



**British  
Geological Survey**  
Expert | Impartial | Innovative

Gateway to the Earth

# Denoising Swarm Geomagnetic Virtual Observatories using Principal Component Analysis

Grace Cox<sup>1</sup>, Will Brown<sup>1</sup>, Ciarán Beggan<sup>1</sup>, Magnus Hammer<sup>2</sup>, Chris Finlay<sup>2</sup>

1. British Geological Survey, UK

2. DTU Space, Technical University of Denmark

# Geomagnetic Virtual Observatories (GVO)

- Point estimates of Earth's core field for a specified location at satellite altitude
- Can be used to infill large gaps left by uneven ground observatory coverage
- Typically use data selection criteria and parameterised models to reduce external magnetic field contamination

## Issues with current GVO data sets:

- External signals remain in the data
- Common selection criteria use <5% of available data
- Monthly GVO have a local time sampling bias of approximately 4 months due to orbital dynamics
- Averaging this bias out reduces temporal resolution

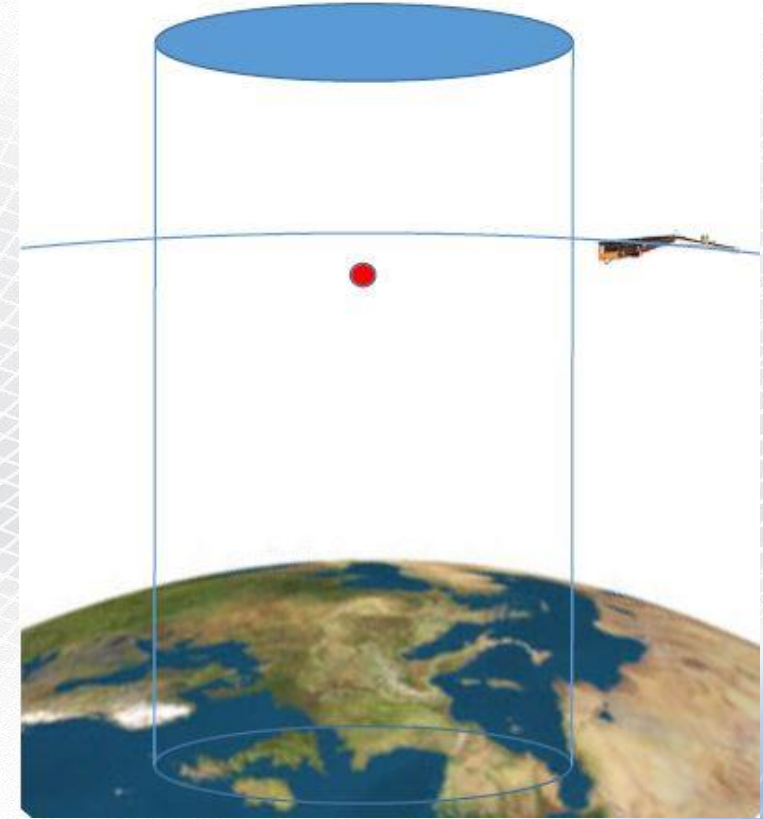


Figure from M. Hammer

# Swarm DISC GVO project

## Project aims:

- Provide the scientific community with new GVO data sets and uncertainties
- Use an alternative method to remove contaminating signals without reducing resolution

## Project pipeline:

- Produce monthly and four-monthly GVO from Swarm vector magnetic measurements and estimate their uncertainties (see [EGU2020-13913](#) by M. Hammer)
- Apply Principal Component Analysis to monthly GVO to remove external magnetic fields and local time biases
- Separate internal/external contributions using Spherical Harmonic Analysis
- Release GVO data products for the core magnetic field

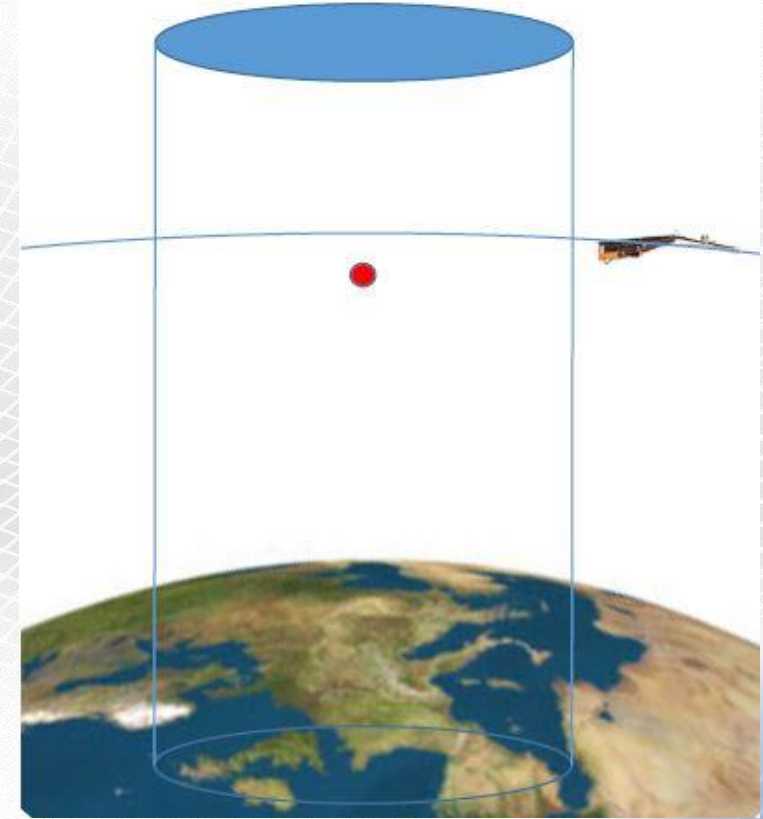
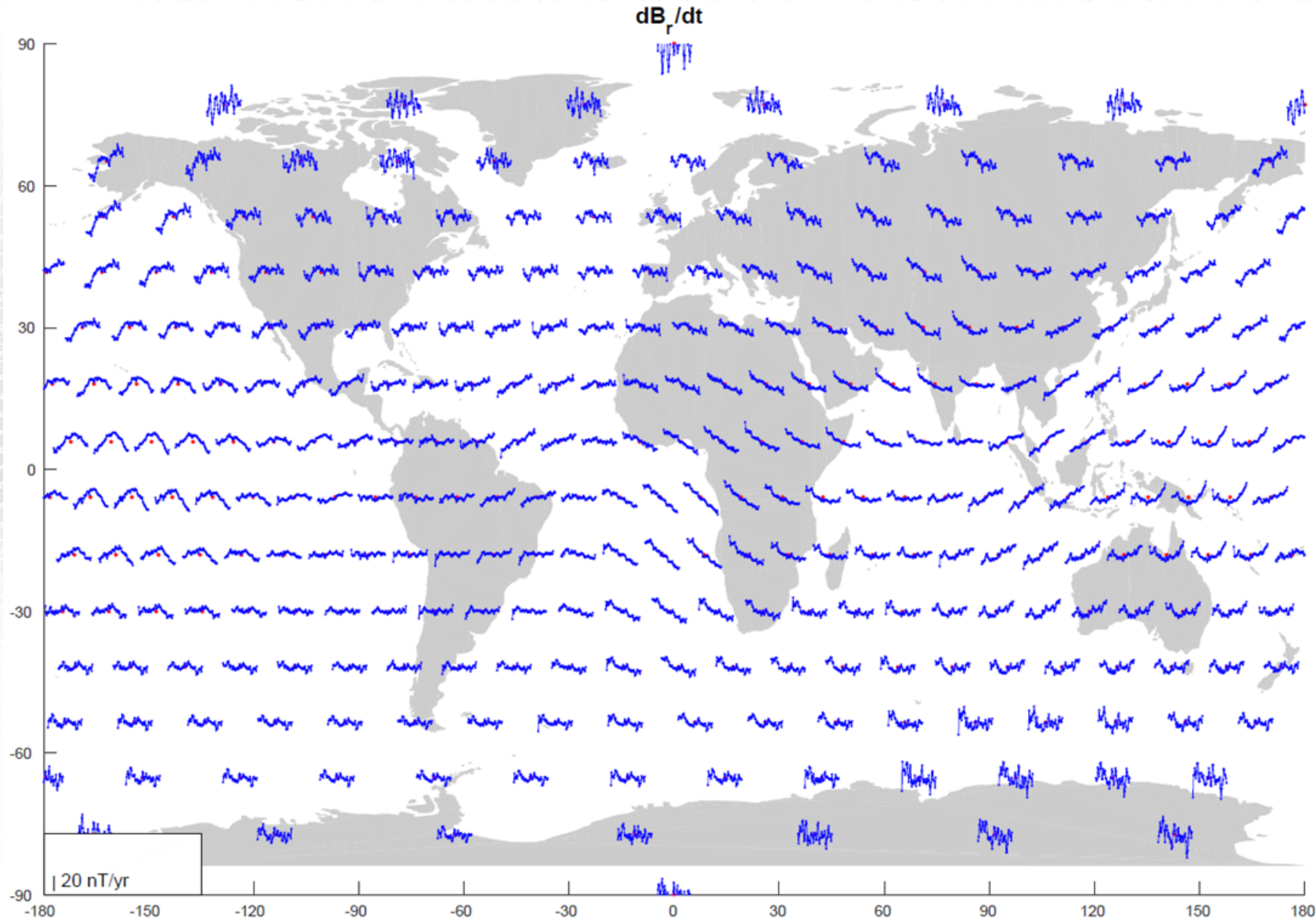


