Proceedings of the 1993 Interface Symposium

Monte Carlo Study of a Non Gaussian Radar Clutter Generator

Jorge Luis Romeu
Department of Mathematics
SUNY-Cortland
Cortland, NY 13045 *

April 3, 1993

Abstract

Generation of multivariate Non Gaussian random variates is of importance in radar clutter studies. For, when the analytical evaluation of a radar clutter distribution is difficult or impossible, it is through computer simulation that such evaluation is attacked and solved. A new statistical method, based on SIRP's (Spherically Invariant Random Processes) allows both fit testing of a multivariate Non Gaussian process and the computer generation of these processes. This theoretical method decomposes the Non Gaussian process into the product of two subprocesses. One of these processes is univariate and drives the distribution of the Non Gaussian process. The other subprocess is multivariate Gaussian. In theory, the new method is correct. However, in practice the method's results may not always reflect its theoretical properties with the required accuracy due to implementation problems. This report describes a Monte Carlo study designed to assess the computer implementation of the theoretical method. Using different sample sizes and number of variates, we generate, via Monte Carlo, two specific SIRP Processes: one univariate and one multivariate. Goodness-of-fit tests are performed on several variables obtained from the Processes as well as on the Processes itself. The case where the covariance matrix of the SIRP Process is known and the case where it is estimated from the data are considered. Results and statistical tests are discussed and compared.

1 Introduction and Motivation

This research is motivated by the need to assess the (computer) implementation of a new theoretical model for generating (simulating) and evaluating Non Gaussian radar clutter. As with many new theoretical methods, the question remains as to whether in practice (i.e. in a digital computer, using constrained amounts of data and estimated parameters) the method will still provide useful results. It is to this empirical assessment of the theoretical method that we refer to as validation.

Such is the situation that we have investigated, via Monte Carlo, for the theory of SIRP, *Spherically Invariant Random Processes*. This new model was developed and presented by Rangaswamy, Weiner and Ozturk (1991 and 1992). It has also been extensively and carefully discussed in the Rome Lab document referred in this paper as the Kaman Report (1992). Our work is based on this latter document. The present report is a revised and enhanced version of our Summer Research (Romeu, 1992d) and CASE Center (Romeu 1992e) reports.

Succinctly, a multivariate (N-variate) SIRP X is defined, via the product ($X = S * Z$) of two independent random processes S, Z. The first one, S, is a univariate process and *drives* the SIRP process X: i.e. S completely determines X. The second process is the Multivariate (N-variate) Gaussian process with mean zero and covariance matrix M, denoted by $Z \sim MVN_N(0, M)$. The process Z is independent of S and remains the same no matter what is the SIRP X.

Such representation of a multivariate process, via the product of a univariate process and a well defined multivariate one (referred to in Rangaswami et al. (1992) as the Representation Theorem) has also been studied in theoretical statistics. See, for example, the work on *Elliptically Contoured Distributions* by Cambanis, Huang and Simons (1981), Johnson, Chiang and Ramberg (1984) and Johnson (1987) among others.

It is convenient to *standardize* the univariate process S so $E(S^2) = 1$. This way, the covariance matrix of the SIRP X is now $\Sigma = M$. Letting $p = X'\Sigma^{-1}X$, the *quadratic form* of the process X, we obtain the conditional density function (pdf) of the SIRP X, given the variable S, denoted $X|S$ as:

$$f_{X|S}(x|s) = (2\pi)^{-N/2}|M|^{-1/2}s^{-N}exp(\frac{-p}{2s^2})$$

From here, the unconditional pdf of the SIRP X becomes

$$f_X(x) = (2\pi)^{-N/2}|M|^{-1/2}h_N(p)$$

$$where \quad h_N(p) = \int_0^\infty s^{-N}exp(\frac{-p}{2s^2})f_S(s)ds$$

The *quadratic form* p of the process X plays a decisive role in the SIRP Theory: $h_N(p)$ provides the pdf of the new random variable p via

$$f_P(p) = \frac{1}{2^{N/2}\Gamma(N/2)}p^{N/2-1}\,h_N(p)$$

From the above relations it follows that any SIRP X is completely specified once we obtain the two key elements: (i) the function $h_N(p)$ of the quadratic form p (through the pdf $f_S(.)$ of the univariate process S) and (ii) the covariance matrix Σ of the SIRP X.

However, such theoretical representation must be empirically assessed (validated) before proceeding further. For, theoretical methods often have two serious types of implementation problems that must be carefully investigated.

One of the problems is that some theoretical methods are asymptotic in nature (i.e. the derived distribution holds for large samples only). Therefore, when the sample sizes are not large the small-sample statistic distributions are only approximations of the (asymptotic) theoretical ones. In such cases, a minimal sample size n^* for the asymptotic values to hold, may be found. And empirical results for samples $n < n^*$, may be obtained. An example of this type of problem investigation and its corresponding adaptive solution is presented in Romeu and Ozturk (1993). Here, it is shown how the two multivariate normality Goodness-of-Fit (GOF) tests of Mardia, both of them asymptotic, require minimal samples of size 200. And how, for smaller sample sizes, empirical critical values are required (and tables provided).

A second serious (computer) implementation problem of theoretical methods occurs with the use of numerical approximations and convergence algorithms. In such cases, results may be heavily dependent on the hardware used. In theoretical derivations, results are often given in closed forms, using certain integrals and derivatives that can only be approximated numerically on a digital computer. An example of this other situation is presented in Romeu (1990). There, the *angles* multivariate normality GOF test of Koziol, which requires the numerical inversion of the sample covariance matrix, is studied. It turns out that Koziol's test yields widely different results when implemented on a sequential machine (versus a parallel computer). This occurs when the multivariate distributions under study have more than four highly correlated components. In both of the mentioned cases, it was only through a Monte Carlo study of the theoretical method that the mentioned problems were uncovered and adaptive solutions were provided. Thus its practical importance.

In the present case, the closed forms of the SIRP distributions (obtained by transformations and products of random variables) are not asymptotic but mathematically convoluted. They include the calculation of Modified Bessel and Gamma functions, with very small (shape) parameters. Even more, some of these functions are in the denominator. Therefore the use of numerical methods and specific hardware may have important effects in the method's implementation. And it is necessary to investigate, through a Monte Carlo experiment, this situation.

2 Objectives of the Study

There are several assertions involved in the theoretical representation of an SIRP X, when it is defined as the product of the two subprocesses S and Z.

First, we want to verify that (i) the desired process X is accurately obtained by the product of S and Z; that (ii) the resulting distribution of the quadratic form p actually depends on the distribution of the univariate process S; that the (iii) marginal distributions F_{X_i}, $i = 1 \ldots N$, of the multivariate process X follow the same (univariate) family of SIRP distributions as the multivariate X and that (iv) the resulting covariance matrix Σ of the multivariate process X is accurately obtained through the univariate process S and the covariance matrix M of the process Z.

To empirically assess that the above assertions are met in the computer implementation of the theoretical method, and to investigate its limitations, we conduct a Monte Carlo experiment. We first investigate for what sample sizes and number of covariates the SIRP process X allows us to accurately identify (i) the univariate process S, (ii) the marginal distributions of X, (iii) the quadratic form p and (iv) the covariance matrix Σ obtained from S and M.

In addition, the theoretical SIRP model assumes that the covariance matrix Σ, of X **is always known**. This seldom occurs in practice, except when simulating radar clutter in the computer. However, such simulations constitute an important application of the SIRP model. For, through them some types of radar clutter with difficult or impossible analytical solutions are studied. Such model application leads to an additional objective of the present research: the verification of several simulation routines written following the SIRP theory, which will be used to study Non Gaussian radar clutter.

Yet another research objective is to perform a limited *degradation study* of the model. We verify whether the resulting SIRP random variate X, the univariate process S and the quadratic form p, can be accurately identified when generated under different experimental settings. These settings include decreasing sample sizes, increasing multivariate intercorrelations ρ and increasing number of variates N in the multivariate SIRP process X under study.

