jprg</i>	++++++		
<i>U2pg</i>	-----0	gAGE/UPS	2 day forecast
<i>Uhrg</i>	++++++		
<i>Upcg</i>	++++++	UPC-IonSAT	raw
<i>Uprg</i>	++++++		raw
<i>Uqrg</i>	++++++		

All maps for the single day were analysed separately. Each element of the map represents the point on the surface with the radius-vector is TEC value:

$$\mathbf{r} = \text{TEC}(\cos \lambda \cos \varphi, \sin \lambda \cos \varphi, \sin \varphi),$$

and λ, φ are its coordinates (longitude and latitude).

No pre-selection criteria were applied to the map list before analysis.

*Choliy.Vasy1@gmail.com

¹ftp://cdis.gsfc.nasa.gov/pub/gps/products/ionex/

Table 2: Random error estimation from HT (in 0.1 TECU).

name	22/09/2014		22/12/2014		21/03/2015		20/06/2015		22/09/2015		22/12/2015	
C1pg	4.40	1.50	7.41	2.64	2.88	1.25	2.11	0.42	1.73	0.81	3.23	1.48
C2pg	4.10	0.70	7.43	2.12	3.13	0.92	2.12	0.41	1.90	0.20	3.08	0.66
Codg	1.86	0.69	3.15	1.37	2.10	0.81	0.65	0.09	0.77	0.27	0.89	0.29
Corg	1.41	0.38	4.32	1.18	1.77	0.56	0.78	0.16	0.79	0.27	1.59	0.56
Ehrg	1.46	0.78	2.61	1.27	1.69	0.72	1.05	0.31	0.66	0.23	1.24	0.63
Emrg	—	—	—	—	—	—	—	—	—	—	3.83	1.09
Esag	1.91	0.54	3.64	1.33	1.94	0.80	0.96	0.26	0.82	0.18	1.53	0.51
Esrg	2.07	0.71	3.64	1.28	2.23	0.65	1.10	0.27	0.96	0.23	1.72	0.48
Igrg	0.53	0.16	0.43	0.15	0.79	0.34	0.19	0.03	0.26	0.04	0.28	0.10
Igsg	0.85	0.38	1.13	0.43	0.96	0.42	0.35	0.07	0.35	0.16	0.54	0.21
Jplg	1.54	0.52	2.69	0.83	1.74	0.43	0.56	0.09	0.76	0.29	1.08	0.24
Jprg	1.54	0.53	2.49	0.77	1.91	0.71	0.62	0.11	0.78	0.30	1.02	0.25
Uhrg	3.34	1.16	4.88	1.79	3.05	0.82	0.90	0.11	1.36	0.20	1.91	0.63
Upcg	1.79	1.05	2.52	1.70	1.51	0.91	0.80	0.23	0.63	0.35	0.84	0.58
Uprg	1.79	1.05	2.52	1.70	1.51	0.91	0.80	0.23	0.63	0.35	0.84	0.58
Uqrg	3.34	1.16	4.88	1.79	3.05	0.82	0.90	0.11	1.36	0.20	1.91	0.63

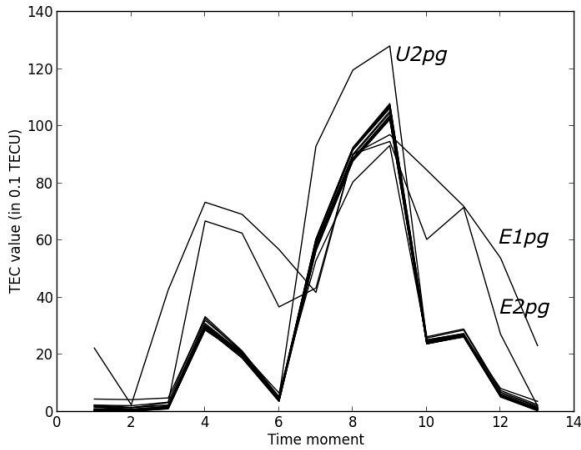


Fig. 1: First step of TEC map analysis for 20/06/2015.