Figure from M. Hammer

# Monthly GVO data set: Radial SV



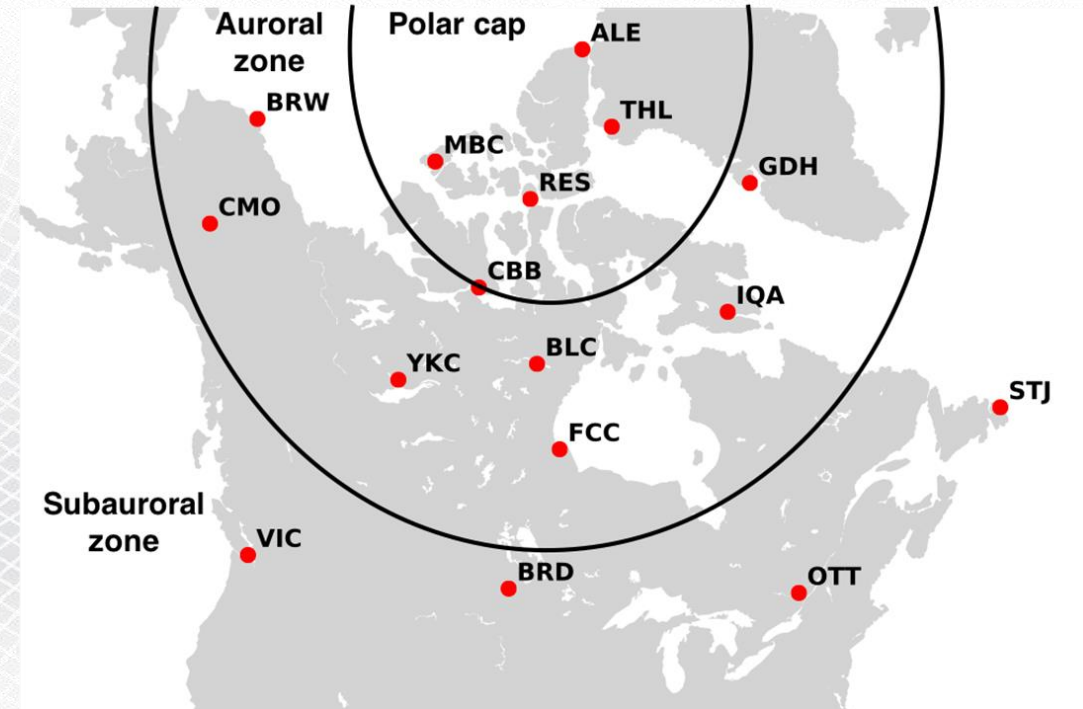
Time series of radial SV at each of the 300 GVO locations in the monthly data set during the Swarm mission.

Interesting features:

1. Noise at high latitudes
2. Geomagnetic jerk in the Pacific region – hard to see in ground data due to lack of coverage

# Principal Component Analysis (PCA)

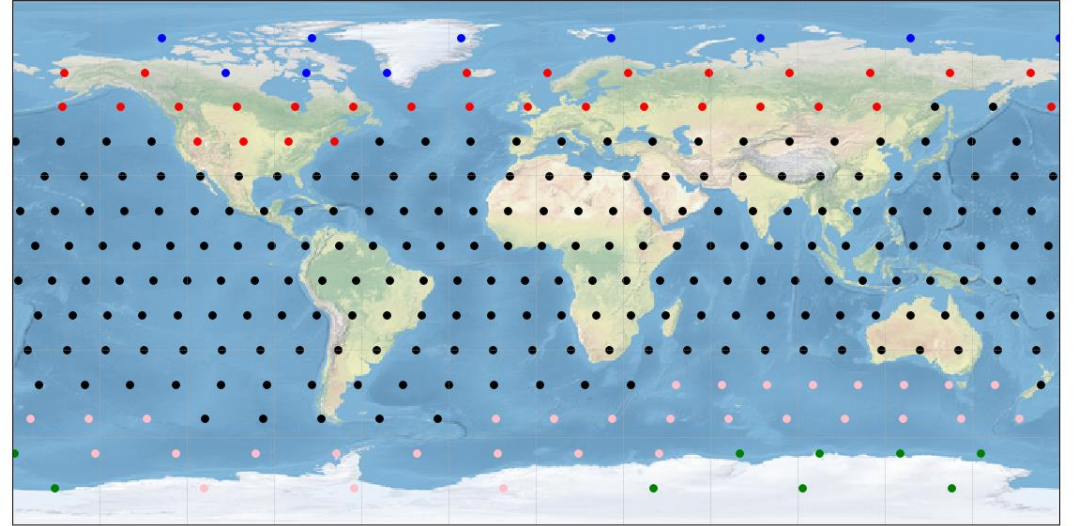
- Uses an open-source code for denoising geomagnetic data (Cox et al. 2018, G<sup>3</sup>)
- Calculate residuals between SV data and an internal field model for  $N$  GVO
- PCA of residual covariance matrix for all GVO considered ( $3N \times 3N$ )
- Residuals projected into the dominant principal components (PCs), and their Discrete Fourier Transforms (DFTs), provide information about:
  1. Unmodelled external magnetic fields
  2. Local time sampling biases
- This method has previously been applied to ground data and CHAMP GVO



Cox et al. (2018, G<sup>3</sup>) applied PCA to ground data and found that the method is most effective when performed on groups of observatories at similar magnetic latitude

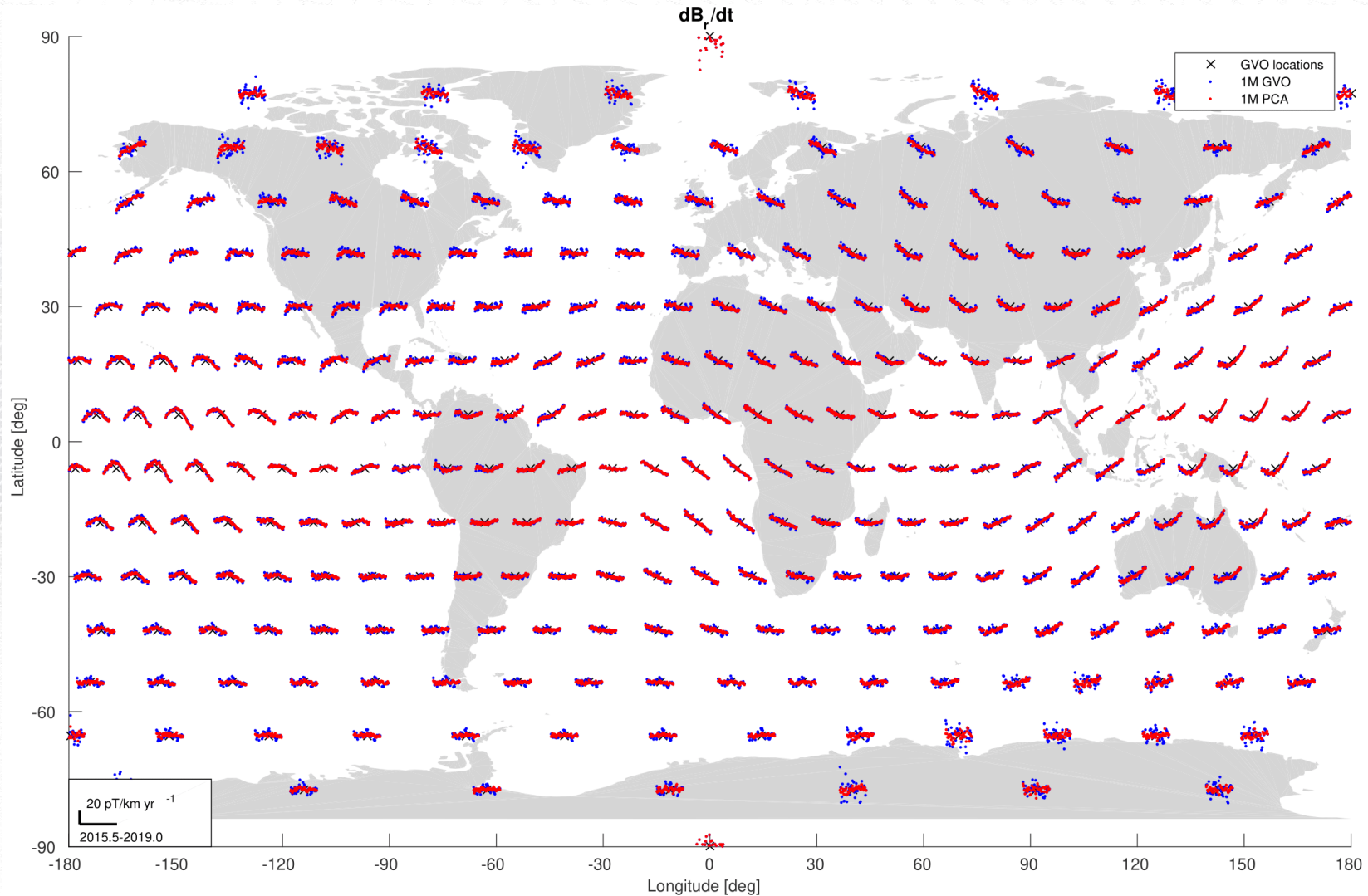
# PCA overview

- Set of 300 GVO split into five magnetic latitude regions (see table and figure)
- All GVO within each region denoised simultaneously
- Denoising involves identifying the content of the dominant PCs and removing those related to external fields or sampling biases
- The geographic poles were excluded



