Bayesian estimation of soil parameters from radar backscatter data

Ziad S. Haddad
Pascale Dubois

Jet Propulsion Laboratory, California Institute of Technology

Abstract

Given measurements ml, $m_2, \ldots m_J$ representing radar cross-sections of a given resolution element at different polarizations and/or different frequency bands, we consider the problem of making an "optimal" estimate of the actual dielectric constant ϵ and the r.m.s. surface height h that gave rise to the particular $\{m_j\}$ observed. To obtain such an algorithm, we start with a data catalogue consisting of careful measurements of the soil parameters ϵ and h, and the corresponding remote sensing data $\{m_j\}$. We also assume that we have used this data to write down, for each j, an average formula which associates an approximate value of m_j to a given pair (ϵ, h). Instead of deterministically inverting these average formulas, we propose to use the data catalogue more fully and quantify the spread of the measurements about the average formula, then incorporate this information into the inversion algorithm. This paper describes how we accomplish this using a Bayesian approach. In fact, our method allows us to

1) make an optimal estimate of ϵ and h

2) place a quantitatively honest error bar on each estimate, as a function of the actual values of the remote sensing measurements

3) fine-tune the inital formulas expressing the dependence of the remote sensing data on the soil parameters

4) take into account as many (or as few) remote sensing measurements as we like in making our estimates of ϵ and h, in each case producing error bars to quantify the benefits of using a particular combination of measurements.

1 Introduction

Several investigators ([1], [5],[7],[8], [9], [10], [1 1],[12], [14], [15]) have studied the use of remote sensing methods to estimate the physical parameters of natural surfaces, namely the soil moisture content and the r.m.s. surface roughness. Because the laws of scattering from randomly rough natural surfaces are quite complicated, especially at microwave frequencies, empirical models have often been used to help express the observed remote sensing measurements as a function of the surface parameters. A typical approach, adopted in ([10]), is to start with a "training set" consisting of a catalogue of carefully collected data: in the case of [10], this catalogue consists of L-, C- and X-band polarimetric radar backscattering measurements for various bare soil surfaces, along wit h laser-profiler and dielectric-probe measurements of the corresponding r.m. s. surface height and dielectric constant values, Guided by the physics that govern electromagnetic scattering, and using the data at hand, a model relating the radar backscatter to the surface parameters can be established. Of course, the model will not agree exactly with the data in every instance. Reasons for mismatches include measurement error, the non-uniformity of the background power distribution, and the inhomogeneit y of the surface within a given resolution element or from one element to the next. Still, if the model is regarded as providing an *approximate* formula that is correct on average, one way to proceed is to disregard the data catalogue from this point on: given a particular radar measurement, a deterministic method can be used to invert the approximate model and retrieve the corresponding surface parameters, The accuracy of the retrieved parameters would naturally depend on the inversion method used, and would be difficult to quantify.

Another approach, that can potentially make fuller use of the data catalogue, is to model not only the approximate dependence of the radar backscatter on the surface parameters, but also the *spread* of the actual data about the approximate model. Indeed, the approximate model can be more or less accurate over certain intervals, Using this information about its accuracy, and how it depends on the values of the surface parameters, as evidenced" by the carefully collected data, can only help in the inversion problem. In fact,

1) a Bayesian approach can indeed use this information to produce an optimal algorithm, i.e. an algorithm which, among all possible algorithms, and on average, makes the smallest error in its estimates of the surface parameters.

2) Moreover, such an approach can quantify the accuracy of its estimates, depending, naturally, on the values of the measured radar backscatter in every case.

3) It also turns out that the approach allows one to fine-tune the initial approximate model to better fit the data.

4), Finally, the approach does not restrict one to a prespecified number of input measurements: indeed, it can use any combination of inputs to produce an estimate of the surface parameters that is based on these inputs. Moreover, it can quantify the uncertainty of these estimates. This is important because it provides a natural means of evaluating the usefulness of using one or another combination of measurements to estimate one or another surface parameter.

Section 2 summarizes the mathematics underlying the Bayesian approach in the case at hand. At the heart of this approach is the problem of modeling the spread of the data in the "training set" about the approximate model. This is described in section 3. Finally, section 4 discusses the results of this approach in the case of ([10]). The quantitative results are quite encouraging, especially when the method is used to estimate the surface parameters from the measured bascatter at all three frequencies simultaneously. The salient results of the first three sections are summarized at the beginning of section 4, for the benefit of the reader who would prefer to skip directly to that section and look at the results before delving into the theoretical details behind them.

2 Mathematical Approach

2.1 Motivation

For definiteness, we start by considering the following specific problem: Given two measurements m and n, representing respectively the ratio of HH to VV L-band radar cross-section and the ratio of HV to VV cross sections, respectively, of a single radar resolution element, we would like to make an "optimal" estimate of the correct pair (ϵ,h) that gave rise to the particular (m,n) observed. By "optimal", we mean that the r.m.s. difference between the optimal estimates and the actual values of ϵ and h should be smallest among all the errors made by any candidate estimators: the optimal method is the one which, on average, i.e. over many (all) observations, makes the smallest error.

How does one go about finding such an optimal procedure ? A natural way to proceed is to look for an expression of the form

$$\begin{aligned} m &= f(\epsilon,h) \\ n &= g(\epsilon,h) \end{aligned} \qquad (1)$$

Once such candidate functions have been identified, a direct inversion method can be used to "solve for ϵ and h" in equation (1). One can then apply this inversion method to a data

catalogue consisting of simultaneous measurements of m, n, ϵ and h, and use the r.m.s. difference between the predicted values of ϵ and h and the measured values to measure once and for all the success of the approach.

Yet it is unlikely that any given candidate functions f or g can make the equalities (1) exactly true, ever, because of noise and other uncertainties inherent to radar data. In fact, modeling the average dependence of m and n on ϵ and h *is* not sufficient, in itself, to allow one to determine which ϵ and h best correspond to given measurements m, n: one must still model the dependence of m and n on the many remaining factors, whose omission from equation (1) is indeed the reason that this equation is never "exactly verified, In other words, until one incorporates some information about the reason for which the right hand side of (1) does not exactly match its left hand side, one would find it difficult to justify a particular method of "solving for ϵ and h". Moreover, the r.m.s. error produced by any "direct inversion" approach can be due as much to the inexactness of the functions f and g used in the model, as to the shortcomings of the inversion method used. Our starting point in this paper is that using all the statistical information available in a reliable "training data set" would give a more complete mathematical approach to solving the problem and interpreting the inevitable shortcomings. Perhaps most important, a probabilistic approach would allow one to calculate not just optimally estimated values for the soil parameters, but also the associated uncertainty in each individual estimate.

2.2 Bayesian approach

It is clear from this discussion that in order to find an optimal procedure, one needs to make an effort to mathematically account for the discrepancy between the left– and right–hand-sides of (1), We would like to replace the deterministic equation (1) by a stochastic equation

$$\begin{aligned} \mathrm{m} &= L_1(\epsilon, h) \\ n &= L_2(\epsilon, h) \end{aligned} \tag{2}$$

where L_1 and L_2 are random variables *whose joint density function* $\mathcal{P}_{(L_1,L_2)}$ *is known* (and depends only on ϵ and h). In fact, we want to represent L_1 as

$$L_1 = f(\epsilon, h) \cdot M_1 , \tag{3}$$

where f represents the deterministic "typical)' or "average" way in which m depends on ϵ and h, and in which M_1 is a random variable that does not depend on ϵ or h and which represents the remaining randomness in m. Similarly, we will represent L_2 as

$$L_2 = g(\epsilon, h) \cdot \mathrm{M}_2 . \tag{4}$$

Once we have succeeded din establishing (2), (3) and (4), andidentifying the joint behavior of the random variables (M_1, M_2), the next step is to compute the conditional density function $\mathcal{P}_{(\epsilon,h)|(m,n)}$ fore, h given the measured values of m and n. F'rom Bayes'stheorem, it follows that the unnormalized version of this conditional density satisfies

$$\begin{aligned}\mathcal{P}_{(\epsilon,h)|(m,n)} &= \mathcal{P}_{(\epsilon,h)} \cdot \mathcal{P}_{(m,n)|(\epsilon,h)} \\ &= \mathcal{P}_{(\epsilon,h)} \cdot \mathcal{P}_{(L_1,L_2)} \\ &= \mathcal{P}_{(\epsilon,h)} \cdot \frac{1}{f(\epsilon,h)} \frac{1}{g(\epsilon,h)} \cdot \left(\frac{m}{f(\epsilon,h)} \cdot \frac{n}{g(\epsilon,h)}\right)\end{aligned} \tag{5}$$

where $\mathcal{P}_{(\epsilon,h)}$ is the a priori joint' density for (c, h), in which one includes all the a priori information about ϵ and h (such as estimates based on other instruments – in case one does not know anything a priori about them except their physical range of values, $\mathcal{P}_{(\epsilon,h)}$ would just be the uniform density function over the product of the corresponding intervals). The density function given by (5) must still be numerically normalized so that its integral is 1,

In any case, using the conditional density given by equation (5), the optimal unbiased estimator $\hat{\epsilon}$ for ϵ that has minimum variance (i.e. that minimizes the r.m.s. error) is the conditional mean

$$\hat{\epsilon} = \mathcal{E}\{\epsilon|(m, n)\} = \int \epsilon\, \mathcal{P}_{(\epsilon,h)|(m,n)}(\epsilon,h)\, d\epsilon\, dh\,. \tag{6}$$

Similarly, the optimal unbiased estimator $\hat{h}$ for h is the conditional mean $\mathcal{E}\{h|(m, n)\}$, Formula (6) is quite easy to discretize (in ϵ and h) and evaluate numerically. Thus, if we can replace the deterministic equation (1) by stochastic equations (2,3,4) in such a way that we also know the joint density function $\mathcal{P}_{(M_1,M_2)}$ of (M_1,M_2), we have a straightforward method of obtaining the optimal estimate, and of calculating all its moments (indeed, we have its whole density function !).

