NASA Contractor Report 156850

Optimization of Satellite Altimeter and Wave Height Measurements

(NASA-CR-156850) OPTIMIZATION OF SATELLITE ALTIMETER AND WAVE HEIGHT MEASUREMENTS Final Report, Oct. 1975 - Mar. 1976 (Technology Service Corp., Silver Spring, Md.) 45 p HC A03/MF A01 CSCL 08J G3/48

N79-16506

Unclas 43647

R. P. Dooley, L. W. Brooks and E. N. Khoury

January 1979

NASA
National Aeronautics and Space Administration

Wallops Flight Center
Wallops Island, Virginia 23337
AC 804 824-3411

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1.0 INTRODUCTION AND SUMMARY

Two techniques for simultaneously estimating altitude, ocean wave height, and signal-to-noise ratio from the GEOS-C satellite altimeter data are described in this report. One technique is based on maximum likelihood estimation, MLE, and the other on minimum mean square error estimation, MMSE. Performance is determined by comparing the variance and bias of each technique with the variance and bias of the smoothed output from the GEOS altimeter tracker. Ocean wave height tracking performance for the MLE and MMSE algorithms is measured by comparing the variance and bias of the wave height estimates with that of the expression for the return waveform obtained by a fit to the average output of the 16 waveform sampling gates.

Both the MLE and MMSE algorithms achieve improved altitude tracking performance over the GEOS trackers with real data in simulated "real time". In fact, as shown in the results (see Section 2.0), both algorithms achieve close to theoretical performance in all parameter estimates (altitude, wave height, signal and noise). It is significant to note that both algorithms out-perform the GEOS tracker despite the fact that the processing was performed on a poor data tape with low sea state and only the 16 leading edge samples could be used to provide the parameter estimates. Thus, it is felt that with good data from an altimeter designed to accommodate the MLE and MMSE algorithms (better range resolution and more leading edge samples as well as usable noise and plateau gate samples), the algorithm could achieve the theoretically predicted factor of four reduction in the standard deviation of the altitude estimated as compared to the standard split-gate tracker at high sea states. Furthermore, with such a properly designed altimeter, the standard deviation of the other parameter estimates (wave height, signal-to-noise ratio, and pointing error) could be expected to be on the order of 1% of the parameters mean value.

A major problem was encountered in using the MLE and MMSE

algorithms for ground processing the GEOS altimeter data. Since in GEOS, the location in range of the 16 sample gates is varied according to the detected altitude error on a per pulse basis, it is necessary to process via the MLE or MMSE algorithms on a per pulse basis. Thus it would be impractical, from a computational standpoint, to utilize either the MLE or MMSE algorithms as an operational GEOS ground data processing system. For this reason, the GEOS data runs that have been processed have been very short and are intended to verify the algorithms performance. However, in future altimeters the range samples can be preaveraged and the algorithms described in this report could be readily implemented as either an on-board data processor or as a general data processing system.

2.0 RESULTS OF GEOS-C DATA PROCESSING

The performance of the MLE and MMSE algorithms is summarized in Table 1 for two different portions of orbit 186. The GEOS split-gate tracker performance is also shown to provide a direct comparison of the standard deviations in the altitude estimates. As shown in the table, the improved tracking algorithms reduce the standard deviation in the altitude estimate by about 25 to 30 centimeters, or 30 percent, compared to GEOS. While this in itself is certainly not considered a significant improvement, it should be noted that the measured performance is very close to the theoretically predicted performance. Thus, while the GEOS altimeter design itself limits the improved performance, achieving near theoretical performance with a properly designed altimeter would result in a factor of four reduction in standard deviation at high sea states. Of course, considerably more data runs (especially at different sea states) would be required to ultimately verify this conclusion.

The following comments are appropriate for a more comprehensive understanding of the data presented in Table 1.

1. The number of pulses averaged in the tracking loops is 7.56 for altitude (MLE and MMSE matched to GEOS) and 40.8 for the other three loops in all cases.

2. The results of two runs are shown; one containing files 69-72 and the other files 10-14. Each runs contains 12 seconds of data (1200 pulses).

3. For each run, a least squares straight line fit is made to the GEOS instantaneous altitude data. The measured altitude standard deviations on a per pulse basis for GEOS, MLE and MMSE are then computed from the residues about this line. The line is also used to determine the biases for the MLE and MMSE relative to GEOS. As an example, the straight line for files 69-72 is $y = 4.46 + 13.5155\ \Delta T$ meters where ΔT is the interpulse period and

TABLE 1

RESULTS OF GEOS-C DATA PROCESSING FOR ORBIT 186

	ESTIMATION ALGORITHM	ALTITUDE CENTIMETERS		WAVE HEIGHT (H 1/3) METERS		SIGNAL LEVEL MILLIVOLTS		NOISE LEVEL MILLIVOLTS	
		Standard Deviation	Bias	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean
Files 69-72	MLE								
	Theoretical	62.3	*2.8**	.22	*4.0**	7.1	*82.0**	.60	*9.34**
	Measured	68.7	8.7	.32	4.1	9.6	81.7	.71	9.42
	MMSE								
	Theoretical	67.2	*2.8**	.39	*4.0**	7.6	*82.0**	.89	*9.34**
	Measured	63.0	1.2	.39	2.6	12.1	76.6	2.39	10.6
	GEOS								
	Theoretical	91.0	Reference		*4.0**		*82.0**		*9.34**
	Measured	94.0	Reference						
Files 10-14	MLE								
	Measured	75.3	13.7	.35	4.2	10.7	81.2	.70	9.48
	MMSE								
	Measured	79.6	11.7	.65	3.9	15.5	77.8	3.03	10.09
	GEOS								
	Measured	105.0	Reference						

NOTE: * Values in italics correspond to a static MLE fit to the data and are used to calculate theoretical performance.

the line is relative to an altitude of 853,300 meters.

4. The MLE and MMSE theoretical performance is based on a numerical evaluation of the covariance matrix of the parameters. To perform the numerical evaluation, it is assumed that the static fit parameter values are the correct values, and that the 16 range samples are independent. The GEOS theoretical performance is based on the previously derived closed form expression for split-gate trackers and the static fit parameter values. For GEOS, it is also assumed that there are two independent samples in both the early and late gates, which is consistent with the 16 independent sample assumption for the MLE and MMSE.

The need for additional data processing at other sea states is readily demonstrated in Figure 1, which shows the measured standard deviation in the altitude estimate versus measured wave height for the MLE and MMSE and versus the static fit wave height for GEOS.

To achieve some degree of confidence in the measured mean parameter estimates (as well as for use in evaluating the theoretically predicted standard deviations), it was decided to perform a static MLE fit to one of the data runs (files 69-72). This was done as follows: The GEOS instantaneous (per pulse) altitude estimates were quantized into 23 range bins, each of 1/4 gate width (23.44 cm), centered about the mean straight line fit to GEOS altitude. The resulting distribution of Gate #10 is shown in Figure 2. Now, the resulting returns (each of the 16 range samples) for each bin were averaged and then a MLE performed on this averaged return for each bin. Figure 3 shows the MLE fit for bin zero. Finally, a weighted (proportional to number of pulses in a bin) average of the 23 MLE estimates is computed to obtain what is referred to as a "static MLE fit" to the data. These are then the parameter values shown in Table 1. Considering the excellent agreement between the static fit values and the measured values and the effort required to perform a static fit to the data, only measured parameter values should be used in any future data runs.