Region	Magnetic Latitude, $\lambda_m$	# PCs removed	Total percentage of variance
Polar North	$70^\circ \leq \lambda_m < 90^\circ$	4	34.8%, 30.1%, 14.5%, 5.7% = <b>85.1%</b>
Auroral North	$50^\circ \leq \lambda_m < 70^\circ$	3	69.2%, 7.8%, 6.3% = <b>83.3%</b>
Low-mid Latitudes	$-50^\circ \leq \lambda_m < 50^\circ$	4	70.8%, 9.4%, 5.1%, 3.3% = <b>88.6%</b>
Auroral South	$-70^\circ \leq \lambda_m < -50^\circ$	3	62.9%, 10.4%, 6.0% = <b>79.3%</b>
Polar South	$-90^\circ \leq \lambda_m < -70^\circ$	4	31.7%, 22.8%, 15.0%, 13.3% = <b>82.8%</b>

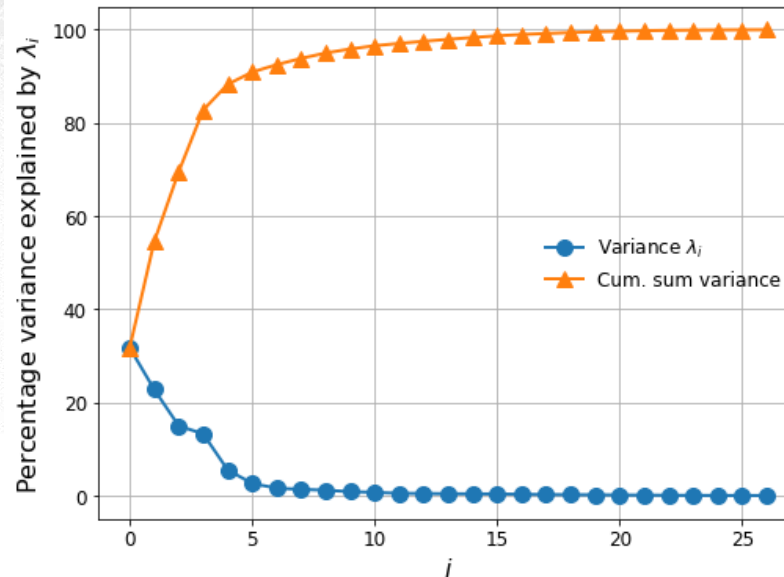
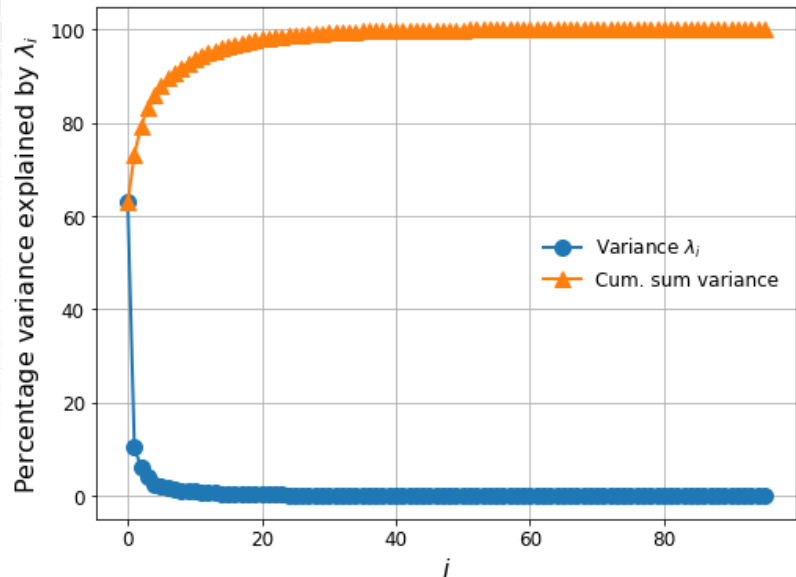
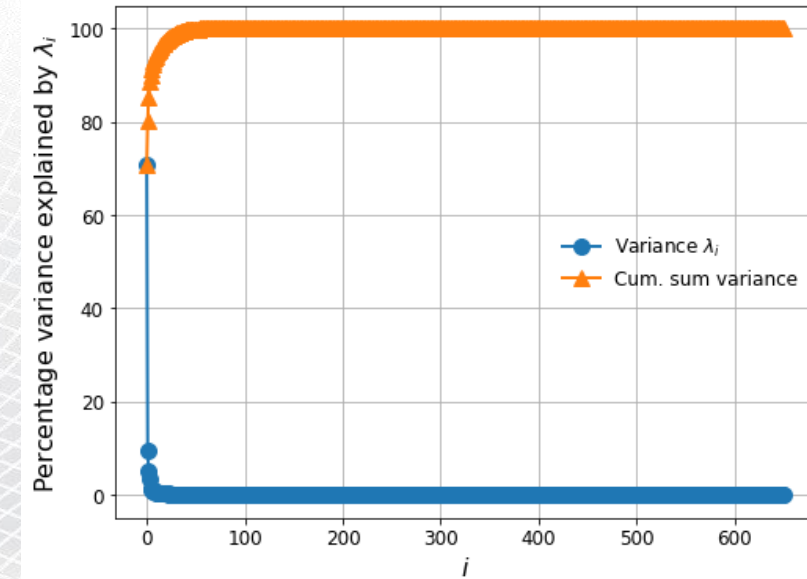
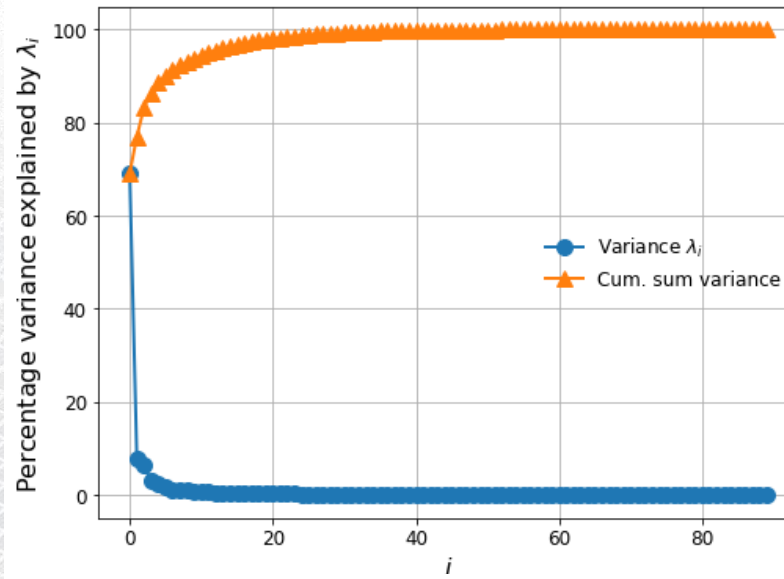
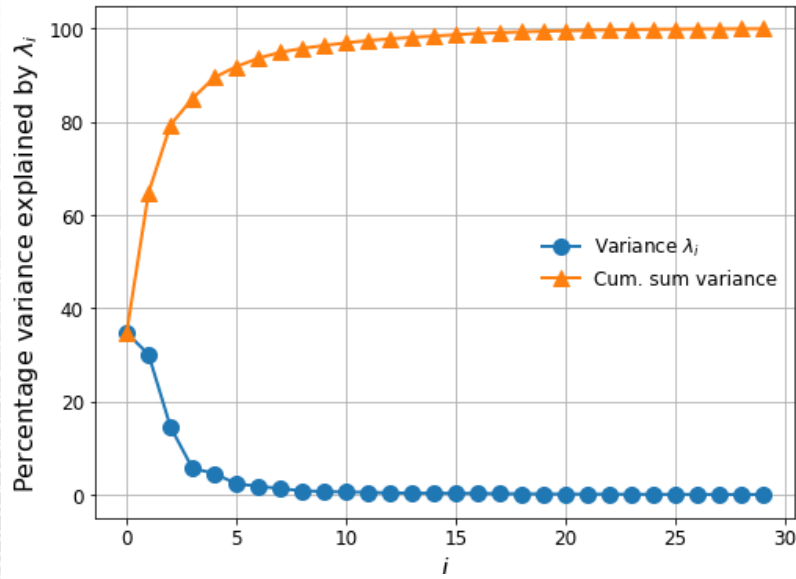
# Radial SV before (blue) and after (red) PCA



Time series of radial SV at each of the 300 GVO locations in the monthly data set during the Swarm mission.

