DOE/ID/12079-86

WEIGHTED LEAST SQUARES ESTIMATES OF THE MAGNETOTELLURIC TRANSFER FUNCTIONS FROM NONSTATIONARY DATA

by

John A. Stodt

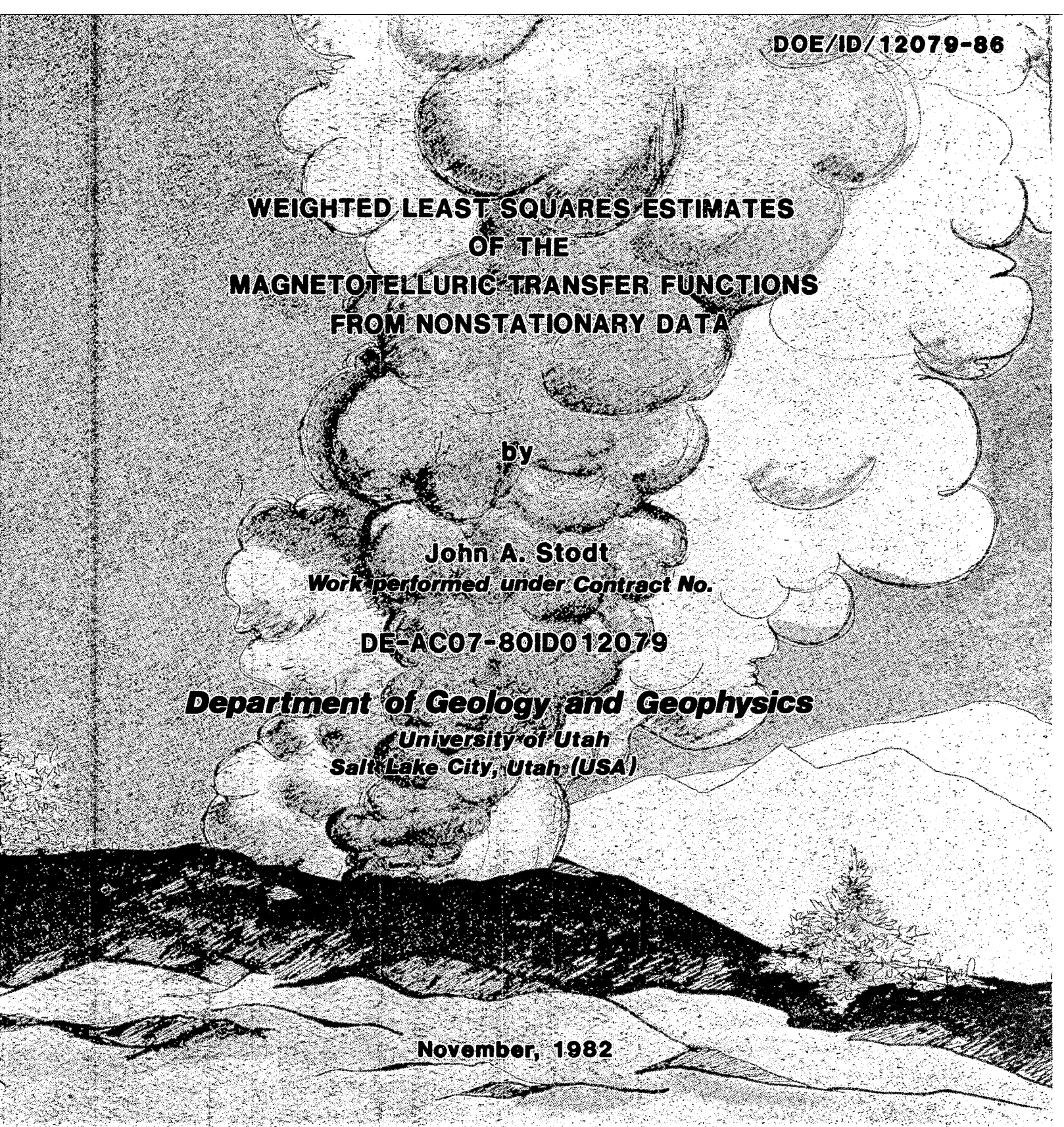
Work performed under Contract No.

DE-AC07-80ID012079

Department of Geology and Geophysics

University of Utah

Salt Lake City, Utah (USA)



November, 1982

Prepared for

DEPARTMENT OF ENERGY

Division of Geothermal Energy

DISCLAIMER

DOE/ID/12079-86

WEIGHTED LEAST SQUARES ESTIMATES OF THE MAGNETOTELLURIC TRANSFER FUNCTIONS FROM NONSTATIONARY DATA

by

John A. Stodt

November 1982

Prepared for the Department of Energy, contract number DE-AC07-80ID12079

NOTICE

This report was prepared to document work sponsored by the United States Government. Neither the United States nor its agent, the United States Department of Energy, nor any Federal employees, nor any of their contractors, subcontractors or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

NOTICE

Reference to a company or product name does not imply approval or recommendation of the product by the University of Utah or the U.S. Department of Energy to the exclusion of others that may be suitable.

TABLE OF CONTENTS

ABSTRACT

Magnetotelluric field measurements can generally be viewed as sums of signal and additive random noise components. The standard unweighted least squares estimates of the impedance and tipper functions which are usually calculated from noisy data are not optimal when the measured fields are nonstationary. The nonstationary behavior of the signals and noises should be exploited by weighting the data appropriately to reduce errors in the estimates of the impedances and tippers.

Insight into the effects of noise on the estimates is gained by careful development of a statistical model, within a linear system framework, which allows for nonstationary behavior of both the signal and noise components of the measured fields. The signal components are, by definition, linearly related to each other by the impedance and tipper functions. It is therefore appropriate to treat them as deterministic parameters, rather than as random variables, when analyzing the effects of noise on the calculated impedances and tippers. From this viewpoint, weighted least squares procedures are developed to reduce the errors in impedances and tippers which are calculated from nonstationary data.

