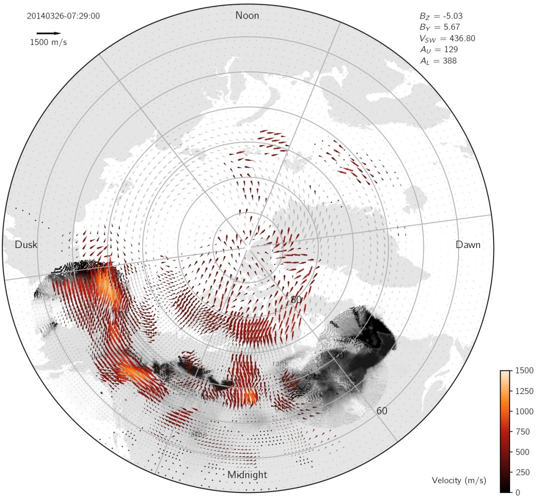
Convection Velocity Covariance Estimated from SuperDARN Observations

Bill Bristow & Matthew Wilcox

June 3, 2025

PennState
College of Earth
and Mineral Sciences

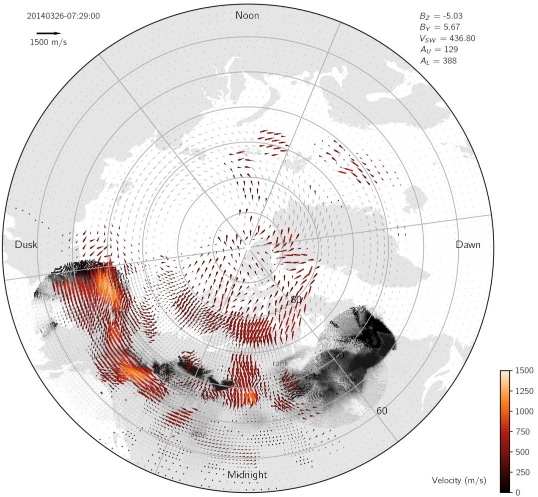
G-LDFF Fits



The LDFF fitting method uses all available LOS observations in conjunction with a climatological forecast model to generate local estimates of the convection velocity. The method uses a Bayesian inversion to combine the PDFs of the observations, the model, and an assumed divergence free condition.

The G-LDFF algorithm carries out the LDFF fitting using a set of simultaneous equations solved in a least squares sense over the entire high-latitude grid.

G-LDFF Fits



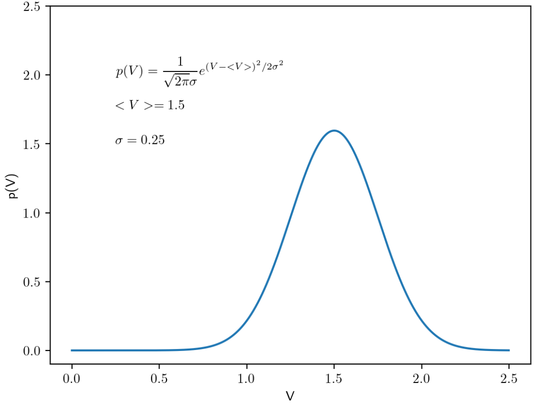
The PDF of the velocity field is determined from the PDFs of the various pieces of available information (assuming that the pieces are independent).

$$p(\mathbf{v}^t) = p(\mathbf{d})p(\mathbf{v}^f)p(\mathbf{D})$$

where:

- $p(\mathbf{v}^t)$ is the PDF of the “true” velocity
- $p(\mathbf{d})$ is the PDF of the data
- $p(\mathbf{v}^f)$ is the PDF of the forecast velocity
- $p(\mathbf{D})$ is the PDF of the divergence

Gaussian distribution of a Scalar



Many processes are well modeled by a Gaussian probability distribution.