A final objective of the present validation effort is to conduct performance studies of the estimation of several parameters of interest. The SIRP theory requires knowledge of several key elements, seldom known in practice. This is the case with the covariance matrix Σ of the SIRP process $X = S * Z$. In practice, Σ is estimated from the data and its estimate Σ^* is used. It is necessary to study the effects, if any, of such a substitution. And it is necessary to study the sample size requirements and the number of variates for which the estimation of the covariance matrix becomes so imprecise as to be useless.

3 The Distribution of Interest

For the above reasons, a Monte Carlo Validation Study that generates the SIRP model $X = S * Z$ as indicated in the previous section is required. However, there are serious problems when undertaking such a study. First, validating this SIRP model requires testing **both** the SIRP process X **and** the quadratic form p.

There are two reasons for checking **both** X and

p. We cannot **assume** that the multivariate X, generated by the product of variates S and Z is correct, especially when we are checking the simulators as part of our task. This is due to the algorithmic problems discussed above, which may have an effect in the computer implementation of the theoretical model.

Second, the SIRP model will be used for two purposes: signal identification and generation of the clutter data. Therefore we need (i) to assess whether we can identify the (correct) distribution of the quadratic form p synthesized from the multivariate SIRP X (signal identification). And (ii) to assess and verify that the multivariate SIRP process X is the one prespecified, given that we are simulating the SIRP model via the product of S and Z (clutter data generation).

Here is where the *Multivariate K-Distribution* comes into play. This distribution has been theoretically identified as the main SIRP distribution of interest in our radar study. The K-Distribution is commonly used for modelling radar clutter pdf's that have extended tails. It is defined, following the Kaman report, through the pdf:

$$f_X(x) = \frac{2b}{\Gamma(\alpha)}\left(\frac{bx}{2}\right)^{\alpha} K_{\alpha-1}(bx)\, u(x)$$

where α and b are the shape and scale parameters, respectively, of the multivariate K-Distribution, K_N is the N^{th} order Modified Bessel Function of the second kind and $u(x)$ is the unit step function.

The Multivariate K-Distribution arises when the product of a Rayleigh and a Chi Square random variables are considered. Still according to the SIRP theory (Kaman, 1992), each K-distributed SIRP process $X = S*Z$ is associated with a *characteristic* pdf of the corresponding univariate process S, which is defined by:

$$f_S(s) = \frac{2}{\Gamma(\alpha)2^{\alpha}}(bs)^{2\alpha-1} exp\{\frac{-b^2 s^2}{2}\} u(s)$$

and also with a function $h_N(p)$ for the quadratic form p which is:

$$h_N(p) = \frac{b^N}{\Gamma(\alpha)} \frac{(b\sqrt{p})^{\alpha-\frac{N}{2}}}{2^{\alpha-1}} K_{\frac{N}{2}-\alpha}(b\sqrt{p})$$

Notice how, in the above functions, the shape parameter α is present in several expressions in the denominator as well as in the (numerically obtained) Modified Bessel Function $K_{\frac{N}{2}-\alpha}$. For long tailed K-Distributed variables, this shape parameter α is very small. And this may also be a potential source of computer implementation problems for the SIRP model.

Another problem is that there are no multivariate GOF tests for the Multivariate K, our distribution of interest. Therefore, we cannot test directly the simulated SIRP $X = S*Z$. If there were such test, then a key reason for developing the SIRP theory (the need for the indirect testing of the multivariate process X via its univariate quadratic function p) would no longer exist. However, we still **need** to assess the simulated SIRP X.

To circumvent the problem of lack of a multivariate test for a K-distributed SIRP X, we approach the validation process through a *two-phase scheme.* In the first phase (taking advantage of some SIRP properties) we implement a special case of an SIRP: the multivariate Gaussian Process $X = 1*Z$. This special SIRP is obtained when S is a constant unit. There are several well investigated, multivariate normality GOF tests, that can be applied on process X. This is a well known case where we can simultaneously test for the (Gaussian) distribution of the multivariate SIRP X and the univariate (Chi Square) distribution of the quadratic form $p = X'\Sigma^{-1}X$.

In the second phase we analyze a **univariate** SIRP. We develop cases of the *univariate K-Distribution*, analytically simple enough to obtain a closed form for its Cumulative Distribution Function (CDF). We need such closed form CDF for the simulations and for the GOF tests. We first generate univariate SIRP's following the $X = S*Z$ factorization model. Then, using the closed form CDF's obtained, we test **both** the univariate X and the univariate p and verify that the distributions agree with the SIRP theory (e.g. passing a GOF tests) for the sample sizes given.

4 Phase I: Multivariate Gaussian SIRP

By letting S be a unit valued constant we obtain that $X = 1*Z$ is an SIRP. By sampling the real multivariate Gaussian $Z \sim MVN_N(0, I_N)$ with covariance identity, we obtain the quadratic function $p = \sum_{i=1}^{N} X_i^2$. When the covariance Σ is known, $p \sim \chi^2_N$, i.e. the quadratic form is distributed as a Chi Square (N) random variable. When Σ is unknown and estimated from the data, $p = X'\Sigma^{*-1}X$ follows a Beta distribution.

The multivariate normal SIRP X is a well known case in the statistical literature. However, we had four reasons to briefly redevelop it here. First, for completeness. Since (i) there are multivariate GOF test for this special case SIRP X and (ii) there is a known and tractable distribution for the quadratic function p (both of which are unavailable in the general case of the multivariate K).

Second, to verify the implementation of the generator in the simulation. Third, to demostrate two newly developed multivariate normality GOF tests, in this new setting, vis-a-vis two well established GOF tests. These new GOF tests will be heavily used in further stages of this study, for performance evaluation of estimation procedures (Romeu, 1992c). Finally, to assess the (potential) degradation of (i) the quadratic form p and (ii) the estimated covariance matrix Σ^*, as a function of the sample sizes and number of covariates.

We apply a battery of four multivariate normality GOF tests to the multivariate SIRP $X = Z$ and two Kolmogorov-Smirnov GOF tests to the two cases of the quadratic form p. One test is used for the case when Σ is known (denoted P KN/KS in the tables) and another when we estimate Σ from the input (radar) data (denoted as P ES/KS in the tables). Finally, we apply an additional Chi Square GOF test to p, for rechecking (denoted P K/CHI in the tables). Notice that the Kolmogorov-Smirnov (KS) GOF test is originally devised for fitting distributions with known parameters. And that we are implementing an adaptive KS procedure for the unknown case (P ES/KS).

Two of the four multivariate normality tests implemented (Ozturk and Romeu, 1990) were recently developed at the CASE Center of Syracuse University and have good power properties when sample sizes are small (CHOLESKI and SIGMA in the tables). The other two multivariate normality GOF tests (M-SKEW and M-KURT in the tables), are Mardia's Skewness and Kurtosis tests (Mardia (1970)). They were studied for the small samples case in Romeu (1992a). Romeu (1990) provided empirical critical values when $n < 200$, which improve their efficiency in the present situation. All four of these tests are scale-location invariant.

A series of FORTRAN routines were written and integrated into a REXX system program that drives the simulation system. Three factors, 9i) the correlated (H_1) and uncorrelated (H_0) multivariate Gaussian SIRP, (ii) the sample sizes of 50, 100, 200 and (iii) the number of variates 2, 4, 8 were simulated for 1000 replications. We denote the uncorrelated case as H_0 (null hypothesis), for we are simulating from a multivariate Gaussian with covariance matrix identity. The correlated case is denoted as H_1 (alternative hypothesis), for we are simulating from $\Sigma \neq I_N$ while *erroneously* assuming $\Sigma = I_N$ to study the *Power* of the tests as a function of sample size.