INTRODUCTION

Magnetotelluric (MT) data are obtained as sets of simultaneous measurements of orthogonal electric and magnetic field components at a given site on the earth's surface. The data sets are Fourier transformed and used to calculate complex transfer functions which relate the field components to each other in the frequency domain at the air-earth interface. When the usual assumptions concerning the plane-wave nature of the source fields are satisfied (e.g. Madden and Nelson, 1964; Swift, 1967), the signal components (subscript s) of the measured fields are related to each other in the following manner:

$$E_{si} = Z_{ix}H_{sx} + Z_{iy}H_{sy}, \quad i = x \text{ or } y \tag{1}$$

and

$$H_{sz} = T_{zx}H_{sx} + T_{zy}H_{sy}. \tag{2}$$

The tensor impedances, Z_{ix} and Z_{iy}, and the tipper functions, T_{zx} and T_{zy}, are functions of frequency and conductivity structure. The goal of MT is to deduce the conductivity structure of the earth from the frequency dependent behavior of the impedance and tipper functions.

Generally, MT field measurements consist of signal components of variable amplitude, contaminated by noise. Noise can be defined in general terms as any components of the processed field measurements which do not satisfy the plane-wave impedance relationships given by equations (1) and (2). This general definition includes systematic errors in addition to additive random noise components. Systematic

errors are caused by deviations from the assumed model, e.g., errors due to sources which are not plane waves, cultural noise, and analogue or digital processing errors from instrument drift, aliasing, or truncation effects. It is important to distinguish between systematic errors and random noise when developing estimation procedures and error analysis for the impedance and tipper functions. The primary goal of this paper is to suggest processing procedures which may lead to improved impedance and tipper estimates when systematic errors are negligible and the field measurements are contaminated by additive random noise.

Impedances and tippers are usually calculated as unweighted least squares estimates. We distinguish between conventional and remote reference impedance and tipper estimates. Conventional estimates are calculated entirely from field measurements obtained at a single base site (e.g. Sims et al, 1971). Two of the horizontal field measurements are used as references with equations (1) or (2) to compute the estimates. In contrast, remote reference estimates (Goubau, et al, 1978; Gamble et al, 1979a) are computed by introducing two reference fields which are measured at a separate location. This is done to avoid correlations between the noises in the base and reference field measurements which introduce bias errors into the estimates. Such bias errors have long been recognized in conventional estimates (e.g. Swift, 1967). Attempts to remove them (Kao and Rankin, 1977; Boehl et al, 1977; Goubau et al, 1978), or to avoid them by using more sophisticated estimation procedures (Jupp, 1978), have met with limited success.

In view of the nonstationary behavior of MT data, it seems

reasonable to attempt to improve the estimates by introducing weighting in the least squares estimation procedure. The problem is to develop quantitative weighting procedures which can be applied automatically to produce estimates of equal or superior quality when compared to the unweighted estimates.

Preliminary sections of the paper provide the background to meet this objective. First, a general linear system representation for the MT process is developed which includes additive random noise in all field components. Then a statistical model of the signal and noise components is developed which is consistent with the viewpoint that signals are related deterministically to each other through the linear system, while the noises are not. Particular attention is paid to the definition of a statistical expectation operator that is appropriate for error analysis of the impedance and tipper estimates. Ordinary and multiple coherence functions, which are associated with the linear system and are used in optimizing the estimation procedure, are defined and their properties are examined within the framework of the statistical model. Then the least squares nature of both conventional and remote reference estimates is examined. Finally, insights gained from these developments are used to formulate weighted averaging procedures which are expected to reduce the effects of noise when either the signals or noises are non-stationary.

LINEAR SYSTEM REPRESENTATION FOR MT

In this section, a general linear model is presented which applies to both the impedance and tipper relations between the signal components of the measured fields. Additionally estimates for the transfer functions of the model are developed.

Equations (1) and (2) can be written in the general form

$$O_{si} = G_{ix}I_{sx} + G_{iy}I_{sy}, \quad i = x, y, \text{ or } z. \tag{3}$$

From the viewpoint of linear system theory, I_{sx} and I_{sy} are input signals which are linearly related to an output signal O_{si} through a dual input, single output linear system with transfer functions G_{ix} and G_{iy}, as shown in Figure 1. With regard to the conventional viewpoint in MT, the inputs are the horizontal components, H_{sx} and H_{sy}, of the magnetic field. The transfer functions G_{ix} and G_{iy} are the tensor impedances when the output is an electric field component, E_{si}, i = x or y. If the output is the induced vertical magnetic field, H_{sz}, then G_{ix} and G_{iy} are the tipper functions. The formal solutions for the transfer functions are

$$G_{ix} = [O_{si1}I_{sy2} - I_{sy1}O_{si2}]/V \quad ,$$

and (4)

$$G_{iy} = [I_{sx1}O_{si2} - O_{si1}I_{sx2}]/V \quad ,$$

where $V = [I_{sx1}I_{sy2} - I_{sy1}I_{sx2}]$ and subscripts 1 and 2 identify Fourier transforms from two different sets of input-output data. The existence of the equations (4) requires that $I_{sx1}/I_{sx2} \neq I_{sy1}/I_{sy2}$, i.e., the inputs from the two data sets must not be linearly dependent. Physically, this condition requires a change in polarization of the inputs.

Our goal is to estimate G_{ix} and G_{iy} as accurately as possible from input and output measurements containing additive random noises which are assumed not to be processed by the linear system, as shown in Figure 2. In this paper, the noise components are designated by lower case letters, while the measured fields are designated by upper case letters without the subscript s. The input and output measurements, I_x, I_y, and O_i, are sums of signal and noise, e.g. $I_x = I_{sx} + i_x$.

Averaging over a number of independent data sets is required to suppress random errors caused by omnipresent noise in the measured input and output data. Commonly, G_{ix} and G_{iy} are calculated as the least squares estimates (subscript m),

$$G_{mij} = W_{mij}/V_m \quad , \tag{5}$$

where

$$W_{mix} = [\overline{O_iA^*}\ \overline{I_yB^*} - \overline{I_yA^*}\ \overline{O_iB^*}] \ ,$$

$$W_{miy} = [\overline{I_xA^*}\ \overline{O_iB^*} - \overline{O_iA^*}\ \overline{I_xB^*}] \ ,$$

and

$$V_m = [\overline{I_xA^*}\ \overline{I_yB^*} - \overline{I_yA^*}\ \overline{I_xB^*}] \ .$$

In equation (5), A and B are the reference fields, whose signal components, A_s and B_s, are linearly related to the input signals, I_{sx} and I_{sy}, i.e., $[A_s\ B_s]^T = \underline{M}[I_{sx}\ I_{sy}]^T$ where $\underline{M}$ is a non-singular transfer matrix. The bar represents averaging, e.g. $\overline{O_i A^*} = N^{-1} \sum_{k=1}^{N} O_{ik} A_k^*$, where the individual products, $O_{ik} A_k^*$, are cross periodograms (e.g. Oppenheim and Schafer, 1975) at a given harmonic, and * indicates complex conjugate. In practice, the averaging can be carried out over a number of adjacent harmonics within a narrow bandwidth of an individual periodogram, or over a number of periodograms. We refer to either type of averaging loosely as "averaging over N harmonics" in subsequent developments. Hereafter the subscripts ij in G_{ij}, G_{mij} and W_{mij} are suppressed except where required to distinguish between estimates.

