J. S. Afr. Inst. Min. Metall., vol. 85, no. 4.
Apr. 1985. pp. 131-135

Statistical aspects of materials balancing in the minerals industry

by C.A. HUNT* and A.L. HINDE†

SYNOPSIS

The problem of materials balancing is reviewed with particular reference to the minerals industry. A broad statistical framework is established that, besides extending the classical estimation problem, allows questions concerning the value of sample information to be addressed. The following are examples of such questions: How much money should be spent on sampling? Which streams in a plant should be sampled? What should be sampled? The approach adopted makes use of Bayesian analysis. The discussion includes an illustrative example involving a simple flotation circuit.

SAMEVATTING

Die probleem van materiaalbalansering word in oënskou geneem met spesiale verwysing na die mineraalbedryf. Daar word 'n breë statistiese raamwerk daargestel wat, benewens die uitbreiding van die klassieke ramingsprobleem, dit ook moontlik maak om aandag aan vrae oor die waarde van monsterinligting te skenk. Die volgende is voorbeelde van sodanige vrae. Hoeveel geld moet daar aan monsterneming bestee word? Van watter strome in 'n aanleg moet daar monsters geneem word? Waarvan moet daar monsters geneem word? Die benadering wat gevolg word, maak gebruik van Bayes-ontleding. Die besprekings sluit ter illustrasie 'n voorbeeld van 'n eenvoudige flottasiekring in.

Introduction

In the minerals industry, there is a constant need to increase mineral recoveries subject to grade and tonnage constraints. The effectiveness of any strategy to meet this objective is clearly dependent on the degree of quantitative information that is available concerning the steady-state balance of material flowrates in a given plant. Obviously, if no information is available, it is impossible to implement a meaningful strategy. On the other hand, the installation of sophisticated on-stream instrumentation and the execution of extensive sampling campaigns on a routine basis can be a very costly exercise. Clearly, a trade-off must exist between the costs of obtaining extra information about the mass balance of a circuit and the net profits that are likely to accrue as a result of this extra information. Unfortunately, the problem of finding an optimal trade-off is aggravated by the fact that mineral processes are inherently stochastic in their behaviour. Accordingly, information obtained about the mass balance of the plant should be interpreted in statistical rather than deterministic terms. This paper provides a basis to meet this requirement, its main aim being to indicate how a statistical approach can be used in the estimation of the monetary value of sample information.

Preliminary Considerations

At the outset, some of the terms that will be used in the discussion are defined. Firstly, the plant will be regarded as a *network* of *nodes* connected by numbered *arcs*. The nodes represent the physical units in the plant such as flotation cells, grinding mills, and hydrocyclones. The arcs represent the pulp flows between units. One is usually interested in certain *species* in the pulp. For example, for Witwatersrand ores, species of interest might be gangue, gold, sulphur, and possibly particle size. In general, product and recycle streams can be sampled to obtain measurements of the concentration of the various species. Estimates of the mass flowrates of each species in these streams, relative to the mass flowrates of the circuit feeds, can then be calculated from these measurements. In practice, it is virtually impossible to extract perfectly representative samples from a plant, let alone maintain the plant under conditions of ideal steady state. It is therefore not surprising that redundant and replicate measurements are usually inconsistent and lead to uncertainty with regard to the estimate of a unique materials balance. The establishment of an acceptable and logical procedure for the adjustment of the raw measurements to give a self-consistent mass balance is referred to as the *materials balance problem.*

The task of adjusting concentrations or assays has been described in many papers, a good review being that by Reid *et al.*[1]. The general approach in these papers is the design of a computer-based algorithm to perform a non-linear minimization of some error criteria, such as the sum of squares of adjustments made to the observed assays. Thus, the algorithm yields an estimate of the 'true' balance at the time samples were taken. Some workers[2] also provide an idea of the error in the estimate.

The emphasis of the present paper is on the statistical interpretation, rather than algorithmic aspects, of the materials balance problem.

