CONTENTS

Estimation of Parameters in Models for Cesium Beam Atomic Clocks

Peter V. Tryon* and Richard H. Jones**

National Bureau of Standards, Boulder, CO 80303

This paper is intended to serve as an introduction to the use of the Kalman filter in modeling atomic clocks and obtaining maximum likelihood estimates of the model paramaters from data on an ensemble of clocks. Tests for the validity of the model and confidence intervals for the parameter estimates are discussed. Techniques for dealing with unequally spaced and partially or completely missing multivariate data are described. The existence of deterministic frequency drifts in clocks is established and estimates of the drifts are obtained.

Key words: atomic clocks; Kalman filter; maximum likelihood; missing observations; random walks; state space; time series analysis; unequally spaced data.

1. Introduction

The recursive updating of least squares estimates that would later become a special case of the Kalman "filter" was apparently known to Gauss [1],[1] and published in the modern statistical literature by Plackett [2]. Kalman's contribution [3] was the generalization to dynamic systems. Kalman's filter has found immense application in diverse areas of engineering. In classical applications, recursive estimates are obtained of the "state" of the system but parameters appearing elsewhere in the state space representation for the system and the mathematical model for the system must be considered known.

In our application, Kalman's recursive equations allow us to compute the likelihood function for given values of parameters occurring anywhere in the state space representation. Nonlinear optimization techniques can then be used to find the maximum likelihood estimates. The recursive residuals, or innovations, may be examined to judge the adequacy-of-fit of the model, and generalized likelihood ratio tests may be used to test the significance of model parameters such as frequency drifts and obtain confidence intervals for the estimated parameters. Kalman filter techniques make it quite easy to deal with unequally spaced and completely or partially missing multivariate data. We describe these methods in this paper.

The next article in this issue, "Estimating Time from Atomic Clocks," by Jones and Tryon [4] describes statistical procedures for detecting clock errors and allowing for the insertion or deletion of clocks in the ensemble during the parameter estimation process. It also describes the development of a time scale algorithm based on the model developed in this paper.

2. Models For Atomic Clocks

A cesium atomic clock is a feedback control device whose frequency locks onto the fundamental resonance of the cesium atom at (ideally) 9,192,631,770 Hz, which defines the second. Frequency bias due to fundamental instrumentation limitations and environmental effects are determined by calibration against primary frequency standards. Stochastic fluctuations arise from shot noise in the cesium beam and the probabilistic nature of quantum

---

mechanical transition rates which cause white noise in the frequency of the clock which is integrated into a random walk in time. In addition, empirical studies have demonstrated that frequency wanders independently as a random walk, introducing the integral of a random walk into time.

Finally, there is considerable international debate over the inclusion of linear drift in frequency to account for break-in and aging. Such drifts could be strictly constant over a long time period, or allowed to change slowly as a third random walk. The methods developed in this paper provide, for the first time, valid statistical tests of these hypotheses.

The proposed stochastic model for a clock's behavior is

$$
\begin{aligned}
x(t) &= x(t-1) + \delta(t)\,y(t-1) + \tfrac{1}{2}\,\delta^2(t)w(t-1) + \varepsilon(t) \\
y(t) &= \hphantom{x(t-1) + } y(t-1) + \hphantom{\tfrac{1}{2}}\delta(t)w(t-1) + \eta(t) \\
w(t) &= \hphantom{x(t-1) + \delta(t)\,y(t-1) + } w(t-1) + \alpha(t)
\end{aligned}
\tag{2.1}
$$

where

$x(t)$ is the clock's deviation from "perfect" time at sample point t in nanoseconds;

$\delta(t)$ is the time interval in days between sample points t and $t-1$;

$y(t)$ is the clock's frequency deviation from perfect frequency at sample point t in nanoseconds/day;

$w(t)$ is the drift in frequency in nanoseconds/day^2;

$\varepsilon(t)$, $\eta(t)$, and $\alpha(t)$ are mutually independent white zero mean Gaussian random variables with standard deviations $\sqrt{\delta(t)}\sigma_\varepsilon$, $\sqrt{\delta(\tau)}\sigma_\eta$, and $\sqrt{\delta(t)}\sigma_\alpha$, respectively, in nanoseconds, nanoseconds/day, and nanoseconds/day^2. The $\sqrt{\delta(t)}$ arises because the variance of a random walk is proportional to the time interval.

If $\sigma_\alpha = 0$, $w(t)$ is a constant, w, representing a deterministic linear trend. If both $\sigma_\alpha = 0$ and $w = 0$, the model is drift-free. The purpose of this study is to evaluate the validity of these models, and to obtain generalized likelihood ratio tests for the significance of the drift parameters and maximum likelihood estimates of the parameters σ_ε, σ_η, and w (or σ_α, if appropriate), for each clock in the ensemble.

3. The Kalman State Space Model for a Clock Ensemble

The combination of clocks into a Kalman state space model will be explained by considering the special case of a three-clock model. The *state transition equation* is

$$
\begin{bmatrix} x_1(t) \\ y_1(t) \\ w_1(t) \\ x_2(t) \\ y_2(t) \\ w_2(t) \\ x_3(t) \\ y_3(t) \\ w_3(t) \end{bmatrix}
=
\begin{bmatrix}
1 & \delta(t) & \tfrac{1}{2}\delta^2(t) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & \delta(t) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \delta(t) & \tfrac{1}{2}\delta^2(t) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \delta(t) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \delta(t) & \tfrac{1}{2}\delta^2(t) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \delta(t) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} x_1(t-1) \\ y_1(t-1) \\ w_1(t-1) \\ x_2(t-1) \\ y_2(t-1) \\ w_2(t-1) \\ x_3(t-1) \\ y_3(t-1 \\ w_3(t-1) \end{bmatrix}
+
\begin{bmatrix} \varepsilon_1(t) \\ \eta_1(t) \\ \alpha_1(t) \\ \varepsilon_2(t) \\ \eta_2(t) \\ \alpha_2(t) \\ \varepsilon_3(t) \\ \eta_3(t) \\ \alpha_3(t) \end{bmatrix}
$$

where subscripts indicate the clocks.

4

In more compact form this is

$$X(t) = \Phi(t)\,X(t-1) + U(t) \ , \ t = 1, 2, ...,$$
(3.1)

where $X(t)$ is the state vector and $\Phi(t)$ is the state transition matrix. The stochastic elements of the transition from one state to the next are contained in $U(t)$.

The clocks are never read directly, but, using sophisticated electronics, the time differences between clocks can be read with a precision of less than a nanosecond. "Perfect" time, which we can never observe, cancels out, leaving an observation of the clock error differences. $z_1(t) = x_1(t) - x_2(t)$ and $z_2(t) = x_1(t) - x_3(t)$. We write the *observation equation* in matrix form as

$$
\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} =
\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}
\begin{bmatrix} x_1(t) \\ y_1(t) \\ w_1(t) \\ x_2(t) \\ y_2(t) \\ w_2(t) \\ x_3(t) \\ y_3(t) \\ w_3(t) \end{bmatrix}
+ \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}
$$

or

$$Z(t) = H(t)\,X(t) + V(t) \ , \ t = 1, 2, ...$$
(3.2)

$Z(t)$ is the vector of observed clock time differences and $H(t)$ is the observation matrix. $V(t)$ is a vector of observational errors. We write $H(t)$ as a time-varying matrix because, in the multivariate case, partially missing data is represented by deleting rows from $H(t)$ thus changing the dimension of $Z(t)$, $H(t)$, and $V(t)$. In some cases it is possible to recover from a problem with the reference clock by defining a new (temporary) reference and changing the patterns of ± 1 and zeros in $H(t)$. This is described in detail in Jones and Tryon [4].

For small $\delta(t)$ the covariance matrix of the transition errors, $U(t)$, is of the diagonal form $\delta(t)Q$ where Q is

$$
Q =
\begin{bmatrix}
\sigma_{\epsilon 1}^2 & & & & & & & & \\
& \sigma_{\eta 1}^2 & & & & & & & \\
& & \sigma_{\alpha 1}^2 & & & & & 0 & \\
& & & \sigma_{\epsilon 2}^2 & & & & & \\
& & & & \sigma_{\eta 2}^2 & & & & \\
& & & & & \sigma_{\alpha 2}^2 & & & \\
& & & & & & \sigma_{\epsilon 3}^2 & & \\
& 0 & & & & & & \sigma_{\eta 3}^2 & \\
& & & & & & & & \sigma_{\alpha 3}^2
\end{bmatrix}
$$
(3.3)

Note that the matrix Q contains all the parameters to be estimated except w.

If the time offset of each clock is measured to the nearest nanosecond and is otherwise error free, the observational error vector $V(t)$ has covariance matrix

$$
R(t) = \begin{bmatrix} 1/12 & 0 \\ 0 & 1/12 \end{bmatrix},
$$
(3.4)

since the variance of a uniformly distributed random variable from -0.5 to 0.5 is $1/12$. The dimension of $R(t)$ will also vary with time if there are partially missing data.

4. The Kalman Recursion

First, we adopt some notation. By $X(t|s)$, $s \leq t$ we denote the best estimate of the state vector at sample time t, based on all the data up to and including sample time s. We define $P(t|s)$ to be the covariance matrix of $X(t|s)$. Similarly, $Z(t|s)$ will be the predicted observation at time t based on all the data up to time s.

The recursion proceeds through the following steps (Kalman [3]).

1. Given the state vector and its covariance matrix at any sample time, say t, we can predict the state vector one observation interval into the future using the transition eq (3.1). The prediction is

$$X(t+1|t) = \Phi(t+1) \, X(t|t) \tag{4.1}$$

since the expected value of the stochastic component of the transition is zero. The covariance matrix of the prediction is

$$P(t+1|t) = \Phi(t+1) \, P(t|t) \, \Phi'(t+1) + \delta(t+1)Q, \tag{4.2}$$

where the first term on the right is the covariance matrix of the linear operation in the state transition and the second term is the covariance of the additive stochastic component.

2. From the observation matrix and the predicted state we can predict the next observation.

$$Z(t+1|t) = H(t+1) \, X(t+1|t). \tag{4.3}$$

3. We next compare the prediction with the actual data (when it becomes available in a real-time application). The difference is the recursive residual, or innovation,

$$I(t+1) = Z(t+1) - Z(t+1|t), \tag{4.4}$$

which has covariance matrix

$$C(t+1) = H(t+1) \, P(t+1|t) \, H'(t+1) + R(t+1). \tag{4.5}$$

The residuals are critically important. Later, they will be used for checking the fit of the model and estimating the parameters. For now, we will use the information in the residual to complete the recursion.

4. Let

$$\Delta(t+1) = P(t+1|t) \, H'(t+1) \, C^{-1}(t+1). \tag{4.6}$$

This is sometimes called the Kalman gain.

Using the Kalman gain we can update the predicted state and its covariance

$$X(t+1|t+1) = X(t+1|t) + \Delta(t+1) \, I(t+1) \tag{4.7}$$

and

$$P(t+1|t+1) = P(t+1|t) - \Delta(t+1) \, H(t+1) \, P(t+1|t). \tag{4.8}$$

This completes the recursion.

5. Maximum Likelihood Modeling and Estimation

In this section we will compare the classic use of the Kalman filter with our method of application. In the classical application, unknown parameters to be estimated can appear only in the state vector. All other quantities which appear in the recursion must be known. A single pass of the recursion through the data (often in real time) provides the latest estimate of the state vector and its covariance at each point in time. This is exactly how the Kalman filter is applied to forming a time scale. Given the model and estimates of the parameters, the recursion gives the updated state vector after each new observation. The first, fourth, and seventh, etc., elements of the state vector are estimates of the clock errors, from which, along with their uncertainties from the state covariance matrix, we can estimate time.

Our interest, however, is in evaluating how well the clock model fits real data, and in estimating the parameters. To do this, we make use of two facts concerning the residuals, $I(t)$, $t = 1, 2, ..., N$.

1. If the model is correct, and the recursion is run through the data with the correct values of the unknown parameters, the residuals will form Gaussian white noise series. Standard statistical tests can then be used to determine if the residuals are in fact Guassian white noise. If not, the structure of the residuals can often give some clues as to the defects in the model.

2. From the residuals and their covariance matrix, obtained by running the recursion with *any* fixed values of the unknown parameters, the likelihood function for those parameter values can be computed. Specifically, the –2 ln likelihood function (ignoring the additive constant) is

$$L = \sum_t [\ln |C(t)| + I'(t) \, C^{-1}(t) \, I(t)]. \tag{5.1}$$

The maximum likelihood estimates of the parameters are those for which L is minimized. Starting from initial guesses, we can use iterative nonlinear function minimization methods to find the estimates. Evaluating the function L for new trial values of the parameters requires one pass of the Kalman recursion through the data. Note that the parameters to be estimated may appear anywhere in the structure of the Kalman recursion.

6. Computational Procedures

It is not necessary to invert the residual covariance matrix, $C(t)$, when computing the likelihood function or updating the predicted state vector and its covariance. If the upper triangular portion of $C(t+1)$ is augmented by $H(t+1)P(t+1|t)$, and $I(t+1)$ to form the partitioned matrix

$$[C(t+1) \mid H(t+1) \, P(t+1|t) \mid I(t+1)], \tag{6.1}$$

a single call to a Choleskey factorization routine (Graybill [5], p. 232) which factors

$$C(t+1) = T'(t+1) \, T(t+1), \tag{6.2}$$

where $T(t+1)$ is upper triangular, will replace the partitioned matrix by

$$[T(t+1) \mid [T'(t+1)]^{-1} H(t+1) \, P(t+1|t) \mid [T'(t+1)]^{-1} I(t+1)]. \tag{6.3}$$

If we call the parts of the new partitioned matrix

$$[T(t+1) \mid B(t+1) \mid D(t+1)], \tag{6.4}$$

the updated state and state covariance matrices become

$$X(t+1|t+1) = X(t+1|t) + B'(t+1) D(t+1) \qquad (6.5)$$

and

$$P(t+1|t+1) = P(t+1|t) - B'(t+1) B(t+1). \qquad (6.6)$$

The -2 ln likelihood function becomes

$$L = \sum_t \left(\sum_i \ln t^2_{ii}(t) + D'(t) D(t) \right) \qquad (6.7)$$

where $t_{ii}(t)$ are the diagonal elements of $T(t)$.

Missing data are easily handled by the Kalman filter procedure. However, there are two distinct situations. First, all the data may be missing at one or several successive observation times, and second, occurring only in multivariate data, only part of the data may be available on a given occasion. For example, one or more, but not all, of the clock differences are missing on that occasion.

The first case is most easily dealt with. In the present model, the unequal spacing feature can account for any size gap in the data by proper choice of $\delta(t)$. This, however, is a feature of random walk based models and may not apply to other models. A more general procedure for missing data is the following: Suppose $k \geq 1$ successive observations are missing. Simply repeat step one in the recursion k additional times to obtain the predicted state, $X(t+k+1|t)$, and its covariance, $P(t+k+1|t)$, for the next observation time for which there is data. Then proceed with the remainder of the recursion. There is no residual and no contribution to the likelihood function when there is no data (Jones [6]).

In the second case, partial loss of data, i.e., missing elements of the data vector $Z(t)$, can be represented by eliminating the corresponding rows of the observation matrix, $H(t)$. Similarly, rows of the observational error vector $V(t)$ and rows and columns of the error covariance matrix $R(t)$ causes no problem with the remaining steps of the recursion, which proceed normally. The information available in the data is propagated through the recursion to correctly update the state vector prediction and its estimated covariance matrix.

The matrix operations in the recursion are, for reasons of efficiency, best coded to take advantage of the specific structure (many zeros) of each matrix. Using available general purpose subroutines to perform the required matrix operations would increase the cost of computer resources enormously. These calculations can be adapted to a time-varying $H(t)$. One approach is to use indicator variables for missing data to form the various reduced matrices directly. Another approach is to note that pre- and/or post-multiplying by $H(t)$ with rows missing is equivalent to performing the operation with the original H matrix and then eliminating the appropriate rows and/or columns of the product matrix and closing it up. For example, the residual covariance matrix

$$C(t+1) = H(t+1) P(t+1|t) H'(t+1) + R(t+1)$$

with, say, the second and fourth elements of the data vector missing can be obtained by using the original H and R but then eliminating the second and fourth rows and columns from $C(t+1)$. It is straightforward to write general purpose subroutines to eliminate rows and/or columns of a matrix and close up the spaces. Again, this can be done in response to indicator variables denoting missing observations.

To start the recursion the initial state $X(0|0)$ and its covariance $P(0|0)$ are needed. Both depend on the application. The time states could be set to zero and the clocks set to agree with International Atomic Time as defined by the BIH. The corresponding variances on the diagonal of $P(0|0)$ would be determined by the accuracy of the intercomparison. If a strictly internal application such as parameter estimation or a time interval measurement was desired, the time states could be set to agree with the difference readings at $t = 0$, with observational error variances set to the appropriate value for roundoff error. The frequency states might be measured by intercomparison with the frequency standards and the variance determined from the measurement uncertainty. When a measurement is not possible a best guess at the frequency offset can be used with a corresponding guess at its uncertainty used for the variance.

8

7. Data Analysis

The data used for this analysis were collected over a period of 333 days beginning on February 16, 1979, during which seven clocks listed below were in essentially uninterrupted normal operation. Clock differences were recorded daily to the nearest nanosecond.

