Specifying the NSCAT MLE Expected Variance Using the After-the-Fit σ_o Residual

By: Frank J. Wentz

The maximum likelihood estimator (MLE) used by the NSCAT wind vector retrieval algorithm has the form

$$Q = \sum_{i=1}^{N} \frac{\sigma_{0i} - F(W, \phi_i, \theta_i, p_i)^2}{\varepsilon_i^2} + \ln(\varepsilon_i^2) \qquad (1)$$

where W is wind speed, ϕ_i is the relative wind direction, θ_i is the incidence angle, p_i is the polarization, and ε_i^2 is the expected variance between the model F and the observations σ_o. The relative wind direction ϕ_i equals the wind direction ϕ_w minus the observation azimuth direction ϕ_{oi}. The objective function Q is minimized by varying the wind speed W and direction ϕ_w until a minimum is found. The function $F(W,\phi,\theta,p)$ is called the 'model function' and is derived so that, on the average, it represents the observed dependence of σ_o on (W,ϕ,θ,p). The summation is over a group of σ_o measurements, typically within a 25-km cell ($N \approx 4$) or a 50-km cell ($N \approx 16$). The variance ε_i^2 is due to such things as σ_o measurement noise, geometry errors (i.e., error in specifying θ), and errors in the specification the geophysical model function F. Secondary geophysical parameters which are not included in the current model for F (i.e., sea-surface temperature, wind fetch, significant wave height, etc.) may contribute to ε_i^2.

The reason for including ε_i^2 in the MLE is to give the proper weight, or confidence, to each of the σ_o observations. For example, if the measurement noise for h-pol were greater than that for v-pol, then the h-pol observations should receive less weight (i.e., the h-pol ε_i^2 should be greater than the v-pol ε_i^2). Neglecting the $\ln(\varepsilon_i^2)$ term, which we think should be removed for the MLE formulation, one see that the important feature of ε_i^2 is its relative size among the observations, not its absolute size. For example, applying a constant scale factor to all ε_i^2 has no effect on the MLE solution once the $\ln(\varepsilon_i^2)$ term is removed. Thus the primary objective in specifying ε_i^2 is to obtain the correct relative weighting for the observations.

The term ε_i^2 can be partitioned into a sensor component (i.e., measurement and geometry errors) and a geophysical component (error in the geophysical model F). In principle, the sensor component can be estimated based on sensor performance specifications. However, there is some uncertainty about NSCAT's actual operational characteristics, and there remain questions about the ε_i^2 sensor component. The geophysical component of ε_i^2 is probably even more difficult to specify. How does one characterize unknown geophysical errors?

In an attempt to better understand the characteristics of ε_i^2, we do a statistical analysis of the NSCAT observations. One can lump all the errors that contribute to ε_i^2 can into a single term ε, and then write

$$\sigma_o = F(W,\phi,\theta,p) + \varepsilon \tag{2}$$

The variance of ε when averaged over a large number of observations is then

$$\langle \varepsilon^2 \rangle = \left\langle \left[\sigma_o - F(W,\phi,\theta,p) \right]^2 \right\rangle \tag{3}$$

where the brackets $\langle\ \rangle$ denote the averaging function. Hereafter, we drop the averaging brackets, and denote $\langle \varepsilon^2 \rangle$ as simply ε^2. The primary obstacle to evaluating the above expression is that the wind speed W and direction ϕ are not precisely known. Errors in specifying W and ϕ will add additional variation to the $(\sigma_o - F)$ difference, and one will obtain too large of a value for ε^2. Buoys, global circulation models, and other satellite wind estimates all lack sufficient accuracy and spatial/temporal coverage to specify W and ϕ to the accuracy required to reliably compute ε^2.