Principles of the Value of Sample Information

This section gives an overview of basic concepts and the general structure of the approach. Some specific ap-

* Anglovaal Limited, P.O. Box 62379, Marshalltown, 2107 Transvaal.
† Gold Exploitation Laboratory, Chamber of Mines of South Africa Research Organization, P.O. Box 91230, Auckland Park, 2006 Transvaal.

SA ISSN 0038-223X/$3.00 + 0.00. Paper received 28th August, 1984.

plications of the approach are outlined.

Variance Costs Money

There are many sources of error when a plant is sampled to establish its steady-state composition. These sources include random process disturbances, imperfect sampling, and assaying or instrumentation errors. Gy[3] gives a full treatment of the systematic and random errors that are likely to occur in practice.

Control action is usually taken as a result of information obtained about the materials balance. If no action is ever taken, there would of course be little point in sampling at all. Control action may take a variety of forms, such as a change in the set-point of an automatic controller or a change in reagent dosage to a bank of flotation cells. If the balance is exactly true (with an error variance of zero), correct action can be taken and no loss is incurred. However, the less precise (with a higher error variance) the materials balance, the greater the economic loss because, on average, the resulting control actions will depart from the ideal. It follows that the error distribution of the estimated materials balance determines the expected economic loss.

In the quantification of the expected loss, it is important that emphasis should be placed on the variables that are critical for the control of the circuit. In this paper, it is assumed that the solids recoveries at each separator node, given by the vector $\boldsymbol{\theta}$, are the critical variables about which information is sought. Thus, the estimation of the error variance of $\boldsymbol{\theta}$ will determine the expected loss of a given sampling situation. This is a convenient rather than a necessary assumption. A simple assumption along these lines would be as follows:

$$\text{Loss} = k(w_1 \sigma_1^2 + w_2 \sigma_2^2 + \ldots + w_n \sigma_n^2), \ldots\ldots (1)$$

where n is the number of separator nodes, σ_i^2 is the error variance of the estimated solids recovery at node i, and w_i is the relative weight attached to σ_i^2. The factor k translates loss into monetary terms and depends on many items, such as the economic value of the plant products, the response of the plant to various design and operational changes, and the nature of the control strategy. From Equation (1) it follows that perfect information (with sampling-error variances of zero) would imply zero loss; but the greater the error variances, σ_i^2, the greater the economic loss.

Even if no idea about the loss function is available, the theory of this paper can still be used to provide useful information concerning confidence regions for $\boldsymbol{\theta}$ under different sampling schemes, which can be compared on a qualitative basis. This approach is used for convenience in the illustrative example given later, where it is assumed that loss is proportional to the area of the 80 per cent confidence region for $\boldsymbol{\theta}$.

Error Distribution of Solids Recoveries

In the calculation of the error distribution of solids recoveries at each node, it is important that there should be an error model for the sampling and assaying errors of the original data. It will be assumed that the errors in the measured assays follow a multivariate normal distribution with zero means and covariance matrix Σ. The specification of the model in this case reduces to a specification of the elements of the covariance matrix Σ, or equivalently the error variance of each individual assay and the covariance between each pair of assays. This information can be obtained by the taking of replicate samples from each stream. Often the model can be simplified if the relative standard error is allowed to be constant for a given species. However, it should be appreciated that the covariance terms are not necessarily zero; if two metal species have the same mineralogical origin, one would expect their sampling errors at the same sampling point to be positively correlated. Similarly, if size gradings are used, the errors in different size classes at the same sample point will be negatively correlated.

Once the sampling-error model has been derived, the error distribution for the estimated solids recoveries[4], $\boldsymbol{\theta}$, can be calculated. Factors that are taken into account include the structure of the network and actual values of the assays. This latter point is illustrated by the fact that a species that has the same assay value in all streams gives no information at all about recoveries. In general, the more markedly a species concentrates around a node, the more information it gives about recoveries in that node.

