MagPySV: a Python package for processing and denoising geomagnetic observatory data

Grace Cox[1], Will Brown[2], Laurence Billingham[2], & Richard Holme[1]

1. University of Liverpool, UK
2. British Geological Survey, UK

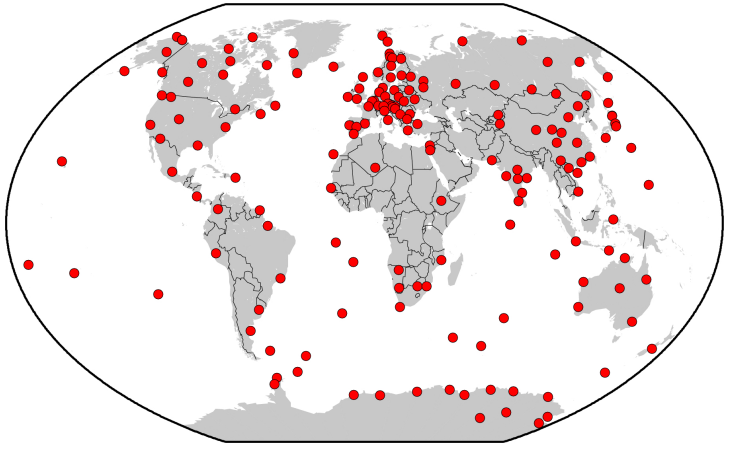
Observations of the recent geomagnetic field

Data collection

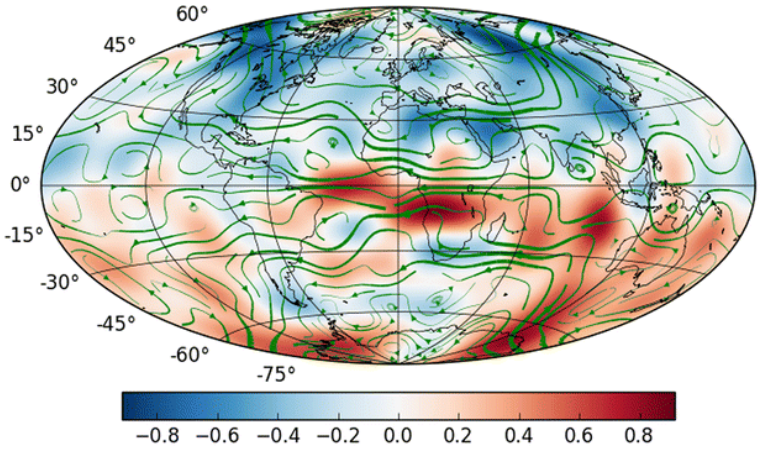
- Ground-based observatories
- Satellites

Common data uses

- Global field models for scientific and industrial purposes
- Time changes of Earth's internally generated magnetic field (secular variation, SV)
- Probing outer core dynamics and properties



Map of currently operating magnetic observatories [BGS].

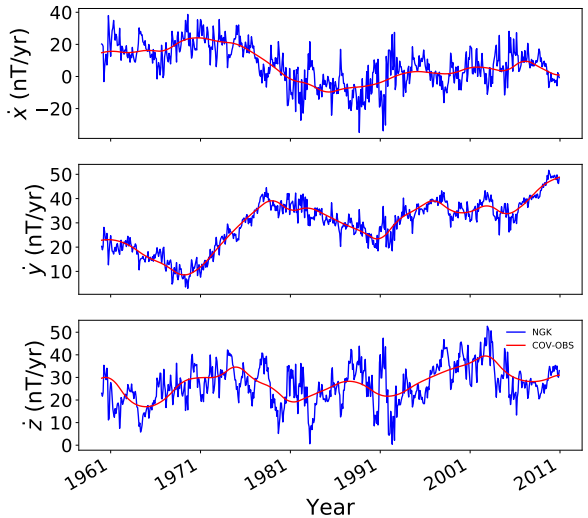


Radial geomagnetic field at the core-mantle-boundary (in mT) according to CHAOS6 model. Green lines are tracers of instantaneous core flow [Finlay et al. (2016), EPS].

Calculating secular variation

Observatory annual means

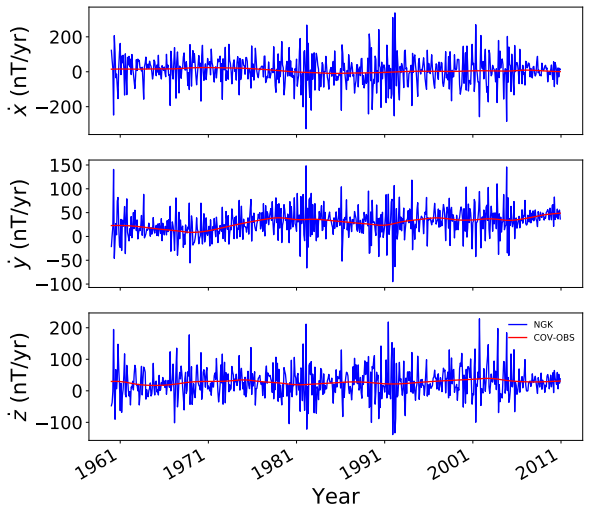
- First differences are widely used, e.g. field modelling
- Averaging reduces effects of external field contamination
- Resolution too low for studying rapid core dynamics and jerks



Annual differences of monthly means at NGK, Germany.