Figure 1

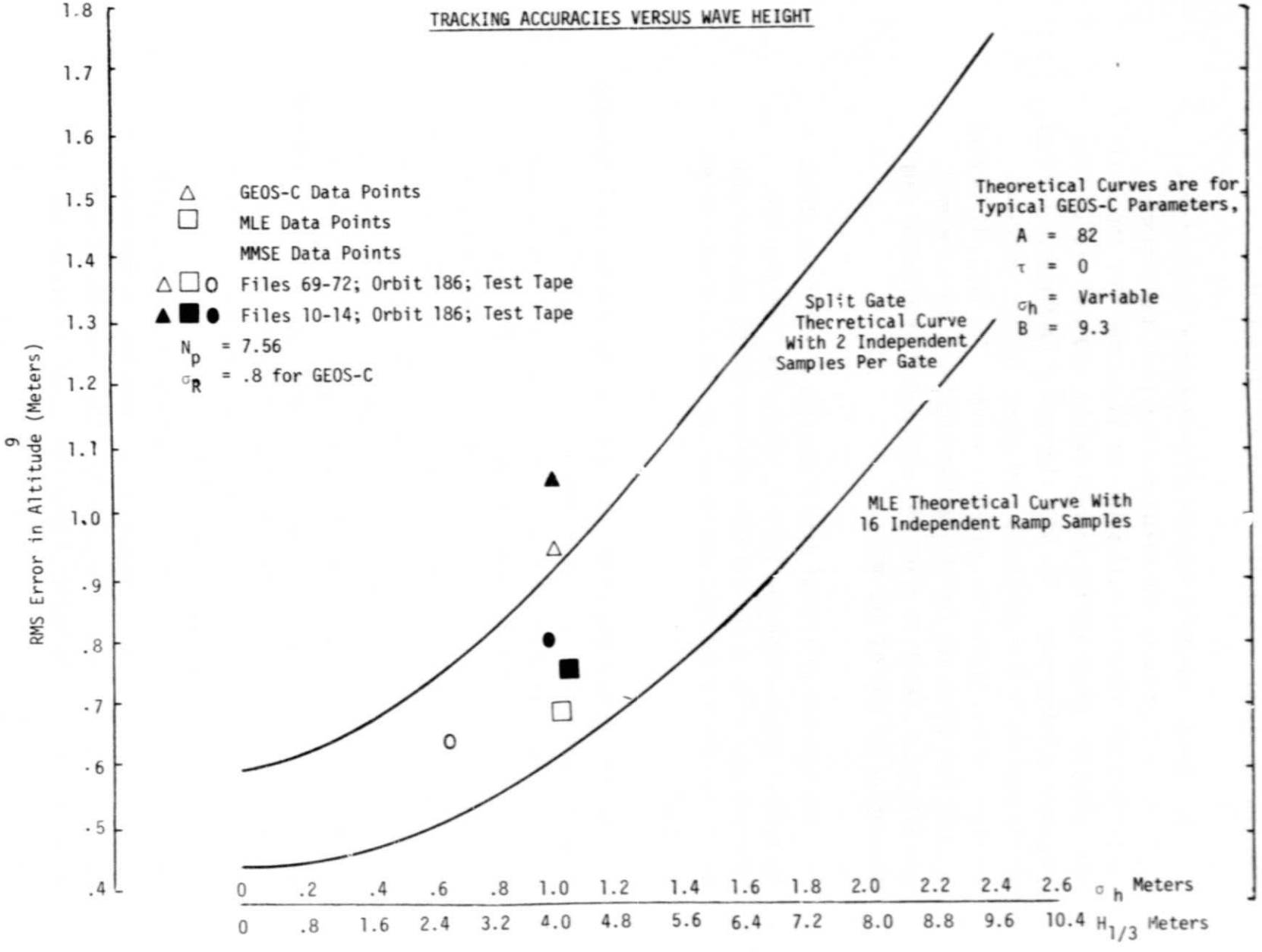

TRACKING ACCURACIES VERSUS WAVE HEIGHT
RMS Error in Altitude (Meters) σ
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
△ GEOS-C Data Points
□ MLE Data Points
MMSE Data Points
△ □ O Files 69-72; Orbit 186; Test Tape
▲ ■ ● Files 10-14; Orbit 186; Test Tape
N_p = 7.56
σ_R = .8 for GEOS-C
Split Gate Thecretical Curve With 2 Independent Samples Per Gate
MLE Theoretical Curve With 16 Independent Ramp Samples
Theoretical Curves are for Typical GEOS-C Parameters,
A = 82
τ = 0
σ_h = Variable
B = 9.3
0 .2 .4 .6 .8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 σ_h Meters
0 .8 1.6 2.4 3.2 4.0 4.8 5.6 6.4 7.2 8.0 8.8 9.6 10.4 $H_{1/3}$ Meters

Figure 2

DISTRIBUTION FUNCTION OF GATE 10

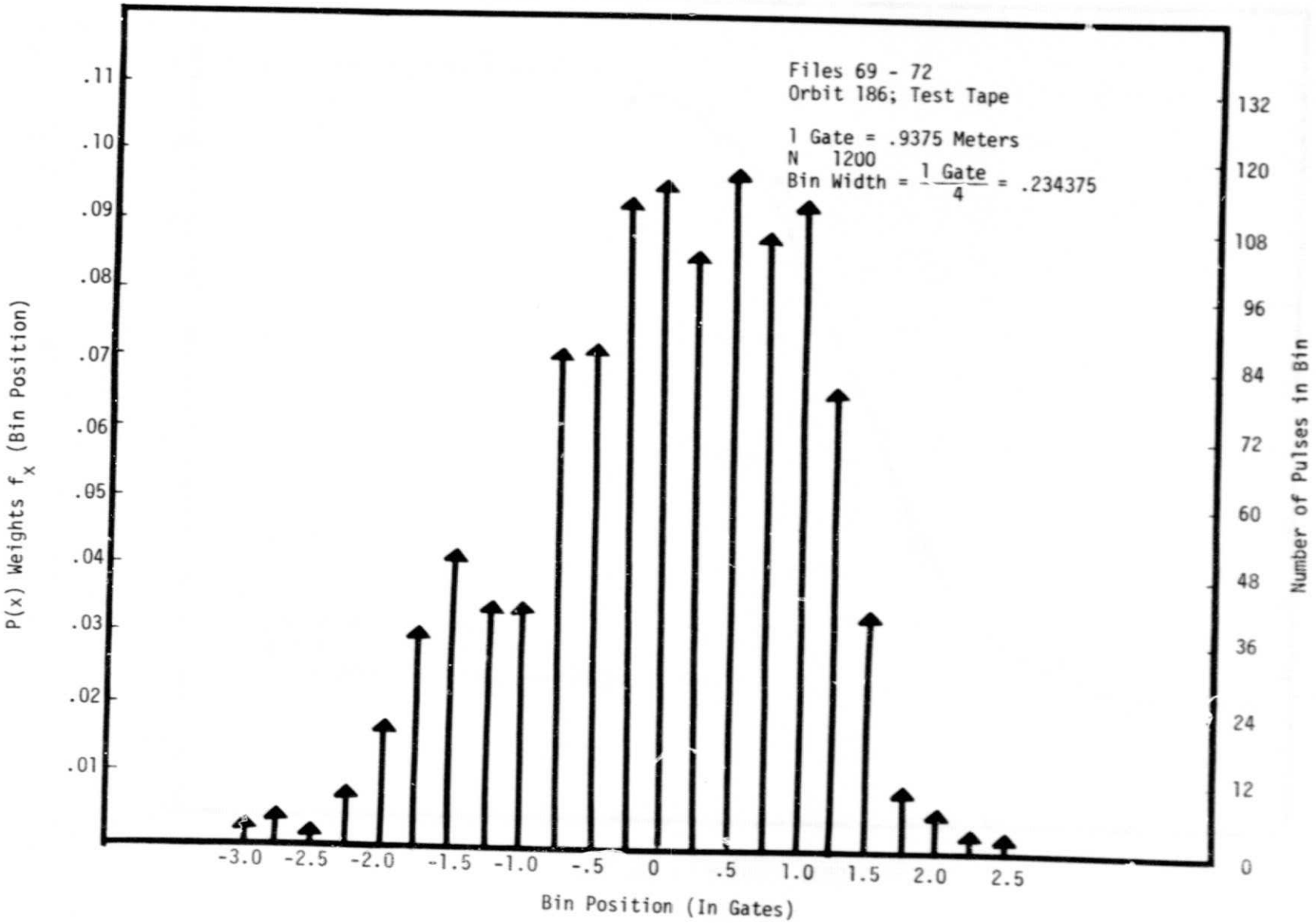
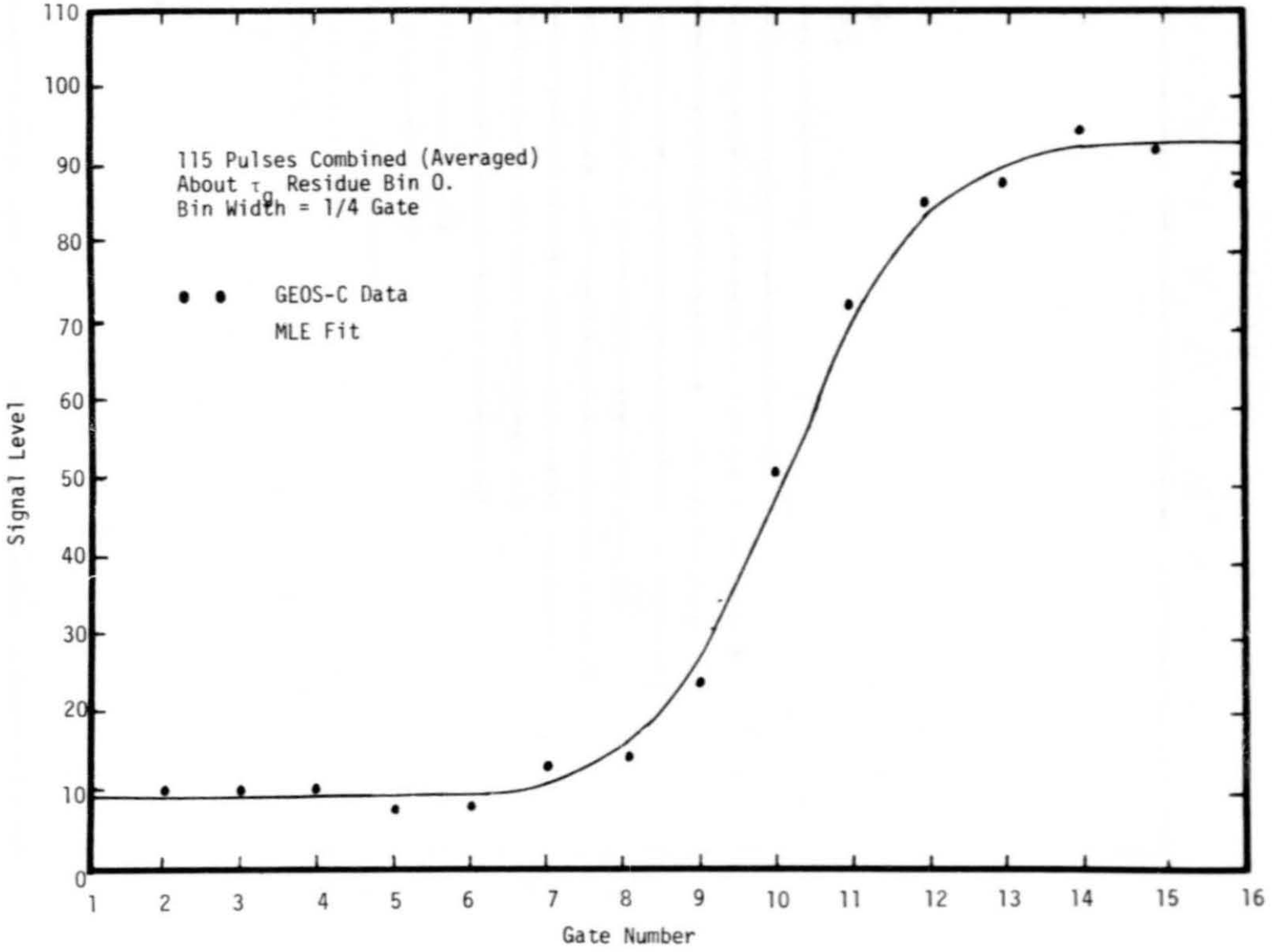


Figure 3

MLE FIT TO AVERAGE GEOS-C DATA



3.0 DEVELOPMENT OF THE MLE AND MMSE ALGORITHMS

Consider the altimeter receiver which can be modeled as (nearly) matched filtering, square law detection and sampling as shown in Figure 4. The description of maximum likelihood estimation* and its properties are simplified with this model, since the problem is now constrained to that of finding the optimum estimates for the sampled video outputs. All that is required is a parametric expression for the mean video power return, $\overline{V}$. In general, $\overline{V}$, is a function of a large number of parameters. These include not only the usual radar parameters required to establish receiver signal-to-noise ratio, but also antenna beamwidth, off-nadir pointing error, range resolution, and the height and slope distributions of the ocean surface. The trick is to reduce the number of parameters required to specify $\overline{V}$ to a minimum, since a joint maximum likelihood estimate of all of these parameters will be necessary. It can be shown that an approximate closed form solution for $\overline{V}$ can be developed that includes all of the above parameters. That is, the average video power return versus range y from the altimeter can be expressed in the form

$$\overline{V}(y|y_o,a,\sigma_e,\mu) \tag{1}$$

where

y_o - is the altitude (range) of the altimeter above the mean sea surface;

a - is the signal-to-noise ratio;

σ_e - is the effective RMS waveheight;

and

μ - is the exponential decay factor.

