SuperDARN velocity errors

Pasha PONOMARENKO

University of Saskatchewan

Outline

- Background
- SuperDARN velocity estimates and related errors
- Revision of error estimates
- Summary

Background

- The SuperDARN radars detect echoes from continuously fluctuating ionospheric plasma, so the backscatter data represent a random signal which characterised by statistical parameters (mean, variance, etc.). The HF radar data are also subject to different kinds of external noise and interference.
- Both sets of factors contribute to errors in estimating ionospheric plasma parameters.

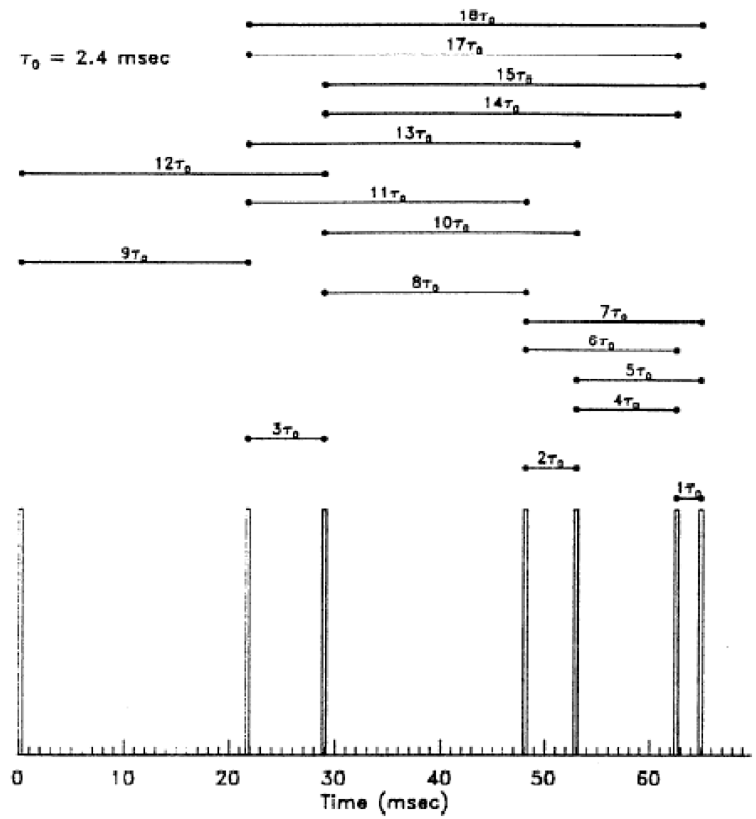
Background & problem formulation (cont.)

- The existing data processing routine, FITACF, deals with some of these factors. However, it currently produces unrealistically small velocity errors ~0.1 m/s. In addition, there is a small but distinct population of echoes characterised by very large errors of ~40,000 m/s.
- In order to improve this situation, we need to analyse what FITACF is actually doing and, if necessary, to go back to the basic statistics.

SuperDARN ACF

SuperDARN emits a sequence of unevenly spaced pulses that are combined into different ACF time lags and averaged .

It is done to resolve the conflicting requirements imposed on the sampling rate due to the desired ranges of Doppler shift and spatial coverage.



BARTHES ET AL.: SUPERDARN HIGH-RESOLUTION SPECTRAL ANALYSIS
Radio Science, Volume 33, Number 4, Pages 1005–1017, July–August 1998

SuperDARN ACF

- A complex ACF is calculated for each beam-range cell (range gate)
- Real part is calculated as a normal ACF

$$\Re\left[R_{xx}(\tau)\right] = \frac{\langle x(t)\ x(t-\tau)\rangle}{\sigma_x^2}$$

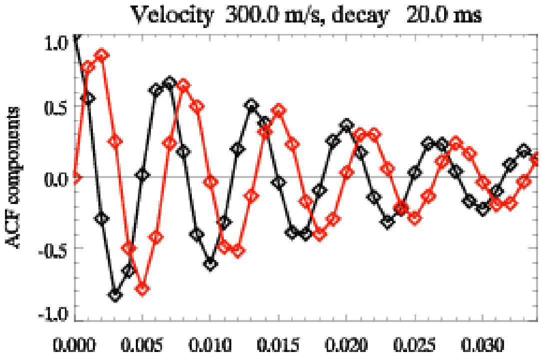
- Imaginary part results from correlating the signal with its Hilbert transform (all spectral components are shifted by 90 deg):

$$\Im\left[R_{xx}(\tau)\right] = \frac{\langle x(t)\ H\left[x(t-\tau)\right]\rangle}{\sigma_x^2}$$

Why to average?

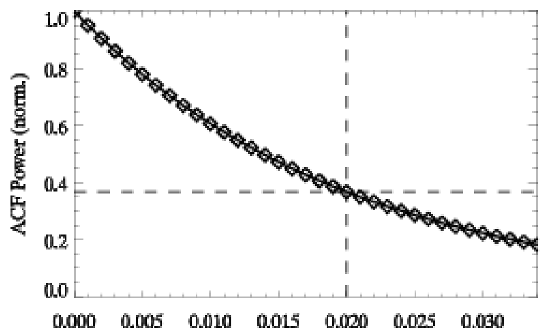
- ACF averaging decreases incoherent contribution from the cross-range interference (CRI) which is an intrinsic feature of multi-pulse radars (see *FITACF tutorial* by K.Baker).
- HF backscatter itself represents a random signal so an accurate estimate of its parameters also requires decent averaging to reduce statistical uncertainty $\sigma \propto 1/N^{1/2}$.

SuperDARN ACF



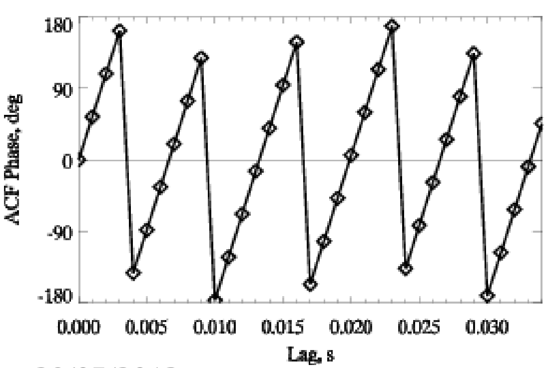
Complex ACF :

$$R(\tau) = \Re[R(\tau)] + j\Im[R(\tau)] = |R(\tau)|e^{-j\varphi(\tau)}$$



ACF power (normalised) :

$$|R(\tau)| = \sqrt{\Re^2[R(\tau)] + \Im^2[R(\tau)]}$$



ACF phase :

$$\varphi(\tau) = \tan^{-1}\frac{\Im[R(\tau)]}{\Re[R(\tau)]}$$

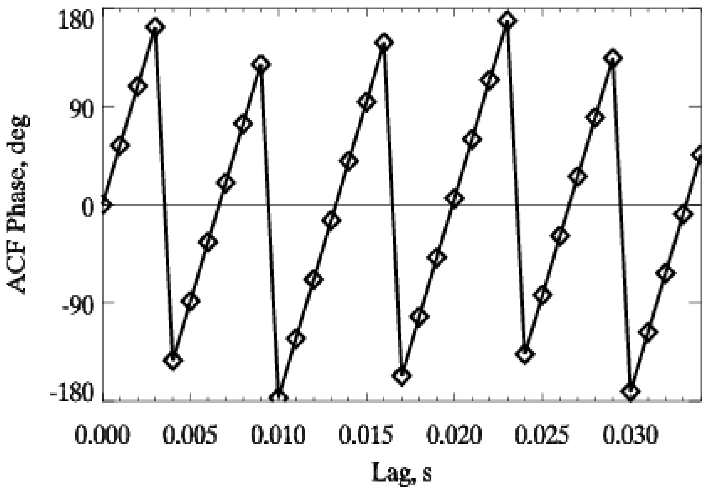
SuperDARN velocity

- SuperDARN velocity measurements assume a uniform plasma drift within the scattering volume (range gate).
- In this case, the ACF phase represents a linear function of the time lag, and the line-of-sight velocity can be estimated from the phase derivative over the time lag.

SuperDARN velocity (cont.)

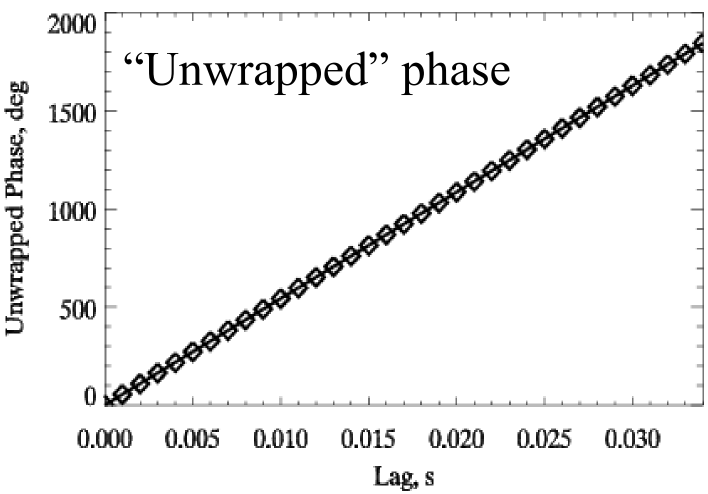
Measured phase (ideal)

φ



τ

φ



τ

Linear phase model:

$$\varphi = b\tau$$

Phase slope:

$$b = \frac{\partial \varphi(\tau)}{\partial \tau} = \frac{\varphi}{\tau}$$

Velocity:

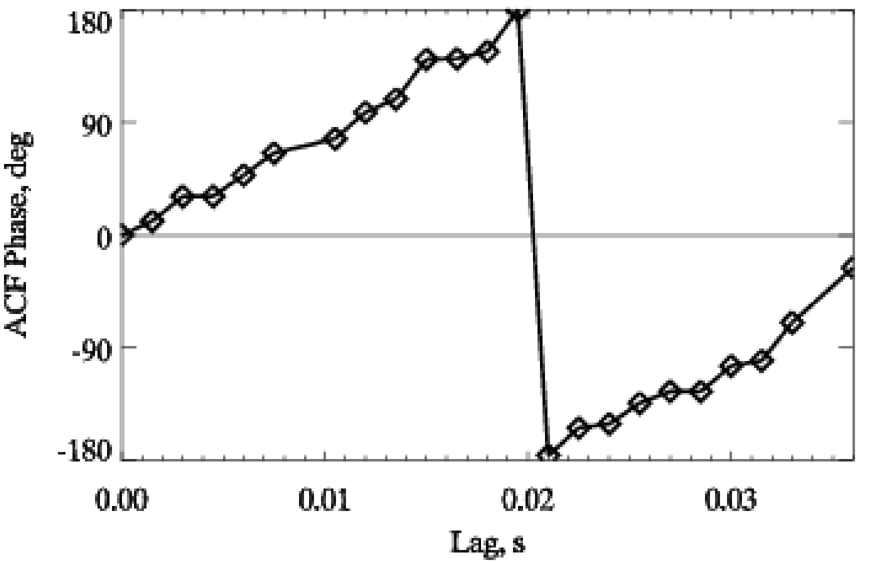
$$v = \frac{\lambda}{4\pi} b$$

Ideally, velocity can be estimated from a single ACF lag, but…

SuperDARN velocity (cont.)