Observatory monthly means

- SV calculated as annual differences or first differences
- Resolution suitable for studying rapid core dynamics and jerks
- Very noisy, significant external field contamination



First differences of monthly means at NGK, Germany.

Available observatory data products

World Data Centre (WDC) at Kyoto:

- Minute means
- Second data

British Geological Survey/WDC at BGS:

- Annual means (baseline data documented, corrections not applied)
- Hourly means (baseline data not included)
- AUX_OBS for Swarm mission (baseline corrections applied)

BCMT Paris:

- Monthly means (baseline data documented, corrections not applied)

Producing SV time series from raw data

Some processing steps:

1. Baseline discontinuity corrections
2. Data resampling/averaging (e.g. low solar activity times)
3. SV calculation
4. Outlier identification and removal
5. Removal of external field contamination ('denoising')

Currently, this processing is:

- Often performed on an ad-hoc basis/by hand
- Often done using piecemeal codes
- Time consuming and not easily reproducible

Producing SV time series from raw data

Some processing steps:

1. Baseline discontinuity corrections
2. Data resampling/averaging (e.g. low solar activity times)
3. SV calculation
4. Outlier identification and removal
5. Removal of external field contamination ('denoising')

Currently, this processing is:

- Often performed on an ad-hoc basis/by hand
- Often done using piecemeal codes
- Time consuming and not easily reproducible

Aim of this work: dedicated open-source tool to process raw WDC data quickly, reproducibly and according to user needs.