In each simulation run we tested the distribution fit for the (i) multivariate (white Gaussian) SIRP X and the quadratic form p obtained (ii) with a known covariance Σ and (iii) with covariance Σ^* estimated from the data.

In Table 1 we show the results for the uncorrelated bivariate Gaussian (H_0) and samples of 200 data points. Results from the seven GOF tests applied to the data (four for multivariate normality of X and three for the univariate quadratic form p) are consistent with their expected values (i.e. percent rejections (α^*) are close to their GOF test nominal significance levels $\alpha = 0.1, 0.05, 0.01$).

Consistency between expected (α) and empirical (α^*) significance levels was assessed by deriving approximate confidence intervals for the true α via the usual Binomial distribution approach:

$$\sigma_\alpha(m) = \sqrt{\frac{\alpha(1-\alpha)}{m}}$$

for $\alpha = 0.1, 0.05, 0.01$ and $m = 1000$ the number of replications in the simulation. For example, for $\alpha = 0.1$ we have:

$$\sigma_{0.1}(1000) = \sqrt{\frac{0.1 * 0.9}{1000}} = 0.0095$$

Then, an approximate 95 percent confidence interval (c.i.) for α is obtained using, as half width, three standard deviations, i.e. $3 * \sigma_\alpha(m)$ and center in α^*.

For example, for large sample sizes (i.e. 200 data points) and $\alpha = 0.1$, the performance of the quadratic form p^* obtained with the *estimated or sample covariance* Σ^*, is acceptable. The distance between the corresponding theoretical and empirical significance levels, $|\alpha - \alpha^*|$, for *P ES/KS*, is $|0.1 - 0.106| = 0.006$, well within a half width of $3 * 0.0095 = 0.028$.

There is one caveat regarding our adaptive procedure for the KS GOF test for the quadratic form p^*. It is known that KS is a conservative test when the parameters are unknown and estimated from the data. We followed the approach in Goel (1982), suggested by Allen (1978). We adjusted the significance level by using four times the nominal level α, to test for that level (e.g. we test at $\alpha = 0.4$ and report at $\alpha = 0.1$). Observe how, for the larger samples ($n = 200$) this approach works well and allows us a criterium to assess the degradation of the estimations as sample sizes shrink.

In Table 2 we report similar simulation results, now for samples of size 100. Most GOF tests are still close to their nominal significance levels. There is one exception: the GOF test for p^* using Σ^*, the covariance matrix estimated from the data (P ES/KS in the tables). For example, its empirical significance level α^*, for $\alpha = 0.1$, has now gone up to 0.162. We can attribute it to a *loss of efficiency* in the estimation of Σ as the sample size decreases. Assessing the efficiency of these estimations (p^*, Σ^*, for decreasing sample sizes, is of practical concern. In field applications the covariance matrix of the SIRP is unknown and must be estimated from the available data.

Results for Table 3 are obtained for samples of size 50. As the sample size decreases, the GOF test for p^*, when the covariance matrix Σ is estimated from the data, continues deteriorating. The empirical significance level α^*, for this test has now gone up to 0.192, for $\alpha = 0.1$.

In Table 4 we present similar results for *four* variates and sample sizes of 200 data points. Observe, for this large sample, that the distance $|\alpha - \alpha^*|$ is still within the half width criteria. The GOF test for p^*, when Σ is estimated, deteriorates faster in the case of four variates ($N = 4$) as sample size shrinks, than the previous $N = 2$ case, measured by the half width criteria.

In Table 5 we observe the same type of results, now for 100 data points. Notice the rapid deterioration of the GOF test for the quadratic form p^*, when Σ is estimated from the data. In Table 6 we show similar results, now for samples of 50 data points. The test for p^*, when Σ is estimated from the data, deteriorates further more.

In Table 7 we show results for eight variates and sample sizes of 200 data points. Observe here that all GOF tests are close to their nominal significance level, except the one for p^*, when Σ is estimated from the data. As expected, when the number of variates N in the multivariate SIRP X increases, the same performance of the GOF test for the quadratic form p^*

requires a larger sample. Hence, where samples of 200 data points were sufficient for the fit of the bivariate Gaussian, it is no longer so, for the Gaussian with $N = 8$ variates. And becomes worse in the case where the covariance matrix Σ is unknown and estimated from the data.

In Table 8, for a sample size of 100 data points and in Table 9, for $N = 8$ variates and 50 data points, the performance of p^*, when the covariance matrix Σ is estimated, is $\alpha^* = 0.199$. The other GOF tests are within a half width of their respective nominal levels α.

The above GOF tests are not joint tests. Therefore, if some of them, *isolated*, depart from their expected value, this is not necessarily indicative of statistical problems. In the long run we will expect some of these tests to reject H_0, by chance, when it is true ($k * \alpha$ times, where k is the number of tests performed and α is the corresponding significance level).

We also explored the problem of correct identification of the SIRP process X, under the alternative hypothesis H_1 (i.e. when the true distribution of the SIRP X is not $MVN_N(0, I_N)$). We simulated the SIRV $X \sim MVN_N(0, \Sigma)$, $\Sigma \neq I_N$. In particular, we simulated a multivariate normal with covariance matrix equal to its correlation matrix, with all non diagonal entries $\rho_{ij} = 0.5$, $i \neq j$ (i.e., with medium correlation). In this case, we assessed (i) the effect of incorrectly specifying the covariance matrix, (ii) the effect of estimating Σ directly from the data and (iii) the power of the SIRP model to identify alternative distributions when they are true.

In Table 10 we show the results when simulating the SIRP X from a bivariate correlated Gaussian, with samples of size 200. We obtained poor agreements between empirical and nominal significance levels for the GOF tests for the quadratic form p, when the test procedure (erroneously) assumes the SIRP has covariance $\Sigma = I_N$. The same occurs in Table 11 and 12, for samples of size 100, 50, respectively. Finally, in Table 13 we show the same problem for Gaussians with eight variables and sample sizes 200.

The distance $|\alpha - \alpha^*|$, between nominal and empirical significance levels can be interpreted in two ways. First, it is an indication that the quadratic form p can actually be used to discriminate between H_0, H_1 (i.e. between distributions of SIRP's) with high Power. Then, the distance $|\alpha - \alpha^*|$ warns about the dangers of ill-specifying a covariance matrix Σ. The first interpretation, discrimination, can be useful in simulation studies, to assess, say, the minimal sample sizes required for identification problems. The second interpretation, effect of estimation of Σ from data, can be useful to assess minimal sample sizes required when performing estimations in the field.

Phase I of our validation study shows how, for the special case of the Gaussian SIRP, (i) the computer implementation of the theoretical model holds; (ii) we can accurately test for the fit of the quadratic form p, for samples down to size $n = 50$ and number of variates up to $N = 8$, when the covariance matrix is correctly specified; (iii) the quadratic form p^*, obtained with the covariance matrix Σ^* estimated from the data, approximates reasonably well that of $p = X'\Sigma^{-1}X$, for large samples (say of size $n = 200$ and above) but not accurately enough for medium ($n = 100$) or small ($n = 50$) ones. Finally, (iv) the new model of the SIRP X can be used to effectively discriminate, through its quadratic function p, an incorrectly specified (alternative) SIRP model.

It is well known that the exact distribution of the quadratic form p^*, obtained when the SIRP process is Gaussian and the covariance matrix Σ is estimated from the data, is not Chi Square but *Beta*. However, we have intentionally used the theoretical χ^2_N distribution assumed by the SIRP model with a specific objective in mind.

In other SIRP processes of interest, especially in the Multivariate K SIRP, we do not know the exact distribution of the quadratic form p^*, arising when the covariance matrix Σ is estimated from the data. Such distribution has only been obtained for the present (Gaussian) case. In the general case, only the theoretical distribution of the SIRP X, which in turn yields the theoretical distribution for the quadratic form p, is available. And we have to make-do with that.