Ideally, one would like to compute ε^2 from just the NSCAT observations alone. One quantity that is easily computed is the 'after-the-fit' residual defined by

$$\hat{\varepsilon}^2 = \frac{N}{N-2} \left\langle \left[\sigma_o - F(\hat{W},\hat{\phi},\theta,p) \right]^2 \right\rangle \tag{4}$$

which is the same as (3) except that the true wind speed and direction have been replaced the estimates $\hat{W}$ and $\hat{\phi}$ coming from the NSCAT wind vector retrieval algorithm. Also (4) contains a leading factor that accounts for the fact that N observations are used to find 2 unknowns, and hence the degrees of freedom are N – 2, not N. For example, if only two σ_o observations were used to retrieve $\hat{W}$ and $\hat{\phi}$, then the wind vector retrieval algorithm would find $\hat{W}$ and $\hat{\phi}$ values that perfectly matched the two σ_o observations (i.e., two observations and two unknowns result in a perfect fit). In this case, one can not estimate the after-the-fit residual. However, for N >2, an estimate can be obtained. Ideally one would like to have a large N, but if N is too large, spatial wind gradients over the ocean become a problem. For this analysis, we use the 50-km NSCAT cell data for which N typically equals 16. We think this is a good compromise. N is large enough to obtain a reliable estimate of $\hat{\varepsilon}^2$, while wind gradients are not be problematic in most cases.

If ε were an uncorrelated, random variable with a mean of zero, then $\hat{\varepsilon}^2$ would be a good estimate of the variance term in the MLE. However, this is probably not the case. The ε values for the N observations are most likely correlated. For example, unmodeled geophysical parameters such as sea-surface temperature, fetch, or significant wave height could possibly bias all N σ_o observations high relative to the standard model F. Spacecraft attitude errors will also produced correlated values of ε. These correlated errors will be 'absorbed' by the wind vector retrieval algorithm and will end up as systematic errors in $\hat{W}$ and $\hat{\phi}$. If ε were such that it increased all N σ_o observations by 5%, then $\hat{W}$ would also be too high so as to zero out the σ_o - F difference. As a consequence, correlated errors do not fully contribute to the after-the-fit residual. In other words, the after-the-fit residual is mostly

a measure of the error component that is uncorrelated among the N observations. With this caveat in mind, we proceed to compute $\hat{\varepsilon}^2$ using equation (4).

The procedure for computing $\hat{\varepsilon}^2$ is to average over the first two months of NSCAT data. (We plan to use the full 10-month data set for the final analysis.) Wind retrievals are done at the 50-km resolution to find $\hat{W}$ and $\hat{\phi}$, and averages are then done according to equation (4). Three separate averages are done for the fore and aft v-pol observations (j=1), for the middle h-pol observations (j=2), and for the middle v-pol observations (j=3). The averages are stratified according to incidence angle and the value of the model function F evaluated at $\hat{W}$ and $\hat{\phi}$. There are 26 incidence angle bins going from 16° to 66° in 2° steps. The F-bins are every 0.1 dB. Binning the averages in this way allows us the specify $\hat{\varepsilon}^2$ as a function of antenna type j, incidence angle θ, and the σ_o value of the model function F. We denote this σ_o variance function by

$$\hat{\varepsilon}^2 = V_j(\theta, F) \tag{5}$$

In addition to the usual quality control that is applied during NSCAT data processing, we place two further restrictions on the data. First, an SSM/I rain flag is computed for each NSCAT 50-km cell. If the SSM/I indicates a high likelihood of rain, the NSCAT cell is skipped. The second restriction pertains to the wind field spatial gradient. If either NSCAT or ECMWF indicate that either the zonal or meridional component of the wind changes by more than 3 m/s across the 50-km cell, then the cell is skipped. These restrictions are intended to remove unwanted σ_o variability that would bias $\hat{\varepsilon}^2$ high.

For NSCAT, the usual form for the variance function has been

$$V_j(\theta, F) = \alpha_j(\theta)F^2 + \beta_j(\theta)F + \gamma_j(\theta) \tag{6}$$

We find that this form fits our table of $\hat{\varepsilon}^2$ values reasonable well, and we will continue to use this form for the σ_o variance function. Analysis of the $\hat{\varepsilon}^2$ values shows that the following forms are suitable representations for the α, β, γ coefficients.

