The Probability Distribution of NSCAT Measurements

Paul E. Johnson and David G. Long
Brigham Young University, MERS Laboratory
459 CB, Provo, UT 84602
801-378-4884, FAX: 801-378-6586 johnsonp@ee.byu.edu

ABSTRACT

Due to the on-board signal processing used by NSCAT, the Gaussian distribution model for the power measurements used in the wind retrieval algorithm is only an approximation to the actual distribution. Working from first principles and the design of the NSCAT signal processor we derive the distribution of the NSCAT measurements as a function of the normalized radar cross section, NRCS, the signal to noise ratio and the cell number. The resulting distribution is skewed relative to the traditional Gaussian model. Simple compass simulations are used to compare the accuracy of winds estimated using the actual and Gaussian model distributions. These results are readily generalizable to other scatterometers and applications involving spectral estimation from averaging modified periodograms.

INTRODUCTION

Wind retrieval with NSCAT requires the estimation of the backscatter from each resolution element or cell. Resolution of the antenna footprint into ocean cells is done by Doppler filtering, in which the periodogram of the received signal is computed and several frequency bins summed [1]. Welch's method for periodogram estimation [2] is used with 50% overlapping data segments and a Hann window applied to minimize spectral leakage [3]. The signal power for a cell is then computed by subtracting a noise-only power estimate from a signal-plus-noise power estimate (the received signal is corrupted by radiometric noise). This signal power, used in the wind estimation, has been assumed to have a Gaussian distribution. However, Welch's periodogram estimate is not Gaussian [4]. In this paper, the probability density function of NSCAT power estimates is found, and the effect on wind retrieval is considered.

The pdf of the sum of frequency bins, based on Welch's method with K overlapping data segments, and an arbitrary data window, is briefly explained, with some simple verifying simulations [4]. The pdfs of a noise-only measurement and of a signal-plus-noise measurement only require simple scale changes from basic pdfs; the pdf of the signal-only measurement is then found as the correlation of these two pdfs. Finally, compass simulation results are reported to suggest the minimal impact of using the incorrect pdf on wind retrieval.

VECTOR SPACE ANALYSIS OF WELCH'S SPECTRUM ESTIMATION

The power in a frequency band of Welch's modified periodogram technique, corresponding to a NRCS cell, can be written as a quadratic form in the random vector $\mathbf{x}$:

$$P = \mathbf{x}^T \Upsilon \mathbf{x} \qquad (1)$$

where the matrix Υ depends on the DFT size, the data window, the frequency bins of interest, the number of periodograms to average, and the amount of overlap of the data segments [4]. The random vector, $\mathbf{x}$, is assumed to be normally distributed $\mathrm{N}(0, \mathbf{R})$ for the scatterometer problem because of the large number of independent scattering centers on the ocean surface.

The moment-generating function of a quadratic form in a Gaussian vector, such as Eq. (1), can be expressed as [5]

$$M(t) = \prod_{i=1}^{D} (1 - t2\eta_i)^{-\frac{\nu_i}{2}} \qquad (2)$$

where the η_i are the D distinct, non-zero eigenvalues of $\mathbf{R}\Upsilon$. Each eigenvalue has multiplicity ν_i. If the multiplicity of all of the non-zero eigenvalues have even multiplicities, $M(t)$ can be expanded with a partial fraction expansion (if not, eigenvalues can be clustered into groups with even multiplicity with minimal numerical error). The density function is then [4]:

$$f_P(p) = g \sum_{i=1}^{D} \sum_{j=1}^{\nu_i/2} A_{ij} \frac{p^{j-1}}{(j-1)!} \exp \frac{-p}{2\eta_i} U(p), \qquad (3)$$

where the A_{ij} are coefficients from the partial fraction expansion, g is an appropriate scale factor (a simple function of the eigenvalues), and $U(p)$ is the step function. Note that changing the power of $\mathbf{x}$ simply scales the pdf; that is, if $\mathbf{x}_0$ has a covariance matrix $\mathbf{R}_0$ and $\mathbf{x}_1$ has a covariance matrix $a\mathbf{R}_0$, then the pdf of the power based on $\mathbf{x}_1$ is simply a times the pdf of the power based on $\mathbf{x}_0$.