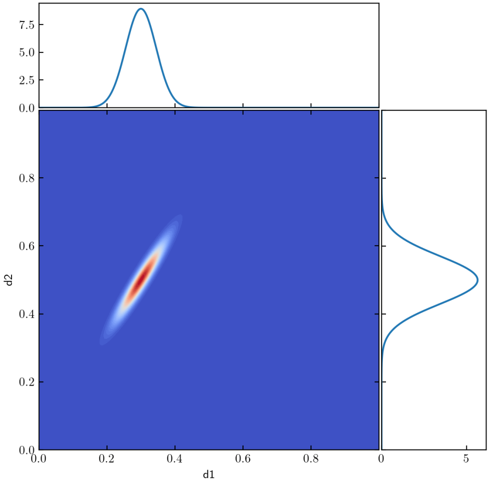
The sample mean is the numerical average of the observed points.

$$\overline{V} = \frac{1}{N} \sum_{i=1}^{N} V_i$$

The sample variance is the average of the square of the value minus the mean:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (V_i - \langle V \rangle)^2$$

Gaussian distribution of a Vector



The expression for a multi-variate Gaussian is a bit more complex.

$$p(\mathbf{v}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|} e^{-\frac{1}{2}\left(\mathbf{v}-\langle\mathbf{v}\rangle\right)^{T}\mathbf{C}^{-1}\left(\mathbf{v}-\langle\mathbf{v}\rangle\right)}$$

Where n is the number of vector elements and $\mathbf{C}$ is the covariance matrix of the vector $\mathbf{v}$, the elements of which for discrete variables are found from:

$$\mathbf{C}_{ij} = \frac{1}{K-1}\sum_{k=1}^{K}\left(\mathbf{v}_{ik} - \langle\mathbf{v}_i\rangle\right)\left(\mathbf{v}_{jk} - \langle\mathbf{v}_j\rangle\right),$$

where i and j indicate the different vector elements, and k represents different instantaneous values of the elements.

Data and Model PDFs

SuperDARN Data

The PDF of data given v^t:

$$p(\mathbf{d}\,|\,\mathbf{v}^t) = A\, e^{-(\mathbf{d}-\mathbf{M}\mathbf{v}^t)^T \mathbf{C}_d^{-1} (\mathbf{d}-\mathbf{M}\mathbf{v}^t)},$$

where $\mathbf{d}$ is a vector of the radar observations, the product $\mathbf{M}\mathbf{v}^t$ gives the linear projection of the true velocity that would produce the observations.

We assume that the variance of the data around its mean is determined by random noise in the scattering and detection process. As such, the fluctuations in each data point are independent of those of all other data points.

With that assumption, the covariance matrix $\mathbf{C}_d$ is diagonal. The elements along the diagonal being the individual data variances determined in FITACF.

Climatological Forecast

The PDF of the model given v^t and F:

$$p(\mathbf{v}^f\,|\,\mathbf{v}^t, F) = A\, e^{-(\mathbf{v}^f-\mathbf{v}^t)^T \mathbf{C}_f^{-1} (\mathbf{v}^f-\mathbf{v}^t)}.$$

The models are formed by observing convection over a long period of time and then binning based on some set of selection parameters (F).

It is not a reasonable assumption that the fluctuations of convection at one location are independent of fluctuations at all other locations. If that were the case, SuperDARN convection patterns would look like model patterns with some random noise on top.

Convection fluctuations have clear patterns that are not represented in any climatology.

$\mathbf{C}_f$ is not diagonal.

Data and Model PDFs

SuperDARN Data

The PDF of data given v^t:

$$p(\mathbf{d} \mid \mathbf{v}^t) = A\, e^{-(\mathbf{d}-\mathbf{M}\mathbf{v}^t)^T \mathbf{C}_d^{-1} (\mathbf{d}-\mathbf{M}\mathbf{v}^t)},$$

where $\mathbf{d}$ is a vector of the radar observations, the product $\mathbf{M}\mathbf{v}^t$ gives the linear projection of the true velocity that would produce the observations.

We assume that the variance of the data around its mean is determined by random noise in the scattering and detection process. As such, the fluctuations in each data point are independent of those of all other data points.

With that assumption, the covariance matrix $\mathbf{C}_d$ is diagonal. The elements along the diagonal being the individual data variances determined in FITACF.