STATISTICAL DESCRIPTION OF THE MEASURED FIELDS

The analysis of the effects of additive noise in the field measurements on the estimates of the transfer functions is facilitated by first developing a clear statistical model of the measured fields. Both the signal and noise components are described statistically by their respective joint probability distributions. Note however, that with regard to estimating the transfer functions G_{ix} and G_{iy} it is appropriate to treat the signal components of each data set as parameters, rather than as random variables which can introduce errors into the estimates. Treating the signals as parameters accounts for the fact that they are related deterministically through G_{ix} and G_{iy}. The signal components can be treated as parameters while treating the noise components as random variables by introducing the concepts of marginal and conditional probability distributions (e.g. Hamilton, 1964, p. 17), and of statistical expectation with respect to the conditional distribution.

The signal and noise components of the measured fields can be viewed as two sets of statistically independent measurements from two different random processes. The two processes are described by their respective joint probability distributions and are assumed to be zero mean, but are otherwise unspecified. In particular, neither process is assumed to be stationary, i.e., the probability distributions which describe the processes can be time dependent. We are interested in

the joint distribution of discrete Fourier transforms of sequences of sampled signal and noise components at a given harmonic. Since Fourier transformation is a linear operation, we have

$$U_s(t) + u(t) \xleftrightarrow{F.T.} Z_s(w) + z(w)$$

where $U = U_s + u$ is any field measurement and $Z = Z_s + z$ is its transform. Since U_s and u are statistically independent with zero means, Z_s and z, which are obtained from U_s and u by linear processing, exhibit these properties as well (e.g. Miller, 1974). Thus, individual harmonics of the discrete Fourier transforms of the sampled field measurements are random variables of zero mean which are sums of statistically independent signals and noises. Let $\underline{Z}_s = \underline{X}_s + j\underline{Y}_s$ be an n-dimensional complex vector of Fourier transformed signal components, and let $\underline{z} = \underline{x} + j\underline{y}$ be the corresponding vector of noise components. Then, since the signals and noises are statistically independent, the joint 4n-dimensional distribution of real and imaginary parts is synthesized from the product

$$h(\underline{X}_s, \underline{Y}_s, \underline{x}, \underline{y}) = f(\underline{X}_s, \underline{Y}_s)g(\underline{x}, \underline{y}) \ ,$$

where $f(\underline{X}_s, \underline{Y}_s)$ and $g(\underline{x}, \underline{y})$ are, respectively, the joint 2n-dimensional distributions of the signal components and the noise components.

With appropriate processing, and assuming plane wave signals, errors in the G_m exist only when noise components are present in the measured fields. Therefore, we treat the signal components as parameters rather than as random variables and the errors in the G_m are governed by the conditional distribution of the noises only,

rather than by the joint distribution of the signals and noises. The conditional density function of the noises is defined as

$$h_c(\underline{x}, \underline{y} \mid \underline{X}_s, \underline{Y}_s) = h(\underline{X}_s, \underline{Y}_s, \underline{x}, \underline{y})/h_m(\underline{X}_s, \underline{Y}_s) ,$$

where $h_m(\underline{X}_s, \underline{Y}_s)$ is the marginal density function defined by

$$h_m(\underline{X}_s, \underline{Y}_s) = \int_{\underline{x},\underline{y}}^{\infty} h(\underline{X}_s, \underline{Y}_s, \underline{x}, \underline{y})\, d\underline{x}d\underline{y} = f(\underline{X}_s, \underline{Y}_s) .$$

The conditional distribution of the noises is the joint probability density function of the 2n noise components, given that the 2n signal components have certain fixed values. On the other hand, the marginal distribution of the signals is obtained by integrating over the range of the 2n noise variables, and represents the probability that the signals lie in given ranges irrespective of the values of the noises. In our case, we have $h_c(\underline{x}, \underline{y} \mid \underline{X}_s, \underline{Y}_s) = g(\underline{x}, \underline{y}) = h_m(\underline{x}, \underline{y})$; the conditional and marginal densities of the noises are equal because the signals and noises are statistically independent.

The statistical expectation operator is used to calculate bias errors and variances of the G_m and it is important to recognize that statistical expectation is defined with respect to an underlying probability distribution. Suppose $q(\underline{Z})$ is some function of the measured fields, $\underline{Z} = \underline{Z}_s + \underline{z}$. If the signal components are treated as parameters, then the expected value $E[q(\underline{Z})]$ is calculated from the conditional distribution, $h_c(\underline{x}, \underline{y} \mid \underline{X}_s, \underline{Y}_s) = g(\underline{x}, \underline{y})$, rather than the joint distribution, $h(\underline{X}_s, \underline{Y}_s, \underline{x}, \underline{y})$.

To illustrate the difference in the two viewpoints, we calculate the expected value and variance of a single field component, $Z = Z_s + z$, treating the signal first as a random variable and then as a

parameter. If both Z_s and z are treated as random variables, then there results (e.g. Miller, 1974)

$$E[Z] = \int_{X_s,Y_s,x,y}^{\infty} Z\, h(X_s,Y_s,x,y)\, dX_s dY_s dxdy = 0. \tag{6}$$

Alternatively, we find

$$E[Z] = \int_{x,y}^{\infty} Z\, g(x,y)\, dxdy = Z_s \tag{7}$$

when the signals are treated as parameters. Continuing, the variance of a complex random variable is defined by (e.g. Miller, 1974)

$$\text{Var } Z = E[|Z|^2] - |E[Z]|^2. \tag{8}$$

Treating Z_s as a random variable, then in view of equation (6), Var Z $= E[|Z|^2]$, where

$$E[|Z|^2] = E[X_s^2 + 2X_s x + x^2 + Y_s^2 + 2Y_s y + y^2].$$

The quantities $E[X_s x]$ and $E[Y_s y]$ are zero because the signals and noises are statistically independent so that, e.g., $E[X_s x] = E[X_s]$ $E[x]$. Thus, we obtain

$$\text{Var } Z = E[|Z_s|^2] + E[|z|^2]. \tag{9}$$

Alternatively, treating Z_s as a parameter, then in view of equation (7), Var Z $= E[|Z|^2] - |Z_s|^2$, where

$$E[|Z|^2] = |Z_s|^2 + E[|z|^2].$$

Now we obtain

$$\text{Var } Z = E[|z|^2]. \tag{10}$$

We shall see in a later section that equations (7) and (10) lead to error analysis which is consistent with the hypothesis that signal components are related deterministically through the transfer functions G_{ix} and G_{iy}. Equations (6) and (9) do not.