- In reality, the phase fluctuates and introduces uncertainty in determining the slope b.
- It becomes necessary to estimate an average slope
- In FITACF, this is done by fitting a linear model to $\varphi(\tau)$.

Realistic phase



$$v = \frac{\lambda}{4\pi}\frac{\partial\varphi}{\partial\tau}$$

Phase fitting

- Fitting is done by minimising a weighted sum of measured phase deviations from the linear model $\varphi(\tau) = b\tau$ (e.g. *Numerical Recipes*, Ch. 15.2),

$$\chi^2 = \sum_{i=1}^{n_{lag}} \frac{\left(\varphi_i - b\,\tau_i\right)^2}{\sigma_{\varphi i}^2}$$

  - n_{lag} is the number of lags (degrees of freedom, DoF)
  - $\sigma_{\varphi\, i}$ is the phase variance at the i^{th} lag

Phase fitting (cont.)

- In our case, we can obtain an analytic solution:

$$\frac{\partial \chi^2}{\partial b} \equiv 0 \rightarrow \boxed{b = \frac{\sum_{i=1}^{n_{lag}} \left(\varphi_i \tau_i \big/ \sigma_{\varphi i}^2 \right)}{\sum_{i=1}^{n_{lag}} \left(\tau_i^2 \big/ \sigma_{\varphi i}^2 \right)}}$$

Fitting errors

- Slope fitting errors can be estimated by propagation of the phase errors

$$\sigma_b^2 = \sum_{i=1}^{n} \sigma_{\varphi i}^2 \left(\frac{\partial b}{\partial \varphi_i} \right)^2 \quad \Rightarrow \quad \boxed{\sigma_b^2 = \frac{1}{\sum_{i=1}^{n_{lag}} \tau_i^2 / \sigma_{\varphi i}^2}}$$

- Importantly, a larger number of lags (degrees of freedom) should lead to smaller errors

Assumed phase statistics

- To proceed further, we need to know σ_φ. In the current data processing package, FITACF, some *ad hoc* assumption have been made about the phase variance (*White Paper* by K. Baker and G. Blanchard). It was assumed that the phase variance is inversely proportional to the ACF power at a given lag,

$$\sigma_\varphi(\tau) = \langle \sigma_\varphi \rangle \frac{\langle P \rangle}{P(\tau)}$$

Here $\langle P \rangle$ is the <u>average</u> ACF power and $\langle \sigma_\varphi \rangle$ is the unknown effective phase variance

- This assumption seems qualitatively reasonable, because with decreasing P the relative contribution from statistical power fluctuations to φ should increase

$$\varphi = \tan^{-1} \frac{\langle \Im[P(\tau)] \rangle + \sigma_P^{\Im}}{\langle \Re[P(\tau)] \rangle + \sigma_P^{\Re}}.$$

FITACF velocity

- Then the expression for the fitted velocity becomes

$$v = \frac{\lambda}{4\pi} b = \frac{\lambda}{4\pi} \frac{\sum_i \varphi_i \tau_i P_i^2}{\sum_i \tau_i^2 P_i^2}$$

- Importantly, the unknown $\langle \sigma_\varphi \rangle$ is not required here.

Fitting errors

- A linear relation between phase and velocity makes it easy to convert slope errors into velocity errors

$$\sigma_v = \frac{\lambda}{4\pi}\sigma_b$$

- However, estimation of the slope variance requires knowledge of the effective phase variance, $<\sigma_\varphi>$, which is not a trivial thing.

$$\sigma_b^2 = \frac{1}{\sum_{i=1}^{n_{lag}} \tau_i^2 / \sigma_{\varphi i}^2} = \frac{\left\langle \sigma_\varphi \right\rangle^2 \left\langle P \right\rangle^2}{\sum_{i=1}^{n_{lag}} \tau_i^2 P_i^2}$$

FITACF phase variance

- In FITACF, it is estimated directly from the measured phase deviations as a power-weighted average of the phase deviation at different lags:

$$\left\langle \sigma_{\varphi} \right\rangle^2 = \frac{n_{lag}}{n_{lag}-1} \frac{\sum_i P_i^2 \left(\varphi_i - b\tau_i \right)^2}{\sum_i P_i^2}$$

Here, the $n_{lag}/(n_{lag} - 1)$ factor reflects a decrease in the number of degrees of freedom due to one free (fitted) parameter, the slope.

Finally, FITACF velocity "error"

- Propagation of errors gives us the respective velocity error:

$$\sigma_v^2 = \left(\frac{\lambda}{4\pi}\right)^2 \sum_i \sigma_{\varphi i}^2 \left(\frac{\partial b}{\partial \varphi_i}\right)^2 = \langle \sigma_\varphi \rangle^2 \frac{\langle P \rangle^2}{\sum_i P_i^2 \tau_i^2} \left(\frac{\lambda}{4\pi}\right)^2$$

$$= \frac{n_{lag}}{n_{lag}-1} \frac{\sum_i P_i^2 (\varphi_i - b\tau_i)^2}{\sum_i P_i^2} \frac{\langle P \rangle^2}{\sum_i P_i^2 \tau_i^2} \left(\frac{\lambda}{4\pi}\right)^2$$

Contemporary FITACF summary

- This is what we have now.
- A close look at the code revealed that there was a missing term of sqrt(N_{lag}), which led to very low error estimates.
- In addition, a “bad quality slope” flag 9999 is interpreted as valid data generating very large error values ~40000 m/s.
- These errors can be easily fixed, but we still need to know how realistic the implemented assumptions are.
- In order to achieve this, we need a more in-depth analysis supplied by realistic simulations, where we can compare a known input with measured output.

Theoretical phase variance

- Now we will go a step further and see what the theoretical RMS phase variance is. According to Bendat & Piersol (*Random data analysis and Measurement Procedures*, John Wiley & Sons, 2000), it is:

$$\sigma_\varphi(\tau) = \sqrt{\frac{\left|R^{-2}(\tau)\right| - 1}{2N}}$$

- Here *N* is the number of averages. Importantly, it also depends on the magnitude of the correlation coefficient at a given lag,

$$|R(\tau)| = P(\tau)/\,P(0)$$

Theoretical phase variance (cont.)

- The variance increases with de-correlation, from zero to infinity

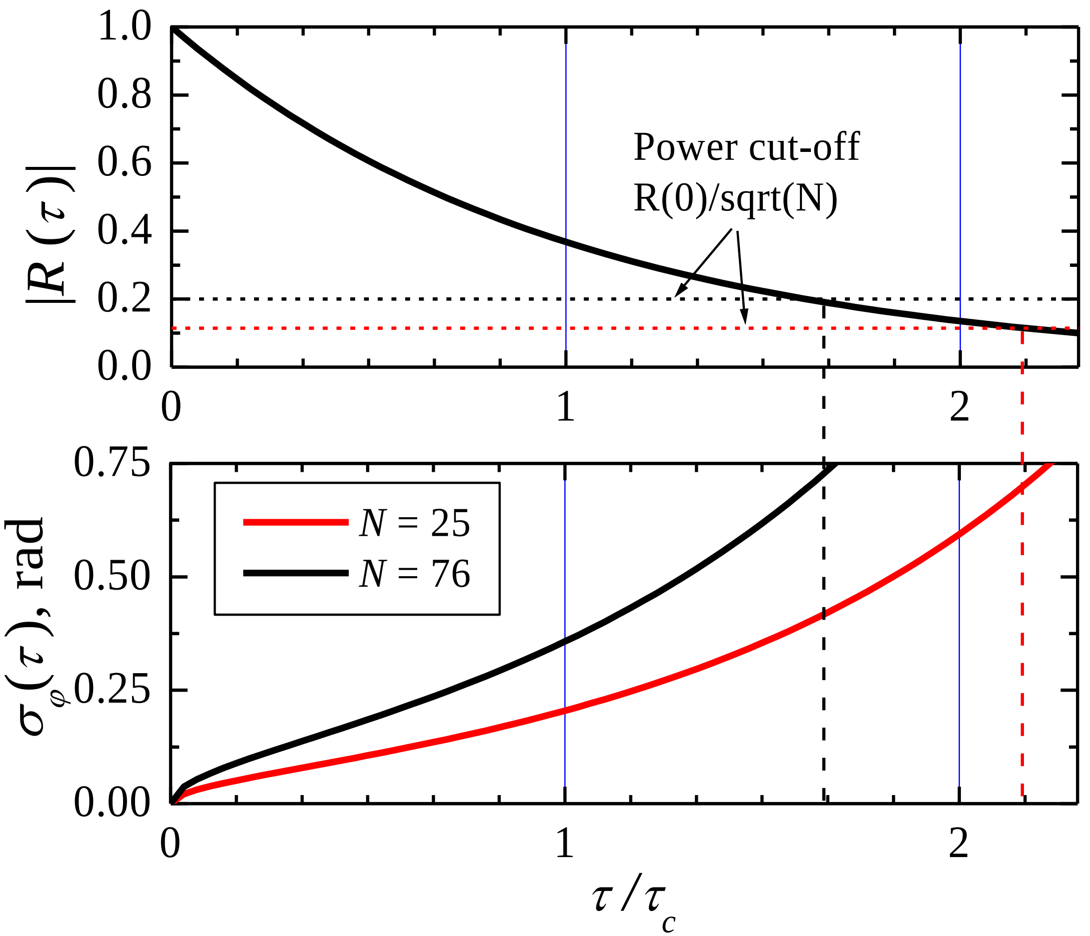
$$\sigma_{\varphi}(\tau)=\sqrt{\frac{\left|R^{-2}(\tau)\right|-1}{2N}}$$

$$|R|=1 \Rightarrow \sigma_{\varphi}=0, \quad |R|=0 \Rightarrow \sigma_{\varphi}=\infty$$