Clock Number	Identification
601 (Reference)	HP model 5061A option 004 serial #601
167	HP model 5060A serial #167
137	HP model 5060A serial #137
1316	HP model 5061A option 004 serial #1316
323	HP model 5061A 004 serial #323
324	HP model 5061A 004 serial #324
8	Frequency and Time System model 400 serial #8

In addition to occasional variability in the time of data collection, two successive days were missing entirely. On three other occasions, single clocks had obvious read errors which when discarded resulted in partially missing data. Several dozen other observations which had been flagged as possible errors by the ad hoc procedure then in use were examined visually, judged to be not serious, and retained in the analysis.

Table 1 gives the results of the computations. First we note that the inclusion of deterministic trends (model II) results in a drop of 41 in the $-2 \ln$ likelihood function value over the no-trend model (model I). The reduction is distributed as $X^2(6)$ under the hypothesis of model I, so this is highly significant. There is less than one chance in a thousand that this is a spurious result. Three of the seven clocks had significant trends. The inclusion of random walk drifts (model III), however, makes no improvement. Furthermore, the estimated values for the σ_a's are all approximately zero and the estimated initial values of the drift, $w(0)$, are about the same as the deterministic drifts. The conclusion is that several frequency trends are present in this data and that they are constant over the year, rather than wandering as a random walk. It should be noted that one of the drifting clocks had just begun operation while the remaining two were nearing their end of life and failed shortly after the test period. The remaining clocks were in their "middle years." The standard errors of the parameter estimates given in the table are obtained from the estimated covariance matrix as approximated by twice the inverse of the Hessian of the $-2 \ln$ likelihood function, L.

In model II it is necessary to constrain the solution so that the sum of the drifts is zero. This means that we cannot detect a common drift from clock difference readings. Because of this constraint, a standard error for clock #8 is not available. Due to the large cost of computing the 28 x 28 Hessian for model III, which we have rejected anyway, the standard errors of the estimates have not been computed.

Figures 1a-f are integrated periodograms of the six series of residuals. The parallel lines represent five percent significance test limits for white noise. That is, only five percent of actual white noise series so tested will wander outside the lines. In this data set, only one pair of clocks (601 minus 167 two) show a slight deviation from white noise. We conclude that this model fits very well. Figures 2a-f are histograms of the six residual series. Also given are three different tests for normality. All the series are well approximated by the normal (Gaussian) distribution which verifies that the assumed likelihood function was reasonable.

We note in conclusion that, having selected the constant drift model, we could reduce the dimension of the state space model by one-third by rewriting the clock model as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 1 & \delta(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(t-1) \\ y(t-1) \end{bmatrix} + \begin{bmatrix} \delta^2(t)/2 \\ \delta(t) \end{bmatrix} w + \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}$$

The state prediction in the first step of the recursion then becomes

$$X(t+1|t) = \Phi(t+1) X(t|t) + A(t+1)$$

9

where

$$A(t) = \begin{bmatrix} \dfrac{\delta^2(t)\ w_1}{2} \\[2mm] \delta(t)w_1 \\[2mm] \dfrac{\delta^2(t)\ w_2}{2} \\[2mm] \delta(t)w_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} ,$$

and $X(t)$ and $\Phi(t)$ are redefined from the clock model in the obvious way. This model would be suitable for estimation purposes but not for a time scale algorithm. With the drift in the state vector we can enter the uncertainty of its estimation in the initial state covariance matrix. It is then propagated through to the uncertainty in time. The simplified model above assumes w is known without error.

TABLE 1. Results of data analysis.

Clock Number	Model I No Trend $L = 10609.1$				Model II Deterministic Trend $L = 10567.8$				Model III Random Walk Drift $L = 10567.7$	
	Param. Est.*		S.E		Param. Est.*		S.E.		Param. Est.*	
601	$\hat{\sigma}_\epsilon$	=	7.42	.32	$\hat{\sigma}_\epsilon$	=	7.46	.32	$\hat{\sigma}_\epsilon$	= 7.47
	$\hat{\sigma}_\eta$	=	.86	.24	$\hat{\sigma}_\eta$	=	.44	.26	$\hat{\sigma}_\eta$	= .48
					$\hat{w}$	=	.152	.038	$\hat{w}(0)$	= .152
									$\hat{\alpha}_a$	= 4.1×10^{-5}
167	$\hat{\sigma}_\epsilon$	=	13.45	.56	$\hat{\sigma}_\epsilon$	=	13.45	.56	$\hat{\sigma}_\epsilon$	= 13.46
	$\hat{\sigma}_\eta$	=	1.15	.39	$\hat{\sigma}_\eta$	=	1.11	.36	$\hat{\sigma}_\eta$	= 1.07
					$\hat{w}$	=	.052	.061	$\hat{w}(0)$	= .054
									$\hat{\sigma}_a$	= 3.7×10^{-5}
137	$\hat{\sigma}_\epsilon$	=	10.03	.45	$\hat{\sigma}_\epsilon$	=	10.04	.45	$\hat{\sigma}_\epsilon$	= 10.06
	$\hat{\sigma}_\eta$	=	1.71	.36	$\hat{\sigma}_\eta$	=	1.60	.36	$\hat{\sigma}_\eta$	= 1.57
					$\hat{w}$	=	.179	.081	$\hat{w}(0)$	= .178
									$\hat{\sigma}_a$	= 2.2×10^{-4}
1316	$\hat{\sigma}_\epsilon$	=	3.61	.24	$\hat{\sigma}_\epsilon$	=	3.62	.25	$\hat{\sigma}_\epsilon$	= 3.59
	$\hat{\sigma}_\eta$	=	1.29	.24	$\hat{\sigma}_\eta$	=	1.36	.24	$\hat{\sigma}_\eta$	= 1.38
					$\hat{w}$	=	-.017	.070	$\hat{w}(0)$	= -.015
									$\hat{\sigma}_a$	= 1.3×10^{-4}
323	$\hat{\sigma}_\epsilon$	=	3.27	.24	$\hat{\sigma}_\epsilon$	=	3.53	.22	$\hat{\sigma}_\epsilon$	= 3.52
	$\hat{\sigma}_\eta$	=	1.54	.21	$\hat{\sigma}_\eta$	=	.73	.20	$\hat{\sigma}_\eta$	= .70
					$\hat{w}$	=	-.313	.046	$\hat{w}$	= -.314
									$\hat{\sigma}_\epsilon$	= 4.9×10^{-5}
324	$\hat{\sigma}_\epsilon$	=	3.30	.25	$\hat{\sigma}_\epsilon$	=	3.30	.25	$\hat{\sigma}_\epsilon$	= 3.30
	$\hat{\sigma}_\eta$	=	1.42	.23	$\hat{\sigma}_\eta$	=	1.40	.22	$\hat{\sigma}_\eta$	= 1.41
					$\hat{w}$	=	.035	.072	$\hat{w}(0)$	= .035
									$\hat{\sigma}_\epsilon$	= 7.9×10^{-5}
8	$\hat{\sigma}_\epsilon$	=	9.08	.43	$\hat{\sigma}_\epsilon$	=	9.09	.43	$\hat{\sigma}_\epsilon$	= 9.09
	$\hat{\sigma}_\eta$	=	2.68	.39	$\hat{\sigma}_\eta$	=	2.65	.39	$\hat{\sigma}_\eta$	= 2.66
					$\hat{w}$	=	-.088	–	$\hat{w}(0)$	= -.090
									$\hat{\sigma}_r$	= 3.9×10^{-5}

*Units are in nanoseconds.

8. Conclusion

The Kalman filter is a powerful tool for maximum likelihood model fitting and parameter estimation in time series analysis. We have established the validity of the proposed models for clock behavior and obtained precise estimates of clock parameters. The existence of deterministic frequency drifts over a period of a year has been demonstrated in clocks at the beginning and end of their life span.

The primary disadvantage of these methods is the cost of function evaluation in the nonlinear optimization process. The analysis of a year of data on seven clocks takes 10 minutes on a CDC Cyber 170/750 computer and 10 hours on the DEC 11/70. Use of state-of-the-art optimization codes is essential. We use an optimization package especially designed for maximum likelihood applications. It is adapted from the package written by Weiss [7], based on algorithms by Dennis and Schnabel [8]. The code used on the CDC Cyber 170/750 is part of a preliminary version of the National Bureau of Standard's STARPAC library [9].

---

9. References

[1] Gauss C. F. Theoria combinationis observationum erroribus minimis obnoxiae. Werke, 4. Gottingen. 1821. (Collected works 1873)

[2] Plackett, R. L. Some theorems in least squares. Biometrika, 19: 149-157; 1950.

[3] Kalman, R. E. A new approach to linear filtering and prediction problems. ASME Transactions, Part D, Journal of Basic Engineering, 82: 35-45; 1960.

[4] Jones, R. H.; Tryon, P. V. Estimating time from atomic clocks. Second Symposium on Atomic Time Scale Algorithms, June 23-25, 1982, Boulder, CO.

[5] Graybill, F. A. *Theory and application of the linear model. North Scituate:* Duxbury Press; 1976.

[6] Jones, R. H. Maximum likelihood fitting of ARMA models to time series with missing observations. Technometics, 22: 389-395; 1980.

[7] Weiss, B. E. A modular software package for solving unconstrained nonlinear optimization problems. Master's thesis, University of Colorado at Boulder; 1980.

[8] Dennis, J. E.; Schnabel, R. B. *Numerical methods for unconstrained optimization and nonlinear equations.* New Jersey: Prentice Hall; 1982.

[9] Donaldson, J. R.; Tryon, P. V. STARPAC: The standards time series and regression package. Unpublished—Documentation available from Janet P. Donaldson, NBS-714, 325 Broadway, Boulder, CO 80303.
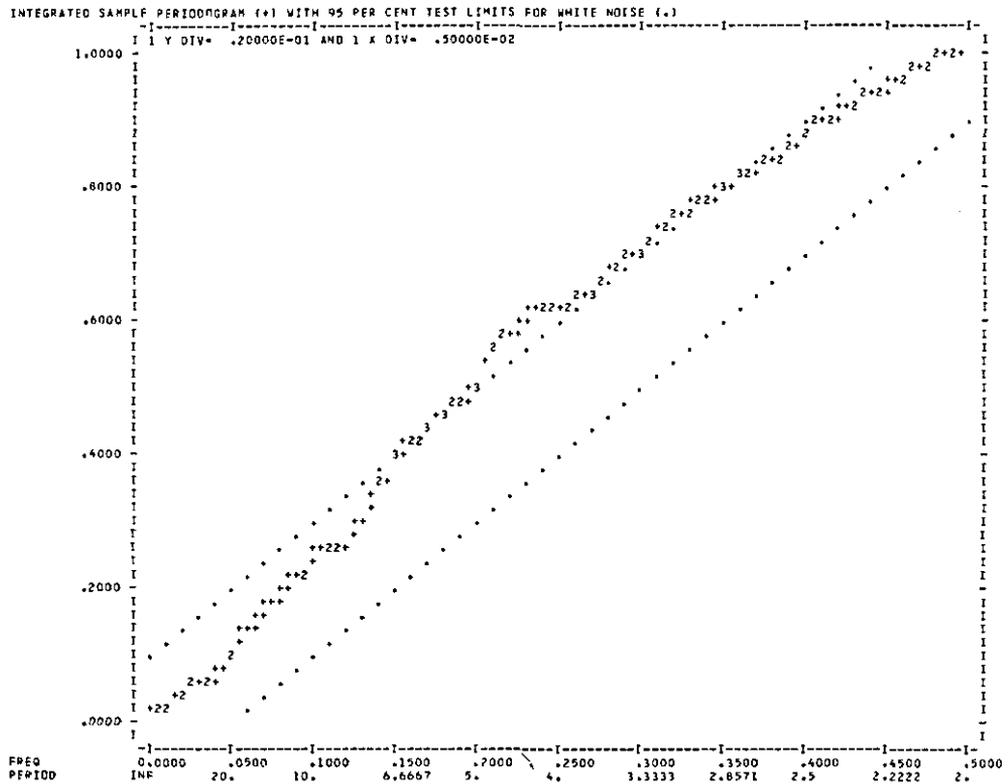
Figures 1a-f and 2a-f follow.

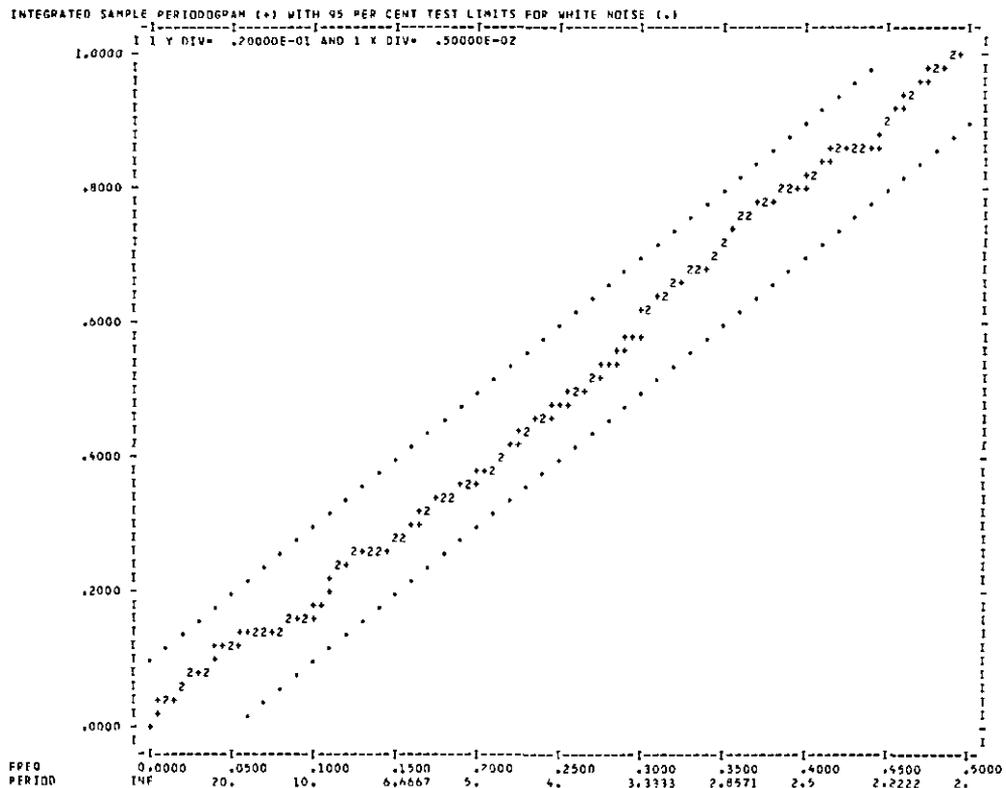FIGURE 1a. Residuals from clock 601 minus clock 167.

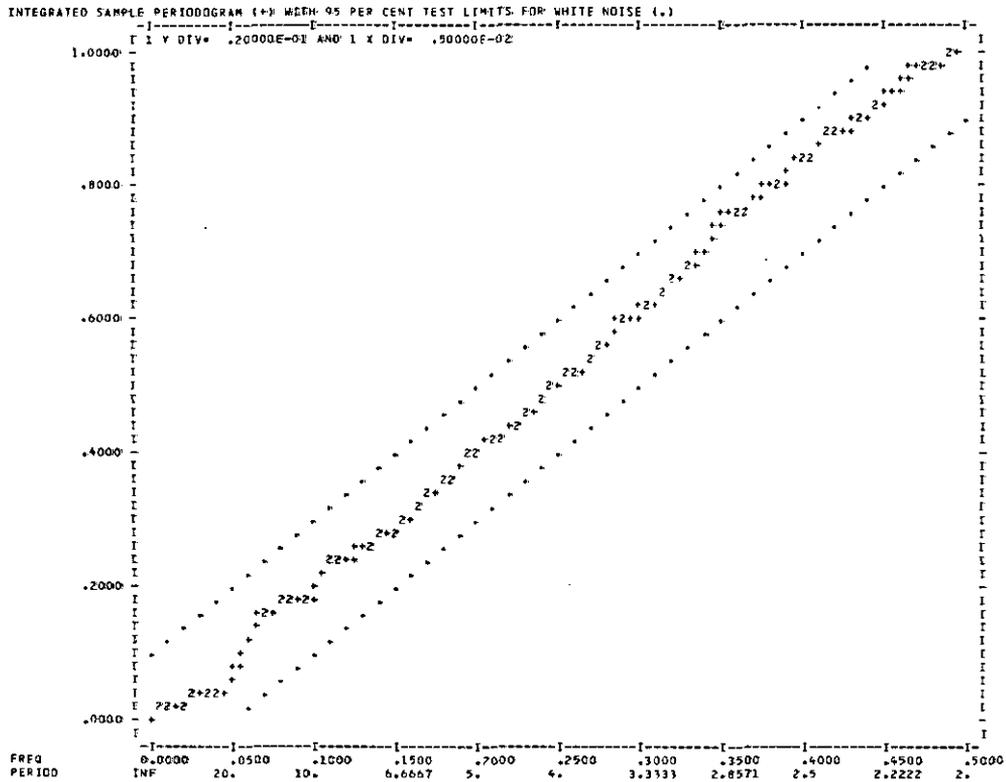FIGURE 1b. Residuals from clock 601 minus clock 137.