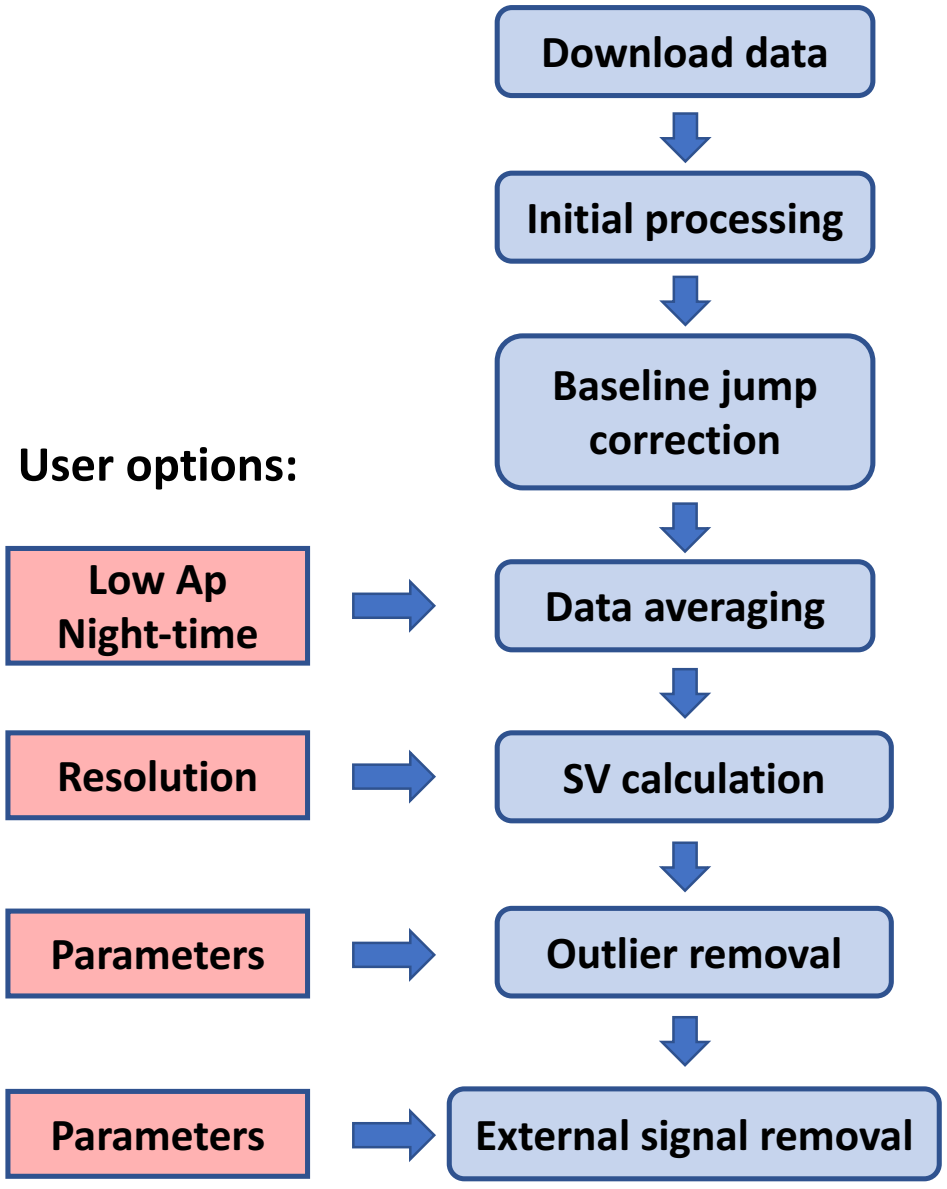
Proposed solution: MagPySV

Open-source Python code designed to:

- Easily create suitable datasets from raw hourly data
- User specifies data (e.g. all hourly, low Ap index, night-time)
- Denoise data, particularly useful for monthly means
- Automate the processing pipeline wherever possible to improve reproducibility and consistency
- Reduce the time and effort required to reproduce, or modify, datasets underlying publications

Preliminary results and application presented here. **Request contributions and suggestions from geomagnetism community.**
Code available at `https://github.com/gracecox/MagPySV`
Documentation at `http://magpysv.readthedocs.io/en/latest/`

Workflow


Download data
Initial processing
Baseline jump correction
User options:
Low Ap Night-time
Data averaging
Resolution
SV calculation
Parameters
Outlier removal
Parameters
External signal removal

Downloading data

Application developed with Laurence Billingham and Will Brown at BGS

Programmatically download data from the BGS WDC online data portal
`http://wdc.bgs.ac.uk/dataportal/`

User specifies:

- List of observatories
- Start and end dates
- File format - WDC (IAGA-2002 planned)
- Data source - hourly WDC (AUX_OBS planned)

Code available at:
`https://github.com/lbillingham/geomag_wdc_web_app_interface`

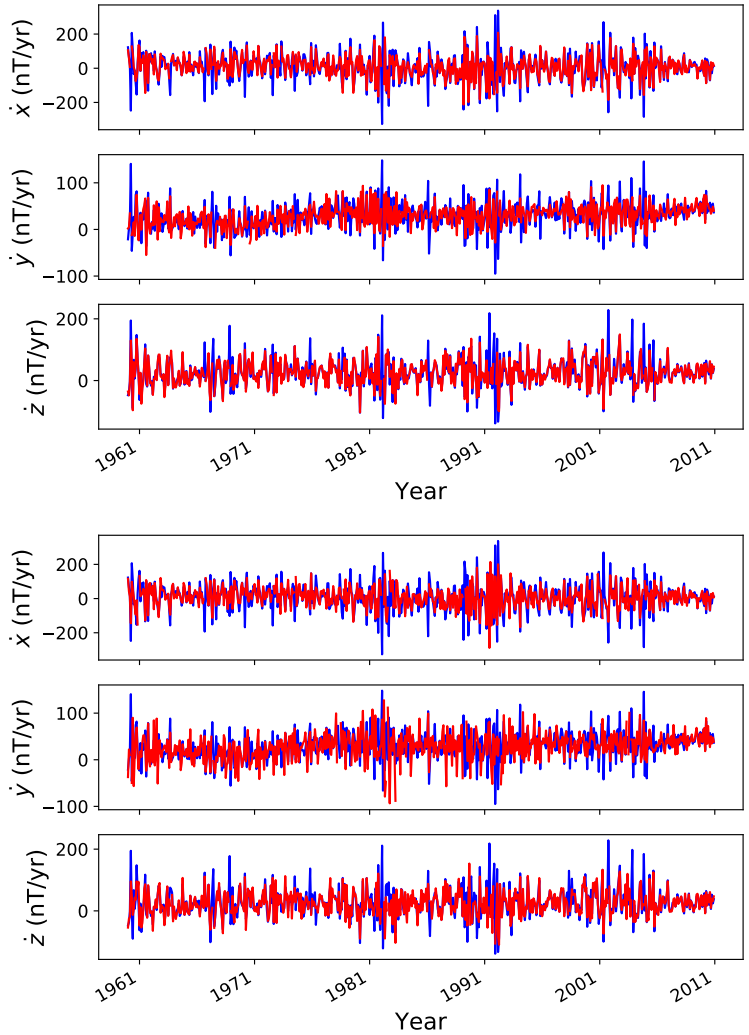
Initial data processing and resampling

Processing

- Parse WDC files
- Convert to X, Y, Z components
- Correct baseline jumps
- Geocentric/geodetic conversion

Resampling/averaging

- All hourly data
- Apply an Ap threshold
- Night time only (planned)



First differences of monthly means at NGK, Germany using all data (blue lines) and with Ap thresholds applied (red lines). Thresholds are 30 (top) and 6 (bottom).