Therefore, here we are investigating the efficiency loss resulting when we use the approximate distribution of p^*, that results from the substitution of Σ^* for Σ. In the Gaussian case developed above, the resulting quadratic form p^* becomes now only approximately distributed Chi Square (instead of exactly Beta). However for samples of size 200 such approximation works well. We assume that when the multivariate SIRP X is K-Distributed, the sample sizes would not be smaller than $n = 200$ either.

5 Phase II: Univariate SIRP's

In this phase we empirically demonstrate, in one dimension, how the SIRP model properties hold, even for very small sample sizes. We implement a (univariate) K-Distributed SIRP model via Monte Carlo. We perform several statistical transformations that finally provide easily invertible random variables. With these transformations we (i) generate and (ii) evaluate de CDF of these random variables in a quick and easy way suitable for a Monte Carlo procedure. The numerical evaluation of Bessel Functions (for the inversion and CDF evaluation of the SIRP X and the resulting quadratic function p) do pose significant and techncal round-off problems in simulation programs. We have avoided such problems in the current one dimension, via the convenient set of statistical transformations presented below.

A univariate SIRP X is just a special case of the general SIRP for $N = 1$. Hence, all properties of the theoretical model should also hold, as with $N > 1$. We still define $X = S * Z$. Only now $Z \sim N(0, 1)$, is a standard normal random variable and X is also univariate. We investigate the case of the K Distribution, following the Kaman roadmap given in the Introduction section above, through a special case: the **univariate Laplace**. This (univariate) K-Distribution is easily invertible and hence suitable for a Monte Carlo study. A random variable X is Laplace distributed if:

$$f_X(x) = \frac{1}{2\lambda} exp(-\frac{|x-\mu|}{\lambda}), \quad \lambda > 0$$

To obtain a Laplace univariate SIRP X, let the random variable $w \sim exp(1)$, i.e. exponential with mean unit. Making the transformation $y = \sqrt{2w}$ we obtain a Rayleigh distributed random variable y, with $E(y^2) = 2$ and density function:

$$f_Y(y) = y\, exp(-\frac{y^2}{2}); \quad y > 0$$

However, the resulting covariance matrix Σ_X of the SIRP $X = y * Z$ is, by definition :

$$\Sigma_X = E(y * Z)(y * Z)' = E(y^2)\Sigma_Z \neq \Sigma_Z$$

Hence, such a Rayleigh distributed y is not convenient for our simulation study since $\Sigma_X \neq \Sigma_Z$. We seek an SIRP X with the same covariance matrix Σ_Z as SIRP Z. Therefore, we transform our original random variable y to one with a unit expectation by redefining:

$$s = \frac{y}{\sqrt{2}} = \sqrt{w}$$

The resulting random variable s, has now expectation $E(s) = 1$, as assumed in the SIRP theory. This yields an SIRP process $X = S*Z$, with covariance matrix $M = \Sigma_Z = \Sigma_X$. The density of the transformed variable s is now:

$$f_S(s) = 2s\, exp\,(-s^2)$$

To obtain the distribution of the quadratic form $p = x'x = x^2$, following the SIRP model in the Kaman report, we substitute $f_S(.)$ in $h_N(.)$ for $N = 1$:

$$h_N(p) = \int_0^\infty s^{-N} exp(\frac{-p}{2s^2}) f_S(s) ds = \sqrt{\pi} exp(-\sqrt{2p})$$

From the SIRP theory, the distribution of the quadratic form p is then:

$$f_P(p) = \frac{1}{\sqrt{2\pi}\sqrt{p}} h_1(p) = \frac{1}{\sqrt{2p}} exp(-\sqrt{2p})$$

This is still not a simple enough distribution for testing goodness of fit in a Monte Carlo study. It is more convenient to find an equivalent, well known variable with an easily invertible distribution. We perform the transformation $t = \sqrt{2p}$ and obtain the random variable $t \sim exp(1)$, exponentially distributed with mean unit, easily invertible for CDF evaluation.

We can then test that the quadratic form p is distributed according to the SIRP model above derived (H_0), by testing that the distribution of the transformed variable $t = \sqrt{2p}$ is exponential with unit mean.

Hence, process $x = s * z$, with s, z and $h_N(p)$ defined above, is just a (univariate) SIRP. Following the theory developed in the Kaman Report, we obtain the distribution of the resulting SIRP X as:

$$f_X(x) = \sqrt{2\pi}|\Sigma|^{-1/2} h_N(p) = \frac{1}{\sqrt{2}} exp\,(-\sqrt{2}|x|)$$

and we recognize it as a Laplace Distribution with $\lambda = \frac{1}{\sqrt{2}}$.

However, the distribution of this resulting SIRP process X can still be simplified to perform efficiently and quickly in a Monte Carlo study. Making the transformation $u = \sqrt{2}|x|$, we find an equivalent random variable with a better suited distribution. We can easily invert this variable u in our simulation, for it follows a univariate exponential distribution with mean unit.

We sampled for $n = 25, 50, 100, 200$ from the univariate SIRP $x = s * z$ discussed above. The variables x, s, p were tested, as in Phase I, for their GOF. The empirical significance levels α^* were obtained for $\alpha = 0.1, 0.05, 0.01$ (Tables 14 through 17). As in the previous section, we assess the model by the distandce between the theoretical α and its corresponding empirical significance level α^*, again using the half width criteria of the previous section.

In Table 14 we show the simulation results for 5000 replications of batches of large sample sizes (e.g. 200). Observe the close agreement between α and α^* for significance levels $0.1, 0.05, 0.01$, obtained using as half width the same $3{*}\sigma_\alpha(m)$ criterium used before. Agreement is obtained in all three GOF tests: for the SIRP process $x = s * z$, for the quadratic form $p = x^2$ and for the Rayleigh distributed variable S, which drives the SIRP.

Table 15 shows similar results, now for 10,000 replications and samples of size 100. We can still observe a close agreement between α and α^*, for significance levels $0.1, 0.05, 0.01$.

Tables 16 and 17 show the same type of results for 10,000 and 20,000 replications and samples of sizes 50, 25, respectively. We observe (as measured by $|\alpha - \alpha^*|$) some deterioration in the empirical efficiency of the GOF tests, as sample sizes decrease, and for statistics x, p.

We also considered alternative distributions H_1 to the null hypothesis H_0 and assessed the efficiency of an SIRP to reject such false hypothesis ($Power$). Instead of the previously used SIRP $X = s * z$, distributed Laplace with parameter $\lambda = \frac{1}{\sqrt{2}}$ (H_0), we generated the alternative SIRP X, using $X = y * z$ instead (H_1). This way, we obtained a related K-Distributed SIRP, but with different parameter. In what follows, we investigate the sample size requirements to differentiate one SIRP from the other. In Tables 17 through 20 we show our simulation results, for 5000 replications and for $n = 100, 50, 25, 10$, respectively.

In Table 18 we show, for samples of size ($n = 100$), how this second SIRP X is correctly identified (is it no longer distributed Laplace with parameter $\lambda =$

$\frac{1}{\sqrt{2}}$). For example, the GOF test applied to the original SIRP X, achieves an empirical significance level (Power) $\alpha^* = 0.46$, several times larger than its nominal $\alpha = 0.1$ (i.e. rejects the false hypothesis H_0 46 percent of the time). However, the GOF test applied to the *quadradic* form p, is two times larger than that of X, yielding a (Power) empirical significance $\alpha^* = 0.84$ (i.e. the test on p is much more *Powerful*).