---

* Later it will be shown that the MMSE estimation algorithm is the same as the MLE except that the return difference is weighted by unity instead of its variance. Thus the following discussion is appropriate to both the MLE and MMSE algorithms.

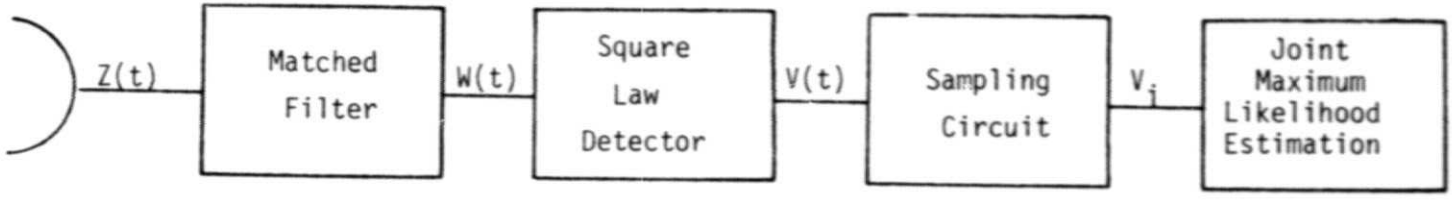


Figure 4. Functional Model of the Altimeter Receiver

Explicit expressions for a, σ_e and μ in terms of physical parameters are given in Appendix A.

3.1 Estimation Equations

The procedure to be followed in developing the maximum likelihood estimation equations will be to find a suitable approximation to the likelihood function (i.e., the joint probability density) of the video outputs in terms of the four parameters y_o, a, σ_e and μ. Then for any set of observed video samples, the joint maximum likelihood estimates are the values (y_o, a, σ_e, μ) which maximize the logarithm of the likelihood function. The maximum, of course, occurs at a point in the parameter space for which the four partial derivatives with respect to the parameters are zero.

The form of the likelihood function for the sampled video outputs of Figure 1 is simple, if it is assumed that the samples are separated by at least one range resolution cell. In this case, the correlation between samples is negligible, and since the underlying process is Gaussian, it is reasonable to assume them to be independent. Thus, the likelihood function has the form:

$$\Lambda(V_{11},V_{12},\ldots V_{nk}/y_o,a,\sigma_e,\mu) = \prod_{i,k} \overline{V}_i^{-1} \exp(-V_{ik}/\overline{V}_i) \qquad (2)$$

where:

V_{ik} is the sampled video output from the $i^{\underline{th}}$ range cell on the $k^{\underline{th}}$ pulse.

and

$\overline{V}_i$ is the expected value of the video from the $i^{\underline{th}}$ range cell.

That is, the square law detected video outputs form an independent exponential process. The dependence of the likelihood function on the

four parameters y_o, a, σ_e, μ is contained entirely in the variation with range of the mean value $\bar{V}_i$. Thus taking the logarithm of (2), and differentiating, the maximum likelihood estimates must satisfy the four equations

$$0 = \frac{\partial \log \Lambda}{\partial \alpha} = - \sum_k \sum_i (\bar{V}_i - V_{ik}) \bar{V}_i^{-2} \frac{\partial \bar{V}_i}{\partial \alpha} \qquad (3)$$

where α stands for any of the four parameters y_o, a, σ_e, μ.

3.2 Implementation

The problem of determining the joint maximum likelihood estimates of the four parameters describing the altimeter video power return requires solving Equations (3). One technique for achieving this is to use the negative of the partial derivatives in Equations (3) as inputs to integrating - feedback filters to derive the estimates as shown in Figure 5.

That is, since the estimates are derived from integrating filters, one has:

$$\frac{d\hat{y}_o}{dt} = -k \frac{\partial \log \Lambda}{\partial y_o}$$

$$\frac{d\hat{a}}{dt} = -k \frac{\partial \log \Lambda}{\partial a} \qquad (4)$$

$$\frac{d\hat{\sigma}_e}{dt} = -k \frac{\partial \log \Lambda}{\partial \hat{\sigma}_e}$$

$$\frac{d\hat{\mu}}{dt} = -k \frac{\partial \log \Lambda}{\partial \hat{\mu}}$$

where k is a gain factor associated with the integrators. Now, the time rate of change of the logarithm of the likelihood function is:

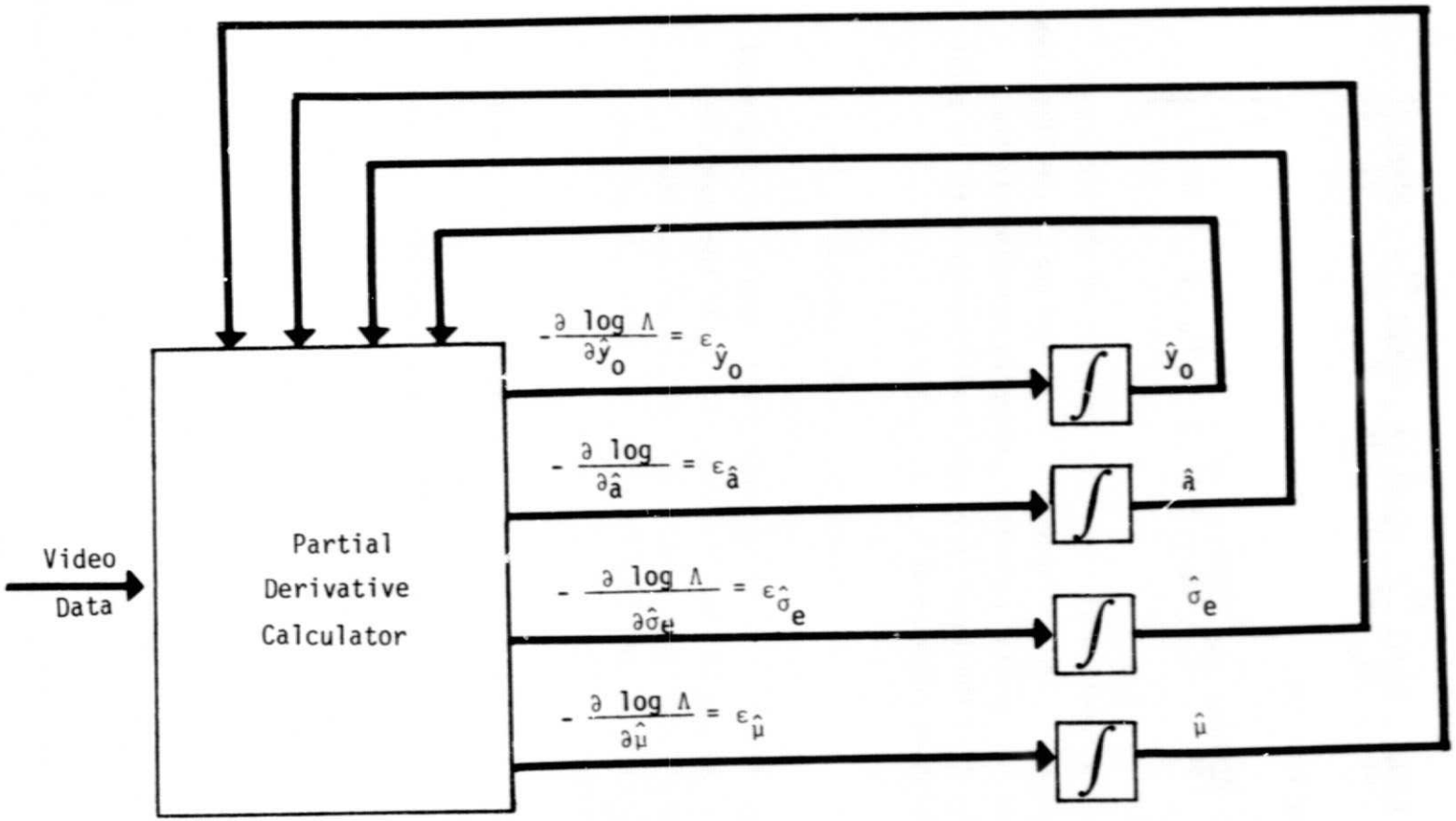


Figure 5. General MLP Implementation

$$\frac{d\log\Lambda}{dt} = \frac{\partial\log\Lambda}{\partial\hat{y}_o}\frac{d\hat{y}_o}{dt} + \frac{\partial\log\Lambda}{\partial\hat{a}}\frac{d\hat{a}}{dt} + \frac{\partial\log\Lambda}{\partial\sigma_e}\frac{d\sigma_e}{dt} + \frac{\partial\log\Lambda}{\partial\hat{\mu}}\frac{d\hat{\mu}}{dt}$$

of from Equations (9),

$$\frac{d\log\Lambda}{dt} = -K\left\{\left(\frac{\partial\log\Lambda}{\partial\hat{y}_o}\right)^2 + \left(\frac{\partial\log\Lambda}{\partial\hat{a}}\right)^2 + \left(\frac{\partial\log\Lambda}{\partial\sigma_e}\right)^2 + \left(\frac{\partial\log\Lambda}{\partial\hat{\mu}}\right)^2\right\} \leq 0. \tag{5}$$

Thus, as a function of time, the logarithm of the likelihood ratio must decrease and converge to a local minimum of the likelihood function. If the initial parameter estimates are close to the correct ones, then the loops in Figure 5 will converge to the joint maximum likelihood estimate of y_o, a, σ_e and μ.