12

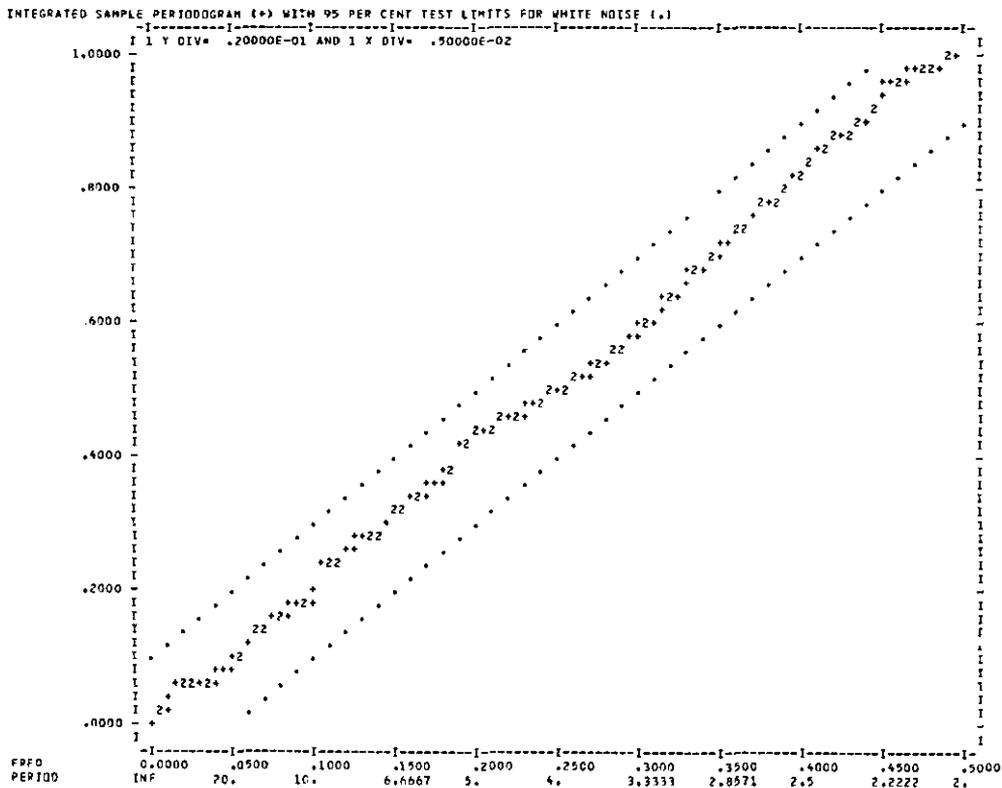**FIGURE 1c.** Residuals from clock 601 minus clock 1316.

FIGURE 1d. Residuals from clock 601 minus clock 323.
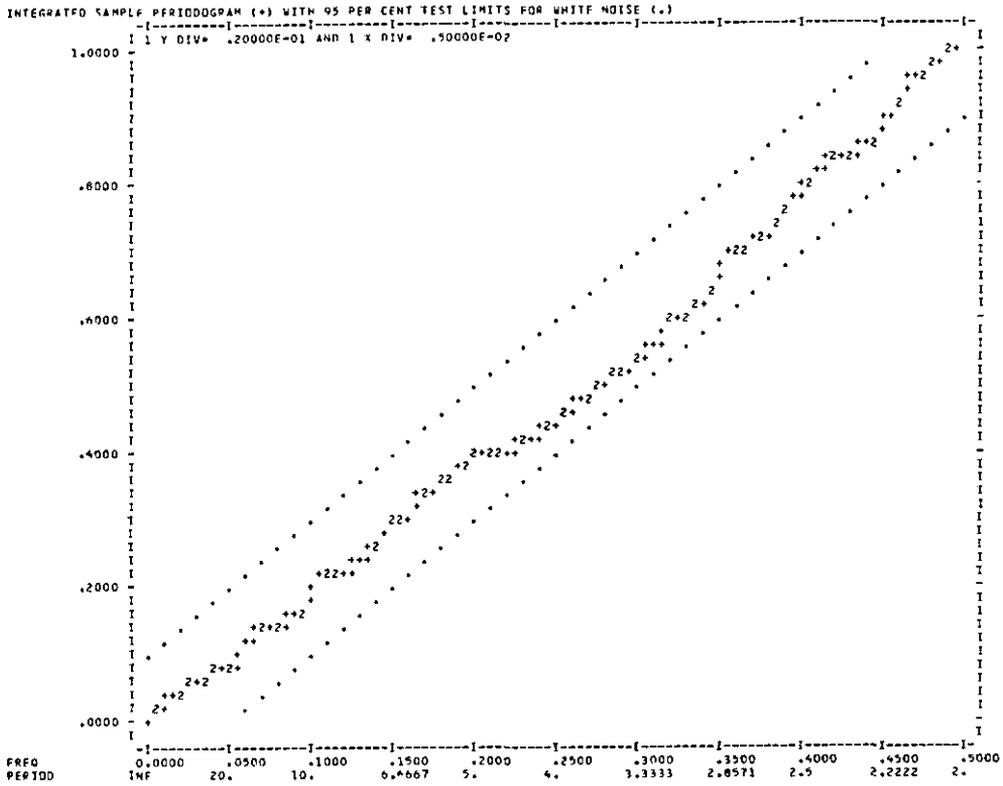
13

FIGURE 1e. Residuals from clock 601 minus clock 324.

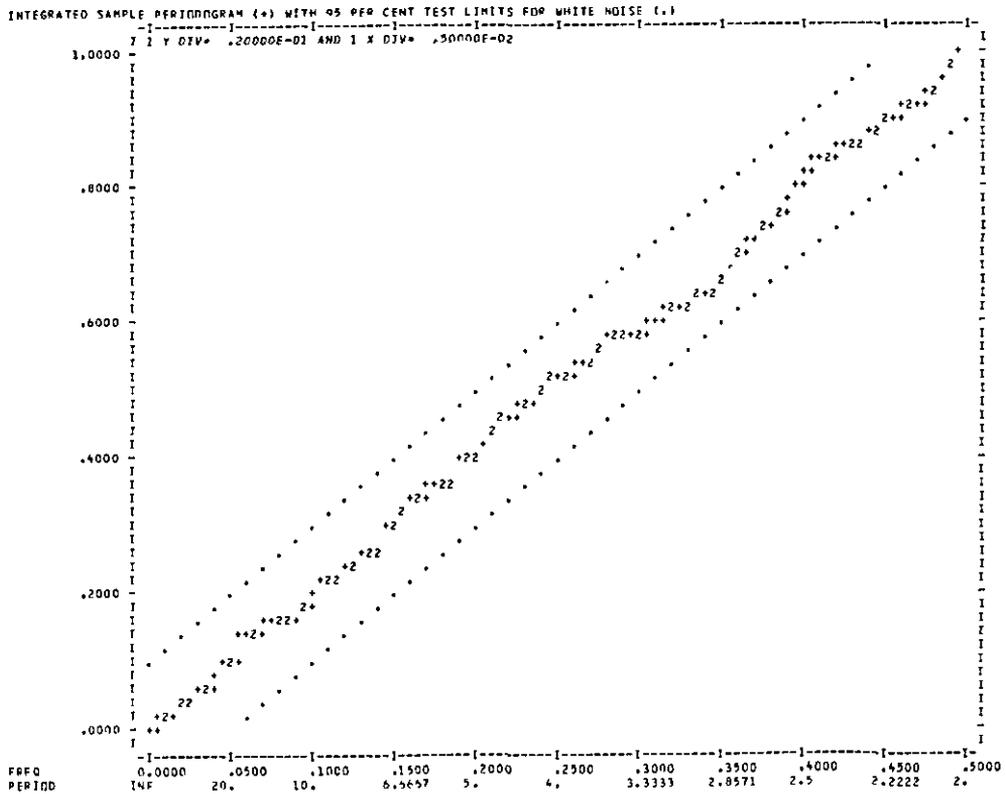FIGURE 1f. Residuals from clock 601 minus clock 8.

14

```
            NUMBER OF OBSERVATIONS  =              332
            MINIMUM OBSERVATION     = -.43333971BE+02
            MAXIMUM OBSERVATION     =  .428583010E+02

            HISTOGRAM LOWER BOUND   = -.450000000E+02
            HISTOGRAM UPPER BOUND   =  .450000000E+02

            OBSERVATIONS USED       =              332       25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
            MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
            MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
            MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY 0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
 INTERVAL    CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
 MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015   1.000   .015    5    +++
 -30.0000    .042    .985   .027    9    +++++
 -20.0000    .181    .958   .139   46    ++++++++++++++++++++++++++
 -10.0000    .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++
 0.          .636    .605   .241   80    +++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
 20.0000     .931    .184   .114   38    +++++++++++++++++++++++
 30.0000     .991    .069   .060   20    ++++++++++++
 40.0000    1.000    .009   .009    3    ++
```

FIGURE 2a. Histogram—residuals from clock 601 minus clock 167.

```
            NUMBER OF OBSERVATIONS  =              332
            MINIMUM OBSERVATION     = -.43333971BE+02
            MAXIMUM OBSERVATION     =  .428583010E+02

            HISTOGRAM LOWER BOUND   = -.450000000E+02
            HISTOGRAM UPPER BOUND   =  .450000000E+02

            OBSERVATIONS USED       =              332       25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
            MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
            MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
            MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY 0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
 INTERVAL    CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
 MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015   1.000   .015    5    +++
 -30.0000    .042    .985   .027    9    +++++
 -20.0000    .181    .958   .139   46    ++++++++++++++++++++++++++
 -10.0000    .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++
 0.          .636    .605   .241   80    +++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
 20.0000     .931    .184   .114   38    +++++++++++++++++++++++
 30.0000     .991    .069   .060   20    ++++++++++++
 40.0000    1.000    .009   .009    3    ++
```

FIGURE 2b. Histogram—residuals from clock 601 minus clock 137.

```
            NUMBER OF OBSERVATIONS  =              332
            MINIMUM OBSERVATION     = -.433339718E+02
            MAXIMUM OBSERVATION     =  .428583010E+02

            HISTOGRAM LOWER BOUND   = -.450000000E+02
            HISTOGRAM UPPER BOUND   =  .450000000E+02

            OBSERVATIONS USED       =              332       25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
            MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
            MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
            MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY 0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
 INTERVAL    CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
 MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015   1.000   .015    5    +++
 -30.0000    .042    .985   .027    9    +++++
 -20.0000    .181    .958   .139   46    ++++++++++++++++++++++++++
 -10.0000    .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++
 0.          .636    .605   .241   80    +++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
 20.0000     .931    .184   .114   38    +++++++++++++++++++++++
 30.0000     .991    .069   .060   20    ++++++++++++
 40.0000    1.000    .009   .009    3    ++
```

FIGURE 2c. Histogram—residuals from clock 601 minus clock 1316.

15

```
            NUMBER OF OBSERVATIONS    =         332
            MINIMUM OBSERVATION       = -.433339718E+02
            MAXIMUM OBSERVATION       =  .428583010E+02

            HISTOGRAM LOWER BOUND     = -.450000000E+02
            HISTOGRAM UPPER BOUND     =  .450000000E+02

            OBSERVATIONS USED         =         332         25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED     = -.433339718E+02     STANDARD DEVIATION   =  .160738390E+02
            MAX. OBSERVATION USED     =  .428583010E+02     MEAN DEV./STD. DEV.  =  .805496712E+00
            MEAN VALUE                = -.153504202E+00     B1**0.5              =  .145368914E-02
            MEDIAN VALUE              = -.691321082E+00     B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
INTERVAL    CUM.   1-CUM.   CELL    NO.                            CELL FRACTION
MID POINT   FRACT. FRACT.   FRACT.  OBS.  0.00     .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015  1.000     .015     5   +++
 -30.0000    .042   .985     .027     9   +++++
 -20.0000    .181   .958     .139    46   ++++++++++++++++++++++++++++++++
 -10.0000    .395   .819     .214    71   +++++++++++++++++++++++++++++++++++++++++++++++++
 0.          .636   .605     .241    80   ++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816   .364     .181    60   ++++++++++++++++++++++++++++++++++++++++++
 20.0000     .931   .184     .114    38   ++++++++++++++++++++++++++
 30.0000     .991   .069     .060    20   +++++++++++++
 40.0000    1.000   .009     .009     3   ++
```

FIGURE 2d. Histogram—residuals from clock 601 minus clock 323.

```
            NUMBER OF OBSERVATIONS    =         332
            MINIMUM OBSERVATION       = -.433339718E+02
            MAXIMUM OBSERVATION       =  .428583010E+02

            HISTOGRAM LOWER BOUND     = -.450000000E+02
            HISTOGRAM UPPER BOUND     =  .450000000E+02

            OBSERVATIONS USED         =         332         25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED     = -.433339718E+02     STANDARD DEVIATION   =  .160738390E+02
            MAX. OBSERVATION USED     =  .428583010E+02     MEAN DEV./STD. DEV.  =  .805496712E+00
            MEAN VALUE                = -.153504202E+00     B1**0.5              =  .145368914E-02
            MEDIAN VALUE              = -.691321082E+00     B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
INTERVAL    CUM.   1-CUM.   CELL    NO.                            CELL FRACTION
MID POINT   FRACT. FRACT.   FRACT.  OBS.  0.00     .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015  1.000     .015     5   +++
 -30.0000    .042   .985     .027     9   +++++
 -20.0000    .181   .958     .139    46   ++++++++++++++++++++++++++++++++
 -10.0000    .395   .819     .214    71   +++++++++++++++++++++++++++++++++++++++++++++++++
 0.          .636   .605     .241    80   ++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816   .364     .181    60   ++++++++++++++++++++++++++++++++++++++++++
 20.0000     .931   .184     .114    38   ++++++++++++++++++++++++++
 30.0000     .991   .069     .060    20   +++++++++++++
 40.0000    1.000   .009     .009     3   ++
```

FIGURE 2e. Histogram—residuals from clock 601 minus clock 324.

```
            NUMBER OF OBSERVATIONS    =         332
            MINIMUM OBSERVATION       = -.433339718E+02
            MAXIMUM OBSERVATION       =  .428583010E+02

            HISTOGRAM LOWER BOUND     = -.450000000E+02
            HISTOGRAM UPPER BOUND     =  .450000000E+02

            OBSERVATIONS USED         =         332         25 PCT TRIMMED MEAN = -.333716985E+00
            MIN. OBSERVATION USED     = -.433339718E+02     STANDARD DEVIATION   =  .160738390E+02
            MAX. OBSERVATION USED     =  .428583010E+02     MEAN DEV./STD. DEV.  =  .805496712E+00
            MEAN VALUE                = -.153504202E+00     B1**0.5              =  .145368914E-02
            MEDIAN VALUE              = -.691321082E+00     B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
INTERVAL    CUM.   1-CUM.   CELL    NO.                            CELL FRACTION
MID POINT   FRACT. FRACT.   FRACT.  OBS.  0.00     .05      .10      .15      .20      .25
-----------------------------------------+---------+---------+---------+---------+---------+
 -40.0000    .015  1.000     .015     5   +++
 -30.0000    .042   .985     .027     9   +++++
 -20.0000    .181   .958     .139    46   ++++++++++++++++++++++++++++++++
 -10.0000    .395   .819     .214    71   +++++++++++++++++++++++++++++++++++++++++++++++++
 0.          .636   .605     .241    80   ++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 10.0000     .816   .364     .181    60   ++++++++++++++++++++++++++++++++++++++++++
 20.0000     .931   .184     .114    38   ++++++++++++++++++++++++++
 30.0000     .991   .069     .060    20   +++++++++++++
 40.0000    1.000   .009     .009     3   ++
```

FIGURE 2f. Histogram—residuals from clock 601 minus clock 8.

16

Estimating Time From Atomic Clocks

Richard H. Jones* and Peter V. Tryon**

National Bureau of Standards
Boulder, CO 80303

A Kalman recursive algorithm for estimating time from an ensemble of atomic clocks has been developed. The algorithm allows for the addition or deletion of clocks at any time, and provides automatic error detection and correction. The observations consist of time differences between clocks and may be taken at unequally spaced time points. Maximum likelihood estimates of the unknown parameters are obtained with confidence intervals, as well as hypothesis tests to determine whether the estimated parameters are significantly different from zero. The program is operational on the National Bureau of Standards' Time and Frequency Division's PDP 11/70.

Key words: atomic clocks; Kalman recursion; maximum likelihood estimation; missing observations; nonlinear estimation; state space; time series analysis; unequally spaced data.

1. Introduction

Cesium beam atomic clocks have an accuracy of a few parts in 10^{14} over a period of a day; however, they are not deterministic and undergo stochastic variations in both time and frequency. It was shown by Tryon and Jones [1] that the actual frequency of a clock behaves as a random walk which, over a time period of a day, has a standard deviation of less than four nanoseconds per day. The effect of this random walk in frequency on time measurements is that the individual frequency steps are summed, producing an integrated random walk.

In addition, frequency has a white noise component which integrates into an independent random walk in time with a standard deviation of up to 15 nanoseconds over a time period of a day. Therefore, the deviation of clock time from true time behaves as a random walk plus the sum of a random walk. The possibility of a frequency drift also exists.

When estimating time from an ensemble of clocks, the only observations possible are measurements of time differences between clocks. In addition, observations may occur at unequally spaced time points, and various types of errors are possible. The most common errors are *read errors*, where a single measurement is incorrect but subsequent measurements are the same as if the read error did

not occur; *time steps*, where a jump in time occurs and the clock remains at the new position; and *frequency steps*, where the frequency takes a step and remains at the new position. This last error appears as a change in the rate of gain of the clock.