In Table 19 we show similar results, for sample sizes of $n = 50$. It is still plain that the (erroneously) hypothesized distribution of the SIRP X (H_0) is correctly rejected even with such small sample size. The empirical level $\alpha^* = 0.59$ for the fit of p is still twice as large as that ($\alpha^* = 0.26$) of the GOF test for X, at $\alpha = 0.1$. In Table 20 we show similar results for samples of size $n = 25$. These simulation results show that (*i*) it is possible to differenciate two closely related but different SIRP, even with such reduced sample size. And that (*ii*) the empirical performance (power) of the quadratic form statistic p ($\alpha^* = 0.34$) is better than that of the original SIRP X ($\alpha^* = 0.18$) for such differenciation purposes.

It is not until the sample size decreases to 10, in Table 21, that it is possible to confound these two closely related (but different) SIRP processes. Table 21 shows, for 5,000 replications, how the nominal $\alpha = 0.1$ is now close to the emprical $\alpha^* = 0.13$, when we are testing for the GOF, using the original SIRP X directly. However, the GOF test performed through the quadratic function p yields an empirical $\alpha^* = 0.19$ (i.e. it is still capable of differenciating between the two distributions).

Such better empirical small sample power of the quadratic form p is a strong and positive result in favor of the SIRP model. It shows that a GOF test on an SIRP is more powerful if performed through the quadratic function p (univariate) than if performed directly on the original (possibly multivariate) SIRP X, even when such a multivariate GOF test is available.

From Phase II, we conclude that (*i*) it is possible to test for the GOF fit of a (univariate) SIRP, directly on X or through its quadratic form p, derived following the SIRP theoretical model. Also (*ii*) that it is possible to perform this equivalent test with excellent results, for samples as small as 50 data points. Finally, that (*iii*) when the postulated (H_0) model is not true, even when it is closely related as in the above case, it is possible to detect and reject such a false hypothesis. Even more, that such discrimination can be performed through the quadratic form p even more efficiently (i.e. with a larger Power) than through the GOF test performed directly on the original SIRP process X.

6 Conclusions

This Monte Carlo study has shown, for the general case of a (univariate) K-Distributed process X, that the computer implementation of a the theoretical *SIRP* model X works as intended. Also, that we can use the quadratic form p as a powerful statistic to test whether the distribution of the SIRP X is the one theoretically specified (e.g. that an incoming radar signal *is* of a pre-specified type) for samples as small as 50 data points.

Moreover, we have shown how it is also possible to discriminate an erroneously postulated SIRP model (e.g. that an incoming signal is *not* of a pre-specified type) on the basis of the GOF test for the univariate quadratic form p. And we have shown how such a GOF test based on the quadratic form p is **more powerful** than the GOF test directly based on the original (multivariate) SIRP process X, were such multivariate test available.

These results may be of particular importance in radar modelling studies. For, if the (Non Gaussian) distribution of certain types of radar clutter can be prespecified, a GOF test using the statistic p may be implemented to identify these patterns more effectively.

Our results verify two important issues involving the implementation of the new theoretical SIRP model. First, the accuracy of the small sample *generation* of Non Gaussian radar clutter via the product $X = S * Z$. Then, the assessment of the small sample Power of the quadratic form p as a GOF test statistic for the **correct identification** of the distribution of the SIRP X.

Future research on process identification and on degradation of estimations of p^*, Σ^*, for higher dimensional K-Distributed SIRP processes, is currently under preparation.

In addition, we showed for $N > 1$ dimensions and the special case of a multivariate Gaussian SIRP X, that the quadratic form p can be effectively used to discriminate between different SIRP when the covariance Σ is known. And that the distribution of statistic p^*, obtained when the covariance matrix Σ is unknown and the sample sizes are large (say $n > 200$), can be approximated by the (theoretical) distribution of the statistic p. Similar investigations, pursuing identification and degradation studies for general (K-Distributed) multidimensional SIRP processes, are currently also in preparation.

7 Acknowledgements.

The author gratefully acknowledges the cooperation and interaction with SIRP model developers Drs. Don Weiner and Muraly Rangaswamy (Syracuse University) and with Dr. Jim Michels (Rome Lab Summer Research Focal Point). We appreciate the computer support of the CASE Center of Syracuse University.

We also appreciate the valuable comments of several colleagues who read the original manuscript. Among them, Dr. J. Michels, of Rome Lab, Dr. W. Smith, of Texas A. and M. and two of his CIS editors and Dr. B. Schmeiser of Purdue University. Their valuable suggestions greatly improved the presentation of this document. Responsibility for the results rests with the author.

This work was performed while the author was a Summer Research Associate for the Air Force Office of Scientific Research, at Rome Lab, Griffiss Air Force Base, New York. A contract for the continuation of this project in the directions above described, was recently awarded.

MONTE CARLO INVESTIGATION

OF A MODEL FOR THE

GENERATION OF NON GAUSSIAN RADAR CLUTTER.

Jorge Luis Romeu

Department of Mathematics

SUNY-Cortland

jromeu@suvm.acs.syr.edu

Presented at the 25th

Symposium on the Interface Between

Statistics and Computer Science.

San Diego, California

April of 1993.

Acknowledgement.

This research was performed while the author was an Air Force Summer Research Associate at Rome Lab, Griffiss AFB, NY. This Presentation is supported by a Summer Research Extension Program Grant from the AFOSR awarded to pursue this work.

OUTLINE:

========

* Motivation of the Study

* Introduction and Background

* Objectives of the Study.

* Statistical Distribution of Interest

* Implementation Problems

* Generation of Multivariate Alternatives

* The Univariate SIRP

* The Quadratic Form

* Current Developments

* Conclusions

I. MOTIVATION OF THIS STUDY:

* Need for Empirical Study of Engineering Models

- *Example: Circuit Design in Electrical Engineering*
- *Theoretical Problems: Ideal Conditions in Models*
- *Implementation Problems: Building the Circuit Board*
- *Discrepancy Between Theoretical/Empirical Performance*

* Need for Monte Carlo Study of Statistical Models

* Theoretical (Ideal) Conditions: Assymptotic Methods

- *Minimal Sample Size Requirements* (n^*)
- *Empirical Critical Values for* $n < n^*$
- *Determination of Other Influencing Factors*
- *Performance Degradation*
- *Example: Mardia's Multivariate Normality Tests*

* Implementation Problems: Numerical Approximations.

- *Theoretical Closed Form Solutions*
- *Numerical Implementation on a Digital Computer*
- *Eigenvalues/Eigenvectors and Convergence Methods*
- *Effect of Hardware*
- *Example: Koziol's Multivariate Normality Test*

* Effect of the Estimation of Parameters

- *Sample size effects*
- *Small Sample Distribution*

II. INTRODUCTION AND BACKGROUND

* SIRP Definitions

- *The Conditional Distribution of the Process* $X = S * Z$
- *The Distribution of the Quadratic Form* p
- *The Distribution of the Process Driver* S
- *The Covariance Matrix of the Process* X

III. OBJECTIVES OF THE MONTE CARLO STUDY

* Assessment of the Numerical Implementation

* Accuracy of the Generated Data: Fit

* Simultaneous Control of Both, SIRP X and p

* Parameter Estimation Effects:

- *Covariance Matrix of the SIRP X*
- *Resulting Distribution if Unknown*
- *Sample Size Requirements*
- *Effects of Number of Covariates*
- *Effect of Inter-Covariate Correlation*
- *Other Possible Factors Producing Degradation*

* Assessment of the Quadratic Form p

- *Distribution of* p *Under Known Covariance*
- *Distribution of* p *Under Unknown Covariance*
- *Approximations of the Distribution*
- *Sample Size Revisited*

IV DISTRIBUTION OF INTEREST

*** Radar Clutter Modeling Distributions**

*** The Multivariate K-Distribution**

- *The Problem of Lack of a Fit Test*
- *The Modified Bessel Function and Computer Implementation*

*** Implementation Problems**

- *Simultaneous Control of the Fit of X and p*
- *Generation of the Multivariate Gaussian SIRP*
- *Unknown Covariance Matrix and Distribution of p*
- *Generation of the Special Case of K-Distribution*
- *Required Transformations for Testing X, S, p*
- *The Power of the Quadratic Form p*
- *Practical Importance: Uses of p*