$$\alpha = (a_0 + a_1\theta)^2 \tag{7}$$

$$\beta = A_{0,7}(\theta)(b_0 + b_1\theta + b_2\theta^2)^2 \tag{8}$$

$$\gamma = A^2_{0,7}(\theta)(c_0 + c_1\theta + c_2\theta^2)^2 \tag{9}$$

where we have dropped the subscript j and it is understood that a different set of coefficients is found for each antenna type. The function $A_{0,7}(\theta)$ is the A_0 coefficient is the σ_o model function evaluated at W=7.5 m/s. The v-pol (h-pol) A_0 is used for the v-pol (h-pol) antenna. This function is included because we find that to first order the variation of β (γ) versus θ is similar the A_0 (A_0^2) variation with θ. Figures 1, 2, and 3 show the following three polynomials: $a_0 + a_1\theta$, $b_0 + b_1\theta + b_2\theta^2$, $c_0 + c_1\theta + c_2\theta^2$. Note that the beta polynomial for mid v-pol has been linearly extrapolated for θ > 48° (a small cosmetic fix). The alpha polynomial for the mid v-pol is the same as that for the fore,aft v-pol.

Figures 4, 5, and 6 show the following two quantities Y_1 (red curve) and Y_2 (black curve) ploted versus X.

$$Y_1 = \frac{\sqrt{\hat{\varepsilon}^2}}{A_{0,5}(\theta)} \tag{10}$$

$$Y_2 = \frac{\sqrt{V_j(\theta,F)}}{A_{0,5}(\theta)} \tag{11}$$

$$X = \frac{F}{A_{0,5}(\theta)} \tag{12}$$

All three quantities are normalized by the A_0 value at 5 m/s. The red Y_1 curves show the increase in the after-the-fit residual with F. The black Y_2 curves are our best fits to the after-the-fit residuals, i.e., equations (6) through (9). There are 26 black curves corresponding to the 26 θ-bins. Not all θ-bins have data, and hence some black curves do not have red curves associated with them. In order to show all curves simultaneously, the curves are offset by the amount 0.05(k-1), where k is the θ-bin number that goes from 1 to 26. As can be seen, the polynomial expressions fit the observed residuals reasonably well.

A 3×26 table of α, β, and γ are produced. The first dimension corresponds to the antenna type (fore,aft v-pol, mid h-pol, and mid v-pol), and the second dimension corresponds to incidence angle (16°, 18°, …, 66°). These tables have been sent to the NSCAT project for analysis and testing.

The question remains as to what should be done about specifying $\varepsilon^{2.}$ for the MLE. Currently, the sensor component for ε^2 is first computed, and then a 16% geophysical modeling error is added (i.e., α is multiplied by 1.0256). In computing the sensor component, the geometry error (i.e., K_{PR}) is 14%, but probably a smaller value around 7% would be a better choice (Wu-Yang, private communication). The 16% modeling error is a guesstimate based on a brief analysis of aircraft data. The estimate of ε^2 coming from the after-the-fit residuals has the advantage of being an objective, statistical measurement, but it is an underestimation of the true ε^2 because correlated errors go undetected. However, it may be better to use the after-the-fit residuals and acknowledge the underestimation problem rather than continuing to use a 16% geophysical modeling error that dominates the errors statistics and has little basis of support.

We would like to make one final comment on how the results from the after-the-fit residuals compared to the assumption of a constant 16% geophysical modeling error. To obtain an estimate of the modeling error from the after-the-fit residuals, the ε^2 sensor component must be subtracted. We estimate the ε^2 sensor component using the standard NSCAT K_P algorithm, but we reduced K_{PR} from 14% to 7%. This sensor component is then subtracted (in an rss sense) from the after-the-fit residuals. A brief analysis of these results reveals the following. The resulting modeling error varies considerably with F and θ. At high F (i.e., high winds), the modeling error is usually small (< 5%). For moderately high θ and for F values corresponding to winds near 5 m/s, a modeling error near 16% seems reasonable. Meanwhile, for low F (3 to 5 m/s) and high θ, the h-pol modeling error is quite

large (40%). All of these results, of course, depend on the ε^2 sensor component being correct, which may not be the case. In any event, it does appear likely that the modeling error has a significant dependence on F and θ and cannot be simply represented by a single number.