To validate the pdf of Eq. 3, Welch's method was applied to a series of pseudo-random data sequences, a sample pdf of the power in a frequency band was determined. For each case, data segments of length $L = 2^{16}$ were used. Figure 1 displays comparisons between the sample density functions and the pdf computed based on Eq. (3). The simulations used more than 6500 estimates to obtain

0-7803-4403-0/98/$10.00 © 1998 IEEE

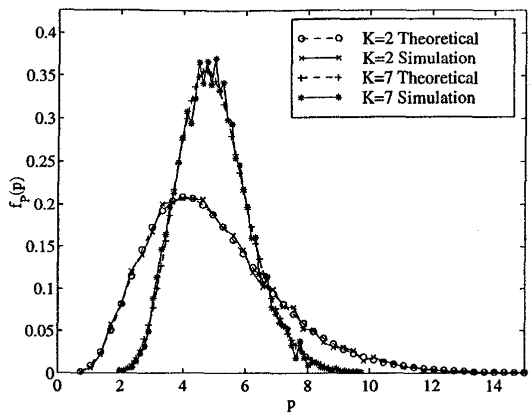

Figure 1: *Comparison between simulations and the derived probability density functions for the power in 5 frequency bins, using a Hann window with 50% overlap. The simulations compare very well with the theoretical densities and show considerable skew.*

the empirical density functions. Each of the cases demonstrates strong correlation between the simulation and the theory, and considerable skew in the densities.

PDF OF NSCAT MEASUREMENTS

For NSCAT, two power estimates are made, a signal-plus-noise measurement and a noise-only measurement. The power in the signal is estimated as the difference between these two power estimates: $P = P_1 - P_2$. Because P_1 and P_2 are independent random variables, distributed as described in the previous section, the pdf of the signal-only power is the convolution of the first pdf with that of the second pdf with a negative argument.

Based on the NSCAT processing system, estimation of P_1 from a particular cell can be based on 2, 3 or 7 overlapping segments and from 5 to 20 frequency bins, and is improved by averaging 25 pulses [3]. Similarly, the noise power estimate, P_2, averages 4 'pulses' (measurements with no transmitted power) and sums more than 200 frequency bins. Such averaging reduces the variance of the noise estimates to nearly zero. The signal power measurement can be written (approximating the noise as a constant) as $P = \text{NRCS}\left(P_0 - \frac{1}{\text{SNR}}\right)$.

The basic pdf for P_0 requires the eigenvalues of $\mathbf{R\Upsilon}$. However, because the number of overlapping segments and frequency bins included change from cell to cell, these eigenvalues also change. As an example, the Table 1 lists the eigenvalues, along with their multiplicities for 2 overlapping segments and 5 frequency bins.

Eigenvalue	Multiplicity
0.5932	2
0.5860	2
0.4252	2
0.4052	2
0.2578	2
0.1495	2
0.0613	2
0.0185	2
0.0030	2
0.0002	2

Table 1: *Example of eigenvalues and their multiplicities for 2 overlapping data segments and 5 frequency bins.*

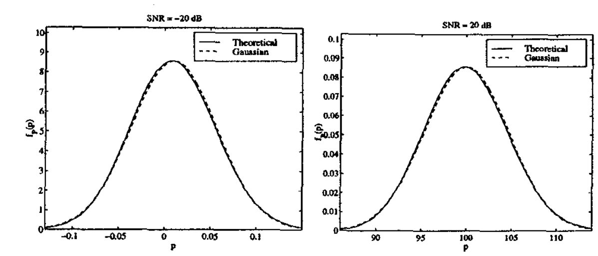

Figure 2: *Sample pdfs for the signal-only power estimate based on NSCAT processing of a near-swath cell for a low SNR (low wind speed) and a high SNR (high wind speed).*

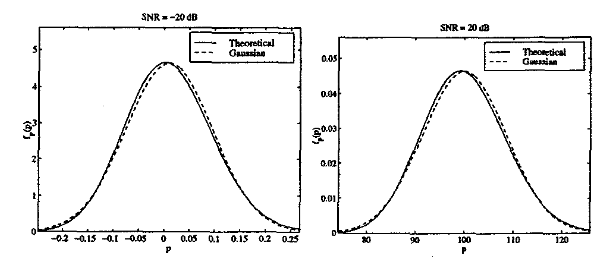

Figure 3: *Sample pdfs for the signal-only power estimate based on NSCAT processing of a far swath cell for a low SNR (low wind speed) and a high SNR (high wind speed).*

Figures 2 and 3 display several examples of the density functions of signal power estimates. The near and far swath cells use 7 and 2 overlapping data segments, respectively, and 5 frequency bins. Note that the SNR simply scales the power estimate for each case. The dashed line indicates a Gaussian density with the same mean and variance. It is clear that the actual distribution is skewed toward low σ^o values. For near swath cells, the true distribution is closer to Gaussian since more overlapping segments are used.

