

TESTING FOR NONGAUSSIAN FLUCTUATIONS IN GRAIN NOISE

Steven P. Neal
Department of Mechanical and Aerospace Engineering
University of Missouri-Columbia
Columbia, MO 65211

Kevin D. Donohue
Department of Electrical Engineering
University of Kentucky
Lexington, KY 40506

INTRODUCTION

In ultrasonic nondestructive evaluation (NDE), grain noise corrupts the scattered wave field from a flaw in a polycrystalline material. Many probabilistic approaches associated with flaw detection and characterization utilize stochastic models in which grain noise is assumed uncorrelated and zero-mean Gaussian distributed. Typically, the Gaussian assumptions is justified via heuristic arguments based on the central limit theorem. This paper presents the kurtosis test and the Shapiro-Wilk W test as methods to quantitatively test time domain noise ensembles for deviations from Gaussian statistics. We will establish, through the application of these hypothesis tests to grain noise, a quantitative tool which can be used to consider "how Gaussian" grain noise signals must be for Gaussian noise based signal processing procedures to out perform alternative approaches.

A review of the literature reveals that Neal and Thompson [1] addressed the measurement of grain scattering signals and evaluation of grain noise as a random variable. They documented measurement procedures and, for a limited number of cases, showed qualitatively in the time and frequency domain that grain noise is uncorrelated and zero mean Gaussian distributed. In addition to their work, the NDE literature is replete with publications addressing materials characterization or grain noise suppression where grain scattering signals have been measured. However, other than the research briefly reviewed below, the focus of these projects has not been the measurement and analysis of grain noise, and there have been no qualitative or quantitative assessments of the grain noise included within these publications.

Researchers at the Center for Nondestructive Evaluation (CNDE) at Iowa State University have been studying backscattered grain noise as part of a program aimed at the detection of defects in titanium engine component [2-5]. Comparisons of grain noise distributions with the Gaussian distribution were done qualitatively by comparing the probability density functions for the observed grain noise with Gaussian probability density functions. Comparisons were also done quantitatively, but without using hypothesis testing, by comparing the ratios of the peak noise value to the root-mean squared noise level for the observed distribution and a Gaussian distribution. They used experimental measurements of grain noise [3], Monte Carlo calculations of synthetic noise signals [2], and a combination of the Monte Carlo approach with random walk theory [4] to study statistical models for grain

noise. Whenever possible, the grain noise distribution was studied as a function of the number of contributing scatterers. Using each of the three approaches, they showed that grain noise tends to be Gaussian for large numbers of scatterers. As the number of scatterers becomes small, significant deviations from Gaussian behavior were seen. Experimentally, nongaussian behavior was observed near the focal region of a focused transducer where the number of sonified scatterers is relatively small. More recently they have found that in some cases, using the K-distribution (see below) yields improved results over a Gaussian based model when modeling the largest extreme value distribution associated with grain noise in a gated peak-detection environment [5].

Jakeman and Pusey [6] explained non-Rayleigh fluctuations for microwave sea echoes as resulting from a small number of effective scatterers being present in the resolution cells, thereby making the application of the central limit theorem invalid. A class of probability density functions (K-distributions) were derived based on a finite average number of scatterers per resolution cell. The K-distributions exhibit tails which are fatter than the tails for the Gaussian distribution. Fat tails are consistent with the observed distribution associated with echoes from the resolution cells with limited numbers of effective scatterers.

Sleeve and Lele [7] derived a scatterer number density estimator which is based on the ratio of the second and fourth moments of the ultrasonic backscattered signal and includes compensation for system effects. The estimator operates on the basic principle that moment ratios consistent with Gaussian statistics imply an infinite number of scatterers in the volume. As the number of scatterers becomes smaller (finite), the moment ratio deviates from the expected Gaussian statistic. The estimator maps this deviation into a scatterer number density value. Estimator performance was shown to be very good when tested on tissue phantoms.

Whether using a central limit theorem or a maximum entropy principle perspective, a heuristic argument can be constructed which justifies the Gaussian assumption for grain noise. The central limit theorem indicates that under very general conditions, the distribution associated with the sums of independent random variables tends to a Gaussian distribution as the number of summed independent random variables goes to infinity. Therefore, grain noise will approach Gaussian behavior provided that 1) a large number of *significant* scattering sites are involved, and 2) no regular structure exists between the significant scatterers.