V. CURRENT DEVELOPMENTS OF THE STUDY

*** Further Studies on the Multivariate K-Distributions**

- *Effect of Shape Parameters on Accuracy*
- *Other Possible Factors*

*** Further Studies on Estimation Procedures**

- *Other Estimation Methods*
- *Comparative Study and Selection*

VI. CONCLUSIONS

*** Need to Investigate Implementations via Monte Carlo**

- Power of Quadratic Function p

- Degradation Studies

*** Final Note From a Teacher's Perspective for an**

*** APPLIED STATISTICS GRADUATE COURSE**

- Illustrates Transformation Applications

- Illustrates use of Large Computers in Statistics

- Illustrates Statistical Consulting

* * * * *

Proceedings of The 1997 Interface Symposium. Rice U. - Houston, TX. May/97

Further Monte Carlo Investigations of a Model for Non Gaussian Radar Clutter Generation

Jorge Luis Romeu
Department of Mathematics
SUNY-Cortland, NY 13045
email: jromeu@cat.syr.edu

Abstract

In a previous Interface (Romeu, 1993) the first part of this validation study was presented. There, the generation and evaluation of a Spherically Invariant Random Process (SIRP) X, was studied for the unidimensional case. Here, we complete such investigation. First we study the distribution identification of a particular Bi-Dimensional SIRP X, for which we can obtain the closed form of the distribution. Then, we study the general multivariate case via Monte Carlo. We assess the distribution identification performance via this method, for specific sample sizes, number of variates and intercorrelation. Finally, we study the effects of these factors and of covariance estimations on the SIRP identification procedure.

1 Introduction

For this author, a radar is a black box that receives as input a sample of n iid vectors $X_i, i = 1, ..., n$, of dimension N (i.e. the SIRP X representing clutter). The radar then extracts, from this sample, an estimation of the covariance matrix Σ (say Σ^*) of the SIRP X. Finally, with the aid of Σ^*, the radar identifies the distribution of X (i.e. the radar assesses whether the incoming signal actually represents clutter or not).

It is theoretically known (Rangaswamy, 1992; Rangaswamy et al., 1991, 1992; Kaman, 1992) that an SIRP X can be decomposed into $X = S * Z$ (where S is the univariate process driver, Z is multivariate normal with covariance Σ). We can then obtain the quadratic form $p = X'\Sigma^{-1}X$, whose density $f_P(*)$ is:

$$f_P(p) = \frac{1}{2^{N/2}\Gamma(N/2)} p^{N/2-1} h_N(p)$$

$$where\, h_N(p) = \int_0^{\infty} s^{-N} exp(\frac{-p}{2s^2}) f_S(s) ds$$

Therefore, it is possible for our radar to estimate, from the multivariate input $(X_1, ..., X_n)$, both the process covariance Σ^* and a sequence of univariate, iid, quadratic forms $p_i^* = X_i'\Sigma^{*-1}X_i$, $i = 1, ..., n$, and through these identify the distribution of the incoming SIRP X.

However, for $N > 1$, the expression for $h_N(p)$ typically does not have a mathematically simple form. Hence, the corresponding $f_P(p)$ is analytically difficult to obtain, even after performing approximations, changes in variables, and other mathematical manipulations.

On the other hand, we can adapt the Ozturk approach to distribution identification (see Section 6.4 of Kaman, 1992). Ozturk's approach implements the graphical Q_n goodness of fit (GOF) test of Ozturk and Dudewicz (1990) on the quadratic form p, to assess the true distribution of X via distribution charts.

Our proposed approach is a modification of the above. First, it simulates a different reference distribution than normal. Then, it assumes that the (radar input) SIRP X is known (for it represents clutter) and hence also the quadratic form p that is obtained from it. Finally, the procedure tests (via the univariate quadratic function p) whether or not the distribution identification of the SIRP X is correct.

In addition, one can use this modified approach in two other areas. First, as a tool to evaluate several factors that affect the assessment of an SIRP: (i) the sample size n, (ii) the number of variates N (vector size), (iii) its inter-correlation ρ and (iv) the interaction among them. Then, to compare the performance of different covariance matrix estimators Σ^*. We do this by evaluating the performance of statistic Q_n.

We use the above described approach to complete the model validation in Romeu (1993). First, we study the special case of a Bi-Dimensional K-Distributed SIRP X. Then, we study the general case, using the Q_n statistic to test the GOF of the quadratic form p, obtained from SIRP X for prespecified settings (n, N, ρ).

In the rest of this paper we discuss the three sequential phases of our research:

(i) Estimation of the Empirical Distribution of Q_n

(ii) Validation of the Empirical Parameters Obtained

(iii) Evaluation of Factors and Covariance Estimators

2 Estimation of the Empirical Distribution of Q_n

Romeu (1993), for the special K-Distributed case $N = 1$, showed how the quadratic form p provides a powerful statistic for correctly identifying an SIRP, when Σ is known and even when Σ is estimated from sufficient data. Hence, we will also use p for the $N > 1$ case.

We obtain, via Monte Carlo, all necessary parameters to implement the Q_n test on the quadratic form p. To assess these Monte Carlo estimations we use the conveniently tractable, special case of the Bivariate K-Distributed ($N = 2$) SIRP X. We use shape parameter $\alpha = N/2 - 0.5$ and scale parameter $b = 1$. Its quadratic form p has now the following pdf:

$$h_N(p) = \frac{\sqrt{p}^{N/2-0.5-N/2}}{\Gamma(N/2-0.5)2^{N/2-0.5-1}} \times K_{N/2-N/2+0.5}(\sqrt{p})$$

$$= \sqrt{\frac{\pi}{p}} \frac{exp(-\sqrt{p})}{\Gamma(\frac{N-1}{2})2^{\frac{N-2}{2}}}$$

$$K_{\frac{N}{2}-\alpha}(p) = K_{0.5}(p) = \sqrt{\frac{\pi}{2}} \times \sqrt{\frac{1}{x}} exp(-\sqrt{p})$$

$$f_P(p) = \frac{1}{2^1\Gamma(1)} \times p^{1-1}h_2(p) = \frac{1}{2}p^{-1/2}exp(-\sqrt{p})$$

Such pdf is still too complex for us to work with, directly. And certain transformations are required to simplify our generation and testing work.

Under the transformation $\frac{w^2}{2} = \sqrt{p}$, the resulting w is distributed Rayleigh and is easy to test for GOF. It also allows (i) to validate our Monte Carlo experiment and (ii) to compare the Power of the empirical Q_n test with an exact GOF test for p. For, if $p \sim F_P$ then $w \sim$ Rayleigh. We implement the latter as our exact test, for its ease and computational speed.

Accordingly, the driver random variable S, for this special case of K-Distributed SIRP $X = S * Z$ is:

$$f_S(s) = \frac{2}{\Gamma(\alpha)2^\alpha} s^{2\alpha-1} exp(\frac{-s^2}{2}) = \sqrt{\frac{2}{\pi}} exp(-s^2/2)$$

which, under the transformation $y = s^2/2$ becomes:

$$f_Y(y) = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{y}} exp(-y) = \frac{1}{\Gamma(1/2)} y^{\frac{-1}{2}} exp(-y)$$

i.e., a Gamma distribution with parameters $\lambda = 1, r = 1/2$. This allows an easier and faster way to generate X: generate a Gamma y, obtain $s = \sqrt{2y}$ and multiply by bivariate normal Z.