SV calculation and outlier removal

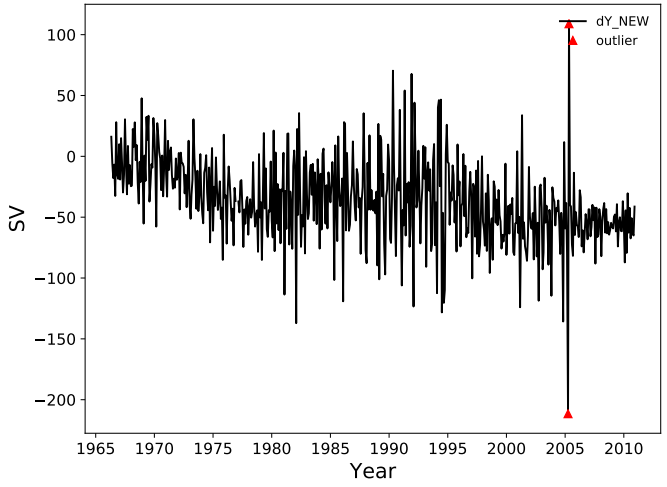
Typical SV studies

- Annual means: first differences
- Monthly means: annual differences
- Monthly means: first differences

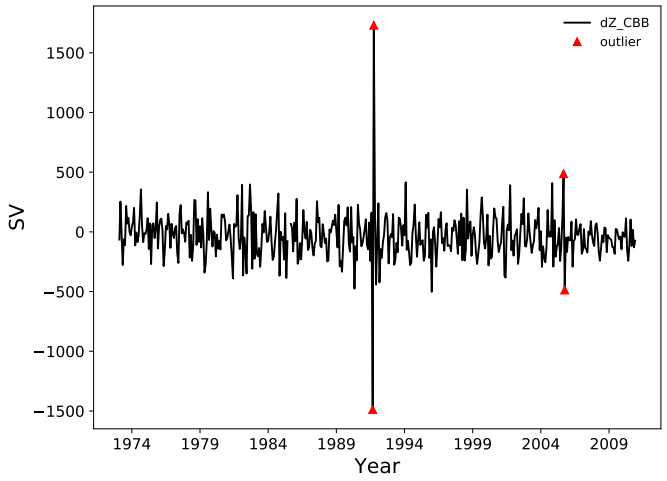
Outlier identification

(signal - median) $>$ threshold $\times$ std

- Use running median and running standard deviation
- Window length and threshold value specified by user
- Outliers are set to NaN



