Extraction of the daily quiet variation from the geomagnetic field observations with the principal component analysis

Anna Morozova, Rania Rebbah

1 Centre for Earth and Space Research of University of Coimbra, Department of Physics, University of Coimbra, 3000-456 Coimbra, Portugal
2 Department of Physics, University of Coimbra, 3000-456 Coimbra, Portugal
Abstract

- Geomagnetic field (GMF) variations from external sources are classified as regular (diurnal) or occurring during periods of disturbances.
- The most significant regular variations are the quiet solar daily variation (Sq) and the disturbance daily variation (S_D).
- These variations have well recognized daily cycles and need to be accounted for before the analysis of the disturbed field.
- Preliminary analysis of the GMF variations shows that the principal component analysis (PCA) is a useful tool for extraction of regular variations of GMF; however, the requirements to the data set length, geomagnetic activity level etc. need to be established.
- Here we present preliminary results of the PCA-based Sq and S_D extraction procedure based on the analysis of the Coimbra Geomagnetic Observatory (COI) measurements of the geomagnetic field components H, X, Y and Z between 2007 and 2017.
Data

- H, X, Y and Z components of the geomagnetic field
- Measured at the Coimbra Geomagnetic Observatory (COI), Portugal
  - 40° 13′ N, 8° 25′ W, 99 m asl
- Hourly series
- Only December months from 2007 to 2017

The month-long hourly series of each component was analyzed
- for individual month of each of 11 years
- for all 11 years together

Due to the location of the COI observatory, $H \approx X$ (since $D \approx -4°$)

COI is located near or slightly north to the mean Sq vortex focus position for European sector ($\leq 40°$ N) (e.g., Yamazaki and Maute, 2017)
Methods for Sq & $S_D$ extraction

1. Standard approach using quietest days of a month

2. Principal component analysis (PCA)

- Correlation analysis
  - Similarities between series were analyzed using the correlation coefficients ($r$) and their statistical significances ($p$ value)
  - Statistical significance ($p$ value) was estimated using the Monte Carlo approach with artificial series constructed by the “phase randomization procedure” (Ebisuzaki, 1997).
Method 1: Sq & $S_D$ – standard approach

- “daily quiet” (Sq):
  - calculated as the mean daily variation of the 5 most quiet days of a month
  - international quiet days – IQD, estimated by the GFZ-Potsdam from Kp
  - local quiet days – LQD, estimated from the local K-index
  - ionospheric origin
  - Source: electric current vortex in the sunlit hemisphere
  - contamination from magnetospheric currents (mostly in polar regions)

- “daily disturbed” ($S_D$):
  - calculated as the mean daily variation of all days of the month (S) minus Sq
  - the name comes from the similarity of shapes of the $S_D$ and Dst variations
  - magnetospheric origin
Method 1:
Problems of the standard approach

- IQD are days that are only relatively quiet comparing to others days of a month
- They can be disturbed on the absolute scale
- Final IQD definition is lagged by 1-2 yr
- Observations for certain IQD day at a particular observatory can be missing
- There is a single curve for all days of a month without accounting for variability
  - in the ionosphere and magnetosphere,
  - for the position of the Sq-generating vortex
  - for the shape of the Sq-generating vortex
- A number of studies (Xu and Kamide, 2004; Chen et al., 2007; Yamazaki et al., 2016) showed the need for methods of Sq (and $S_D$) extraction which take into account day-to-day variability of the ionospheric conditions.
Sq “ideal” shape for a mid-latitudinal station (north of the Sq vortex focus)

adapted from
Chapman and Bartels (1940)
COI data: $S_{IQD}$ – individual December series
COI data: $S_{IQD}$ – individual December series

- $H \approx X$ components
  - Mean $Sq$ is far from the “ideal $Sq$” for a station located north of the $Sq$ vortex focus, i.e.
    - either there is contamination by disturbances
    - or for most of these IQD days COI was located near the $Sq$ vortex centre
- High months-to-month variability of the $S_{IQD}$ shape:
  - the shapes of $S_{IQD}$ for December of 2010, 2011, 2014, 2015 are similar to the “ideal $Sq$”
  - the shapes of $S_{IQD}$ for December of 2008, 2012, 2013, 2017 are close to the “ideal $Sq$”
  - the shapes of $S_{IQD}$ for December of 2007, 2009, 2016 are strongly affected by disturbances/$Sq$ vortex shape and position
- $Y$, $Z$ components
  - Both mean $Sq$ and $Sq$ for individual months are similar to the “ideal $Sq$”
  - Low month-to-month variability of the $S_{IQD}$ shape
\( S_D \) “ideal” shape for a mid-latitudinal station

adapted from
\footnotesize{Chapman and Bartels (1940)}
COI data: $S_D$ – individual December series
COI data: $S_D$ – individual December series