Illustrations for exponential decay

$$\left|R(\tau)\right| = \exp(-\tau/\tau_c)$$

$$\sigma_\varphi(\tau) = \sqrt{\frac{\left|R^{-2}(\tau)\right| - 1}{2N}}$$



Comparison with FITACF

FITACF

$$\sigma_\varphi(\tau) = \langle \sigma_\varphi \rangle \frac{\langle P \rangle}{P(\tau)} \propto \frac{\langle P \rangle}{P(\tau)}$$

Bendat & Piersol

$$\sigma_\varphi(\tau) = \sqrt{\frac{1 - \left|R^2(\tau)\right|}{2N\left|R^2(\tau)\right|}}$$

$$\left|R^2(\tau)\right| << 1 \Rightarrow \sigma_\varphi(\tau) \propto \frac{P(0)}{P(\tau)}$$

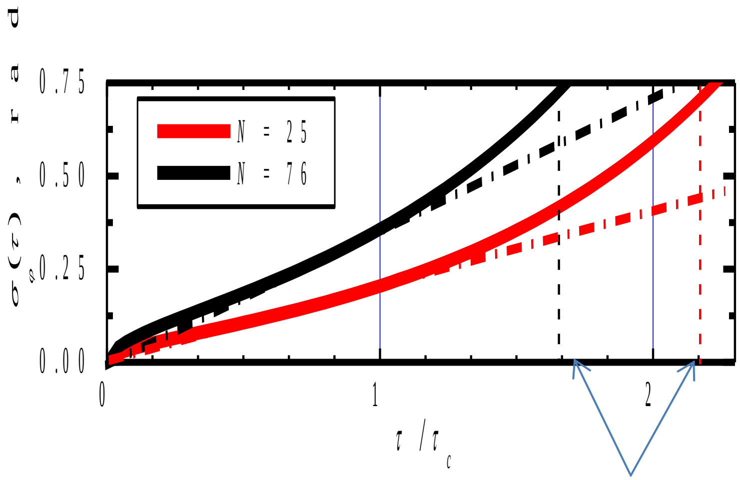
The theoretical dependence of phase variance on ACF power coincides with that used in FITACF for relatively low levels of correlation $\tau \geq \tau_c$, but assigns relatively larger weight (i.e. lesser variance) to the highly correlated lags.

Also, FITACF underestimates variance at all lags for smaller τ_c (larger *W*), because the ratio $<P>/P(0)$ decreases with increasing τ_c .

Velocity variance

A zero-order estimate of the phase slope variance (STD) can be obtained by linearising the phase error's dependence on the time lag, using the decay time value as a reference point:

$$\sigma_b^{lag} \approx \Delta\varphi / \Delta\tau \approx \sigma_\varphi(\tau_c) / \tau_c$$



Power cut-offs

The respective velocity STD is proportional to spectral width and to $1/\sqrt{N}$

$$\sigma_v^{lag} = \frac{\lambda}{4\pi}\sigma_b \approx \frac{\lambda}{4\pi}\frac{1}{\tau_c}\sqrt{\frac{1-\left|R^{-2}(\tau_c)\right|}{2N}}$$

$$= w\sqrt{\frac{1-e^{-2}}{8N}} \approx \frac{w}{\sqrt{N}}$$

Here we use the relation between spectral width and decay time,

$$w = \frac{\lambda}{2\pi}\frac{1}{\tau_c}$$

Velocity variance (cont.)

The expression obtained for the velocity STD has very clear physical meaning: a wider Doppler spectrum corresponds to a larger uncertainty in the location of its maximum. On the other hand, a larger number of averages produces a smoother spectrum. Importantly, this value is independent of the radar frequency.

$$\sigma_v^{lag} \approx \frac{w}{\sqrt{N}}$$

$$w = 100\,\mathrm{m/s}$$

$$N = 25 \rightarrow \sigma_v^{lag} \approx 20\,\mathrm{m/s}$$

$$N = 76 \rightarrow \sigma_v^{lag} \approx 11.5\,\mathrm{m/s}$$

Fitting errors

These numbers are further improved by fitting a linear function to the phase at n_{lag} available ACF lags, so that the uncorrelated phase fluctuations at different lags partially compensate each other and decrease the overall velocity uncertainty $\propto 1 \big/ \sqrt{n_{lag}}$

$$\sigma_b = \left(\sum_{i=1}^{n} \tau_i^2 \big/ \sigma_{\varphi i}^2 \right)^{-\frac{1}{2}}$$

$$\sigma_{\varphi i} \approx \frac{\tau_i}{\tau_c} \sigma_\varphi(\tau_c) \rightarrow \sigma_b \approx \frac{\sigma_\varphi(\tau_c)}{\tau_c} \frac{1}{\sqrt{n_{lag}}}$$

$$\boxed{\sigma_v = \frac{\lambda}{4\pi} \sigma_b \approx \frac{w}{\sqrt{N}\sqrt{n_{lag}}}}$$

$$w = 100\,\mathrm{m/s}$$

$$n_{lag} = 22$$

$$N = 25 \rightarrow \sigma_v \approx 4.3\,\mathrm{m/s}$$

$$N = 76 \rightarrow \sigma_v \approx 2.5\,\mathrm{m/s}$$

Numerical simulations

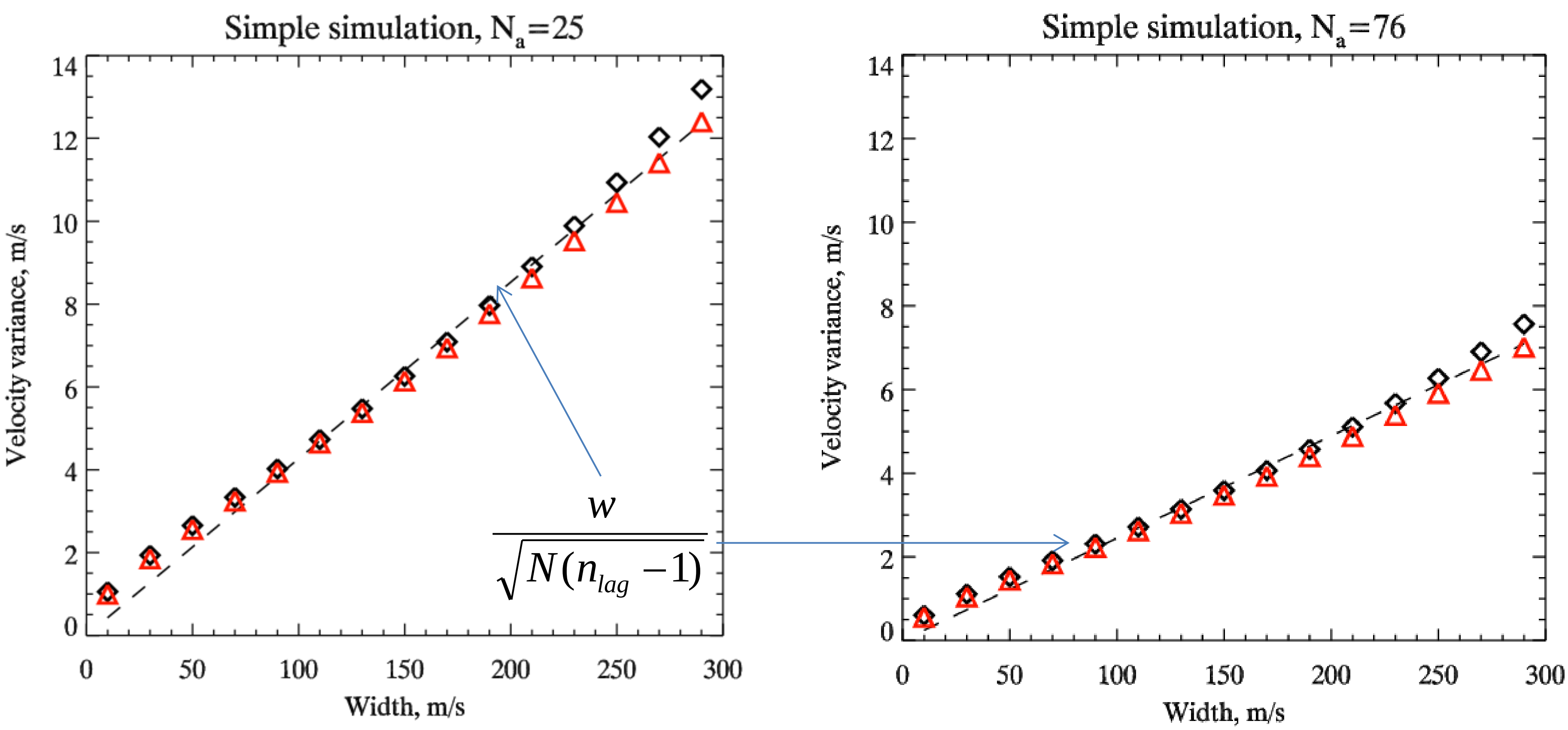
The theoretical estimates have to be tested against real data to make sure that our statistical model is reasonable. However, there is a problem: in real data, we do not know the "input" velocity value to be able to estimate how much its measured value deviates from the "real" value. However, a zero-order test can done using simulated data where the input is known.

Simple Simulation

- The simplest simulation was performed by generating independent normally distributed phase variations at all lags, with variance magnitudes defined by the theoretical dependence on $|R|$ and N
- These simulations agreed well with the theory for both velocity value and its variance.

Simple simulation (cont.)