Note that the geographic poles were not denoised. Details of PCA in each magnetic latitude region are on the following slides

# Eigenvalue spectra: the first few PCs account for ~90% of the variance in the residuals

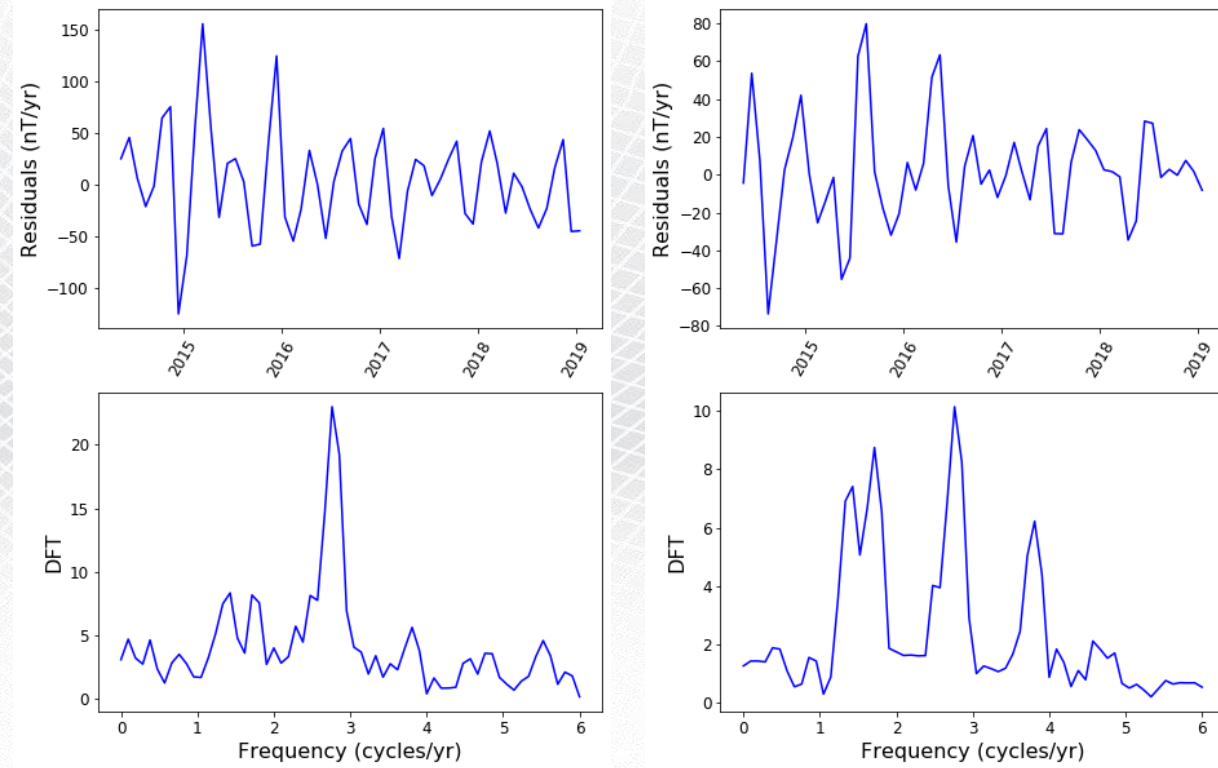


Top row (left to right): Polar North, Auroral North, Low-to mid-latitudes  
Bottom row: Auroral South, Polar South



# Low-mid magnetic latitudes

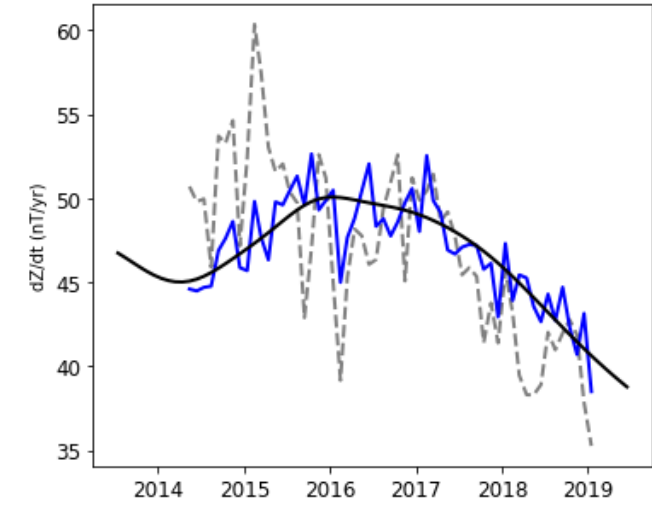
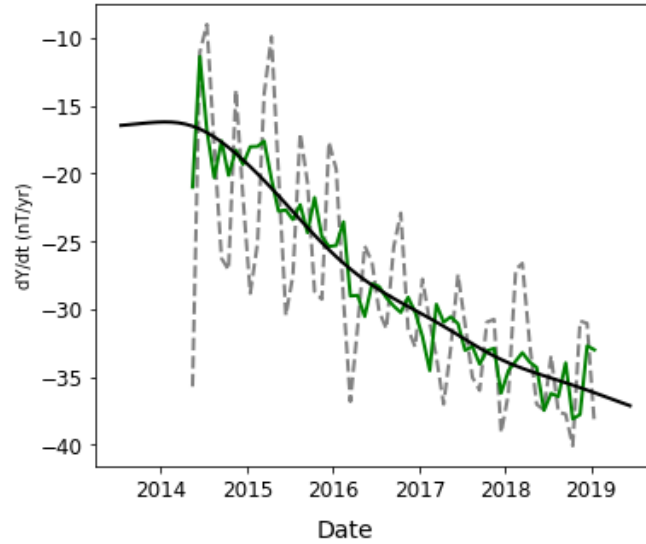
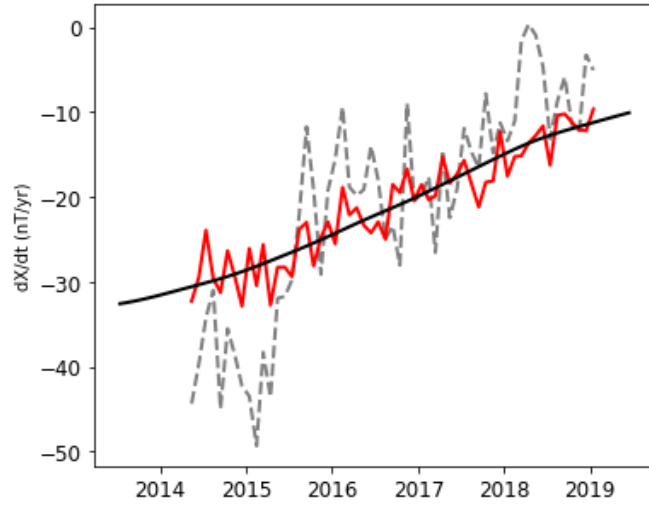
- PC0: identified as an external field (magnetosphere) based on correlations with time derivatives of Em (0.84), AE (0.81) and Dst (0.70)
- Predominantly in the X component, consistent across all GVO
- PC1-PC3: No strong correlations to magnetic indices
- Local time (LT) sampling effects, based on peaks in DFTs at ~4.2 months (LT sampling bias period)



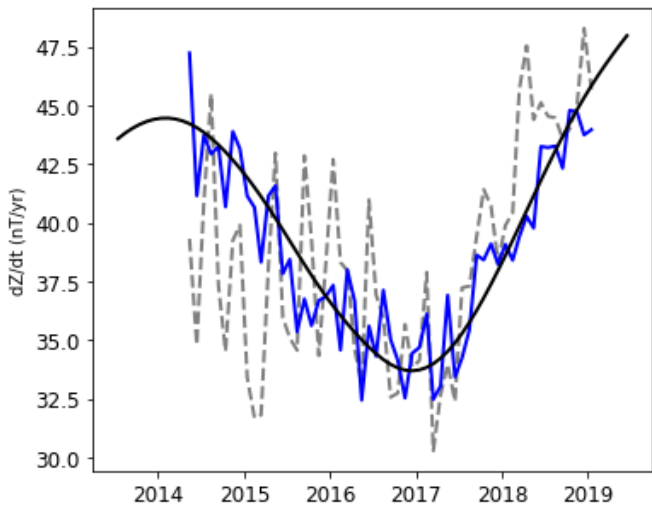
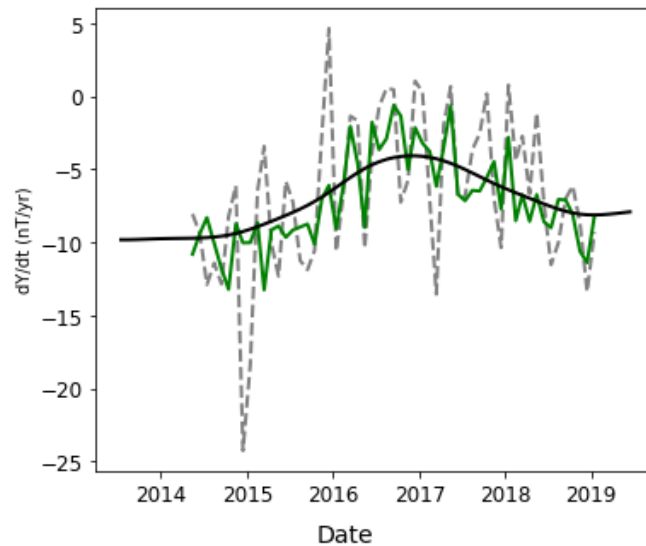
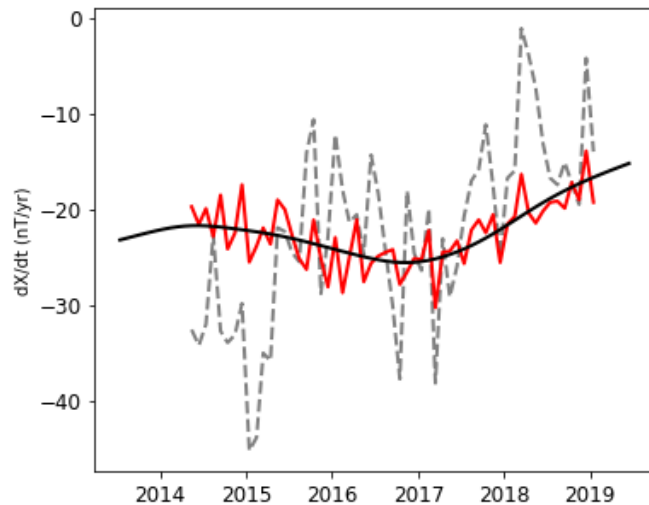
PC1 (left) and PC3 (right): Top: Residuals projected onto PCs; bottom: DFTs of projected residuals (peaks at ~4.2 months)

