Workshop on Objective Analysis
Analysis Strategies for Field-Dependent Estimation

M.A. Pedder, University of Reading, England
H.J. Thiébaux, Dalhousie University, Halifax, Nova Scotia, Canada
P.R. Julian, Editor

ATMOSPHERIC ANALYSIS AND PREDICTION DIVISION
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
BOULDER, COLORADO
Abstract

This Technical Note presents a distillation of the discussions that took place at an informal workshop on "Analysis Strategies for Field-Dependent Estimation" held at NCAR in August 1983 during which participants examined, comparatively, various objective analysis schemes.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>vii</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. SIGNAL AND NOISE CONCEPTS IN METEOROLOGICAL ANALYSIS</td>
<td>5</td>
</tr>
<tr>
<td>M. A. Pedder</td>
<td></td>
</tr>
<tr>
<td>III. IDENTIFYING REQUIREMENTS FOR ANALYSES</td>
<td>31</td>
</tr>
<tr>
<td>IV. DISCUSSION OF ACCURATE FIELD-DEPENDENT ESTIMATION PROCEDURES</td>
<td>36</td>
</tr>
<tr>
<td>V. REVIEW OF PRACTICAL APPLICATIONS AND THEORETICAL STRUCTURE OF STATISTICAL ESTIMATION SCHEMES</td>
<td>40</td>
</tr>
<tr>
<td>H. J. Thiébaux</td>
<td></td>
</tr>
<tr>
<td>VI. NOTES ON FIELD-DEPENDENT STATISTICAL OBJECTIVE ANALYSIS</td>
<td>77</td>
</tr>
<tr>
<td>H. J. Thiébaux</td>
<td></td>
</tr>
</tbody>
</table>
Preface

This Technical Note presents a distillation of the discussions that took place at an informal workshop on "Analysis Strategies for Field-Dependent Estimation" held at NCAR in August 1983. The impetus for the workshop was provided by H. J. Thiébaux, Dalhousie University, and M. A. Pedder, University of Reading. Their original suggestions for content and emphasis were received with interest and enthusiasm, and were scarcely modified before or during the workshop.

Summarizing the workshop was not an easy task because of the informal, unstructured nature of the sessions and the extent of the contribution made to the outcome by discussion among the participants. Contained herein are summaries of the structured presentations made by Jean Thiébaux and Mike Pedder, together with my own synthesis of the discussions. The order in which topics were covered has been preserved by-and-large. However some liberty has been taken with content, principally because subsequent inquiry has been made into analysis objectives in mesoscale modeling.

A word concerning terminology and the workshop subtitle is in order. 'Field-dependent estimation' is the term given to an objective analysis process which specifies the spatially coherent, or bulk-state, component of meteorological variables (at grid points) by making use of locally occurring observations in a data-adaptive sense. The term 'bulk-state' components is covered by M. Pedder's presentation and includes consideration of the goals of the analysis; namely, what the analysis is meant to represent. The term 'field-dependent' intends to
viii

convey the idea that analysis schemes should consider proximal observations as specifying the bulk-state components in as characteristic a fashion as possible. Various objective schemes differ as to the manner in which unique information contained in observations is transferred to analysis locations. Commonly used statistical schemes, which otherwise have many advantages, make use of ensemble statistical information which may fail to accommodate much of the unique information in the observations employed. Nonstatistical schemes can perform the desired transfer, but suffer other shortcomings.

Thus, the Workshop participants examined, comparatively, the various objective analysis schemes, and discussed ways in which the proper match of analysis purpose and analysis design can be achieved.

P. R. Julian

Participants

Edward Barker, Naval Environmental Prediction Research Facility
Richard Franke, Naval Postgraduate School
Paul Julian, National Center for Atmospheric Research
Robert Kistler, National Meteorological Center
Michael Pedder, University of Reading
Gary Rasmussen, PROFS, NOAA
Thomas Schlatter, PROFS, NOAA
H. Jean Thiébaux, Dalhousie University
David Williamson, National Center for Atmospheric Research
There is quite a wide variety of objective techniques, from which one may choose, for converting observations of meteorological variables to an analysis for any specific purpose. It is important to emphasize that before adopting an analysis strategy and selecting a particular scheme, some thought must be given to the goals of the analysis. The analysis should represent a desired signal, according to some set of criteria, which the observed data are presumed to contain in addition to unwanted noise. This concept is implicit in the commonly held view that an objective analysis scheme is essentially a smoothing device. Some published material on objective analysis has considered the smoothing characteristics of particular schemes, but not always has thought been given to identifying the desired signal. The major issue is whether a specific scheme produces the desired signal through the form of the analysis. The desired signal is also called the bulk-state field.

There is a three-way interaction between 1) the goals to be achieved by the analysis scheme, 2) the characteristics of the data, and 3) the details of the response or behavior of the analysis algorithm, that will influence the strategy of erecting the analysis scheme. As a first step, the requirements of the analysis must be identified. Next, the characteristics of the observational data must be understood, and the system of equations comprising the objective analysis analyzed to determine the response of the scheme to the data. The accuracy of the analysis will be relative to the prescribed expectations of the analysis and will depend upon the spatial and temporal errors of the observed
data, and the relative spacings of the analysis grid and the observation mesh.

A comparison of the various objective analysis schemes can be made which considers the foregoing analytical framework. As a prerequisite, it is useful to consider some fundamental properties of the schemes which have been developed and used, either in an operational or in a research context.

All schemes can be grouped into three categories by considering the manner in which the resulting analysis is viewed: viz, deterministic, statistical, and mixed schemes. In deterministic schemes the bulk-state field is not considered a stochastic variable with concomitant uncertainty. All uncertainty 'error' is assumed to be observation error and is specified to occur at the observation locations. Statistical schemes, on the other hand, implicitly or explicitly view the bulk-state field as stochastic and the concept of error is more comprehensive. Mixed schemes, as the name implies, utilize deterministic basis functions to represent the analysis, but with statistical formulation of error or uncertainty. In very simplified terms, deterministic schemes are basis function fitting schemes, and statistical schemes are smoothing, interpolative devices.

Further, statistical schemes may be subdivided into those that treat the observations influencing the local bulk-state variables in a unique fashion and those that make use of aggregate or ensemble information. Perhaps the best way to elaborate this point is to compare a theoretically complete and optimal analysis scheme, viz., Kalman
filtering, with the so-called optimum interpolation scheme. The following brief description is taken from Ghil et al. (1981) and Rasmussen and Schlatter (1983).

Let $X_t$ be an $n$-vector of the descriptive (i.e., bulk-state) variable, and $Y_t$ an $m$-vector of observations of $X_t$. The Kalman system of equations includes: 1) a model, or system, equation describing the evolution of $X_t$,

$$
X_t = \Phi_{t-1} X_{t-1} + W_{t-1}
$$

where $W_t \sim N(0, Q_t)$

The $\Phi$ denotes a known system matrix which specifies changes in the $X$.

2) an analysis equation describes how $X_t$ is estimated from the observations.

$$
Y_t = H_t X_t + V_t
$$

where $V_t \sim N(0, R_t)$

and $H$ denotes a known matrix describing how the observations are related to the dependent variable vector $X$. With the assumption that expected covariance of the error terms is zero: $E(W_t, V_t) = 0$ for all $t$ and $s$, a coupled system of iterative equations may be written as follows.

Model time-step:

$$
\hat{X}_t(-) = \Phi_{t-1} \hat{X}_{t-1}(+) 
\hat{C}_t(-) = \Phi_{t-1} \hat{C}_{t-1}(+) \Phi_{t-1}^t + Q_{t-1}
$$

and

$$
K_t = \hat{C}_t(-) H_t [H_t \hat{C}_t(-) H_t^t + R_t]^{-1} .
$$
Analysis step: \[ \hat{X}_t(+) = \hat{X}_t(-) + K_t [Y_t - H_t \hat{X}_t(-)] \]

\[ \hat{C}(+) = [I_t - K_t H_t] \hat{C}_t(-) \]

In this notation (-) and (+) denote the matrix before and after the addition of the information from the observations, and a circumflex denotes the matrix transpose.

Two interlocking error matrices are carried and updated each time-step: \( K_t \) is the Kalman gain matrix specifying how the observations determine the analysis \( \hat{X}_t(+) \) from the observations and the 'guess' or forecast \( \hat{X}_t(-) \); and \( C_t \) is the system matrix specifying the model (or forecast) error.

The two basic quantities appearing in the expressions for \( K_t \) and \( C_t \) are \( Q_t \), the model error variance; and \( R_t \), the analysis error variance. Optimal Statistical Objective Analysis (O.S.O.A.) or, more commonly, O.I. (optimum interpolation) can be viewed as a simple Kalman scheme wherein the terms appearing in the expression for \( K_t \), viz. \( CH' \) and \( HCH' \), are parameterized once for all rather than being iteratively updated. Thus traditional O.S.O.A. is an 'ensemble' scheme: specifically, model and analysis behavior from a statistical sample are used to determine the parameterization. In this sense it is sub-optimal rather than optimal, but is also considerably simpler than the Kalman technique.
II. SIGNAL AND NOISE CONCEPTS IN METEOROLOGICAL ANALYSIS

M. A. Pedder

In the Introduction we emphasized the importance of identifying the nature of the signal component of an observed field in addressing the problem of choosing a suitable analysis strategy. Before reviewing some alternative analysis strategies, it seems appropriate to examine the concepts of 'signal', the bulk-state variable; and 'noise', the observation error variable, and identify some ambiguities that may arise unless a proper distinction is made between what we shall refer to as the 'state' and 'measurement' components of observation error. For convenience, we examine these concepts in the context of a conventional O.I. analysis strategy, though similar considerations also apply to other types of spatial analysis.

In the context of meteorological analysis, the use of the term "observation error" usually implies the combined effect of 'measurement' (= 'instrument' plus 'sampling') error and small scale variability (or 'atmospheric noise') on the extent to which observations are considered representative of the space-coherent, 'bulk-state' component of observed spatial variation. Gandin (1963) and many workers since have regarded observation errors as spatially uncorrelated random variables, and therefore that the only difference between the space-lagged covariance structure of 'bulk-state' and 'observed' variables is at zero separation, where the zero-lagged covariance of the latter exceeds the former by the variance of the observation error. This assumption leads to considerable simplification in obtaining models for the space-lagged
covariance structure of a synoptic field variable from the estimated second order moment statistics of observations, as required, for example, in an optimal linear interpolation (O.I.) analysis of the type developed by Gandin.

Unfortunately, the above concept of observation error is ambiguous, due to its dependence on 'analysis goals'. We suggest here that the assumption that errors are spatially uncorrelated might not be justified when either (a) an analysis incorporates predicted variables with error variances much smaller than the variance of the bulk-state variables, or (b) the analysis imposes constraints on the bulk-state field structure which are not necessarily satisfied by the true atmospheric structure of the observed field. We imply that analysis strategies for dealing with prediction error data should take into account the possibility that observation errors are often not spatially uncorrelated.