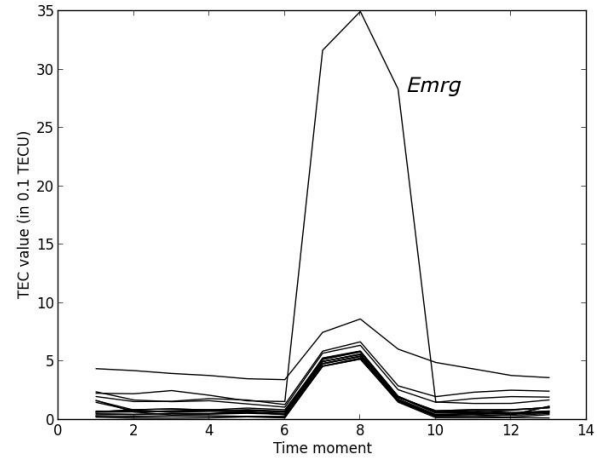


Fig. 2: Next step of TEC map analysis for 20/06/2015.

HELMERT TRANSFORM

However for HT such a preliminary analysis is necessary, but however it may be done with the help of HT itself. Let us show it with three steps of the analysis for 20/06/2015, where the analysis is the most illustrative. All figures (Figs. 1–3) have TEC (in units of 0.1 TECU) along Y axis, but in different scales. Abscissas are the map sequential number, or, in other words, the time from 0^hUTC of the date in two-hour intervals. Different lines on the same figure are for different files, say, different processing centres. Only some of the maps were marked, as these figures (Figs.1–3) outlines the pre-selection process and the figures are only for illustration.

Graph in Fig. 1 is essential for the case of patchy data. At this step the most shifted maps (E1pg, E2pg) and the map U2pg (declared by its authors as a preliminary one) were removed from the analysis.

It gave much better but still not quite good picture with the only shifted map (Fig.2). It is Emrg map. Fig.3 was left for further analysis despite of two clearly distinct graphs for C2pg and C1pg (two uppermost lines from top). These are AUIB forecasts for two and one days. They demonstrate a little worse precision as compared with another ones, but still are quite good for forecasts. Another two (bottommost) are combines map Igrg and Igsg. All other maps' precision lies around and lower than ~ 0.1 TECU.

This situation is repeatable for another dates where some of the maps must be removed due to their shifts and uncompensated errors, not analysed here.

Comparison results in TECU units are given in Table 2. Each cell contains daily averaged estimation of the random precision of the TEC map and its standard deviation.

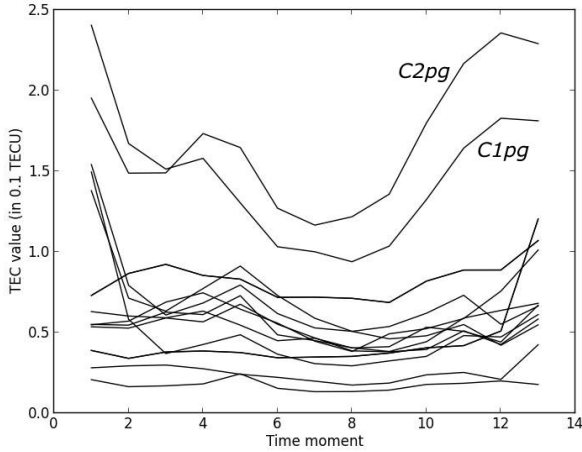


Fig. 3: Final step of TEC map analysis for 20/06/2015.

TWO DIMENSION SSA

There are some specific features of 2dSSA. Let us shortly outline them here. The algorithm of one dimension SSA clearly explained in original works [6, 7], was used by author for the analysis and forecasting time series of pole coordinates. 2dSSA and principal component analysis are widely used in the analysis of photo and video, e.g. [8]. Applying 2dSSA to the precision analysis of ionospheric maps, most likely, is attempted here for the first time.

At this stage, the TEC map is black and white image, where brightness of its pixels is defined by TEC value. In general the SSA algorithm does not undergo any changes and consists of the steps:

- building of the trajectory matrix \mathbb{X} ;
- singular value decomposition of the matrix $\mathbb{S} = \mathbb{X} \cdot \mathbb{X}^T$, say, in the sum of the principal components;
- grouping of the eigenvalues and eigenvectors of the matrix \mathbb{S} to subdivide principal components into deterministic and random ones;
- restoration of the random (noisy) principal components to determine their standard deviation;
- restoration of the deterministic part of the map for its analysis (for example, for forecasting).