2.3 Joint density function

So how do we go about modeling L_1 and L_2, or, equivalently, M_1 and M_2 jointly, and how can wc determine the corresponding joint density function ? One obvious source of noise in (1) is measurement error. Another is the fact that f and g can only represent the mean values of the radar cross section ratios. The fact that the actual background power distribution is not a delta-function is another reason for (1) to be inexact. In fact, these three sources of error have been extensively studied. In particular, the results of [6] imply that if m represented the HH cross section by itself, and if f represented the true mean of the background distribution

of m, then the density function PM of the random variable $M = m/f(\epsilon, h)$ is

$$\mathcal{P}_M(x) \qquad \frac{2(N\mu)^{(N+\mu)/2}}{\Gamma(N)\Gamma(\mu)} - x^{(N+\mu-1)/2} K_{\mu-N}(2\sqrt{N\mu x}) \tag{7}$$

where K denotes the modified Bessel function of the second kind. In (7), we have also assumed that N radar looks were used to produce m, that the fading has Rayleigh characteristics, and that the background power level is I'-distributed with relative variance $1/\mu$. If we were to use this result in our case, where m and n are ratios of backscattering cross-sections, we would need to assume that each of M_1 and M_2 in equations (3,4) is the ratio of two random variables distributed according to (7). There are several reasons not to do this directly. First, the assumption that the background power level is I'-distributed is not always necessarily valid ([1 3]). Other distributions would produce different expressions for *PM*. Second, the presence of a Bessel function in expression (7) makes it unnecessarily difficult to use in practice. In fact, in the two extremes where μ is very large (corresponding to the case where the variance of the background level is very small) or N is very large (corresponding to the case where a large number of looks are used to average out uncertainties in the backscattered power), (7) reduces to a Γ-distribution. Finally, the candidate functions f and g may well turn out to be poor approximations of the true means. Indeed, even if one used a very accurate method to estimate the sample mean, one remains vulnerable to measurement error, and to contamination of the measurements by unknown scatterers on the surface ('(debris", etc).

Taking all these considerations into account, it is neither unreasonable nor arbitrarily restrictive to assume that the measurements m and n are related to ϵ and h by equations (2,3,4), in which

- in the case where the measured variable is the received power (at the HH–, HV– or VV-polarization), each M_i is itself I'-distributed, these distributions being mutually independent.
- in our case, where the measured variables are *ratios* of received powers, each M_i is distributed like the ratio of the two corresponding (independent) Γ distributions.

To determine the joint density function $\mathcal{P}_{(M_1,M_2)}$ for (M_1, M_2), *we* will use a reliable set of simultaneous measurements of m, n, ϵ and h, and test if this data is consistent with our assumptions about M, The model for the M_i's is still not completely specified: indeed, the parameters of the I'-distributions involved must still be chosen. As we shall see in the next section, there is an optimal way to determine these parameters from the data catalogue. Moreover, it will turn out that the a priori assumptions that we make about the exact form of the distribution of the M_i's are ultimately not crucial: any final expression that we settle on for the density function of the M_i's can and will be tested for

goodness-of-fit with the measurements ([2]). The problem of determining the parameters in the I'-distributions and testing the consistency of the resulting density function $\mathcal{P}_{(M_1,M_2)}$ with the data is addressed in detail in the next section,

3 Joint density for the Michigan model

Let us apply the procedure described above to the case where the model and data to be used are the University of Michigan Radiation Laboratory Model [1 O], and the corresponding set of radar cross-sections measured by the LCX polarimetric scatterometer POLARSCAT ([10]). Specifically, we assume that m = the ratio of HH to VV L-band radar cross-sections, n = the ratio of HV to VV cross-sections, and

$$f(\epsilon, h) = \left(1 - \left(\frac{2\theta}{\pi}\right)^{1/3\Gamma_0} e^{-kh}\right)^2$$
$$g(\epsilon, h) = 0.23\sqrt{\Gamma_0}(1 - e^{-kh}) \tag{8}$$

where θ is the incidence angle of the radar beam, k is the wave number, and $\Gamma_0 = \Gamma_0(\epsilon) = \left|\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right|^2$ is the Fresnel reflectivity of the surface at normal incidence, We further assume that the functions f, g model m, n in the sense that

$$\frac{m}{f(\epsilon, h)} = M_1 = \frac{N_1}{N_3}$$
$$\frac{\text{n}}{g(\epsilon, \mathbf{h})} = M_2 = \frac{N_2}{N_3}$$

where N_1, N_2 and N_3 are independent, I'-distributed random variables, Practically, this means that

$$\mathcal{P}_{(M_1,M_2)}(x, y) = \frac{1}{\xi\nu} \cdot \frac{(x/\xi)^{\alpha-1}(y/\nu)^{\beta-1}}{(x/\xi + y/\nu + 1)^{\alpha+\beta+\gamma}} \cdot \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \tag{9}$$

where the parameters $\alpha, \beta, \gamma, \xi$, v are to be determined,

We can determine these parameters and assess how appropriate this model is for the error, all at once, using the Michigan data [1 O]. Indeed, given this catalogue of reliable data consisting of quadruples (m, n, ϵ, h), in order to test for goodness-of-fit, we compute all the ratios $m/f(\epsilon, h)$, $n/g(\epsilon, h)$ and apply the χ^2 test to verify that the ratio values are consistent with the assumption that the joint distribution of (M_1, M_2) *is as* in equation (9). But our

proposed distribution depends on five parameters. In this case, the χ^2 test is still valid as long as we replace the parameters by their maximum-likelihood values computed from the data, and decrease the number of degrees of freedom for our χ^2 test by the number of parameters, i.e. by five ([2]).

Yet, a priori, it seems that we are using too many parameters. Indeed, the density function in (9) does not change very quickly as the parameters vary. We tried to reduce the number of parameters by trying to estimate them first from the marginal density functions corresponding to M_1 and M_2 separately. Using the University of Michigan set of scatterometer data for m only, the maximum-likelihood values for α, γ and ξ were $\alpha = 7.08$, $\gamma = 7.12$ and $\xi = 1.06$, with a fairly flat likelihood function, especially along the "ridge" $\alpha = \gamma, \xi = 1$. Similarly, the ML values for β, γ and ν, based on the data for n only, were@= 4.275, $\gamma = 4.26$ and $\nu = 0{,}85$, again with a fairly flat likelihood function along the "ridge" $\beta = \gamma$, v = 1. We are thus justified in making the simplifying hypothesis that (M_1, M_2) obeys (9) with $\alpha = \beta = \gamma$, and set out to determine the maximum-likelihood values for γ, ξ and ν based on the joint measurements for M_1 and M_2. The likelihood function to be maximized is

$$(\gamma - 1) \cdot \left(\frac{1}{J} \sum_j \log\left(\frac{M_1^{(j)}}{\xi}\right) + \frac{1}{J} \sum_j \log\left(\frac{M_2^{(j)}}{\nu}\right) \right) - 3\gamma \cdot \left(\frac{1}{J} \sum_j \log\left(\frac{M_1^{(j)}}{\xi} + \frac{M_2^{(j)}}{\nu} + 1\right) \right)$$
$$+ \log \Gamma(3\gamma) - 3 \log \Gamma(\gamma) \qquad \text{(lo)}$$

where J is the number of measurements at hand, The values at which the maximum is achieved are

$$\begin{aligned} \gamma (= \alpha = \beta) &= 5, \\ \xi &= 1.04, \\ \nu &= 0.82. \end{aligned} \qquad (11)$$

We now test our answer for goodness-of-fit. The joint distribution function can be obtained by integrating (9) directly. One finds that

$$\begin{aligned} \mathrm{pr}\{M_1 < A \text{ and } M_2 < B\} &= 1 - \frac{1}{(1+a)^5} - \frac{1}{(1+b)^5} + \frac{1}{(1+a+b)^5} \\ &- 5\frac{1}{(1+a)^6} - 5\frac{1}{(1+b)^6} + 5\frac{1}{(1+a+b)^6} - 15\frac{a'}{(1+a)^7} - 15\frac{b^2}{(1+b)^7} + 15\frac{(a+b)^2}{(1+a+b)^7} \\ &- 35\frac{a^3}{(1+a)^8} - 35\frac{b^3}{(1+b)^8} + 35\frac{(a+b)'}{(1+a+b)'} - 70\frac{a'}{(1+a)^9} - 70\frac{b'}{(1+b)^9} + 70\frac{(a+b)'}{(1+a+b)^9} \\ &+ 630\frac{b'a + 2b3a2 + 2b2a3 + ba^4}{(1+a+b)^{10}} + 1050\frac{3b^4a^2 + 4b3a3 + 3b2a4}{(1+a+b)]'} \end{aligned}$$

$$\frac{b^4a^3 + b3a4}{\text{‘}11550(1 + a + b)^{12}} + 34650(1+ a + b)^{13} \frac{b^4a^4}{} \tag{12}$$

where $a = A/\xi$, and $b = B/\nu$. We segment our data into eight disjoint events, and, using equation (12), compute the predicted frequencies for each event. The ratios of predicted-to-observed counts is summarized in table 1. The χ^2 statistic for these counts is 9.9. For

Ml \ M_2	$M_2 < 0.5$	$0.5 < M_2 < 1.i$	$1.1 < M_2 < 1.7$	$1.7 < M_2$
M_1 <1.1	7 / 9.77	18 / 14,88 ‘	7 / 4.03	3 / 1,27
$1.1 < M_1$	5 / 2.76	6 / 10.35	4 / 6.77	6 / 6,17

Table 1: Ratios of predicted-to-observed counts for (M_1, M_2)

a X2-variable with 5 degrees of freedom, the cut-off value for the critical region of size 0.05 is 11.1. Thus our value is well within the acceptable region, and we conclude that it is reasonable to assume that (9), with the parameters as in (11), is indeed the joint density function for (M_1, M_2) in the case of the Michigan model,

4 Results

Before describing our results, let us summarize the approach described in the previous sections. Write m for the ratio of HH to VV L-band radar cross-sections, n for the ratio of HV to VV cross-sections. We start with the model (refer to equations 2,3,4, 8, and to [10])

$$\begin{aligned} m &= f(\epsilon, h)\, Ml = \left(1 - \left(\frac{2\theta}{\pi}\right)^{1/3\Gamma_0} e^{-kh}\right)^2 M_1 \\ n &= g(\epsilon, h)\, M_2 = 0.23\, \sqrt{\Gamma_0}(1 - e^{-kh})\, M_2 \end{aligned} \tag{13}$$

where θ is the incidence angle, k is the wave number, $\Gamma = |(1-\sqrt{\epsilon})/(1+\sqrt{\epsilon})|^2$, and where the pair (M_1, M_2) of random variables has joint density function (refer to equations 9,11)

$$\mathcal{P}_{(M_1,M_2)}(x, y) = \frac{1}{(0.82) \cdot (1.04)} \cdot \frac{(x/1.04)^4\, (y/0,82)4}{(z/1.04 + y/.082 + 1)^{15}} \cdot \frac{14!}{(4!)^3} \tag{14}$$