Implementation and Applications

The user must first contribute a loss function and an error model as described above. In some instances it may be possible for known information about $\boldsymbol{\theta}$ to be specified before any samples are taken. One obvious assumption might be that the recoveries must lie between zero and unity. In Bayesian[5,6] analysis, this information is expressed in the form of a prior distribution for $\boldsymbol{\theta}$. The quality of the inputs, of course, will determine the quality of the outputs. The net output, after the sample data, $\mathbf{x}$, have been taken into account, is a posterior distribution for $\boldsymbol{\theta}$, written $f(\boldsymbol{\theta} \mid \mathbf{x})$. This distribution summarizes available information about the unknown 'true' value of $\boldsymbol{\theta}$.

Whatever loss function is specified, the Bayesian estimator for $\boldsymbol{\theta}$ is that value $\hat{\boldsymbol{\theta}}$ which minimizes the posterior expected loss; the minimum thus obtained is the *Bayesian risk* for the situation, and will depend on the sampling scheme and the observed sample $\mathbf{x}$. The Value of Sample Information (VSI) is defined as the reduction in Bayesian risk achieved by the incorporation of the sample information. In other words, it is the difference between prior Bayesian risk and posterior Bayesian risk. Thus, VSI compares two states of information: before the sample is taken, and after. VSI depends on the actual values of the data. In an otherwise identical situation, one sample may yield more information than another. The Expected VSI (EVSI) is the expected value of VSI over a prior distribution for the sample results, $\mathbf{x}$. EVSI is thus more appropriate in a comparison of the average performance of different sampling schemes.

The following are some useful applications of the above procedure.

- The planning of a sampling scheme for a plant at the design state. Some of the inputs would have to come from knowledge gained from previous experience, or from short-term research projects on plants of similar design.

- The modification of sampling schemes on existing plants. EVSI will help one to decide, for example, whether to introduce new sampling points or even drop all assays of a certain species on economic grounds. The same statistical framework could also be adapted for an evaluation of the advantages of direct estimation of mass flowrates by the use of magnetic flowmeters and nuclear density gauges.

Other applications that merit consideration include the detection of outlying assays and the designing of tests for bias at a particular sample point.

Illustrative Example

This example (Fig. 1), which is adapted from Smith and Ichiyen[2], serves to illustrate some of the concepts described in the previous sections. Fig. 1 shows the network under consideration. This network comprises a zinc rougher-flotation cell in closed loop with a cleaner cell. Assays are taken on all the arcs (numbered from 1 to 6), and in each arc the percentages of zinc, copper, and iron are determined. This circuit is parameterized in terms of $\boldsymbol{\theta} = (\theta_1, \theta_2)$, where θ_1 is the solids recovery to the concentrate for the rougher and θ_2 is the corresponding recovery for the cleaner. The data used are summarized below:

Stream	*Cu, %*	*Zn, %*	*Fe, %*
1	0,409	9,31	13,47
2	0,462	11,73	13,56
3	0,406	8,72	13,45
4	0,964	38,85	14,48
5	1,009	37,01	14,47
6	0,649	51,70	14,56