Measurement error

Let $P(x,y,z,t)$ be the atmospheric state value of a variable $P$ at a point $(x,y,z,t)$, and $p_j$ a 'coincident' observed value of $P$ using measuring device $j$.\(^1\) The measurement error on $p_j$ is defined here as

$$\delta p_j = (p_j - P) = s_j + \varepsilon_j$$

\(^1\) The term 'coincident' is here meant to imply that $p_j$ is assumed to be coincident with $P(x,y,z,t)$. If $p_j$ is not actually coincident with $P(x,y,z,t)$, then the effect of using $p_j$ as an estimate of $P(x,y,z,t)$ is to introduce a sampling error which is related to the interpretation of the observation rather than to any device-dependent source of measurement error.
where \( s_j \) and \( \varepsilon_j \) represent systematic (bias) and random error components such that

\[
\begin{align*}
E\{\delta p_j\} &= s_j \quad (2a) \\
E\{\varepsilon_j\} &= 0 \quad (2b) \\
E\{\delta p_j^2\} &= E\{\varepsilon_j^2\} + s_j^2 \quad (2c)
\end{align*}
\]

In (2), \( E\{\cdot\} \) denotes conditional expectation in the sense of averaging over all possible realizations of the state \( P_0 \) as observed using a single 'device'.

The statistics defined by (2) are specific to one device. By considering colocated devices (physically impossible, but conceptually convenient), the concept of ensemble averaging is extended to include an ensemble of devices. In this case, measurement error can be considered to have the statistical properties

\[
\begin{align*}
E\{\delta p\} &= E\{s\} = E\{\varepsilon\} = 0 \quad (3a) \\
E\{\delta^2\} &= \sigma(0) = E\{\varepsilon^2\} + E\{s^2\} \quad (3b)
\end{align*}
\]

In (3), \( E\{x\} \) without a subscript on \( x \), denotes the expectation over all possible realizations of \( P_0 \) and over all devices. Note that (3) implies that, with respect to an ensemble of devices, \( s \) and \( \varepsilon \) have symmetric probability distributions centered on zero means, which is an acceptable simplification in most meteorological applications.
The covariance between measurement errors on colocated devices is given by
\[
E\{(\delta p_j \cdot \delta p_k)\} = 0, \ j \neq k \quad (4a)
\]
\[
= \sigma(0), \ j = k \quad (4b)
\]
and the covariance between measurement errors on separate devices which are not necessarily colocated by
\[
\sigma(r) = E\{(\delta p_j, \delta p_{j,m})\} = 0, \ j \neq k \text{ for all } \ell, m \quad (4c)
\]
where subscripts \(\ell, m\) index location, and \(r\) is the separation between locations \(\ell\) and \(m\).

If we consider pairs of realizations, with device \(j\) being used to measure \(P\) at first one location \((\ell)\), then a second location \((m)\), then
\[
\sigma(r) = E\{(\delta p_j, \delta p_{j,m})\} = s_j^2, \ \ell \neq m, \quad (5a)
\]
\[
= E\{\varepsilon_j^2 + s_j^2\}, \ \ell = m, \quad (5b)
\]
the form of which is represented schematically by the location-lagged covariance model shown in Fig. 1a.

The covariance function defined by (5) is suggestive of the situation common to both satellite retrieved temperatures, and radiosonde retrieved temperature and humidity data. In both cases one instrument supplies observations which are 'nearly' simultaneous with respect to time \(t\), but which are quite distinct with respect to one or more space coordinate locations. However, the location-lagged covariances of measurement errors on derived state variables may be somewhat
different. For example, in deriving geopotential heights (Z) from sonde-retrieved temperature and humidity data, the vertical-lagged correlations in the measurement error on Z depend not only on the errors in temperature and humidity, but also on vertical separation. This results in the type of measurement error covariance function represented schematically in Fig. 1b.

It can be argued that neither Fig. 1a nor Fig. 1b is physically realistic, since for very small separations it is unlikely that any source of measurement error can be regarded as completely random. However, we can reasonably speak of an effective discontinuity in \( \epsilon(r) \) at \( r = 0 \) corresponding to a rapid decrease in \( \epsilon(r) \) by \( E[\epsilon^2] \) over a separation scale which is smaller than that which is actually observed. This provides a convenient method of recognizing the different effects in observation error statistics which result from using the 'same' as distinct from 'separate' observing devices.

**Atmospheric Noise**

The local atmospheric state \( P_z \) can be considered as the response to a spatially distributive process over an effectively infinite continuum of all space and time scales. With respect to the problem of analyzing the instantaneous spatial structure of \( P \), not all these scales can be represented unambiguously via a linear transformation of the discretely observed state, nor is this a desirable goal for most practical purposes. Viewed from the spectral domain, spatial variations in \( P \) associated with spatial frequencies above the resolvable limit tend to be aliased into those which can, in principle, be resolved from the
Fig. 1. Schematic representation of measurement error covariance function \( \sigma(r) \) when all observations are obtained from the same device: (a) No error dependence on lag; (b) error depends on lag.
spatial net of observations. Thus in any one realization of the 'spatial process' responsible for the observed state, an unambiguous representation of both amplitude and phase characteristics of frequency components in P is only possible for spatial wavelength which are much greater than the separation scale characterizing observations.\textsuperscript{2} This problem has been examined quantitatively by Stephens and Stitt (1970) in relation to the frequency response characteristics of analysis by the 'successive corrections' method (Cressman, 1959). These authors noted that any attempt to resolve too small a scale structure through linear interpolation of discrete observations could result in a catastrophic decrease in the 'signal-to-noise' variance ratio in the analysis field, and they concluded that it is probably better to 'oversmooth' rather than 'undersmooth' the observed state when the analysis goal is to represent faithfully those global features of the observed state which can be resolved unambiguously by a randomly distributed network of observation. This implies that the analysis goal should not be to estimate P\textsubscript{\text{g}}, but rather a bulk-state or signal variable \( \tilde{P}_g \), where the latter involves some form of averaging over a finite location domain centered on \((x,y,z,t)_{\text{g}}\). The averaging scale implied by \( \tilde{P}_g \) might be determined by station separation, as considered by Stephens and Stitt.\textsuperscript{2}

\textsuperscript{2} As has been demonstrated by Stephens (1971), the minimum resolvable scale depends on whether or not spatial derivatives of the state as well as the state itself are being observed. Thus 'resolvable scale' is not necessarily a simple function of station separation. However, for univariate analysis schemes applied to a single scalar field variable, the minimum resolvable wavelength is considered to be typically of the order of 5 \times the mean station separation.
However, a larger averaging scale (possibly including time) might be more appropriate if certain scales which could be represented unambiguously in the analyzed state were nevertheless redundant as far as the goals of a diagnostic or prognostic analysis were concerned. The signal variable concept is thus related to analysis goals, and it follows that the statistics of variables related to $\tilde{P}$ are themselves dependent on these analysis goals.

For a prescribed averaging scale, the difference $\delta P_L = (P_L - \bar{P}_L)$ can be considered as a state error variable associated with atmospheric noise, with the properties

$$E[\delta P_L] = 0$$  \hspace{1cm} (6a)

$$E[\delta P_L^2] = \eta(0)$$  \hspace{1cm} (6b)

$$E[\delta P_L \delta P_m ] = \eta(r)$$  \hspace{1cm} (6c)

where $E[x]$ again implies ensemble averaging over all possible realizations of the local state. (6b) and (6c) thus define the variance and separation-lagged covariance function of the state error variable $\delta P$.  

Now consider some unbiased predictor $P^f_L$ of $\tilde{P}_L$, such that

$P^f_L$ could be the local output value of P as estimated by a numerical (deterministic) prediction scheme, an estimate of a climatological mean for P, or an estimate on time or space filtering of observations. The source of the predictor $P^f$ is of no concern here, provided that $P^f_L$ satisfies (7a).
(7b) and (7c) define the variance and lagged covariance function of a signal prediction-error or increment variable $\tilde{P}'$ which we regard as containing the 'signal component' of a corresponding state increment variable $P' = (P - P^f)$. The statistics of the state increment variable are defined by

\[
E\{P'_\ell\} = 0 \quad (8a)
\]

\[
E\{P'_\ell^2\} = C(0) \quad (8b)
\]

\[
E\{P'_\ell P'_m\} = C(r) \quad (8c)
\]

(8a) following from (6a) and (7a).

Now by definition

\[
P' = P - P^f = (\tilde{P} - P^f) + (P - \tilde{P}) = \tilde{P}' + \delta P \quad (9a)
\]

and in the above sense of $E\{X\}$,

\[
e\{P^f_\ell \delta P_m\} = e\{\tilde{P}_\ell \delta P_m\} = 0 \quad \text{for all } \ell, m \quad (9b)
\]
It follows that by substituting (9a) into (8a), the covariance functions defined by (8c), (6c) and (7c) are related linearly by

\[ C(r) = \tilde{C}(r) + \eta(r) \quad (10) \]

This relationship is represented schematically in Fig. 2. Note that as suggested by this graphical representation, the noise process associated with the state error variable \( \delta P \) is not regarded as purely random in the sense of having zero covariance values at all \( r > 0 \). This is because it has been admitted that \( \delta P \) is associated with the effect of small but finite length scale structure in \( P \).

Although the precise nature of the predictor \( P^f \) is irrelevant in the above analysis, it is worth noting here that for a fixed averaging scale on \( \tilde{P} \), \( C(0) \) and \( C(r) \) will both tend to decrease in magnitude as the accuracy of the predictor \( P^f \) increases. However, \( \eta(r) \), not being a function of \( P^f \), remains constant. Thus \( C(r) \) also tends to decrease in magnitude as prediction accuracy improves, until in the limit it approaches \( \eta(r) \). We later demonstrate that this property is important in relation to optimal interpolation of the signal increment field using linearly weighted combinations of observed prediction error variables.

**Observation Error**

Within the context of a prescribed analysis goal, the total error in an observation \( p \) is the sum of measurement and state errors. Thus

\[ \delta p^0 = \delta p + \delta P = p - \tilde{P} \quad (11) \]
Fig. 2. Schematic representation of the relationship between the covariance functions for state error ($\eta(r)$), signal increment ($C(r)$), and state increment ($\tilde{C}(r)$).
where $\delta P^0$ is the observation error variable. If $\delta P$ and $\delta p$ are considered as independent random variables with first and second order moments independent of location, then the statistical properties of $\delta P^0$ can be expressed by

\begin{align}
E\{\delta P^0\} &= 0 \quad (12a) \\
E\{\delta P^0 \delta P^0\} &= \eta^0(0) = \eta(0) + \sigma(0) \quad (12b) \\
E\{\delta P^0 \delta P^0_m\} &= \eta^0(r) = \eta(r) + \sigma(r) \quad (12c)
\end{align}

where $\eta^0(0)$ and $\eta^0(r)$ are the variance and lagged covariance function for the observation error.

A schematic representation of $\eta^0(r)$ which results from adding a state error covariance (curve (a) in Fig. 2) to the type of measurement error covariance shown in Fig. 1b, is given in Fig. 3. This is the type of observation error covariance structure which could characterize observations using a single instrument at different locations when these observations are assumed to be temporally simultaneous. In the 'separate-device' case, $\eta^0$ would be the same as $\eta(r)$ for all $r > 0$.