More work needs to be done to partition the after-the-fit residuals into a sensor error component and a modeling error component. Also, we need to better understand the consequences to the MLE of neglecting the correlated errors in ε^2. Finally, there is the interesting possibility that one could directly compute the error correlation among the various antennas using the following expression.

$$\hat{\varepsilon}^2_{ij} = \frac{N}{N-2}\left\langle \left[\sigma_{oi} - F(U,\hat{\phi}_i,\theta_i,p_i)\right]\left[\sigma_{oj} - F(U,\hat{\phi}_j,\theta_j,p_j)\right]\right\rangle \qquad (13)$$

where subscripts i and j denote two separate antennas. This may be a way to bring correlated errors into our estimate of ε^2. If the error correlation can be inferred from after-the-fit residuals, then the MLE will need to be modified to include the error covariance matrix. Fortunately, this is a straightforward modification. There will still remain of the problem of a bias error (i.e., a correlated error that is the same for all antennas). However, the impact of bias errors on the MLE may be of secondary importance compared to the inter-antenna correlation errors.

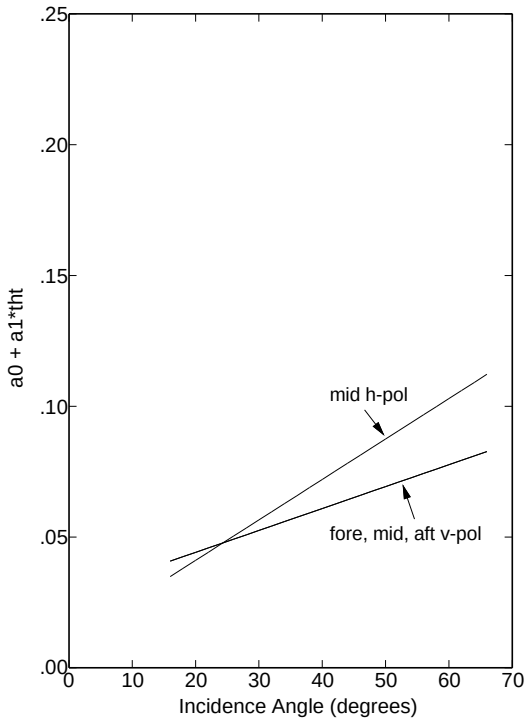

Figure 1. Polynomial fit for the alpha term in the variance function

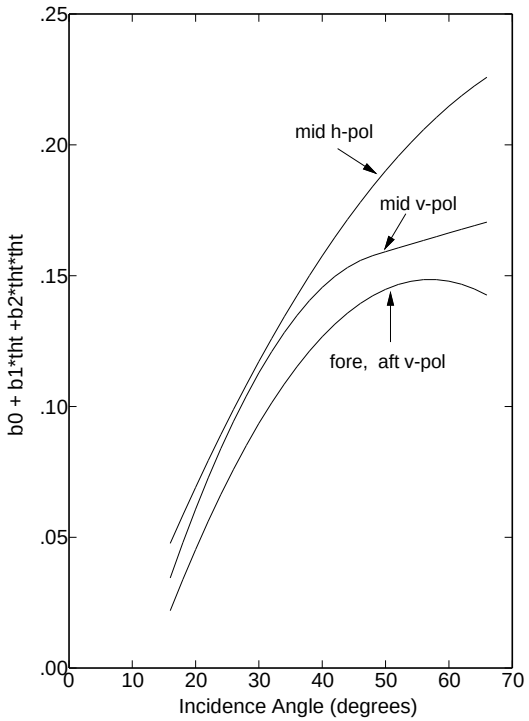

Figure 2. Polynomial fit for the beta term in the variance function

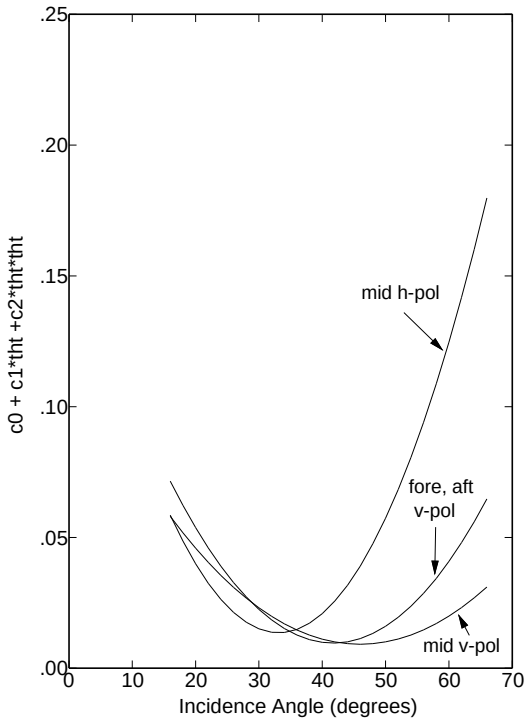

Figure 3. Polynomial fit for the gamma term in the variance function

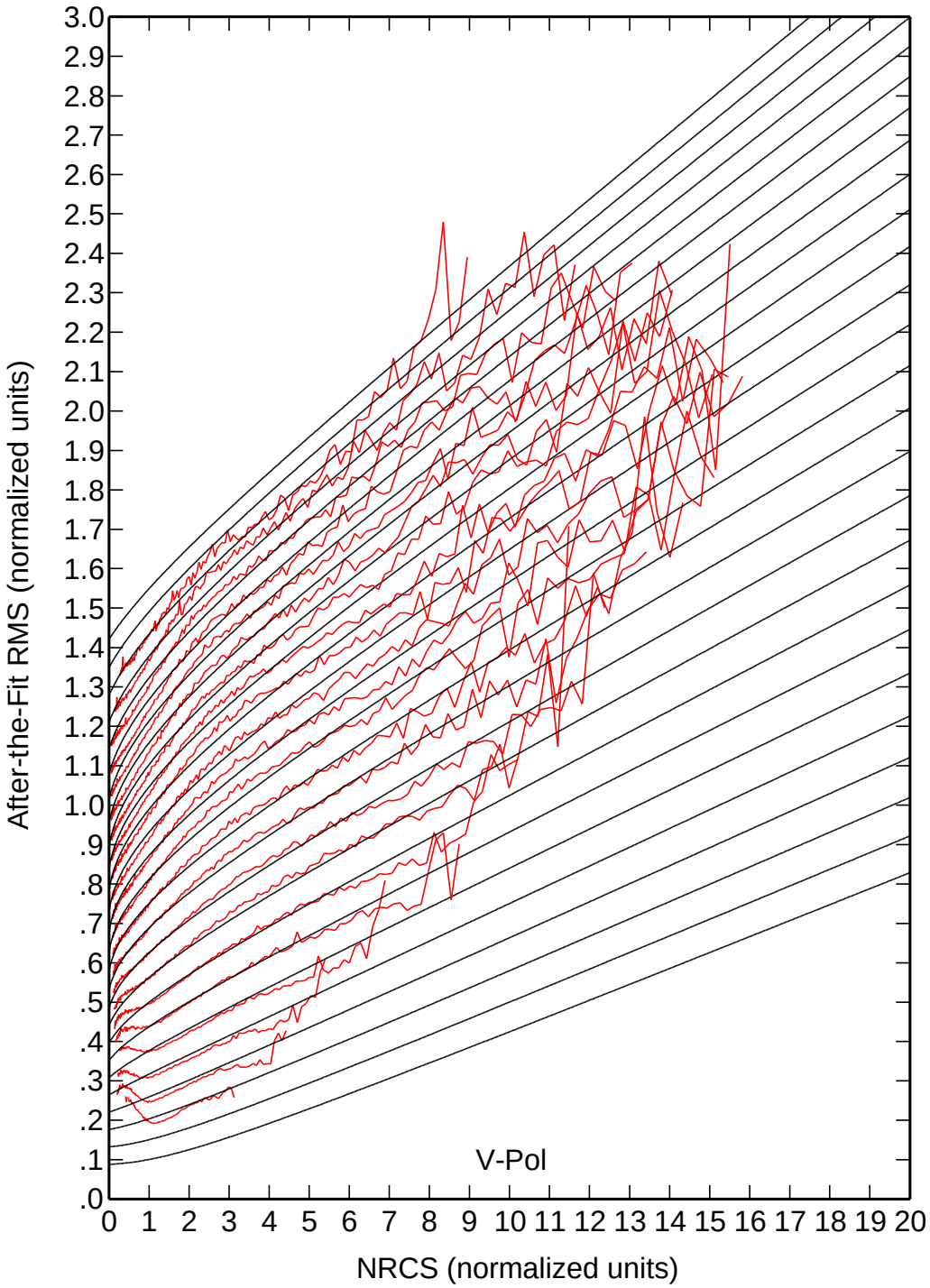

Fig. 4. After-the-fit rms residuals plotted versus σ_o for the fore and aft v-pol antennas. The red curves are the observations and the black curves are best fits. Each black curve corresponds to an incidence angle, going from 16° (bottom) to 66° (top) in 2° steps. The curves are offset by 0.05 so that all incidence angles can be shown simultaneously. Both the residual and σ_o have been normalized by the σ_o value for a 5 m/s wind.