This paper describes a state space algorithm for estimating time from an ensemble of cesium beam atomic clocks with unequally spaced observations subject to various errors. Clocks can be added to or deleted from the ensemble. Using iterative calculations over a recent history of the ensemble, maximum likelihood estimates of the unknown variances can be obtained by nonlinear optimization. Confidence intervals on the estimated parameters and test of hypotheses, such as whether parameters are significantly different from zero, can also be obtained from the likelihood function.

2. Mathematical Model

The model for a single clock in state space form is

$$
\begin{bmatrix} x(t) \\ y(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} 1 & \delta(t) & \delta^2(t)/2 \\ 0 & 1 & \delta(t) \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(t-1) \\ y(t-1) \\ w(t-1) \end{bmatrix} + \begin{bmatrix} \varepsilon(t) \\ \eta(t) \\ \alpha(t) \end{bmatrix} \quad (2.1)
$$

where $t = 1, 2, 3, \ldots$ is an index of the observation points, $x(t)$ is the difference between the clock's time and "true" time, and $y(t)$ is the difference between the clock's frequency and the fundamental resonance of the cesium atom which defines the second, 9,192,631,770 Hz, and

*Division of Biometrics, Box B-119, School of Medicine, University of Colorado, Denver, CO 80262.
**Deceased. Dr. Tryon served with the Center for Applied Mathematics, National Engineering Laboratory.

is expressed in units of nanoseconds per day. $w(t)$ represents a possible frequency drift in units of nanoseconds/day², and $\delta(t)$ is the time interval between t–1 and t in days. $\varepsilon(t)$, $\eta(t)$ and $\alpha(t)$ are random variables with zero mean, uncorrelated with each other, and uncorrelated in time. These are the input to the random walks, and for small $\delta(t)$, their variances are proportional to the length of the time intervals between observations, $\delta(t)$,

$$\text{Var}\{\varepsilon(t)\} = \delta(t)\,\sigma_\varepsilon^2$$

$$\text{Var}\{\eta(t)\} = \delta(t)\,\sigma_\eta^2 \qquad (2.2)$$

$$\text{Var}\{\alpha(t)\} = \delta(t)\,\sigma_\alpha^2$$

If $\sigma_\alpha^2 = 0$, the drift, $w(t)$, does not change with time.

The state equation for an ensemble of m clocks is obtained by concatenating the three state elements of each clock into a column vector of length 3m. The state transition matrix, $\Phi(t)$, is a 3m by 3m matrix consisting of 3 by 3 blocks on the main diagonal corresponding to the state transition matrix in eq (1), and zero blocks off the main diagonal. This equation of state for m clocks will be written.

$$X(t) = \Phi(t)\,X(t-1) + U(t). \qquad (2.3)$$

The random input vector, $U(t)$, is of length 3m and consists of the $\varepsilon(t)$, $\eta(t)$, and $\alpha(t)$ for each clock. The covariance matrix of $U(t)$, $\delta(t)$ Q, is a 3m by 3m diagonal matrix with diagonal elements consisting of the σ_ε^2, σ_η^2 and σ_α^2 for each clock. While reasonable guesses are available for the σ_ε^2 and σ_η^2, it is necessary to obtain better estimates of these variances from data. The values are characteristics of each clock and will be different for each clock. It is also necessary to determine whether the σ_α^2 are significantly different from zero. If the σ_α^2 are not significantly different from zero it would indicate that any drifts that exist are deterministic rather than random walks. In this case the $w(t)$ will be constant with respect to time and these can be tested to determine if the drifts are significantly different from zero.

When observations are taken, one of the clocks is used as a reference clock, and the time differences between this clock and each of the other clocks are recorded. This gives an observation vector $Z(t)$, of length m–1, and an observation equation of the form

$$Z(t) = H(t)X(t) + V(t), \qquad (2.4)$$

where $H(t)$ is a matrix indicating which clock differences are observed. If the first clock is the reference clock, $H(t)$

is of the form

$$H(t) = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \end{bmatrix} \qquad (2.5)$$

and has dimension $(m$–1$)$ by $3m$. $H(t)$ can be time dependent since if observations are missing or deleted because of error detection, rows of $H(t)$ are eliminated. $V(t)$ is a vector of observational errors. These errors are very small when reading clock differences. If the data are truncated to the nearest nanosecond, the variances of the elements $V(t)$ would be

$$r = 1/12 \text{ (nanoseconds)}^2. \qquad (2.6)$$

When observations are taken with higher precision than the nearest nanosecond, r can be reduced and its value determined from the instrumentation. The observational error variance matrix, R, will be a diagonal matrix and diagonal elements r, assumed to be equal and known. While these numbers are very small, the inclusion of R in the recursion does add numerical stability.

3. The Data

In a companion paper [1] Tryon and Jones used 333 daily observations from seven clocks starting February 16, 1979 to estimate clock parameters. These data were chosen since they consisted of a fairly long record with few errors. Some preprocessing was necessary to remove several outliers giving a set of data that could be considered error-free. These calculations were carried out on a CDC Cyber 170/750 computer.

This paper reports the new statistical procedures and techniques that were incorporated when the algorithm was rewritten to run on the National Bureau of Standards Time and Frequency Division's PDP-11/70. The algorithm was initialized on March 31, 1981 with the first set of data collected on April 1, 1981. At the beginning of each month the daily observations from the previous month are passed to the algorithm without preprocessing. While the data are nominally collected at the same time each day, for various reasons, the time of the observations are sometimes off by several hours giving unequally spaced data. Individual clocks may have read errors, time steps, and frequency steps. In addition, clocks may be added or deleted from the ensemble.

18

4. The Kalman Recursion With Error Detection

Let $X(t|s)$, $s \leq t$ be the best estimate of the state vector at time t given observations up to time s, and let $P(t|s)$ be its covariance matrix. The recursion begins by specifying $X(0|0)$ and $P(0|0)$. First a one step prediction of the state vector is calculated [3]

$$X(t+1|t) = \Phi(t+1) \ X(t|t), \quad (4.1)$$

as is its covariance matrix

$$P(t+1|t) = \Phi(t+1) \ P(t|t) \ \Phi'(t+1) + \delta(t+1)Q. \quad (4.2)$$

The predicted values of the observations at time $t+1$ are

$$Z(t+1|t) = H(t+1) \ X(t+1|t). \quad (4.3)$$

The difference between the vector of actual observations and predicted observations is the *innovation* vector,

$$I(t+1) = Z(t+1) - Z(t+1|t), \quad (4.4)$$

which has covariance matrix

$$C(t+1) = H(t+1)P(t+1|t) \ H'(t+1) + R(t+1). \quad (4.5)$$

It is at this point in the recursion that a statistical method of error detection is employed. A simple method would be to divide each element of the innovation vector by the square root of corresponding diagonal element of its covariance matrix giving a standard normal variable under the null hypothesis of no error. However, since clock differences are being measured and the elements of the innovation vector are intercorrelated, this is not an optimal test.

If there is an error in the reference clock, a constant bias will appear in every reading. This can be written as a regression equation with correlated errors

$$I(t+1) = A_1 \beta + \varepsilon, \quad (4.6)$$

where A_1 is a column vector of ones

$$A_1' = [1, 1, \ldots, 1], \quad (4.7)$$

and the error vector, ε, has covariance matrix $C(t+1)$ from eq (4.5). If there is an error in a clock which is not the reference clock, the error will appear only in the measurement involving that clock. For clock i, the model is

$$I(t+1) = A_i \ \beta + \varepsilon, \quad (4.8)$$

where

$$A_i' = [0 \ 0 \ \cdots \ -1 \ \cdots \ 0], \quad (4.9)$$

the minus one appearing in the position of the clock. Although there are only $m-1$ measurements, m tests, one for each clock, are possible, but they are not statistically independent tests. Letting A represent any of the A vectors, the minimum variance weighted least squares estimate of β is

$$b = A'C^{-1}(t+1)I(t+1)/A'C^{-1}(t+1)A, \quad (4.10)$$

and has standard error

$$\text{s.e.}(b) = (1/A'C^{-1}(t+1)A)^{1/2}. \quad (4.11)$$

If the test statistic

$$z = b/\text{s.e.}(b) \quad (4.12)$$

has absolute value larger than some value such as 3.0, the corresponding clock is assumed to have an error. It must be remembered that this test is only approximate if guesses are being used for the random walk variances in the Q matrix, and will be much more accurate after maximum likelihood estimates are obtained for these variances. The details of the above calculations are given in the Appendix.

An overall test for errors is possible by calculating the quadratic form

$$Q = I'(t+1) \ C^{-1}(t+1) \ I(t+1) \quad (4.13)$$

which will be distributed as chi-square with $m-1$ degrees of freedom under the null hypothesis of no errors in the observations. Although this statistic is calculated, it is no longer being used for error detection since it is felt that this test lacks power against the most common alternative of an error in a single clock.

If an error is detected, the reading for the clock in question is eliminated from the observation vector. If more than one error is detected the clock with the largest absolute value of z is eliminated and the tests recalculated for the remaining clocks. If the clock to be eliminated is the reference clock, the reference clock is eliminated by choosing a clock still in the model to be the new reference clock and forming a new data vector by subtracting the reading of the difference between the previous reference clock and the new reference clock from each of the other readings. This is shown schematically as:

Previous   New

$$\textcircled{1} - \textcircled{2}$$

$$\textcircled{1} - \textcircled{3} - \left[\textcircled{1} - \textcircled{2}\right] = \textcircled{2} - \textcircled{3}$$

$$\textcircled{1} - \textcircled{4} - \left[\textcircled{1} - \textcircled{2}\right] = \textcircled{2} - \textcircled{4} \qquad (4.14)$$

$$\textcircled{1} - \textcircled{5} - \left[\textcircled{1} - \textcircled{2}\right] = \textcircled{2} - \textcircled{5}.$$

When all remaining clocks pass the test, the recursion continues. Let the Kalman gain be

$$\Delta(t+1) = P(t+1|t) \, H'(t+1) \, C^{-1}(t+1). \quad (4.15)$$

If clocks have been eliminated due to errors, the rows of $H(t+1)$ corresponding to the eliminated clocks are eliminated as are the corresponding rows and columns of $C(t+1)$. In the cases where the reference clock is changed, the positions of the ones in the $H(t+1)$ matrix must be moved to correspond to the new reference clock. The estimated state vector is then updated,

$$X(t+1|t+1) = X(t+1|t) + \Delta(t+1) \, I(t+1), \quad (4.16)$$

as is its covariance matrix

$$P(t+1|t+1) = P(t+1|t) - \\ \Delta(t+1) \, H(t+1) \, P(t+1|t). \quad (4.17)$$

This completes the recursion, but the effects of any detected errors must be resolved. Because the measurement errors of the data are so small, the final updating eq (4.16) essentially sets the time states to agree with the actual readings of the differences. This will not be the case for any clock that was eliminated at this step since its reading has been eliminated. That clock will simply remain at its position which was predicted from the previous time point. Since the most common type of error is a time step (a clock change which persists into the future), an administrative decision was made to correct the time state of a clock with a detected error so that the difference between the time of the reference clock and the clock with an error agreed with the actual observation. If the error was a read error which does not persist into the future, an error will probably be detected at the next time point which will have opposite sign and a reverse correction applied.

The remaining problem is the possibility of a frequency step. If a very small frequency step occurs, the Kalman algorithm has the flexibility to slowly adjust the estimated frequency of the clock to the new value. If the frequency step causes a time change large enough to be detected, it is not known in one interval whether this is a time error or frequency step. In case a frequency step did occur, it is desirable to allow the frequency state to adjust quickly to the new value. In order to accomplish this, the diagonal element of the state covariance matrix, $P(t+1|t+1)$ corresponding to $y(t+1)$ of the clock in question is increased. The value that is added to this diagonal element is the minimum of $(2*c/\delta(t))^2$ and $\delta(t)10^6$, where c is the clock time correction. This allows the recursion to adjust to a new frequency within a few time steps. Whether or not a frequency step occurred, the P matrix returns to its previous order of magnitude within a few time steps.

The estimate of "true" time is obtained from the state vector $X(t|t)$ which contains the estimates of $x(t)$ for each clock. The estimated values of $x(t)$ for each clock are in relative agreement since their differences agree with the observed differences at time t. A relatively crude clock within the computer is used to determine the approximate time of the measurement.

Each of the clocks runs freely and independently of the others. Thus the ticks of the clocks each second are randomly distributed. If the tick on the reference clock is used as the indicator to read the other clocks, the readings are the time differences between the ticks in nanoseconds. At some time in the past the ensemble was initialized relative to some definition of "true" time (for example, international Atomic Time) and the uncertainty expressed in the initial covariance matrix $P(0|0)$. At that time each clock's tick was a certain number of nanoseconds off of "true" time. As data are collected and the algorithm progresses through its steps, all the clocks remain in relative agreement, but the ensemble drifts as a whole. The variance of this ensemble drift is given by the diagonal elements of $P(t|t)$ corresponding to the $x(t)$ for each clock. Since observational error is so small, these diagonal elements are essentially all the same. True time is estimated by correcting the tick of any one of the clocks by an amount equal to its estimated time state, $x(t)$, which is an estimate of the clock's deviation from true time. The variance of time estimate is the corresponding diagonal element of $P(t|t)$.

Sample output from the algorithm starting with the initialization, which was output from the previous run, is shown in figure 1. An example of a detected error with an adjustment is shown in figure 2. The algorithm also has the capability to adjust any clock administratively by a given amount or apply a frequency adjustment to all clocks in order to steer a time scale. These are software adjustments which do not affect the physical clocks but enable a time scale to trace another time scale such as International Atomic Time. The National Bureau of Standards has several of these paper scales.

```
INITIALIZATION.  9 CLOCKS

MJD= 45059    TIME= 1200
```

CLOCK	INITIAL STATE			STANDARD DEVIATION			ROOT Q		
	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT
1 1316	270534.6	9.20	0.000	1356.2	5.02	0.000	4.14	0.80	0.000
2 167	-44523.9	549.80	0.000	1356.2	6.72	0.000	13.52	1.11	0.000
3 137	11731.0	52.97	0.000	1356.2	7.29	0.000	11.31	2.49	0.000
4 113	-207978285.4	939.05	0.000	1356.2	7.60	0.000	9.48	3.18	0.000
5 PHM4	-254618838.5	660.64	0.000	1356.2	4.81	0.000	0.65	0.77	0.000
6 352	-78346.9	568.72	0.000	1356.2	7.57	0.000	8.85	3.32	0.000
11 1375	3038.1	-96.37	0.000	1356.2	6.20	0.000	10.71	1.48	0.000
14 601	12014.9	41.48	0.000	1356.2	4.61	0.000	2.13	0.06	0.000
15 8	-25018.5	-966.53	0.000	1356.2	7.05	0.000	8.65	2.76	0.000

```
MJD= 45060    TIME= 1200
```

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	314518.6	314517.9	0.7	15.1	0.0
1- 3	258754.8	258759.8	-5.0	13.4	-0.4
1- 4	208247878.0	208247890.2	-12.2	12.1	-1.0
1- 5	254888722.0	254888721.7	0.3	4.8	0.1
1- 6	348332.6	348322.0	10.6	11.6	0.9
1-11	267581.0	267602.1	-21.1	12.4	-1.7
1-14	258488.5	258487.5	1.0	5.1	0.2
1-15	296546.0	296528.8	17.2	11.1	1.5

CLOCK	UPDATED STATE			STATE CHANGES			STANDARD DEVIATIONS		
	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT
1 1316	270544.8	9.31	0.000	0.9	0.11	0.000	1360.7	5.02	0.000
2 167	-43973.8	549.80	0.000	0.2	0.00	0.000	1360.7	6.61	0.000
3 137	11790.0	54.08	0.000	6.0	1.11	0.000	1360.7	7.29	0.000
4 113	-207977333.2	942.67	0.000	13.1	3.62	0.000	1360.7	7.59	0.000
5 PHM4	-254618177.2	660.74	0.000	0.6	0.10	0.000	1360.7	4.81	0.000
6 352	-77787.8	565.58	0.000	-9.7	-3.14	0.000	1360.7	7.57	0.000
11 1375	2963.8	-93.56	0.000	22.0	2.80	0.000	1360.7	6.20	0.000
14 601	12056.3	41.47	0.000	-0.1	0.00	0.000	1360.7	4.61	0.000
15 8	-26001.2	-971.05	0.000	-16.2	-4.52	0.000	1360.7	7.05	0.000

FIGURE 1. Example of output of algorithm starting with the initialization from the previous run. Under INITIALIZATION and UPDATED STATE, the standard deviations are the square roots of the diagonal elements of the state covariance matrix, $P(t|t)$. Under $MJD = 45060$, the standard deviations are the square roots of the diagonal elements of the innovation covariance matrix, $C(t)$.