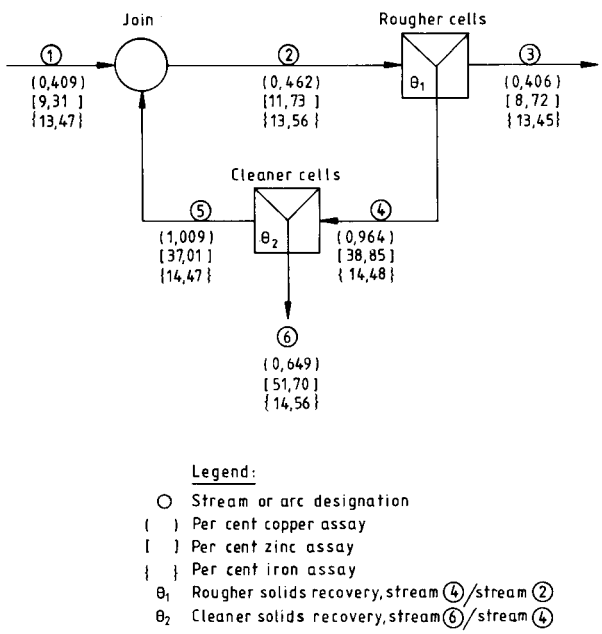

Fig.1—Details of the circuit for the illustrative example

These data have already been smoothed and correspond to the solids recoveries $\boldsymbol{\theta} = (0{,}1 \quad 0{,}125)$. Smoothed data are used because the actual smoothing process is not central to the theme of this paper; rather, the information value of the data is being emphasized. One consequence of the use of smoothed data is that the peak posterior probabilities in Figs. 2, 3, and 7 all occur at the value $\boldsymbol{\theta} = (0{,}1 \quad 0{,}125)$, which makes comparison between all the diagrams somewhat easier.

It is assumed that the economic loss is proportional to the area of the 80 per cent joint confidence region for θ_1 and θ_2. The greater the area of an 80 per cent joint confidence region, the vaguer, and hence the less valuable, is the information about $\boldsymbol{\theta}$. The figure of 80 per cent is merely a convenient choice. In Figs. 2 to 7, several contours of the joint distribution of θ_1 and θ_2 are given to help the reader visualize the surface. The contours are drawn at fractions of the peak height; thus the 0,6 contour connects all points having a probability density equal 60 per cent of the peak probability density. The exact probability content of any particular contour can be determined by numerical integration, but, by analogy with the bivariate normal distribution, the approximate content of the k-contour is $1 - k$. Thus, the approximate content of the 0,2 contour is 80 per cent.

For the example the following assumptions were made.

- No prior information about $\boldsymbol{\theta}$ is assumed.
- Variances of assays were taken as known with a relative standard error (coefficient of variation) of 10 per cent for copper, and 5 per cent for both iron and zinc.
- Except where otherwise stated, no correlations between assay errors were assumed.

From the above information, the posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{x})$ can be calculated, where $\mathbf{x}$ represents the data used. The details of this calculation require a Bayesian approach to multivariate statistics[4,7]. A summary of the procedure for the calculation of $f(\boldsymbol{\theta} \mid \mathbf{x})$ is outlined below.

- For each $\boldsymbol{\theta}$ and each species, write the mass balance equations in the form

  $$B(\boldsymbol{\theta})\mathbf{x} = 0,$$

  where $B(\boldsymbol{\theta})$ is the matrix

  $$\begin{bmatrix} 0 & -1 & 1-\theta_1 & \theta_1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1-\theta_2 & \theta_2 \\ 1-\theta_1(1-\theta_2) & -1 & 0 & 0 & \theta_1(1-\theta_2) & 0 \end{bmatrix}$$

  and $\mathbf{x}$ is the vector:

  $$\mathbf{x}' = (x_1, x_2, x_3, x_4, x_5, x_6)$$

  representing the true assay x_i in each stream i.

- Find an orthonormal basis W (a 6 × 3 matrix) for the feasible space of $\mathbf{x}$, such that

  $$BW = 0$$

  and $W'W = I$, the 3 × 3 identity matrix.

  Although B and W both depend on $\boldsymbol{\theta}$, this fact is suppressed to make the notation clearer.