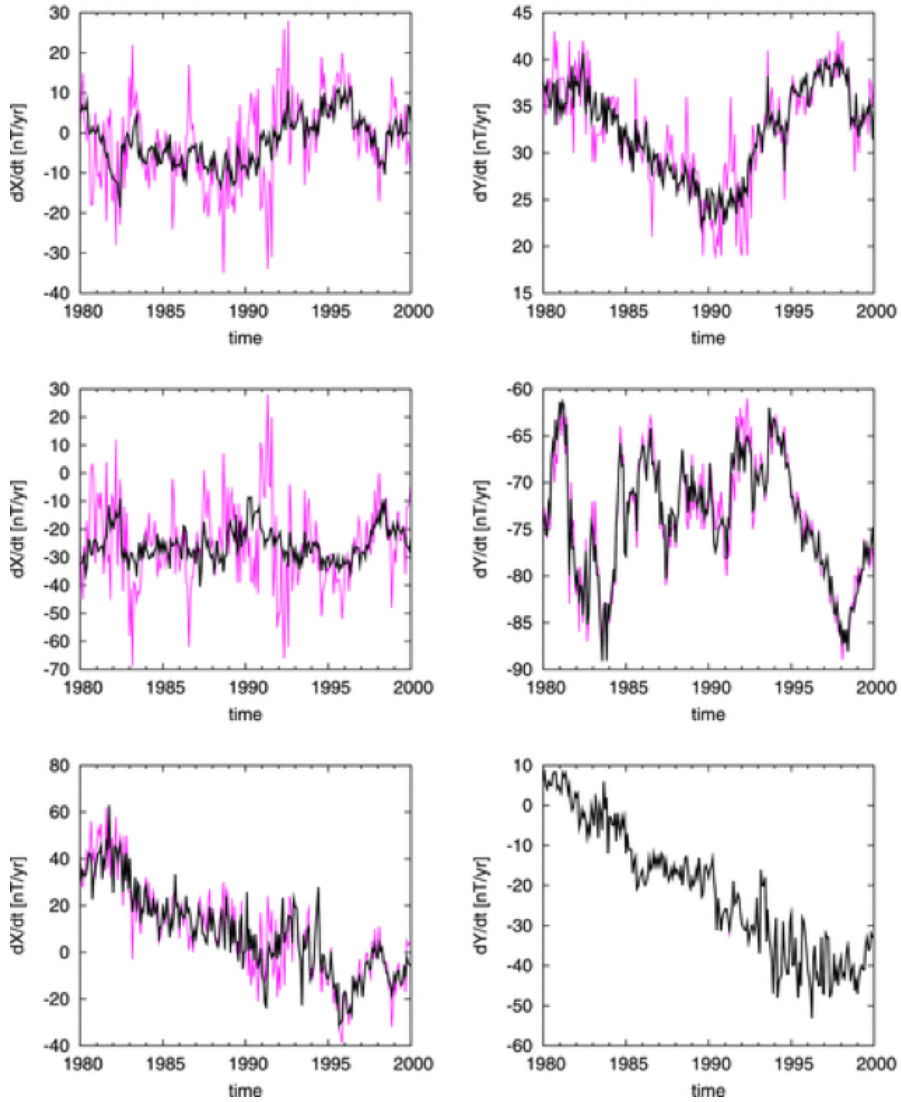
caption



Example time series of first differences of monthly means in nT/yr (black lines) with identified outliers (red triangles).

Current denoising methods

1. Averaging
2. 'Quiet' time data/Ap thresholds
3. Night-time only data
4. Parameterised models
5. Eigensystem method:
   - Calculate residual between data and field model
   - Eigenvalue/vector analysis of residual covariance matrix
   - Residuals in 'noisy' eigendirection give proxy for external signal



Annual differences of monthly means before (pink) and after (black) denoising at NGK, Germany (left) and MBour, Senegal (right), [Wardinski & Holme (2011), GJI].

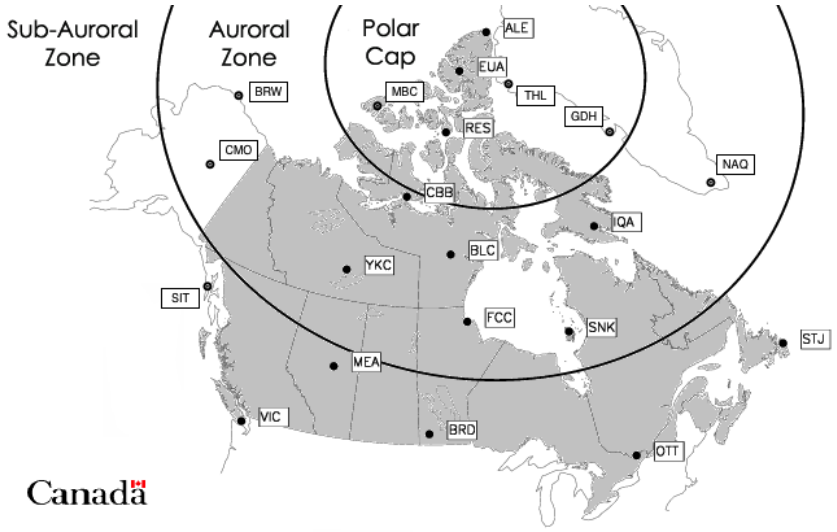
MagPySV denoising: extension to Wardinski & Holme method

Denoised with several observatories simultaneously (user specifies list of IAGA codes)

Case study: high latitude observatories

- Very noisy, influenced by auroral electrojet
- Noise characteristics differ from lower latitudes
- Different noise characteristics in different high latitude zones