We also investigated other potentially interesting cases of SIRP, for $\alpha = N/2 - 0.5$ and $b = 1$:

$$For\, N = 3: \; f_P(p) = \frac{\sqrt{\pi}}{4\Gamma(1.5)} e^{-\sqrt{p}}$$

$$For\, N = 4: \; f_P(p) = \frac{\sqrt{p\pi}}{8\Gamma(2)} e^{-\sqrt{p}}$$

$$For\, N = 8: \; f_P(p) = \frac{p^2\sqrt{p\pi}}{3!2^7\Gamma(3.5)} e^{-\sqrt{p}}$$

none of which yielded interesting pdf cases.

As shown above, we lack in general the closed form of the CDF, F_P and/or the pdf, f_P. However, F_P is completely determined by S, h_N, Σ. And we can obtain any required empirical result by Monte Carlo. For purposes of testing with Q_n, we will circumvent the lack of F_P through Monte Carlo simulation.

Let the $Q_n = (U_n, V_n)$ GOF test, be defined:

$$U_n = \frac{1}{n}\sum_i^n cos\theta_i\,|Z_i|$$

$$V_n = \frac{1}{n}\sum_i^n sin\theta_i\,|Z_i|$$

$$with\, \theta_i = \pi\int_{-\infty}^{m_{i:n}} f_P(t)dt = \pi F_P(m_{i:n})$$

where $m_{i:n}$ is the i^{th} order statistic from the ordered sample $p_1 < p_2 < \ldots < p_n$. Let these n samples be obtained from the simulated SIRP $X = S * Z$. And let F_P, f_P be, respectively, the distribution and density functions of quadratic form p. In addition $Z_i = \frac{p_i - p_{avg}}{S_p}$, are the standardized p_i, for $i = 1, \ldots, n$.

To obtain the angles θ_i, $i = 1, \ldots n$, which yield the endpoints (U_n, V_n) we require the distribution function F_P. Since, in general, we cannot obtain F_P analytically, we approximate $F_P^*(*)$, via Monte Carlo.

First, we generate NTOT samples of size n each, from the same SIRP X, defined above. Then, for $j = 1, \ldots NTOT$, we then obtain the ordered sample $p_{1,j} < \ldots < p_{n,j}$ of quadratic forms. Finally, we calculate:

$$m_{i:n}^* = \frac{1}{NTOT}\sum_j^{NTOT} p_{i,j},\; for\, i = 1, \ldots n$$

From the empirical $m^*_{i:n}$ values, we get the empirical CDF's $F^*_P(m^*_{i:n})$, also by Monte Carlo:

$$F^*_P(m^*_{i:n}) = \frac{pp < m^*_{i:n}}{NTOT}, i = 1, \dots n$$

where $pp < m^*_{i:n}$ is the number of simulated quadratic forms p, out of the total NTOT generated in the Monte Carlo experiment, that are smaller than the corresponding order statistics $m^*_{i:n}$.

We thus evaluate $F^*_P(*)$ at each of the empirical order statistic $m^*_{i:n}$ of the quadratic form p, of a given sample size n. From these values, the empirical angles $\theta^*_i = \pi F^*_P(m^*_{i:n})$ are easily obtained. Since we also have, for the Bivariate K-Distributed SIRP X, the exact distribution $F_P(*)$ of the quadratic form p, we can compare them to assess our Monte Carlo procedure.

Using these values we obtain, through a second Monte Carlo experiment, the empirical estimators of the parameters $E(U_n), E(V_n), \sigma^2_u, \sigma^2_v, \rho_{uv}$, required for implementing the Q_n GOF procedure.

The set of empirical values $F^*_P(*), m^*_{i:n}, \theta^*_i$, $i = 1 \dots n$ are calculated only once for each parameter setting (n, N, ρ), via Monte Carlo.

3 Validation of the Empirical Distribution

A total of $NTOT = 10,000$ samples of size $n = 25, 50, 100, 200$ of an N-variate SIRP X were generated on Syracuse University's IBM 3090, using the IMSL statistical library. We used vector size values of $N = 2, 4, 8$ with covariance matrix Σ having unit in the diagonal and ρ elsewhere. We used $\rho = 0.0, 0.5, 0.9$. The experiments for $N = 4, 8$ were implemented at the Cornell Supercomputer (vector facility) given their extensive run times.

For each simulated sample (i.e. each prespecified setting (n, N, ρ)) we calculated the quadratic forms p of the SIRP X and obtained the empirical estimations of (i) the corresponding order statistics $m_{i:n}$, (ii) distributions $F_P(m_{i:n})$, and (iii) angles θ_i, for $i = 1, \dots n$. These three sets of estimations $(m^*_{i:n}, F^*_P, (m^*_{i:m}), \theta^*_i)$ were used in the calculation of NTOT Monte Carlo $Q^*_n = (U^*_n, V^*_n)$ statistics. Finally, these NTOT Q^*_n values provided the empirical mean, variance and correlation of the bivariate distribution of statistic Q_n.

We assessed the empirical estimations of $F_P(*), m_{i:n}, \theta_i$, $i = 1, \dots, n$, and of the mean, variance and correlation of the bivariate distribution of Q_n. The GOF statistic Q_n was obtained with reference to F_P, the known distribution of the quadratic form p. We also assess the validity of our proposed statistic Q_n by comparing the Monte Carlo derived Q^*_n GOF test results with those from the special case of the Bivariate K-Distributed SIRP X.

Finally, by generating from the same SIRP X used before, we obtain the statistic $h(Q_n) = h(U_n, V_n)$:

$$\frac{1}{1-\rho^2_{uv}}\{\frac{(U_n - E(U_n))^2}{\sigma^2_u}$$

$$-2\rho_{uv}\frac{(U_n - E(U_n))(V_n - E(V_n))}{\sigma_u \sigma_v} + \frac{(V_n - E(V_n))^2}{\sigma^2_v}\}$$

distributed approximately as a χ^2_2.

We implement the empirical GOF tests for p via $h(Q_n)$, using Monte Carlo estimations to get the expected values and variances of (U_n, V_n), (i.e. $E(U_n), E(V_n), \sigma^2_u, \sigma^2_v$ and intercorrelation ρ_{uv}).

To compare the performance of both GOF test for SIRP X (the Q_n GOF empirical test on p with the **Exact** GOF test on the transformation w) we used the corresponding percent rejections (P_r). We thus validated our proposed Q_n testing approach for $N = 2$.

We investigated intensively these two GOF tests (Exact and empirical). A graphical analysis, based on the settings (n, N, ρ), confirms the existence of a small Bias for Q_n, even for a large n. A regression analysis confirms that (i) sample size is a significant factor, but that (ii) intercorrelation ρ is not. And that (iii) the effect of sample size (n) decreases with n. The Bias detected for a large sample size (n) was estimated at about 15 percent of the nominal level α. We conclude that the proposed Q_n GOF test of p is adequate for assessing the fit of a multivariate SIRP X.

4 Evaluation of Factors and Estimators

The Q_n GOF test can also be used as a research tool to assess different characteristics of an SIRP.

We can implement the Q_n GOF test to (i) *assess the performance of the estimator* p^* and to (ii) *analyze the effects of specific characteristics* (e.g. n, N, ρ) of the SIRP X. We perform another Monte Carlo experiment. from the same SIRP X used before. This time we compare the two quadratic forms: $p = X'\Sigma^{-1}X$ and $p^* = X'\Sigma^{*-1}X$ (obtained using covariance (Σ) or its maximum likelihood estimator) to assess the performance of Σ^*.

In Table 1 we show the percent rejections (P_r) for an experiment implemented on various settings (n, N, ρ). We first compare the performance of the Q_n test (using the known covariance Σ) against that of the Exact

GOF (for the Bivariate K-Distributed SIRP). To investigate the effects of estimating the covariance matrix (Σ^*) in the empirical Q_n test, we used statistics p, p^* (denoted **Known** and **Estim** in Table 1). We implemented multiple regression and ANOVA on the model: $P_r = f(n, N, \rho)$. These analyses results reconfirmed that (i) sample size is a significant factor and that (ii) intercorrelation ρ is not significant in the performance of the Q_n GOF test. Also, that (iii) as the sample size increases, the performance of Q_n (obtained with the estimated covariance matrix) gets closer to the one obtained using the known covariance matrix. Finally, we obtained estimates of the adequate sample sizes required ($n = f(N)$) to safely implement the adapted Q_n GOF approach using Σ^*. These were found to be: (i) for $N = 2$ we need $n > 25$, (ii) for $N = 4$ we need $n > 100$, (iii) for $N = 8$ we need $n > 400$.