- If Σ is the covariance matrix for the species, write

$$\mathbf{z} = \Sigma^{-\frac{1}{2}}\mathbf{x}$$

and $U = \Sigma^{-\frac{1}{2}}W$,

where $\Sigma^{-\frac{1}{2}}$ is a symmetric square root of Σ^{-1}. Then, the posterior density can be calculated as

$$f(\boldsymbol{\theta} \mid \mathbf{x}) \propto |U'U|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\mathbf{z}'(I - U(U'U)^{-1}U')\mathbf{z}].$$

This calculation must be done for each species, and the product of the resulting posterior densities formed, on the assumption that the errors for each species are independent of those for other species. For each of Figs. 2 to 7, $f(\boldsymbol{\theta} \mid \mathbf{x})$ was calculated at a grid of points, and approximate contours were drawn in.

Figs. 2, 3, and 7 show the posterior distribution for $\boldsymbol{\theta}$ based only on assays for copper, zinc, and iron respectively. It can be seen that the zinc assays contain far more information about $\boldsymbol{\theta}$ than the copper assays do, while the iron assays are virtually useless. The uselessness of the iron assays arises from the fact that these assays are almost uniform throughout the network.

The fact that zinc assays give more information than copper assays is partly due to the assumption in the error model that the copper assays have a greater relative error, and partly due to the fact that zinc separates more selectively in the network.

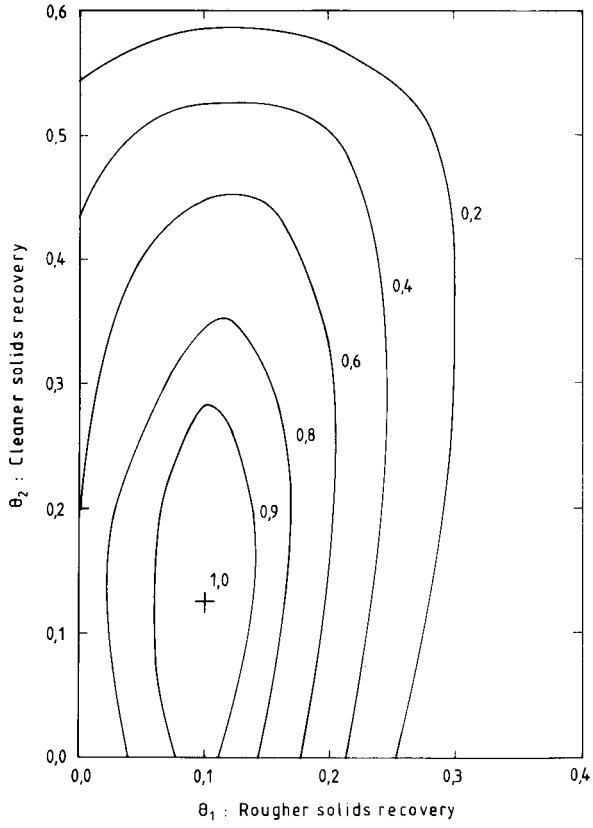

Fig. 2—Contours of posterior distribution for solids recoveries when only copper assays are used

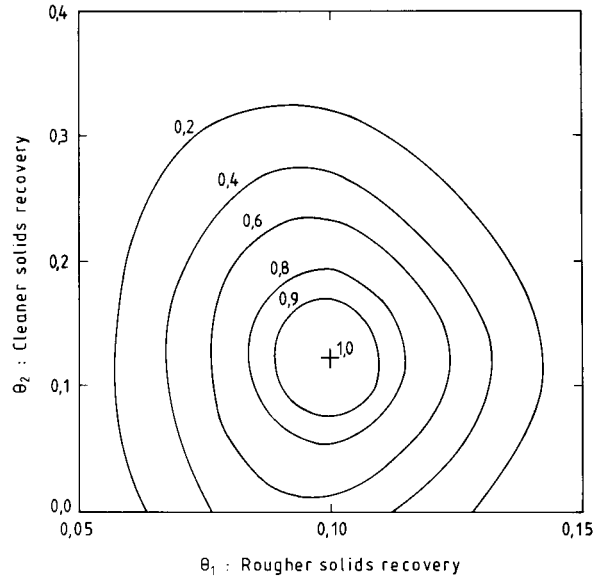

Fig. 3—Contours of posterior distribution for solids recoveries when only zinc assays are used

It is important to notice that, while the error model allows for small relative errors in the assays, the relative standard errors for the solids recoveries (especially θ_2) are relatively large. Thus, there is more uncertainty in the estimated recoveries than in the original assays. Another way of looking at this is that small adjustments to the assays do not imply that the corresponding recoveries are close to the 'true' values.