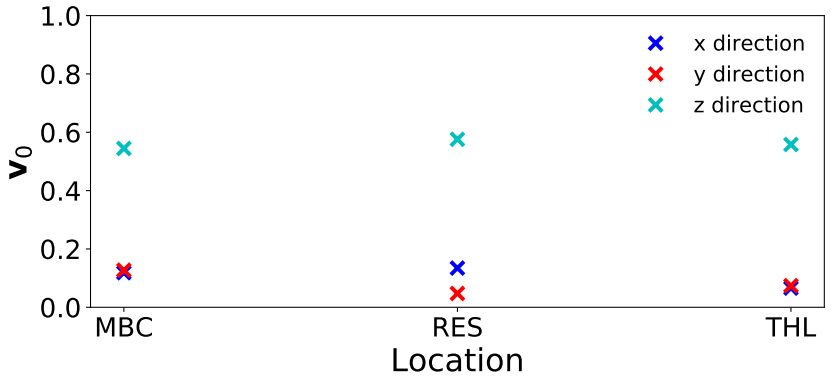
High latitude regions



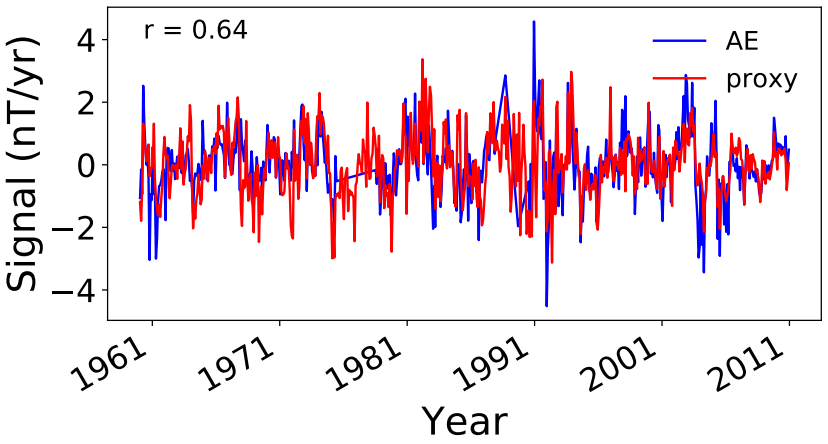
Different zones at high northern latitudes according to location relative to the auroral electroject [Geological Survey of Canada].

Application: clean SV series needed for studies of core dynamics inside the tangent cylinder.

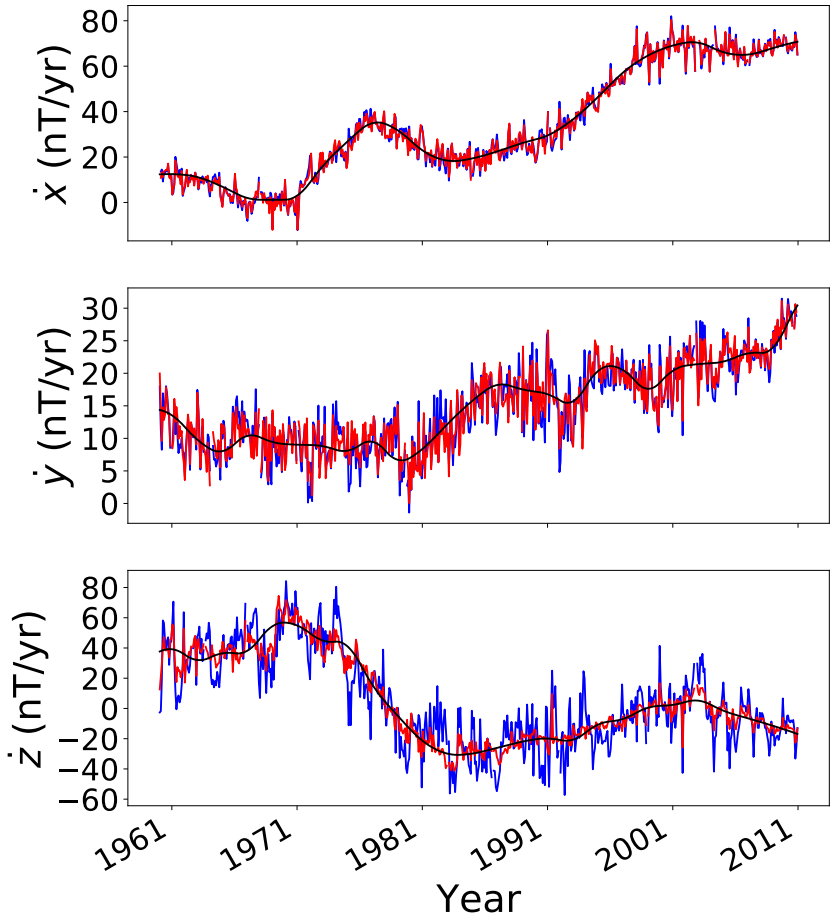
High latitude noise: polar zone



Contributions of the X, Y and Z directions at each observatory to the noisy residuals.



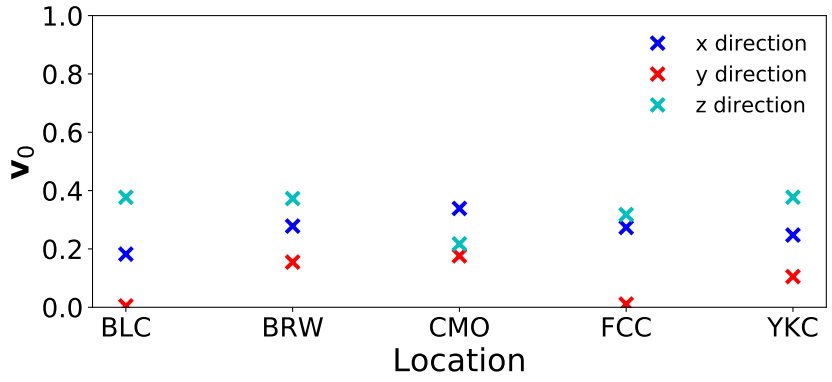
Zero-mean, unit-variance AE index (blue) and proxy for unmodelled external contamination (red).