Given specific values for (m, n), our optimal estimates $\hat{\epsilon}$ and $\hat{h}$ are then obtained using the ‘formulas (see equations 5)

$$\hat{\epsilon} = \int\int \epsilon\, \mathcal{P}(\epsilon, h)\, \frac{1}{f(\epsilon, h)}\, \frac{1}{g(\epsilon, h)}\, \mathcal{P}_{(M_1,M_2)}\left(\frac{m}{f(\epsilon, h)}, \frac{n}{g(\epsilon, h)}\right) d\epsilon\, dh \tag{15}$$

$$\hat{h} = \int\int h\,\mathcal{P}(\epsilon,h)\frac{1}{f(\epsilon,h)}\frac{1}{g(\epsilon,h)}\,\mathcal{P}_{(M_1,M_2)}\left(\frac{m}{f(\epsilon,h)},\frac{n}{g(\epsilon,h)}\right)d\epsilon\,dh \quad (16)$$

and t he corresponding error bars are given by

$$\sigma(\hat{\epsilon})^2 = \int\int (\epsilon-\hat{\epsilon})^2\,\mathcal{P}(\epsilon,h)\frac{1}{f(\epsilon,h)}\frac{1}{g(\epsilon,h)}\,\mathcal{P}_{(M_1,M_2)}\left(\frac{m}{f(\epsilon,h)},\frac{n}{g(\epsilon,h)}\right)d\epsilon\,dh \quad (17)$$

$$\sigma(\hat{h})^2 = \int\int (h-\hat{h})^2\,\mathcal{P}(\epsilon,h)\frac{1}{f(\epsilon,h)}\frac{1}{g(\epsilon,h)}\,\mathcal{P}_{(M_1,M_2)}\left(\frac{m}{f(\epsilon,h)},\frac{n}{g(\epsilon,h)}\right)d\epsilon\,dh\ . \quad (18)$$

The a priori density function $\mathcal{P}(\epsilon,h)$ will typically be assumed uniform over a rectangle in the (c, h)-plane. The integrals can be computed numerically. We are finally ready 'to describe the results of this approach.

4.1 Optimal estimates:

Figure 1 (resp. 2) is a contour plot of the optima.] estimate $\hat{kh}\,(\text{resp.}\,\hat{\epsilon})$ of $kh\,(\text{resp.}\,\epsilon)$, as a function of the cross section ratios m and n at $\theta = 40°$. The values of $\hat{kh}$ and $\hat{\epsilon}$ were obtained using our Bayesian approach, and starting with an a priori density function $\mathcal{P}(\epsilon,$ h) that is uniform over the rectangle $2 \leq \epsilon \leq 20$, $0 \leq kh \leq 1$. Overlaid on the contour plot of figure 1 (resp. 2) are those samples of the Michigan data that were collected at 40° incidence, each accompanied by the value of kh (resp.ϵ) computed according to the inversion algorithm proposed in [1 O], as well as the measured values, At 40°, the value of kh calculated by the direct inversion algorithm falls within 25% of the measured value in four out of the eight samples, but it misses by 100% in three cases. Our approach is not visibly more accurate. Similarly, the estimates of ϵ do not appear to be very accurate,

4.2 Error bars:

One way to measure the accuracy of the model functions f and g quantitatively is to look at the variance of the conditional density function, Indeed, these variances quantify the uncertainties in the estimates obtained by using the optimal approach. Figures 3 and 4 show the r.m.s. uncertainty in the estimates of kh and ϵ respectively. In this case, the model consisting of the function f and g of equation (8) can be considered "useful" if the r.m.s. uncertainty of the estimates that we obtain with it is smaller than the a priori uncertainty made by assuming that ϵ and kh are uniformly distributed, i.e. if the r.m.s. uncertainty in $\hat{\epsilon}$ is smaller than $(20-2)/\sqrt{12} \simeq 5{,}2$, and if the r.m. s. uncertainty in kh is smaller than

$(1-0)/\sqrt{12} \simeq 0.29$. Figures 3 and 4 show that this is indeed achieved everywhere. While it is encouraging to verify that using the model is an improvement over chosing values for ϵ and kh at random, it should be noted that the relative uncertainty in the estimates is quite large, Indeed, figure 3 shows that if the absolute value of the HH-to-VV ratio n is less than 21 dB, the 1. σ uncertainty in the estimate of kh exceeds 0.2, But according to figure 1, when n is around -20 dB, the optimal estimate of kh is itself around 0.3. Its 1. σ uncertainty is therefore close to 100% ! Overall, one can see from figures 1 and 3 that the smallest that the relative uncertainty in the estimate of kh gets is about 30%, when $|n|$ is less than 13 dB. *On* the other hand, figures 2 and 4 imply that the. r.m.s. relative uncertainty in the case of ϵ is never worse than 50%. Although still somewhat high, this value seems encouraging. Yet, in the case of c, one is typically interested in differentiating between "wet" and "dry" soil. Figure 5 compares the estimates for ϵ based on the actual measurements in the Michigan data, showing side-by-side four pairs of samples. Each pair represents two measurements at the same site, one when the soil was "wet" , followed by another when it was "dry". Note that the estimates for the first pair would erroneously indicate drier conditions on the first day. The "spread" for the remaining cases between the optimal estimates on the dry and wet cases is not very significant in comparison with the size of the r.m.s. uncertainty. In fact, the wet–dry difference is typically less than 1/4 of the r.m.s. uncertainty in the estimate of ϵ. 'l'his implies that the model will typically allow little discrimination between "wet" and "dry" soil.

4,3 Fine-tuning the initial model:

The results above illustrate the direct application of our estimation approach, and its ability to quantify the uncertainty in the estimates one obtains using a particular model. In this case, it turned out that this uncertainty is rather large. Let us now try to use our method to help improve the estimates. So far, we have been using the model expressed in equation (8). One way to improve our estimates is to tune the parameters in that model to the situation at hand. Specifically, one can postulate a model of the form

$$f(\epsilon, h) = \left(1 - \left(\frac{2\theta}{\pi}\right)^{a/\Gamma_0} e^{-kh}\right)^2$$

$$g(\epsilon, h) = b\Gamma_0{}^c (1 - e^{-kh}) \qquad (19)$$

together with a probability density function $\mathcal{P}_{(M_1,M_2)}$ for the observed ratios $(m/f$, n/g) of the form

$$\mathcal{P}_{(M_1,M_2)}(x,y) = \frac{x^{N-1}\, y^{N-1}}{(x+y+1)^{3N}} \frac{\Gamma(3N)}{\Gamma(N)^3} \qquad (20)$$

as before, then go on to determine a, b, c and N in order to maximize the likelihood of observing the radar cross sections reported in the Michigan data. In section 2, we derived the equation relating the density function $\mathcal{P}_{(L_1,L_2)}(m,n)$ for the observed HH-to-VV and HV-to-VV cross section ratios (m, n)to the density function $\mathcal{P}_{(M_1,M_2)}$. In fact, equation (5) states that

$$\mathcal{P}_{(L_1,L_2)}(m,n) = \frac{1}{f(\epsilon,h)}\frac{1}{g(\epsilon,h)}\mathcal{P}_{(M_1,M_2)}\left(\frac{m}{f(\epsilon,h)},\frac{n}{g(\epsilon,h)}\right) \tag{21}$$

The Michigan data include 32 observations at L-band with incidence angles between 30 and *60* degrees. Calling these observed ratios $\{(m_j, n_j), j = 1, \ldots, 32\}$, we are thus lead to find the values of a, b, c, N which maximize

$$\begin{aligned}
\prod_{j=1}^{32}\mathcal{P}_{(L_1,L_2)}(m_j,n_j) &= \prod_{j=1}^{32}\frac{1}{f(\epsilon_j,h_j)}\frac{1}{g(\epsilon_j,h_j)}\,\mathcal{P}_{(M_1,M_2)}\left(\frac{m_j}{f(\epsilon_j,h_j)},\frac{n_j}{g(\epsilon_j,h_j)}\right)\\
&= \prod_{j=1}^{32}\frac{1}{f(\epsilon_j,h_j)}\frac{1}{g(\epsilon_j,h_j)}\,\frac{(m_j/f(\epsilon_j,h_j))^{N-1}\,(n_j/g(\epsilon_j,h_j))^{N-1}\,\Gamma(3N)}{(m_j/f(\epsilon_j,h_j)+n_j/g(\epsilon_j,h_j)+1)^{3N}\,\Gamma(N)^3}\\
&= \prod_{j=1}^{32}\frac{1}{m_j n_j}\,\frac{\Gamma(3N)}{\Gamma(N)^3}\left(\frac{(m_j/f(\epsilon_j,h_j))\,(n_j/g(\epsilon_j,h_j))}{(m_j/f(\epsilon_j,h_j)+n_j/g(\epsilon_j,h_j)+1)^3}\right)^N\\
&= \left(\prod_{j=1}^{32}\frac{1}{m_j n_j}\right)\cdot\left(\frac{\Gamma(3N)}{\Gamma(N)^3}\right)^{32}\cdot\left(\prod_{j=1}^{32}\frac{(m_j/f(\epsilon_j,h_j))\,(n_j/g(\epsilon_j,h_j))}{(m_j/f(\epsilon_j,h_j)+n_j/g(\epsilon_j,h_j)+1)^3}\right)^N
\end{aligned} \tag{22}$$

It is apparent from this last equation that the problem of finding the optimal values of a, b, c decouples from the problem of finding the most suitable N. Indeed, to maximize (22), we must find the values of a, b, c which maximize

$$\prod_{j=1}^{32}\left(\frac{(m_j/f(\epsilon_j,h_j))\,(n_j/g(\epsilon_j,h_j))}{(m_j/f(\epsilon_j,h_j)+n_j/g(\epsilon_j,h_j)+1)^3}\right) \tag{23}$$

and that value of N which then maximizes

$$\left(\frac{\Gamma(3N)}{\Gamma(N)^3}\right)^{32} K^N \tag{24}$$

where K is the value of expression (23) when a, b, c take on their optimal values. This decoupling is in fact expected: the problem of finding the "right" values of the parameters $a, b,$ c should consist in matching the postulated form for f and g to the sample mean of the data at hand, while the problem of finding the "right" value of N involves quantifying how closely the "best" model then fits the data,