# Low-mid latitudes: SV examples pre- (grey) and post-PCA (coloured lines). Black line: CHAOS-7.

GVO 47



GVO 228



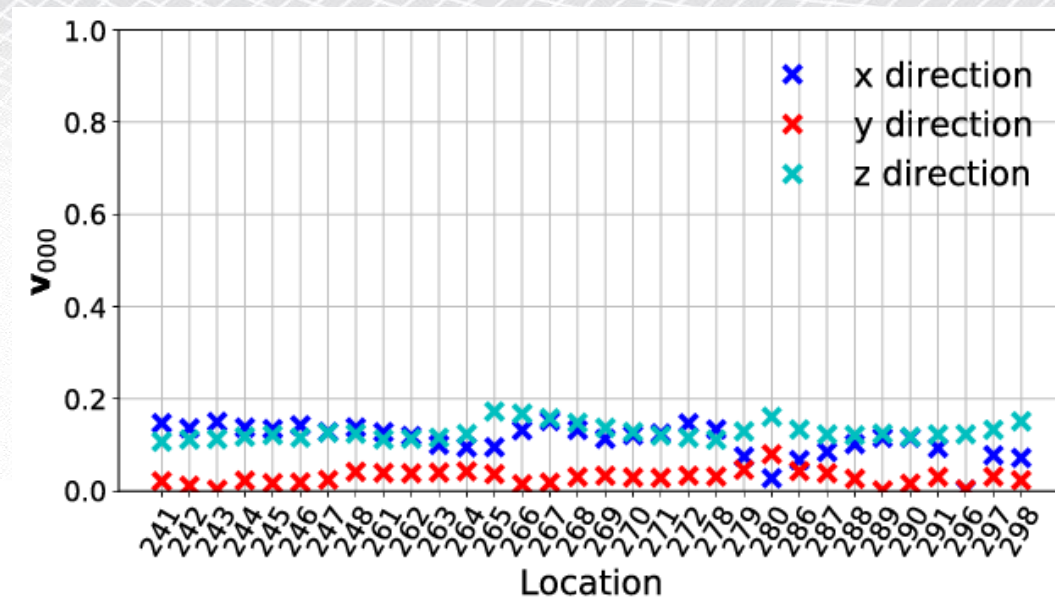
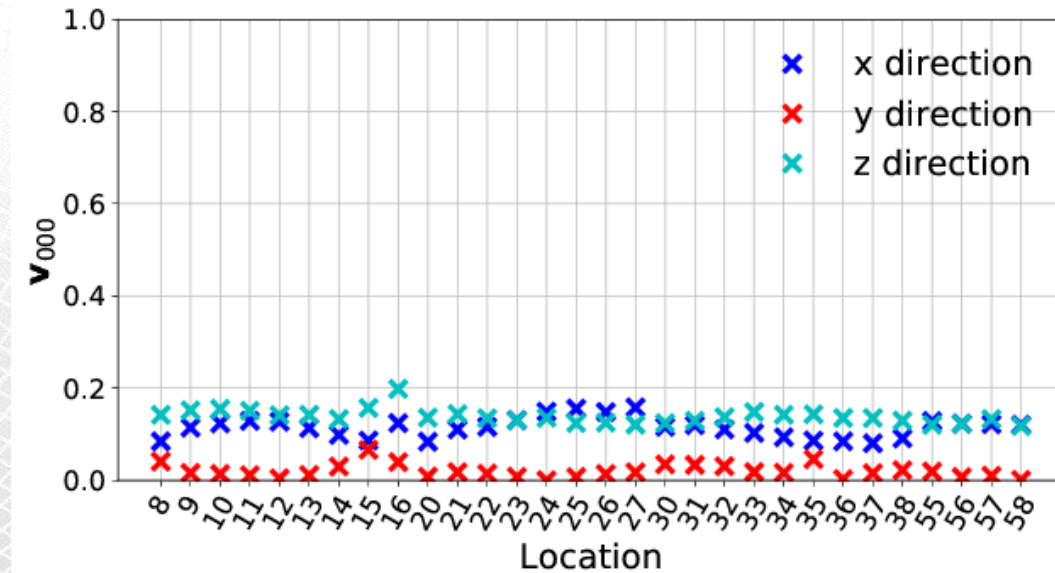
# Auroral zones

## North

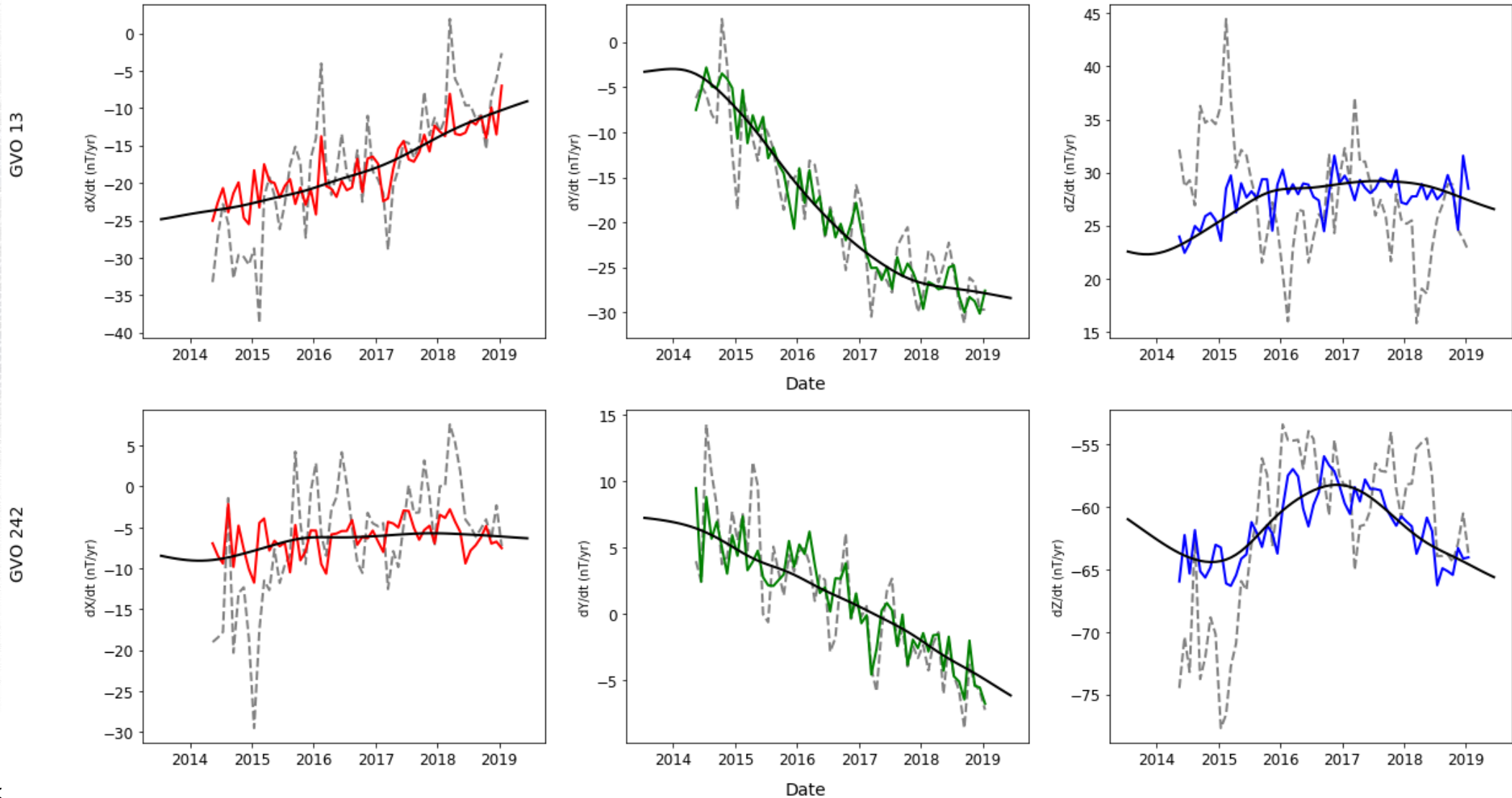
- PC0: correlations with time derivatives of Em (0.74), AE (0.73) and Dst (0.57)
- In the X and Z components (top right)
- PC1-2: peaks in DFT at ~4.2 months

## South

- PC0: correlations with time derivatives of Em (0.76), AE (0.72) and Dst (0.63)
- In the X and Z components (bottom right)
- PC1-2: peaks in DFT at ~4.2 months