Dashed line – theory, **black diamonds** – measured deviation, red triangles – FITACF errors



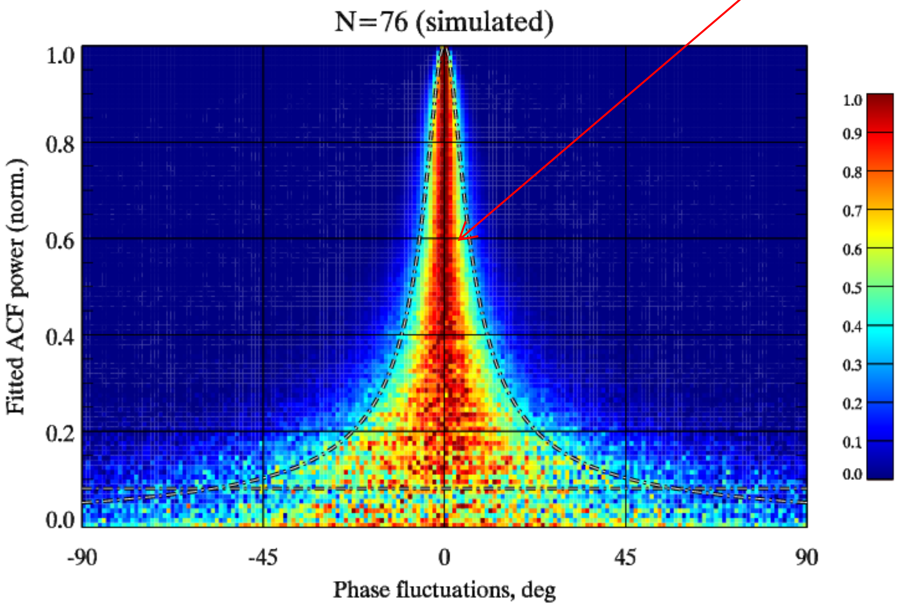
More realistic simulation

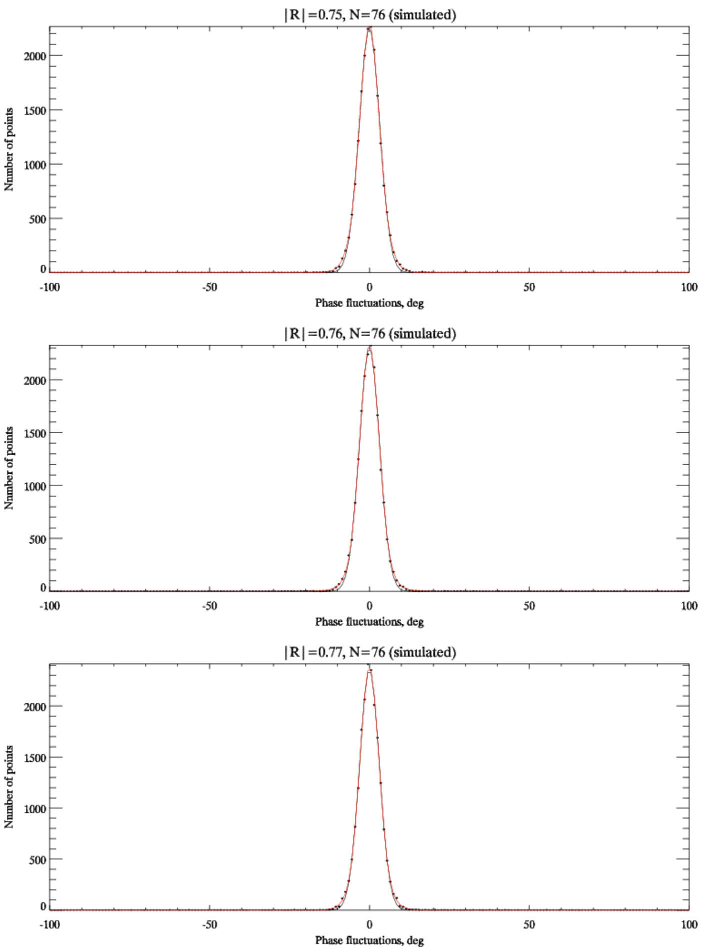
- Collective scatter approach + SuperDARN pulse sequence

  (Ribeiro et al, *Radio Science* [2013])
  - Large number of point targets inside the range gate
  - Input parameters:
    - Amplitude (power)
    - Lifetime (spectral width)
    - Background velocity
    - Number of averages (fluctuation level)
  - Output – standard RAWACF data

Simulated phase variance (N=76)

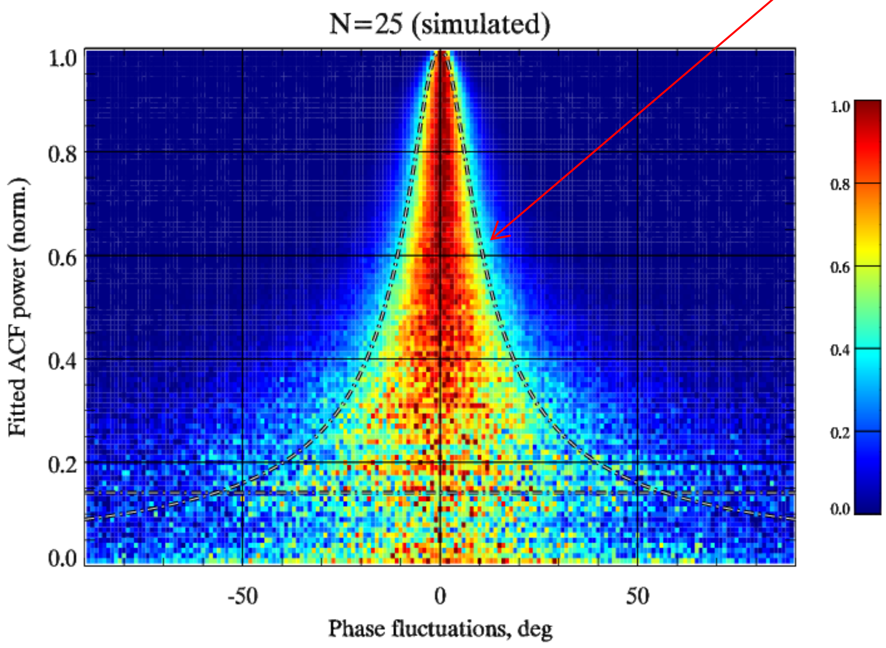
$$\sigma_{\varphi}(\tau)=\sqrt{\frac{\left|R^{-2}(\tau)\right|-1}{2N}}$$

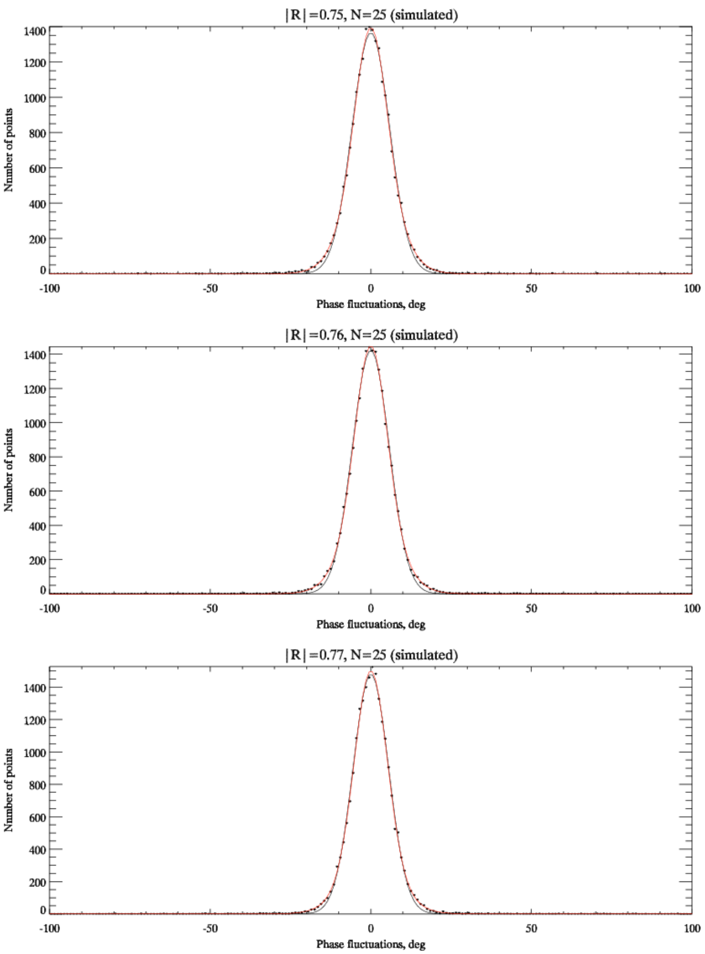




Simulated phase variance (N=25)

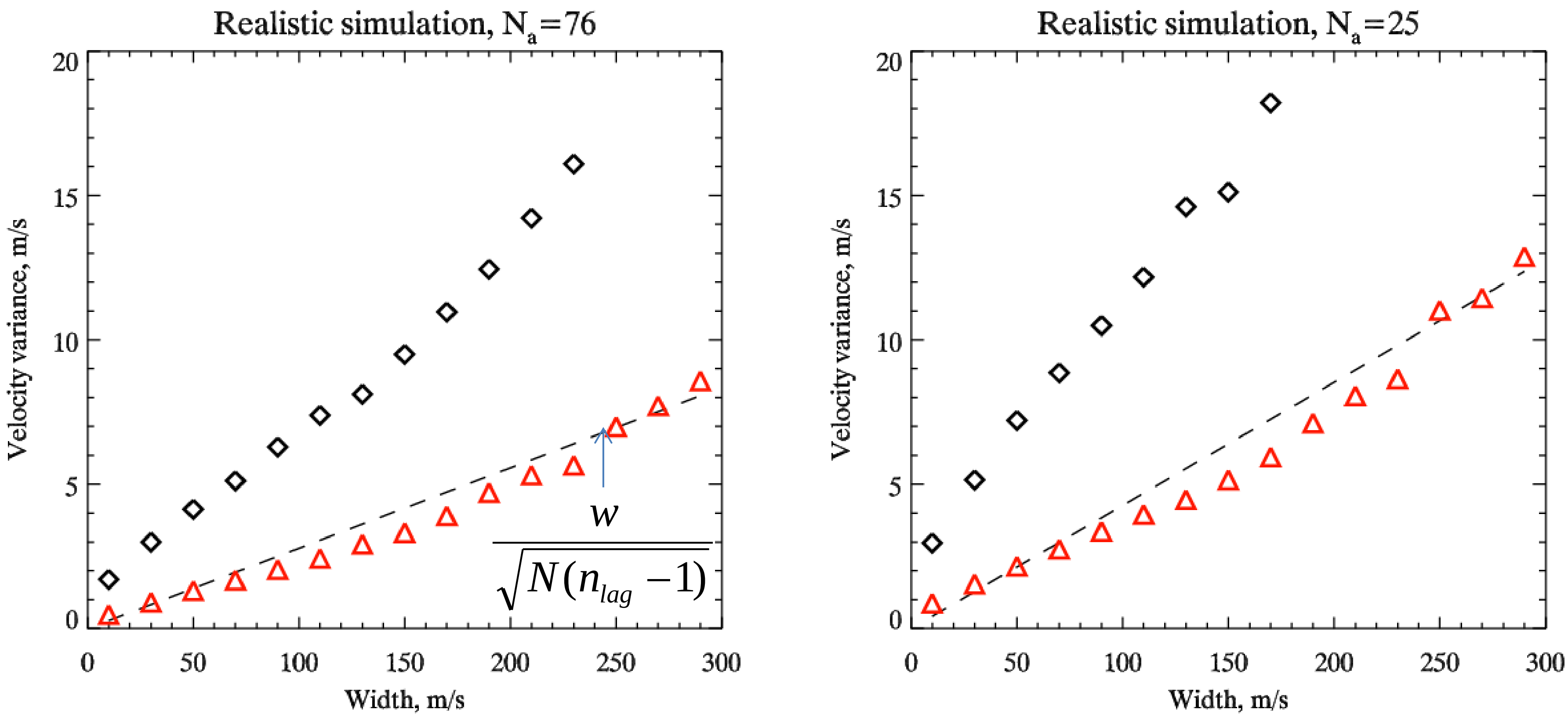
$$\sigma_{\varphi}(\tau) = \sqrt{\frac{\left|R^{-2}(\tau)\right| - 1}{2N}}$$