'Observed' versus 'Signal' Increments

Now consider the relationship between 'observed' and 'signal' increment covariance functions which results when the previously defined ensemble averaging concepts are applied to the sum of prediction and observation error variables. Two different relationships emerge, depending on whether or not pairs of observations are obtained using separate instruments.
(a) 'Separate-devices'

If simultaneous observations of variable P are recorded by different instruments, then the covariance of the observed variable $P'^{(0)} = (p - p'^{(f)})$ is the sum of the signal covariance $C(r)$, state error covariance $\eta(r)$, and measurement error covariance $\sigma(r)$. In this case the latter is zero for all $r > 0$, so that

$$C^{O}(0) = \tilde{C}(0) + \eta(0) + \sigma(0) = C(0) + \sigma(0)$$

but

$$C^{O}(r) = \tilde{C}(r) + \eta(r) = C(r) \quad \text{for } r > 0,$$

where $C^{O}(r)$ is the covariance function of $P'^{(0)}$. This relationship is represented schematically in Fig. 4.

Many conventional O.I. schemes calculate the weights to be applied to observed increments using a functional model for $C(r)$ based on estimates of $C^{O}(r)$, the latter being obtained by time averaging $(P'^{(0)}_l, P'^{(0)}_m)$ when $l$ and $m$ index a fixed pair of station locations. Suppose such a model is used to estimate $P'^{(a)}_l$ from predictions $P'^{(f)}_l$, $P'^{(f)}_m$ and one observation $P'^{(m)}_m$ at distance $r$ from location $l$. The conventional O.I. analysis variable $P'^{(a)}$ is then calculated as

$$P'^{(a)} = P'^{(f)}_l + \frac{C(r)}{C(0) + \sigma(0)} P'^{(m)}_m$$

(13)

where $P'^{(m)}_m = P'^{(m)} - P'^{(f)}$, $C(0)$ is obtained from the intercept of $C(r)$ at $r = 0$ and $\sigma(0)$ is a known or assumed value for measurement error variance. However, the best linear unbiased estimate for $P'^{(a)}$ is actually given by
Fig. 3. Schematic representation of the relationship between the covariance functions for observation error ($\eta_0(r)$), state error ($\eta(r)$) and measurement error ($\sigma(r)$).

Fig. 4. Schematic representation of the relationship between the covariance functions for observed increment ($C_0(r)$), signal increment ($\tilde{C}(r)$), and state error ($\eta(r)$), assuming measurement error covariance zero for all $r > 0$. 
\[ \hat{p}_g = p^e_g + \left[ \frac{C(r)}{C(0) + \sigma(0)} \right] p_m^0 \] (14)

It follows that (13) results in an analysis bias \( \beta(r) \) given by

\[ \beta(r) = (p^a - \hat{p}_g) = \left[ \frac{\eta(r)}{C(0) + \sigma(0)} \right] \frac{p_m^0}{p_m^0} \] (15)

If \( \eta(r) \) is assumed to be associated only with small scale variability in the true state variable \( P \), then \( \eta(r) \) and therefore \( \beta(r) \) will become negligible at sufficiently large separation. However, if \( r \) is small, then (15) implies too much weight being given to an observed increment. This means that the analysis (13) undersmooths the observed increment field.

(b) 'Same device'

When spatially distinct observations are assumed to be temporally simultaneous, but are nevertheless obtained using the same observing device, then the analysis \( p^a \) calculated using (13) leads to an analysis bias given by

\[ \beta(r) = \left[ \frac{\eta(r) + \sigma(r)}{C(0) + \sigma(0)} \right] \frac{p_m^0}{p_m^0} \] (16)

Thus in this case both measurement and state errors contribute to analysis bias if the model for \( C(r) \) is based on pairs of observed increment data obtained from the same observing device. This effect is clearly significant when rawinsonde data from different height levels are assimilated by an O.I. scheme which calculates \( p^a \) at different
points on a single height surface, and in this application some schemes attempt to allow for the measurement error correlation term \( \sigma(r) \) in calculating the weights applied to observed increments (Lorenc et al., 1977). However, this procedure can still not reduce the bias to less than that given by (15).

If we limit our attention to the 'separate-device' example (15), then the analysis bias variance as a function of separation \( r \) is given by

\[
\text{var}[\beta(r)] = \frac{\eta^2(r)}{C(0) + \sigma(0)} .
\] (17)

Dividing this by \( \bar{C}(0) \) and taking the square-root leads to a dimensionless measure of analysis bias relative to r.m.s. signal increment variability given by

\[
F_{\beta}(r) = \frac{\eta(r)}{\bar{C}(0)[\bar{C}(0) + \eta(0) + \sigma(0)]}^{1/2} .
\] (18)

If \( \eta(r) \) is associated with small-scale fluctuations in \( P \), then \( F_{\beta}(r) \) clearly tends to zero as \( r \) increases. \( F_{\beta}(r) \) is also small for all \( r \) provided that \( C(0) \) is large compared with \( \eta(0) \). The latter circumstance probably applies in meteorological analysis when \( p^f \) is a crude predictor for \( \bar{P} \), such as when climatological means or space-averaged observations are used to define the predicted field; there is then no point in distinguishing between 'signal' and 'observed' increment covariances in the O.I. analysis formulation. However, if \( p^f \) is an accurate predictor of \( \bar{P} \) in the sense that \( \bar{C}(0)/\text{var}[\bar{P}] \ll 1 \), then (18)
suggests that the fractional analysis bias \( F_B(r) \) increases roughly in proportion to the accuracy \( \tilde{C}(0)^{-1} \) of the predictor \( P^f \). This has obvious implications for large scale assimilation of increment field data in relation to numerical forecasting, but could also be important in relation to mesoscale analysis of special data sets when the predictor \( P^f \) is based on the output from some larger-scale background analysis. (We discuss possible strategies for eliminating this source of analysis bias in the concluding part of this section.)

**Sample Calculations**

The previous discussions are only relevant to the problem of meteorological field analysis using O.I. methods provided that there are significant differences between the space-lagged covariance functions of 'signal' and 'observed' increment variables. This difference depends on constraints imposed by analysis goals on the estimated signal field structure. Here we consider the case where a height increment analysis is required to result in a band-limited variance spectrum for the signal increment when the spectral representation of an observed increment covariance model is not itself band-limited. The motivation for this example is that a global O.I. analysis is unlikely to resolve unambiguously the amplitude and phase of spatial wavelength components smaller than some multiple of mean station separation, and that these scales should therefore be eliminated by the analysis.

To obtain typical spectral representations of observed increment data we use some correlation models fitted to RAOB-minus-predicted
geopotential correlations detailed by Thiébaux (1980). These are isotropic, second order autoregressive models of the form

\[ \rho(r) = \rho(0) \left[ \cos(\alpha r) + \left( \frac{\gamma}{\alpha} \right) \sin(\alpha r) \right] \exp(-\gamma r), \quad (19) \]

where in our notation \( \rho(r) = C(r)/C^0(0) \). Thiébaux tabulates values for \( \alpha, \gamma \) and \( C^0(0) \) at standard pressure surfaces between 1000 and 100 mb. Thus an equivalent space-lagged covariance model for \( C(r) \) can be computed from

\[ C(r) = C^0(0) \rho(r). \quad (20) \]

The curve of \( C(r) \) versus \( r (<1000 \text{ km}) \) for 500 mb height increment data is shown in Fig. 5a.

The spectral representation of \( C(r) \) is given by the Hankel transform \( H(\omega) \) of \( C(r) \):

\[ H(\omega) = \int_0^\infty r J_0(\omega r) C(r) \, dr \quad (21) \]

where \( \omega \) is the magnitude of a spatial wavenumber \( 2\pi/\lambda \), and \( J_0(*) \) the zero order Bessel function of the first kind (Sneddon, 1951). The inverse transform is

\[ C(r) = \int_0^\omega H(\omega) J_0(\omega r) \, d\omega. \quad (22) \]

With \( r = 0 \), the state increment variance is therefore given by
so that the variance contribution per unit wavenumber at wavenumber \( \omega \) is

\[
S(\omega) = \omega H(\omega).
\]

The upper curve in Fig. 5b shows \( S(\omega) \) versus \( \omega \) for the height covariance model derived from (19) using values for \( \alpha, \gamma \) and \( C^0(0) \) at 500 mb (Fig. 5a).

To simulate a feasible spectral representation for the signal increment covariance \( \tilde{C}(r) \), we assume an empirical relationship between \( S(\omega) \) and its signal counterpart \( \tilde{S}(\omega) \) of the form

\[
\tilde{S}(\omega) = \omega \tilde{H}(\omega) = f(\omega) S(\omega) \tag{24}
\]

where \( f(\omega) = \exp(-\beta \omega^4) \).

With \( \beta = 1.5 \times 10^9 \text{ km}^4 \), this model effectively eliminates the variance contribution of \( C(0) \) for wavelengths less than the order of 1500 km (see lower curve in Fig. 5b). This is equivalent to assuming that the resolvable part of the observed increment field contains contributions from zonal wavenumbers less than about 30 in middle latitudes, and that the higher wavenumber contributions correspond mainly to state-error fluctuations in the observed increment data used to derive the \( C(r) \) model.
Fig. 5. The relationship between models for state and signal increment processes as represented by

(a) the difference in lagged covariance functions, and

(b) the distribution of field variance with respect to spatial frequency
Having obtained a model for the spectral representation of a signal covariance function \( \tilde{C}(r) \), \( \tilde{C}(r) \) itself is obtained via the inverse Hankel transform (22) when \( \tilde{H}(\omega) \) replaces \( H(\omega) \).

The result for the case of 500 mb height increment data is shown as the dashed curve in Fig. 5a. Not surprisingly, the difference between \( \tilde{C}(r) \) and \( C(r) \) is small at large separation scales. However, there are significant differences for \( r < 300 \) km. The difference \([C(0) - \tilde{C}(0)]\) estimates the state-error variance \( \eta(0) \). At 500 mb, our model for \( \tilde{C}(0) \) thus estimates \( \eta(0) \approx 83 \) gpm\(^2\), which is roughly half the measurement error variance \( \sigma(0) \) estimated as \( C^0(0) - C(0) \). At first sight, this result suggests that the influence of state error is relatively small compared with the influence of measurement error. This may be true in the case of the height error correlation function \( \rho(r) \). However, an important difference between \( \tilde{C}(r) \) and \( C(r) \) at small separation is that the derivative \( \partial \tilde{C}(r) / \partial r \) is considerably less than \( \partial C(r) / \partial r \). This may be particularly significant with respect to the problem of modeling the 'bulk-state' wind component correlation and cross-correlation functions by assuming that these are related geostrophically to an observed height increment correlation model (Buell, 1972).

Some indication of the maximum O.I. analysis bias which is likely to result from using the model for \( C(r) \) in place of a signal covariance \( \tilde{C}(r) \) can be obtained by calculating \( F_\beta(0) \) using (18) with \( r = 0 \). Results are shown in Table I, where we have assumed that the same relationship between signal and state increment variance spectra applies at all levels (Eq. (23)). According to these calculations, the
potential risk of O.I. analysis bias is small below 500 mb and above 200 mb. However, the relatively large values of $F_\beta(0)$ between 500 and 200 mb suggest that an inappropriate choice of signal covariance model could lead to appreciable analysis errors at levels of relatively high spatial variability in the bulk-state field.