There are some features of the 2dSSA making difference in some of its steps from 1d variant. In 1d the SSA window is applying to the time series and the only window parameter is its length L . In 2d the window is rectangular and has width w and height h . The image itself has size $W \times H$. The SSA window is scanning the image from upper left to bottom right corners like the beam in television tube and in

every of its position it generates new column of the trajectory matrix:

$$\begin{pmatrix} \lceil I_{1,1} & I_{1,2} & I_{1,3} \rceil & I_{1,4} & \dots \\ I_{2,1} & I_{2,2} & I_{2,3} & I_{2,4} & \dots \\ \lfloor I_{3,1} & I_{3,2} & I_{3,3} \rfloor & I_{3,4} & \dots \\ I_{4,1} & I_{4,2} & I_{4,3} & I_{4,4} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \rightarrow \begin{pmatrix} I_{1,1} & \lceil I_{1,2} & I_{1,3} & I_{1,4} \rceil & \dots \\ I_{2,1} & I_{2,2} & I_{2,3} & I_{2,4} & \dots \\ I_{3,1} & \lfloor I_{3,2} & I_{3,3} \rfloor & I_{3,4} & \dots \\ I_{4,1} & I_{4,2} & I_{4,3} & I_{4,4} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \rightarrow \dots \rightarrow \begin{pmatrix} \dots & \lceil I_{1,w-3} & I_{1,w-2} & I_{1,w-1} \rceil & I_{1,w} \\ \dots & I_{2,w-3} & I_{2,w-2} & I_{2,w-1} & I_{2,w} \\ \dots & \lfloor I_{3,w-3} & I_{3,w-2} & I_{3,w-1} \rfloor & I_{3,w} \\ \dots & I_{4,w-3} & I_{4,w-2} & I_{4,w-1} & I_{4,w} \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \rightarrow \dots \rightarrow \begin{pmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & I_{h-3,w-3} & I_{h-3,w-2} & I_{h-3,w-1} & I_{h-3,w} \\ \dots & I_{h-2,w-3} & \lceil I_{h-2,w-2} & I_{h-2,w-1} & I_{h-2,w} \rceil \\ \dots & I_{h-1,w-3} & I_{h-1,w-2} & I_{h-1,w-1} & I_{h-1,w} \\ \dots & I_{h,w-3} & \lfloor I_{h,w-2} & I_{h,w-1} & I_{h,w} \rfloor \end{pmatrix},$$

which leads to the trajectory matrix like this:

$$\mathbb{X} = \begin{pmatrix} I_{1,1} & I_{1,2} & \dots & I_{1,w-2} & \dots & I_{h-2,w-2} \\ I_{1,2} & I_{1,3} & \dots & I_{1,w-1} & \dots & I_{h-2,w-1} \\ I_{1,3} & I_{1,4} & \dots & I_{1,w} & \dots & I_{h-2,w} \\ I_{2,1} & I_{2,2} & \dots & I_{2,w-2} & \dots & I_{h-1,w-2} \\ I_{2,2} & I_{2,3} & \dots & I_{2,w-1} & \dots & I_{h-1,w-1} \\ I_{2,3} & I_{2,4} & \dots & I_{2,w} & \dots & I_{h-1,w} \\ I_{3,1} & I_{3,2} & \dots & I_{3,w-2} & \dots & I_{h,w-2} \\ I_{3,2} & I_{3,3} & \dots & I_{3,w-1} & \dots & I_{h,w-1} \\ I_{3,3} & I_{3,4} & \dots & I_{3,w} & \dots & I_{h,w} \end{pmatrix}. \quad (1)$$

Thereby the trajectory matrix owns the block-hankel structure. It has $(H - h + 1) \times h$ blocks of $(W - w + 1) \times w$ size. It should be noted when building the transform:

$$(\text{image}) \leftrightarrow (\text{trajectory matrix}).$$

Let us index the block with uppercase letters:

$$I = 0, \dots, H - h, \quad J = 0, \dots, h - 1,$$

and pixels inside the block with the lowercase ones:

$$i = 0, \dots, W - w, \quad j = 0, \dots, w - 1.$$

Now we can explain block-hankel structure via block and inner block indices. Any point within the image with the coordinates (R, C) , where R is a row

number (from top to bottom), C is a column number (from left to right), will participate in the blocks of the trajectory matrix where:

$$I + J = R,$$

moreover, in the upper leftmost point of any block there is the point with $C = 0$, in the upper rightmost the one with $C = W - w$, and in lower rightmost the point with $C = W - 1$. It is easily recognizable from (1). Using that information one can select the starting point for hankelization and then follow the anti-diagonal, skip the unnecessary blocks, do the averaging.