However, velocity variance is too large!

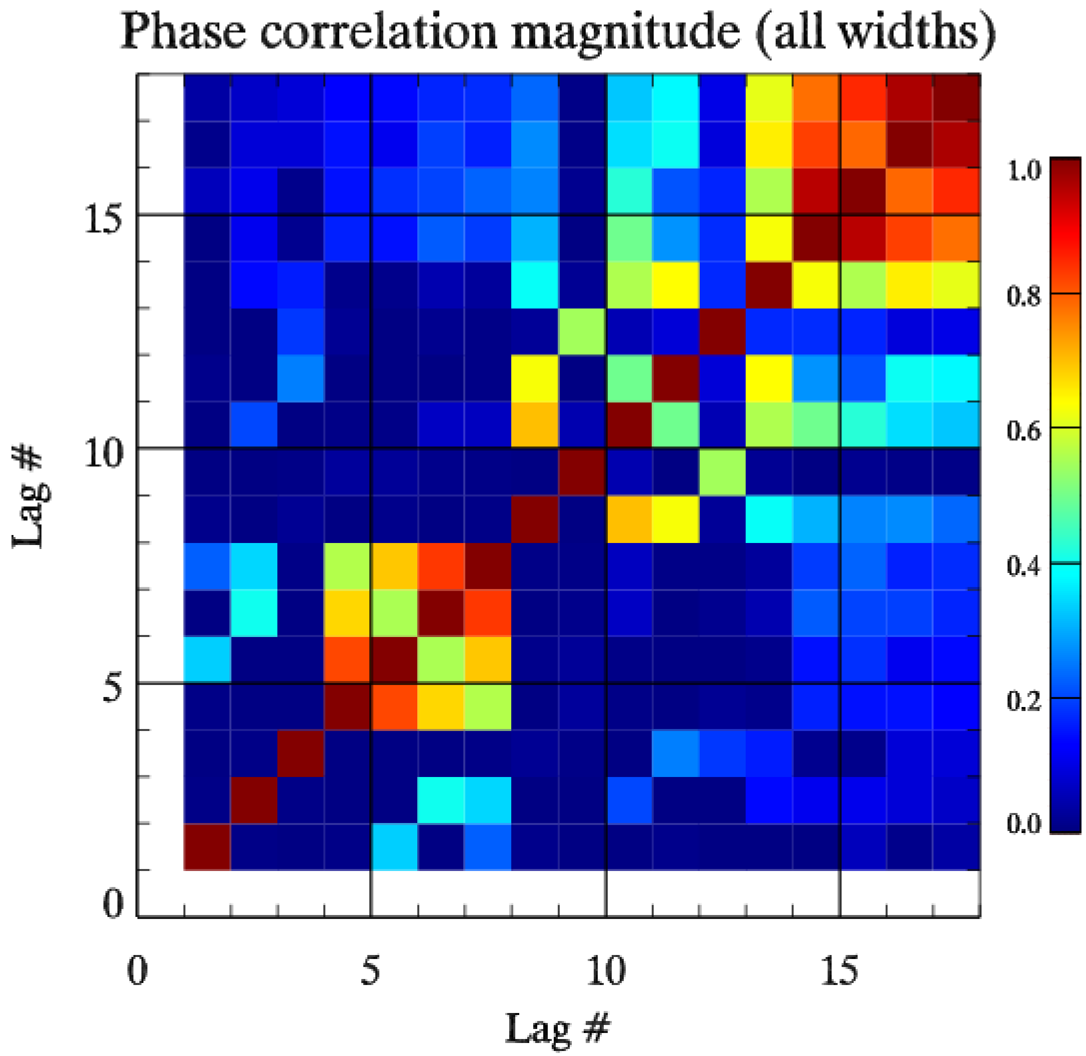
Black diamonds – measured deviation from input velocity, **red triangles** – FITACF errors



Correlated phase variations?

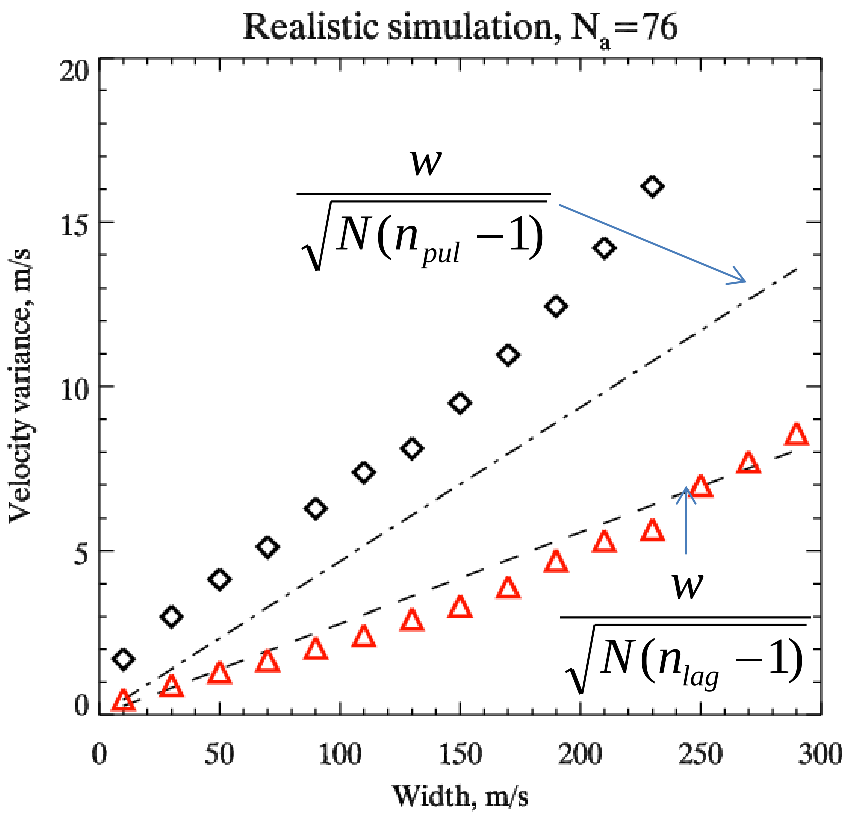
The explanation for this discrepancy is that we assumed uncorrelated phase variations at different lags. In fact, this may not be the case because the ACF lags are combined from only 7 or 8 independent samples (pulses).

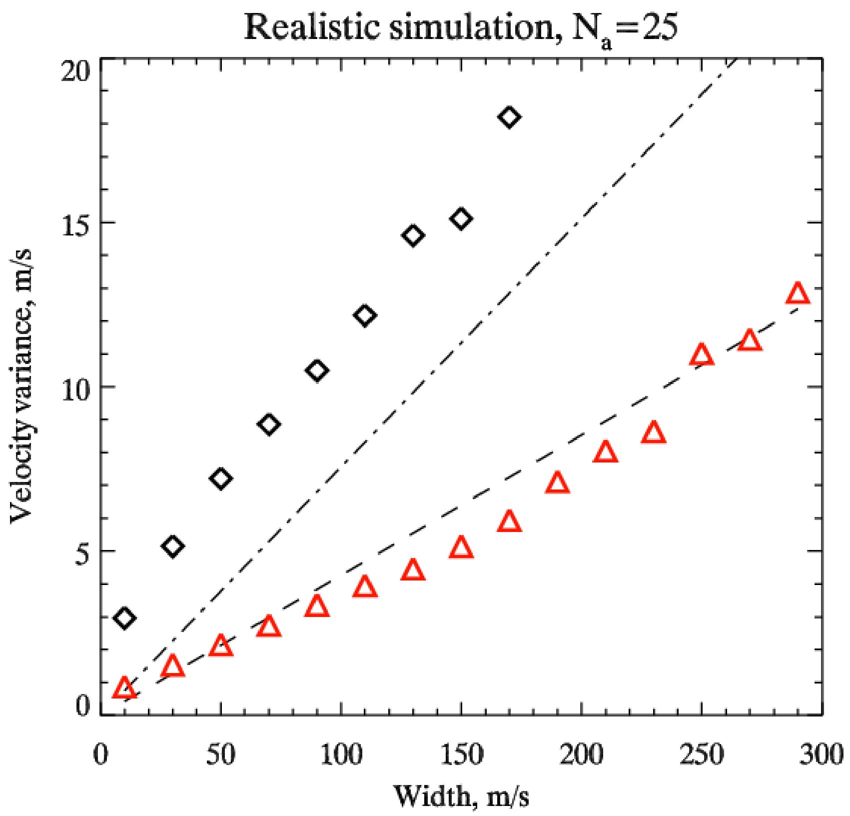
Therefore, the actual number of DoF is determined by the number of pulses, n_{pul}, but not the number of lags, n_{lag}. This assumption is supported by calculating correlations between the simulated phase variations at different lags.