Implications for Analysis Strategies

We have suggested that in global O.I. analysis schemes involving a relatively accurate predictor for bulk state analysis variables, covariance models derived from observed increment statistics do not lead to the best weight formulation for interpolating the bulk-state increment field. However, we also have to admit that it may not be easy to arrive at more suitable covariance models, since the bulk-state field is not itself observable. One obvious solution is to carry out O.I. gridpoint analysis using the observed-increment covariance models, and rely on gridpoint filtering techniques to remove unwanted or ill-defined scales of variability from the analysis field. In effect, this procedure would be similar to current data assimilation schemes in which gridpoint analysis fields are 'initialized' prior to insertion into a numerical prediction model. One drawback of this approach in diagnostic applications is that realistic details of the analysis field may be removed from data-rich regions of the global analysis domain. This problem could be overcome using filtering techniques in which the degree of smoothing depends on the local density of observations. This approach would be difficult to incorporate into conventional O.I. schemes, since these are based on an essentially global approach to analysis. However,
the concept of data-adaptive filtering is certainly not new to meteorology; it is effected by the successive approximation method developed originally by Cressman (1959), the filter-response aspects of which have been studied by Stephens and Stitt (1970) and Stephens and Polan (1971). Cressman's successive-approximation approach cannot in general result in optimal (minimum variance) estimation of the bulk state increment variables, and for this reason O.I. formulations are likely to be preferred for most meteorological applications. We note that the 'predictor-corrector' approach proposed by Ghil et al. (1981) may be capable of effecting variable resolution of the observed field. This approach has the advantage of being optimally formulated, though some assumptions regarding the covariance properties of both measurement and state error variables are still required, and the method has yet to be proved in realistic analysis situations. (Optimally formulated 'conditional probability' approaches to data-adaptive analysis are considered further in Sections V and VI.)

In summary, we suggest that O.I. analysis strategies for field dependent estimation of meteorological prediction error variables should take into account the difference between 'signal' and 'state' components of observed variation in arriving at suitable covariance models for optimal weight formulation. Failure to distinguish between signal and state structure is likely to lead to sub-optimal estimation of the signal field when (i) the variance of a signal increment variable is appreciably less than the variance of the signal variable itself, and (ii) the signal analysis is expected to satisfy constraints with respect to its spatial structure which are not satisfied by the state structure sampled by observations.
Table 1  
Columns 2 to 6 give estimated variances for observed increments \(C^0(0)\), state increments \(C(0)\), signal increments \(\tilde{C}(0)\), measurement error \(\sigma(0)\), and state error \(\eta(0)\), based on height prediction error statistics published by Thiébaux (1980) and the assumed relationship between state and signal increment variance spectra represented by Eq. (23). Column 7 gives the maximum signal analysis bias relative to r.m.s. signal increment variation which results from using \(C(r)\) in place of \(\tilde{C}(r)\) in Eq. (14).

<table>
<thead>
<tr>
<th>Level (mb)</th>
<th>(C^0(0))</th>
<th>(C(0))</th>
<th>(\tilde{C}(0))</th>
<th>(\sigma(0))</th>
<th>(\eta(0))</th>
<th>(F_B(0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>330</td>
<td>175</td>
<td>157</td>
<td>155</td>
<td>18</td>
<td>0.08</td>
</tr>
<tr>
<td>850</td>
<td>270</td>
<td>162</td>
<td>155</td>
<td>162</td>
<td>7</td>
<td>0.03</td>
</tr>
<tr>
<td>700</td>
<td>310</td>
<td>170</td>
<td>162</td>
<td>140</td>
<td>8</td>
<td>0.04</td>
</tr>
<tr>
<td>500</td>
<td>615</td>
<td>462</td>
<td>378</td>
<td>153</td>
<td>83</td>
<td>0.17</td>
</tr>
<tr>
<td>400</td>
<td>920</td>
<td>717</td>
<td>560</td>
<td>203</td>
<td>157</td>
<td>0.22</td>
</tr>
<tr>
<td>300</td>
<td>1225</td>
<td>1030</td>
<td>743</td>
<td>195</td>
<td>287</td>
<td>0.30</td>
</tr>
<tr>
<td>250</td>
<td>1355</td>
<td>1030</td>
<td>788</td>
<td>325</td>
<td>242</td>
<td>0.23</td>
</tr>
<tr>
<td>200</td>
<td>1445</td>
<td>910</td>
<td>650</td>
<td>535</td>
<td>260</td>
<td>0.23</td>
</tr>
<tr>
<td>150</td>
<td>1650</td>
<td>875</td>
<td>816</td>
<td>775</td>
<td>59</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>2420</td>
<td>1112</td>
<td>1080</td>
<td>1308</td>
<td>32</td>
<td>0.02</td>
</tr>
</tbody>
</table>
REFERENCES


III. IDENTIFYING REQUIREMENTS FOR ANALYSES

Large-Scale Numerical Weather Prediction

The most numerous applications of objective analysis techniques have been in the field of large-scale numerical weather prediction. Over the few decades of the use of techniques for translating observations of meteorological variables into analyses specified at mesh points, the goals of such procedures have become clearer. At the outset, the presently perceived goals for large-scale analyses were either not necessary, not recognized, or only vaguely stated. The successive approximation technique (Bergthórsson and Déjos, 1955, and Cressman, 1959) was developed as a smoothing, interpolative device, but with little attention given to prescribing expectations of the analyses per se. The statistical optimum interpolation scheme (Gandin, 1963) offers a different approach with many appealing features; however the least squares minimization criterion which defines O.I. analyses still needs to be carefully considered in terms of analysis goals. With the advent of numerical forecasting models based upon the primitive equations, the analysis and the initialization procedures have gradually merged. At the present time, a basic paradigm exists for large-scale analysis for numerical weather prediction and the goals can be stated with some clarity. Simply put, the translation from observations to analysis needs to specify those modes which allow the model equations to simulate the low frequency time evolution of the atmosphere. The first attempt to do this was the Hough analysis scheme. The combining of analysis and initialization procedures has focused attention on the
need for clearly defining the goals, and makes possible discussion of analysis strategies and variants of analysis procedures to meet stated goals. Examination of advantages and limitations of both the analysis and the initialization procedures can be carried out once these are understood. For example, Phillips (1982) has pointed out some interesting consequences of combining the analysis and initialization steps. Among them are: 1) the absence of a need for dynamic constraints to be built into the analysis scheme if all the data are used at each mesh point, and 2) that the proper statistical structure to be incorporated into a statistical O.I. or Kalman technique is that of the forecast error of the desired ('slow') modes. This does not mean that such an analysis-initialization procedure has been achieved in practice, since there are many problems to be solved. However, at the least, the problems are identifiable.

Analysis for Diagnostic Purposes

Although diagnostic calculations or phenomenological studies have commonly used objective analyses produced by operational weather forecasting procedures, recent experience has suggested that these purposes really have a different set of objectives. [In special circumstances, such as the GATE, either subjective analyses were used, or special objective techniques were developed.] The gradual merging of analysis and initialization techniques has created assimilation procedures for forecasting purposes, for which the goals are basically different from those specified by diagnostic purposes. While the objective of the analysis specified for forecasting purposes is to be in terms of the
slow modes of the forecast model, a more appropriate goal for purposes of diagnostic work might be an analysis of the slow manifold of the atmosphere or, at least, of the most energetic modes. In brief, the reason for considering analysis goals in this framework is that, with the present state of the art, the slow manifolds of atmosphere and model are not the same. Ultimately, of course, the objectives of analyses for forecast and diagnostic purposes should converge.

Requirements for Mesoscale Analysis

At present there seem to be two different genera of techniques termed mesoscale forecasting. The simpler of the two, from the point of view of analysis, is the carrying out of a forecast on a computational grid or on a scale which is significantly smaller than that in global weather prediction. The initial state for such a mesoscale forecast model is one predicted from, or prescribed by, a global model. For these types of mesoscale models the requirement that initial model conditions be balanced dictates that the initial conditions come from a much larger scale model; and the role of the assimilation of observations is secondary. The raison d'être for the 'mesoscale' of the model is the reduced computational mesh and the sophistication of the model physics. In this type of mesoscale modeling, the information on scales between the observation scale (relatively large) and the computation scale (smaller) is generated by the model.

The principal characteristic which distinguishes the second type of mesoscale model is the scale of the observational material. While the observation mesh is the conventional World Weather Watch system for the
first type of mesoscale model, here the observations are assumed to be on a smaller or denser scale. Furthermore, the field variables themselves might not be restricted to the conventional large-scale variables, and the model coordinates might be different. For example, isentropic coordinates might be used. To date, little if any thought appears to have been given to the implications of these model features for the structure of the analysis scheme.

The inclusion of field variables not normally treated by conventional objective analysis techniques may be appropriate for some mesoscale analysis goals. Among those variables appearing as candidates are various boundary layer parameters, e.g., boundary layer height, and mesoscale parameters such as cloud indices and rainfall rate. In addition, since the mesoscale model mesh is capable of resolving them in a more adequate fashion than a global model mesh, such phenomenological features such as fronts and squall lines might be included. The very physical basis for employing a mesoscale model strongly suggests that the model physics be more detailed than for global scale models. Thus, there are additional demands placed upon the analysis procedure. A 'mesoscale analyst' must first come to grips with the goals of such an analysis and then determine how best to achieve them in this, more rigorous setting.
REFERENCES


IV. DISCUSSION OF ACCURATE FIELD-DEPENDENT ESTIMATION PROCEDURES

PRESENT STRENGTHS AND LIMITATIONS OF ANALYSIS SCHEMES

Deterministic Schemes

In deterministic schemes the estimated bulk-state field is not considered as a stochastic variable. The estimation schemes involve the fitting of preselected or empirical basis functions by minimizing the squared difference between the observations and the function values. The only operational analysis scheme in this category is the Hough function scheme employed by the National Meteorological Center of NOAA. The apex of these deterministic schemes was probably the method developed by Dixon et al. (1972), using orthogonal polynomials, although it is no longer in use. The advantage of deterministic schemes, if it is that, is the continuous form of the resulting analysis which makes it simple to evaluate a variable field at grid points. The disadvantages are that the fitting procedure with observational meshes which vary in space and time becomes time consuming and clumsy. And, perhaps more importantly, the very fact that the bulk-state field is considered as deterministic and the 'analysis error' is considered to be entirely 'observation error' is a disadvantage: No consideration is thus given to the characteristics of the field away from the observation points.

Statistical Schemes

These schemes, by contrast, consider the bulk-state field as a stochastic variable. The following schemes use statistical information of various sorts in varying degrees of explicitness.
Successive approximation or 'distance weighting' schemes, which also go under the name of Cressman-type schemes, represent a large class of procedures of varying degrees of sophistication. They all involve propagating or interpolating differences between a guess field and observations to a grid mesh. An empirical weighting function, or functions, is used to interpolate; the interpolation procedure may be repeated recursively, using different weighting functions. It is important to note that in the successive approximation schemes no minimization procedure is used, nor are statistics of the field variable(s) being analyzed made use of. The statistical content of these schemes would seem to lie in the fact that the empirical nature of the form of the distance weighting makes possible a scale selectivity so that any desired spatial frequency response may be produced. This is, then, an advantage. The disadvantages of the successive approximation schemes are that their behavior tends to be highly dependent on the exact local spatial distribution of the observations, and considerable tinkering with the algorithms is necessary to obviate these difficulties. Furthermore, since these schemes are essentially just smoothing schemes, they are incapable of producing interpolated values which are outside the range of the observed values.