In this paper we extend the research of Neal and Thompson [1] from a qualitative analysis of grain noise which considered only reasonably Gaussian grain noise cases to a quantitative analysis which shows both Gaussian and nongaussian grain noise behavior using simulated and measured signals. We introduce the kurtosis test and the Shapiro-Wilk W test as the quantitative tools for assessing the Gaussian assumption for grain noise.

HYPOTHESIS TESTING

The application of a hypothesis test begins with the statement of a null hypothesis. In our case, the null hypothesis, H_0 , is that the noise is Gaussian. The purpose of a hypothesis test is then to provide a basis for either accepting or rejecting the null hypothesis. The basic procedure for using hypothesis testing to test for Gaussian statistics is as follows:

1. Measure or create (using a simulation) an ensemble of N independent sampled A-scans from a region with statistically stationary scattering properties.
2. Extract N samples over the ensemble corresponding to the same time in each A-scan.
3. From the N samples, compute a statistic whose distribution is known for the case in which the samples are Gaussian distributed. Repeat this procedure for each time sample along the A-scan. As discussed below, we will compute the kurtosis and the W -statistic for each time sample.

4. From the statistic's distribution, determine the threshold for a chosen significance level, α , and determine how many computed statistics cross the threshold. A noise distribution is said to deviate from Gaussian statistics when the fraction of statistics, denoted $\bar{\alpha}$, which cross the threshold is significantly greater than α . By significant, we mean that α is outside of the 95% confidence limits on $\bar{\alpha}$.

Kurtosis

The first of the two tests that we introduce to the analysis of grain noise is based on the kurtosis [8,9]. The kurtosis is defined as the ratio of the fourth central moment to the square of the second central moment as given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad (1)$$

The kurtosis has a value of 3 for the Gaussian distribution. The hypothesis test based on the kurtosis is a directional test which is particularly sensitive to deviations in the tails of the distribution. This makes the test well suited to the detection of fat tails in the distribution which are indicative of nongaussian grain noise fluctuations due to a small numbers of scatterers. Also note that the computations for sample kurtosis include the sample mean (for the central moment computations) and the sample variance for normalization. This characteristic is critical for the analysis of measured grain noise which can show non-zero mean behavior due to front surface ringing and which may be nonstationary in both its mean and variance.

Shapiro-Wilk W Statistic

The Shapiro-Wilk test [8,10] is an omnibus test. That is, it is sensitive to deviations from Gaussian statistics which occur anywhere in the distribution, thus providing a good compliment to the kurtosis test. The Shapiro-Wilk test is based on order statistics. Following the notation of Mardia [8], a vector, \mathbf{x} , containing n independent samples of some random process can be represented as

$$\mathbf{x} = [x_1, x_2, x_3 \dots x_n]^T \quad (2)$$

The order statistic vector, \mathbf{x}_{os} , is defined as

$$\mathbf{x}_{os} = [x_{(1)}, x_{(2)}, x_{(3)} \dots x_{(n)}]^T \quad (3)$$

where $x_{(1)} \leq x_{(2)} \leq x_{(3)} \dots \leq x_{(n)}$. The expected value of the order statistic vector for a standard normal distribution with sample size n is given by \mathbf{c} , with covariance matrix \mathbf{V} . The expected value of the order statistic vector for any Gaussian distribution with mean μ and variance σ^2 is given by

$$E[\mathbf{x}_{os}] = \mu \mathbf{c}_0 + \sigma \mathbf{c} \quad (4)$$

where \mathbf{c}_0 is an n by 1 vector of ones. The best linear unbiased estimate of σ in Equation (4) comes from [8,10]

$$\tilde{\sigma}^2 = \frac{\mathbf{c}' \mathbf{V}^{-1} \mathbf{x}_{os}}{\mathbf{c}' \mathbf{V}^{-1} \mathbf{c}} \quad (5)$$