COHERENCE FUNCTIONS

Ordinary and multiple coherence functions can be defined (e.g. Bendat and Piersol, 1971) which provide measures of the linear relationships between the field components associated with the linear system in Figures 1 and 2 and are thus useful in noise analysis. In this section we define ordinary and multiple coherence functions within the framework of the statistical model presented in the previous section. The properties of the coherence functions and their estimates must be understood in order that they may be used intelligently in subsequent developments. Therefore, we digress to examine them.

The ordinary coherence γ^2_{AB} is a measure of the correlation between two field components in the frequency domain, and its estimate, C^2_{AB} , is

$$C^2_{AB} = \frac{\overline{|AB^*|^2}}{P_A P_B} , \qquad (11)$$

where the quantities $P_A = \overline{AA^*}$; $P_B = \overline{BB^*}$, and $\overline{AB^*}$ are auto and cross spectral density estimates calculated from discrete Fourier transforms of the two field components. The multiple coherence $\gamma^2_{I_x I_y - O_i}$ measures the frequency domain correlation of the measured output to the measured inputs of the linear system in Figure 2 and is estimated from

$$C^2_{I_xI_y\text{-}O_i} = \frac{|\overline{I_xO_i^*}|^2 P_{I_y} + |\overline{I_yO_i^*}|^2 P_{I_x} - 2\mathrm{Re}[\overline{I_x^*O_i}\ \overline{I_yO_i^*}\ \overline{I_xI_y^*}]}{P_{O_i} P_{I_x} P_{I_y} [1 - C^2_{I_xI_y}]} . \qquad (12)$$

Theoretical coherencies are defined for stationary random processes of zero mean by replacing the spectral density estimates in equations (11) and (12) with their theoretical values, which are obtained using expectation defined with respect to the joint distribution of the signals and noises. However, different definitions are required under the statistical model of the previous section. We define the theoretical ordinary coherence between two field measurements $A = A_s + a$ and $B = B_s + b$ as

$$\gamma^2_{AB} = \frac{E[\overline{AB^*}]}{E[P_A]E[P_B]} = \frac{\overline{A_sB_s^*} + E[\overline{ab^*}]}{(P_{A_s} + E[P_a])(P_{B_s} + E[P_b])} ,$$

where expectation is taken with respect to the conditional distribution of the noises, a and b. The theoretical multiple coherence is defined similarly. Both theoretical coherencies satisfy the inequality $0 \leq \gamma^2 \leq 1$, where the value 1 signifies perfect linear correlation at a given frequency. Note that $\gamma^2_{I_{sx}I_{sy}\text{-}O_{si}}$ is unity by definition. In practice, the multiple coherence between the signals could be less than unity if the transfer functions G_{ix} and G_{iy} varied significantly over the bandwidth of the calculated signal spectra (Foster and Guinzy, 1967). However, we have already assumed implicitly that such variations are insignificant so that G_{ix} and G_{iy} can be calculated accurately from noise-free data.

The coherence estimates given by equations (11) and (12) also range in value from zero to one but they contain random and bias errors which are decreased as the number, N, of harmonics averaged to

obtain the spectral estimates increases. Useful discussions of the errors in ordinary coherencies estimated from stationary Gaussian random processes are provided by Benignus (1969) and Carter et al (1973). Foster and Guinzy (1967) provide experimental evidence which suggests that the assumption of normality is not critical in the error analysis. For our purposes, the behavior of the bias error is particularly important. In general, $E[C_{AB}^2]$ is greater than γ_{AB}^2 when γ_{AB}^2 is less than unity. Carter et al (1973) show that an upper bound on the bias error is given by $(1 - \gamma_{AB}^2)/N$. It is important to understand the behavior of this bias error since it influences our interpretation of the coherence estimates and limits our subsequent use of them. It is well known that when N is unity, the bias error is $(1 - \gamma_{AB}^2)$, and C_{AB}^2 is unity regardless of the true relationship between the processes. This bias error behaves as $O(N^{-1})$ as N increases and it exists because C_{AB}^2 is a nonlinear function of A and B. Thus, although the results are derived for coherencies calculated from stationary random processes, this general behavior is also to be expected when A and B are nonstationary. The behavior of the estimates of the multiple coherence is similar to that of the estimates of ordinary coherence (e.g. Bendat and Piersol, 1971).

The relationship of coherencies between the signal components to coherencies between the measured fields is of interest. For example, if the noises a and b are independent, then for large N the relationship between C_{AB}^2 and $C_{A_sB_s}^2$ is given by

$$C_{AB}^2 = \frac{C_{A_sB_s}^2}{[1 + \frac{P_a}{P_{A_s}}][1 + \frac{P_b}{P_{B_s}}]} ,$$

so that $C^2_{AB} \leq C^2_{A_sB_s}$. Note that the inequality becomes larger as the ratio of noise power to signal power in A or B increases.

The multiple coherence estimate also exhibits this behavior, so that it is less than unity when I_x, I_y, or O_i are noisy. If noise exists only in the output measurement, O_i, the equality $C^2_{I_{sx}I_{sy}-O_i} = P_{O_{si}}/P_{O_i}$ is valid for large N. Then asymptotically unbiased estimates of the output noise power and of the ratio of output noise power to output signal power are obtained from the expressions

$$P_{o_i} = P_{O_i}[1 - C^2_{I_{sx}I_{sy}-O_i}] \quad , \tag{13}$$

and

$$\frac{P_{o_i}}{P_{O_{si}}} = \frac{[1 - C^2_{I_{sx}I_{sy}-O_i}]}{C^2_{I_{sx}I_{sy}-O_i}} \, . \tag{14}$$

If the input measurements I_x or I_y are noisy then $C^2_{I_xI_y-O_i}$ is less than $C^2_{I_{sx}I_{sy}-O_i}$, and it is easy to verify from equation (12) that the inequality becomes larger as the ratio of the noise power to the signal power of either input increases. Note that the estimates of output noise power and of noise to signal ratio, given by equations (13) and (14), are inflated by this behavior. These expressions, and their behavior when the inputs are noisy, are exploited subsequently.