CLOCK	UPDATED STATE			STATE CHANGES			STANDARD DEVIATIONS		
	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT
1 1316	270698.8	5.80	0.000	1.3	0.29	0.000	1477.2	5.03	0.000
2 167	-29561.0	555.26	0.000	-13.8	-1.07	0.000	1477.2	6.09	0.000
3 137	13186.1	58.85	0.000	14.7	2.97	0.000	1477.2	7.29	0.000
4 113	-207952798.7	939.05	0.000	0.9	0.38	0.000	1477.2	7.59	0.000
5 PHM4	-254600975.5	662.08	0.000	-1.8	-0.68	0.000	1477.2	4.82	0.000
6 352	-63184.8	560.60	0.000	9.0	2.94	0.000	1477.2	7.57	0.000
11 1375	417.2	-96.96	0.000	-9.3	-1.16	0.000	1477.2	6.21	0.000
14 601	13135.5	41.47	0.000	0.0	0.00	0.000	1477.2	4.62	0.000
15 8	-51091.2	-965.30	0.000	6.1	1.79	0.000	1477.2	7.06	0.000

```
MJD= 45087    TIME= 1405
```

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	299666.2	299662.6	3.6	14.9	0.2
1- 3	257496.6	257455.0	41.6	13.7	3.0
1- 4	208222466.2	208222483.2	-17.0	12.4	-1.4
1- 5	254870967.7	254870961.0	6.7	4.9	1.4
1- 6	333255.5	333280.6	-25.1	11.9	-2.1
1-11	270444.9	270393.3	51.6	12.5	4.1
1-14	257535.0	257524.5	10.5	5.2	2.0
1-15	322841.2	322845.4	-4.2	11.4	-0.4
*****QUAD=	40.6	CLOCK 11	-44.7	11.7	3.8

DIFF	OBSERVATION	PREDICTED OBS	RESIDUAL	STANDARD DEVIATION	T TEST
1- 2	299666.2	299662.6	3.6	14.9	0.2
1- 3	257496.6	257455.0	41.6	13.7	3.0
1- 4	208222466.2	208222483.2	-17.0	12.4	-1.4
1- 5	254870967.7	254870961.0	6.7	4.9	1.4
1- 6	333255.5	333280.6	-25.1	11.9	-2.1
1-14	257535.0	257524.5	10.5	5.2	2.0
1-15	322841.2	322845.4	-4.2	11.4	-0.4

```
CLOCK 11 ADVANCED BY          -44.7
```

CLOCK	UPDATED STATE			STATE CHANGES			STANDARD DEVIATIONS		
	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT	TIME	FREQ	DRIFT
1 1316	270713.7	7.39	0.000	10.6	1.58	0.000	1482.1	5.02	0.000
2 167	-28950.5	555.74	0.000	7.1	0.48	0.000	1482.1	6.07	0.000
3 137	13219.1	51.98	0.000	-31.0	-6.87	0.000	1482.1	7.23	0.000
4 113	-207951750.5	946.44	0.000	27.7	7.39	0.000	1482.1	7.50	0.000
5 PHM4	-254600252.0	662.59	0.000	3.9	0.51	0.000	1482.1	4.81	0.000
6 352	-62539.8	571.19	0.000	35.6	10.59	0.000	1482.1	7.48	0.000
11 1375	270.8	-96.70	0.000	3.7	0.26	0.000	1482.1	89.55	0.000
14 601	13180.8	41.45	0.000	0.1	-0.02	0.000	1482.1	4.62	0.000
15 8	-52125.5	-961.81	0.000	14.8	3.49	0.000	1482.1	6.99	0.000

FIGURE 2. Example of computer output with an error detected in clock 11. The listing starts with the final results from the previous day. QUAD refers to the overall test which is not actually used for detecting errors but is printed out so that errors are easily visible. The optimal estimate of the clock error is –44.7 with standard deviation 11.7.

5. Estimation of Parameters

Assuming Gaussian errors, -2 ln likelihood is calculated from

$$L = \sum_{t} [ln|C(t)| + I'(t)\, C^{-1}(t)\, I(t)]. \qquad (5.1)$$

This gives a number which depends on the values of the parameters of the model. Nonlinear optimization routines can then be used to minimize this function giving maximum likelihood estimates of the parameters. Each function evaluation is a pass of the recursion through all the data.

When error detection is used, a new problem arises. The dimensions of $C(t)$ and $I(t)$ are reduced if a clock is eliminated at time t. When the nonlinear optimization routine varies the parameters, it is possible that at some point a clock that was previously eliminated is suddenly not eliminated, or that a clock that was not eliminated *is* eliminated. This would cause a discontinuity in the likelihood function. To avoid this problem, a file of indicator variables, indicating which clocks are deleted at which time points, is produced. The nonlinear optimization is then run conditional on fixed clocks being eliminated at certain time points. Upon convergence the parameters in the original model are replaced by their optimized value and the initial program rerun generating a new array of indicator variables. If the value of the likelihood function changes it indicates that the error detection procedure has detected a different set of errors so the nonlinear optimization is repeated. Experience has shown this procedure converges in two or three iterations.

The optimization was carried out on one year of data, April 1, 1981 to March 31, 1982. *During this time 12 clocks were used*, but since clocks were added and deleted administratively from the algorithm (removed because of problems with the clock, not error detection) a maximum of 10 clocks and a minimum of eight clocks were available at any one time. Figure 3 shows the times at which various clocks were present and the number of actual observations used for each clock. The number missing is the number of observations eliminated by error detection.

In the first nonlinear optimization run, σ_ϵ and σ_η were varied for each clock and $w(t)$ and σ_α were set equal to zero. This is a drift-free model, and with 12 clocks gives 24 parameters to be estimated. Standard deviations rather than variances are used for optimization since if they go negative their squares are still positive. If variances were used directly, it would be possible for the nonlinear optimization routine to try negative values which would be meaningless. The results of this optimization procedure are shown in table 1.

CLOCKS PRESENT IN STUDY

Clock		No. Present	No. Missing
601		298	2
8		360	5
113		354	5
NBS4		66	3
1375		357	8
323		255	5
352		354	11
PHM4		203	6
61		67	4
137		358	7
167		361	4
1316		364	1

DAYS (0 to 365)
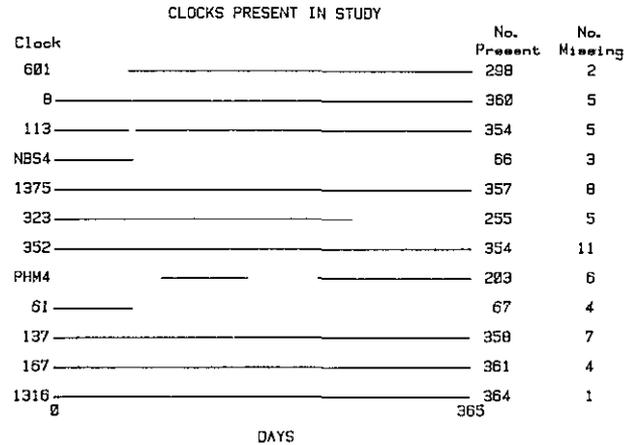
FIGURE 3. A schematic diagram showing the time when various clocks are available for use in the algorithm. The number present are the actual number used after error detection. The numbers missing are the number of errors detected. The sum of the number present and number missing is the number of observations available to the algorithm.

TABLE 1. *Estimated values of σ_ϵ and σ_η and 95% confidence intervals.*

Clock	n	$\sigma_\epsilon \sim$ nanoseconds			$\sigma_\eta \sim$ nanoseconds/day		
		Lower Limit	Est	Upper Limit	Lower Limit	Est	Upper Limit
1316	364	3.81	4.14	4.53	0.53	0.80	1.23
167	361	12.58	13.52	14.67	0.57	1.11	2.07
137	358	10.41	11.31	12.27	1.76	2.49	3.56
61	67	5.48	6.77	8.43	1.53	2.80	4.83
352	354	8.12	8.85	9.74	2.42	3.32	4.41
323	255	2.06	2.37	2.74	0.63	0.94	1.34
1375	357	9.93	10.71	11.64	0.96	1.48	2.25
NBS4	66	0	0.88	1.86	0.72	1.34	2.16
113	354	8.73	9.48	10.38	2.49	3.18	4.11
8	360	7.98	8.65	9.49	2.11	2.76	3.66
601	298	1.89	2.13	2.41	0	0.06	0.52
PHM4	203	0	0.65	1.19	0.55	0.77	1.09

When m clocks are in the model, it is only possible to estimate $m-1$ drift terms when the observations are time differences. Tryon and Jones [1] handled this problem by estimating m drifts with the constraint that the sum of the drifts was zero. Here, a different approach is used. A clock in the model is assumed to have zero drift and the drifts of the $m-1$ remaining drifts are estimated. Clock 601 was chosen to be the clock with no drift since the standard deviation of its frequency random walk was not significantly different from zero. Since clock 601 was not available for the entire year, the drift for clock 323 was also set equal to zero. These two clocks had overlapping data intervals which included the entire year. Clock 323 was also chosen because it is known to be a particularly

stable clock. The drifts for two other clocks, 61 and NBS4, were also set equal to zero because of their short lengths of data. The remaining eight clocks were tested for drift by a 24-parameter nonlinear optimization run. For each of the eight clocks, the frequency random walk standard deviation, σ_η, the initial value of the drift, $w(O)$, and the drift random walk standard deviation, σ_α, were allowed to vary. All of the time random walk standard deviations, σ_ε, were fixed at their optimized values from the previous run, as were the parameters for the remaining four clocks in the algorithm. This approach of fixing some of the parameters was necessitated by the storage available on the PDP 11/70 which allowed a maximum of 25 parameters to be optimized simultaneously. Changes in σ_ε that occur when drift is included in the model, while small, do in fact exist [1].

The estimated values of all eight σ_α were very small, so their significance was tested by fixing their values at zero and rerunning the previous optimization with 16 parameters. Statistical tests based on changes in $-2 \ln$ likelihood have asymptotic chi-square distributions with degrees of freedom equal to the number of parameters added or deleted from the model or constraints introduced [2]. In this case, eight parameters were fixed at zero but $-2 \ln$ likelihood only increased by 0.2 indicating that the σ_α are not significantly different from zero. This agrees with the results obtained on the earlier data set [1]. These results, which were obtained on a CDC Cyber 170/750 computer, were checked using the new algorithm on the PDP 11/70 and duplicated exactly. At this point it is assumed that any frequency drifts which may exist can be considered to be constant over the length of the data span.

When the eight drift terms were eliminated from the model, the change in $-2 \ln$ likelihood was not significant, indicating that the inclusion of drift terms did not improve the model significantly. This did not agree with the previous study where there were significant trends for clocks 601, 137, and 61, but there are explanations for this. In the first study clock 601 was nearing the end of its life, and the cesium beam tube was replaced between the two studies making it essentially a different clock. Clock 137 was introduced as a new clock on the first day of the first study. Clocks are known to drift near the beginning and end of their lives and be more stable during midlife. Clock 61 was only available for 69 days in the second study so any drift that may have existed could not be detected. If a common drift exists in all clocks, this could not be estimated, and the uncertainty introduced into the time scale would not appear in the $P(t|t)$ matrix.

It is of interest to compare the confidence intervals obtained on two different sets of data. Table 2 shows the results for the clocks that were common to the two studies. The estimates and confidence intervals are in agreement except for clock 601 in which the cesium beam tube was replaced.

TABLE 2. *1979 and 1981 clock characterization 95% confidence intervals.*

Clock	Study	$\sigma_\varepsilon \sim$ nanoseconds			$\sigma_\eta \sim$ nanoseconds/day		
		Lower Limit	Est	Upper Limit	Lower Limit	Est	Upper Limit
601	1	6.88	7.46	8.14	0	0.44	0.96
	2	1.89	2.13	2.41	0	0.06	0.52
167	1	12.44	13.45	14.61	0.56	1.11	1.96
	2	12.58	13.52	14.67	0.57	1.11	2.07
137	1	9.26	10.04	10.96	1.03	1.60	2.26
	2	10.41	11.31	12.27	1.76	2.49	3.56
1316	1	3.17	3.62	4.05	0.97	1.36	1.82
	2	3.81	4.14	4.53	0.53	0.80	1.23
323	1	3.11	3.53	3.94	0.41	0.73	1.10
	2	2.06	2.37	2.74	0.63	0.94	1.34
8	1	8.32	9.09	9.95	2.01	2.65	3.48
	2	7.98	8.65	9.49	2.11	2.76	3.66

6. Conclusion

A major advantage of a Kalman algorithm for estimating time is its flexibility when changes are needed. An example is the introduction of frequency calibrations from a frequency standard. A frequency calibration, together with its precision, can easily be introduced into the state vector and state covariance matrix. Experimentation with different Kalman models are possible since changing the state transition matrix is not difficult. However, the major advantage is that the Kalman algorithm is based on a mathematical model which fits the data. The innovations can be tested for whiteness, and standard statistical methods can be used for obtaining confidence intervals and hypothesis tests.

Acknowledgment

References

[1] Tryon, P. V. and Jones, R. H. (1982). Estimation of parameters in models for cesium beam atomic clocks. Second Symposium on Atomic Time Scale Algorithms, June 23-25, 1982, Boulder, CO.

[2] Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the theory of statistics.* Third edition. New York: McGraw-Hill.

[3] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME (Journal of Basic Engineering), 82D, 35-45.

[4] Graybill, F. A. (1976). *Theory and application of the linear model.* North Scituate: Duxbury Press.

Appendix

For computational efficiency and numerical stability, a procedure involving a Cholesky decomposition is used in the Kalman recursion [4]. The tests for errors are also simplified. The matrix product $H(t+1) P(t+1|t)$ in eq. (4.5) is calculated without actually performing a matrix multiplication. Since each row of $H(t+1)$ contains a one and a minus one with the rest of the elements zero, each element of the matrix product is simply the difference between two elements of $P(t+1|t)$. The matrix $C(t+1)$ is calculated and augmented as follows

$$[C(t+1)\vdots H(t+1)\ P(t+1|t)\vdots I(t+1)\vdots A_1\vdots A_2\vdots \ldots \vdots A_m]. \quad (A.1)$$
$$(m-1) \qquad\quad 3m \qquad\quad 1 \quad\ 1\ \ 1\ \cdots\ 1$$

This matrix has $m-1$ rows with the number of columns shown under each partition. Only the upper triangular position of $C(t+1)$ need be stored. The third partition is the vector of innovations from eq (4.4) followed by the vectors to be used for testing for clock errors. One of these vectors will correspond to the reference clock eq (4.7) and the others to the remaining clocks eq (4.9).

The Cholesky decomposition factors $C(t+1)$ into

$$C(t+1) = T'(t+1)\ T(t+1), \quad (A.2)$$

where $T(t+1)$ is upper triangular. The algorithm as given in [4] works on the entire augmented matrix and is equivalent to premultiplying the matrix by $[T'(t+1)]^{-1}$. Let

$$B(t+1) = [T'(t+1)]^{-1}\ H(t+1)\ P(t+1|t)$$

$$D(T+1) = [T'(t+1)]^{-1}\ I(t+1) \quad (A.3)$$

$$G_i = [T'(t+1)]^{-1}\ A_i.$$

The test for each clock, eq (4.10) becomes

$$b = G'D(t+1)/G'G, \quad (A.4)$$

and eq (4.11) becomes

$$s.e.(b) = (1/G'G)^{1/2} \quad (A.5)$$

The quadratic form in eq (4.13) reduces to

$$Q = D'(t+1)\ D(t+1), \quad (A.6)$$

and eq (4.15) to

$$\Delta(t+1) = B'(t+1)\ [T'(t+1)]^{-1}. \quad (A.7)$$

The final updating equations are (4.16)

$$X(t+1|t+1) = X(t+1|t) + B'(t+1)\ D(t+1), \quad (A.8)$$

and (4.17)

$$P(t+1|t+1) = P(t+1|t) - B'(t+1)\ B(t+1). \quad (A.9)$$

Measurement Assurance Program Transmittance Standards for Spectrophotometric Linearity Testing:* Preparation and Calibration

Kenneth L. Eckerle,** Victor R. Weidner,** Jack J. Hsia,** and Karen Kafadar†

National Bureau of Standards, Washington, DC 20234

A Measurement Assurance Program for spectrophotometry is being established in order to assist laboratories involved in spectrophotometric calibrations. This paper deals with the preparation and calibration of neutral density glass filters for checking the linearity of photometric response, as applied to spectral transmittance measurements. Several sets of filters were prepared from suitable neutral glass to provide nominal transmittances of 92, 70, 50, 25, 10, 1, and 0.1% at a wavelength of 548.5 nm. These filter sets will be available in three sizes: these are, 38 mm diameter aperture in 51 x 51 mm holder, 25 mm diameter aperture in 51 x 28 mm holder, and 30 x 8 mm aperture in a cuvette holder. The filters were calibrated for spectral transmittance on the NBS Reference Spectrophotometer for high accuracy transmittance measurements. Measurements were made with a 1.5 nm passband collimated sample beam. The filters were checked for uniformity and measurements were made to determine the effects of sample beam polarization. The transmittance data for the wavelength range of interest were analyzed by statistical methods to determine the effects of passband for a range of 1.5 nm to 10.5 nm passband. The results of these measurements are presented in tabular and graphical detail for the master filter set.

Key words: experimental design; filter uniformity; linearity testing; measurement assurance program; neutral density glass; passband effects; polarization effects; polynomial fitting; spectrophotometry; statistics; transmittance standards.

1. Introduction

This paper deals with the preparation and calibration of neutral density glass filters for checking the linearity of photometric response, as applied to spectral transmittance measurements.

The application of absolute techniques, such as the double-aperture method [1],[1] for checking the linearity of a spectrophotometer's response, is difficult to realize on many instruments because of problems involving sample and reference beam geometry or detector configurations. The other approach to checking linearity is through the use of a set of transmittance standards. Several sets of glass filters have been prepared and calibrated for this purpose. A Measurement Assurance Program (MAP) is being established through which these sets of transmittance standards will be used for the purpose of improving the accuracy of spectrophotometer measurements in laboratories participating in the program.