In short words the 2dSSA hankelization is the averaging of the values for:

- any blocks where $I + J = R$,
- along anti-diagonal, starting in every appropriate blocks at the point:

$$\begin{aligned} 0 \leq C < w - 1 & : (C, 0), \\ w \leq C < W & : (W - w, W - 1 - C). \end{aligned}$$

Transformation from (R, C) to $(i, j) + (I, J)$ is used on the first step of 2dSSA, but backward algorithm — on the two last steps, during principal component restoration. We left implementation issues beside the article scope.

Any other steps of the analysis does not diverge from the 1d variant.

MAP ANALYSIS WITH 2DSSA

To determine the applicability of 2dSSA algorithms to the postulated task we first run some numerical experiment. For this purpose the maps from 22/09/2014 were analysed by 2dSSA with different windows, starting from 6×5 window (6 longitudinal points by 5 latitudinal ones), and ending with the 36×35 window. The size of the window varied with step 6 for longitude and 5 for latitude. The smallest window led to eigenvalue spectrum of 30 points, the greatest one (it is nearly the quarter of the whole map) to the spectrum of 1260 values.

Typical overview of the eigenvalue spectra (decimal logarithm of the eigenvalue against its sequential number after descending sort) is shown in Fig. 4. Three regions at the spectra are clearly recognizable. Each of them can be approximated with linear function of different slopes. In our opinion, there can be an interpretation of these region. Region 3 of the spectra is mostly due to computing rounding errors. Whole bunch of region 3 principal components can explain a very tiny ($10^{-6} \sim 10^{-8}$) portion of the image standard deviation. Region 1 is for the main (deterministic) part of the map. Then the region 2 is for random portion of the TEC map.

Some additional arguments for this point of view can be found in Fig. 5, where the dependence of the

total standard deviation of the first m principal components for different $6 \times n$ windows are presented. There was no sense to present lines for every window, as the figure become too crowded. We gave only some of them. For another windows the general view of the graph is quite similar. For low n the map deviations was not accounted in full. Greater n better represents the standard deviation of the map and asymptotically $n \rightarrow w \times h$ it tends to the standard deviation of the whole map. Position of the breaking point on the graph in Fig. 5 is like the A point in Fig. 4.

That is why we interpret the A point as a border between the deterministic and random parts of the TEC map. Doing precision analysis one can limit itself to the region 2, as with the region 3 principal components only very little part of the total standard deviation may be explained. It was proven by direct calculations.

Additional analysis with linear approximations of the spectra and counting of the principal components led us to the conclusion that in the region 1 there is a small number of the first principal components. Sometimes it is the only one of them. Empirical rule for amount of the principal components in region 1 is given in Table 3. There are window sizes along vertical and horizontal table limits. There are data lacks in the Table as it was too much time consumed to calculate them. In Table 4 there are standard deviations of the region 2 principal components for Igrg map calculated according to Table 3 recommendations. Despite the variability, the results are similar.

Total standard deviations of the region 2 principal components for all TEC maps of 20/06/2016 were calculated. After that the random error estimations as a part of the standard deviation in its full value were calculated and presented in Table 5. We used the greatest possible window with the smallest possible amount of principal components in region 1 — namely one. This is 10×12 window. The coincidence of these values (given in Table 5) with the data in Table 2 is poor. The correlation between them never exceeds 0.5. It means that 2dSSA can be used for the task, but the results stay uncertain. The only item we should point on now is that 2dSSA estimations are much more smooth which is very essential for the results based on the same raw data and nearly the same processing strategy.

The “classical” variant of the subdivision of principal components by the analysis of correlation functions shows consistency with the proposed methodology.

CONCLUSION

Singular spectrum analysis method may be used with some success for determination of the formal precision of the TEC maps if left behind the scene the fact of the calculation time. In real applications

Table 3: Recommended number of region 1 principal components.

	6	12	18	24	30	36
5	1	1	2	2	2	2
10	1	1	2	2	2	2
15	1	2	3	2	2	3
20	2	2	3	3	3	-
25	2	3	4	-	-	-
30	3	4	4	-	-	-
35	3	4	-	-	-	-

Table 4: Standard deviations estimated for region 2 for 20/06/2016, Igrg map.