Climatological Forecast

The PDF of the model given v^t and F:

$$p(\mathbf{v}^f \mid \mathbf{v}^t, F) = A\, e^{-(\mathbf{v}^f-\mathbf{v}^t)^T \mathbf{C}_f^{-1} (\mathbf{v}^f-\mathbf{v}^t)}.$$

The models are formed by observing convection over a long period of time and then binning based on some set of selection parameters (F).

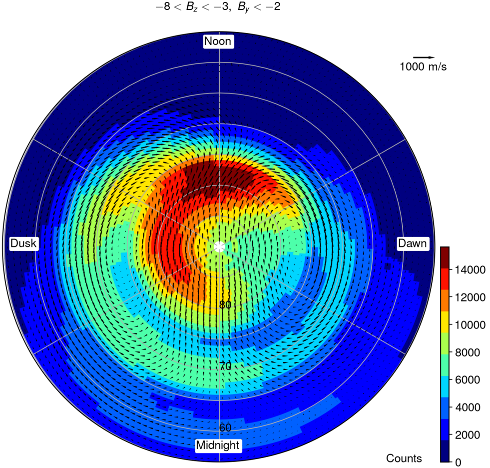
It is not a reasonable assumption that the fluctuations of convection at one location are independent of fluctuations at all other locations. If that were the case, SuperDARN convection patterns would look like model patterns with some random noise on top.

Convection fluctuations have clear patterns that are not represented in any climatology.

$\mathbf{C}_f$ is not diagonal.

Calculating a SuperDARN Covariance Matrix

Number of Points in Database



Ran the G-LDFF code on the SuperDARN database over the years 2013-2017 (a little of 2018).

Used a 1-degree in latitude equal area grid (delta-longitude changes with latitude)

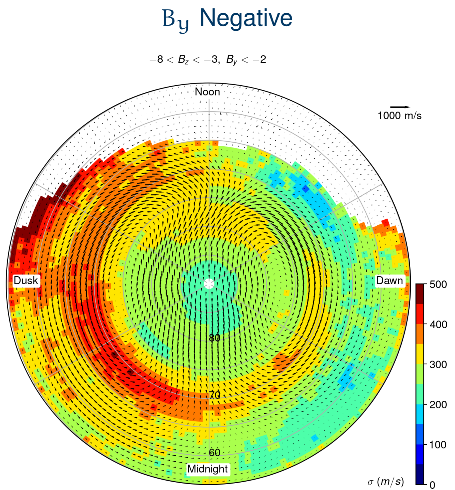
Used the Thomas-Shepherd 2018 climatology as the forecast.

Used IMF from Omni to select model. (didn't worry about dipole tilt Evan)

Created database with entries for each grid cell containing the velocity vector and associated Omni record.

Required more than 500 pts in a cell to plot covariance.

Velocity Standard Deviation (the diagonals)

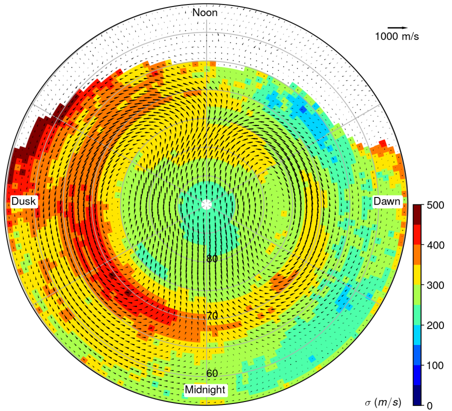


Even if one considered the model covariance matrix to be diagonal, the variances must be calculated. The figure shows that the variance depends on the state.

Velocity Standard Deviation TS18 vs Database

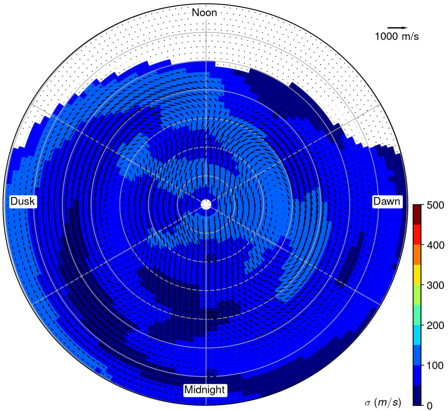
Database

$-8 < B_z < -3,\ B_y < -2$



TS18 Model

$-8 < B_z < -3,\ B_y < -2$



For the range of IMF used to create the plots, the TS18 model does not show significant variance.