The practicality of implementing the MLP depends on the complexity of the partial derivative calculator in Figure 5. The steps required to compute the four error signals are diagrammed in Figure 6. As can be seen from that figure, if the mean video versus range plus the four partial derivatives are available, then the computations required to derive the error signals are relatively minor.

Figures 5 and 6 give a fairly pleasing intuitive picture of the operation of the MLP. In Figure 6, it is seen that the error signals for the tracking loops in Figure 5 are derived as follows: First, an error signal versus range, $\hat{\bar{V}} - V$, is generated which represents the difference between the measured return as a function of range, $\hat{\bar{V}}$, and the estimated mean return, V. Secondly, this error signal is weighted inversely proportional to its variance*. That is:

$$E(V - \bar{V})^2 = \bar{V}^2 \tag{6}$$

Thus, after normalization, the signal may be loosely described as having uniform information content. Finally, the normalized error-versus-range signal is "gated" by the partial derivative of $\hat{\bar{V}}$, with respect to

*For MMSE estimation, the error signal weighting is unity instead of its variance.

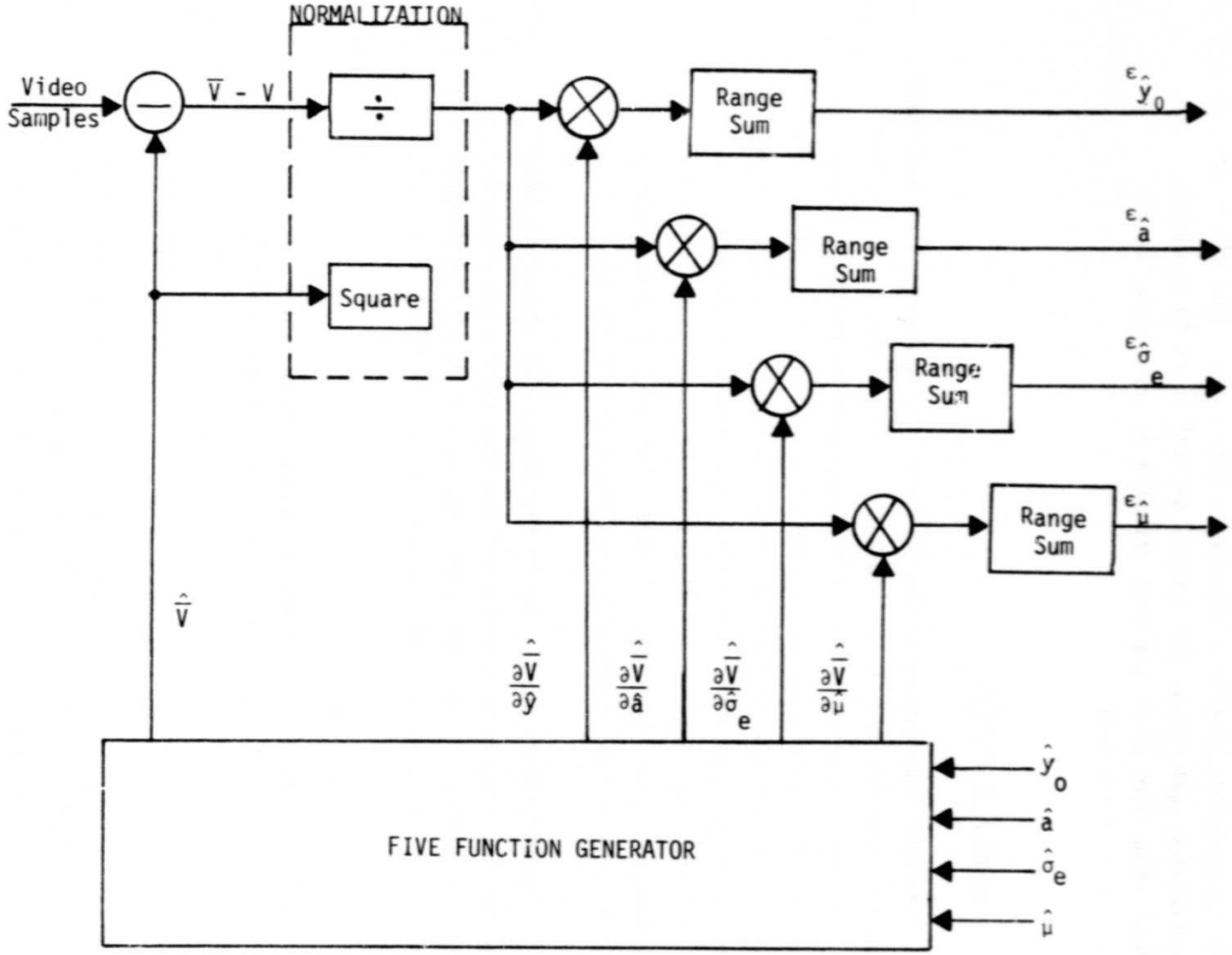


Figure 6. Generation of the Error Signals for the MLP

the parameter of interest, and summed over range. The effect of multiplying (gating) by the partial derivative is to emphasize those range bins which are most affected by variations in the parameter of interest.

3.3 Decoupling Matrix

A major shortcoming of the feedback loop implementation described above is that the loop error signals are very highly cross-coupled. As a result, an error in one of the parameter estimates can (and usually does) produce errors in the other estimates. Furthermore, experience with the implementation has shown the convergence time of the loops to be excessively long as a result of the error cross-coupling. These difficulties may be essentially eliminated by decoupling the loop error signals. Fortunately, a decoupling matrix can be implemented with a fairly simple technique.

The underlying assumption in the decoupling matrix approach is that the loop error signals and the true errors are linearly related. However, since the loops are very highly cross-coupled, the linear relation must be described by a full matrix transform:

$$\underline{\varepsilon} = K\underline{\Delta\alpha} \tag{7}$$

where $\underline{\varepsilon}$ - is a 4 x 1 vector of error signals which are generated by the true errors $\underline{\Delta\alpha}$ ($\underline{\Delta\alpha}$ is a 4 x 1 vector).

and K - is a 4 x 4 matrix describing the linear relation.

The basic problem in determining the matrix K is that it depends on the (4 x 1 vector) of true parameter values, $\underline{\hat{\alpha}} = [a, y_o, \sigma_e, \mu]^T$ which are, of course, unknown. An obvious approach to solving this problem would be to evaluate the matrix K at the estimated parameter values, $\underline{\hat{\alpha}}$ and use the estimated matrix to decouple the loops. The usefulness of this approach will, of course, depend on the region over which a linear relation, Equation (7), holds. If the region is very small,

then $\hat{\underline{\alpha}}$ would have to be very close to the true value $\underline{\alpha}$. Here again, experience with the implementation has shown that over an acceptably large region about the true values, a linear relation exists between an input error in one parameter and the corresponding errors produced in the other three parameters.

Next, consider the problem of estimating the matrix K. Suppose there is an error in a single parameter, say $\Delta\mu$. If the error signals are linearly related to the true error, then:

$$\begin{bmatrix} \epsilon_a^{\Delta\mu} \\ \epsilon_{y_o}^{\Delta\mu} \\ \epsilon_{\sigma_e}^{\Delta\mu} \\ \epsilon_u^{\Delta\mu} \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta_\mu \end{bmatrix} \tag{8}$$

$$= \begin{bmatrix} k_{14} \\ k_{24} \\ k_{34} \\ k_{44} \end{bmatrix} \Delta\mu$$

Where the notation $\epsilon_a^{\Delta\mu}$ is used to represent the error signal produced in the "a" loop by an error of $\Delta\mu$ in the μ parameter. From the above equation, it is seen that the fourth column of the K matrix is given by

$$\begin{bmatrix} k_{14} \\ k_{24} \\ k_{34} \\ k_{44} \end{bmatrix} = (\Delta\mu)^{-1} \begin{bmatrix} \epsilon_a^{\Delta\mu} \\ \epsilon_{y_o}^{\Delta\mu} \\ \epsilon_{\sigma_e}^{\Delta\mu} \\ \epsilon_\mu^{\Delta\mu} \end{bmatrix} \tag{9}$$

Hence the K matrix can be computed according to the relation

$$K = \begin{bmatrix} (1/\Delta a) \begin{bmatrix} \epsilon_a^{\Delta a} \\ \epsilon_{y_o}^{\Delta a} \\ \epsilon_{\sigma_e}^{\Delta a} \\ \epsilon_{\mu}^{\Delta a} \end{bmatrix} \cdots (1/\Delta\mu) \begin{bmatrix} \epsilon_a^{\Delta\mu} \\ \cdot \\ \cdot \\ \epsilon_{\mu}^{\Delta\mu} \end{bmatrix} \end{bmatrix} \qquad (10)$$

The procedure for decoupling the loops is as follows.

1) After the initial large transients have died down ($\sim$ 4 iterations) compute a simulated return based on the current parameter estimates.

2) Using the simulated return as the correct return, perturb the parameter estimates one at a time and compute the corresponding error signals.