At L-band, we found that the optimal values for a, b and c were

$$\begin{aligned} a &= 0.33675 \\ b &= 0.12344 \\ c &= 0, \end{aligned} \tag{25}$$

and N = 15. We can now apply our Bayesian estimation approach using the previous model with these new values. Figure 6 shows a contour plot of the optimal estimate $\hat{kh}$ using our modified model with incidence angle O = 40°, as a function of m and n, overlaid, as before, with those samples of the Michigan data that were collected at 40° incidence, each accompanied by the value of kh computed according to the inversion algorithm proposed in [10], as well as the measured values. This time, the measured values of kh never fall farther than 20% away from the corresponding contour line, a very encouraging sign. Figure 7 shows the 1, σ uncertainty of the optimal estimates. Comparing figures 6 and 7, one sees that the worst-case relative r.m.s. uncertainty is now around 50%, a significant improvement over the original model, 'l'he results for ϵ are less encouraging. Figure 8 compares the estimates for ϵ with the values obtained by direct deterministic inversion and with the actual measurements in the Michigan data, showing, as before, the four wet-dry pairs of samples at 40° incidence. The estimates for the first pair still erroneously indicate drier conditions on the first day. The accuracy, along with the spread between the optimal estimates on the dry and wet cases for the other samples, have improved. Figures 9 and 10 show the case where the incidence angle is 30° and 50° respectively. The results at the steeper incidence angle are visibly better, as was predicted in [3]. Yet the variance of the estimate is still large compared with the difference between the wet and dry cases,

4.4 Incorporating additional measurements:

The last application of our method will consist in trying to reduce the uncertainty in our estimates, this time by trying to fuse data collected from different instruments. Indeed, in the case at hand, the data come from nine independent channels: HH, HV and VV polarizations at L-, C- and X-bands. To keep the notation simple, we shall index our variables with 1, 2 or 3 according to whether they relate to band L, C or X. To make use of the three polarizations from the three bands simultaneously, we must first replace kh and ϵ by two frequency-independent parameters, say h itself and the moisture content μ. We then need to determine model functions f_j, g_j, $j = 1,2,3$ that give a suitable approximation for the ratios (m_j, n_j), and we must determine how each pair (f_j, g_j) fits the data in its respective band. This is accomplished as before using a maximum likelihood process. The results in the case at hand are a_2 = 0.252, b_2 = 0,1399, C_2 = O and N_2 = 20 for C-band, and

$a_3 = 0.198$, $b_3 = 0.13, c_3 = 0.035$ and $N_3 = 29$ for X-band. Given 6 measurements of the HH-to-VV and HV-to-VV ratios in these three bands, the expression for the conditional density function $\mathcal{P}$ for (μ, h), conditioned on these observations, can be derived from equations (5):

$$\mathcal{P}(\mu,h) = \left(\prod_{j=1}^{3} \mathcal{P}_j\left(\frac{m_j}{f_j(\mu,h)}, \frac{n_j}{g_j(\mu,h)}\right) \frac{1}{f_j(\mu,h)\, g_j(\mu,h)}\right) \cdot \mathcal{P}_0(\mu,h) \qquad (26)$$

where $\mathcal{P}_0$ is the a priori density function for (μ,h), which we took to be uniform over the rectangle $0 < \mu < 0.4$, $0 < h < 3.2$ cm, and where $\mathcal{P}_j(x,y) = (\Gamma(3N_j)/\Gamma(N_j)^3)(xy)^{N_j-1}/(x + y + 1)^{N_j}$. The conditional means of $\mathcal{P}$ would then be the optimal estimates for μ and h given

actual h	actual μ	$\hat{\mu}_{30°}$	$\sigma(\mu_{30°})$	$\hat{\mu}_{40°}$	$\sigma(\mu_{40°})$	$\hat{\mu}_{50°}$	$\sigma(\mu_{50°})$	$\hat{\mu}_{60°}$	$\sigma(\mu_{60°})$
0,4 cm	0.29	0.17	0.063	0.25	0,064	0.24	0.060	0.28	0.061
0.4 cm	0.14	0.12	0.057	0.18	0.058	0.21	0.069	0.23	0.072
0.32 cm	0,30	0.23	0.062	0,32	0.053	0.34	0.047	0.30	0.059
0.32 cm	0.09	0.08	0.049	0.13	0.057	0.13	0.063	0.22	0,084
1.12 cm	0.31	0.21	0.104	0.30	0.074	0.31	0.067	0.33	0.056
1,12 cm	0.15	0.15	0.090	0.23	0.092	0.17	0.103	0.21	0,101
3,02 cm	0.19	0.16	0,109	0.23	0.107	0.24	0.102	0.28	0.088
3.02 cm	0.16	0.16	0.106	0,17	0.109	0.21	0.107	0.25	0.095

Table 2: Soil moisture estimates using HH/VV and HV/VV ratios at L, C and X bands

the six measurements at hand, Table 2 summarizes the results for μ. In order to use the model functions f and g, we needed to convert μ to a corresponding value of ϵ. To do this, we used the semi-empirical formula

$$\epsilon = \mathrm{Re}\left(\left(1.686 + \mu^{1.16}\left(4.9 + \frac{74.1}{1 + if/18.64}\right)^{0.065}\right)^{1.538}\right) \qquad (27)$$

similar to the one used in ([4]), and in which f is the frequency in GHz, and $i = \sqrt{-1}$. The resulting estimates are quite good. The estimated r.m s. uncertainty almost never reaches 50% of the estimated moisture content anymore. In fact, except in the roughest case, the wet—dry spread is typically as large as the uncertainty in the estimate, whereas in the L-band-only case it was very much smaller. This is very important because it implies that the combination of L-, C- and X-bands does allow one to estimate the soil moisture content accurately enough to discriminate between wet and dry conditions, Table 3 shows the results

actual h	$h_{30°}$	$\sigma(h_{30°})$	$h_{40°}$	$\sigma(h_{40°})$	$h_{50°}$	$\sigma(h_{50°})$	$h_{60°}$	$\sigma(h_{60°})$
0.40	0.29	0.065	0.36	0.081	0.36	0.079	0.38	0.076
0.40	0,29	0.063	0.38	0.091	0.60	0.160	0.66	0.165
0.32	0.22	0.044	0.36	0.077	0.35	0.069	0,43	0,089
0.32	0.29	0.061	0,47	0.121	0.73	0.226	1.21	0.408
1.12	1.39	0.494	1.09	0,365	1.17	0,354	1.02	0.246
1.12	0.79	0.261	1.13	0.434	2,26	0,582	2.33	0.557
3.02	2.59	0.467	2.68	0.416	2.70	0.407	2.70	0.391
3.02	1,97	0.615	2.38	*0.550*	*2,44*	*0.526*	*2.55*	*0.466*

Table 3: Soil r.m.s. height estimates (in cm) using HH/VV and HV/VV ratios at L, C and X bands

for the r.m.s. height h. The best results arc the ones at 40° incidence, although the remaining estimates are quite good too.

This example does demonstrate the ability to produce sharper estimates by combining data from different channels using our approach, We intend to test this procedure more extensively on SIR-C / X-SAR data. For now, our purpose in describing this example is merely to illustrate a feature unique to our Bayesian approach: indeed, a deterministic inversion approach would be hard pressed to compute the values of two variables, given a greater number of equations (six in our case), let alone estimate the error in the result at the same time.

5 Conclusions

The method presented above allows one to use any model for the dependence of radar cross-section measurements on soil parameters in order to obtain improved ("(optimal") estimates of these parameters, along with the variance in the estimates based on the model. The method achieves this by calculating the entire joint conditional density function for the problem. When one has a model with a priori unknown parameters, the method can still be applied in combination with a maximum-likelihood approach to estimate the unknown parameters.

While there may be more direct (exact or approximate) ways to compute the conditional

mean, calculating the conditional density itself is quite interesting and useful. Indeed, as was demonstrated in the case of the combination of L- C- and X- band measurements, this approach can be used to quantify the improvement afforded by incorporating a particular measurement, by comparing the conditional (minimal) variance of the density function that is conditioned on this particular measurement with the unconditional one to get a first-order measure of the utility of using the measurement in question. Perhaps equally important, the last application demonstrated how one can use the conditioned density as a new *a priori* density function and incorporate observations from additional instruments by applying this algorithm repeatedly, each time updating the density function from the previous step in the data fusion. We intend to test this approach more extensively on the data that will be gathered by the SIR-C / X-SAR Space Shuttle experiment.

We are currently evaluating different models for the dependence of active radar measurements (at various frequencies and polarizations) on the soil parameters. More specifically, we use our method to compare the variance of the optimal estimates obtained using the various models under consideration. In addition, the results in the bare-soil case presented this paper encourage us to believe that the method can be extended to account for more complex sources of randomness in the measurements, such as the presence of vegetation. Finally, we intend to apply our approach to optimally fuse passive radiometric measurements together with the active radar data to obtain estimates of the soil parameters that, it is hoped, will have a correspondingly smaller variance.

6 Acknowledgements

We wish to thank Professors Fawwaz Ulaby and Kamal Sarabandi, and the Radiation Laboratory of the Department of Electrical Engineering and Computer Science at the University of Michigan for graciously sharing their data with us. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

[1] A.A. Chukhlantsev and S,1. Vinokur: *Radar sensing of soil and vegetation cover,* Soviet J. Rem. Sens. 9, 1991, pp. 570-579,

[2] R.D'Agostino and M.A. Stephens, Eds.: *Goodness of fit techniques,* M. Dekker, New York, 1986.

[3] M.C. Dobson and F.T. Ulaby: *Active microwave soil moisture research,* I.E.E.E. Trans. Geossci. Rem. Sens. 24, 1986, pp. 23-28.

[4] M.T. Hallikainen, F.T. Ulaby, M.C. Dobson, M.A. el-Rayes and L. Wu: *Microwave* dielectric *behavior of wet soil-I,* I. E.E.E. Trans. Geossci. Rem. Sens. 23, 1985, pp. 25-34.

[5] T.J. Jackson, K.G. Kostov and S.S. Saatchi: *Rock fraction effects on the interpretation of microwave emission from soils, I. E. E. F.,* Trans. Geosci. Rem. Sens. 30, 1992, pp. 610-616.

[6] D.J. Lewinski: *Nonstationary probabilistic target and clutter scattering models,* I. E.E.E. Trans. Ant. Prop. 31, 1983, pp. 490-498.

[7] H.H. Lim, A.A. Swartz, H.A. Yueh, J.A. Kong, et al: *Classification of earth terrain using polarimetric synthetics aperture radar images,* J. Geophys. Res. 94, 1989, pp. 7049-7055.

[8] M. Moghaddadm and S.S. Saatchi: *An inversion algorithm applied to SAR data to retrieve surface parameters,* Proc.IGARSS '93, Tokyo, pp. 587-589.

[9] E.G. Njoku and Y.H. Kerr: *A semiempirical model for interpreting microwave emission from semi-arid land surfcaces as seen from space,* I. E.E. E. Trans. Geosci. Rem. Sens. 28, 1990, pp. 384-393.