Covariance matrix

The elements of the covariance matrix for discrete variables are found from:

$$\mathbf{C}_{ij} = \frac{1}{K-1} \sum_{k=1}^{K} \left(\mathbf{v}_{ik} - \langle \mathbf{v}_i \rangle\right)\left(\mathbf{v}_{jk} - \langle \mathbf{v}_j \rangle\right),$$

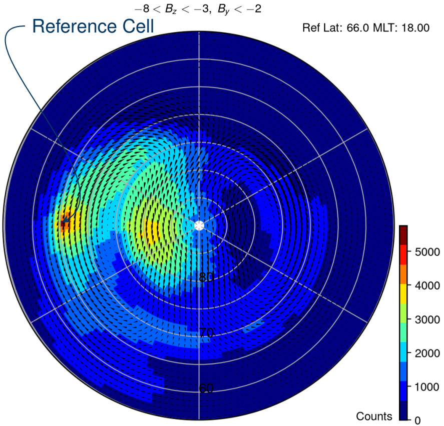
where i and j indicate the different vector elements, and k represents different instantaneous values of the elements.

When $\mathbf{v}_i$ itself is a 2-d vector, there are actually four separate covariances: $C_{\nu_{ei}\nu_{ej}}$, $C_{\nu_{ei}\nu_{nj}}$, $C_{\nu_{ni}\nu_{ej}}$, $C_{\nu_{ni}\nu_{ej}}$.

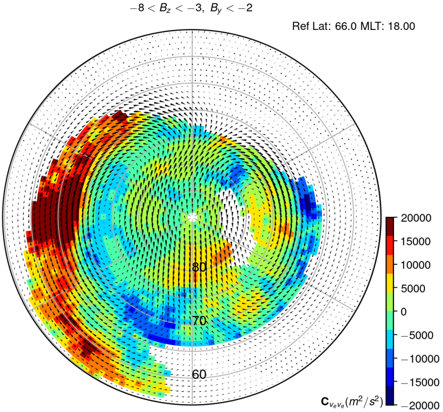
To visualize the covariance, a reference latitude-longitude cell is chosen and the covariance of the components of the velocity in that cell with those in every other cell is plotted.

Velocity Covariance For v_e in 66° MLT 18 with v_{ej} :

Number of Points

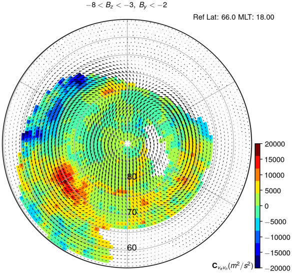


Covarance

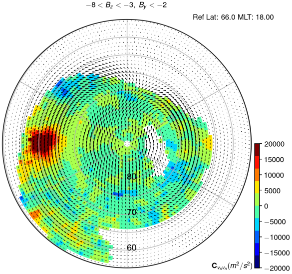


Calculating the covariance requires that simultaneous data to be available in both the reference cell and the target cell. The probability of meeting that requirement drops off with distance between the two cells.

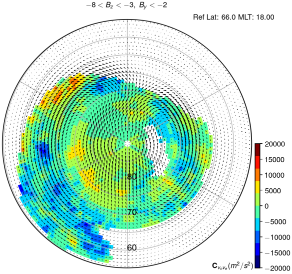
Velocity Covariance For v in 66° MLT 18 with v_j :