Real number of DoF

Black diamonds – measured deviation from input velocity, **red triangles** – FITACF errors

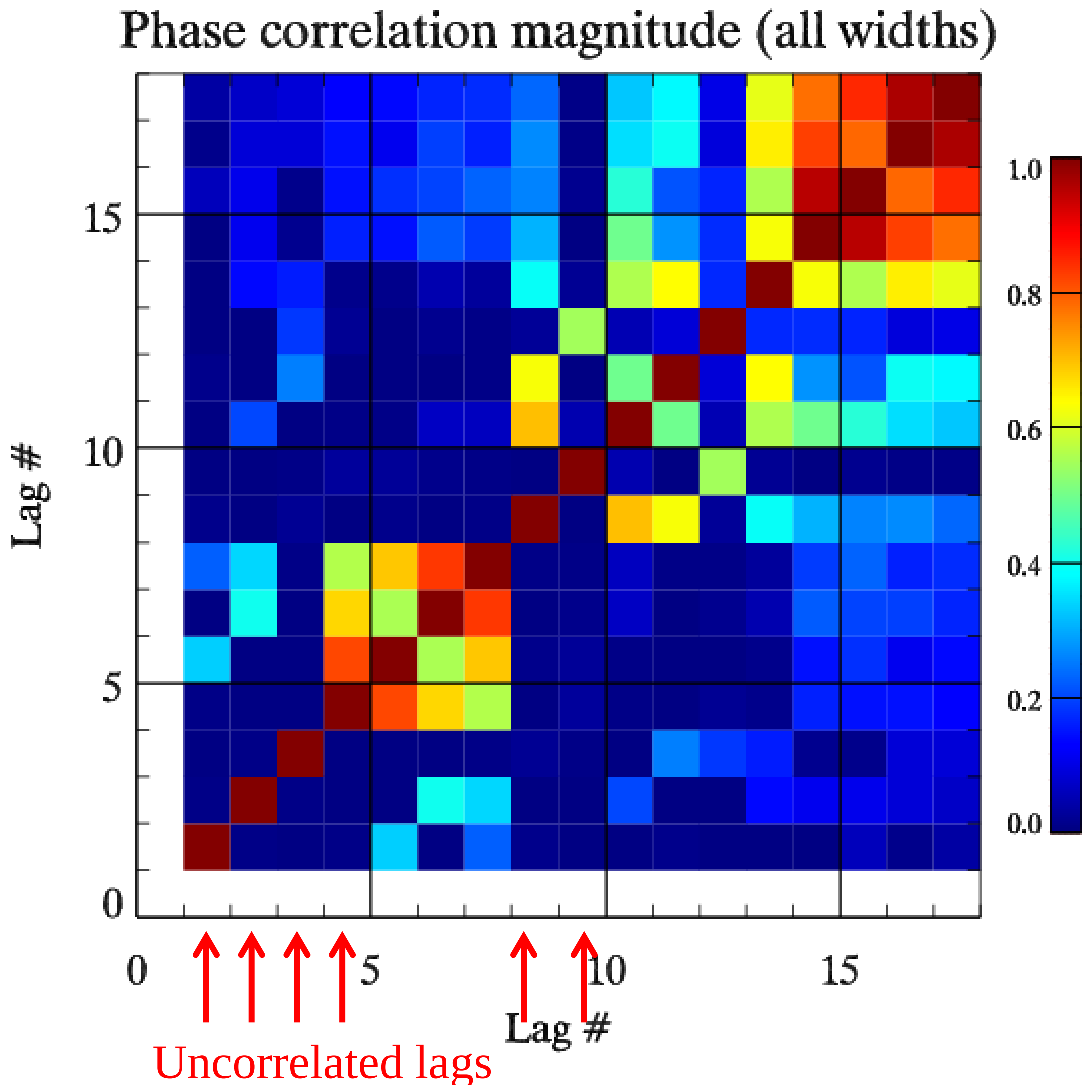




Looking for uncorrelated lags

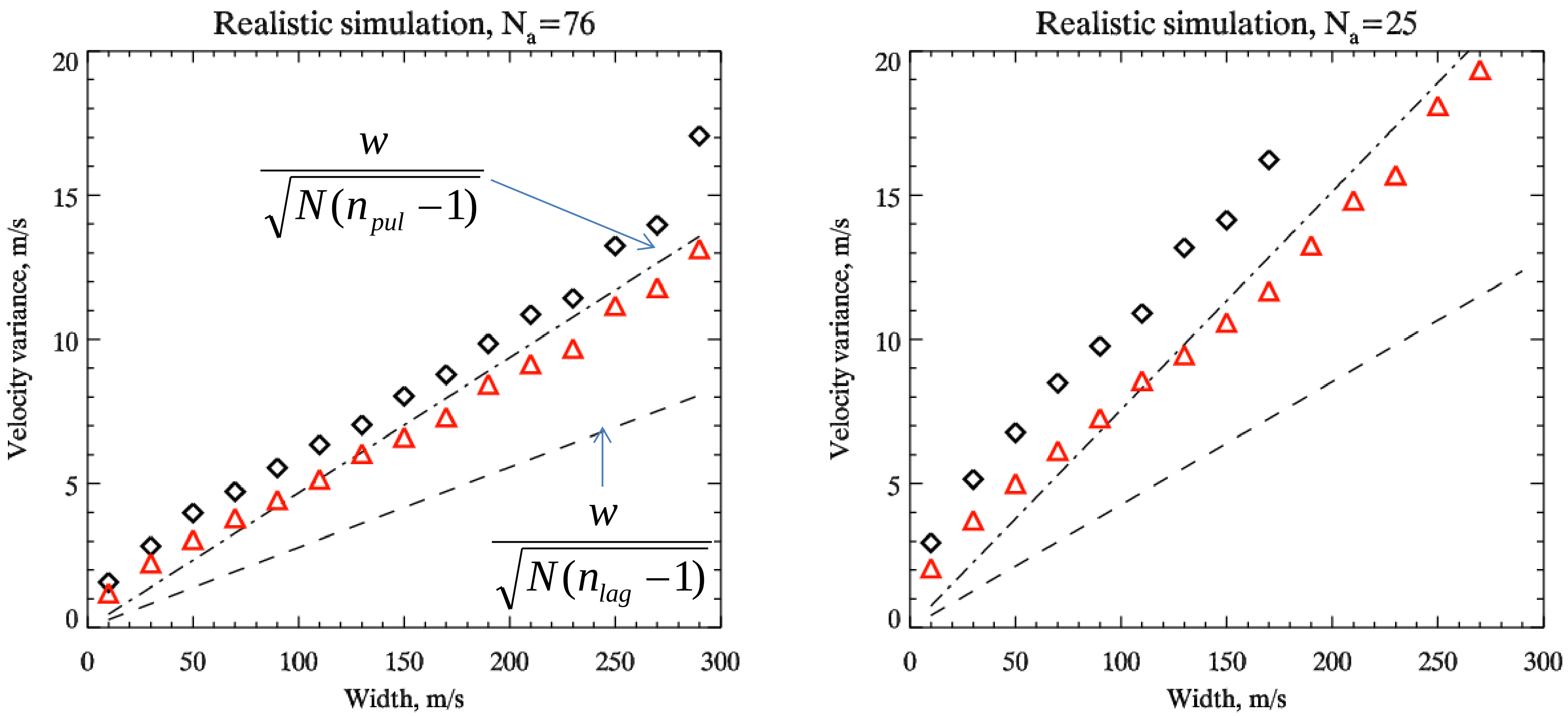
A close look at the phase correlation matrix reveals that there are n_{pul}-1 lags which are uncorrelated with each other. For the 7-pulse sequence these are lags 1,2,3,4,8 and 9.

What if we use only them for fitting?

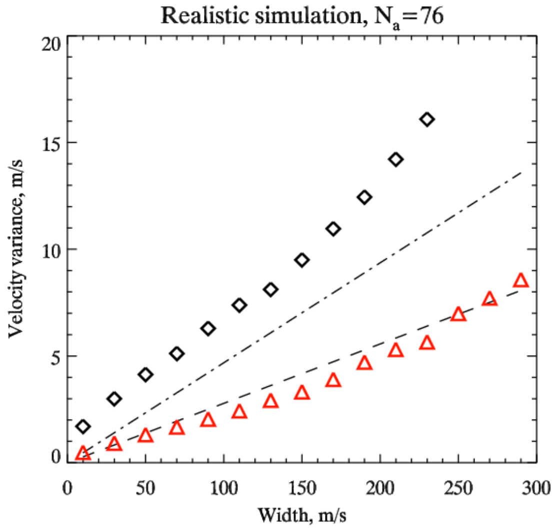


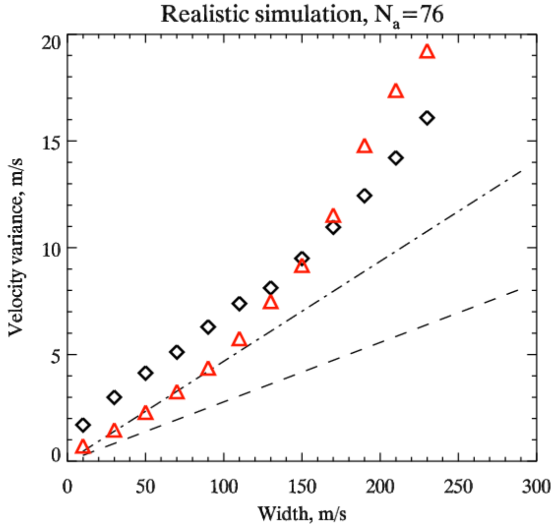
Uncorrelated lags only

Black diamonds – measured deviation from input velocity, **red triangles** – FITACF errors



How to fix the problem?