Fig. 4 arises from the combination of copper and zinc assays. When Figs. 3 and 4 are compared, it can be seen that not much extra information (VSI) about $\boldsymbol{\theta}$ is provided by the inclusion of copper assays when zinc assays are already available.

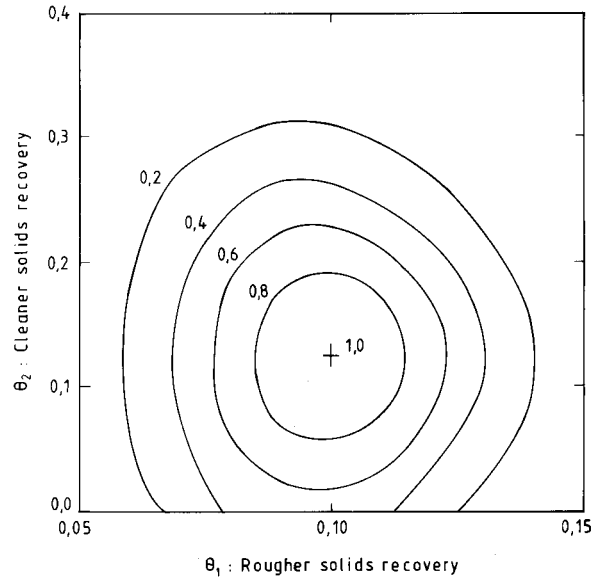

Fig. 4—Contours of posterior distribution for solids recoveries when copper and zinc assays are used and no error correlation is assumed

Fig. 5 gives an analysis for zinc assays only, with no sample taken from arc 3 (rougher tailings). When Figs. 3 and 5 are compared, it can be seen that, by the omission of sample 3, the information is decreased substantially (and hence the expected loss is increased). Interestingly, because the contours are extended mostly in the direction $\theta_1 = \theta_2$, the information about $\theta_1 + \theta_2$ is affected more drastically than is the information about $\theta_1 - \theta_2$.

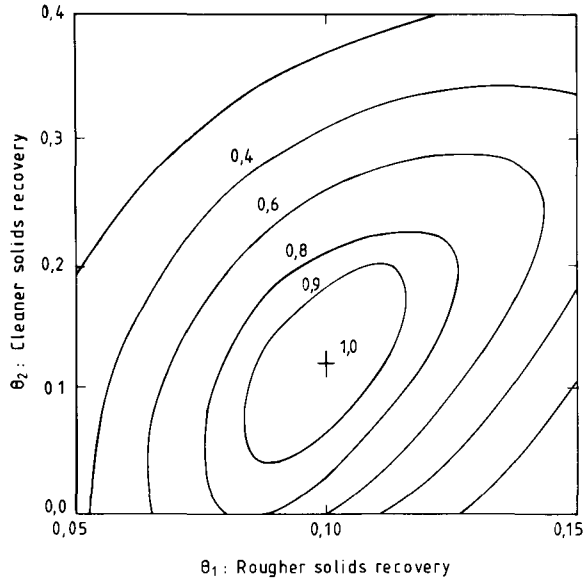

Fig. 5—Contours of posterior distribution for solids recoveries when only zinc assays are used and no sample is taken from stream 3