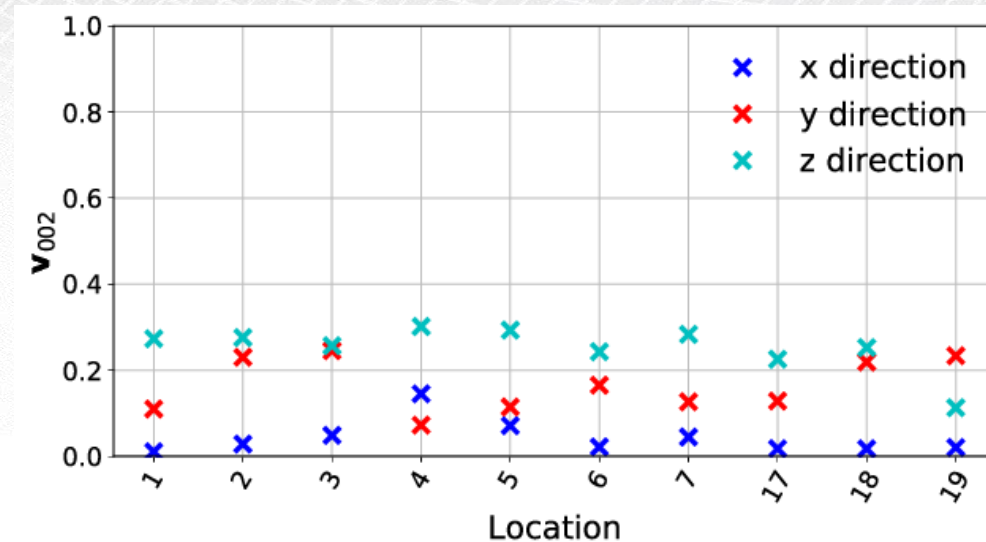
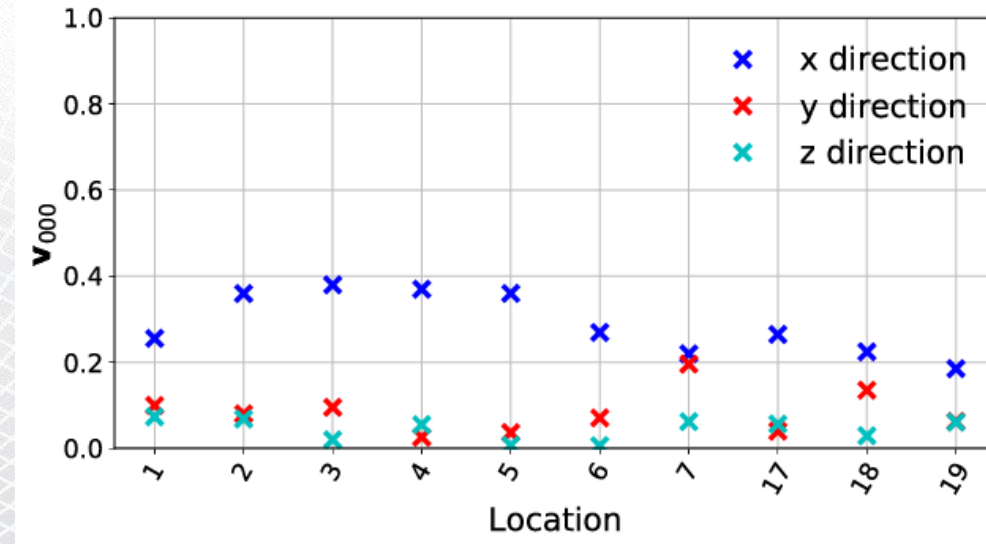


# Auroral zones: SV examples pre- (grey) and post-PCA (coloured lines). Black line: CHAOS-7.

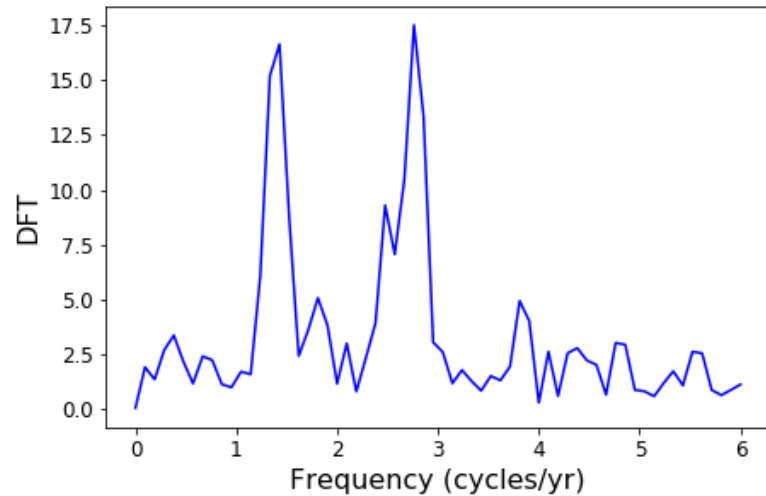
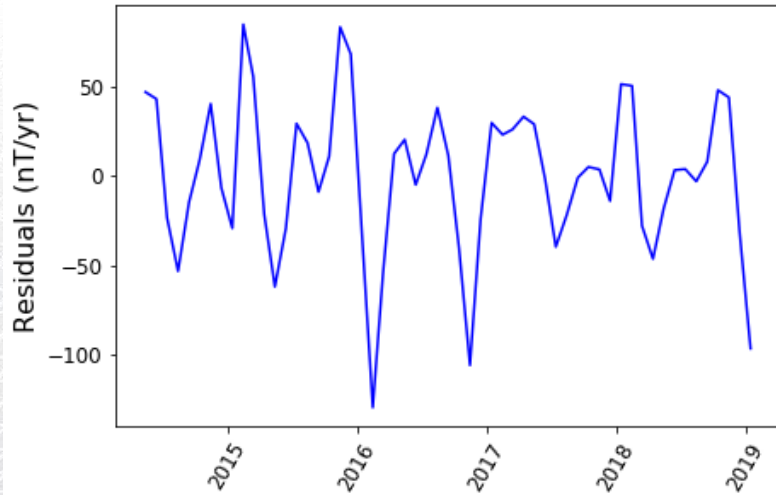


# Polar North

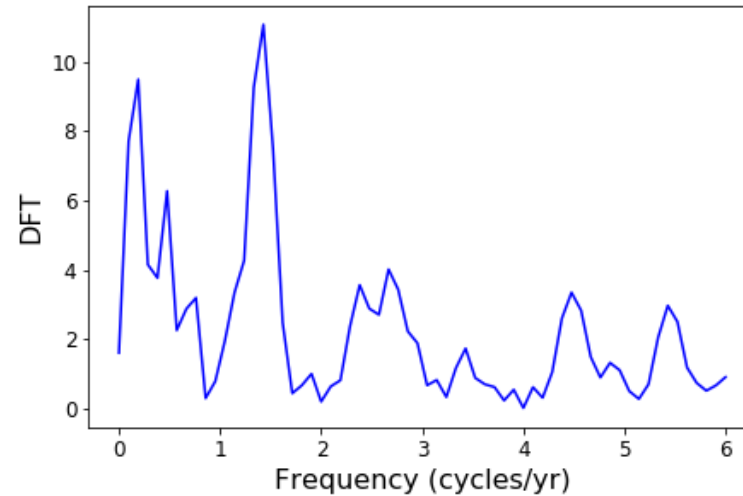
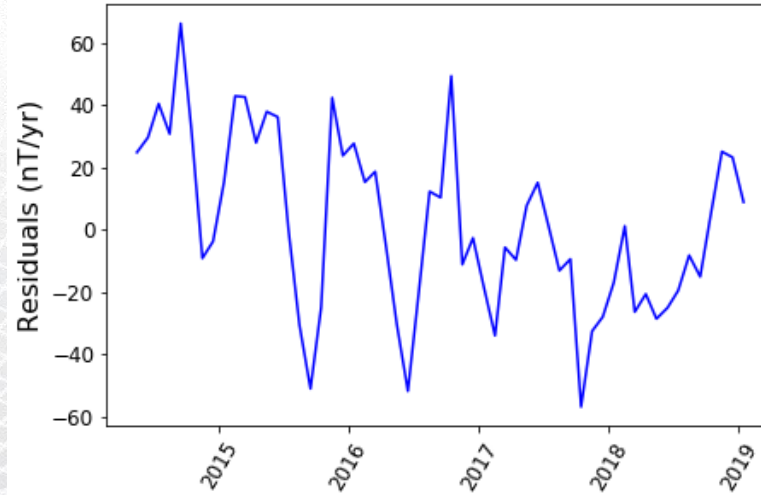
- PC0: Mostly X components (top right). Weak correlations (Em: 0.26, AE: 0.31, PCN: 0.02, PCS: 0.16)
- Local time (LT) bias frequency peak and a second frequency peak
- PC1: LT bias only, dominant Y components
- PC2: Correlations with time derivatives of Em (0.59, AE (0.57) and PCN (0.58)
- PC3: similar correlations/frequencies to PC2
- “Mystery” DFT peak at around 1.5 cycles/yr. Perhaps an artefact of dense polar sampling of external fields (see the next slide)