Note that the values of c and V used in this estimator imply Gaussian distributed data. As the data in x_{OS} deviates from the Gaussian distribution, the accuracy of the estimate degrades. The sample variance, on the other hand, makes no assumption on the distribution of the data, and is given by

$$\bar{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = s^2 \tag{6}$$

where \bar{x} is the sample mean. For Gaussian data, $\bar{\sigma}^2$ approximately equals $\tilde{\sigma}^2$, and they are statistically independent of one another. From Shapiro and Wilk [10], the W -statistic is given by

$$W = \frac{\tilde{\sigma}^2}{\bar{\sigma}^2} \leq 1 \tag{7}$$

Significance tables for critical values of W and values of the coefficients for the estimator in Equation (5) are given by Shapiro and Wilk [10].

RESULTS

The threshold for each statistic is a function of the chosen significance and the sample size. For a significance, $\alpha = 0.05$, and, for discussion purposes, a sample size, $N = 30$, the threshold for the kurtosis is 4.12 [9], and the threshold for the W -statistic is 0.983 [10]. Further, for discussion purposes, assume that each A-scan contains 1000 time samples. For the kurtosis the procedure would be to calculate the kurtosis at each of the 1000 time samples, and determine the fraction, $\bar{\alpha}$, of kurtosis values which are higher than the 4.12 threshold. The null hypothesis is then rejected if $\bar{\alpha}$ is significantly greater than 0.05 (i.e., if significantly more than 50 out of 1000 kurtosis values are greater than 4.12). For the W -statistic the procedure would be to calculate the W -statistic for each of the 1000 time samples, and determine the fraction, $\bar{\alpha}$, of W -statistic values which are lower than the 0.983 threshold. For the Shapiro-Wilk test, the null hypothesis is rejected if $\bar{\alpha}$ is significantly greater than 0.05. Again, we take significant to mean that $\bar{\alpha}$ minus the 95% confidence limit is still greater than 0.05. For each test, the 95% confidence limits on $\bar{\alpha}$ are calculated as 1.95 times the maximum likelihood estimate of the standard error for the statistic which is given by the square root of $\alpha(1-\alpha)/N$ where N is the number of samples (equivalently, the number of A-scans).

Results are presented below as plots of $\bar{\alpha}$ with 95% confidence limit error bars versus either the number of scatterers per resolution cell or versus the material type. In all cases, the value of $\alpha = 0.05$ is indicated with a dashed line.

Simulation Results

The importance of the simulation is that the average number of scatterers per resolution cell can be controlled. Consider in more detail the problem of estimating the number of *significant* scatters per resolution cell. For any chosen time within a grain noise A-scan and assuming single scattering, the incident pulse length and time-of-flight considerations dictate the material volume, called the resolution cell, in which all scattering must occur in order to contribute to the A-scan at the chosen time [11]. If we think of an idealized situation in which all of the randomly placed scatterers within the resolution cell are identical and the scatterers are bathed with a uniform sonifying field, then if the number of scatterers per unit volume and the size of the resolution cell are known, the number of contributing scatterers can be quantified in a meaningful way. Unfortunately, the number of *significant* contributing scatterers can only be clearly determined for grain noise signals

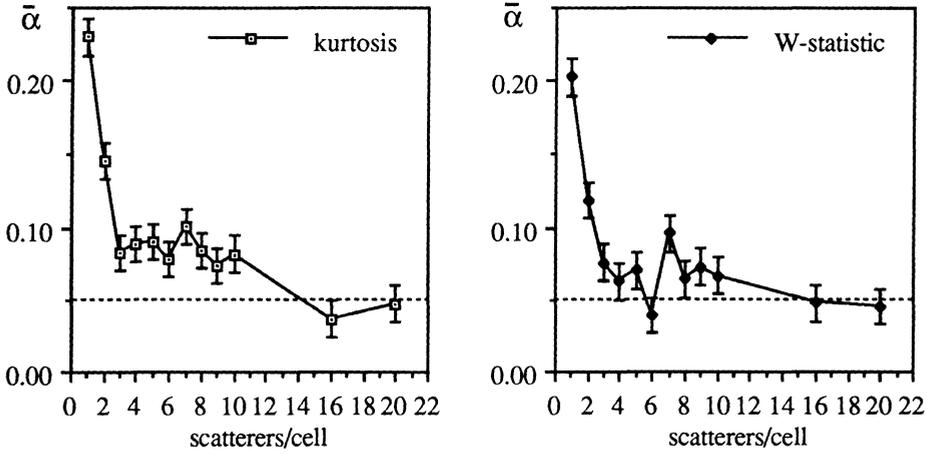


Figure 1. Simulation results.