East - North

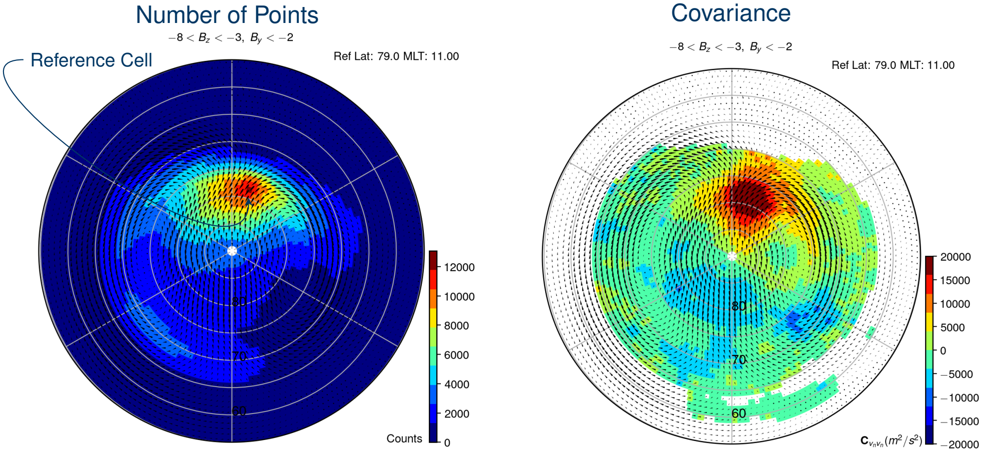


North - North



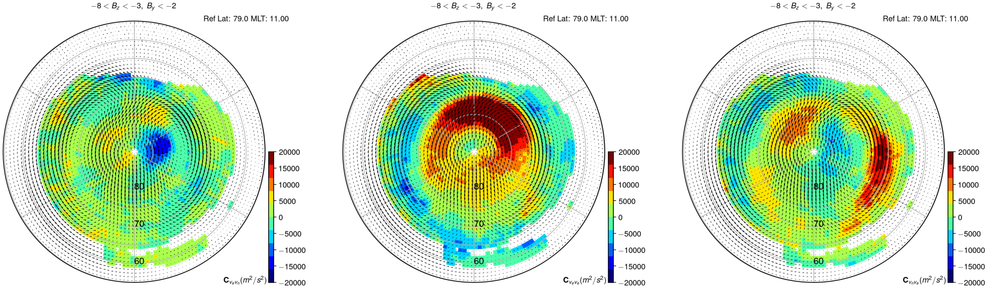
North - East

Velocity Covariance For v_n in 79° MLT 11 with v_{nj} :



Calculating the covariance requires that simultaneous data to be available in both the reference cell and the target cell. The probability of meeting that requirement drops off with distance between the two cells.

Velocity Covariance For v in 79° MLT 11 with v_j :

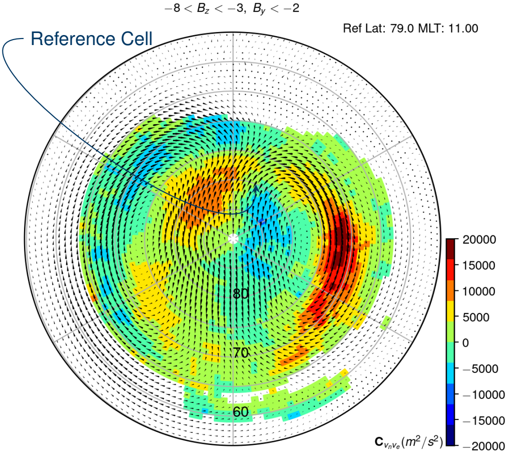


East - North

East - East

North - East

Velocity Covariance For v_n in 79° MLT 11 with v_{ej} :



North - East

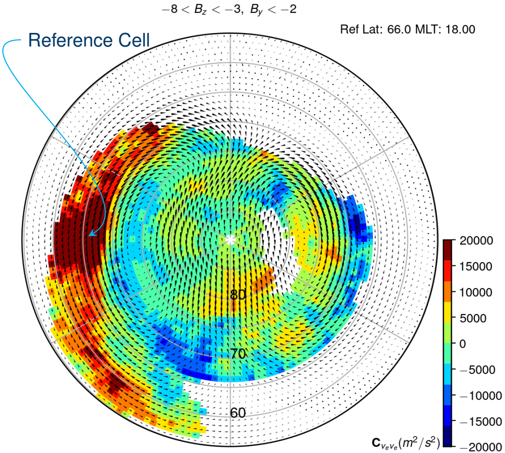
All of the covariance patterns show some interesting features.