2. Selection and Preparation of the Standards

Some of the factors to be considered in selecting suitable filters for transmittance standards are: available range of transmittances, light scattering properties, uniformity, stability, passband sensitivity, and wavelength sensitivity. Schott[2] neutral density glass types NG-4, NG-9, and NG-11 were selected on the basis of these characteristics and properties as well as previous experiences with Schott glass by NBS in the preparation of Standard Reference Material (SRM) 930 [2].[3] Two levels of transmittance were prepared from each of the glass types as follows:

---

[2]Certain commercial materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards nor does it imply that the materials identified are necessarily the best available for the purpose.

[3]SRM 930 consists of three filters with nominal transmittances of 10%, 20%, and 30% calibrated at several wavelengths. Originally these SRM filters were designed for use by clinical chemists and are sold by the set. Our goal with the present filters was to provide a larger dynamic range so that the filters could be used to measure linearity in applications when other methods are difficult if not impossible. Also, these filters are not for sale, but are to be retained by NBS and issued only temporarily as part of a MAP measurement service.

Glass Type	Nominal Transmittance	Equivalent Transmission Density
NG-11	70%	0.15
NG-11	50%	0.3
NG-4	25%	0.6
NG-4	10%	1.0
NG-9	1%	2.0
NG-9	0.1%	3.0

An additional filter of borosilicate crown glass having a nominal transmittance of 92% (transmission density 0.036) is included in the set.

The filter sets were prepared in three sizes in order to accommodate different instruments. The largest filters are mounted in 51 × 51 mm holders with a filter aperture of 38 mm diameter. The intermediate size filters are mounted in 51 × 38 mm holders with a filter aperture of 25 mm. The smallest filters are mounted in cuvette holders and provide a filter aperture of 30 × 8 mm.

The preparation of the filters from the stock glass was done in the NBS Optical Shop. The tolerances on parallelism of the two faces of the filters was maintained at approximately 0.01°. They were polished to a flatness of 3 fringes or better.

Calculations of the required thicknesses were made from internal transmittance data supplied by the manufacturer. The internal transmittance of 1 mm path length for the three glass types is as follows:

Glass Type	1 mm path
NG-4	0.32
NG-9	0.035
NG-11	0.78

The values of transmittance for a wavelength of 500 nm were computed from these internal transmittance values. These transmittance values were converted to equivalent transmission density. Since the transmission density is linear with thickness, it is possible to determine, by graphical techniques, the required thickness for a desired transmittance value. Figure 1 shows the relationship of transmission density to thickness for the glass types NG-4, NG-9, and NG-11. The degree of accuracy in determining the required thickness for a given

transmittance value by this graphical technique is dependent on the initial accuracy of the internal transmittance values used to determine the transmittance of 1 mm and 2 mm thicknesses. The actual transmittance values achieved by this technique closely approximated the desired nominal transmittance values for the filter set.
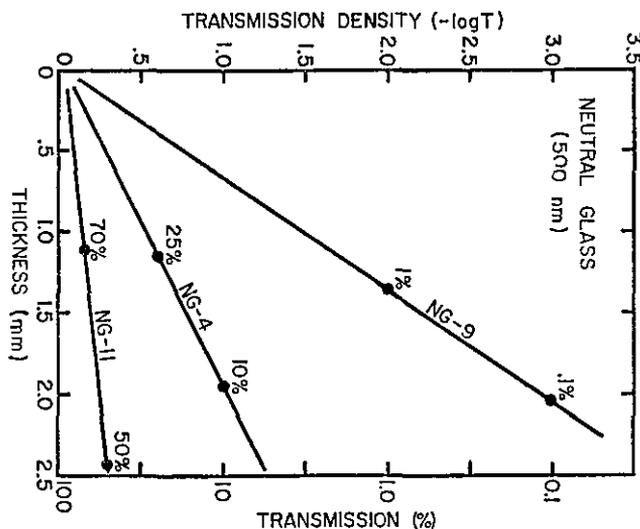
FIGURE 1. Transmission density versus thickness of Schott NG-4, NG-9, and NG-11 neutral glasses as derived from internal transmission data for 1 mm and 2 mm thicknesses.

3. Transmission Analysis

3.1 Light Scattering

Scattering of a collimated sample beam transmitted through the neutral filters was quantitatively assessed by the ASTM recommended method [3] for measuring haze percentage. The measurements were made on the reference hazemeter [4] using the ASTM recommended geometry and methods. This instrument has a well collimated circular incident beam having a color temperature of approximately 6800 K. A visual response filter at the detector modifies its response to give the hazemeter a peak sensitivity at approximately 550 nm. The instrument measures total forward scattering relative to the total transmission within the definition of this response and for the ASTM recommended geometry. The haze percentage determined for a set of filters was found to be as follows:

26

Filter	Haze Percentage	Nominal Transmittance
1-1	0.05%	92%
1-2	0.02%	70%
1-3	0.05%	50%
1-4	0.06%	25%
1-5	0.04%	10%
1-6	—*	1%
1-7	—*	0.1%

* The 1 and 0.1% transmitting filters were too low in transmittance for analysis by the hazemeter. It was concluded from these results that the filters scattered less than 0.1% of the transmitted sample beam.

3.2 Survey Transmittance Measurements

The purpose of a survey study of the neutral filter transmittance was to assess their over-all spectral properties in the visible spectrum and select a specific wavelength for detailed analysis and calibration. Since the filters are not perfectly neutral, it was desirable that a wavelength be selected for calibration, at which the transmittance values of the filters would be least sensitive to variation in passband and wavelength scale errors associated with various spectrophotometers.

With the exception of the borosilicate crown glass filter, the filters in the set are of a similar type glass with a common dye in three different concentrations for the NG-4, NG-9, and NG-11 glass types. The spectral transmission density curves of these filters in the spectral range between approximately 440 nm and 620 nm shows that the absorptions appear at the same wavelengths for the different concentrations associated with these glass types. Figures 2 through 8 illustrate the results of a 1-nm interval survey of the spectral transmittance of the filters between 440 and 620 nm. These measurements were made with a commercial spectrophotometer. The slit-width was maintained at approximately 0.1 mm (0.37 nm passband). Further measurements were made of the 10% filter with slit-widths of 1.0, 2.0, and 3.0 mm. These slit-widths correspond to passbands of 3.7, 7.4, and 11.1 nm, respectively.

There are four wavelengths within the wavelength range 440 to 620 nm corresponding to peaks and valleys that could be used for the purpose of establishing a calibrated set of photometric scale standards. However, only one wavelength is needed and the absorption peak at approximately 548.5 nm was selected as being the most suitable. The spectral peak at approximately 464 nm and the absorption valleys at approximately 510 and 591 nm do not show a significant change in transmittance for the range of passbands used in this survey. A

photometric resolution of approximately ±0.01% is required in order to detect the effect of passband on the spectral transmittance measured at these wavelengths.
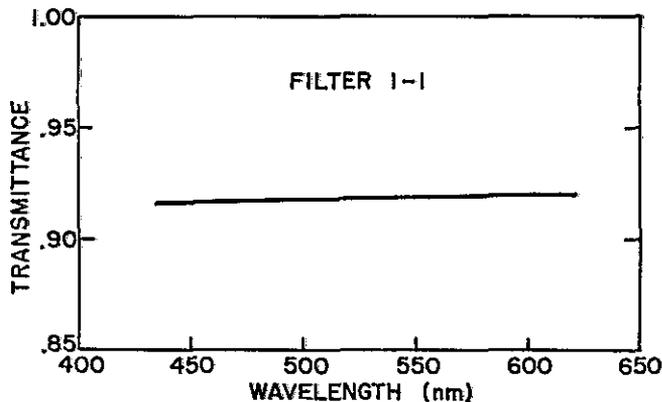
FIGURE 2. Spectral transmittance of Master Filter 1-1.
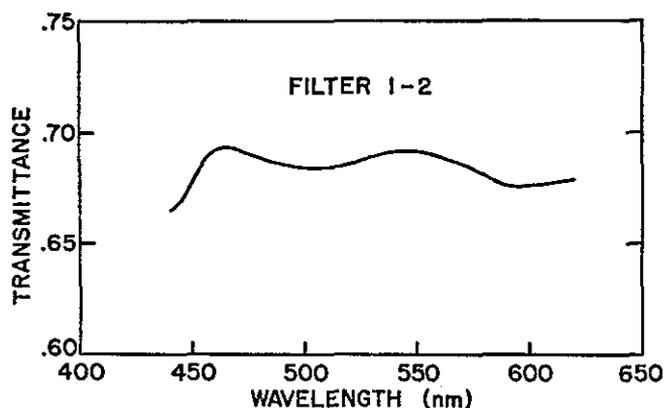
FIGURE 3. Spectral transmittance of Master Filter 1-2.
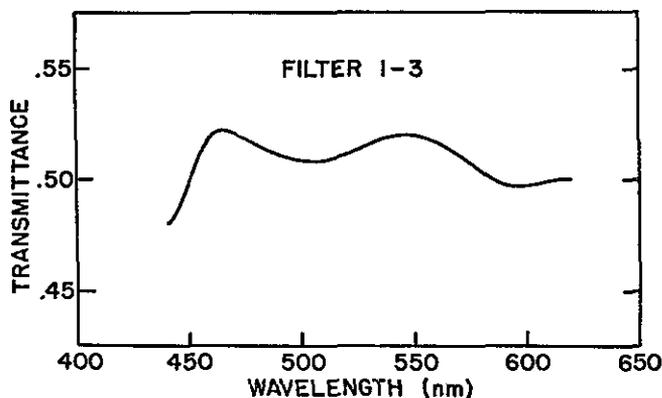

FIGURE 4. Spectral transmittance of Master Filter 1-3.
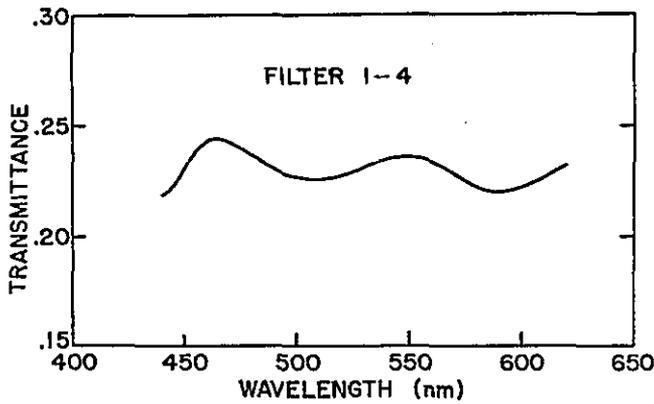
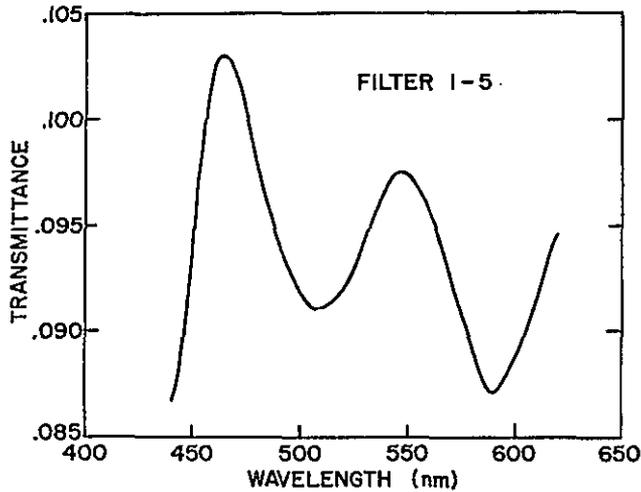FIGURE 5. Spectral transmittance of Master Filter 1-4.

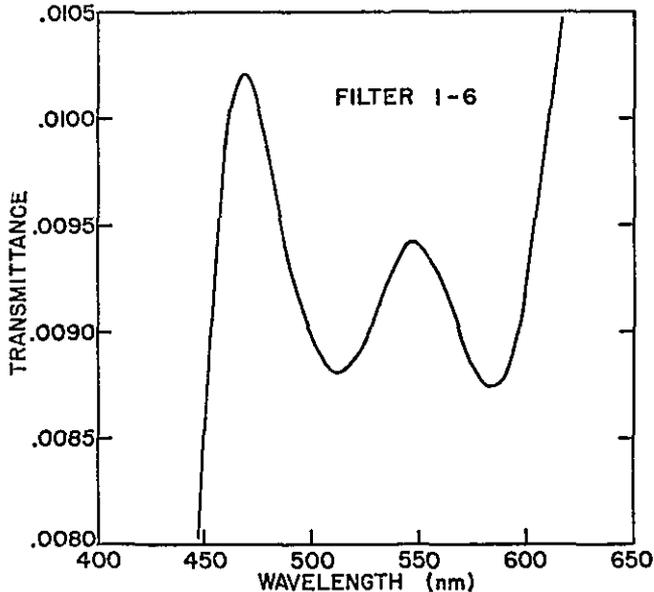FIGURE 6. Spectral transmittance of Master Filter 1-5.

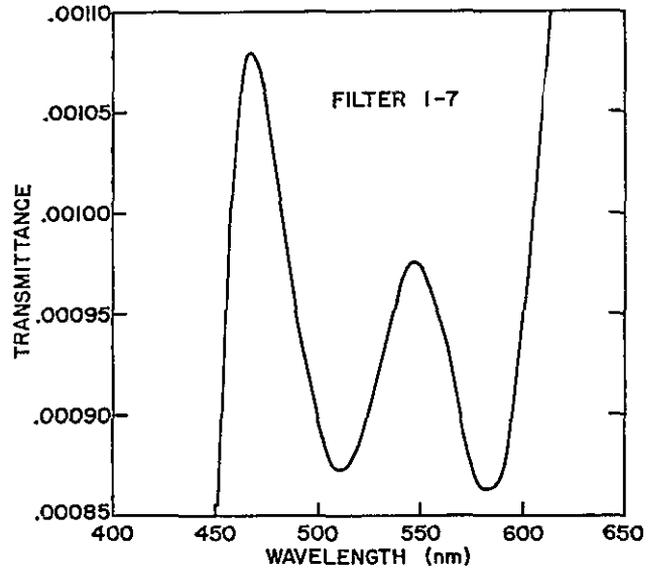FIGURE 7. Spectral transmittance of Master Filter 1-6.

FIGURE 8. Spectral Transmittance of Master Filter 1-7.

3.3 High Accuracy Transmittance Measurements

Calibration of the spectral transmittance of the master set of filters and three other sets was done on the NBS Reference Spectrophotometer [5,6,7] for measuring high accuracy transmittance. The measurements were made with a passband of 1.5 nm and a beam diameter of 10 mm. Measurements were made for perpendicularly (s) and horizontally (p) polarized incident sample beams at 1.5 nm intervals from 545.5 to 557.5 nm. Each filter was scanned for uniformity at 547 nm. The uniformity measurements were made at three locations; center, 2 mm horizontally from center, and 2 mm below center. Transmittances for master set #1 filters are listed in table 1. The results of the uniformity scan at 547 nm are listed in table 2. The transmittance values listed in table 1 are an average of the two polarizations. The values for both polarizations are listed for the uniformity measurements made at 547 nm. The instrumental uncertainties in the values listed in tables 1 and 2 are estimated to be ±0.04% for filter 1-1 to ±0.0005% for filter 1-7. A complete description of the measurement sequence and data reduction for the high accuracy measurements and an analysis of the errors associated with such measurements is given in refs. [5-7].

4. Statistical Analysis

A statistical analysis of the data obtained on the spectral transmittance of the neutral filters was used to determine the magnitude of change in transmittance values due to changes in passband and also the magnitude of

28

TABLE 1. *Spectral Transmittance of Master Set No. 1 for a Passband of 1.5 nm.*

Wave-length (nm)	Filter 1-1	Filter 1-2	Filter 1-3	Filter 1-4	Filter 1-5	Filter 1-6	Filter 1-7
536.5	0.9164	0.6902	0.5177	0.2339	0.0955	0.00919	0.000915
538.0	.9165	.6905	.5181	.2345	.0959	.00922	.000922
539.5	.9165	.6908	.5185	.2350	.0962	.00926	.000925
541.0	.9165	.6910	.5187	.2353	.0964	.00928	.000929
542.5	.9166	.6912	.5190	.2357	.0966	.00931	.000933
544.0	.9166	.6913	.5192	.2360	.0969	.00932	.000935
545.5	.9166	.6913	.5192	.2362	.0970	.00933	.000936
547.0	.9166	.6914	.5194	.2363	.0971	.00934	.000938
548.5	.9168	.6913	.5193	.2363	.0971	.00934	.000938
550.0	.9167	.6912	.5191	.2363	.0971	.00934	.000940
551.5	.9167	.6912	.5190	.2362	.0971	.00933	.000937
553.0	.9167	.6911	.5190	.2361	.0970	.00931	.000934
554.5	.9166	.6908	.5187	.2358	.0968	.00930	.000934
556.0	.9168	.6908	.5184	.2355	.0966	.00928	.000929
557.5	.9168	.6904	.5180	.2351	.0964	.00924	.000923

TABLE 2. *Transmittance Uniformity of Master Set No. 1 at 547 nm.*

Location, Polarization	Filter 1-1	Filter 1-2	Filter 1-3	Filter 1-4	Filter 1-5	Filter 1-6	Filter 1-7
1, p	0.9166	0.6913	0.5193	0.2362	0.0971	0.00934	0.000939
1, s	.9166	.6913	.5193	.2362	.0971	.00935	.000940
2, p	.9166	.6914	.5193	.2361	.0971	.00934	.000937
2, p	.9167	.6914	.5193	.2362	.0971	.00934	.000938
3, p	.9164	.6913	.5192	.2362	.0971	.00934	.000938
3, p	.9165	.6913	.5192	.2362	.0971	.00933	.000939

change due to location of the sample beam (filter uniformity).