created via simulation or for backscattered signals measured from a well characterized phantom. For the actual situation of the interrogation of a polycrystalline material with sound generated by a piston source transducer, the grains have varying sizes and scattering strengths and the sonifying field is highly non-uniform within the resolution cell. For this situation, it is not clear that there is a meaningful way to quantify the number of *significant* contributing scatterers. In fact, for the cases in which we observe nongaussian grain noise, we are actually looking at the sum of a dominate nongaussian grain noise component, a significantly smaller Gaussian grain noise component, and a Gaussian electronic noise component. To clarify this comment, note that the intensity of the sound field generally decreases with increasing distance from the transducer axis; therefore, the number of scatterers sonified by a given pressure level increases as the pressure level decreases. If we invoke the central limit theorem and claim that grain noise is always Gaussian since there are always large numbers of grains in any resolution cell, we may draw an erroneous conclusion since many of the scatterers are significantly off of the transducer axis and are *insignificant* contributors to the measured grain noise. This situation is most acute when considering a resolution cell centered about the focal region of a focused transducer.

Simulated A-scans were modeled as the superposition of distorted, scaled, and time-shifted versions of the illuminating pulse as follows

$$y(t) = \sum_{i=1}^p h_i(r_i, \theta_i, \phi_i, t) * a_i(t - r_i / v_i) \quad (8)$$

where p denotes the number of scatterers in a resolution cell, r_i, θ_i, ϕ_i represent the radius and the orientation of the i^{th} scatterer relative to the beam axis, $a_i(t - r_i / v_i)$ denotes the scattering function for the i^{th} scatterer, v_i denotes the average velocity of sound between the i^{th} scatterer and the transducer, and $h_i(r_i, \theta_i, \phi_i, t)$ represents the measurement system response function for the illuminating pulse including propagation path attenuation but ignoring beam diffraction.

For the simulation, thirty A-scans with 1150 time samples each were generated for twelve different scatterer number densities specified in terms of scatterers per resolution cell. Figure 1 shows $\bar{\alpha}$ based on the kurtosis, β_2 , and on the W -statistic, respectively, versus the number of scatterers per resolution cell. Both tests show clear nongaussian behavior for 1

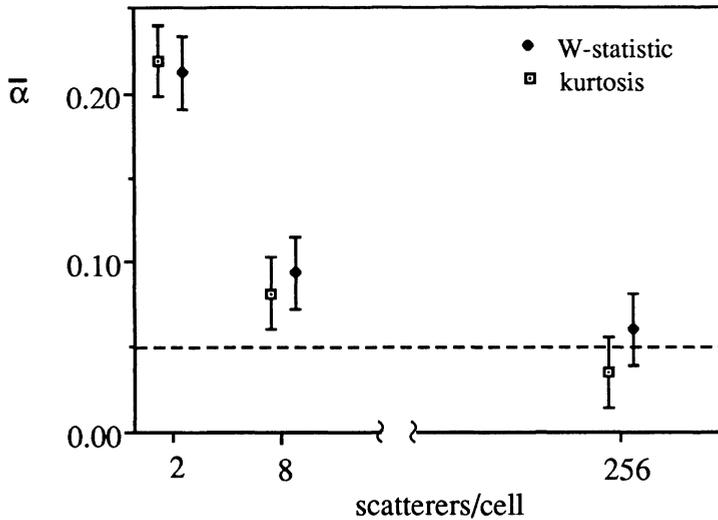


Figure 2. Phantom results.

scatterer per cell and for 2 scatterers per cell with $\bar{\alpha}$ in all cases having a value much greater than 0.05. In the 3 to 10 scatterers per cell range, the kurtosis test clearly identifies nongaussian behavior. The Shapiro-Wilk test demonstrates less *power* as a hypothesis test [8] than the kurtosis test for the detection of this type of nongaussian behavior (fat tails) but still generally indicates nongaussian behavior in the 3 to 10 scatterers per cell range. Both tests dictate that the null hypothesis of Gaussian behavior should be accepted for scatterer number densities above approximately 15 scatterers per cell.