This example shows that a velocity fluctuation near noon at 79° strongly covaries with flow in the low-latitude return region near dawn.

There is also a negative covariance in the return flow in the dusk cell (negative eastward flow for positive northward flow), though the magnitude is smaller.

If this covariance matrix was used in a convection pattern estimation for an interval when flow was elevated in the cusp, but no observations were available in the dawn-side return flow region, the solution would boost the values in that region. I.e. it wouldn't just return the model values.

Velocity Covariance For v_e in 66° MLT 18 with v_{ej} :



East - East

Is there interesting physics in the covariance?

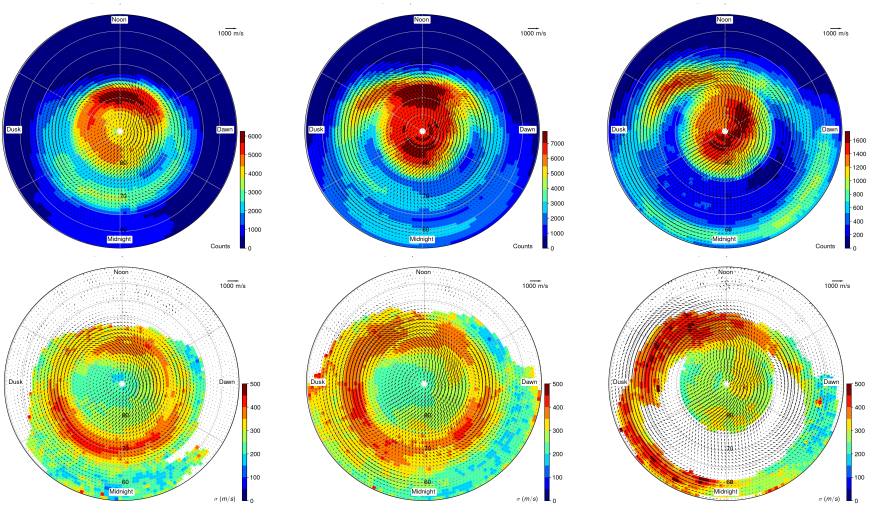
There is a region centered on the reference location that extends over about 5° in latitude and about an hour of MLT where the covariance exceeds 15 000 $m^2\,s^{-2}$. Beyond that localized region, there is a region of positive covariance extending from about 2 MLT to about 22 MLT where the value remains relatively constant between about 10 000 $m^2\,s^{-2}$ and 15 000 $m^2\,s^{-2}$.

Does the pattern give a way of estimating correlation lengths for the mechanisms responsible for the regions of elevated convection velocity? (Short-range “fluid-mechaincal” forces, vs long-range coherent magnetospheric variations)

Velocity Variance: Binning by A_l

Counts

Std Dev



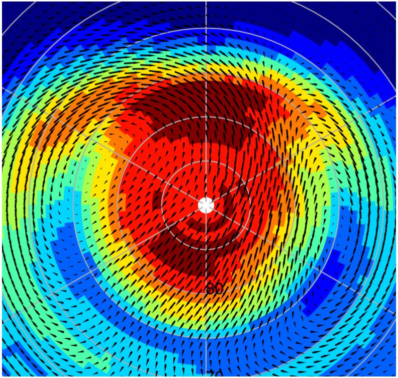
$A_l > -150$ | $-150 \geqslant A_l > -500$ | $-500 \geqslant A_l$

Summary

- An estimate of the covariance matrix is needed for all components used in data assimilation.
- When the fluctuations of elements in a vector random variable are independent, the covariance matrix is diagonal, with the matrix elements being the variances of the vector elements.
- When the vector element fluctuations are not independent, the matrix is not diagonal.
- Velocity fluctuations observed by SuperDARN at different locations are not independent.
- Long-range correlation of the velocities are observed.
- The patterns of covariance depend on the magnetospheric state.

Summary

- We are continuing to build the database of convection maps (slowly).
- A manuscript describing the covariance is forthcoming (soon)
- An implementation of the G-LDFF algorithm that uses the model covariance matrix is in the works.



Thanks!