LEAST SQUARES NATURE OF THE ESTIMATES

Sims et al (1971) have shown that the estimates G_m satisfy various minimum mean squared error criteria when the references are chosen as any pair of the horizontal field measurements at a base site. Insight into the general estimation problem is gained by developing the estimates within a least squares framework. Suppose N independent harmonics are available to estimate G_{ix} and G_{iy}. Then we can write the system of equations,

$$\begin{bmatrix} I_{x1} & I_{y1} \\ \vdots & \vdots \\ I_{xN} & I_{yN} \end{bmatrix} \begin{bmatrix} G_{ix} \\ G_{iy} \end{bmatrix} = \begin{bmatrix} O_{i1} \\ \vdots \\ O_{iN} \end{bmatrix} .$$

Using the matrix notation, $\underline{I}\,\underline{G} = \underline{O}$, then the solution,

$$\underline{G} = [\underline{I}^{*T}\underline{I}]^{-1}\underline{I}^{*T}\underline{O} , \qquad (15)$$

is the unweighted least squares solution to the system of equations. The solutions given by equation (15) are identical to those given by equation (5) when the references (A,B) are chosen as (I_x, I_y).

Equation (15) can be generalized to allow for other reference field pairs, i.e.

$$\underline{G} = [\underline{R}^{*T}\underline{I}]^{-1}\underline{R}^{*T}\underline{O} , \qquad (16)$$

where $\underline{R}$ is the matrix of N-dimensional column vectors of the reference field measurements, $\underline{R} = [\underline{A}\ \underline{B}]$. The signal components of any

orthogonal pair of reference fields are related linearly to the input signals through a transfer matrix $\underline{M}$ at the earth's surface, i.e., $\underline{R}_S^T = \underline{M}\,\underline{I}_S^T$. If both the inputs and the references are noise-free, we have

$$\underline{G} = [\underline{M}^* \underline{I}_S^{*T} \underline{I}_S]^{-1} \underline{M}^* \underline{I}_S^{*T} \underline{O} = [\underline{I}_S^{*T} \underline{I}_S]^{-1} \underline{I}_S^{*T} \underline{O},$$

and the result is identical to equation (15). However, if in addition to output noise, there is noise in either the input or the reference field measurements, then the solutions from equation (16) are not the same as those from equation (15). We defer treatment of this general case to the next section.

In the remainder of this section, we assume that only the output measurements $\underline{O}$ are noisy. Then the classical theory of linear least squares can be applied to determine solutions for G_{ix} and G_{iy} which are optimal in the sense that they are unbiased and have minimum variance. From the Gauss-Markov theorem (e.g., Miller, 1974), these estimates are given by

$$\underline{G} = [\underline{I}_S^{*T} \underline{V}_O^{-1} \underline{I}_S]^{-1} \underline{I}_S^{*T} \underline{V}_O^{-1} \underline{O} \ , \tag{17}$$

where $\underline{V}_O$ is the covariance matrix of the elements of the vector $\underline{O}$. We assume for simplicity that the individual harmonics O_{ik}, k = 1,2,...,N are statistically independent so that $\underline{V}_O$ is a diagonal matrix, even though certain processing techniques may have been utilized which result in some harmonics being largely, but not totally, independent of each other (Welch, 1967; Wight et al, 1977). The diagonal elements of $\underline{V}_O$ are the variances,

$$\text{Var } O_{ik} = E[|o_{ik}|^2], \tag{18}$$

where $E[|o_{ik}|^2]$ is the noise power in the kth harmonic. The variance is calculated from equation (8), with expectation defined with respect to the conditional distribution of the noises.

When the output noise is nonstationary, approximately optimal solutions can be obtained in practice by partitioning the data, $\underline{I}_s$ and $\underline{O}$, into M subsets, $\underline{I}_{sl}$ and $\underline{O}_l$, $l = 1,2,\ldots,M$ each containing N/M harmonics. The average noise power (and thus the variance) in each data vector $\underline{O}_l$ can then be estimated from equation (13), i.e.

$$(P_{o_i})_l = (P_{O_i}[1 - C^2_{I_{sx}I_{sy}-O_i}])_l \ . \tag{19}$$

In practice, the number of subsets, M, is chosen large enough to take advantage of the nonstationary behavior of the noise, but not so large that bias errors in $C^2_{I_{sx}I_{sy}-O_i}$ invalidate equation (19). The optimal solution given by equation (17) is then calculated approximately as

$$\underline{G} = [\sum_{l=1}^{M} W_{rl}\underline{I}^{*T}_{sl}\underline{I}_{sl}]^{-1} [\sum_{l=1}^{M} W_{rl}\underline{I}^{*T}_{sl}\underline{O}_l], \tag{20}$$

where the W_{rl} are relative weights defined as $W_{rl} = W_l/\overline{W}$, with $\overline{W}$ defined as the average, $M^{-1}\sum_{l=1}^{M} W_l$. The individual weights are the reciprocals of the average variances of the output vectors $\underline{O}_l$, i.e. $W_l = (1/P_{o_i})_l$.

The covariance matrix of the solution vector $\underline{G}$ is also of interest. From least squares theory, the covariance matrix $\underline{V}_G$ of the optimal solution calculated from equation (17) is given by

$$\underline{V}_G = [\underline{I}^{*T}_s \underline{V}^{-1}_o \underline{I}_s]^{-1} \ . \tag{21}$$

Similarly, the covariance matrix of the approximately optimal solution calculated from equation (18) is estimated from

$$\underline{V}_G = \bar{W}^{-1}\left[\sum_{l=1}^{M} W_{rl}\, \underline{I}_{sl}^{*T}\underline{I}_{sl}\right]^{-1} . \qquad (22)$$

where $\bar{W}^{-1}$ estimates the average variance of the output vectors, $\underline{O}_{il}$. When the output noise is stationary, the relative weights are all unity and equations (21) and (22) are equivalent. Then, by calculating $P_{o_i}[\underline{I}_s^{*T}\underline{I}_s]^{-1}$ and normalizing the numerator and denominator of the elements by N^2, it can be shown that the variances have the form

$$\text{Var } G_{mij} = \frac{P_{o_i}/P_{I_{sj}}}{N[1 - C^2_{I_{sx}I_{sy}}]} , \qquad (23)$$

where $C^2_{I_{sx}I_{sy}}$ is the sample coherence between the inputs I_{sx} and I_{sy}. For purposes of comparison, it is straightforward to show that the variance expression derived by Gamble et al (1979b) reduces to equation (23) under identical assumptions.