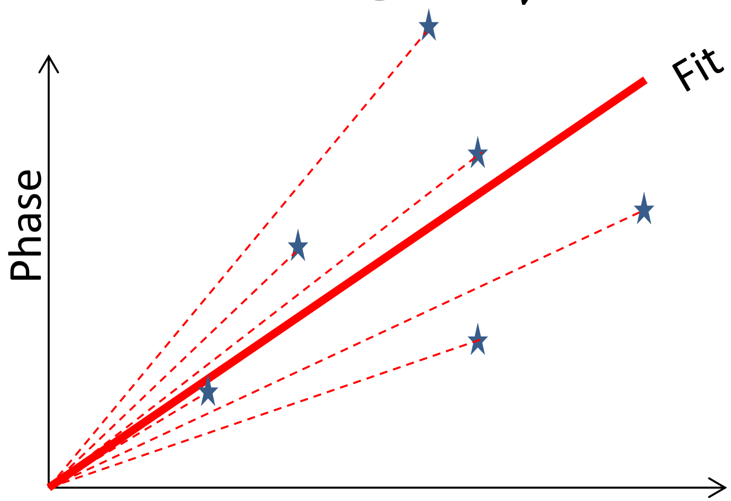
To use the effective number of DoF based on the number of pulses but proportional to the number of available lags

$$n_{DoF} = n_{pul} \frac{n_{lag}^{good}}{n_{lag}^{\max}} - 1$$

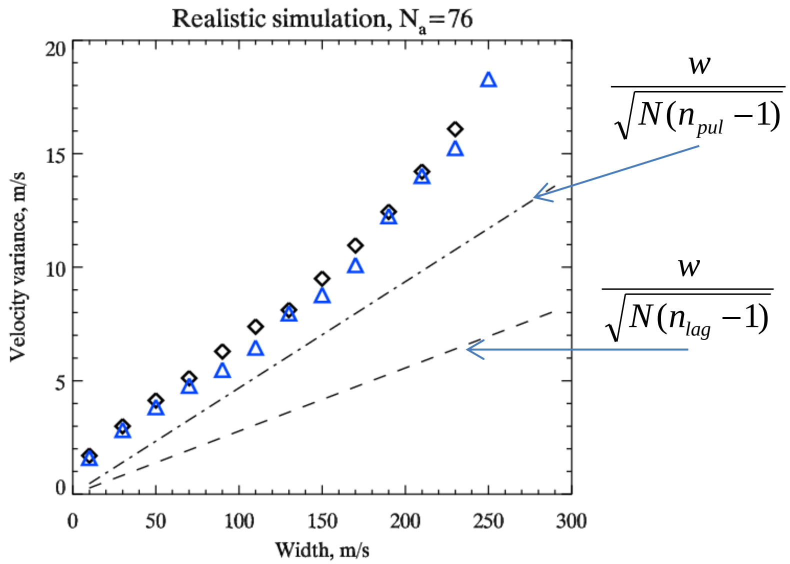
and use the maximum power instead of the mean power as a norm.

Direct way of measuring σ_v

We can also calculate velocity variance for each ACF directly from the slope estimate for each lag. Due to an approximately linear relation between the phase variance and time lag for $\tau \leq \tau_c$, we don't need to worry much about weighting.



In this case we also need to use n_{eff} instead of n_{lag}. The resulting estimate seems to be closer to the measured errors. In fact, it coincides with them if a factor of $\sqrt{2}$ is applied!!!

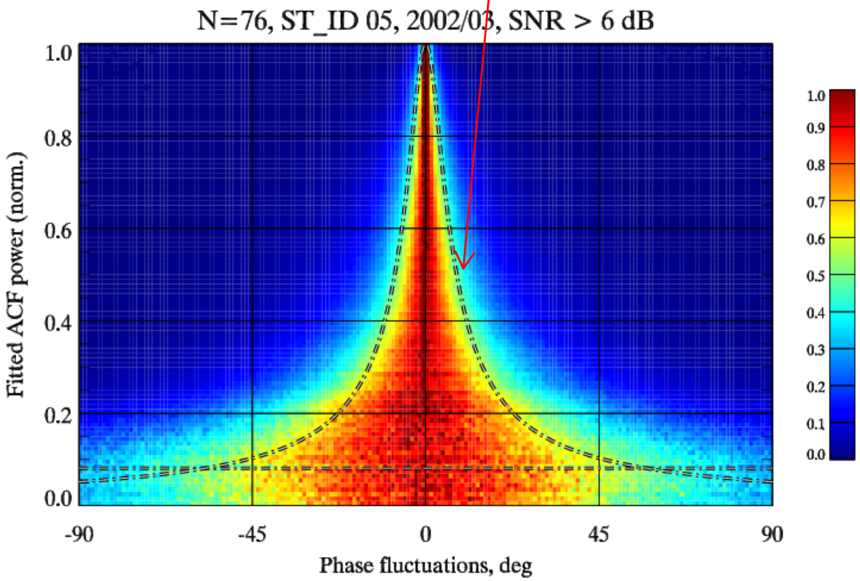


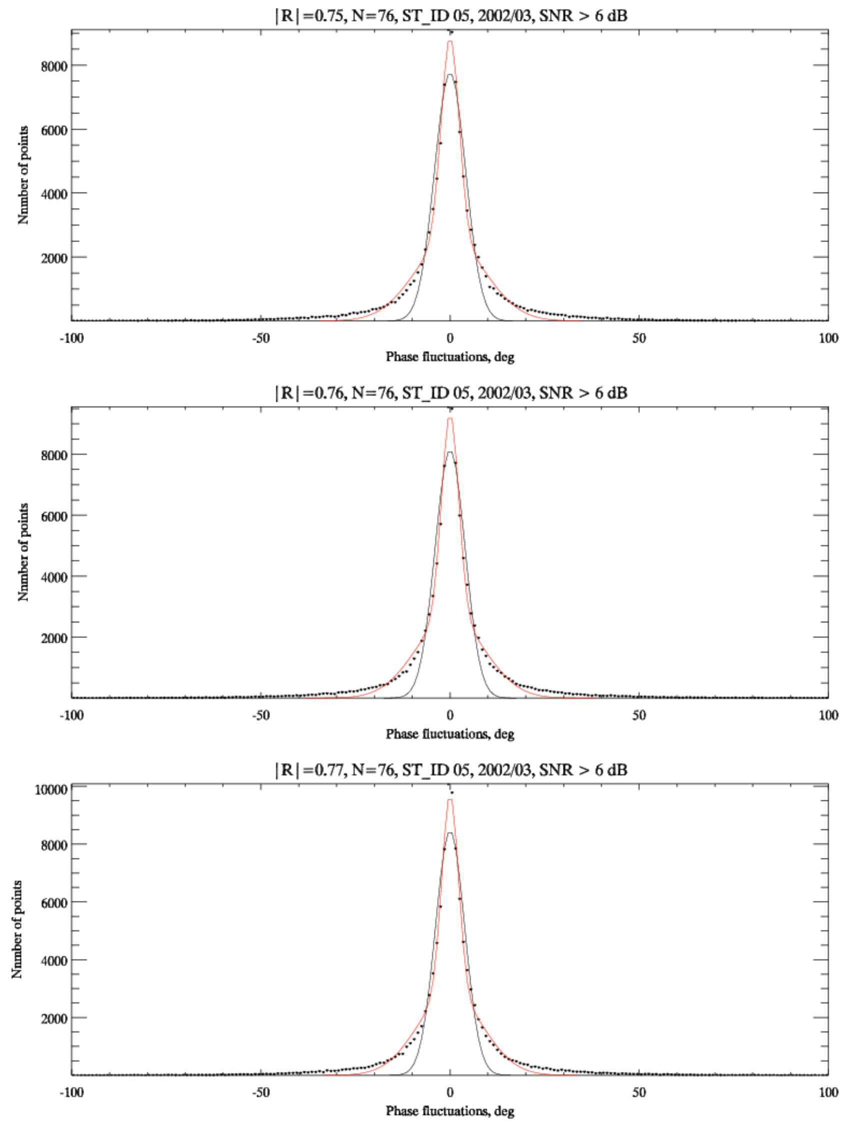
Phase variance in real data

- To test the theoretical expression for phase variance, we analysed deviations from the phase fit at all "good" lags for two month of Saskatoon data, March 2002 (N=76) and March 2012 (N=25). These data also differ by the pulse sequence: original (7 pulses) for 2002 and *katscan* (8 pulses) for 2012.

Real data: March 2002 (N=76)

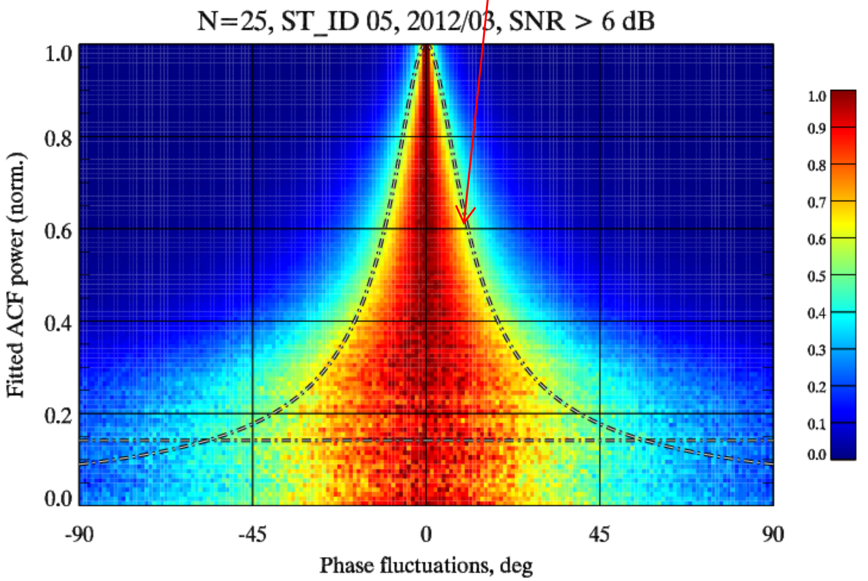
$$\sigma_{\varphi}(\tau)=\sqrt{\frac{\left|R^{-2}(\tau)\right|-1}{2N}}$$