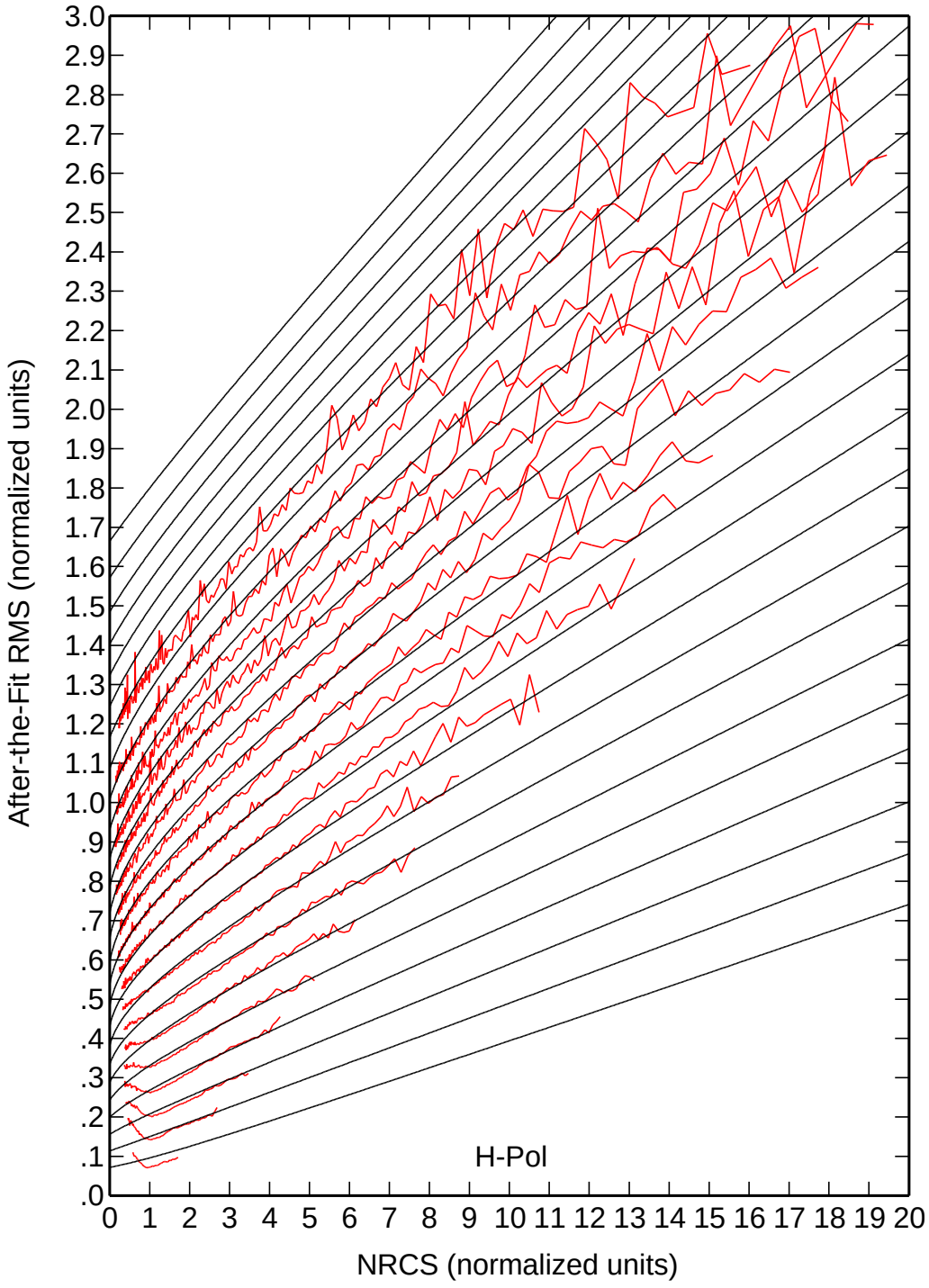

Fig. 5. Same as figure 4 except this figure is for the middle h-pol antennas.