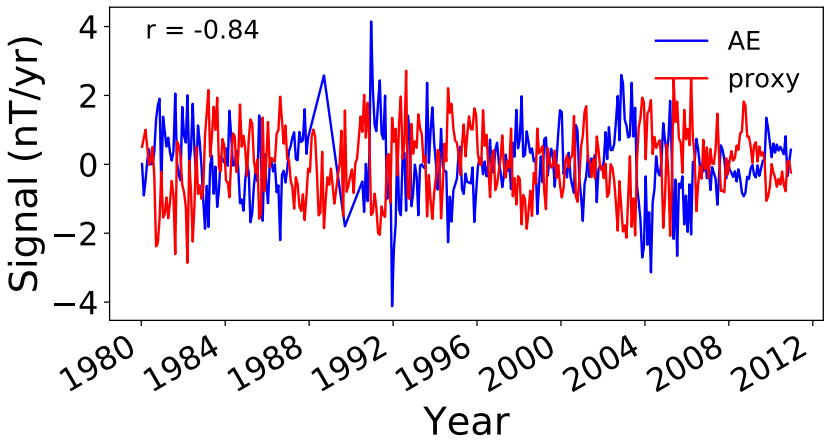
Annual differences of monthly means at Thule, Greenland, before (blue) and after (red) denoising.

Some correlation with the auroral electrojet (AE index) and a weaker anticorrelation with the ring current (Dcx index, r=-0.48).

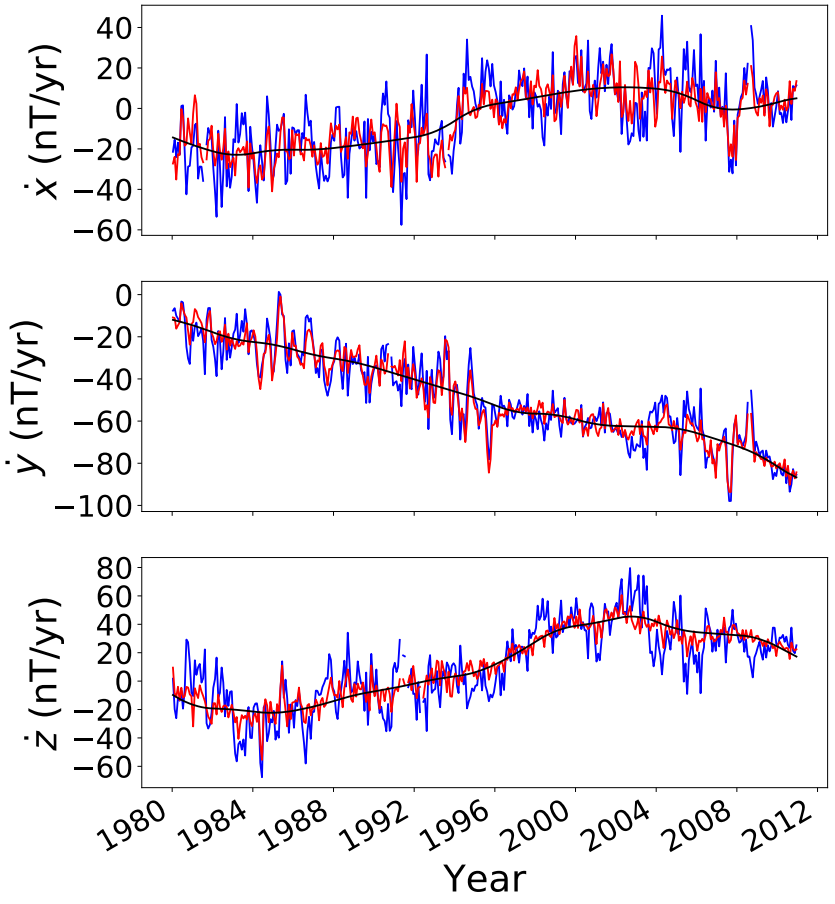
High latitude noise: auroral zone



Contributions of the X, Y and Z directions at each observatory to the noisy residuals.



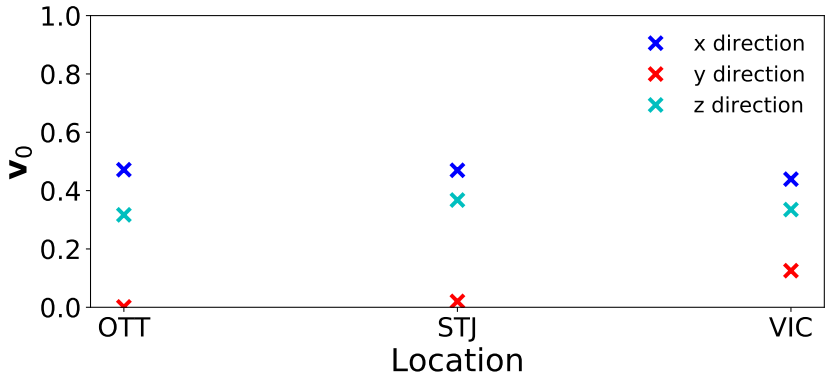
Zero-mean, unit-variance AE index (blue) and proxy for unmodelled external contamination (red).