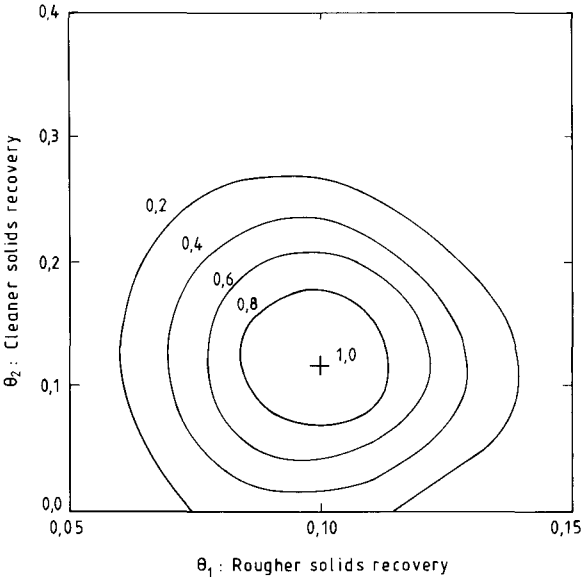

Fig. 6—Contours of posterior distribution for solids recoveries when zinc and copper assays are used and an error correlation of 0,60 is assumed

Figure 6 shows the effect of a change in the error model to allow 0,6 correlation between the error in the copper and zinc assays at the same sample point. Figs. 4 and 6 show that in this case only a small difference in the VSI occurred. Thus, although correlations should be included in the model if they are known to exist, their exclusion may not necessarily change any conclusions from the analysis.

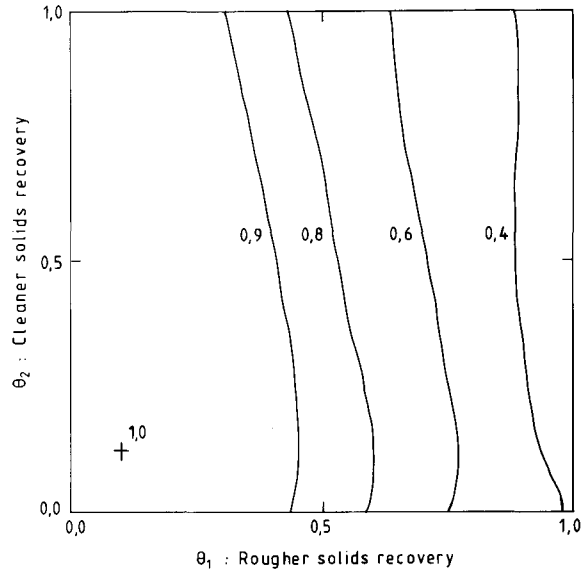

Fig. 7—Contours of posterior distribution for solids recoveries when only iron assays are used

Concluding Remarks

An analytical framework has been proposed for an approach to a wide range of questions related to sampling schemes and the estimation of a materials balance for a plant. This framework makes use of Bayesian statistical theory, which is particularly useful in the analysis of plant economics and the use of prior information. In practice, however, care is needed in the specification of the loss function required by the theory. Plant personnel have important contributions to make, as do statisticians.

Acknowledgements

The authors acknowledge the permission of the Chamber of Mines of South Africa to publish this paper. They also thank Dr L.P. Fatti for many valuable contributions.

References

1. Reid, K.J., Smith. K.A., Voller, V.R., and Cross, M.A. Survey of material balance computer packages in the mineral processing industry. *Proceedings of the 17th APCOM Symposium.* New York, Society of Mining Engineering of AIME, 1982.
2. Smith, H.W., and Ichiyen, N. Computer adjustment of metallurgical balances. *CIM Bull.*, vol. 66. 1973. pp. 97–100.
3. Gy, P.M. The sampling of particulate materials—a general theory. *Int. J. Miner. Process.*, vol. 3. 1976. pp. 289–312.
4. Hunt, C.A. PhD. thesis (in preparation).
5. Berger, J.O. *Statistical decision theory: Foundation concepts and methods.* New York, Springer Verlag, 1980.
6. Winkler, R.L. *Introduction to Bayesian inference and decision.* Holt, Rinehart and Winston, 1972.