[10] Y. Oh, K. Sarabandi and F.T. Ulaby: *An empirical model and an inversion technique for radar scattering from bare soil surfaces,* I.E.E.E. Trans. Geosci. Rem. Sens. 30, 1992, pp. 370-381.

[11] C. Schmullius and R. Furrer: Some *critical remarks on the use of C-band radar data for soil moisture detection,* Intl. J. Rem, Sens. 13, 1992, pp. 3387-3390.

[12] L. Tsang, J.A. Kong, E.G. Njoku, D.H. Staelin and J.W. Waters: *Theory for microwave thermal emission from a layer of cloud or rain,* I. E.E. E. Trans. Ant. Prop, 25, 1977, pp. 650-657.

[13] F.'l'. Ulaby and M.C. Dobson : *Handbook of radar scattering statistics for terrain,* Artech House, Norwood MA, 1989.

[14] J.R. Wang and B.J. Choudhury : *Remote sensing of soil moisture content over bare field at 1.4 GHz frequency,* J. Geophys. Res, 86, 1981, pp. 5277-5282.

[15] J.J. van Zyl and C.F. Burnette : *Bayesian classification of polarimetric SAR images using adaptive a-priori probabilities,* Intl. J. Rem, Sens. 13, 1992, pp. 835-840.

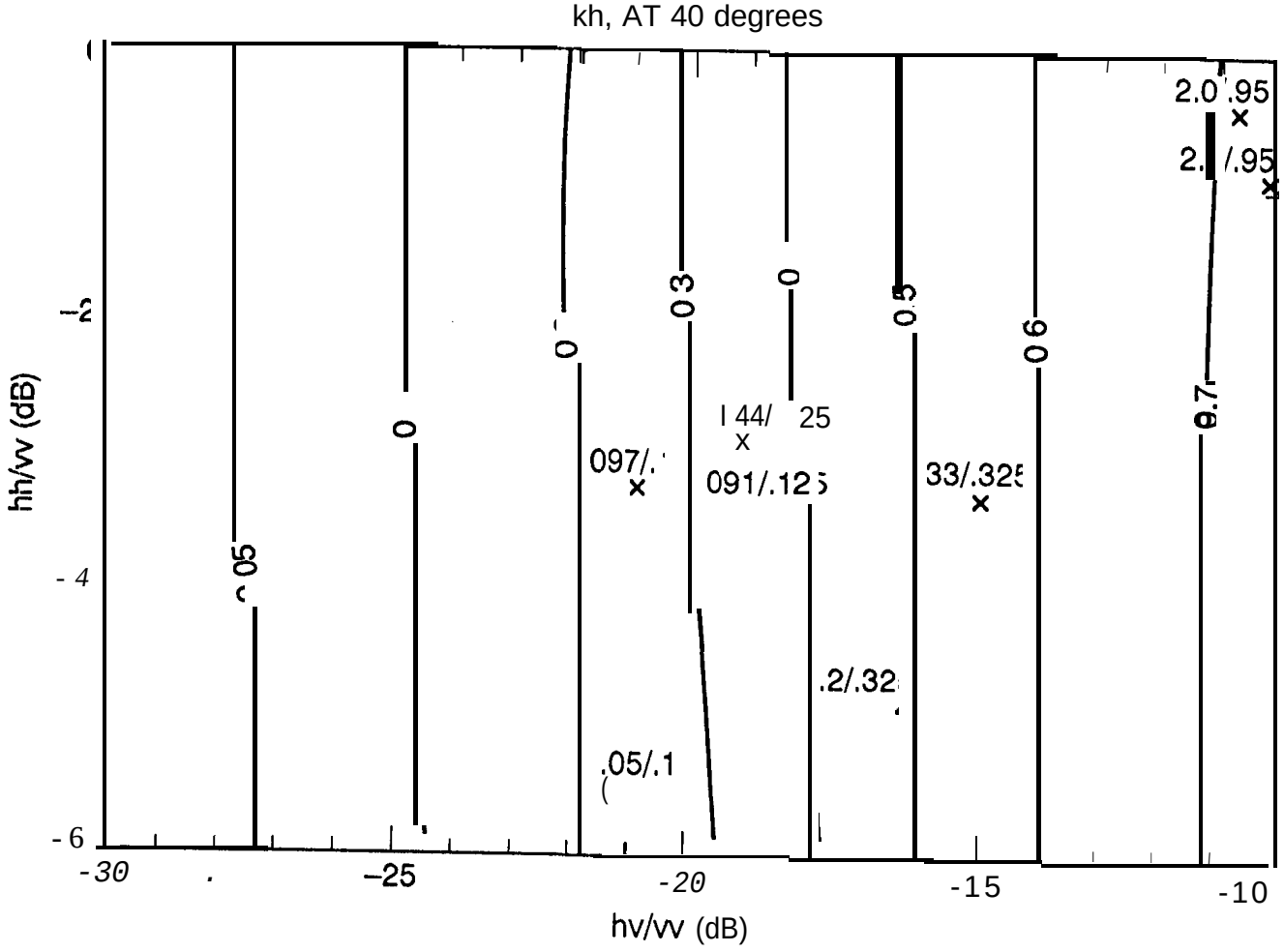
kh, AT 40 degrees
hh/vv (dB)
-4
-6
-30
-25
-20
-15
-10
hv/vv (dB)
2.0 .95
2. /.95
0 3
0
0 5
0 6
0 7
0
I 44/ 25
091/.12
33/.325
.2/.32
.05/.1

Figure 1

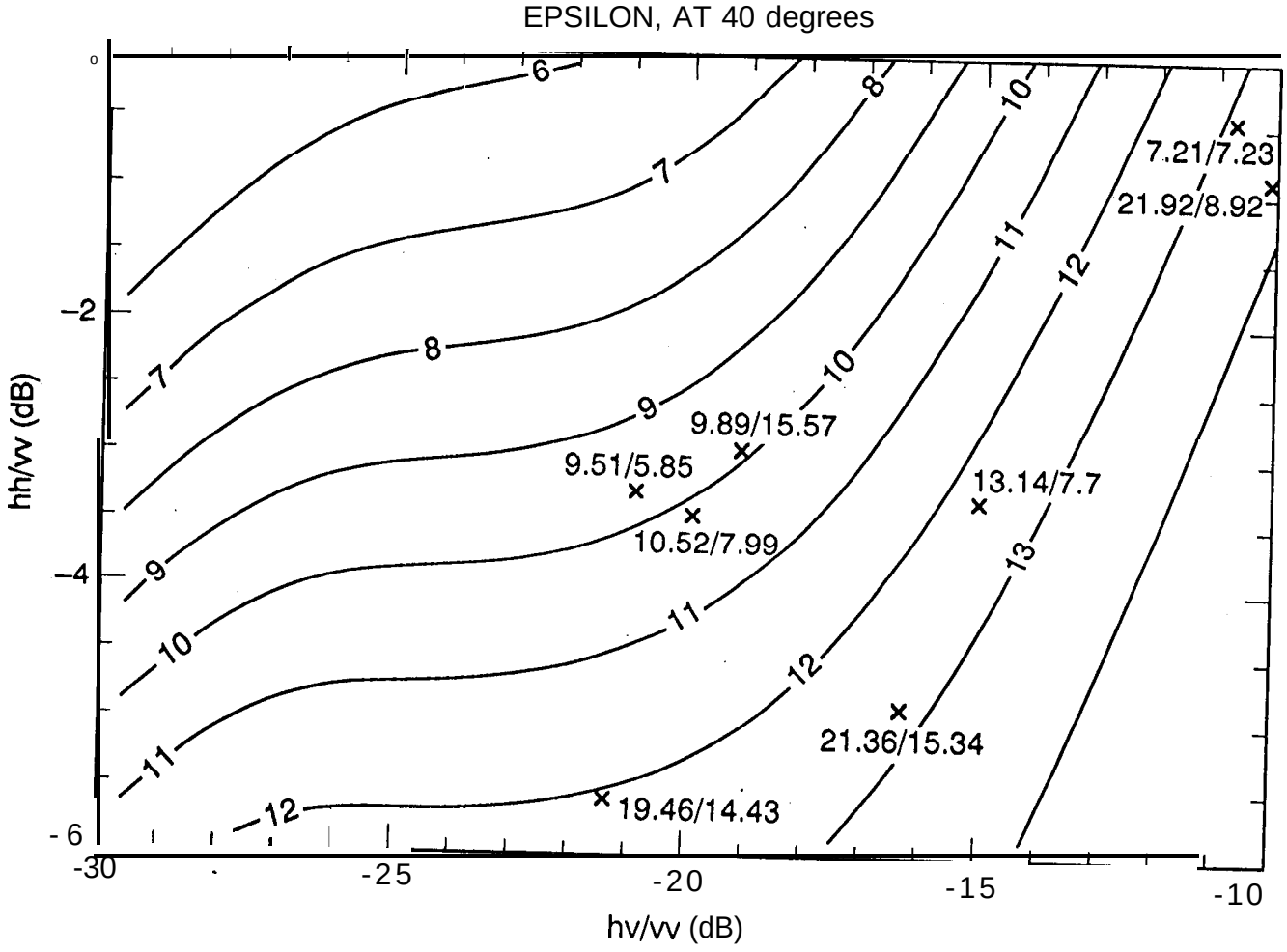
EPSILON, AT 40 degrees
hh/vv (dB)
hv/vv (dB)
0
-2
-4
-6
-30
-25
-20
-15
-10
6
7
8
9
10
11
12
13
7.21/7.23
21.92/8.92
9.89/15.57
9.51/5.85
10.52/7.99
13.14/7.7
21.36/15.34
19.46/14.43