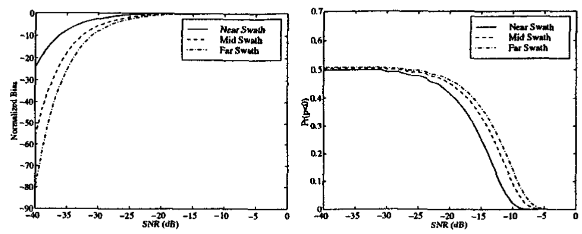

Figure 4: *Two useful measures of the probability density function of signal power estimates: the normalized bias (normalized difference between the mode and mean) and the probability of negative estimates.*

In order to quantify the difference between the computed probability function and the Gaussian assumed in current processing, we introduce two useful measures: the normalized bias and the probability of negative measurements. We define the normalized bias as the difference between the mode and the mean of the computed pdf, normalized by the mean of the distribution. Also of interest is the probability of a negative power estimate, computed as the integral of the pdf from $-\infty$ to 0. These measures are displayed in Fig. 4 as functions of the SNR for three representative cell cases.

The concern is that because the actual density is skewed, the wind estimation may be biased. A compass simulation shows that this bias is small, though non-zero. Using a particular measurement geometry and noise parameters (taken from an NSCAT L1.7 file) simulated backscatter measurements for a given wind vector cell are used to estimate the wind in a traditional compass simulation. The simulated measurements were then biased, according to Fig. 4, and the wind is estimated with this set of biased measurements. Note that a more precise method would be to use the correct pdf, rather than simply a shifted Gaussian distribution; but to obtain a reasonably accurate result with minimal code modification for this preliminary study, this approach was adopted. Figures 5 and 6 display the results of compass simulations for cells in the near and far swaths, respectively. The plots show the difference between the wind estimates made without the correcting bias and the wind estimates with the correcting bias for both speed and direction. As the simulated wind speed increases, the difference between the two estimates increases, while the direction difference is minimal.

CONCLUSION

The probability structure of NSCAT signal estimates shows considerable skew from the frequently assumed Gaussian. Initial observations suggest that this has negligible impact on the estimated wind direction, though the resulting wind speed bias should be investigated further. The

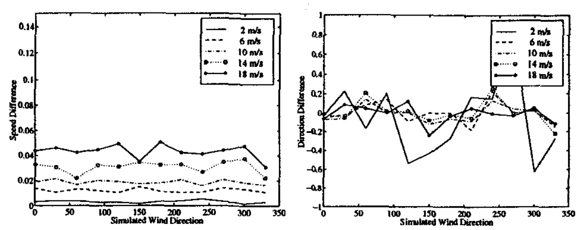

Figure 5: *Impact on wind retrieval with bias error in the pdf; near swath.*

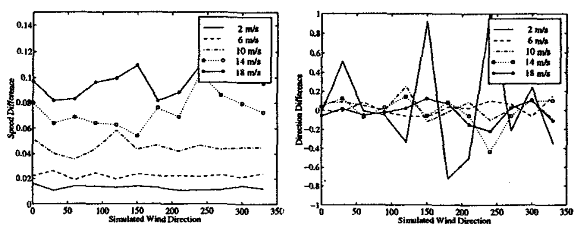

Figure 6: *Impact on wind retrieval with bias error in the pdf; far swath.*

skew of the probability density function of the power measurements causes current wind retrieval techniques (which assume measurements of Gaussian random variables) to be biased high. This effect is ameliorated to some extent by models like NSCAT1, which are tuned to the retrieved wind.

REFERENCES

[1] F. M. Naderi, M. H. Freilich, and D. G. Long, "Spaceborne radar measurement of wind velocity over the ocean—an overview of the NSCAT scatterometer system," *Proceedings of the IEEE*, vol. 79, pp. 850–866, June 1991.

[2] P. D. Welch, "The use of the fast Fourier transform for the estimation of power spectra," *IEEE Trans. Audio Electroacoustics*, vol. AU-15, pp. 70–73, June 1967.

[3] D. G. Long, C.-Y. Chi, and F. K. Li, "The design of an onboard digital doppler processor for a spaceborne scatterometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 869–878, November 1988.

[4] P. E. Johnson and D. G. Long, "The probability density of Welch's modified periodogram estimates," *Submitted to IEEE Transactions on Signal Processing*, 1997.

[5] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis.* Addison-Wesley Publishing Company, 1991.