In concluding this section, we note that when the output noise is stationary, the diagonal elements of $\underline{V}_o$ all have the same value, P_{o_i}. Then, regardless of the behavior of the signals, the unweighted least squares estimate is optimal and any differential weighting will actually increase the variance of the estimates. In particular, it is not appropriate to weight by either the output noise to signal ratio, or the total power, $P_{O_i} = P_{O_{si}} + P_{o_i}$.

GENERAL WEIGHTING PROCEDURES

The classical least squares theory presented in the previous section is not optimal when noise is present in either the input or reference field measurements. In particular, the presence of such noise introduces errors into the coefficient matrix, $[\underline{R}^{*T}\underline{I}]$, in equation (16). In this section we examine the effects of such errors on the estimates. Then weighting procedures are devised which may produce improved estimates when the signal or noise components of the measured fields are nonstationary.

Assuming for the moment that the output measurements are noise-free, then the effect of errors in the coefficient matrix can be examined from the matrix equation $[\underline{A} + \underline{\Delta A}](\underline{G} + \underline{\Delta G}) = \underline{B}$, where $\underline{G} = \underline{A}^{-1}\underline{B}$. The errors $\underline{\Delta G}$ satisfy the sharp bound (e.g. Forsythe and Moler, 1967 p. 23)

$$\frac{||\underline{\Delta G}||}{||\underline{G} + \underline{\Delta G}||} \leq K \frac{||\underline{\Delta A}||}{||\underline{A}||} \tag{24}$$

where $||*||$ is any matrix norm and $K = ||\underline{A}|| \cdot ||\underline{A}^{-1}||$ is the condition number of $\underline{A}$. The bound given by equation (24) is the sharpest bound which can be specified for arbitrary matricies $\underline{A}$ and $\underline{\Delta A}$, since equality is possible for certain choices of $\underline{A}$ and $\underline{\Delta A}$. We desire a bound on the ratio $||\underline{\Delta G}||/||\underline{G}||$. By substituting the inequality $||\underline{G} + \underline{\Delta G}|| \leq ||\underline{G}|| + ||\underline{\Delta G}||$ into equation (24) and rearranging terms, we obtain the bound

$$\frac{||\Delta G||}{||G||} \leq K \frac{||\Delta A||}{||A||} \left[1 - K \frac{||\Delta A||}{||A||}\right]^{-1} , \qquad (25)$$

which is valid provided $K||\Delta A||/||A||$ is less than unity.

If we use the l_2 norm, then the norm of an n-dimensional complex vector X is defined by

$$||X|| = \left[\sum_{i=1}^{n} |X_i|^2\right]^{1/2} ,$$

where the X_i are the components of X, and the norm of an n x n complex matrix C is defined by

$$||C|| = \max_{||X||=1} ||C\,X|| = u_1 ,$$

where u_1 is the largest singular value of C. The singular values are defined as the square roots of the eigenvalues of the Hermitian matrix, $C^{*T}C$. In our case, A and ΔA are 2 x 2 matrices. When n = 2, it can be shown that the singular values of C are given by

$$u_1, u_2 = \frac{S \pm (S^2 - 4D)^{1/2}}{2} , \qquad (26)$$

where S is the trace and D is the determinant of the matrix $C^{*T}C$. Applying these results, we obtain the expression

$$K \frac{||\Delta A||}{||A||} = ||A^{-1}|| \cdot ||\Delta A|| = \frac{\max(u_{\Delta A})}{\min(u_A)} .$$

Note from equation (26) that the largest singular value of ΔA, $\max(u_{\Delta A})$, is related directly to the magnitude of the largest component of ΔA, while the smallest singular value of A approaches zero as the determinant of $A^{*T}A$ approaches zero, i.e. as the linear dependence of the system of equations $A\,G = B$ increases.

An approximate expression for $K||\Delta A||/||A||$ can be derived which yields further insight by making use of the bound (Forsythe and Moler,

1967 p. 4)

$$\max|C_{ij}| \leq ||\underline{C}|| \leq n \max|C_{ij}|, \; i,j = 1,2,\ldots n,$$

where the C_{ij} are the elements of $\underline{C}$. Applying this bound, we obtain the relation

$$K \frac{||\underline{\Delta A}||}{||\underline{A}||} \leq 2K \frac{\max|\Delta a_{ij}|}{\max|a_{ij}|} . \qquad (27)$$

Additionally, it can be shown that the condition number K of the 2 x 2 matrix $\underline{A}$ is given by

$$K = p + (p^2 - 1)^{1/2}, \qquad (28)$$

where p is given by

$$p = \frac{|a_{11}|^2 + |a_{12}|^2 + |a_{21}|^2 + |a_{22}|^2}{2|\text{Det } \underline{A}|} .$$

Equations (25), (27), and (28) together show clearly that the relative errors in the solution vector, $\underline{G} + \underline{\Delta G}$, depend on the noise to signal ratio of the elements of $[\underline{A} + \underline{\Delta A}]$ and the degree of linear dependence of the system $\underline{A}\,\underline{G} = \underline{B}$.

Equations (24) and (25) suggest that $\underline{\Delta G}$ can be reduced by weighting which reduces the noise to signal ratio $||\underline{\Delta A}||/||\underline{A}||$ more than it increases the condition number K. It is likely that such weighting is possible when the signals are nonstationary, even if the noises are stationary. Thus, although no general prescription for a weighting procedure which is optimal in any sense is available, it is clear from equations (24) and (25) that subsets of the data should be weighted to reflect the ratios of noise to signal in the input and reference field measurements, in addition to the absolute noise level

in the output measurements.