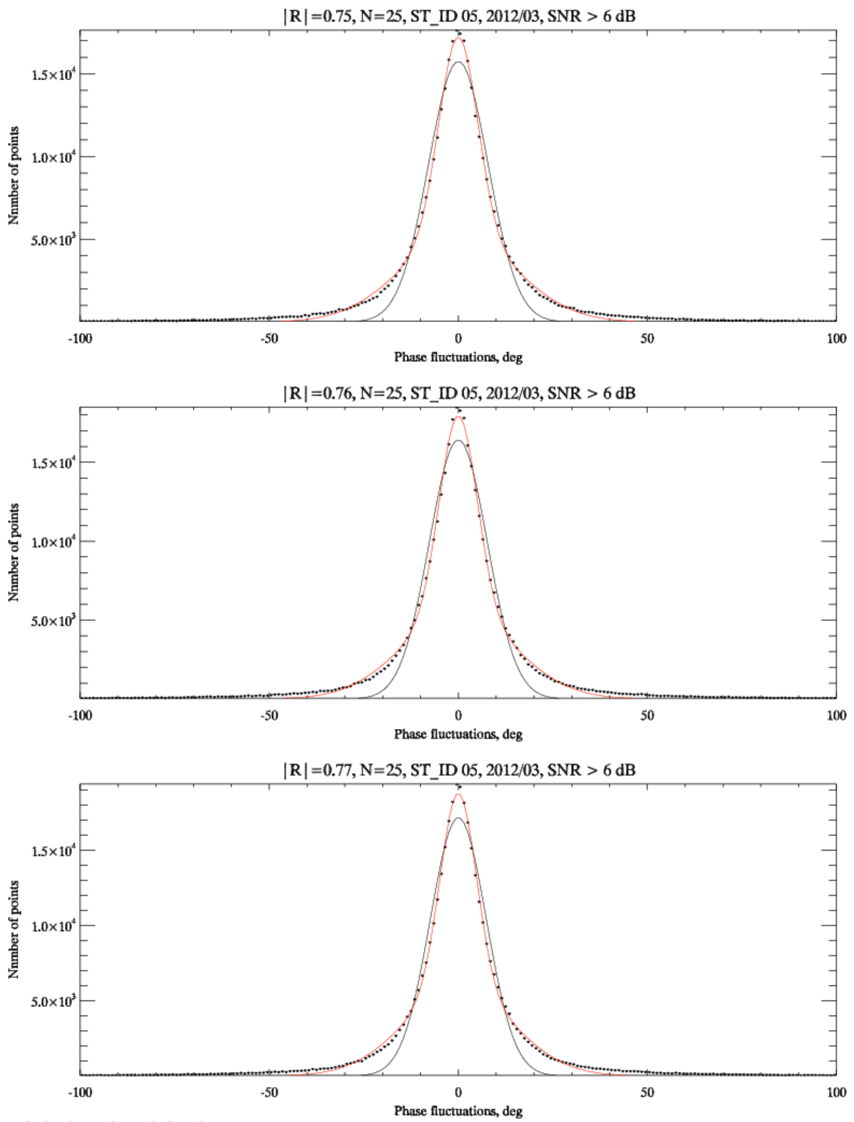




Real data: March 2012 (N=25)

$$\sigma_\varphi(\tau)=\sqrt{\frac{\left|R^{-2}(\tau)\right|-1}{2N}}$$





Reality generally agrees, but…

The experimental phase fluctuations agree with their theoretical estimates surprisingly well and allows us to assume that the theoretical description of the measured ACF phase fluctuations is adequate.

However, the phase distributions also exhibit long “tails” which indicate an additional source of variability.

Cross-range interference?

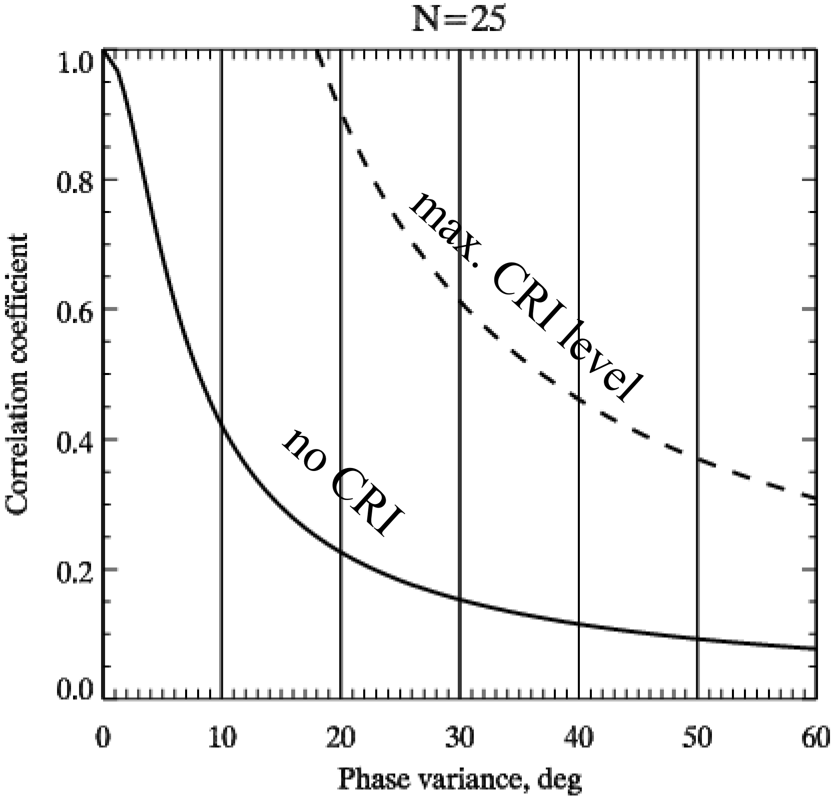
Incoherent noise contribution to phase variance

(*Bendat & Pierslol*)

$$\sigma_{\varphi i} = \sqrt{\frac{\alpha_i^{-2}\left|R(\tau_i)\right|^{-2} - 1}{2N}}$$

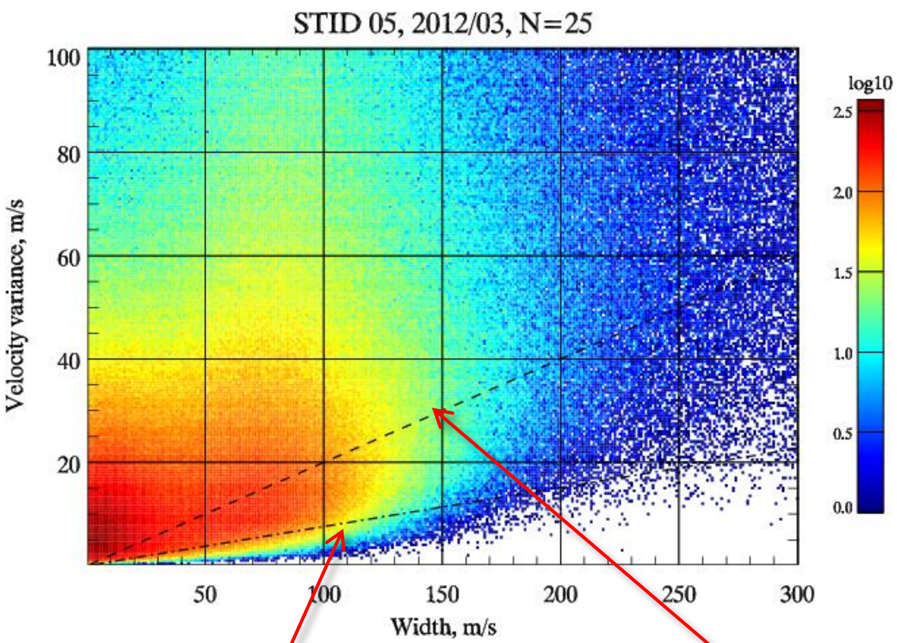
$$\alpha^2 = \frac{P_s^2}{(P_s + P_{\mathrm{int}1})(P_s + P_{\mathrm{int}2})}$$

$$\boxed{P_{\mathrm{int}1}, P_{\mathrm{int}2} \le P_s}$$

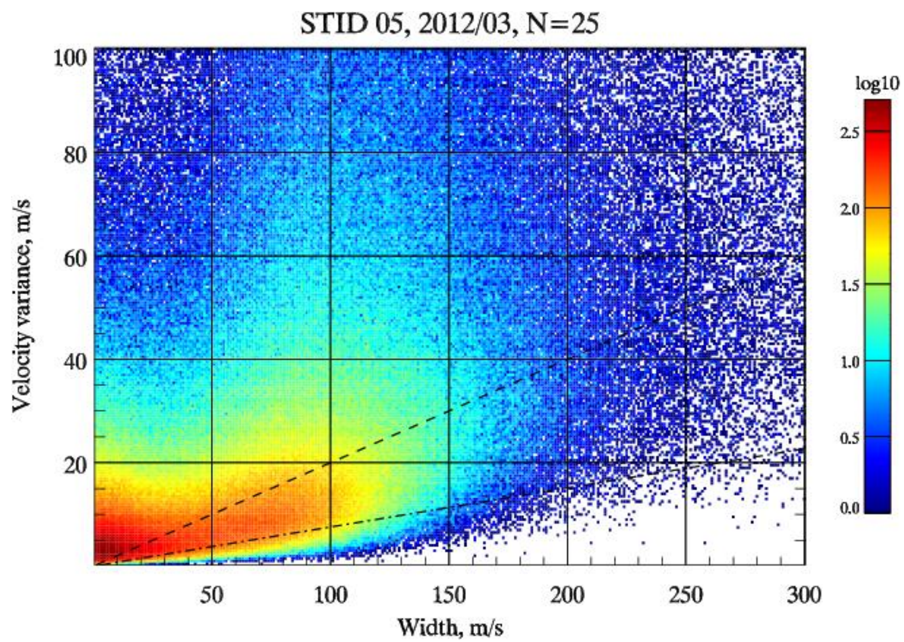


Effect of CRI on real data

Currently allowed CRI level (100%)



Reduced CRI level (10%)



$$\frac{w}{\sqrt{N(n_{pul}-1)}}$$

$$\frac{w}{\sqrt{N}}$$

Summary

- Current FITACF velocity errors (uncertainties) are significantly underestimated due to
  - errors is the code
  - incomplete description/understanding of the phase fluctuations.
- Two major methodology issues to address are:
  - **correlated phase fluctuations** at different lags (effective number of DoF)
  - **cross-range interference** effects on the phase (extra variance)
- It is also desirable to use the optimal (theoretical) weighting phase variance in actual phase fitting and error estimates.
- In a meantime, we can use $\sigma_v \approx W/\mathrm{sqrt}(N)$ as a "rule of thumb".
- Special measures have to be taken to minimise CRI.