Figure 2

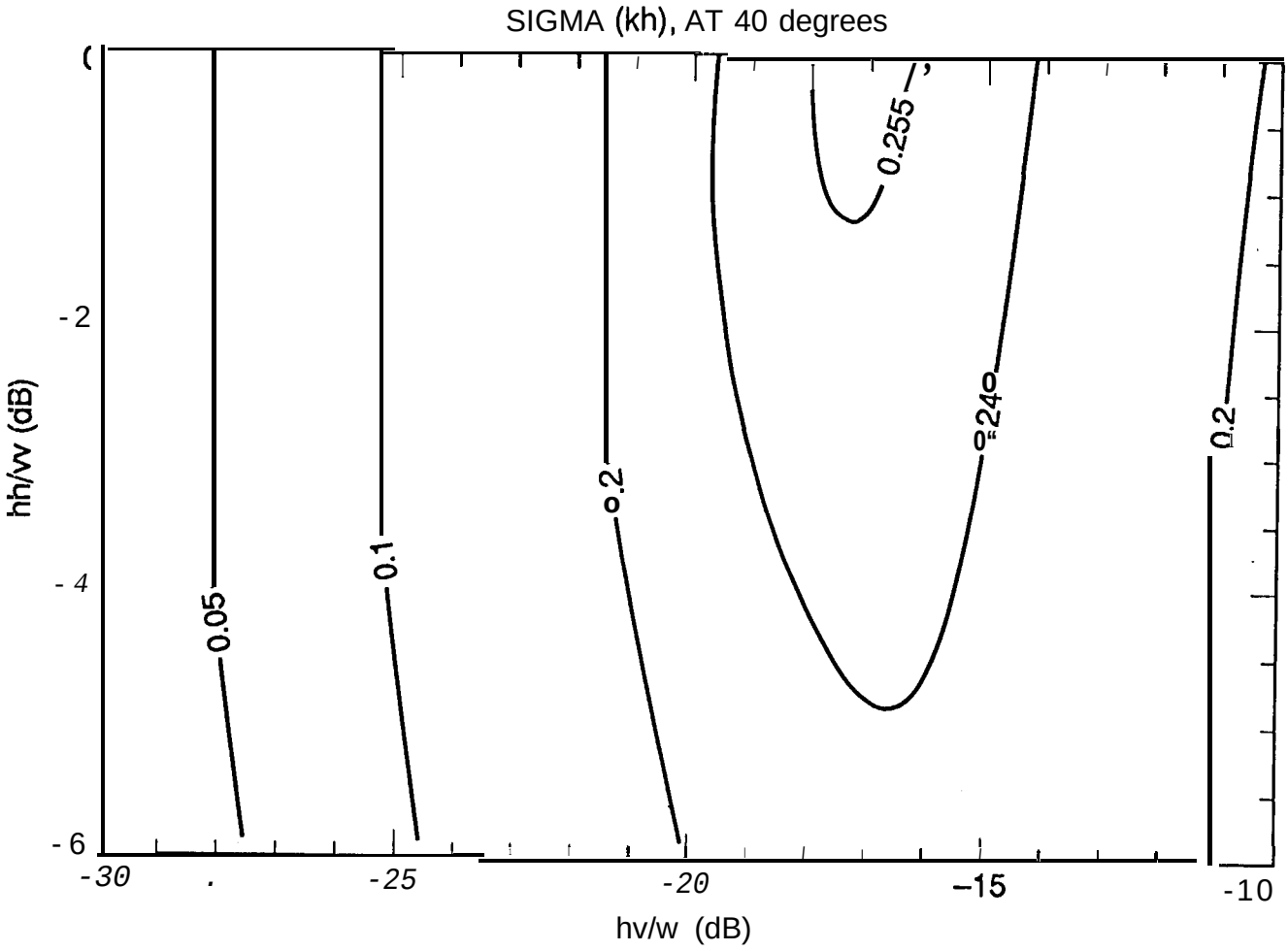
SIGMA (kh), AT 40 degrees
hh/vv (dB)
- 2
- 4
- 6
-30
-25
-20
-15
-10
hv/w (dB)
0.05
0.1
0.2
0.255
0.240
0.2

Figure 3

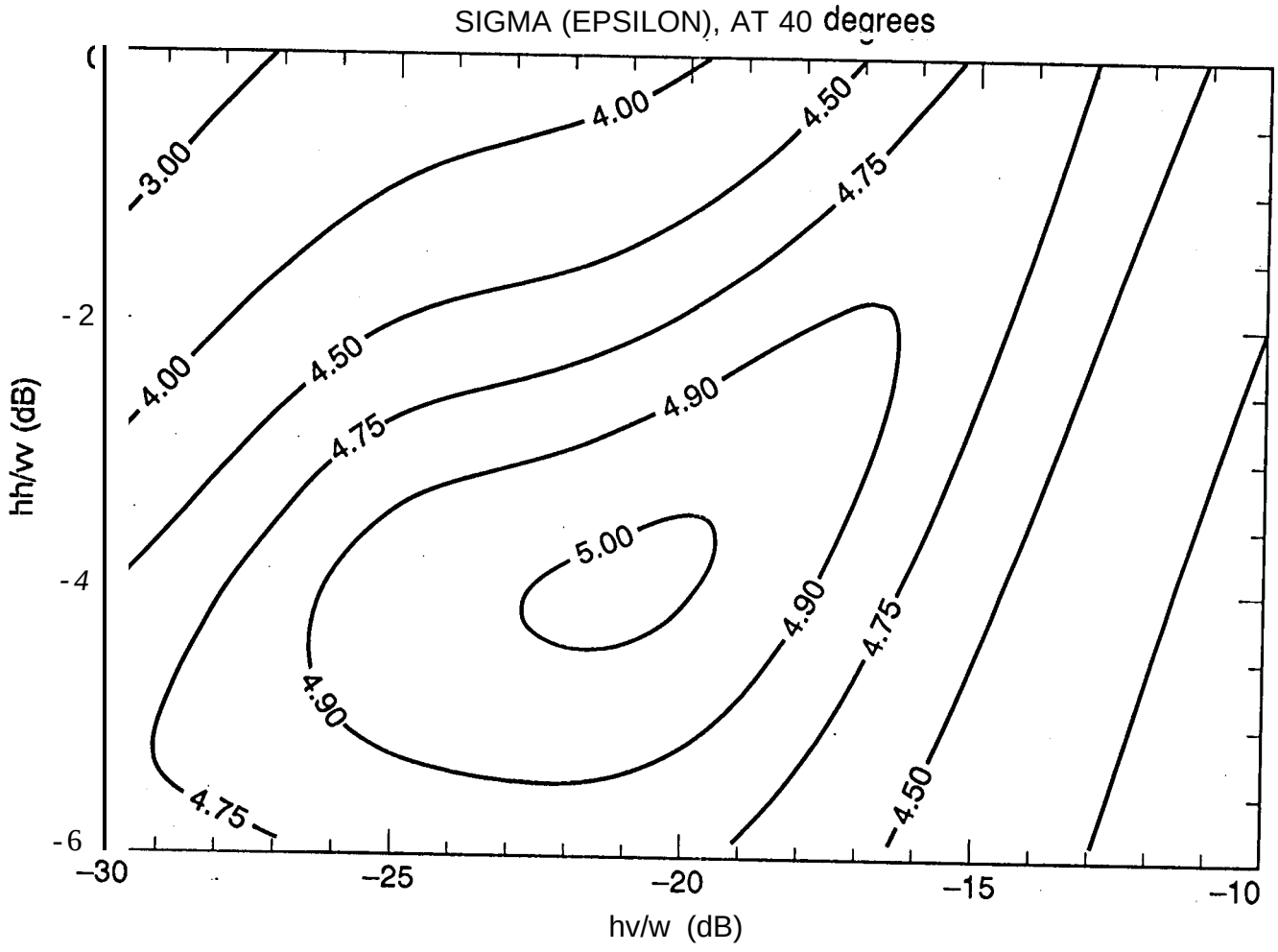
SIGMA (EPSILON), AT 40 degrees
3.00
4.00
4.50
4.75
4.00
4.50
4.75
4.90
5.00
4.90
4.75
4.90
4.75
4.50
-2
-4
-6
hh/vv (dB)
−30
−25
−20
−15
−10
hv/w (dB)

Figure 4

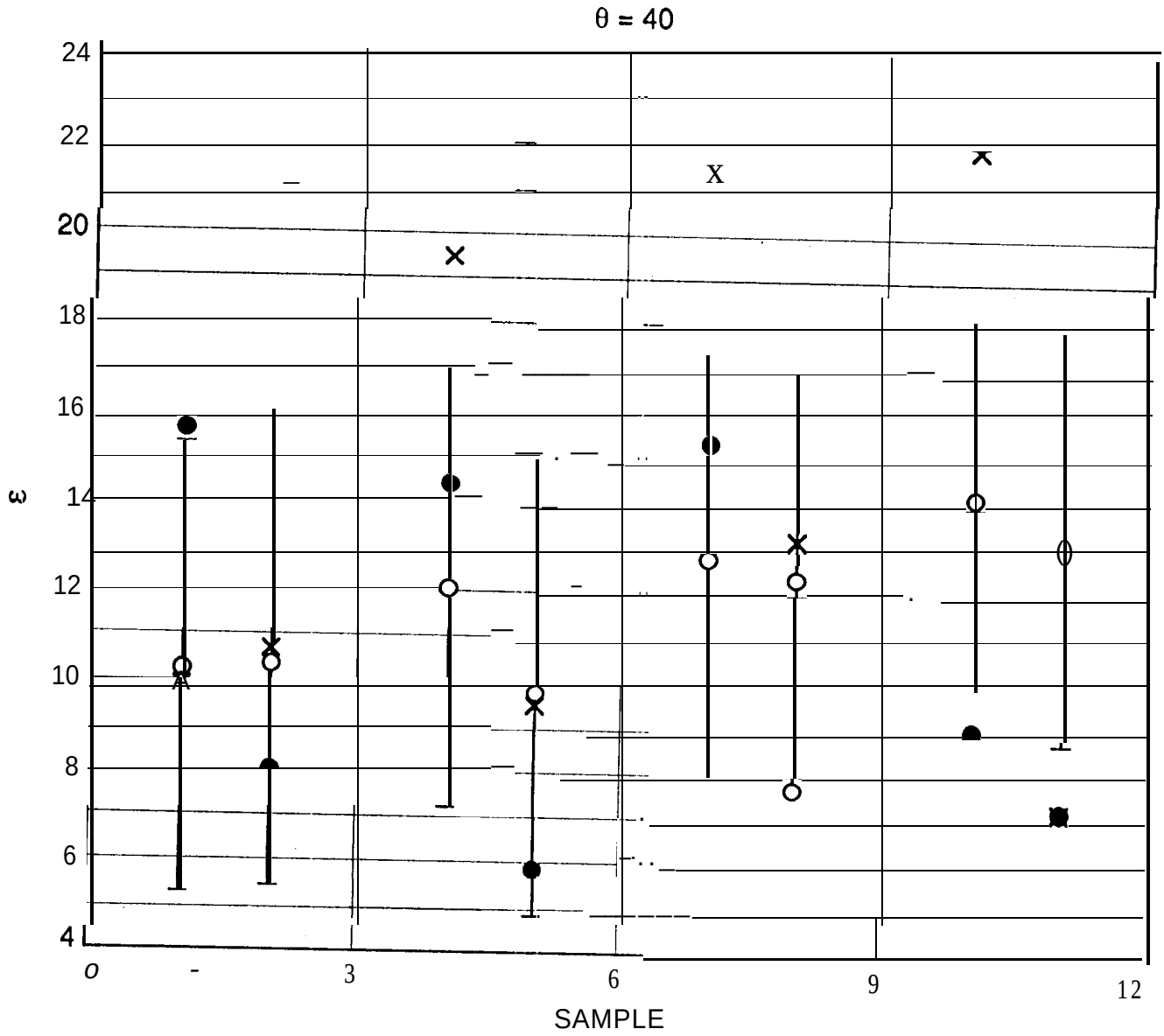
θ = 40
24
22
20
18
16
14
12
10
8
6
4
ε
o
3
6
9
12
SAMPLE
EPSILON
MICHIGAN ESTIMATE
OUR ESTIMATE