- $H \approx X$ components
  - Mean $S_D$ is similar to the “ideal $S_D$”
  - The shapes of $S_D$ for individual months can deviate from the “ideal $S_D$”, sometimes significantly (e.g., December 2007)
  - High month-to-month variability of the $S_D$ shape

- $Y, Z$ components
  - Mean $S_D$ are similar to the “ideal $S_D$”
  - The shapes of $S_D$ for individual months can deviate from the “ideal $S_D$” shape
  - Moderate month-to-month variability of the $S_D$ shape
Method 2: Principal components analysis (PCA)

- Previous studies show (Xu and Kamide, 2004; Chen et al., 2007) that the principal component analysis (PCA) is a useful tool for the extraction of regular variations of GMF.

- PCA is a widely used method to extract independent modes of variability when a number of series of the same parameter of, e.g., different stations or days is used.
Principal components analysis (PCA)

- Input data ⇒ covariance matrix ⇒ eigenvalues & eigenvectors.
- Eigenvalues ⇒ explained variances of the extracted modes
- Eigenvectors ⇒ principal component (PC) & empirical orthogonal function (EOF).
- PCs = daily variations of different types
- EOFs = amplitudes of daily variations (PCs) for each of the analyzed days
- PC# & EOF# ⇒ mode#

PCA input matrix for COI data:
- each column contains 24 observations (every 1 h)
- number of columns:
  - 31 for an individual December (PCA for an individual month)
PCA results:

Sq

all Decembers 2007-2017

- Each of the following plots shows
  - $S_{IQD}$ calculated for each of 11 Decembers – colored thin lines
  - $S_{IQD}$ calculated for December of all 11 years – black thick line
  - $S_{PCA}$: PC1 (Y & Z) and PC2 (H & X) obtained for the whole data set (11 years) – blue and red thick lines, respectively
COI PCA: $S_{\text{IQD}}$ vs $S_{\text{PCA}}$ – all Decembers 2007-2017

$S_{\text{H}}$, 2007-2017, IQD vs H.PC2

$S_{\text{X}}$, 2007-2017, IQD vs X.PC2

$S_{\text{Y}}$, 2007-2017, IQD vs Y.PC1

$S_{\text{Z}}$, 2007-2017, IQD vs Z.PC1
## PCA results: explained variances

all Decembers 2007-2017

<table>
<thead>
<tr>
<th>Components</th>
<th>PC1</th>
<th>Identified as...</th>
<th>PC2</th>
<th>Identified as...</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>54%</td>
<td>$S_D$</td>
<td>18%</td>
<td>Sq</td>
</tr>
<tr>
<td>X</td>
<td>54%</td>
<td>$S_D$</td>
<td>19%</td>
<td>Sq</td>
</tr>
<tr>
<td>Y</td>
<td>67%</td>
<td>Sq</td>
<td>12%</td>
<td>$S_D$ ?</td>
</tr>
<tr>
<td>Z</td>
<td>71%</td>
<td>Sq</td>
<td>10%</td>
<td>$S_D$ ?</td>
</tr>
</tbody>
</table>
COI PCA: $\text{Sq}_{\text{PCA}}$ - all Decembers 2007-2017

- H ≈ X components
  - $\text{Sq}_{\text{PCA}}$ is identified as PC2 and is similar to the “ideal Sq” without notable contamination by disturbances
  - $\text{Sq}_{\text{PCA}} \neq \text{Sq}_{\text{IQD}}$

- Y, Z components
  - $\text{Sq}_{\text{PCA}}$ is identified as PC1 and is similar to the “ideal Sq” without notable contamination by disturbances
  - $\text{Sq}_{\text{PCA}} = \text{Sq}_{\text{IQD}}$
PCA results:

\( S_D \)

all Decembers 2007-2017

Each of the following plots shows

- \( S_D^{IQD} \) calculated for each of 11 Decembers – colored thin lines
- \( S_D^{IQD} \) calculated for December of all 11 years – black thick line
- \( S_D^{PCA} \): PC2 (Y & Z) and PC1 (H & X) obtained for the whole data set (11 years) – red and blue thick lines, respectively
COI PCA : \( S_D \) vs \( S_D \) PCA - all Decembers 2007-2017

EGU2020                                                                D1158: Morozova et al.                                                        EMRP2.3 /ST4
COI PCA : $S_D^{\text{PCA}}$ - all Decemers 2007-2017

- **H \approx X components**
  - $S_D^{\text{PCA}}$ is identified as PC1 and is similar to the “ideal $S_D$”
  - $S_D^{\text{PCA}} \approx S_D$

- **Y, Z components**
  - $S_D^{\text{PCA}}$ is identified as PC2 and is similar to the “ideal $S_D$”
  - $S_D^{\text{PCA}}$ is similar to $S_D$
To test the effect of the data set length on the quality of the PCA-based method of the Sq extraction we applied PCA to the 1-month long data sets of 11 individual Decembers (H component only).