	6	12	18	24	30	36
5	0.1155	0.4580	0.1136	0.2120	0.3458	0.3596
10	0.3366	0.5797	0.2650	0.3265	0.4167	0.4550
15	0.5772	0.4523	0.1919	0.6244	0.6704	0.6976
20	0.2665	0.5321	0.2611	0.3374	0.4175	-
25	0.5679	0.4650	0.6160	-	-	-
30	0.3013	0.4091	0.5564	-	-	-
35	0.3044	0.3780	-	-	-	-

Table 5: Random error for TEC map from 2dSSA (in 0.1 TECU).

Name	C1pg	C2pg	Codg	Corg	Ehrg	Esag	Esrg	Igrg	Igsg	Jplg	Jprg	Uhrg	Upcg	Uprg	Uqrg
precision	0.71	0.55	0.86	0.65	0.67	0.62	0.59	0.58	0.78	0.72	0.66	0.78	0.62	0.62	0.78

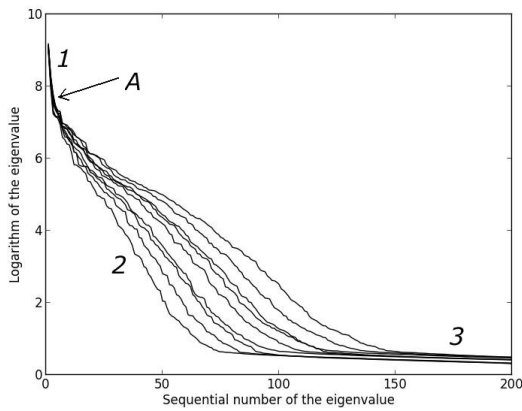


Fig. 4: General view of the eigenvalue spectra for different assorted windows.

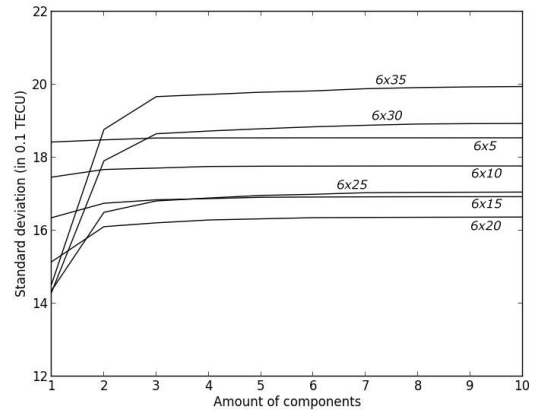


Fig. 5: Standard deviation via principal component amount.

the necessary time is huge as compared with HT. By the way, if one has the only field to be analysed, SSA may be the only possibility, since other methods need intercomparison of the data sets (in the case of HT there are three sets necessary).

Comparison of our results with [9] shows that our estimations are lesser than ones from [9]. Unfortunately we cannot estimate the level of random errors introduced by processing methods, but they may be different and influence the estimations.

There is another difference in HT and 2dSSA, which may cause the difference in the results. HT builds the systematic difference model according to some predefined formulae. It is affine transform including rotation, shift and possible inclination. In 2dSSA the systematic difference is build from the scratch for each of the data sets and may be different for different maps. It needs additional investigation.

REFERENCES

- [1] Choliy V. Ya. 2015, *Kosmichna Nauka i Tehnologiya*, 21, 1, 70
- [2] Choliy V. Ya. 2015, *Kinematics and Physics of Celestial Bodies*, 31, 4, 205
- [3] Choliy V. Ya. 2014, *Kinematics and Physics of Celestial Bodies*, 30, 6, 304
- [4] Choliy V. Ya. & Tkachuk V. V. 2013, *Advances in Astronomy and Space Physics*, 3, 141
- [5] Choliy V. Ya. 2014, *Advances in Astronomy and Space Physics*, 4, 15
- [6] Golyandina N., Nekrutkin V. & Zhigljavsky A. 2001, *'Analysis of time series structure'*, Chapman, New York
- [7] Golyandina N. & Zhigljavsky A. 2013, *'Singular spectrum analysis for time series'*, Springer, Berlin
- [8] Rodrigues-Aragon L. & Zhigljavsky A. 2010, *Statistics and its Inference*, 3, 419
- [9] Rovira-Garcia A., Juan J.M., Sanz J., González-Casado G. & Ibáñez D. 2016, *J. Geodesy*, 90, 229