Figure 5

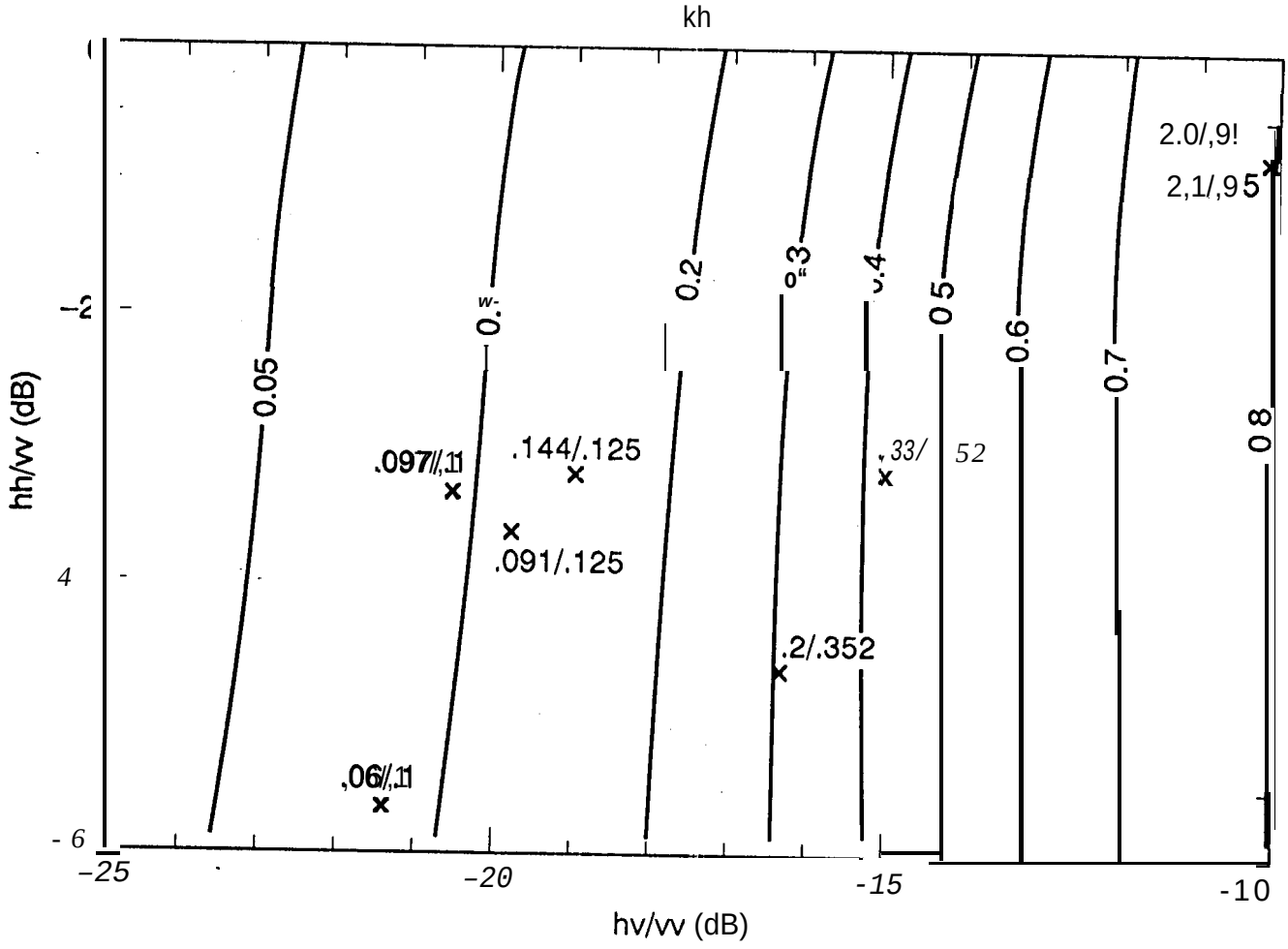
kh
0.05
0.2
0.5
0.6
0.7
0.8
.144/.125
.091/.125
.2/.352
2.0/,9!
2,1/,95
hh/vv (dB)
hv/vv (dB)
-25
-20
-15
-10

Figure 6

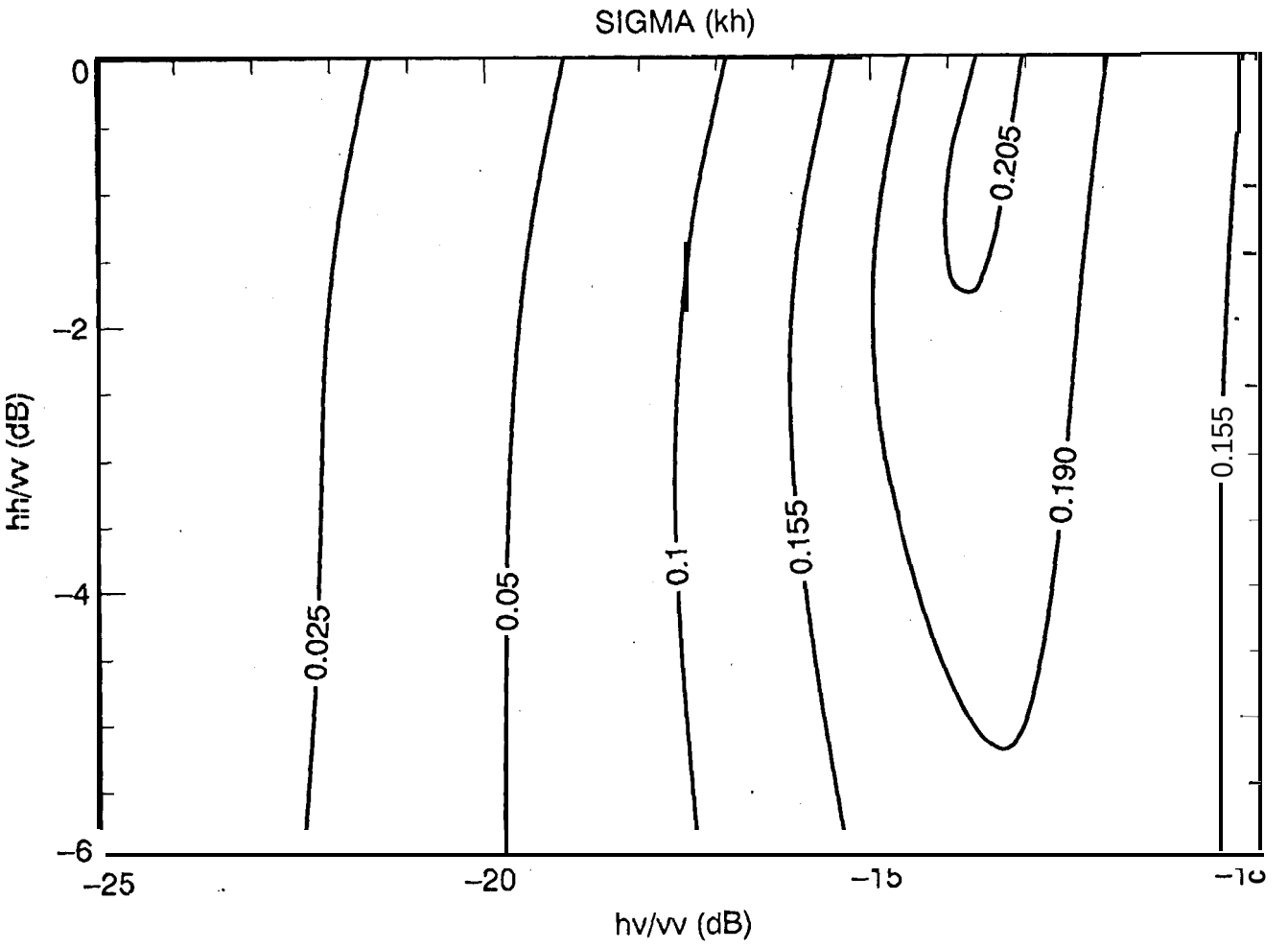
SIGMA (kh)
0
-2
-4
-6
hh/vv (dB)
0.025
0.05
0.1
0.155
0.205
0.190
0.155
-25
-20
hv/vv (dB)