Experimental Results: Phantom Data

Three ATS phantoms were scanned with an ATL Ultramark 9 system. Details of the experiments and nature of the phantom are given by Weng et al. [12]. Each phantom consisted of suspended glass beads with average scatterer number densities of 2, 8, and 256 scatterers per resolution cell, respectively. For each phantom, 40 A-scans were measured with 220 time samples each. Since the measurement procedure yielded adjacent A-scans which show an unacceptable level of correlation, the 40 signals were treated as 2 sets of 20 uncorrelated A-scans by analyzing sets which contained every other A-scan. Figure 2 shows $\bar{\alpha}$ based on β_2 and W , respectively, versus the number of scatterers per resolution cell. For 2 and 8 scatterers per cell, both test clearly indicate nongaussian behavior, consistent with the result for the simulation (see Fig. 1). As expected, at 256 scatterers per cell, both tests indicate acceptance of the null hypothesis that the backscattered signals are Gaussian.

Experimental Results: Grain Noise Data

Backscattered grain noise signals were analyzed which were measured from a number of materials and by a number of investigators over the past ten years. Materials / probe combinations which clearly showed Gaussian grain noise included ASTM8 stainless steel interrogated by a 15 MHz planar probe, ASTM5 stainless steel interrogated by a 15 MHz focused probe, aluminum with 2% porosity (328 μm average pore diameter) interrogated with a 15 MHz planar probe, and beryllium-copper bar interrogated with a 15 MHz planar probe. In Fig. 3 we present hypothesis test results for three material / probe combinations which showed nongaussian behavior. In each case, the material was interrogated by a 15 MHz focused probe. Again, the kurtosis test is the more powerful of the two tests, indicating nongaussian behavior in all three cases. The Shapiro-Wilk test shows Gaussian

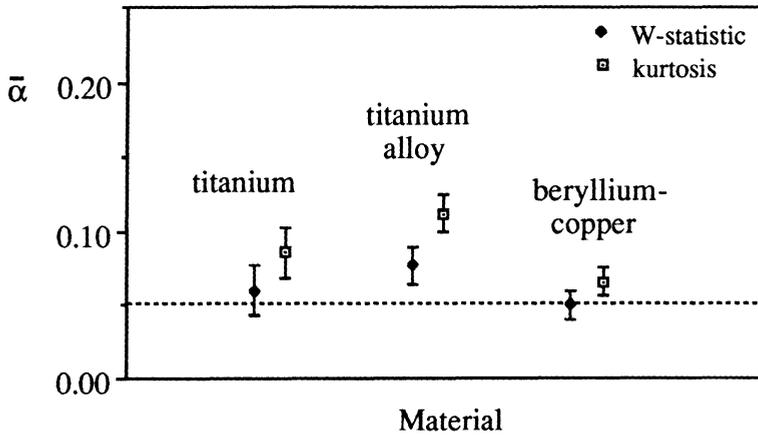


Figure 3. Grain noise results.

behavior for the beryllium-copper, border-line behavior for the titanium, and clear nongaussian behavior for the titanium alloy. The grain noise signals from the titanium alloy were shown to be nongaussian by Margetan et al. [3] using the ratios of the peak noise value to the root-mean squared noise level and more recently using the kurtosis [5]. They showed that deviations from Gaussian statistics were particularly prevalent near the focal zone where the resolution cell is relatively small resulting in a smaller number of dominant scatterers than in the regions away from the focal zone. We have also seen this behavior by plotting β_2 versus time and noting that there is a high density of β_2 values in the portion of the signal which corresponds to scattering from the focal region which break the appropriate threshold for the 0.05 significance level.