Statistical optimum interpolation schemes, now widely used in operational global numerical prediction units, contain in their structure information on the statistical structure of the fields analyzed. They handle the varying location and density of the
observations in superior fashion and are most easily configured to accommodate varying observational error and multi-variable analyses. The disadvantage is that, for practical reasons, they must be applied locally rather than globally; this results in sub-optimal behavior. Thus it prevents any intended constraints on the bulk-state analysis from holding exactly.

Kalman filtering is the most sophisticated theoretically of all suggested objective analysis techniques and the one with which there is the least experience. It permits the accommodation of statistical information both on the field variables and on the model generated guess fields, by continually updating a forecast error covariance model and an analysis error covariance model. At present, the perceived disadvantages are that it is too time consuming and that, after initial start up, it behaves very similarly to the conventional optimum interpolation scheme. Not enough experience has been gained with full-blown realistic assimilation algorithms to be able to know how to handle the various system matrices that must be specified and iteratively updated.

The classification referred to as 'mixed schemes' exists because of the extension of piece-wise spline techniques (which otherwise would be deterministic) to accommodate various quantitative smoothing and local data adaptive techniques. The work of Wahba (1979), and Wahba and Wendleberger (1980) are examples.
REFERENCES


A COMPARISON OF OBJECTIVE ANALYSIS SCHEMES

Having discussed some of the problems involved with identifying analysis requirements and strategies, we now briefly review some alternative analysis schemes and one published comparison between their performance when the analysis goal is to reproduce the value of an observation which itself has not been assimilated by the analysis.

The following is a critical review of Creutin and Obled's (1982) study of objective analysis and mapping technique for rainfall fields. Although their results are not altogether relevant to the analysis of all meteorological fields, it is nevertheless the only comprehensive, comparative study of its kind accessible from the geophysical literature.† The paper does illustrate some of the practical aspects of analysis strategy which are relevant to meteorological analysis. Creutin and Obled present a comparison of techniques to estimate point or average precipitation values at ungauged sites, including:

† Ed. note. In the ensuing time since the workshop, the following study, which is similar to and extends Creutin and Obled, has appeared: Franke, R. (1984). Sources of error in objective analysis. Naval Postgraduate School, NPS-53-84-0003, Monterey, CA 93943, 50 pp.
1. Description of techniques.
2. Discussion of practical aspects.
3. Case study for a region with unusually high rainfall variability.

The latter provided a good sample of events for testing the sensitivity of the estimation to the different methods used.

**Estimation Techniques Compared**

a. The nearest neighbor method (with automated version by Diskin, 1970).

b. The arithmetic mean over a given region surrounding the site.


The notation specifies

\[ Z(t^i) \equiv \text{measured rainfall at station with coordinates } t^i = (x_i, y_i) \]

with rainfall depth for the kth of \( p \) storm events denoted by \( Z_k(t^i) \), for \( i = 1, \ldots, N \).
Thus the entire data set is an \( N \times p \) array, for that region indicated with the rectangle over the mid-south of France, in Figure 1.

Figures 1, 2, 3, 4 and 5 and Table 1 in the following pages are from Creutin, J. D. and C. Obled, 1982: Objective analysis and mapping techniques for rainfall fields. *Water Resources Research*, 18, 413-431, copyright by the American Geophysical Union.
Spline surface fitting is achieved with a technique for minimizing the bending energy of a thin elastic sheet and is uniquely described by:

\[ S(t) = \alpha + \beta t + \sum_{i=1}^{N} \psi_i K(t^i, t) \]

with

\[ K(t^i, t) = ||t-t^i||^2 \log ||t-t^i||^2 \]

interpolating observed points \( S(t^i) = Z(t^i) \) and minimizing

\[ \int_{\Omega} |\nabla S(t)|^2 \, dt \]

for

\[ \nabla S(t) = \frac{\partial^2 S(t)}{\partial x^2} + 2 \frac{\partial^2 S(t)}{\partial x \partial y} + \frac{\partial^2 S(t)}{\partial y^2} \]

where the \( \psi_1, \ldots, \psi_N, \alpha \) and \( \beta \) are solutions of

\[
\begin{bmatrix}
K(t^1, t^j) & 1 & t^1 & \cdots & t^j \\
& & & \vdots & \vdots \\
& & & 1 & t^N \\
1 & \ldots & \ldots & 1 & 0 & 0 \\
t^1 & \ldots & \ldots & t^N & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\psi_1 \\
\vdots \\
\psi_N \\
\alpha \\
\beta \\
\end{bmatrix}
= \begin{bmatrix}
Z(t^1) \\
\vdots \\
Z(t^N) \\
0 \\
0 \\
\end{bmatrix}
\]

Gandin's optimal interpolation estimates the values at ungauged sites as linear combinations of values recorded by \( n \) surrounding rainguage stations as

\[ Z^*(t_0) = \sum_{i=1}^{n} \lambda_i Z(t^i) \]
where the $\lambda$'s are the solutions of

$$\sum_{i=1}^{n} \lambda_i C(t^i, t^j) = C(t^j, y) \quad j = 1, \ldots, n$$

This requires a covariance model; and relies on three hypotheses:

i. Rainfall patterns do not depend systematically on weather type.

ii. The stochastic component of the field is weakly stationary.

iii. The mean field $E\{Z(t)\}$ and standard deviation $\sigma(t)$ are available separately.

Matheron's Kriging method follows the same basic principle as Gandin's, but does not assume $E[Z^*(t^0) - Z(t)] = 0$. Rather it estimates rainfall at an ungauged site with

$$Z^*(t^\prime) = \sum_{i=1}^{n} \lambda_i Z(t^i)$$

where the $\lambda_i$ here are solutions of

$$\sum_{i=1}^{n} \lambda_i C(t^i, t^j) + v = C(t^j, t^0) \sum_{i=1}^{n} \lambda_i = 1.$$ 

This places a constraint on the weights in order to account for the erstwhile bias of the estimate. Creutin and Obled decompose the field as

$$Z_k(t) = m_k(t) + Z_k^*(t);$$
assume \( m_k(t) \) is a slow systematic trend, spatially, for each event; and estimate it with a low-order polynomial in the neighborhood.

EOF expansions represent the field as

\[
Z(t) = \sum_{\ell=1}^{M} Y_{\ell} \phi_{\ell}(t)
\]

which is a truncation of

\[
\sum_{\ell=1}^{\infty} Y_{\ell} \phi_{\ell}(t).
\]

Expansions are orthogonal in the functions, with coefficients that are uncorrelated random variables. For the kth rainfall event the rainfall amount at location \( t \) is represented as

\[
Z_k(t) = \sum_{\ell} Y_{k\ell} \phi_{\ell}(t)
\]

and the coefficients are obtained by orthogonal projection:

\[
\int_{\Omega} Z_k(t) \phi_{\ell}(t) dt = Y_{k\ell}
\]

since

\[
\int_{\Omega} \phi_{\ell}(t) \phi_{m}(t) dt = \delta_{\ell m}
\]

The eigenfunctions satisfy

\[
\int_{\Omega} r(t,t') \phi_{\ell}(t) dt = \mu_{\ell} \phi_{\ell}(t')
\]
where \( r(t,t') \) is the correlation function and \( \mathbb{E}\{Y_i Y_m\} = \mu_i \delta_{im} \). In practice a problem arises in the estimation of the \( \phi_\ell(t) \). However the technique has the following advantages:

i. It does not assume stationarity.

ii. Interstation correlations suffice without a correlation model.

iii. The truncation effects some (desirable) global smoothing.

Considerations in the Practical Applications of Techniques

a. As Creutin and Obled use Gandin's interpolation technique, lag-covariances are computed using location-specific climatological mean values \( \mathbb{E}\{Z(t_i)\} \) estimated from the 81 rainfall events (see Figure 2). Correlograms (shown in Figure 3) are fitted with \( \cos(\beta d) r(d) = e^{-\alpha d} \) for which \( \beta = 0 \). The required separate estimates of the

i. climatological mean field and the

ii. variance field

were obtained by fitting spline surfaces to the point values estimated at the 73 data-input stations, and evaluating at the 26 verification locations.

b. For the EOF expansion

\[
Z(t) = \sum_{\ell} Y_\ell \phi_\ell(t)
\]
Figure 2. Climatological mean contoured within the study region.
Figure 3. Empirical correlograms for the spatial array of gauge stations of the study region.

the $\phi_2(t)$ were evaluated for the observing net and estimated for the test station net. The first three eigenfunctions are illustrated in Figure 4.

If the $\phi_2(t^i)$ were assumed constant over surface elements $\Delta \omega_i$ and the nearest value used at $t^0$, $Z^*(t^0)$ would be equivalent to the nearest neighbor estimate. Smoothing the $\phi_2(t)$ surfaces was achieved using spline functions as in spline surface interpolation, and the smoothed values used at the locations $t^0$. 
Figure 4. First three eigenfunctions: on the left for a Markov field and on the right for the rainfall data.
Validation Procedure

73 stations were used to estimate values for 26 other raingauge sites for all 1956-1976 autumn rainfall events over 50 mm (81 storms). For each of the 81 events the rainfall depth measures were estimated for the 26 locations of the test sites; and the estimated values were compared with the recorded measurements.

Results are presented in three ways, two of which are described here:

a. For each of the 81 rainfall events the correlation between estimated and observed rainfall amounts was computed, using the 26 stations of the study network. The 81 correlation values were then plotted versus their ranks, giving what the authors call an "efficiency curve" for the estimation procedure. The latter was obtained for each of the six estimation procedures; and the results are compared in Figure 5.

b. Table 1 shows contingency tables: estimated versus observed rainfall amounts (categorized, for a 20 x 20 table), for four of the estimation procedures. These have an advantage over the "efficiency curves" of Figure 5, in permitting display and comparison of over and under estimations.

In summary it is clear that, as they are employed by Creutin and Obled, the Kriging, spline, and Gandin interpolation methods of estimation have little to distinguish them—except the complexity of their operational deployment. The authors elect the interpolation technique of Gandin as their preferred technique, in light of its
favorable performance and the relative ease of use. In fact they have very similar components in their practical application.

Figure 5. Efficiency curves for the six methods.
Creutin and Obled's study was concerned with the problem of a purely spatial analysis, and their analysis goal was defined as the ability of the analysis to predict values of 'missing observations'. However, meteorological analysis for numerical weather prediction is not purely spatial, since both forecast and observed data contribute to the analysis. Moreover, meteorological analysis goals, as implied in our earlier discussions, are not necessarily solely identified with the ability of the analysis to predict missing observations if the value of a signal ('bulk state') variable is not necessarily identical with the value of an atmospheric state variable at a point.