3) From these four sets of error signals, build up the K matrix according to Equation (10).

4) Invert K, and there after decouple the error signals according to the formula: $\underline{\Delta\hat{\alpha}} = K^{-1}\ \underline{\epsilon}(\hat{\alpha})$

In practice, the true parameter values will be changing slowly, so that this procedure must be repeated at periodic intervals. Figure 7 shows a functional block diagram of the resulting feedback loop implementation of the MLP when the error decoupling matrix is included.

3.4 Theoretical Accuracies

Finally, it is useful to derive the general expressions for the theoretical accuracies (asymptotic variance) of the joint maximum likelihood estimates. These accuracies are given in terms of

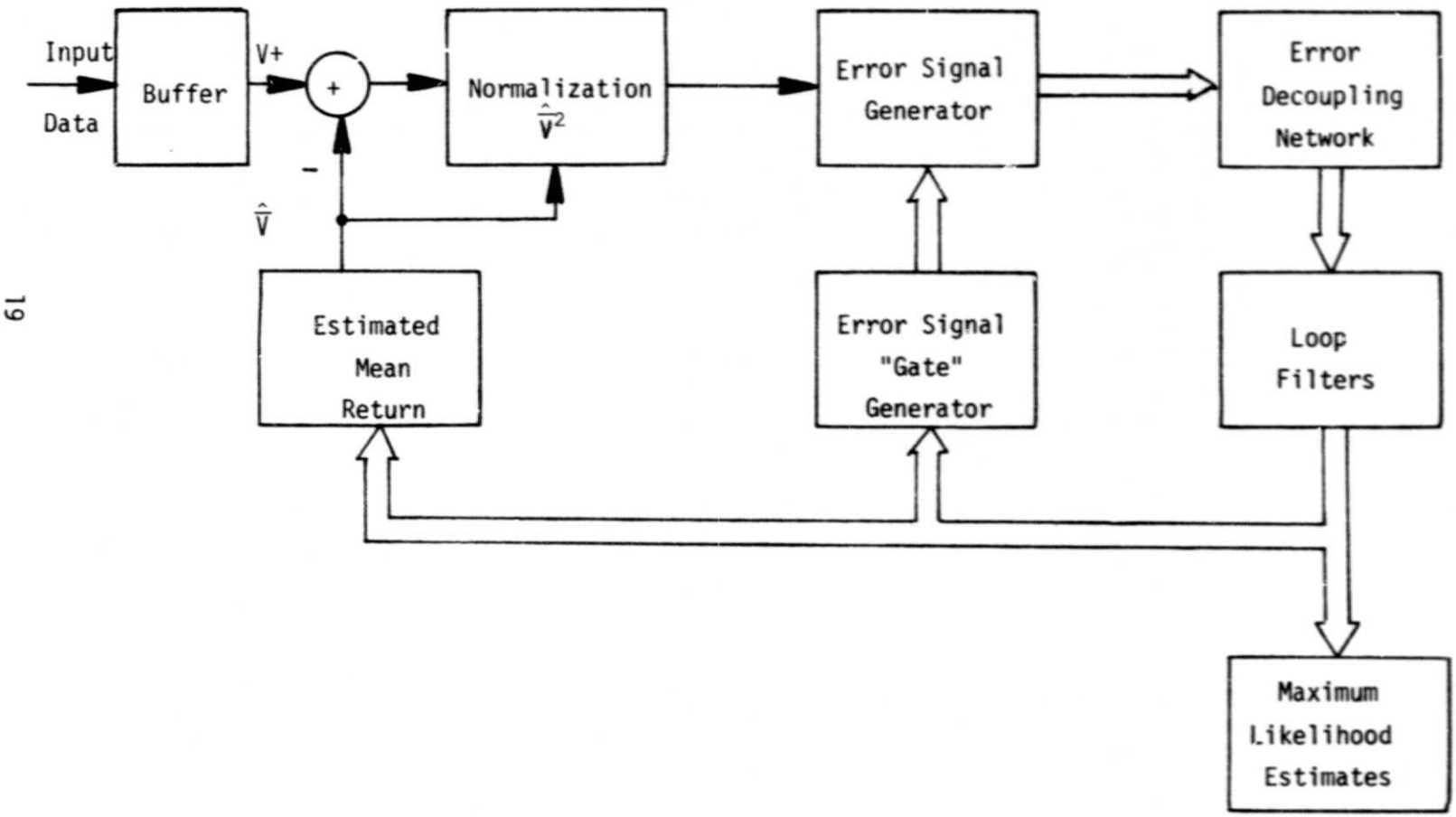


Figure 7. Functional Block Diagram of the MLP

the standard deviation of the estimates due to random fluctuations only. System errors (such as clock and timing errors) as well as errors associated with spatial averaging are not included in the evaluation.

From the theory of maximum likelihood estimation (see for example Wilks [1], Chapter 12), it is known that for regular distributions, the MLE is, asymptotically, unbiased and has minimum variance. The covariance matrix of the estimation errors is given by

$$C = B^{-1} \tag{11}$$

where C is the covariance matrix defined by:

$$C_{\ell m} = \text{Covar}\,(\alpha_\ell \alpha_m)$$

with α_ℓ, α_m standing for any of the parameter estimates and B is the information matrix defined by:

$$B_{\ell m} = E\left\{\frac{\partial \log \Lambda}{\partial \alpha_\ell} \cdot \frac{\partial \log \Lambda}{\partial \alpha_m}\right\} \tag{12}$$

where, in (12) Λ stands for the likelihood ratio evaluated at the true parameter values.

Substituting (3) into (12) and recalling that the video samples V_{ik} are independent, yields

$$B_{\ell m} = \sum_k \sum_i E(\bar{V}_i - V_{ik})^2 \, \bar{V}_i^{-4} \, \frac{\partial \bar{V}_i}{\partial \alpha_\ell} \, \frac{\partial \bar{V}_i}{\partial \alpha_m} \tag{13}$$

But for V_{ik} exponentially distributed,

$$E(\bar{V}_i - V_{ik})^2 = \bar{V}_i^2 \tag{14}$$

and the information matrix becomes:

$$B_{\ell m} = N \sum_i \bar{V}_i^{-2} \frac{\partial \bar{V}_i}{\partial \alpha_\ell} \frac{\partial \bar{V}_i}{\partial \alpha_m} \tag{15}$$

where N is the total number of pulses processed. Thus, the estimation accuracies can be determined from the mean return model given in Appendix A. Although analytical expressions can be developed for the partial derivatives and even for the information matrix for a particular set of parameters, the resulting equations give little insight to the behavior of the errors, nor do they significantly simplify the calculation of B. Therefore, both the partial derivatives and the computation of the covariance matrix are evaluated numerically. Examples for the GEOS altimeter parameter estimates are given in Appendix B.

Reference

[1] S. S. Wolks, "Mathematical Statistics," John Wiley & Sons, New York, 1963.

APPENDIX A
MEAN POWER RETURN

In this appendix, the radar equation for a distributed target is derived and applied to the particular case of a satellite altimeter. The target is modeled as a continuous distribution of independent scattering points distributed in the three dimensions (τ,ν,u). Here, τ is range delay, ν is the Doppler shift, and u is the sine of the angle to the scattering point measured from the antenna boresight. Since the scattering points are assumed to be incoherent, the total power is simply the integral over delay, Doppler shift, and angle of the differential power received from each point. The altimeter equation has been derived in various forms elsewhere (e.g., Barrick [1,2], Harger [3] or Brown [4], This derivation is included for completeness, and to define the notation used to obtain an approximate, closed-form solution for the shape of the mean power return.

Theoretical Development

From the theory of high resolution radar (e.g., Rihaczek [5] or Deley [6], the power received from a mismatched filter can be written in terms of the cross-ambiguity function. That is, let the received narrowband signal be represented in complex form as:

$$s_r(t) = \sqrt{2E_r}\, g(t-\tau)\, \exp[j2\pi(f_o-\nu)(t-\tau)] \qquad (A.1)$$

where τ is the round-trip delay to the scatter,

ν is the Doppler shift produced by the scatterer,

f_o is the carrier frequency,

$g(t)$ is the complex modulation impressed on the carrier,

and E_r is the received signal energy.

In the above expression it is assumed that the complex modulation has

been normalized to unit energy.

If the receiver is matched to a complex modulation $w(\xi)$, a Doppler frequency ϕ, and a delay, t, then the video power out of the filter at time t is given by:

$$\bar{V}(t) = E_r \, G_o \, B \left| X_{gw}(\tau - t, \nu - \phi) \right|^2 \qquad (A.2)$$

where $\left| X_{gw} \right|^2$ is the normalized cross-ambiguity function defined by:

$$X_{gw}(\tau, \nu) = \int_{-\infty}^{\infty} g(\xi) \; w^*(\xi + \tau) \exp[-j2\pi\nu\xi] \, d\xi \qquad (A.3)$$

and

$$G_o B = \int \left| W(f) \right|^2 df = \int \left| w(t) \right|^2 dt$$

is the product of receiver gain, times its noise bandwidth. Note that X_{gw} has been normalized so that

$$\left| X_{gw} \right|^2 \leq 1$$

If the scattering point has differential cross-section, $d\sigma$, and is located at angular position u (in sine theta space) then the signal energy can be computed from the radar equation (Skolnik [7]). That is:

$$E_r = \frac{E_T G A(u)}{4\pi R^2} \quad \frac{d\sigma}{4\pi R^2} \quad \frac{\lambda^2 G A(u)}{4\pi} \quad L \qquad (A.4)$$

where E_T is the transmitted energy,

G is the antenna gain on boresight,

A(u) is the one way loss factor when the target is not on boresight,

λ is the R. F. wave length

$d\sigma$ is the differential cross-section at the scattering point, and is generally a function of all three parameters (τ, ν, u),

R $c\tau/2$ is the range to the scatterer,

and L is the total system losses (except for processing mismatch, which is included in the cross-ambiguity function).