Preliminary data from the commercial instrument indicated that the central wavelength of interest was at approximately 547 nm. However, a further analysis of the data from the high accuracy instrument indicated that 548.5 nm was a better choice.

Data are presented in table 3 to illustrate the effect of passband on the transmittance at the central wavelength for the master set #1 filters. These values are based on calculation for a triangular passband, using continuous integration. Figures 9 through 15 show these results plotted with the standard error limits. The data listed in table 4 illustrate the magnitude of change due to position location for the master set of filters.

Remeasurement of the spectral transmittances of the neutral filters was made one year after initial calibrations. The results of this second calibration for master set #1 are compared with the original calibration and listed in table 5 for wavelength 548.5 nm. Differences in spectral transmittances of the filters shown in table 5 are too small to be clearly interpreted as changes with the possible exception of filter 1-1. Here the apparent change in transmittance was −0.0009. Further measurements at longer time intervals will be required to confirm any real changes in spectral transmittance due to aging.

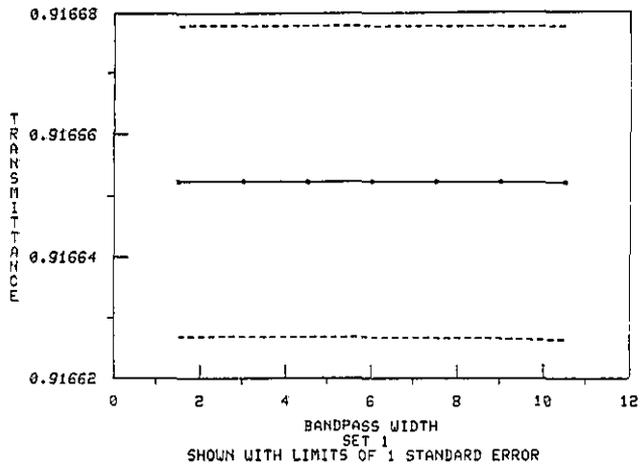Additional details of the statistical analysis are given in appendix A.

TABLE 3. *Transmittance of Master Set No. 1 at 548.5 nm as a Function of Passband for a Triangular Passband.*

Passband (nm)	Filter 1-1	Filter 1-2	Filter 1-3	Filter 1-4	Filter 1-5	Filter 1-6	Filter 1-7
1.5	0.91665 (3)	0.69133 (2)	0.51928 (2)	0.23632 (2)	0.09713 (1)	0.009342 (2)	0.000939 (4)
3.0	.91665 (3)	.69132 (2)	.51927 (2)	.23632 (2)	.09711 (1)	.009341 (2)	.000939 (4)
4.5	.91665 (3)	.69130 (2)	.51924 (2)	.23628 (2)	.9709 (1)	.009338 (2)	.000938 (4)
6.0	.91665 (3)	.69128 (2)	.51921 (2)	.23624 (2)	.09706 (1)	.009335 (2)	.000938 (4)
7.5	.91665 (3)	.69125 (2)	.51916 (2)	.23618 (2)	.09703 (1)	.009332 (2)	.000937 (4)
9.0	.91665 (3)	.69121 (2)	.51911 (2)	.23616 (2)	.09698 (1)	.009327 (2)	.000937 (4)
10.5	.91665 (3)	.69117 (2)	.51905 (2)	.23604 (2)	.09693 (1)	.009321 (2)	.000936 (4)

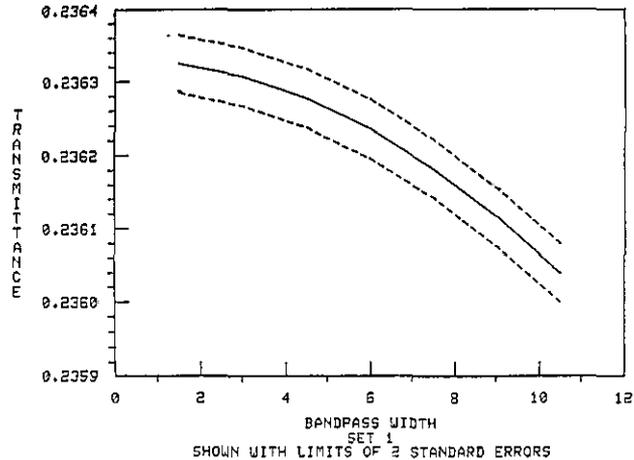Note: uncertainties (one standard deviation) are in parentheses

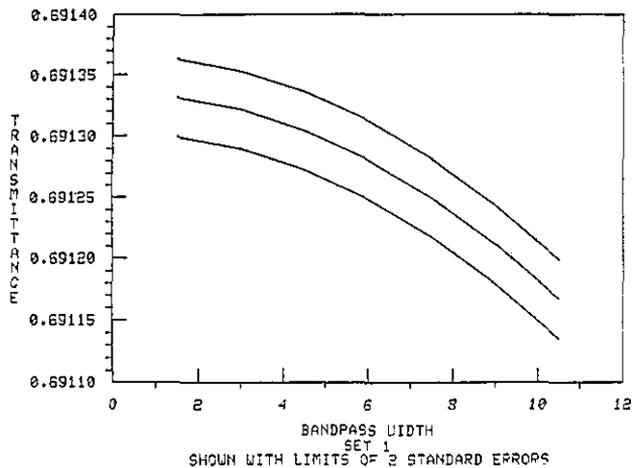FIGURE 9. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-1.

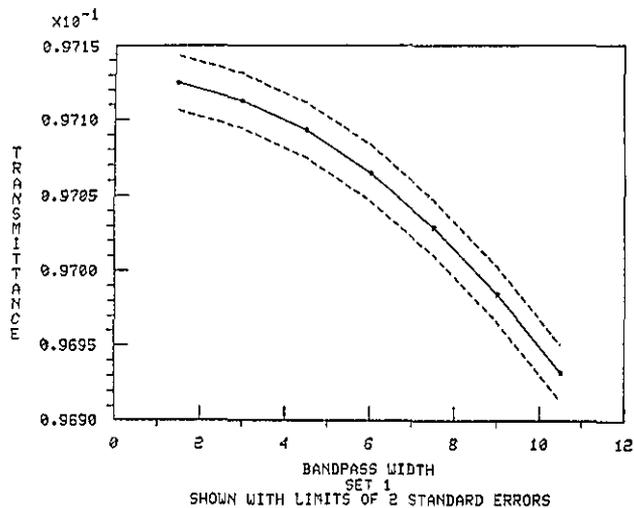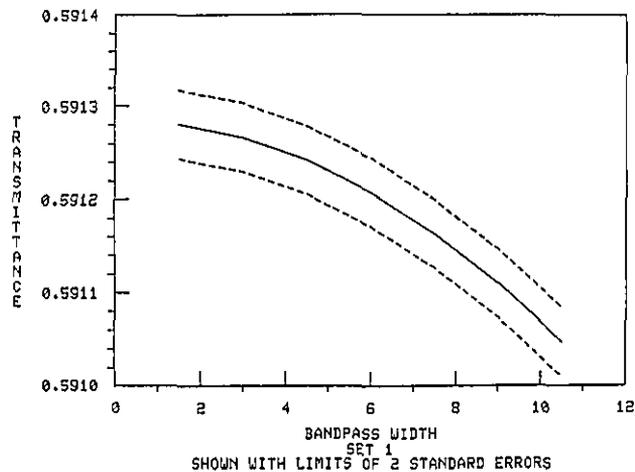FIGURE 10. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-2.

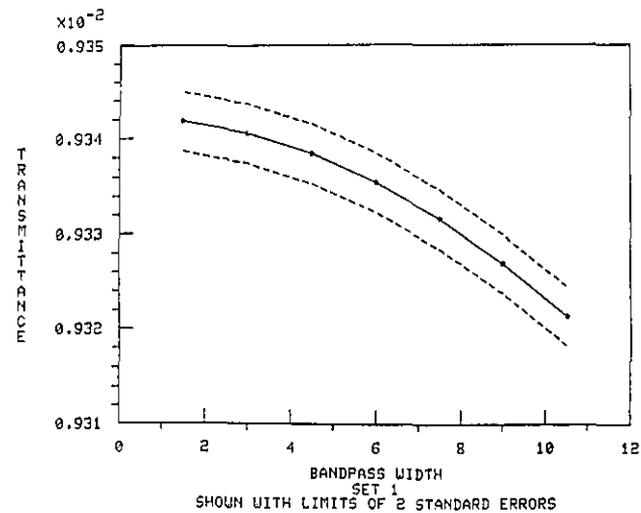FIGURE 11. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-3.

FIGURE 12. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-4.
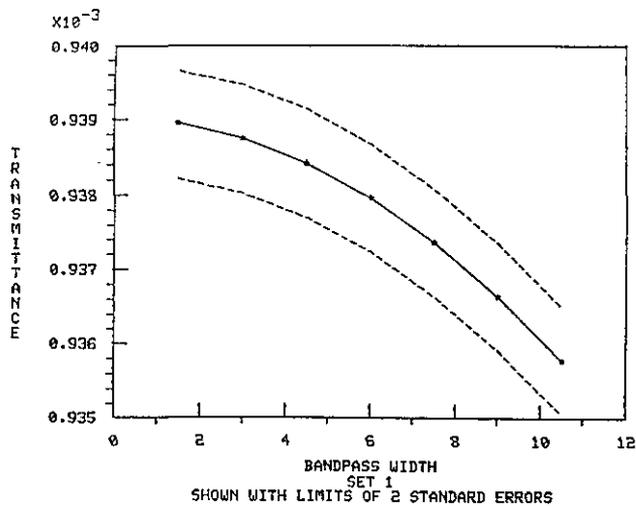

FIGURE 13. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-5.

FIGURE 14. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-6.

30

FIGURE 15. Spectral transmittance at 548.5 nm versus bandpass for Master Filter 1-7.

TABLE 4. *Magnitude of Filter Nonuniformity*

Filter No.	Location	Transmittance at 548.5 nm	Location Difference	
1-1	1	0.91658	(1)–(2) =	–0.00008
1-1	2	.91666	(2)–(3) =	.00018
1-1	3	.91648	(1)–(3) =	.00010
1-2	1	.69132	(1)–(2) =	– .00006
1-2	2	.69138	(2)–(3) =	.00007
1-2	3	.69131	(1)–(3) =	.00001
1-3	1	.51926	(1)–(2) =	– .00002
1-3	2	.51928	(2)–(3) =	.00009
1-3	3	.51919	(1)–(3) =	.00007
1-4	1	.23624	(1)–(2) =	.00009
1-4	2	.23615	(2)–(3) =	– .00004
1-4	3	.23619	(1)–(3) =	.00005
1-5	1	.09709	(1)–(2) =	.00001
1-5	2	.09708	(2)–(3) =	.00000
1-5	3	.09708	(1)–(3) =	.00001
1-6	1	.009341	(1)–(2) =	– .000001
1-6	2	.009342	(2)–(3) =	.000011
1-6	3	.009331	(1)–(3) =	.000010
1-7	1	.000939	(1)–(2) =	.000001
1-7	2	.000938	(2)–(3) =	.000001
1-7	3	.000937	(1)–(3) =	.000002

5. Summary

One master set and three working sets of neutral density glass filters have been calibrated for use as spectral transmittance standards for checking the photometric scale linearity of spectrophotometers. Each set consists of seven filters ranging in transmittance from 0.1 to

92%. These filters will be used in a Measurement Assurance Program (MAP). The purpose of this program will be to assist laboratories wishing to maintain a high level of confidence in the accuracy of their spectrophotometric measurements.

Data are presented that show the spectral transmittances of these filters at 548.5 nm to be relatively insensitive to variations in passband and slight errors in instrument wavelength setting. The filters are of sufficient uniformity for use as transmittance standards.

TABLE 5. *Spectral Transmittance of Master Set No. 1 for a Passband of 1.5 nm at 548.5 nm.*

Filter No.	(Apparent changes in transmittance)		Apparent change
	First Calibration (Oct. 1980)	Second Calibration (Oct. 1981)	
1-1	0.9168	0.9159	–0.0009
1-2	.6913	.6914	.0001
1-3	.5193	.5196	.0003
1-4	.2363	.2365	.0002
1-5	.09712	.09722	.00010
1-6	.009337	.009350	.000013
1-7	.0009376	.0009395	.0000019

6. References

[1] Mielenz, K. D. and Eckerle, K. L. Spectrophotometer Linearity Testing Using the Double-Aperture Method, Appl. Opt. 11 (10), 2294 (Oct. 1972).

[2] Mavrodineanu, R. and Baldwin, J. R. Glass Filters as a Standard Reference Material for Spectrophotometer—Selection, Preparation, Certification, Use SRM 930. NBS Special Publication 260-51 (November 1975).

[3] ASTM D1003, Haze and Luminous Transmittance of Transparent Plastics, ASTM Subcommittee D20, 40 on Optical Properties of Plastics,

[4] Weidner, V. R. and Hsia, J. J. NBS reference hazemeter: its development and testing. Appl. Opt. 18 (10) 1619 (May 15, 1979).

[5] Mielenz, K. D. and Eckerle, K. L. Design, Construction, and Testing of a New High Accuracy Spectrophotometer. Natl. Bur. Stand. (U.S.) Tech. Note 729; June 1972. 55 p.

[6] Eckerle, K. L. Modification of an NBS Reference Spectrophotometer. Natl. Bur. Stand. (U.S.) Tech. Note 913; July 1976, 39 p.

[7] Mielenz, K. D., Eckerle, K. L. Madden, R. P., and Reader J. New Reference Spectrophotometer. Appl. Opt. 12 (7), 1630 (July 1973).

[8] Davis, Philip J., *Interpolation and Approximation*. New York: Blaisdell; 1965.

[9] Federov, V.V., *Theory of Optimal Experiments*. New York: Academic Press; 1972.

[10] Keifer, J. and Studden, W. J. Optimal designs for large degree polynominal regression, Ann. Statist. 4, 1113-23 (1976).

[11] Karlin, Samuel and Studden, William. Optimal experimental designs. Ann. Math. Statist. 37, 783-815 (1966).

Appendix

Statistical Design and Analysis for Calibrating the Transmittance of Filters for Spectrophotometer Linearity Testing

For four sets of filters, each set containing seven filter types of various transmittances, an assigned transmittance value needed to be determined when a triangular passband of given width is superimposed. Additionally, the polarization and location of the spot size on the filter had an unknown influence on the transmittance value. Although a commercial instrument can scan the entire spectrum at little cost, the measurements are an order of magnitude less accurate than those obtained using the high accuracy instrument. For this reason, a design was developed to determine:

(1) The values to be measured for purposes of calibration, and
(2) The existence of a possible effect due to polarity or the effect of location on the filter.

The motivation behind the choice of design is discussed in section A below. In the event of (2), the calibration must be reconsidered. On the basis of these extremely precise measurements, an effect due to location was statistically significant. This suggests that either:

(1) The calibrated values can be used only for the central location on each filter; or
(2) Additional measurements will be required to determine the magnitude of changes in transmittance due to spot location.

Since the observed differences are small (<0.2%; see table 4), the location effect may not interfere with practical usage of these filters. However, measurement as near to the central location on each filter as possible is recommended. The conclusions regarding location and polarity effects are presented in section B, and the estimation of the transmittance curve leading to the calibration values is shown in sections C and D. The method of extension for general passbands is given in section E.

A. The Design

Most organizations using these filters will have triangular passbands on their measuring equipment. Therefore, adequate determination of the transmittance using this passband is essential. However, other passband shapes may be used, and for these, a more general approach is taken. The problem then becomes one of estimating the transmittance curve with high precision. Since the measurements on the high accuracy instrument are time-consuming, a design was chosen so as to permit adequate curve estimation with relatively few points.

1. Preliminary analysis: Choice of Wavelength.

The choice for a wavelength to be calibrated required consideration of:

(i) Varying passband widths, and
(ii) The potential for imprecise wavelength specification.

We consider neighborhoods of 21 nanometers, since the widest passband that is most commonly used weights frequencies within a range of this length. The chosen wavelength is:

$$\lambda_o = 548.5 \text{ nm.}$$

This subsection explains this choice in light of considerations (i) and (ii).

For a passband having width w and functional form $h_w(\lambda)$, the transmittance at wavelength λ_o is given by the convolution integral

$$\tau_w(\lambda_o) = \int_{\lambda_o-\omega}^{\lambda_o+\omega} h_w(\lambda)f(\lambda-\lambda_o)d\lambda \Big/ \int_{\lambda_o-\omega}^{\lambda_o+\omega} h_w(\lambda)d\lambda,$$

where $f(\lambda)$ is the spectrum (transmittance curve). If the passband is normalized so that

$$\int_{\lambda_o-\omega}^{\lambda_o+\omega} h_w(\lambda)d\lambda = 1,$$

then this simplifies to

$$\tau_w(\lambda_o) = \int_{\lambda_o-\omega}^{\lambda_o+\omega} h_w(\lambda)f(\lambda-\lambda_o)d\lambda . \qquad (1)$$

In all that follows, we will assume that the passband h_w has been normalized in this way.

In order for this average value to be relatively insensitive to wavelength specification, $f(\lambda)$ needs to be a smooth, slowly-varying function. In mathematical terms, we search for a neighborhood where the first two derivatives of the spectrum are fairly small. This suggests a region where $f(\lambda)$ is nearly constant, or, at worst, a quadratic having very small degree of curvature. A locally linear spectrum with a noticeably nonlinear slope would yield a seriously biased estimate, whereas a locally quadratic spectrum, centered at a peak or trough, would have a negligible linear term and therefore a bias which depends primarily on the second-order term (hopefully

small). Preliminary readouts on a commercial instrument suggested a central wavelength of 547.0 nm for most filters. On the basis of the first set of measurements on the high accuracy instrument, however, this was later amended to $\lambda_o = 548.5$ nm.