Relative weights which satisfy these requirements can be devised by exploiting the behavior of the multiple coherence estimates. We recall from previous discussion that when I_x and I_y are noisy, the inequality $C^2_{I_xI_y\text{-}O_i} \leq C^2_{I_{sx}I_{sy}\text{-}O_i}$ becomes larger as the ratios of noise to signal in the inputs increase. Thus the quantity, $p = (1 - C^2_{I_xI_y\text{-}O_i})P_{O_i}$, calculated for each subset of data, has the property, $E[p] > P_{O_i}$, and reflects both the nonstationary character of the output noise and the changing noise to signal ratios of the inputs. Note also that since the reference signals, A_s and B_s, are linearly related to the input signals, I_{sx} and I_{sy}, the noise to signal ratios of the references can be estimated from $u = (1 - C^2_{I_xI_y\text{-}A})/C^2_{I_xI_y\text{-}A}$ and $v = (1 - C^2_{I_xI_y\text{-}B})/C^2_{I_xI_y\text{-}B}$ when the references, A and B, are different from I_x and I_y. The estimates, u and v, are also inflated as the ratios of noise to signal in I_x and I_y increase. Thus, p, u, and v exhibit the different types of functional behavior we wish to incorporate in a general weighting scheme.

We can combine p, u, and v in various ways to obtain weights with appropriate behavior. For example, one choice of weighting for the lth subset of data which incorporates all three terms and reduces to the least squares weight when the inputs and references are noise free is given by

$$W_l = p_l^{-1}[1 + u_l + v_l]^{-n} , \quad n > 0 . \tag{29}$$

No claim is made that this choice of weighting is optimal in any sense when input or reference field noise introduces noise into the coefficient matrix, $\underline{A} = [\underline{R}^{*T}\underline{I}]$. In particular, the exponent, n,

which leads to estimates with the smallest error must be chosen from experience. The goal is to choose the value of n which produces the maximum reduction in $K||\underline{\Delta A}||/||\underline{A}||$. Note that the weights given by equation (29) eventually become large as the noise in the base fields becomes small, regardless of the noise in the reference fields. This is desirable behavior, since Var G_{ij} is zero when the base fields are noise free, regardless of the reference field noise (Goubau et al, 1978).

The weights given by equation (29) are used in practice to calculate weighted estimates which are similar to equation (20), i.e.,

$$\underline{G} = [\sum_{l=1}^{M} W_{rl} \underline{R}_l^{*T} \underline{I}_l]^{-1} [\sum_{l=1}^{M} W_{rl} \underline{R}_l^{*T} \underline{O}_l], \qquad (30)$$

where the W_{rl} are relative weights defined as for equation (20). Note that the elements of the matrices in equation (30) are obtained easily as weighted averages of the spectral estimates which are computed for each subset of data. In general, three different sets of weighted spectra are required, corresponding to the three different outputs, $O_i = E_x$, E_y, or H_z. Finally, the variances and covariances of the weighted estimates can be calculated easily by substituting these weighted spectra for the unweighted spectra in the expressions for the variances and covariances derived by Gamble et al (1979b).

DISCUSSION AND CONCLUSIONS

In this paper, we have developed a weighted averaging technique for MT data which is to be applied at the spectral level. The estimate given by equation (30) is approximately optimal when only the output measurements are noisy, but no prescription for optimal weighting is available when in addition to output noise, either the inputs or references are noisy. Instead of weighting at the spectral level, we might consider estimating the transfer functions G_{ml}, $l=1,2,\ldots M$ directly from each subset of the data and then computing the weighted average of the M estimates,

$$\bar{G}_m = \sum_{l=1}^{M} a_l G_{ml},$$

where the weights are given by

$$a_l = \frac{1/\text{Var } G_{ml}}{\sum_l 1/\text{Var } G_{ml}}.$$

We require that $\sum_l a_l$ be unity to avoid introducing bias errors into $\bar{G}_m$ from the weighting procedure. When the estimated variances, Var G_{ml}, adequately represent the dispersions of the distributions of the estimates, this weighting procedure is optimal in the sense that Var $\bar{G}_m$ is minimized (e.g., Hamilton, 1964 p. 41). However, variance and bias errors in the individual estimates, G_{ml}, can be severe due to nonlinear propagation of errors. Weighting at the spectral level, although not guaranteed optimal, can lead to estimates

of the transfer functions with smaller errors than $\overline{G}_m$ because nonlinear error propagation is reduced.

The weighting procedure described in the previous section can be applied to either conventional or remote reference data, but the best value for the parameter n may depend on which type of data is being collected. In particular, it may be appropriate to use larger values of n with conventional data in order to provide stronger rejection of data sets with noisy references, since noise in the references generates bias errors in conventional estimates which can be severe. In any case, weighted averaging can be expected to yield improved estimates of the transfer functions only if the output noise is nonstationary, or if the input or reference signal to noise ratios are nonstationary. Finally, since the objective of weighted averaging is to take advantage of any nonstationary behavior in the signals or noises, weighted averaging schemes might be more successful when applied to data processed using decimation techniques (Wight et al, 1977), instead of FFT processing with subsequent averaging of adjacent harmonics, since the higher harmonics of an FFT represent averages of signal and noise over a larger time window than is necessary for adequate spectral resolution.

ACKNOWLEDGEMENTS

The author acknowledges the efforts of Doctors Stanley H. Ward, William R. Sill, and Gerald H. Hohmann who reviewed this manuscript and offered numerous suggestions for improvement. The work was sponsored under Department of Energy contracts DE-AC07-80ID12079 and DE-AC07-78ET28392.

REFERENCES

Bendat, J. S., and Piersol, A. G., 1971, Random data: analysis and measurement procedures: New York, John Wiley & Sons, Inc., 407 p.

Benignus, V. A., 1969, Estimation of the coherence spectrum and its confidence interval using the fast Fourier transform: IEEE Trans. Audio and Electroacoustics, v. AU-17, p. 145-150.

Carter, G. C., Knapp, C. H., and Nuttall, A. H., 1973, Estimation of the magnitude squared coherence function via overlapped fast Fourier transform processing: IEEE Trans. Audio and Electroacoustics, v. AU-21, p. 337-344.

Forsythe, G. E., and Moler, C. B., 1967, Computer solution of linear algebraic systems: Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 148 p.

Foster, M. R., and Guinzy, N. J., 1967, The coefficient of coherence: its estimation and use in geophysical data processing: Geophysics, v. 32, p. 602-616.

Gamble, T. D., Goubau, W. M., and Clarke, J., 1979a, Magnetotellurics with a remote reference: Geophysics, v. 44, p. 53-68.

_____1979b, Error analysis for remote reference magnetotellurics: Geophysics, v. 44, p. 959-968.

Goubau, W. M., Gamble, T. D., and Clarke, J., 1978, Magnetotelluric data analysis: Removal of bias: Geophysics, v. 43, p. 1157-1166.