Combining Equations (A.2) and (A.4), gives the differential power received from a scattering point located at (τ, ν, u):

$$d\bar{V}(t) = \frac{E_T G^2 \lambda^2 L G_o B}{(4\pi)^3} \frac{A^2(u)}{R^4} \left| \chi_{gw}(\tau - t, \nu - \phi) \right|^2 d\sigma\ (\tau, \nu, u) \qquad (A.5)$$

Integrating this equation over the three variables, and adding the receiver noise, yields the total average video power for a distributed target:

$$\bar{V}(t) = \frac{E_T G^2 \lambda^2 L G_o B}{(4\pi)^3} \int_{u=0}^{1} \int_{\nu=-\infty}^{\infty} \int_{\tau=-\infty}^{\infty} \frac{A^2(u)}{R^4} \left| \chi_{gw}(\tau - t, \nu - \phi) \right|^2 d\sigma\ (\tau, \nu, u) \qquad (A.6)$$

$$+ \quad kT_o F_N G_o B$$

where k is Bolzman's Constant,

T_o is the reference temperature,

and F_n is the receiver noise figure.

Equation (A.6) is a fairly general representation of the power received from a distributed target. To apply it to the satellite altimeter, one must now compute an expression for the differential cross-section $d\sigma(\tau, \nu, u)$ which is appropriate to the altimeter. The geometry being considered is shown in Figure A.1.

Consider a scattering point located at height, h, above the

But the second term is a semidefinite quadratic form so that,

$$D^T B^T (BB^T)^{-1} BD \geq 0 \tag{B.19}$$

and

$$Q(\theta_{n+1}) \leq Q(\theta_n) \tag{B.20}$$

Thus, the iterative solution (eq. (B.16)) converges to a local minimum for small enough k. Note that although a rather crude approximation was made to obtain eq. (B.16), the above discussion shows that the only criterion (for local convergence) is that k be chosen small enough. Experience with the altimetry problem indicates k on the order of .5 to .8 is often times sufficiently small.

APPROXIMATE VARIANCE OF THE ESTIMATE

An approximate estimate of the covariance of the parameter estimates, $\hat{\theta}$, is obtained by expanding eq. (B.5) about the true parameter value θ. Then in analogy to equation (B.13), one has

$$\hat{\theta} \approx \theta - [\nabla \epsilon^T(\theta)]^{-1} \epsilon(\theta) \tag{B.21}$$

Where the above equation holds asymptotically for large numbers of samples. Again, assuming many samples, $\nabla \epsilon^T(\theta)$ is approximated by its mean value:

$$\begin{aligned} \nabla \epsilon^T(\theta) &\approx E \nabla \epsilon^T(\theta) \\ &= BB^T \end{aligned} \tag{B.22}$$

where B is evaluated at the true parameter value, θ. Note, to obtain eq. (B.22), it has been assumed that

$$EV_j = \bar{V}_j(\theta) \tag{B.23}$$

Hence

$$\hat{\theta} \approx \theta - (BB^T)^{-1} BD \tag{B.24}$$

Again, assuming $EV_j = \bar{V}_j$, $\hat{\theta}$ is (asymptotically) unbiased, and

$$E\,\hat{\theta} = \theta \tag{B.25}$$

The covariance matrix of the parameter estimates is:

$$\Sigma_\theta = E(\hat{\theta}-\theta)\,(\hat{\theta}-\theta)^T$$

$$= (BB^T)^{-1} B\,\Sigma_D\,B^T(BB^T)^{-1} \tag{B.26}$$

For independent samples, the covariance of D is given by:

$$\Sigma_D = \text{Diag}\,(\sigma_{V_j}{}^2/\nu_j{}^2) \tag{B.27}$$

where "Diag" denotes a diagonal matrix.

In the special case, of an MLE estimate for an exponential distribution,

$$\sigma_{V_j} = \bar{V}_j = \nu_j \tag{B.28}$$

and equation (B.26) becomes:

$$\Sigma_\theta = (BB^T)^{-1} \tag{B.29}$$

SUMMARY

The important equations in the above discussion are:

Iterative Solution

$$\theta_{n+1} = \theta_n - kWD \tag{B.16}$$

$$W = (BB^T)^{-1}\,B \tag{B.30}$$

$$B = (\nu^{-1}\nabla\bar{V}_1 , \cdots \nu_K^{-1}\nabla\bar{V}_K) \tag{B.9}$$

$$D = ((\bar{V}_1 - V_1)\nu_1^{-1} \cdots (\bar{V}_K - V_K)\nu_K^{-1})^T \tag{B.10}$$

where for an exponential distribution maximum likelihood estimate:

$$\nu_j = \bar{V}_j \tag{B.7}$$

but for a MMSE,

$$\nu_j = U_j \tag{B.8}$$

where U_j^{-2} is the weight applied to each sample residue squared. In the above expressions, B and D are evaluated at the current estimate, θ_n.

Approximate Error Analysis

$$\Sigma_{\hat{\theta}} = (BB^T)^{-1} B\Sigma_D B^T(BB^T)^{-1} \tag{B.26}$$

with

$$\Sigma_D = \text{Diag}\ (\sigma_{V_j}^2/\nu_j^2) \tag{B.27}$$

Special Case

For an exponential distribution,

$$\sigma_{V_j} = \bar{V}_j \tag{B.28}$$

and finally for an exponential distribution MLE, $\sigma_{V_j} = \nu_j = \bar{V}_j$ and

$$\Sigma_{\hat{\theta}} = (BB^T)^{-1} \tag{B.29}$$

mean sea surface. Then, if $\sigma^\circ(\theta)$ is the mean cross-section per unit surface area (at angle θ relative to radar) and if $p_h(h,\sigma_n)$ is the probability density function of the distribution of scatterers with height, (scaled to RMS wave height θ_h) then the differential cross-section at that point is:

$$d\sigma = \sigma^\circ(\theta)\, p_h(h, \sigma_h)\, dV \tag{A.7}$$

where dV is the differential volume at the point of interest. Now, in the spherical coordinate system (R,θ,ψ) located at the satellite, the differential element of volume is:

$$dV = R^2 \sin\theta\, dR\, d\theta\, d\psi \tag{A.8}$$

The final step, then is to compute τ, ν, u and h in terms of R, θ and ψ, substitute the results into Equation (A.6) and then substitute Equation (A.8) into Equation (A.6) to obtain the final expression. But from Figure A.1, it is seen that

$$\tau = \frac{2R}{c}$$

$$\nu = \frac{2V}{\lambda} \sin\theta \cos\psi$$

$$u = \sin\theta \tag{A.9}$$

and

$$h = \left((H_o - R\cos\theta + a_e)^2 + R^2 \sin^2\theta \right)^{1/2} - a_e \tag{A.10}$$

where V is the satellite velocity,

H_o is the satellite altitude,

and a_e is the radius of the earth.

Before making the indicated substitutions, the equations will be simplified by making some approximations. That is, it is assumed that the beamwidth of the altimeter is narrow (less than 5°) so that small angle, and related consistent approximations can be made.

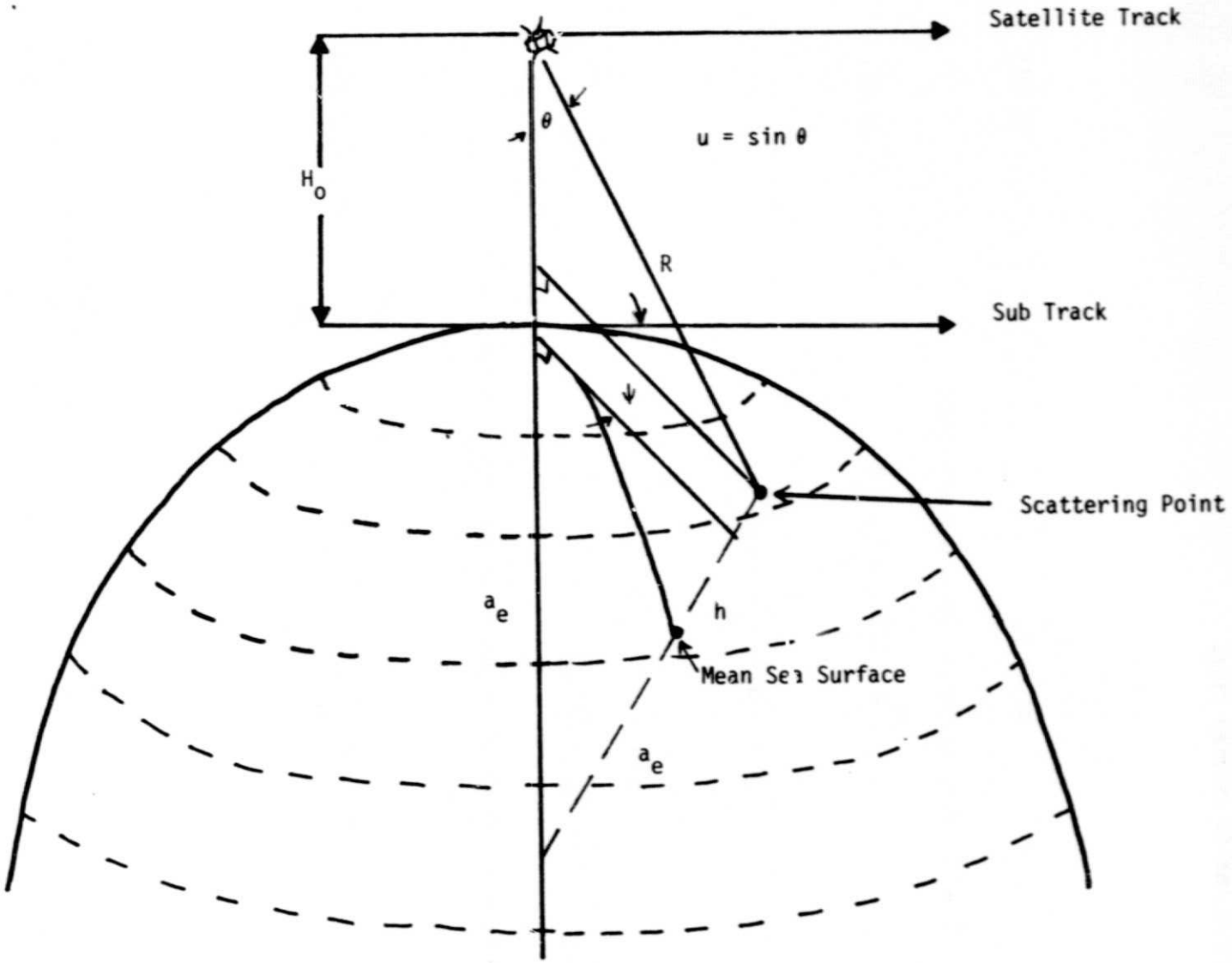


Figure A.1 Geometry for Computing Mean Power Return

Thus, it is assumed that:

$$u = \sin\theta << 1,$$
$$R = H_o + \Delta R,$$
$$\Delta R << H_o$$
$$\Delta R << a_e \qquad (A.11)$$