Here we review some theoretical concepts which apply to the estimation of a signal field from observations of a space-time distributive process, and examine the relevance of these theoretical concepts to analysis model formulation. The following is based on a rather comprehensive account of 'system transfer models' and 'dynamic parameter mapping' given in Bennett (1979). The purpose of our review is to draw attention to the considerable body of existing theory concerning space-time distributive processes (which will not be familiar to most meteorologists), and examine to what extent some theoretical models for such processes provide guidance in developing optimally formulated, linear prediction models assimilating both time and space series.
System transfer function model

A system is regarded as being composed of a vector input variable,
\[ U_t = \begin{bmatrix} U_{t1} \\ \vdots \\ 0_{tN} \end{bmatrix} \]
which is a time sequence of values at N locations, a vector output variable,
\[ Y_t = \begin{bmatrix} Y_{t1} \\ \vdots \\ Y_{tN} \end{bmatrix} \]
which is a second time sequence of values for the same N locations, and a system transfer function, which transfers sequence
\[ \{U_t\} \text{ into } \{Y_t\}: Y_t = S U_t. \]

Quoting Bennett:

"The transfer function of a system is the system since it completely defines the processes or changes induced by the operation of the system", and

"The properties of the transfer function, properly defined, are the properties of the process or system under study".
The simplest model for univariate input and response for \( N \) regions is

\[
\begin{bmatrix}
Y_{t1} \\
\vdots \\
Y_{tN}
\end{bmatrix} = \begin{bmatrix}
S_{11} & \ldots & S_{1N} \\
\vdots & \ddots & \vdots \\
S_{N1} & \ldots & S_{NN}
\end{bmatrix} \begin{bmatrix}
U_{t1} \\
\vdots \\
U_{tN}
\end{bmatrix}
\]

where some elements of \( S \) called system parameters may be 0, corresponding to no regional interaction or coupling. Diagramatically, a simple two-region system may be represented as

![Diagram of a two-region system](image_url)
A multivariate input/response model has a rather more complex representation as:

\[
\begin{align*}
Y_t' & = S \times u_t' \\
Y_{tN} & = S \times u_{tN}
\end{align*}
\]

However the mathematical representation remains as:

Here each region's input is an \( mx_l \) vector and output is an \( nx_l \) vector.
In many physical situations a more realistic representation is a multivariate autoregressive, lag-distribution model with a random component, because it admits a stochastic element to the model. Thus:

\[(A_0 + A_1Z + \ldots + A_pZ^p) Y_t = (B_0 + B_1Z + \ldots + B_qZ^q)U_t + \varepsilon_t.\]

Alternatively we may write

\[Y_t = A^{-1}(Z) B(Z) U_t + \varepsilon_t\]

with system transfer function matrix

\[S = A^{-1}(Z) B(Z) .\]

Bennett has defined Z to be the unit shift operator:

\[Z Y_t = Y_{t-1}, \ldots, Z^p Y_t = Y_{t-p} \text{ etc.}\]

so that A(Z) is a polynomial matrix of autoregressive parameters and B(Z) is a polynomial matrix of distributed lag parameters. Consequently

\[(A_0 + A_1Z + \ldots + A_pZ^p) Y_t = \]

\[
\begin{bmatrix}
A_{11} & \ldots & A_{1N} \\
\vdots & \ddots & \vdots \\
A_{NL} & \ldots & A_{NN}
\end{bmatrix}
\begin{bmatrix}
Y_{t1} \\
\vdots \\
Y_{tN}
\end{bmatrix}
+ \]

\[
\begin{bmatrix}
A_{11} & \ldots & A_{1N} \\
\vdots & \ddots & \vdots \\
A_{NL} & \ldots & A_{NN}
\end{bmatrix}
\begin{bmatrix}
Y_{(t-1)1} \\
\vdots \\
Y_{(t-1)N}
\end{bmatrix}
+ \]

\[
\begin{bmatrix}
A_{11} & \ldots & A_{1N} \\
\vdots & \ddots & \vdots \\
A_{NL} & \ldots & A_{NN}
\end{bmatrix}
\begin{bmatrix}
Y_{(t-p)1} \\
\vdots \\
Y_{(t-p)N}
\end{bmatrix}
\text{ etc.}\]
In the complete space-time system transfer function model

\[ Y_t = A^{-1}(Z) B(Z) U_t + \varepsilon_t \]

the stochastic term \( \varepsilon_t \) acknowledges that the deterministic component does not account for all influences. We illustrate this with a superposition of an observed process on a time-evolutionary signal component, where the wiggly line illustrates the variation about the signal due to the stochastic "noise" element of the process.
The general shift-polynomial representation enables inclusion of variations in gradients and curvatures of impulse and autoregressive influences on output.

Example of simple input + response: precipitation (U) + runoff (Y)

Bennett's general representation is an elegant extension of Box-Jenkins time series models to space-time representations. In both there is a uniqueness failure which bothers me, for a number of reasons. In practice, I believe we obtain the same result if we simply write the model as an autoregression. With this we obtain a representation which is simpler to convert to a continuous space model with its equivalent linear filter representation. These lead easily to derivation of corresponding lag-covariance functions and to examination of spectral properties of the system modeled.
Parameter Mapping

Bennett considers tracking time-evolutionary system behavior; and introduces a change of notation (which we, in turn, adjust slightly to serve our purposes).

In the methodology of dynamic systems analysis two equations are used: one, called the system model by Bennett, relates the observations, $Y_t$, to the state of nature that is relevant but unknown, $\theta_t$:

$$ Y_t = F \theta_t + e_t $$

The second equation, which Bennett calls the parameter model, describes how the state of nature evolves:

$$ \theta_t = G \theta_{t-1} + HV_t + \Gamma n_t $$

In the latter the $V_t$ are the so-called exogenous variables originating outside the system and unaffected by it; and the $e_t$ and $n_t$ are stochastic or noise variables pertaining to the case in which neither equation is deterministic. $F, G, H$ are matrices that are known and describe the relationships between the appropriate dependent variables.

In order to translate the dynamic systems analysis, described by Bennett, into the present context of objective analysis of meteorological variables, it is essential to pause and sort out
which 'model' is which. We begin with the atmosphere itself which we shall regard as the 'ultimate model,' and whose behavior it is that we wish to describe and track. To assist us in doing so we have available a host of models, deterministic or statistical, which are intended to describe and predict the real atmosphere. In the case of deterministic models, the models are formulated in terms of dependent variables which are the meteorological quantities we will use to describe the atmosphere. However, we note that we cannot observe precisely those dependent variables. Rather, our observations are of quantities which are neither the dependent variables of a deterministic model nor a measure of the actual, true state of the atmosphere. Nonetheless we can use a combination of an approximating model of the atmosphere and observations of atmospheric variables to estimate atmospheric structure and behavior. This is exactly what is done in data assimilation procedures, using an analysis and a forecast cycle with an assimilating (forecast) model. Consequently what we need to do here is to make the proper correspondence between the language of dynamic systems analysis and the more familiar language of data assimilation.
Looking at our illustration (adapted from Bennett, page 355), the discrete flow diagram above the double line represents The Atmosphere. The dependent variable vector, $\Theta_t$, is governed by physical and thermodynamical laws which are designated by the matrix $G$. The time evolution, or time dependence, of $\hat{\Theta}_t$ is represented here by time shift operator $ZI$. Exogenous variables which may influence the evolution of $\Theta_t$ are represented by the vector $V_t$ operated on by a matrix $H$ which specifies the physics of the influence of $V_t$ on $\Theta_t$. (For example, for a forecast model, sea-surface temperature is an
exogenous variable.) Below the double line we have represented an approximating model of The Atmosphere. It assumes a number of things which we should take care to consider; namely, that we know the true separation of exogenous and dependent variables, that we have a reasonable model, $G_0$ of the true physics of the atmosphere, and that we can compute its evolution in time reasonably accurately. A number of additional 'boxes' or transformations appear here, which do not appear above the line: The matrix $F$ accounts for the fact that our model estimates of $\Theta_t$, namely $\hat{\Theta}_t$, do not correspond to the actual variables that are measured by our observing system. (See Chapter II.) The matrix $IK$ which is the crux of the system, designates the mechanism for combining the observations of $\Theta_t$, namely the $Y_t$, with our estimates of $\Theta_t$ (from the approximating model) to produce new estimates (for input to the approximating model). Simply put, our objective analysis algorithm and assimilation forecast model together comprise 'The Observer.' The Observer's contact with reality comes from observations. We assume the assimilating forecast model is a reasonably good approximation to the real atmosphere.

Some comments are also in order on Bennett's 'parameter model' equation: First, in our frame of reference it refers not to parameters used in a statistical or mathematical sense but to our dependent variable vector. Second, the fact that this 'parameter model' equation has a specific linear form is of no particular
consequence as far as 'The Observer' is concerned. The atmosphere is certainly not linear; and, although we may approximate its physics with a (vector) linear representation for such purposes as calculating 'The Observer' matrices: \( F, G, \) and \( H, \) in the context of interest to us the assimilation forecast model we use in 'The Observer' is non-linear. We may extend Bennett's representation to accommodate this, simply by rewriting the equation in schematic non-linear form.

"Mapping" the behavior of the system refers to tracking the patterns of system behavior under different inputs and over different time periods. The response of a system is dependent as much on the initial conditions as on the form of the system equation; and Bennett's "global analysis" enables us to study the behavior of the system equations over the entire range of system initial conditions. There are various ways of doing time-dependent sequential parameter estimation, corresponding to various assumptions with regard to the system equation and parameter equation. Three distinct cases are characterized as follows:

a. Dynamic Parameter Reconstruction, otherwise known as "the Luenberger Observer",

b. Stochastic Parameter Reconstruction, alternatively called "The Kalman Observer",

c. Predictor-Corrector Estimates of the stochastic system equation, which is now well known as "the Kalman Filter".

We will briefly review these three approaches.
a. **dynamic parameter reconstruction** models the system and parameter equations as

\[ Y_t = F\Theta_t \]

and

\[ \Theta_{t+1} = G\Theta_t + H V_t, \]

i.e., with no noise contribution to either. Here the estimates of unknown \( \Theta_t \) are constructed from (exact) output \( Y_t \) and known system and parameter equations, via a model of the system which may be diagrammed as
The Luenberger observer $S_2$ is a construct of the individual researcher's choice and produces an estimate:

$$
\hat{\theta}_t = \tau \Theta_t + \epsilon_t
$$

via predictor-corrector equation:

$$
\hat{\Theta}_{t+1} = G \hat{\Theta}_t + \text{H}_{t+1} \gamma_t + \kappa (\gamma_t - F \hat{\Theta}_t)
$$

$$
= (G - \kappa\text{H}F) \hat{\Theta}_t + \kappa \gamma_t + \text{H}_{t+1} \gamma_t.
$$

Thus we have updated estimates of the parameters at each point of time, using known matrices $F$, $G$ and $H$ observed values of $\gamma_t$ and the correction term. The difference $\gamma_t - F \hat{\Theta}_t$ is the difference between the observed system output and that predicted by the previous parameter estimate. The weight matrix $\kappa$ is called "Luenberger gain" and is chosen by the model builder. Thus this system model seems analogous to the successive approximation schemes.