SUMMARY AND DISCUSSION

In summary, both the kurtosis test and the Shapiro-Wilk test have identified nongaussian behavior in the simulated noise signals, the phantom signals, and in the measured grain noise. Based on the simulated signals, the tests showed a clear distinction between one scatterer and two scatterers per resolution cell. The kurtosis test was also able to clearly detect nongaussian behavior for scatterer density in the 3 - 10 scatterers per resolution cell range. The kurtosis test is the more powerful than the Shapiro-Wilk test when the deviations from Gaussian statistics are manifest as fat tails in the distribution caused by a small number of significant scatterers per resolution cell. However, we feel that it is prudent to run both tests. The kurtosis test, which is in general a directional test, was applied as a one-sided test in order to test for fat tails (i.e., tested for kurtosis values greater than 4.12). Therefore, it is essentially insensitive to deviations in the central region of the distribution, and it is also insensitive to skewness. The Shapiro-Wilk test is an omnibus test which is sensitive to deviations from Gaussian behavior anywhere in the distribution. While the kurtosis test is more powerful, the results from the two tests should be in the same range. In particular, if the Shapiro-Wilk test shows nongaussian behavior when the kurtosis test does not, this is indicative that there may be a problem with the data, possibly improper experimental or signal processing procedures.

We have introduced the application of well established hypothesis tests to the problem of testing for nongaussian fluctuations in grain noise. We have presented results which show three cases in which measured grain noise shows significant deviation from Gaussian behavior. We have established hypothesis testing as a vehicle for establishing when a

nongaussian modeling or signal processing procedure should be used in place of a procedure which assumes Gaussian noise statistics.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Frank J. Margetan from the CNDE at Iowa State University and Dr. Mark D. Russell from the University of Missouri-Columbia (MU) for providing grain noise data and for technical discussions related to this work. The authors would also like to thank Dr. Paul L. Speckman from the Statistics Department at MU for his advice on hypothesis testing. The phantom data resulted from measurements made in conjunction with a project supported in part by N.I.H. Research Grant No. HL30045 from the National Heart, Lung, and Blood Institute.

REFERENCES

1. S. P. Neal and D. O. Thompson, "The measurement and analysis of grain noise as a random variable," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1990), Vol. 9A, pp. 625-632.
2. I. Yalda-Mooshabad, F. J. Margetan, and R. B. Thompson. "Monte-Carlo simulation of ultrasonic grain noise," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1993), Vol. 12B, pp. 1727-1734.
3. F. J. Margetan, R. B. Thompson, I. Yalda-Mooshabad, and H. Kim Han. "Detectability of small flaws in advanced engine alloys," Center for Nondestructive Evaluation, Iowa State University, Ames, IA. July 15, 1993.
4. I. Yalda-Mooshabad, F. J. Margetan, and R. B. Thompson. "Applying the 2D random walk formalism to predict ultrasonic grain noise distributions," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1994), Vol. 13B, pp. 1713-1720.
5. I. Yalda, F. J. Margetan, Y. K. Han, and R. B. Thompson. "Survey of ultrasonic grain noise characteristics in jet-engine titanium," these proceedings.
6. E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Transactions on Antennas and Propagation* **AP-24**(6), 806-814 (1976).
7. G. L. Sleaf and P. P. Lele, "Tissue characterization based on scatterer number density estimation," *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control* **UFFC-35**(6), 749-757 (1988).
8. K. V. Mardia, "Tests of univariate and multivariate normality," in *Handbook of Statistics*, edited by P. R. Krishnaiah (North-Holland, Amsterdam, 1980), Vol. 1, pp. 279-320.
9. E. S. Pearson and H. O. Hartley, "Biometrika tables for statisticians," *Biometrika* **1**, Table 34, 208 (1972).
10. S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika* **52**(3,4), 591-611 (1965).
11. M. D. Russell, S. P. Neal, and E. J. Boote, "Experimental estimation of the longitudinal-wave backscatter coefficients for ultrasonic interrogation of weak scattering materials," *J. Acoust. Soc. Am.* **93**(3), 1267-1276 (1993).
12. L. Weng, J. M. Reid, P. M. Shankar, and K. Soetanto, "Ultrasound speckle analysis based on the K distribution," *J. Acoust. Soc. Am.* **89**(6), 2992-2995 (1991).