With these approximations, one can write

$$\tau = \frac{2H_o}{c} + \frac{2\Delta R}{c} = \tau_o + \frac{2\Delta R}{c} \qquad (A.12)$$

$$h \approx -\Delta R + \frac{H_o\theta^2}{2}\left(1+\frac{H_o}{a_e}\right)$$

$$d\sigma \approx \sigma^o(\theta)\, p_h\,(h, \sigma_h)\, H_o^{\,2}\, \theta\, d\Delta R\, d\theta\, d\psi$$

and $$\nu \approx \frac{2V\theta}{\lambda}\cos\psi$$

Thus, the average video power can be written finally as:

$$V(t) = \frac{E_T G^2 \lambda^2 L \sigma^o G_o B}{(4\Pi)^3 H_o^{\,3}(1+H_o/a_e)} \int_{-\infty}^{\infty} d\Delta R \int_o^{\Pi/2} [H_o(1+H_o/a_e)\theta\, d\theta]$$

$$\gamma(\theta) A^2(\theta) p_h\left(-\Delta R + \frac{H_o\theta^2}{2}(1+H_o/a_e), \sigma_h\right)$$

$$\int_o^{2\Pi} d\psi \left| \chi_{gw}\left((\tau_o + \frac{2\Delta R}{c} - t), \left(\frac{2V\theta}{\lambda}\cos\psi - \phi\right)\right)\right|^2$$

$$+ kT_o F_n G_o B \qquad (A.13)$$

where $\gamma(\theta)$ is defined by $\sigma^o(\theta) = \sigma^o\gamma(\theta)$ and defines the variation of $\sigma^o(\theta)$ with angle.

Approximate Evaluation

A closed form solution for the mean power return can now be obtained from the triple integral of Equation (A.13). To achieve such a solution requires some additional approximations and simplifying assumptions. First assume that the cross ambiguity function is essentially constant over the doppler variation of the scattering points, then

$$\int_0^{2\pi} d\psi \left| \chi_{gw}\left((\tau_o + \frac{2\Delta R}{c} - t), \frac{2V\theta}{\lambda} \cos\psi - \phi \right) \right|^2 \tag{A.14}$$

$$\approx 2\pi \left| \chi_{gw} \left(\tau_o + \frac{2\Delta R}{c} - t, 0 \right) \right|^2$$

Further assume that the variation of the cross ambiguity function with range delay can be approximated by a Gaussian function, thus

$$\left| \chi^2_{gw}(t_1) \right| \simeq \exp\left[-\frac{1}{2} \left(\frac{t_1}{\sigma_\tau} \right)^2 \right] = \sqrt{2\pi}\ \sigma_\tau\ \eta(t_1, \sigma_\tau) \tag{A.15}$$

where $\eta(.,\sigma_\tau)$ is the normal density function and σ_τ = RMS pulse width.

Now from Barrick's [1] model, the wave height density is also normal, hence

$$p_h\ (h, \sigma_h) = \eta(h, \sigma_h) \tag{A.16}$$

Letting

$$H = H_o\ \left(1 + \frac{H_o}{a_e}\right) \tag{A.17}$$

and

$$x = \frac{H\theta^2}{2} \tag{A.18}$$

$$dx = H\theta d\theta$$

Substituting Equations (A.14) through (A.18) into the triple integral of Equation (A.13), denoted I_3, yields,

$$I_3 = (2\pi)^{3/2}\,\sigma_R \int_0^{\frac{H\pi^2}{8}} dx\,\gamma(x)A^2(x) \int_{-\infty}^{+\infty} d\Delta R \qquad (A.19)$$

$$\eta(x - \Delta R,\ \sigma_h)\ \eta\left(\Delta R - (t - \tau_o)c/2,\ \sigma_R\right)$$

where $\sigma_R = \dfrac{c\sigma_\tau}{2}$ = RMS range resolution.

Thus the integral over ΔR is simply the convolution of two normal densities having different means and variances. As may be shown, this convolution produces a normal density with a mean equal to the sum of the means, and a variance equal to the sum of the variances. Letting

$$y = \tfrac{1}{2}ct = \text{range at time } t,$$

and $$y_o = \tfrac{1}{2}c\tau_o = \text{range to the mean sea surface} \qquad (A.20)$$

and performing the indicated convolution of Equation A.19 gives

$$I_3 = (2\pi)^{3/2}\,\sigma_R \int_0^{\frac{H\pi^2}{8}} \gamma(x)A^2(x)\,\eta(y - y_o - x,\ \sigma_e) \qquad (A.21)$$

where $$\sigma_e^2 = \sigma_R^2 + \sigma_h^2 \qquad (A.22)$$

Now assuming that the two-way antenna pattern function can be approximated by a Gaussian function

$$A^2(\theta) \simeq \exp[-(4 \ln 2)(\theta/\theta_{BW2})^2] \qquad (A.23)$$

where θ_{BW2} = the two-way 3 dB beamwidth,

and from Barrick's [1] model, the surface shaping function is

$$\gamma(\theta) \simeq \exp[-(4 \ln 2)(\theta/\theta_{BW2})^2] \tag{A.24}$$

where θ_s = the 3 dB spread of the slope distribution.

Then, utilizing the change of variable given by Equation A.18, the product of the surface shaping function and beam pattern can be written as

$$\gamma(x)A^2(x) = \exp(-\mu x) \tag{A.25}$$

where

$$\theta_e = \left[\frac{1}{\theta_{BW2}^2} + \frac{1}{\theta_s^2}\right]^{-1/2} \tag{A.26}$$

is the effective 3 dB beamwidth, and

$$\mu = (8 \ln 2)/(H\theta_e^2) \tag{A.27}$$

is the exponential decay factor parameter.

The remaining integral of Equation A.21 can now be expressed as

$$F(z) = \int_0^{\frac{H\pi^2}{8}} \gamma(x)A^2(x)\,\eta(z - x,\ \sigma_e)\,dx \tag{A.28}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_e} \int_0^{\infty} \exp\left[-\tfrac{1}{2}\left(\frac{z - x}{\sigma_e}\right)^2\right] \exp[-\mu x]\,dx$$

where it has been assumed that $H\pi^2/8 \gg \sigma_e$. The remaining integral can be evaluated by completing the square in the exponent. This yields:

$$F(z) = \exp[\tfrac{1}{2}(\sigma_e\mu)^2 - z\mu] \frac{1}{\sqrt{2\pi}\sigma_e} \int_0^\infty \exp\left[-\tfrac{1}{2}\left(\frac{x-(z-\sigma_e{}^2\mu)}{\sigma_e}\right)^2\right] dx \quad (A.29)$$

So that finally

$$F(z) = \exp[\tfrac{1}{2}(\sigma_e\mu)^2 - z\mu] \, \Phi\left(\frac{z}{\sigma_e} - \sigma_e\mu\right) \quad (A.30)$$

where Φ is the cumulative normal distribution function. Thus, $F(z)$ describes the shape of the mean power return as a function of range measured from the mean sea surface.

The final expression for the mean return may be obtained by substituting (A.30) into (A.21) and the result into (A.13). Before doing this, note that the peak signal-to-noise ratio at the output of the receiver is given by:

$$a = \frac{(2\pi)^{3/2} E_T G^2 \lambda^2 L \sigma_R \sigma^o}{(4\pi)^3 H_o{}^3 \left(1 + \frac{H_o}{a_e}\right) k T_o F_n} \quad (A.31)$$

Thus, the mean return may be written:

$$\bar{V}(y) = k T_o F_n G_o B \left(a F(y - y_o) + 1\right) \quad (A.32)$$

where the function $F(\cdot)$ is given by Equation (A.30). Note that if the noise power is known ($kT_oF_nG_oB$), then the mean return depends explicitly on the four parameters:

a - signal-to-noise ratio
y_o - altitude to the mean sea surface
σ_e - "effective" wave height
and μ - exponential decay factor.