# Polar North: “mystery” frequency peak at ~1.5 cycles/yr)



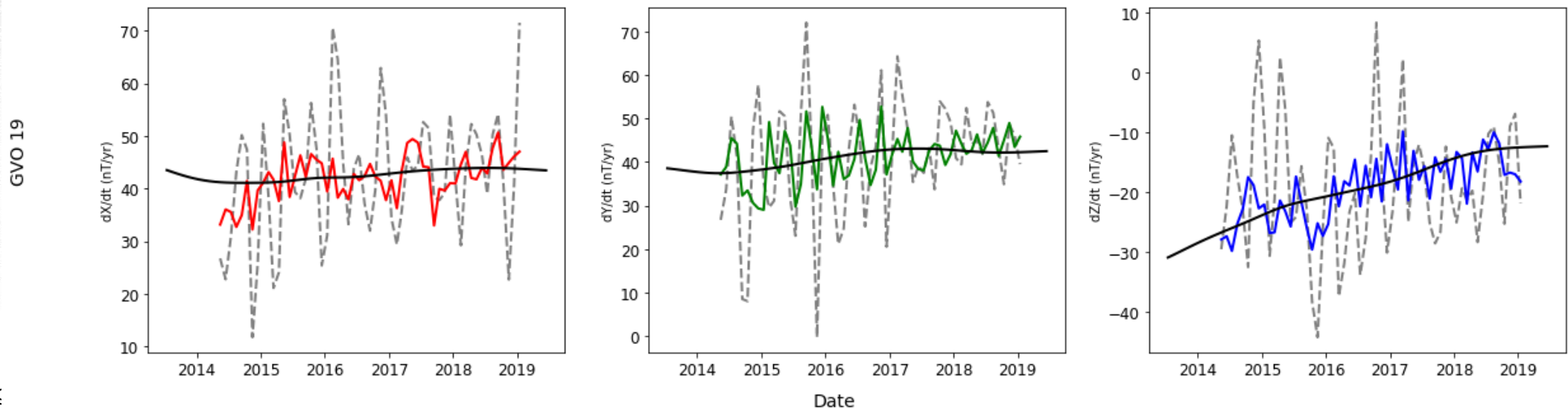
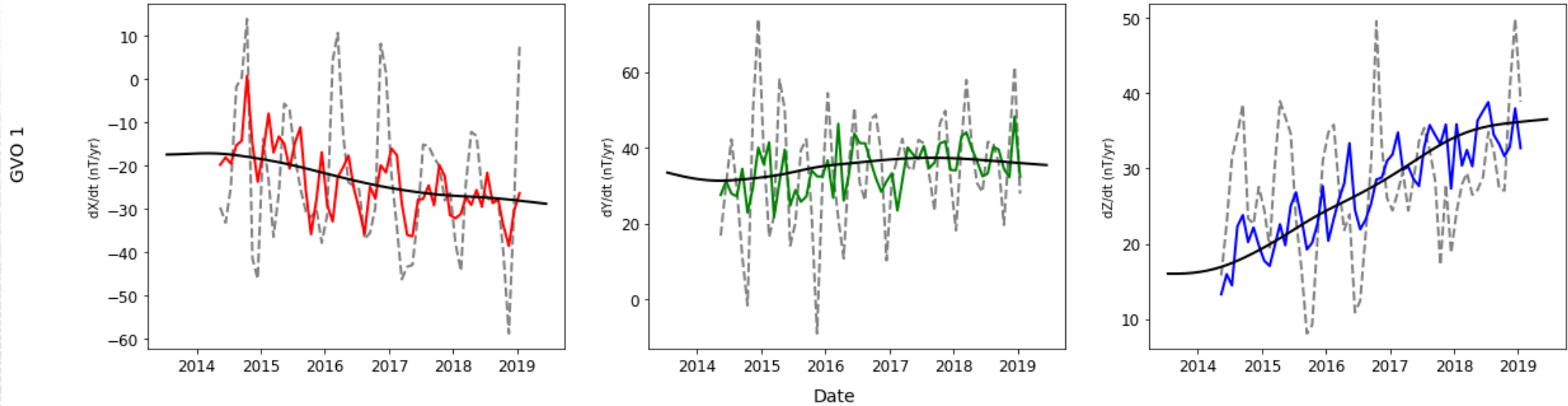
PC0: Dominant X components.  
Not strongly correlated to indices.



PC2: Dominant Z components.  
Strongly correlated to indices.

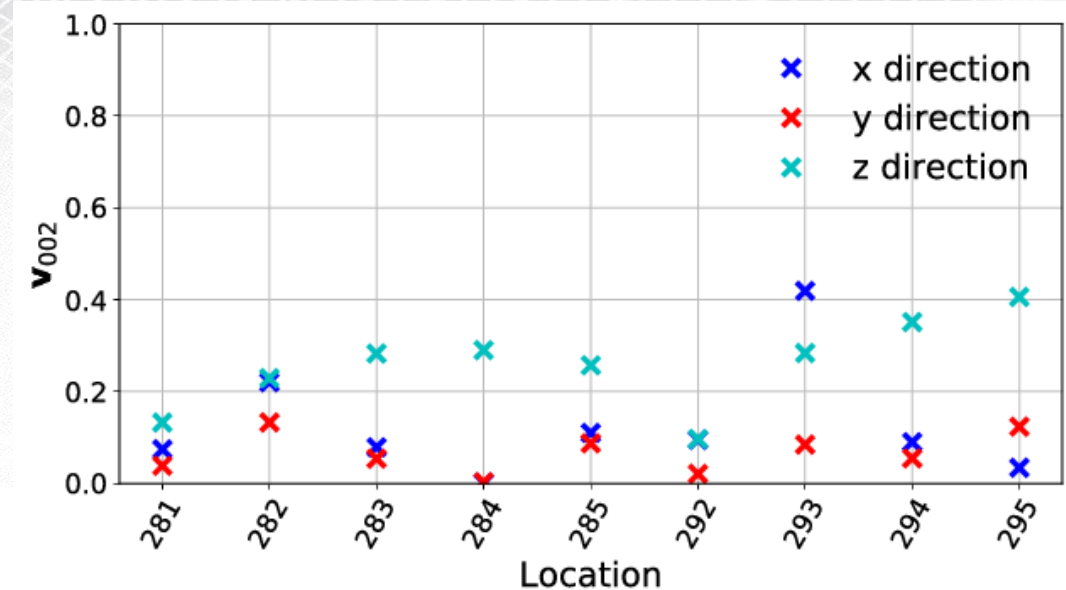
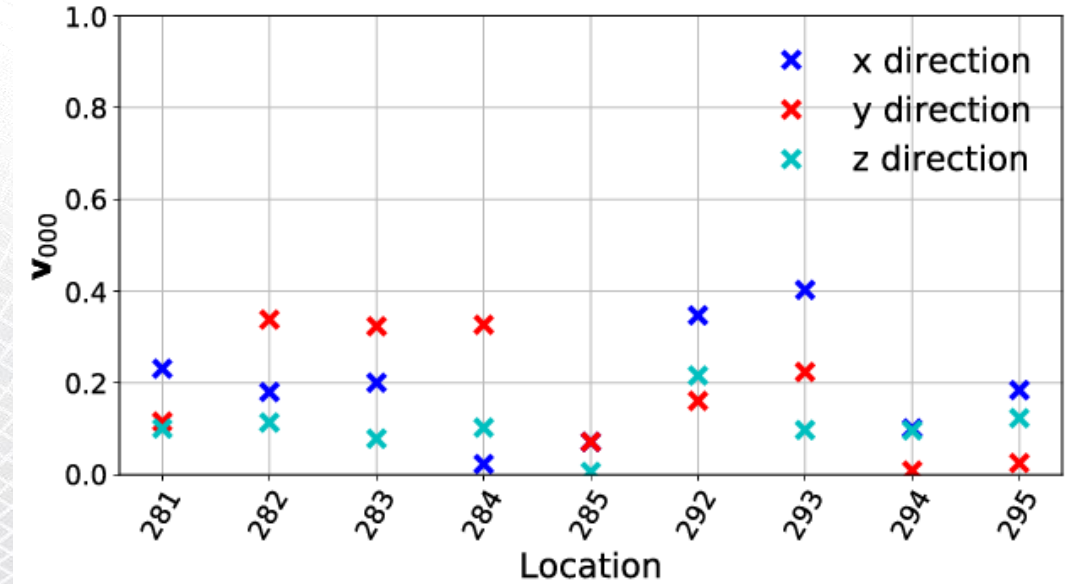
- Appears in PCs in all regions, but is most prominent at high latitude
- Sometimes the PCs are correlated to indices
- Sometimes it appears along with LT bias period
- Removed during PCA denoising because this is not a known internal signal and indices are not necessarily representative of all external sources