The choice of the gain matrix $\kappa$ is clearly crucial. The only restriction is that the eigenvalues of the observer $(G - \kappa HF)$ must be negative. Preferably they should be chosen to be more negative than the system eigenvalues (of $G$) for faster convergence. However, very large negative values make the observer highly sensitive to noise or specification errors. Luenberger suggests as a rule of thumb, choosing the eigenvalues slightly smaller than the system eigenvalues.
Example. Identify an optimal observer for a two-parameter system:

\[
\begin{bmatrix}
\theta_{t+1}^1 \\
\theta_{t+1}^2
\end{bmatrix} = \begin{bmatrix}
-3 & 1 \\
0 & -1
\end{bmatrix} \begin{bmatrix}
\theta_t^1 \\
\theta_t^2
\end{bmatrix} + \begin{bmatrix}
0 \\
1
\end{bmatrix} v_t
\]

and

\[
y_t = \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
\theta_t^1 \\
\theta_t^2
\end{bmatrix}
\]

In this system

\[
G = \begin{bmatrix}
-3 & 1 \\
0 & -1
\end{bmatrix}, \quad W = \begin{bmatrix}
0 \\
1
\end{bmatrix}, \quad F = \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\]

The observed matrix is

\[
G_t - KF = \begin{bmatrix}
-(3 + K_{11}) & 1 \\
-K_{21} & -1
\end{bmatrix}
\]

with characteristic equation

\[
0 = \det(G_t - KF - \lambda I) = -(3 + K_{11} + \lambda) \quad \lambda^2 + \lambda(4 + K_{11}) + (3 + K_{11} + K_{21})
\]

If the eigenvalues of the observer matrix are chosen to be -2 and -3 (the eigenvalues of the system matrix \(G\) are -1 and -3) then

\[
0 = (\lambda + 2)(\lambda + 3) = \lambda^2 + 5\lambda + 6 \quad \text{and thus} \quad 4 + K_{11} = 5, \quad K_{11} = 1 \quad \text{and} \quad 3 + K_{11} + K_{21} = 6, \quad K_{21} = 2.
\]
Hence the \textit{spatial estimates} of the parameters influenced by the exogenous variables $V_t$ and the observed system output are

$$
\hat{\theta}_{t+1} = (G-KF)\hat{\theta}_t + K V_t + H V_t
$$

$$
= \begin{pmatrix} -4 & 1 \\ -2 & -1 \end{pmatrix} \hat{\theta}_t + \begin{pmatrix} 1 \\ 2 \end{pmatrix} Y_t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} V_t
$$

Bennett shows that the response dynamics of the Luenberger observer are equivalent to an exponential smoothing model, with the degree of weighting depending on the choice of the gain:

$$
\hat{\theta}_{t+1} - \tau \hat{\theta}_{t+1} = \tau \hat{\theta}_t + [\exp(G-KF)](\hat{\theta}_0 - \tau \hat{\theta}_0)
$$

b. \textbf{stochastic parameter reconstruction} models the system and parameter equations as

$$
Y_t = F \theta_t
$$

and

$$
\hat{\theta}_{t+1} = G \hat{\theta}_t + H V_t + \Gamma n_t
$$

Hence the parameter equation is stochastic while the system equation is deterministic. In "the Kalman observer" it is assumed that

$$
E[n_t n_{-t}'] = Q \delta(t-\tau) \quad E[\theta_t' n_t] = 0
$$

where $F, G, H, V_t$ are known. With an overbar denoting expectation we have:
\[ \Theta_{t+1} - \Theta_{t+1} = \mathcal{G}(\Theta_t - \Theta_t) + \mathcal{H} \mathcal{V}_t - \mathcal{H} \mathcal{V}_t + \Gamma(n_t - n_t) \]

and the variance-covariance matrix is

\[ E\{(\Theta_{t+1} - \Theta_{t+1}) (\Theta_{t+1} - \Theta_{t+1})'\} \]

\[ = \mathcal{G} E\{(\Theta_t - \Theta_t) (\Theta_t - \Theta_t)'\} \mathcal{G}' + \Gamma E\{n_t n_t'\} \Gamma'. \]

Thus the mean of \( \Theta_{t+1} \) is \( \mathcal{G} \Theta_t + \mathcal{H} \mathcal{V}_t \) and the variance-covariance matrix is

\[ P_{t+1} = \mathcal{G} P_t \mathcal{G}' + \Gamma Q \Gamma. \]

Accordingly:

i. We observe \( Y_t = \mathcal{W} \Theta_t \)

ii. When parameter noise is zero-mean normally distributed, we have the conditional distribution of \( \Theta_{t+1} \) given \( \Theta_t \).

Thus the evolutions of parameter distributions may be studied via

\[ p(\Theta_{t+1}) = p(\Theta_{t+1} | \Theta_t) p(\Theta_t), \]

and Bennett illustrates this with a simple example.

c. predictor-corrector estimates of stochastic system equation, otherwise known as "the Kalman filter" (here we omit for simplicity the exogenous variable vector):

\[ Y_t = \mathcal{W} \Theta_t + \epsilon_t \]

\[ \Theta_{t+1} = \mathcal{G} \Theta_t + \Gamma n_t \]
with noise components in the system and parameter equations assumed to satisfy

\[ E[\epsilon_t \epsilon'_t] = R \delta(t-\tau) \]
\[ E[n_t n'_t] = Q \delta(t-\tau) \]

and to be normal with zero means. The Kalman Filter is a Bayesian solution: the posterior conditional p.d.f. \( p(\theta_t | y_{t-1}) \) derived from the previous stage in parameter estimation becomes the prior p.d.f. at the next.

The estimation algorithm for the parameters and their variances is analogous to the deterministic structure, but takes account of the errors in the estimates at each time point based on information available up to the previous point in time.

**The update equations for deterministic parameter equation**

\( \theta_{t+1} = G_t \hat{\theta}_t \) with a priori parameter variation prediction equations are

\[
\hat{\theta}_{(t+1)|t} = G_t \hat{\theta}_t \quad \text{and} \quad P_{(t+1)|t} = G_t P_t G'_t
\]

\[
\hat{\theta}_{t+1} = \hat{\theta}_{(t+1)|t} - P_{(t+1)|t} \left[ F' P_{(t+1)|t} F + \sigma^2 \right]^{-1} (F \hat{\theta}_{(t+1)|t} - y_{t+1})
\]

\[
P_{t+1} = P_{(t+1)|t} - P_{(t+1)|t} \left[ F' P_{(t+1)|t} F + \sigma^2 \right]^{-1} F' P_{(t+1)|t}
\]
The update equations for stochastic parameter equation

\[ \theta_{t+1} = \mathcal{G}_t \theta_t + \Gamma_n_t \]

are

\[ \hat{\theta}_{(t+1)|t} = \mathcal{G}_t \hat{\theta}_t + \Gamma_n_t \] and \( P_{(t+1)|t} = \mathcal{G}_t P_t \mathcal{G}_t' + \Gamma \cdot \Omega \cdot \Gamma' \)

for prediction, and

\[ \hat{\theta}_{t+1} = \hat{\theta}_{(t+1)|t} - P_{(t+1)|t} [\mathcal{F} \Omega_{(t+1)}|t - \mathcal{F} \hat{\theta}_{(t+1)}|t - \mathcal{F} \mathcal{F}']^{-1} \]

\[ P_{t+1} = P_{(t+1)|t} - P_{(t+1)|t} [\mathcal{F} \Omega_{(t+1)}|t - \mathcal{F} \mathcal{F}']^{-1} \mathcal{F} P_{(t+1)|t} \]

Thus the Kalman filter derives extrapolations of the parameter values at the next (future) stage according to

\[ p(\theta_{t+1} | \gamma_0, \gamma_1, \ldots, \gamma_t, \gamma_{t+1}) = p(\theta_{t+1} | \gamma_{t+1}, \lambda_t) \]

where \( \gamma_{t+1} \) are the vector system outputs at time \( t+1 \) and \( \lambda_t = \{ \gamma_0, \gamma_1, \ldots, \gamma_t \} \) is the complete realization of all outputs up to time \( t \). Specifically

\[ p(\gamma_{t+1} | \theta_{t+1}, \lambda_t) = \frac{p(\gamma_{t+1} | \theta_{t+1}, \lambda_t)}{p(\theta_{t+1} | \lambda_t)} \]

\[ p(\theta_{t+1} | \gamma_{t+1}, \lambda_t) = \frac{p(\theta_{t+1} | \gamma_{t+1}) p(\theta_{t+1})}{p(\gamma_{t+1})} \]

The latter (Bayes) formula combines old information on the parameters with new information derived from the magnitude of the deviation of predicted system output from the observations of the 'actual' system output. Then the optimal estimates are
maximum likelihood

\[ \hat{\theta}_{t+1} = \text{mode} \ p(\theta_{t+1} | \gamma_{t+1}) \]

minimum variance

\[ \hat{\theta}_{t+1} = \mathbb{E}\{\theta_{t+1} | \gamma_{t+1}\} = \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 \ p(\theta | \gamma_{t+1}) d\theta \]

minimum error

\[ \hat{\theta}_{t+1} = \text{median} \ p(\theta_{t+1} | \gamma_{t+1}) \]

For non-linear and non-stationary spatial systems, representations analogous to the foregoing are

\[ \gamma_t = A^{-1}(Z,U_t) B(Z,U_t) U_t \]

with

\[ A(Z,U_t) = 1 + A_1(U_t) Z + \ldots + A_p(U_t) Z^p \]

\[ B(Z,U_t) = B_0(U_t) + \ldots + B_q(U_t) Z^q \]

where the A's and B's are functions of the input terms over the entire N-regional system; and
\[ Y_t = A^{-1}(Z,t) B(Z,t) U_t \]

with

\[ A(Z,t) = 1 + A_1(t) Z + \ldots + A_p(t) Z^p \]

\[ B(Z,t) = B_0(t) + B_1(t) Z + \ldots + B_q(t) Z^q \]

where the polynomials \( A(.,.) \) and \( B(.,.) \) are functions of time variables. With regard to these it is highly relevant to discuss identification and estimation techniques.

Spatial identification problems involve choosing the most appropriate system representation from the family of possible systems that might govern the process under study. The identification of truly spatial processes depends largely upon the same methods as the identification of time series, namely, estimation of spatial correlation and frequency response functions. Bennett gives Besag's \((k,l)\)-order unilateral equation:

\[ Y_{ij} = \sum_{m=1}^{k} a_m Y_{i-m,j} + \sum_{n=1}^{l} b_n Y_{i,j-n} + \epsilon_{ij} \]

and considers lag-correlation structure and parameter estimation for approximations developed by a number of authors whose work he cites. The approximation techniques become inordinately involved, invoking
"local Markov conditions"
"nearest neighbor properties"
"internal site values and boundary values"
and solutions in terms of Bessel functions.

My personal feeling about the foregoing comes in two parts and is based on a number of years work in this field:

i) The "discrete arithmetic representation" which envisages a grid or a regular lattice may be the only practical representation for socioeconomic data which is accumulated and stored in geographic bins: townships, counties, states, etc. Furthermore, in a densely populated, tightly integrated society it may be practical to equate concepts of 1st, 2nd, 3rd...order neighbors with actual geographic configurations of political subdivisions.