The derivation of Equation (A.32) did not include the effects of antenna pointing error. These effects may be included by noting that pointing error causes the two-way beam pattern, A^2 to depend

both on the off-nadir angle, θ, and the azimuth angle ψ. In particular, for an off-nadir angle ξ we have:

$$A^2(\theta,\psi) = \exp\left[-(4 \ln 2)(\beta/\theta_{BW2})\right]^2 \qquad (A.33)$$

where β is the angle between the antenna boresight and a line to the surface scattering point located at angular position (θ,ψ). β is determined by the equation:

$$\cos\beta = \cos\theta \cos\xi + \sin\theta \sin\xi \cos\psi \qquad (A.34)$$

Making small angle approximations yields

$$\beta^2 \approx \theta^2 + \xi^2 + 2\theta\xi \cos\psi \qquad (A.35)$$

Since A^2 now depends on the azimuth angle ψ, it must be factored into the evaluation of the integral over ψ in Equation A.14. Thus, it is necessary to evaluate:

$$\int_0^{2\pi} d\psi\, A^2(\theta,\psi) \left| \chi^2\left((\tau_o + \frac{2\Delta R}{c} - t), \frac{2V\theta}{\lambda}\cos\psi - \phi\right)\right|^2 \qquad (A.36)$$

$$\approx 2\pi \left|\chi\left(\tau_o + \frac{2\Delta R}{c} - t, 0\right)\right|^2 \tilde{A}^2(\theta)$$

where

$$\tilde{A}^2(\theta) = \exp\left[-(4 \ln 2)\left(\frac{\theta^2+\xi^2}{\theta_{BW2}^2}\right)\right] \cdot \frac{1}{2\pi}\int_0^{2\pi} \exp\left(-(8 \ln 2)\frac{\theta\xi}{\theta_{BW2}^2}\cos\psi\right) d\psi$$

$$= \exp\left[-(4 \ln 2)\left(\frac{\theta^2+\xi^2}{\theta_{BW}^2}\right)\right] I_o\left(\frac{(8 \ln 2)\theta\xi}{\theta_{BW2}^2}\right) \qquad (A.37)$$

and I_o is the modified Bessel function of order zero.

Equation (A.37) may be further simplified by noting that for arguments less than 1, I_o may be approximated to about 3 percent accuracy

by:

$$I_o(z) \approx e^{\frac{1}{4}z^2} \tag{A.38}$$

Then,

$$\tilde{A}^2(\theta) \approx \exp\left[-\frac{(4\ln 2)\xi^2}{\theta_{BW2}^2}\right] \exp\left[-(4\ln 2)\left(\frac{\theta}{\tilde{\theta}_{BW2}}\right)^2\right]$$

where

$$\tilde{\theta}_{BW2} = \frac{\theta_{BW2}}{\sqrt{1-(4\ln 2)\left(\frac{\xi}{\theta_{BW2}}\right)^2}} \tag{A.39}$$

The development proceeds identically now, except that $\tilde{A}^2$ is used everywhere A^2 appeared previously. Thus, the mean return model given in Equation (A.32) includes pointing error if the signal-to-noise ratio is modified as:

$$a' = a \exp\left[-(4\ln 2)\left(\frac{\xi}{\theta_{BW2}}\right)^2\right] \tag{A.40}$$

and the modified two-way beamwidth in Equation (A.39) is used in the definition of the exponential decay factor (A.27).

References

[1] D. E. Barrick, "Determination of Mean Surface Position and Sea State from the Radar Return of a Short Pulse Satellite Altimeter," BATTELLE, Columbus Laboratories, October 4, 1971.

[2] D. E. Barrick, "Remote Sensing of Sea State by Radar," Chapter 12, "Remote Sensing of the Troposphere," V. E. Derr (Ed.) National Oceanic and Atmospheric Administration, Boulder, Colorado, August 15, 1972.

[3] R. O. Harger, "Radar Altimetry Optimization for Geodesy Over the Sea," IEEE Trans., Vol. AES-8, No. 6, pp. 728-742, November 1972.

[4] G. S. Brown, "A Closed Form Relation for the Average Return Waveform from a Near-Nadir Pointed, Short Pulse, Satellite Based Radar Altimeter," URSI Symposium, Commission II, Session 2, Radio Oceanography I, Boulder, Colorado, October 1974.

[5] A. W. Rihaczek, "Principles of High-Resolution Radar," McGraw-Hill, New York, 1969.

[6] G. W. Deley, "Waveform Design," Chapter 3, Radar Handbook, M. I. Skolnik (Ed.), McGraw-Hill, New York, 1970.

[7] M. I. Skolnik, "An Introduction to Radar," Chapter 1, Radar Handbook, M. I. Skolnik (Ed.), McGraw-Hill, New York, 1970.

APPENDIX B
MATRIX FORMULATION OF THE MLE AND MMSE ALGORITHMS

INTRODUCTION

In this Appendix, the iterative equations for obtaining the maximum likelihood estimator (MLE) for an exponential distribution and the minimum mean square error (MMSE) estimator expressions for the covariance matrix of the estimators is given.

DEVELOPMENT

As was done in earlier studies (e.g., ref. [1], it is assumed that K independent data samples of a distributed target are taken at the output of a square law detector. Denote these samples as:

$$V_j, \quad j = 1 \cdots K$$

Further, assume that a parametric form of the mean return model is known. That is,

$$\bar{V}_j(\theta), \quad j = 1 \cdots K$$

where θ is a vector of parameter values to be estimated.

Under these assumptions, the logarithm of the likelihood function is given by:

$$\Lambda(\theta) = -\sum_{j=1}^{K} [\ln \bar{V}_j + (V_j/\bar{V}_j)] \tag{B.1}$$

The MLE estimate of the parameters, $\hat{\theta}$, are those values, θ, which maximize eq. (B.1). Equivalently, let the MLE penalty function be:

$$Q_{MLE}(\theta) = -\Lambda(\theta)$$

$$= \sum_{j=1}^{K} [\ln \bar{V}_j + (V_j/\bar{V}_j)] \tag{B.2}$$

and $\hat{\theta}$ minimizes Q_{MLE}.

In a similar manner, if the quadratic penalty function is defined as:

$$Q_{MMSE}(\theta) = \tfrac{1}{2} \sum_{j=1}^{K} ((\bar{V}_j - V_j)/U_j)^2 \qquad (B.3)$$

Then the (weighted) minimum mean square error parameter estimates are those which minimize Q_{MMSE}. (where the weights are U_j^{-2}, j=1....K)

If ∇ is used to symbolize the gradient (w.r.t.θ) operator, then,

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial \theta_1} \\ \vdots \\ \frac{\partial}{\partial \theta_M} \end{pmatrix} \qquad (B.4)$$

A necessary condition on the estimators is that the following M equations be satisfied simultaneously.

$$\mathbf{c}(\theta) = \nabla Q(\theta)$$

$$= \begin{pmatrix} \frac{\partial Q}{\partial \theta_1} \\ \vdots \\ \frac{\partial Q}{\partial \theta_M} \end{pmatrix} = 0 \qquad (B.5)$$

Note that by differentiating eq. (B.2) and eq. (B.3), one may write a unified version of eq. (B.5) as:

$$\mathbf{c}(\theta) = \sum_{j=1}^{K} [(\bar{V}_j - V_j)/(\nu_j^2)] \nabla \bar{V}_j \qquad (B.6)$$

where to obtain a MLE estimate, one lets

$$\nu_j = \bar{V}_j \qquad (B.7)$$

and to obtain a MMSE estimate, one lets

$$\nu_j = U_j \tag{B.8}$$

The matrix form of eq. (B.5) is obtained by letting B be the K by M matrix defined as

$$B = [\nu_1^{-1} \nabla \bar{V}_1, \cdots \nu_K^{-1} \bar{V}_K]$$

$$= \begin{bmatrix} \nu_1^{-1} \frac{\partial \bar{V}_1}{\partial \theta_1} & \cdots & \nu_K^{-1} \frac{\partial \bar{V}_K}{\partial \theta_1} \\ \vdots & & \\ \nu_1^{-1} \frac{\partial \bar{V}_1}{\partial \theta_M} & \cdots & \nu_K^{-1} \frac{\partial \bar{V}_K}{\partial \theta_M} \end{bmatrix} \tag{B.9}$$

and letting D be the K dimensional normalized deviation vector,

$$D = \begin{bmatrix} (\bar{V}_1 - V_1)\ \nu_1^{-1} \\ \vdots \\ (\bar{V}_K - V_K)\ \nu_K^{-1} \end{bmatrix} \tag{B.10}$$

Then the necessary condition becomes

$$\epsilon(\theta) = BD = 0 \tag{B.11}$$

To find an iterative solution to eq. (B.5) or (B.11), let θ_n be the current approximation, and expand eq. (B.5) in a Taylor series about θ_n, thus,

$$0 = \epsilon(\hat{\theta}) \approx \epsilon(\theta_n) + [\nabla \epsilon^T(\theta_n)]\ (\hat{\theta} - \theta_n) \tag{B.12}$$

and

$$\hat{\theta} \approx \theta_n - [\nabla \epsilon^T(\theta_n)]^{-1} \epsilon(\theta_n) \tag{B.13}$$

Differentiating eq. (B.6), one finds:

$$\nabla \epsilon^T(\theta_n) = \sum_{j=1}^{K} \left[\frac{\nabla \bar{V}_j \nabla^T \bar{V}_j}{\nu_j^2} \right] + \sum_{j=1}^{K} (\bar{V}_j - V_j) \nabla \left[\frac{\nabla^T \bar{V}_j}{\nu_j^2} \right] \tag{B.14}$$

Assuming reasonably good starting values, one has $V_j - V_j \approx 0$, and the second term is small compared to the first so that;

$$\nabla \epsilon^T(\theta_n) \approx \sum_{j=1}^{K} \left[\frac{\nabla \bar{V}_j \nabla^T \bar{V}_j}{\nu_j^2} \right] \tag{B.15}$$

$$= (BB^T)$$

Thus the iterative equation becomes

$$\theta_{n+1} = \theta_n - k (BB^T)^{-1} BD \tag{B.16}$$

In the above equation, B and D are evaluated at the current estimate θ_n. The constant k is chosen less than unity. In fact, for k small enough the above procedure converges to a local minimum. That is,

$$Q(\theta_{n+1}) = Q(\theta_n) + [\nabla^T Q(\theta_n)] (\theta_{n+1} - \theta_n) \tag{B.17}$$

$$+ \text{higher order terms}$$

or

$$Q(\theta_{n+1}) \approx Q(\theta_n) + \epsilon^T(\theta_n) [\theta_{n+1} - \theta_n]$$

$$= Q(\theta_n) - k D^T B^T (BB^T)^{-1} BD \tag{B.18}$$