# Polar North: SV examples pre- (grey) and post-PCA (coloured lines). Black line: CHAOS-7.



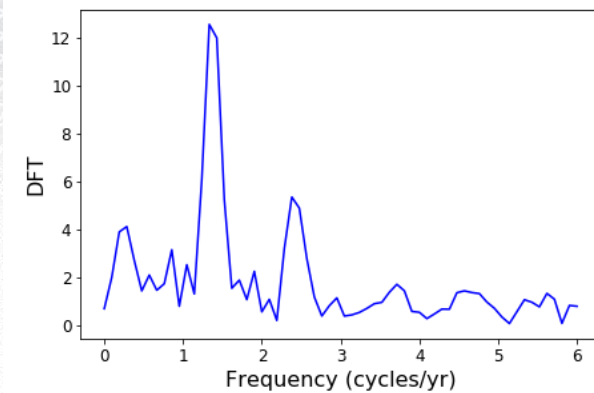
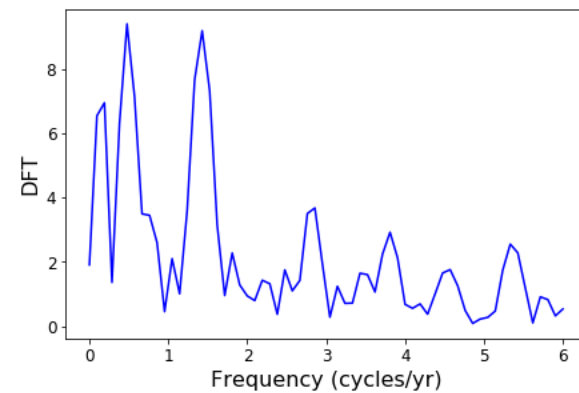
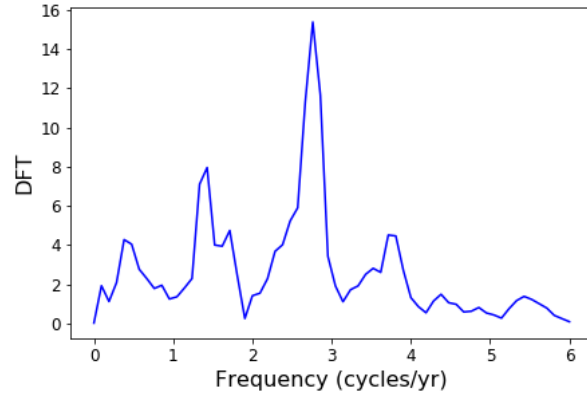
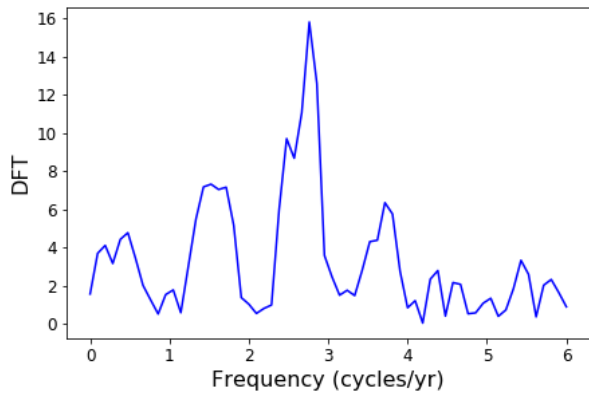
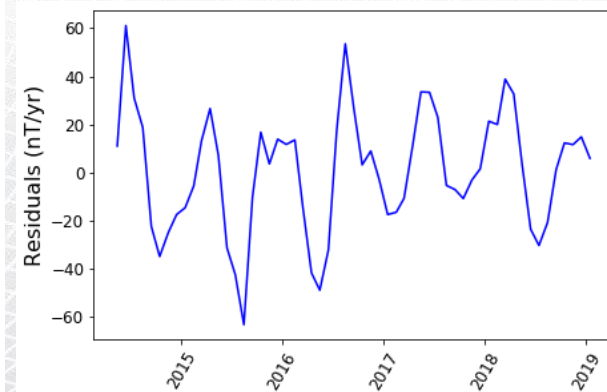
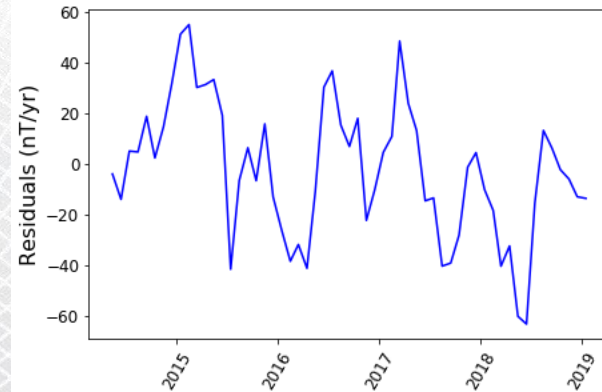
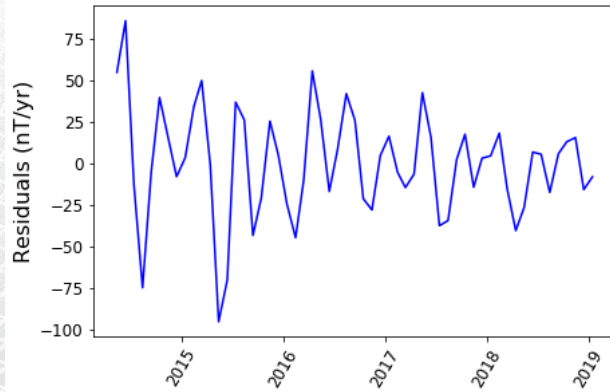
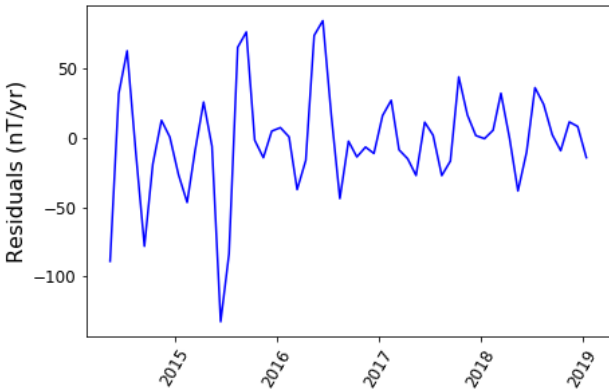
# Polar South

- PC0: Mostly X components (top right). Weak correlations (Em: 0.26, AE: 0.20, PCN: 0.26, PCS: 0.21)
- LT frequency peak and two smaller frequency peaks
- PC1: similar correlations (Em: 0.22, AE: 0.28, PCN: 0.04, PCS: 0.05) and frequencies to PC0
- PC2: Stronger correlations (Em: 0.59, AE: 0.51, PCN: 0.47, PCS: 0.55)
- PC3: no correlations, peak at the “mystery” frequency seen in other regions, especially high latitudes





# Polar South



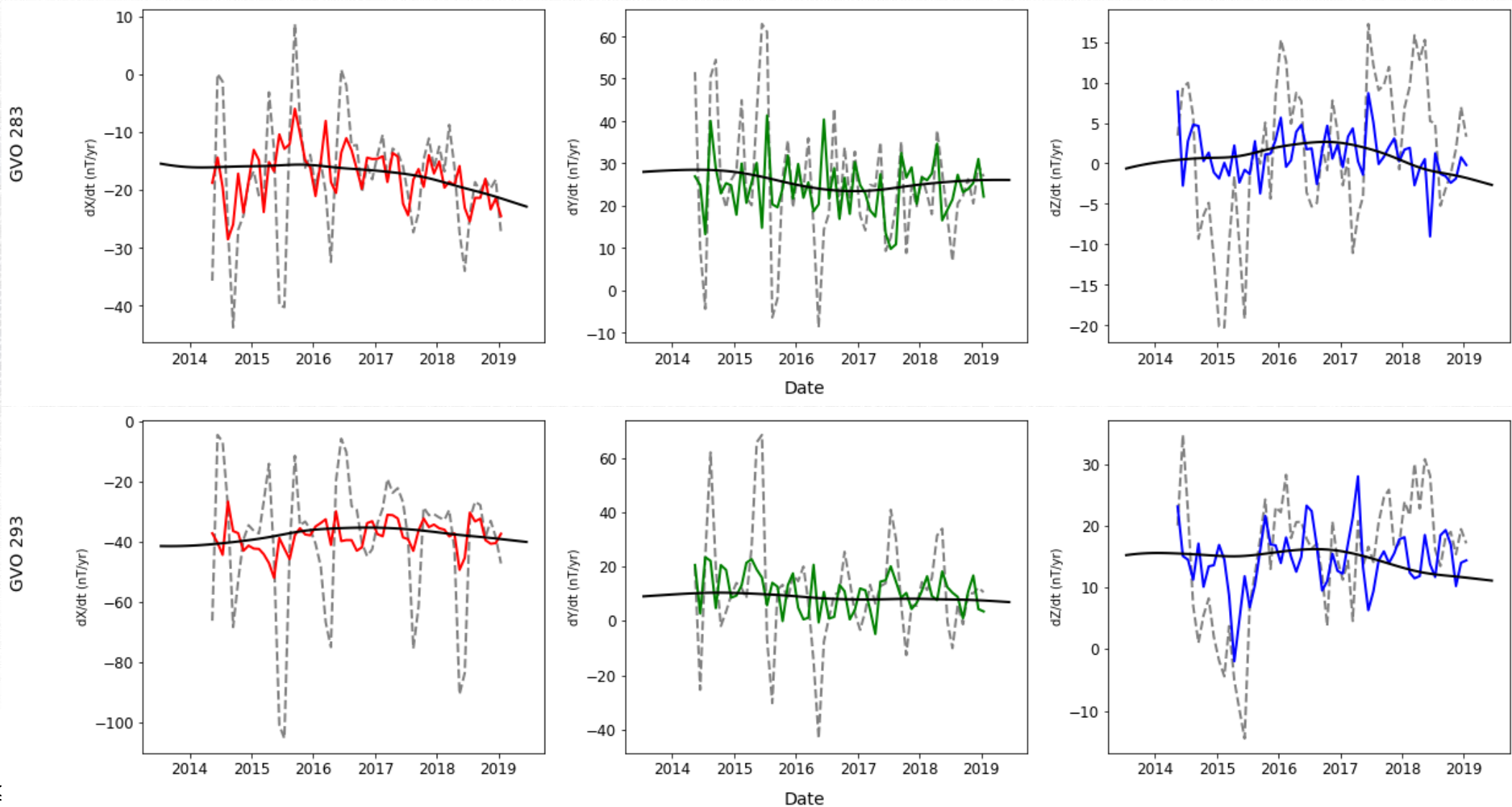
PC0: No strong correlations to magnetic indices

PC1: No strong correlations to magnetic indices

PC2: Strong correlations to magnetic indices

PC3: No strong correlations to magnetic indices

# Polar South: SV examples pre- (grey) and post-PCA (coloured lines). Black line: CHAOS-7.



# Summary

- This work is part of a Swarm DISC project whose aim is to provide new GVO data products for the core field
- We use PCA to remove contaminating signals (external fields and local time sampling biases) from an intermediate GVO data set before producing the final GVO
- Using PCA allows us to retain monthly sampling
- GVO data production will be maintained operationally at BGS and performed regularly
- GVO data will be added to the ESA Vires client for easy visualisation