I don't quarrel with the use or usefulness of such a discrete representation in these contexts (except that they generally follow with an isotropic assumption). However the danger lies in trying to utilize the whole package out of context.

ii. The "continuous arithmetic representation" precludes the arbitrary elements of: definition of geographic (sub)regions, the lumping of observations within sub(regions), and the designation of types of neighbors on the basis of regular (generally istropic) spatial configurations. In addition, the extraction of ensemble relationships and stochastic dynamic parameterizations is generally analytically clean--without requiring unrealistic approximations.
The continuous-arithmetic representations for non-stationary and non-linear processes are the following:

\[
Y_{t,r} = \int_{-\infty}^{t} \int_{R} S(t',r') U_{t',r'} \, dt' \, dr'
\]

\[
Y_{t,r} = \int_{-\infty}^{t} \int_{R} S(U_{t',r'}) U_{t',r'} \, dt' \, dr'
\]

Two very practical points to be noted here, in this regard, are:

1) the system identification phase of analysis, which again relies on estimation of spatial correlation structure, or frequency response functions, does not require equally spaced observations;

and 2) the technique used in modeling spatial structure, with irregular data arrays, minimizes or precludes spatial frequency aliasing.
REFERENCES


VI. NOTES ON FIELD-DEPENDENT STATISTICAL OBJECTIVE ANALYSIS

H. J. Thiébaux

OVERVIEW AND GENERAL CONSIDERATIONS

In section V we discussed, in a general way, the theoretical basis for optimal prediction of one or more variables as a linear function of observed variables which are distributed in both space and time. An important point arising from that discussion was that for a space-time distributive process which is not stationary in time, the analysis model should be capable of allowing for changes in the 'process' linking analyzed variables to both observed, and perhaps, unobserved variables. This leads to the concept of 'dynamic-parameter mapping,' such as represented by some forms of Kalman filtering. The question now is: can such a concept be usefully applied for the analysis of meteorological fields, where some form of 'situation dependent' analysis is often clearly desirable? Some work in this area has already been reported by Ghil et al. (1981), though only in the context of a simple theoretical model of a predictor-corrector system applied to meteorological variables.

In this section, we consider briefly the possibilities for practical field-dependent analysis schemes which could go some way towards meeting the need for a dynamic, or at least 'data-adaptive' approach to spatial field estimation. We concentrate on the type of spatial prediction common in meteorological analysis, where
observations, or their departures from deterministic forecasts, are considered as stochastic variables, so that the spatial or temporal behavior is identified with variations in the joint probability distribution associated with these variables.

Thus we are concerned with a dynamical system evolving in spatial dimensions, which is modeled and forecast with (truncated) equations of dynamics; and our specific concern here will be with that part which is not explained by a given dynamical representation. We will refer to the latter as the anomaly, or first-guess-anomaly, and understand that the stochastic properties of interest are those of the anomaly field—whether the 'first guess' is as refined as that from a twenty-two level spectral model forecast, or as simple as a ten-year climatological field. Eventually we may wish to refer to the latter distinction explicitly. However our initial concerns do not depend on the definition of the first-guess-field.

Notationally we distinguish between values of the field at locations from which we obtain observations: denoted by \( X(s) \) with spatial-array-index, and values of the field at locations for which we construct estimates: denoted by \( Y(s+\Delta s) \). These signify the stochastic anomalies of the spatially-coherent system. The arguments are a convenient shorthand: \( s \) indicating the set of observing locations and \( s+\Delta s \) the distinct set of locations of estimates.

The theory of statistical estimation utilizes the collective co-variability of the whole field in the formulation of an analysis
algorithm. When this is known and can be described in terms of a joint distribution of the vectors $X(s)$ and $Y(s+\Delta s)$, then it is possible to derive the conditional distribution of $Y$ given $X$, say

$$F_{\Delta s}(y|x) = \Pr[Y(s+\Delta s) \leq y | X(s) = x]$$

and from this to determine the mean, or expectation of $Y$, for any given array of observations, as

$$E[Y(s+\Delta s) | X(s) = x] = \int y \, dF_{\Delta s}(y|x).$$

Consequently this is not a static value nor, in principle, is it a static combination of observed values. Rather, the algorithm itself may be conditional on time-coincident knowledge of the field at the points of the observing array. Since the conditioning is on observed values of the field at proximal locations, the formalism of objective analysis by conditional expectation is the spatial analogue of Kalman filtering; this has a good deal to recommend its use to estimate values of the field at points where the field is not observed. At this workshop we have discussed several recent studies which compare the accuracies of differing objective analysis techniques; e.g., Creutin and Obled (1982), Franke and Gordon (1983), Rasmussen and Schlatter (1983). As some of these nicely illustrate there are alternative criteria for evaluating the results of objective analyses. However, insofar as one is concerned with the
traditional goal of analyzing to "minimize the ensemble mean square difference between the estimated field and the observed field", the field-dependent conditional expectation described above is the unique optimal estimator. Typically, the fact that it is unbeatable by this criterion is overlooked in discussions of experiments which compare objective analysis techniques—probably because its operational parameterizations generally give it a form which is insensitive to inhomogeneous stochastic field relationships. That is, the optimal estimator is rendered operationally sub-optimal (in comparison with its own potential) and it becomes technically possible for another objective analysis procedure to better its performance.

The most common use of objective analysis by conditional expectation makes field structure assumptions which imply that the expected value is linear in the proximal observations (or simply restricts algorithm choice to linear combinations of observed values). In this frame, the weighting scheme for the optimal estimator is determined by the field covariance structure, and the estimation algorithm is the so-called optimal statistical objective analysis (o.s.o.a.). Here the key to accurate estimation, in the ensemble mean square sense, is accurate representation of the spatial lag-covariances. The necessity of modeling covariance structure creates considerable latitude for field smoothing which may be counter to the goals of the analysis. This may be appreciated by referring to the following theoretical hierarchy of parameterizations of spatial relationship: \( \text{Cov}[X(s_1), X(s_2)] \)
Since all of the above are models of covariance structure, which must be parameterized by fitting to observed field values, all are approximations; however they are necessarily approximations of decreasing fidelity. The goal commensurate with increasing analysis resolution, i.e., the extent to which the analyzed field represents the true field, is to put as much field structure as is practical into the estimation scheme. We acknowledge that what is "practical" will vary with the analysis objective and with computational requirements for analysis production: two factors which are generally in opposition, due to fiscal and time constraints. However we think it is also important to consider the "trade-off" the other way around, namely, in determining the resource requirements for the production of analyses which will be of practical significance under specified synoptic conditions.
Refinements of o.a. schemes have led to adoption of anisotropic, ensemble average correlation models, in some instances, with concomitant improvement in analysis accuracy. However the concerns of the workshop have centered on an objective with more far-reaching potential impact; namely, a "quantum jump" to the inclusion of field-dependent covariance structure. This would admit the influence of synoptic conditions on the structure of the analysis algorithm, and this has vastly more potential for "getting the analysis right", e.g., with regard to placement of fronts.

ON REPRESENTING COVARIANCE STRUCTURE

Field covariance functions which are analytically derived from spatial stochastic representations of the anomaly fields have demonstrable advantages for multivariate objective analyses. Two notable advantages discussed at the workshop are the implicit positive-definiteness of the full covariance array, required by the o.s.o.a., and improvements brought to large-scale analyses in relatively data sparse regions.

A covariance tensor \( \mathbf{T} \) which is rigorously derived from a spatially autoregressive representative of a multivariate anomaly field necessarily satisfies the requisite properties of true covariance structure—importantly among them, positive definiteness of the \( \mathbf{T} \) tensor. The proof that this is true is simple and direct, and we give it here for completeness.
With $X$ denoting an array of first-guess anomalies and $\langle \cdot \rangle$ indicating the relevant stochastic ensemble average, we may write
\[
T = \left( C(X_j, X_k) \right) = \langle XX^T \rangle.
\]

To show that $B^T T B$ is positive for any non-zero column vector $B$ (if $X$ is not strictly constant in any of its components) we invoke the fact that any linear transformation of $X$ will have positive variance. This includes $B^T T X$ for which the variance is
\[
0 < \langle (B^T X)^2 \rangle = \langle B^T X X^T B \rangle = B^T T B.
\]

Consequently $T$ must be positive definite—provided only that it is the genuine covariance matrix for a stochastic field.

The evidence of improvement brought to large-scale analyses in regions for which routine observation cannot provide information on all variables on all standard pressure levels (principally over oceans) is specific to forecast error analyses. While it comes somewhat indirectly from an analysis error assessment project (Thiébaut, 1980), it is particularly interesting in that it relates to the multivariate aspect of a large-scale objective analysis. The latter project compared assessments of anticipated analysis error changes (associated with the removal of the Ocean Weather Ship P) for different statistical guidance systems. For each of the two basic geopotential correlation models, the auto and cross (wind-wind
and height-wind) correlation models were obtained by geostrophic derivation. The result which is relevant to mention here is the following: While the two geopotential correlation functions (parameterized to fit CMC observed forecast-error correlations) were very similar in appearance over the (substantial) distance ranges of the NE Pacific assessment region, multivariate analysis error differences were markedly different. The nature of the differences clearly suggests a marked advantage to the model derived from the spatial autoregressive representation of the anomaly fields, indicating that its geostrophic derivatives appropriately bring more distant (but significantly correlated) winds and heights into height and wind analyses. Thus, in considering accuracy of covariance representations, we need to consider more than the "master model" in relation to multivariate analyses. This has been discussed elsewhere by researchers concerned with data assimilation for operational forecasting. Specifically, Schlatter (1975), Bergman (1979), and Lorenc (1981) have detailed the geostrophic wind correlation structure implicit in the correlation models of their analysis schemes. In addition Rutherford (1973) considered the separate but parallel issue of the improvement brought to a geopotential analysis by the inclusion of wind data, and vice versa. However they do not compare models in these respects; there does not seem to be more direct evidence of the consequences of particular covariance structure model(s) on the weighting of dissimilar variables in a multivariate analysis.
ON UTILIZATION OF THE UNQUALIFIED OPTIMAL ESTIMATOR

We briefly consider the potential for using conditional expectation of field anomaly values with no linearity restrictions. This returns attention to the unrestricted field-dependent optimal estimator $E[Y(s+\Delta s)|X(s) = x]$. Its utilization clearly requires that stochastic transition structure be identifiable in terms of proximal observed anomalies. Although this is not yet a well developed technology, some initial work with time-transition probability structure (Julian and Thébault, 1983) has been very encouraging in this respect; we have a basic format for extension to the dimensions of spatial arrays. Thus pursuit of the goal of putting synoptic field structure into the objective estimation scheme appears promising.

The preliminary work referred to was designed to investigate the feasibility of using covariance functions which are derived from observed, conditional transition probability structure in order to accommodate known inhomogeneities and thus improve the performance of the linear statistical objective analysis. However, since a parametric stochastic model has been shown to successfully represent field-dependent transition distributions, it is possible to forego the linear estimation algorithm and use in its stead the unqualified, objective analysis by conditional expectation.
OVERVIEW AND CONCLUSIONS

For the purposes of the workshop, we have specifically defined field-dependent estimation as conditional estimation given values of proximal field observations, and focused on estimation schemes as adaptive algorithms which may incorporate time- or space-adjacent information about an observed field into the structure of the algorithm which generates the analysis. Although the need for the development of practical techniques for analysis of meteorological fields, where the analysis algorithms depend on the most recent or proximal observed states, is perhaps best appreciated in relation to mesoscale modeling and nowcasting, it will be a requirement for increasing analysis accuracy, on all scales. Furthermore (and, heuristically, in an inverse sense) field-sensitive stochastic structure models are a requirement of network design. This aspect has received only slight mention in this section; however it is a clear need in the definition of network requirements of analysis definition. Referring back to the opening session of the workshop:

Observing array scales are determinants, in a limiting sense, of the resolution achievable by any analysis algorithm. Consequently, the goals of an analysis, in the detection and representation of weather events, set requirements on time and space scales of observing systems.
Assessment of the requisite balance between analysis resolution and the scales of the observing system that will contribute data to the analysis creates additional, explicit demands for accurate field-sensitive analysis techniques.
REFERENCES