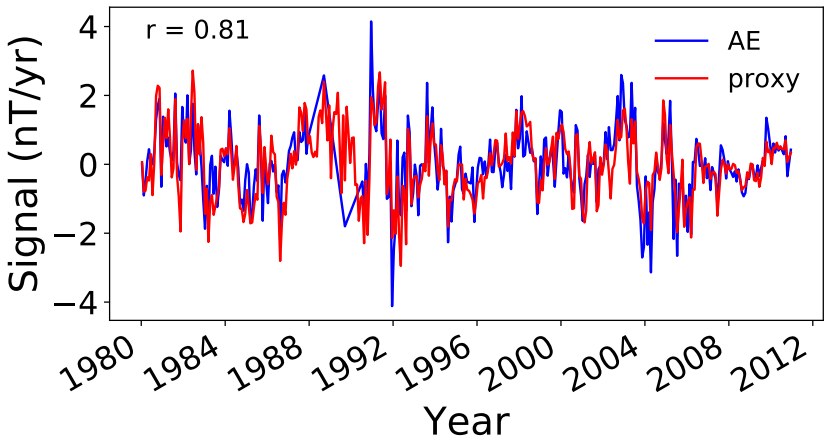
Annual differences of monthly means at Barrow, Alaska, before (blue) and after (red) denoising.

Strong anticorrelation of noise with the auroral electrojet (AE index) and a weaker positive correlation with the Dcx index (r=0.67).

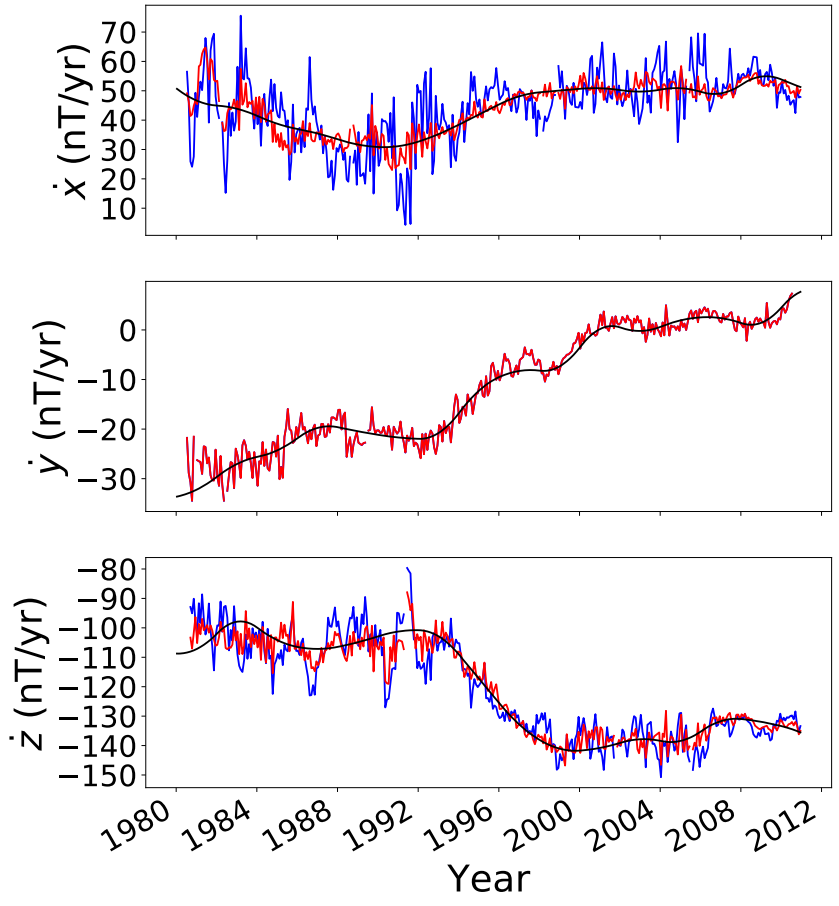
High latitude noise: subauroral zone



Contributions of the X, Y and Z directions at each observatory to the noisy residuals.



Zero-mean, unit-variance AE index (blue) and proxy for unmodelled external contamination (red).



Annual differences of monthly means at Ottawa, Canada, before (blue) and after (red) denoising.

Strong correlation of noise with the auroral electrojet (AE index) and a weaker anticorrelation with the Dcx index (r=-0.72).

Summary

MagPySV:

- An open-source Python tool to download, process and denoise observatory data
- Can be run on the command line or in a Jupyter notebook
- Quickly and reproducibly produces clean SV time series according to user needs

Application presented here: improved resolution data for studying high latitude core dynamics.