Hamilton, W. C., 1964, Statistics in physical science - estimation, hypothesis testing, and least squares: New York, N. Y., The Ronald Press Co., 230 p.

Jupp, D. L. B., 1978, Estimation of the magnetotelluric impedance functions: Phys. Earth Planet. Inter., v. 17, p. 75-82.

Kao, D. W., and Rankin, D., 1977, Enhancement of signal to noise ratio in magnetotelluric data: Geophysics, v. 42, p. 103-110.

Madden, T. R., and Nelson, P. H., 1964, A defense of Cagniard's magnetotelluric method: ONR report, project NR-371-401, Geophysics Lab., MIT.

Miller, K. S., 1974, Complex stochastic processes: an introduction to theory and application: Reading, Mass., Addison-Wesley Publishing Co., Inc., 235 p.

Oppenheim, A. V., and Schafer, R. W., 1975, Digital signal processing: Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 585 p.

Sims, W. E., Bostick, F. X., Jr., and Smith, H. W., 1971, The estimation of magnetotelluric impedance tensor elements from measured data: Geophysics, v. 36, p. 938-942.

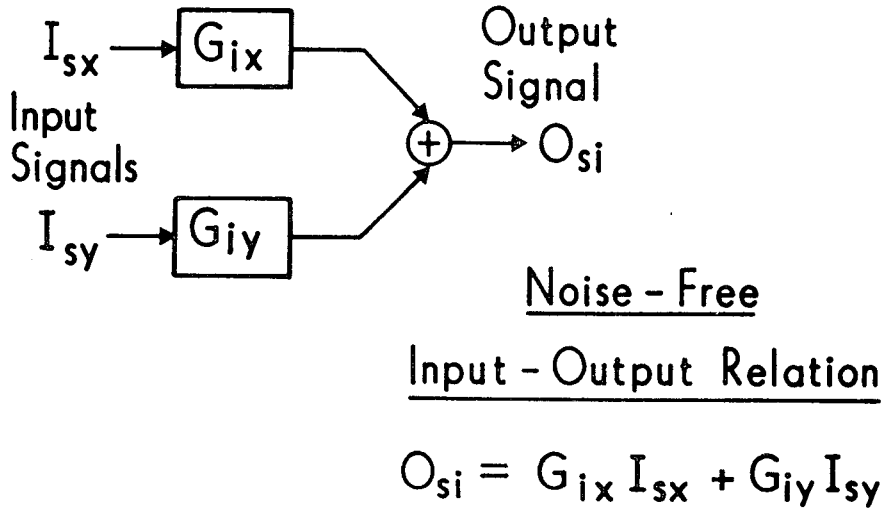
Swift, C. M., Jr., 1967, A magnetotelluric investigation of an electric conductivity anomaly in the Southwestern United States: Ph.D. Thesis, MIT.

Welch, P. D., 1967, The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms: IEEE Trans. Audio and Electroacoust., vol. AU-15, p. 70-73.

Wight, D. F., Bostick, F. X., Jr., and Smith, H. W., 1977, Real time transformation of magnetotelluric data: Tech. rep., Electr. Geophys. Res. Lab., Univ. of Texas, Austin.

LIST OF FIGURES

FIGURE 1

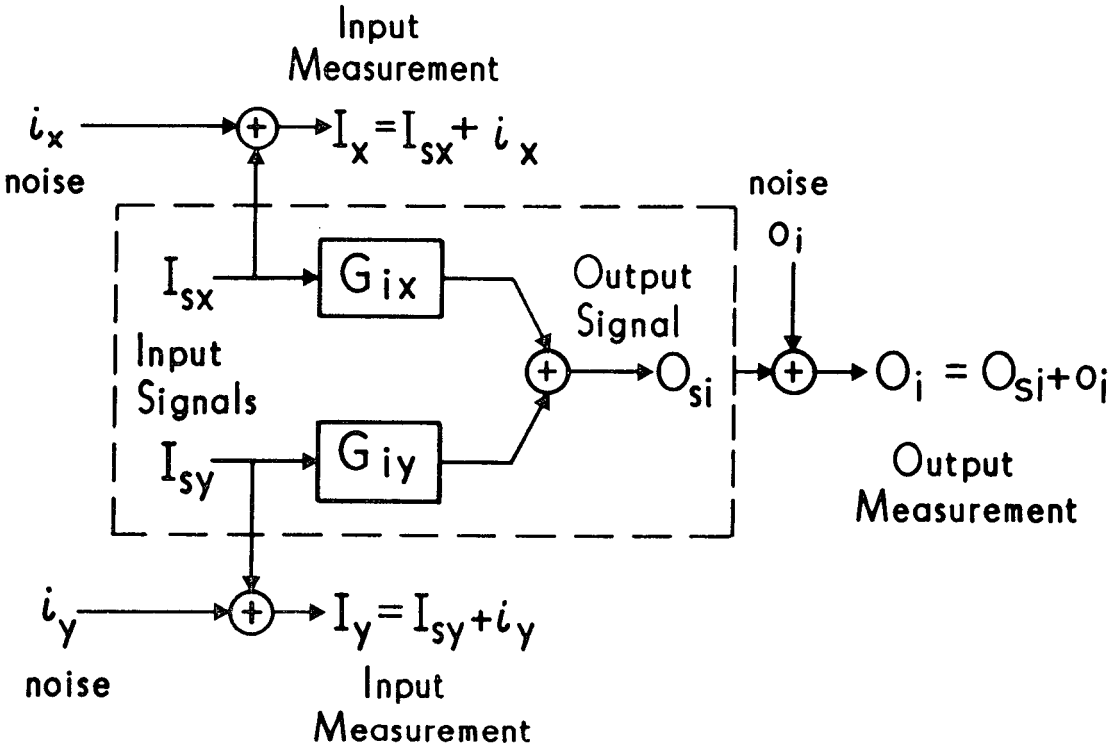


Noise - Free

Input - Output Relation

$$O_{si} = G_{ix} I_{sx} + G_{iy} I_{sy}$$

FIGURE 2


Input
Measurement
i_x
noise
$I_x = I_{sx} + i_x$
noise
o_i
I_{sx}
G_{ix}
Output
Signal
Input
Signals
O_{si}
$O_i = O_{si} + o_i$
I_{sy}
G_{iy}
Output
Measurement
i_y
noise
$I_y = I_{sy} + i_y$
Input
Measurement