Figure 7

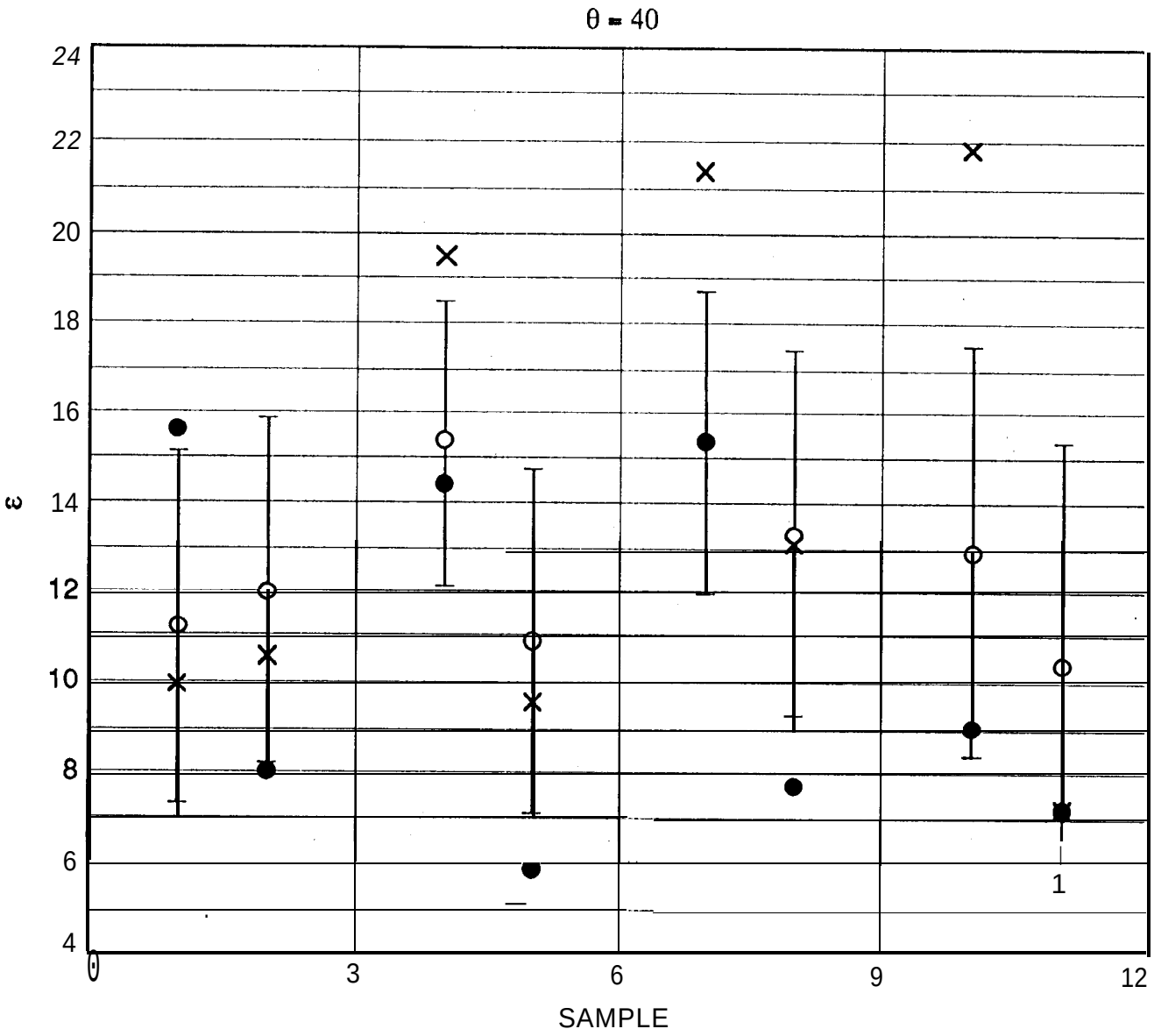
θ = 40
24
22
20
18
16
14
12
10
8
6
4
ε
0
3
6
9
12
1
SAMPLE
● EPSILON
x MICHIGAN ESTIMATE
0 OUR ESTIMATE

Figure 8

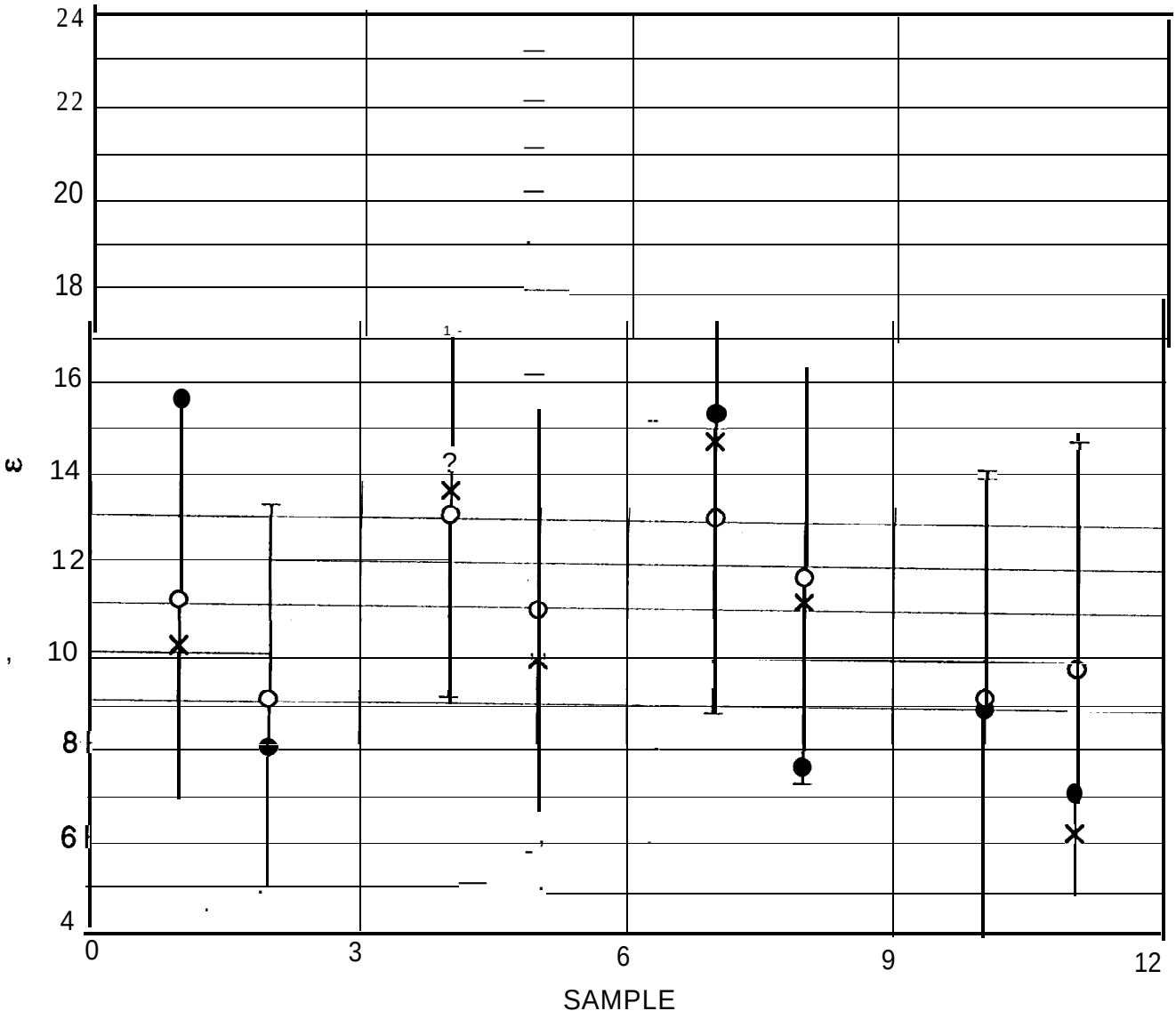
24
22
20
18
16
14
12
10
8
6
4
ε
0
3
6
9
12
SAMPLE
● EPSILON
X MICHIGAN ESTIMATE
0 OUR ESTIMATE

Figure 9

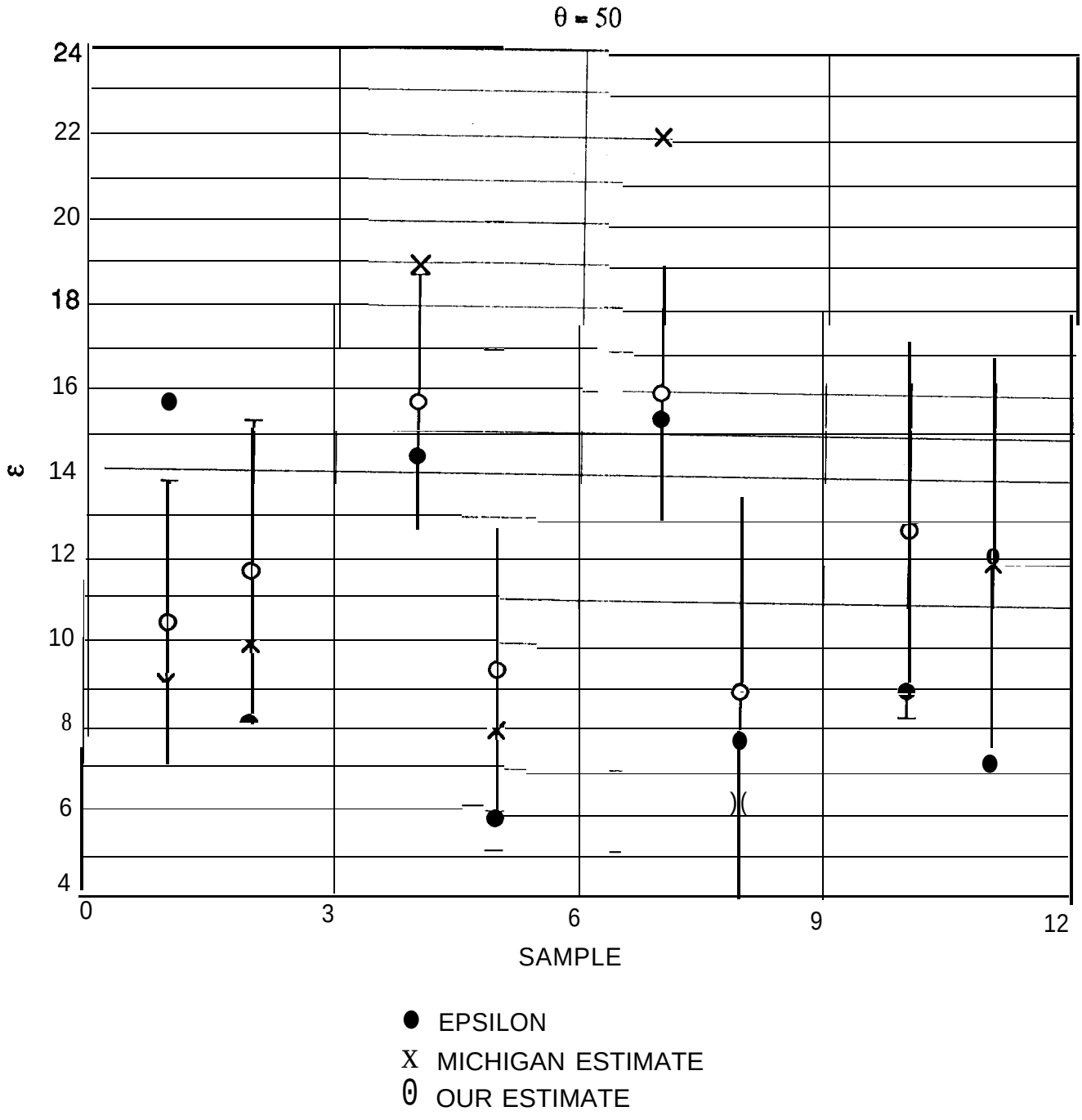
θ = 50
24
22
20
18
16
14
12
10
8
6
4
ε
0
3
6
9
12
SAMPLE
EPSILON
X MICHIGAN ESTIMATE
0 OUR ESTIMATE

Figure 10