5 Summary and Conclusions

We have shown that (i) the theoretical SIRP radar clutter modeling procedure given in Kaman (1992) is valid, computer implementable and fulfills its intended purpose of generating SIRP's and identifying an incoming signal (Romeu, 1992). Also, (ii) that the quadratic form p is a good statistic for testing the GOF for the general ($N > 2$) K-Distributed SIRP (that characterizes clutter). In addition, we have shown that (iii) the Q_n GOF test, a Monte Carlo based testing methodology, is an adequate (though conservative) general procedure for identifying a multivariate SIRP X.

In addition, we have investigated the maximun likelihood estimator of the covariance matrix Σ of an SIRP. If the sample size (n) is adequate for the size (N) of vector X, the estimation is good and Q_n provides good results. We have (iv) investigated the effects of three factors: (n, N, ρ). We have shown how the intercorrelation (ρ) is not significant. The other two factors, can affect the Q_n test when the (sample and vector) sizes are not adequate for its correct implementation.

Finally, the Monte Carlo testing approach developed here has wider applications than just in radar modeling. Any monitoring device with multivariate input, requiring the identification of specific patterns, is a candidate for such an approach. We forsee areas in medicine (life supporting devices) and in industrial quality control where this approach may be successfully used.

Acknowledgements

This work was done while the author was a 1993 AFOSR Summer Research Associate at Rome Lab, Griffiss AFB, NY. We acknowledge the valuable input of Drs. Don Weiner (Syracuse University) and Murali Rangaswamy (Rome Labs; Hanscomb, AFB) and particularly of our Rome Lab Focal Point, Dr. Jim Michels. Computer and logistic support from the CASE Center of Syracuse University is also appreciated.

References

Kaman Sciences (1992), *Contract Report No. F36602-89-C-0082, Task 10,* Rome Lab Technical Report.

Ozturk, A. and E. Dudewicz (1992), "A New Statistical Goodness of Fit Test Based on Graphical Representation," *Bimetrical Journal*, Vol. 34(4), pp. 403-427.

Ozturk, A. and J. L. Romeu (1992), "A New Graphical Test for Multivariate Normality," *Comm. in Statist. (Simula.)*, Vol. 21(1).

Rangaswamy, M. (1992), *Spherically Invariant Random Processes for Radar Clutter Modeling, Simulation and Distribution Identification,* Ph.D. Dissertation. Syracuse University.

Rangaswamy, M.; Weiner, D. and A. Ozturk (1992), "Computer Generation of Correlated Non Gaussian Clutter for Radar Signal Detection," *IEEE Trans. Aerosp. Electr. Sys.*.

Rangaswamy, M.; Weiner, D. and A. Ozturk (1991), "Simulation of Correlated Non Gaussian Interference for Radar Signal Detection," *Proceedings of the 25th Annual ASILOMAR Conference on Signals, Systems and Computers.*

Romeu, J. L. (1993) "Further Monte Carlo Studies of a Theoretical Model for Non Gaussian Radar Clutter Characterization," *AFOSR 1993 Summer Research Report Compendium.*

Romeu, J. L. (1993), "Monte Carlo Investigation of a Model for Generation of Non Gaussian Radar Clutter," *Proceedings of the 25th Anniversary of the Interface Symposium.*

Romeu, J. L. (1994). "Validation of Multivariate Monte Carlo Studies", chapter in *Recent Advances in Statistics and Probability.* Publ. VSP International Sciences. Netherland. 1994.

Romeu, J. L. and A. Ozturk (1993), "A Comparative Study of GOF Tests for Multivariate Normality," *Journal of Multivariate Analysis,* Vol. 46.

Romeu, J. L and A. Ozturk (1996). "A New Graphical Method for Multivariate Normality", chapter in *Multivariate Statistical Analysis in Honor of Prof. M. Siotani.* Publ. American Journal of Mathematical and Management Sciences. NY. 1996.

Table 1. Experimental Results

Test	n	N	ρ	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Exact	25	2	0.0	0.0826	0.0402	0.0064
Q_n	25	2	0.0	0.0930	0.0430	0.0094
Exact	25	2	0.5	0.0774	0.0352	0.0050
Q_n	25	2	0.5	0.0936	0.0448	0.0090
Exact	25	2	0.9	0.0870	0.0414	0.0062
Q_n	25	2	0.9	0.0936	0.0429	0.0102
Exact	50	2	0.5	0.0949	0.0465	0.0094
Q_n	50	2	0.5	0.0880	0.0450	0.0092
Exact	100	2	0.0	0.0978	0.0466	0.0098
Q_n	100	2	0.0	0.0850	0.0440	0.0116
Exact	100	2	0.5	0.0983	0.0489	0.0096
Q_n	100	2	0.5	0.0824	0.0409	0.0109
Exact	100	2	0.9	0.0968	0.0494	0.0095
Q_n	100	2	0.9	0.0828	0.0416	0.0115
Exact	200	2	0.0	0.0984	0.0460	0.0090
Q_n	200	2	0.0	0.0839	0.0442	0.0123
Exact	200	2	0.5	0.0970	0.0482	0.0096
Q_n	200	2	0.5	0.0808	0.0412	0.0109
Exact	200	2	0.9	0.1006	0.0508	0.0103
Q_n	200	2	0.9	0.0813	0.0423	0.0125
Q_n	25	4	0.0	0.0918	0.0426	0.0075
Q_n	25	4	0.9	0.0894	0.0429	0.0079
Q_n	100	4	0.0	0.0917	0.0465	0.0129
Q_n	100	4	0.9	0.0879	0.0453	0.0109
Q_n	25	8	0.0	0.0893	0.0430	0.0078
Q_n	25	8	0.9	0.0898	0.0438	0.0085
Q_n	100	8	0.0	0.0808	0.0408	0.0106
Q_n	100	8	0.9	0.0854	0.0434	0.0113
Σ	n	N	ρ	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Known	10	2	0.5	0.08480	0.03980	0.00980
Estim	10	2	0.5	0.14420	0.07840	0.02060
Known	25	2	0.5	0.09300	0.04100	0.00920
Estim	25	2	0.5	0.10020	0.05780	0.01500
Known	50	2	0.5	0.09120	0.04260	0.00900
Estim	50	2	0.5	0.08300	0.04120	0.01180
Known	100	2	0.5	0.08740	0.04440	0.01240
Estim	100	2	0.5	0.07660	0.03620	0.00820
Known	200	2	0.5	0.10000	0.06100	0.01700
Estim	200	2	0.5	0.09000	0.04100	0.01200
Known	25	4	0.5	0.10020	0.04530	0.00810
Estim	25	4	0.5	0.22920	0.14430	0.05130
Known	100	4	0.5	0.08720	0.04540	0.01136
Estim	100	4	0.5	0.10380	0.05876	0.01724
Known	25	8	0.5	0.08980	0.04200	0.00720
Estim	25	8	0.5	0.94400	0.86200	0.55580
Known	50	8	0.5	0.08648	0.04400	0.00972
Estim	50	8	0.5	0.52872	0.38176	0.17040
Known	100	8	0.5	0.08476	0.04504	0.01160
Estim	100	8	0.5	0.28084	0.18348	0.06648
Known	200	8	0.5	0.08220	0.04160	0.01190
Estim	200	8	0.5	0.16090	0.09290	0.02830