Each of the following plots shows:
- $\text{Sq}_{\text{IQD}}$ calculated for December of this year – black thick line
- PC2 obtained on the whole data set (11 years) – red thick line
- PC1 and PC2 obtained for this particular month – blue and red dashed lines
COI PCA: $S_{\text{IQD}} - S_{\text{PCA}}$

individual Decembers 2007-2012 vs all Decembers
COI PCA: $S_{IQD}$ vs $S_{PCA}$

Individual Decembers 2013-2017 vs all Decembers
### PCA results: explained variances

**individual Decembers from 2007 to 2017**

<table>
<thead>
<tr>
<th>Time interval</th>
<th>PC1</th>
<th>Identified as...</th>
<th>PC2</th>
<th>Identified as...</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2007</td>
<td>47%</td>
<td>$S_D$ ?</td>
<td>21%</td>
<td>?</td>
</tr>
<tr>
<td>December 2008</td>
<td>49%</td>
<td>$S_D$</td>
<td>17%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>December 2009</td>
<td>39%</td>
<td>$Sq$ ??</td>
<td>28%</td>
<td>?</td>
</tr>
<tr>
<td>December 2010</td>
<td>56%</td>
<td>$S_D$</td>
<td>21%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>December 2011</td>
<td>54%</td>
<td>$Sq$</td>
<td>17%</td>
<td>$S_D$ ?</td>
</tr>
<tr>
<td>December 2012</td>
<td>54%</td>
<td>$S_D$</td>
<td>22%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>December 2013</td>
<td>57%</td>
<td>$S_D$</td>
<td>23%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>December 2014</td>
<td>46%</td>
<td>$S_D$</td>
<td>25%</td>
<td>$Sq$ ??</td>
</tr>
<tr>
<td>December 2015</td>
<td>78%</td>
<td>$S_D$ ?</td>
<td>11%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>December 2016</td>
<td>59%</td>
<td>?</td>
<td>12%</td>
<td>?</td>
</tr>
<tr>
<td>December 2017</td>
<td>55%</td>
<td>$S_D$ ?</td>
<td>18%</td>
<td>$Sq$</td>
</tr>
<tr>
<td>Decembers 2007-2017</td>
<td>54%</td>
<td>$S_D$</td>
<td>18%</td>
<td>$Sq$</td>
</tr>
</tbody>
</table>
For 9 out of 11 analyzed individual months PCA extract daily variation that can be identified as Sq
- For 7 out of 11 analyzed months $S_{q,PCA}$ is identified as PC2
- For 2 out of 11 analyzed months $S_{q,PCA}$ is identified as PC1

For 9 out of 11 analyzed individual months PCA extract daily variation that can be identified as $S_D$
- For 8 out of 11 analyzed months $S_{D,PCA}$ is identified as PC1
- For 1 out of 11 analyzed months $S_{D,PCA}$ is identified as PC2
To compare IQD-based and PCA-based Sq curves for individual Decembers and for the whole data set we calculated correlation coefficients between:

- $Sq_{\text{IQD}}$ for individual December and PCs obtained for the whole data set ($PC_{2_{11}}$)
- $Sq_{\text{IQD}}$ for an individual December ($PC_{i_{1}}$: $PC_{1_{1}}$ or $PC_{2_{1}}$)
- $PC_{2_{11}}$ and $PC_{i_{1}}$ which is identified as $Sq$

In the following Table the values in parentheses are p-values.