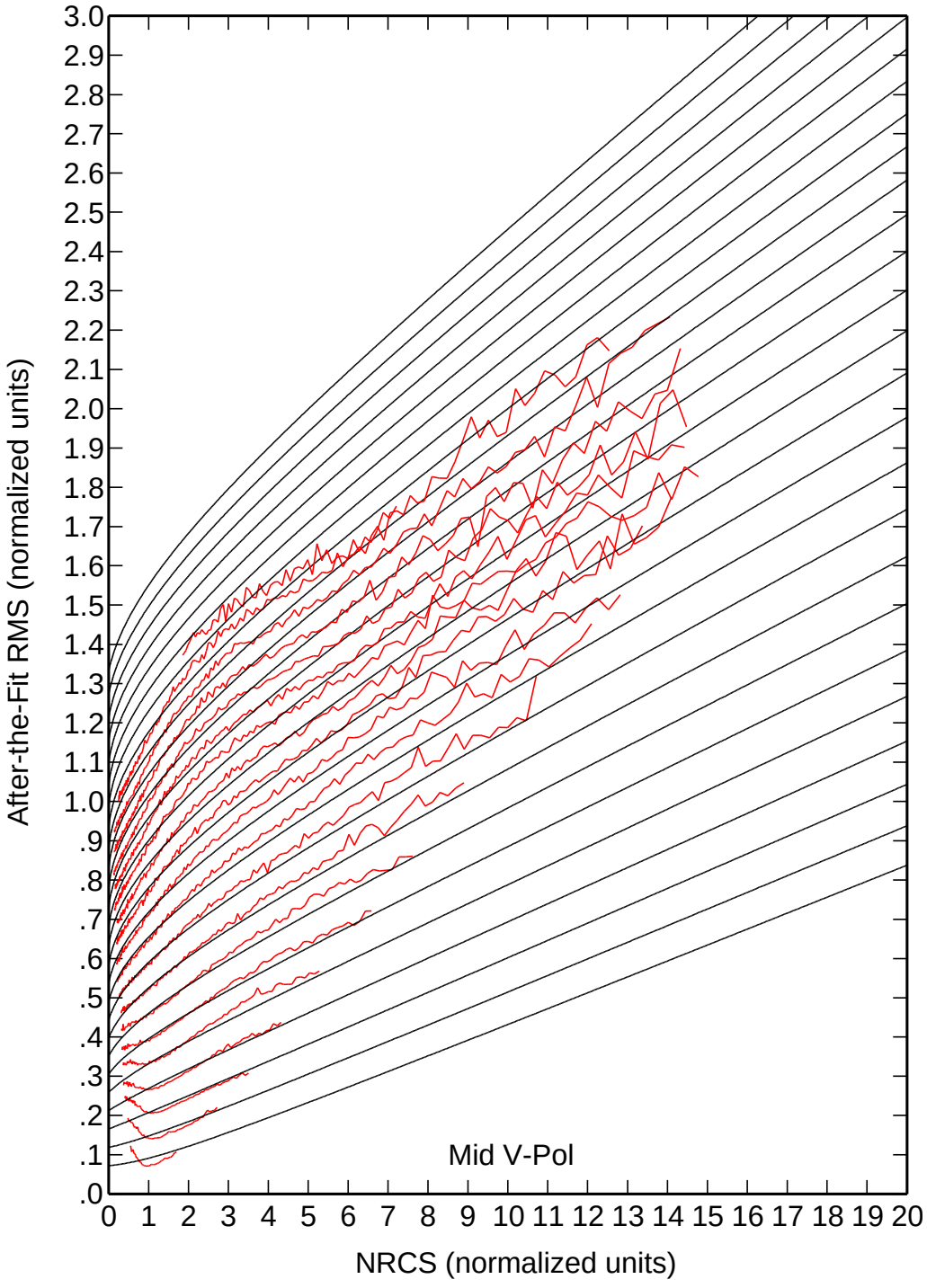

Fig. 6. Same as figure 4 expect this figure is for the middle v-pol antennas.