Only p-values < 0.15 are shown.
Correlation coefficients
$S_{\text{IQD}}$ vs $S_{\text{PCA}}$, individual Decembers vs all Decembers

<table>
<thead>
<tr>
<th>Time interval</th>
<th>$S_{\text{IQD}}$ vs PC$<em>{2</em>{ii}}$</th>
<th>$S_{\text{IQD}}$ vs PC$<em>{1</em>{i}}$</th>
<th>i</th>
<th>PC$<em>{2</em>{ii}}$ vs PC$<em>{1</em>{i}}$</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2007</td>
<td>0.21</td>
<td>0.72 (0.08)</td>
<td>2</td>
<td>0.60</td>
<td>2</td>
</tr>
<tr>
<td>December 2008</td>
<td><strong>0.80 (0.04)</strong></td>
<td><strong>0.65 (0.003)</strong> / 0.56</td>
<td>1 / 2</td>
<td><strong>0.87 (0.04)</strong></td>
<td>2</td>
</tr>
<tr>
<td>December 2009</td>
<td>0.37</td>
<td>0.52</td>
<td>2</td>
<td>0.67</td>
<td>1</td>
</tr>
<tr>
<td>December 2010</td>
<td>0.39</td>
<td>0.79 (0.002)</td>
<td>2</td>
<td>0.64</td>
<td>2</td>
</tr>
<tr>
<td>December 2011</td>
<td><strong>0.80 (0.07)</strong></td>
<td>0.90 (0.005)</td>
<td>1</td>
<td>0.76 (0.14)</td>
<td>1</td>
</tr>
<tr>
<td>December 2012</td>
<td>0.24</td>
<td>0.72 (0.05)</td>
<td>1</td>
<td>0.94 (0.02)</td>
<td>2</td>
</tr>
<tr>
<td>December 2013</td>
<td><strong>0.67 (0.12)</strong></td>
<td>0.52</td>
<td>2</td>
<td>0.90 (0.05)</td>
<td>2</td>
</tr>
<tr>
<td>December 2014</td>
<td><strong>0.90 (0.03)</strong></td>
<td><strong>0.69 (0.07)</strong> / 0.54</td>
<td>1 / 2</td>
<td><strong>0.83 (0.04)</strong></td>
<td>2</td>
</tr>
<tr>
<td>December 2015</td>
<td>0.57</td>
<td>0.47</td>
<td>1 / 2</td>
<td>0.91 (0.002)</td>
<td>2</td>
</tr>
<tr>
<td>December 2016</td>
<td>0.39</td>
<td>0.55</td>
<td>2</td>
<td>0.61 (0.04)</td>
<td>2</td>
</tr>
<tr>
<td>December 2017</td>
<td>0.25</td>
<td>0.53 / <strong>0.67</strong></td>
<td>1 / 2</td>
<td>0.80 (0.07)</td>
<td>2</td>
</tr>
</tbody>
</table>
\( \text{Sq}_{\text{IQD}} \) vs \( \text{Sq}_{\text{PCA}} \)

individual Decembers 2013-2017 vs all Decembers

\( \text{Sq}_{\text{IQD}} \) is highly correlated with \( \text{Sq}_{\text{PCA}} \) for those years when its shape is very similar to the “ideal Sq” shape:

- Exceptions: 2010 & 2017 – years when the time of the daily minimum is shifted to the earlier /later hours (respectfully) resulting in low correlation coefficients

For 7 out of 11 analyzed months \( \text{Sq}_{\text{IQD}} \) is highly correlated with \( \text{Sq}_{\text{PCA}} = \text{PC}_{2_1} \) for this particular month

For 9 out of 11 analyzed individual months \( \text{PC}_{2_{11}} \) is highly correlated with \( \text{PC}_{2_1} \)
Conclusions

- Preliminary results show that PCA can be successfully used for extraction of the Sq and $S_D$ variations from the observations of the geomagnetic field.

- We analyzed H, X, Y and Z components for December months measured at the Coimbra Geomagnetic Observatory (COI) from 2007 to 2017.

- The PCA-based Sq and $S_D$ curves were compared with the standard ones obtained using 5 IQD per month.

- PCA was applied to data sets of different length:
  - either 1 month-long data set for one of the analyzed years
  - or data series for the same month but from all years combined together.
Conclusions

For most of the analyzed years

PC1 was identified as

- \( S_D \) variation for H and X components and
- \( S_q \) variations for Y and Z components.

PC2 was identified as

- \( S_q \) variation for H and X components
- \( S_D \) variations for Y and Z components.

The PCA of the longer series (data for the same month but from different years combined together) produces more reliable results.
References

- Ebisuzaki, W. (1997), A method to estimate the statistical significance of a correlation when the data are serially correlated, J. Clim., 10 (9), 2147-2153.
Acknowledgement

- This study is funded by national funds through FCT (Foundation for Science and Technology, I.P.), under the project MAG-GIC: PTDC/CTA -GEO/31744/2017.
- CITEUC is funded by National Funds through FCT - Foundation for Science and Technology (project: UID/MULTI/00611/2019) and FEDER – European Regional Development Fund through COMPETE 2020 – Operational Programme Competitiveness and Internationalization (project: POCI-01-0145-FEDER-006922).