2. Number of Points

Assuming that the spectrum is locally quadratic and has a maximum around $\lambda_o = 548.5$ nm, our goal is to estimate the three parameters α, β, γ in the approximation:

$$f(\lambda) = \alpha + \beta\lambda + \gamma\lambda^2. \qquad (2)$$

In order to estimate α, β, and γ, by $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, the proposed design specifies a total of *five* points, one at the central wavelength, and one each at equal distances $\pm 2\delta$ from the central wavelength.

Let $x = \lambda - 548.5$, so as to center the relation at the origin. A minimum of three points is required to specify the parameters. This leaves no room for the assessment of error. With more points, the variance of (1) will be a linear combination of the variances and covariances of the parameters, and will be dominated primarily by the variance of the constant term. If our chosen points are symmetrically placed about 0, it can be shown that:

$$\text{Var}(\hat{\alpha}) = \sigma^2/(n - c^2/e), \qquad (3)$$

where

$$c = \Sigma x_i^2 = 2\Sigma_{x_i \geqslant 0} x_i^2$$

$$e = \Sigma x_i^4 = 2\Sigma_{x_i \geqslant 0} x_i^4.$$

As $(\Sigma x_i^4) \leqslant (\Sigma x_i^2)^2$ $(x_i \geqslant 0)$, the smallest value of c^2/e is 2 (all points at the origin except for the two at ± 1). A lower bound on (3) is therefore:

$$\text{Var}(\hat{\alpha}) = \sigma^2/(n-2). \qquad (3')$$

If we concern ourselves with only the first term, then, relative to σ^2, eq (3') gives the reduction in the overall variance that we may hope to gain in our variance of the transmittance value. With $n = 3$, there is no reduction; with $n = 5$, the variance is already reduced to 33% of its value. An extra two points provides only an extra 13% improvement. Thus, the design specifies a minimum of two points on either side of the central wavelength.

3. Location of Points

We consider two criteria in selecting the four wavelengths (two symmetrically placed on either side of the central wavelength):

(i) Small error in quadratic interpolation of the transmittance curve;
(ii) Small variance in the transmittance estimate given by (I).

On the basis of these two criteria, the proposed design specifies measurements to be taken at:

$$\lambda_o, \lambda_o \pm 6 \text{ nm}, \lambda_o \pm 9 \text{ nm}.$$

The reasoning behind this choice is explained by considering (i) and (ii) separately.

(i) Error is quadratic interpolation. A bound for the estimate of error in polynominal interpolation of degree n at the points $a \leqslant x_1, x_2, \ldots x_{n+1} \leqslant b$ is

$$|\varepsilon(\lambda)| \leqslant \max_{a \leqslant \lambda \leqslant b} |f^{(n+1)}(\lambda)| \cdot |\lambda - x_1| |\lambda - x_2| \cdots |\lambda - x_{n+1}|$$

(see, e.g., [8]). The first part of this error depends upon the transmittance curve and led to the choice of $l_0 = 548.5$ in subsection 1 above. Now we concentrate on selecting $x_1, \ldots, x_{n+1}$ so that the error is as small as possible. The answer to this problem is given by the zeroes of the Chebyshev polynomials, namely

$$x_k = (10.5 \text{nm}) \cdot \cos[(2k-1)/(2n+2)], \ k = 1, \ldots, n + 1. \qquad (4)$$

For the problem at hand, $n+1 = 5$; thus:

$$x_1 = -x_5 = 9.986 \text{ nm from } \lambda_o$$
$$x_2 = -x_4 = 6.172 \text{ nm from } \lambda_o$$
$$x_3 = 0 \text{ nm from } \lambda_o.$$

As most passbands are in multiples of 1.5 nm, the closest multiples to these points are ± 6 nm and either ± 9 nm or ± 10.5 nm. However, the nature of the triangular passband which is frequently used in practice would assign zero weight to the values ± 10.5 nm. Since a primary goal is the estimation of the transmittance using this passband, we choose ± 9 nm.

(ii) Minimum variance of transmittance estimate. If we estimate the transmittance curve using a quadratic function, viz.

$$\hat{f}(x) = \hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2, \qquad (5)$$

then the transmittance using passband of width w is

$$\hat{\tau}_w(\lambda_o) = \int_{-w}^{w} h_w(w + \lambda_o)(\hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2)dx.$$

Let

$$c_j = \int_{-w}^{w} x^j\, h_w(x)\, dx$$

Then

$$\hat{\tau}_w(\lambda_o) = \hat{\alpha}c_o + \hat{\beta}c_1 + \hat{\gamma}c_2 = c'\hat{\Theta}$$

where

$$c' = (c_o, c_1, c_2)$$

$$\hat{\Theta}' = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}).$$

Then minimizing the variance of $c'\hat{\Theta}$ is equivalent to

$$\min \mathrm{Var}(c'\hat{\Theta}) = \min c'\, \mathrm{Var}(\hat{\Theta})\, c.$$

A design which minimizes the variance-covariance matrix $\mathrm{Var}(\hat{\Theta})$ of the parameter estimates is given in [9]. The design would place the two observations each at $\lambda_o + 10.5$ nm in addition to the one at λ_o. If our function is truly quadratic, such a design would be optimal.

For many reasons, however, we modify the optimal design which permits more flexibility in our choice of model. Such a design is suggested in [10]. The design recommends two different values rather than repeating them at the endpoints.

Closely related to this design is the one which minimizes the maximum variance of the best linear unbiased estimate of the function $f(x)$ given by eq (5). Reference [11] shows that the five points should be placed at the zeros of the polynomial

$$10.5\,(1 - x^2)\,P_4'(x) ,$$

where P_4' is the derivative of the fourth degree Legendre polynomial

$$P_4'(x) = 17.5x^3 - 7.5x .$$

Hence, the five points are 0, ± 7.1, ± 10.5 nm from λ_o. Again, since our passband applies decreasing weight to $f(x)$ as x is further from the origin, we choose to make these points in towards 0, to ± 6 and ± 9 nm.

4. Tests for Polarity and Spot Location

Additional measurements at the central wavelength are needed to provide tests for difference in polarization and spot location. In addition to the measurement at the central wavelength with polarity 1 (point 1 at location (0,0)), measurements were taken at

Point 2: Spot location (2,0), Polarity 1
Point 3: Spot location (0,2), Polarity 2
Point 4: Spot location (0,2), Polarity 1.

This permits a check for a difference due to polarity by comparing the third and fourth points, and check for location differences by comparing points 1 and 4 (or 1 and 2 or 2 and 4). A more extensive check based on a full 2×3 factorial combination was made on the master set #1.

B. Results: Polarity and Location

The reported results are based on measurements taken on four sets of filters. A complete series of 15 measurements across the 21-nm range at 1.5 nm spacing was taken on the master set, as a check for the adequacy of the five-point design used on the other three sets. In addition, polarity and location were tested on each filter in all sets. The results of these tests are reported in this section; transmittance calculations are reported in section C.

For set 1, six measurements were taken at $\lambda = 547.0$, at both polarity 1 and 2 at the center of the filter (0,0) and at placements of two units to the right (2,0) and above (0,2) the center. This permits a 3×2 factorial analysis for detecting differences due to location and polarity.

For sets 2, 3, and 4, t-tests on location and polarity were calculated. The test on location was deemed significant if the comparison of either (0,0) and (0,2) or (0,2) and (2,0) indicated a chance of less than 1% under the hypothesis of no effect.

None of the tests of polarity was significant at the 0.01 level. (One filter from one set gave significance of 0.025. *Out of 28 such tests, the chance of obtaining one or more spurious significances is more than 50%, so this is hardly surprising.*) For location, however, the following sets showed a significant difference (level given in parentheses):

Filter	Set (level of significance)
2	3 (0.0004)
3	2 (0.013), 3 (0.002), 4(<0.0001)
4	2 (0.0013), 3 (0.0003), 4 (0.01)
5	3 (0.01), 4 (0.0003)
6	1 (0.006), 4 (0.0013) [3 (0.02)]
7	1 (0.0002), 4 (0.0025)

Out of 28 tests of significance, the chance of 13 or more coming up significant at the 0.01 level when in fact location is irrelevant is

$$\sum_{k=13}^{28} \binom{28}{k} (.01)^k (.99)^{n-k} \cong 0 \quad ,$$

and the chance that at least 2 of the 4 sets would show significance on a given filter type is

$$\sum_{k=2}^{4} \binom{4}{k} (0.01)^k (0.99)^{n-k} = 0.0006.$$

For 7 types of filters, the overall level of significance is approximately 0.004.

These tests of significance suggest that it is highly unlikely that location on the filter in measuring transmittance is irrelevant. However, all tests were based on the internal standard errors of the four measurements made within the twenty-minute measurement period. This standard error reflects only the error of the four internal readings but does not reflect the measurement-to-measurement variation caused by apparatus set-up, filter placement in the wheel or polarity switches. As such, the actual error in taking successive measurements may in fact be larger than the reported standard error. However, measurements on these filters as near to the central location as possible is recommended. The transmittance values using triangular passbands have been calculated assuming location is irrelevant (thereby using *all* data in the estimation of the transmittance curve) and assuming location has a significant effect on the transmittance value (using only the data at the center (0,0) spot). Only the latter are given for the master Set #1 in table 3 of the main report.

C. Estimation of Transmittance Curve

As there are four sets which contain all filter types, it will be helpful to use all filters having the same nominal transmittance in estimating the transmittance curve. We fit quadratic functions of the form in eq (5).

Consider one filter type from each set having a nominal transmittance (say, 69%). The average value of transmittance, given by the parameter α, is likely to be specific to each filter in the different sets. However, the curvature parameters β and γ are likely to be common to all 4 filters having a given nominal transmittance. Thus we use all these observations to fit a relation of the form

$$f(x_{ij}) = \alpha_j + \beta \cdot x_{ij} + \gamma \cdot x_{ij}^2 , \qquad (7)$$

$i = 1, \ldots, n_j$ (= # of measurements in jth set)
$j = 1, \ldots, 4$ (= # of sets).

Note that

$$n_1 = 21,$$
$$\{x_{i1}\} = \{-12., -10.5, \ldots, 7.5, 9, -1.5, \ldots, -1.5\} ,$$

since set 1 has the full 15 measurements and 6 additional measurements for polarity and location, and

$$n_j = 8, \{x_{ij}\} = \{-9, -6, 0, 6, 9, 0, 0, 0\}$$

for sets 2, 3 and 4. Note also that the average level, α_j, is specific to the filter in the set, as the level may reflect the amount of dye that is contained in the filter. However, the parameters β and γ are likely to be influenced by the properties of the dye contained in this glass and are thus common to the filters with the same nominal transmittance from all four sets. This joint estimation permits the more accurate calculation of β and γ in sets 2, 3, and 4, where the number of observations is only eight.

Since transmittance may depend on location (section B), we also estimate α, β, γ, for each of the 7 filter types, using only those observations at location (0,0). In this case,

$$n_1 = 17, \{x_{i1}\} = \{-12, -10.5 \ldots, 7.5, 9, -1.5, -1.5\}$$
$$n_j = 5, \{x_{ij}\} = \{-9, -6, 0, 6, 9\}, j = 2, 3, 4.$$

These parameters were not found to differ from those using all the data by more than two standard errors. In light of potential location differences, however, table A1 provides these estimates of the parameters α, β and γ for the master set #1 filters.

D. Estimation of Transmittance Values Using Triangular Passband

As indicated in eq (1), when the transmittance curve is superimposed with a normalized passband centered at λ_o, the observed transmittance can be expressed as a convolution integral. Using a triangular passband of width

35

w, symmetric about λ_o, the result is a weighted average of all transmittances around the central wavelength. This weighted average may be computed either as a discrete sum:

$$\hat{\bar{\tau}}_w(\lambda_o) = \Sigma\ h_w(\lambda_i)y_i \qquad y_i = \text{transmittance at wavelength } \lambda_i;$$
$$= \Sigma\ h_w(x_i+\lambda_o)y_i, \quad x_i = \lambda_i - \lambda_o, \qquad (8)$$

where

$$h_w(x) = w^{-1}[1 - \text{sign}(x)\cdot x/w], \ -w \leqslant x \leqslant w, \qquad (9)$$

or as an integral as in eq (1). The first approach, based on eq (8), is a discrete computation based on the observed values y_i. The second approach, eq (1), uses the values y_i to estimate $f(x)$ (eq (5)) and integrates directly. For sets 2, 3, and 4 where data inside ± 6 nm exists only at the central wavelength, eq (1) is clearly the method of choice. Furthermore, the estimation of $f(x)$ uses all of the data in its estimation (section C), and is therefore likely to remove much of the variability in the values y_i.

The standard error of $\hat{\bar{\tau}}_w$ can be computed using the reported standard errors (s_i) associated with each measurement:

$$SE\ (\hat{\bar{\tau}}_w) = \Sigma[h_w(x_i)]^2\ s_i^2.$$

Using the triangular passband of width w (eq (9)) and the fitted relation (7), we have that the transmittance of a filter from the jth set is

$$\hat{\bar{\tau}}_w(\lambda_o) = \int_{\lambda_o-w}^{\lambda_o+w} h_w(\lambda) \cdot [\hat{\alpha}_j + \hat{\beta}(\lambda-\lambda_o) + \hat{\gamma}(\lambda-\lambda_o)]^2\ d\lambda$$
$$= \hat{\alpha}_j + (w^2/6) \cdot \hat{\gamma} \qquad (10)$$

for which

$$SE(\hat{\bar{\tau}}_w) = [\text{Var}(\hat{\alpha}_j) + (w^4/36)\text{Var}(\hat{\gamma}) + (w^2/3)\text{Cov}(\hat{\alpha}_j,\hat{\gamma})]^{1/2}$$
$$\cong SE\ (\hat{\alpha}_j);$$

TABLE A1. *Coefficients for Estimation of Transmittance Curve for Set #1 (includes data from center location only).*

$$f(x) = \hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2 \quad x = \text{wavelength} -548.5$$

Table gives constant term; all standard errors given in parentheses

	$\hat{\alpha}$	$\hat{\beta} \times 10^5$	$\hat{\gamma} \times 10^5$
Filter 1	.9166525	1.6493	−0.002562
	(200)	(.1877)	(.032669)
2	.6913352	−2.0067	−0.91372
	(117)	(1.5030)	(.02616)
3	.5192856	−2.7205	−1.2973
	(135)	(.1151)	(.0200)
4	.2363304	+1.1966	−1.5858
	(147)	(.0923)	(.0161)
5	.0971290	+0.8657	−1.0699
	(068)	(.0736)	(.0128)
6	.00934226	−.05959	−0.11355
	(115)	(.00979)	(.00170)
7	.000939014	−0.008245	−0.017568
	(266)	(.002893)	(.000504)

since the variance of the curvature parameter $\hat{\gamma}$ and the covariance term are typically two orders of magnitude smaller than the variance of the fitted constant term.

E. Other Passbands

Table 3 of the main report gives transmittances for the master set #1 filters for a triangular passband. For passbands other than those listed in table 3, the transmittance may be calculated directly via eq (10), using the estimated coefficients $\hat{\alpha}_1$, $\hat{\beta}$, and $\hat{\gamma}$ listed in table A1.

Statistical Analysis of Some Gas Chromatography Measurements

Karen Kafadar* and Keith R. Eberhardt*

National Bureau of Standards, Washington, DC 20234

The National Bureau of Standards has certified Standard Reference Materials (SRMs) for the concentration of polychlorinated biphenyls (PCBs) in hydrocarbon matrices (transformer and motor oils). The certification of these SRMs involved measurements of extremely small concentrations of PCBs made by gas chromatography. Despite the high accuracy of the measurement technique, the correlated data cannot be analyzed in a routine independent manner. A linear model for the measurements is described; its complexity encourages the use of simpler exploratory methods which reveal unexpected features and point the way towards obtaining valid statistical summaries of the data.

Key words: exploratory analysis; linear models; median polish; robust estimates; statistical methods; uncertainty statement.

1. Introduction

Exploratory methods in data analysis are used in many fields of application. These methods are typically used on messy data because they are robust in nature; that is, they are insensitive to unexpected departures from an assumed model (e.g. outliers, non-normality). However, they can also provide valuable insight even for so-called "clean" data. This paper discusses an example of a measurement process for which an exploratory approach can reveal particularly interesting or unexpected trends even in extremely precise measurements.

This paper is about the application of some simple exploratory techniques discussed by Tukey [1][1] and others to several sets of measurements obtained by gas chromatography. These data were taken at the National Bureau of Standards between November 1981 and March 1982 and are described in section 2. Despite the high accuracy of the method, the correlated measurements cannot be treated in a routine independent manner. A linear model for the data is proposed in section 3, and a robust analysis of this model, based on exploratory methods, is described in section 4. Conclusions are summarized in section 5.

---

*Statistical Engineering Division, National Engineering Laboratory.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. The Data

2.1 Description of the SRM for PCBs in Oils

The National Bureau of Standards has certified Standard Reference Materials (SRMs) for the concentration of polychlorinated biphenyls (PCBs) in hydrocarbon matrices, namely, transformer and motor oils. PCBs are toxic contaminants; their chemical and thermal stability makes them commercially useful but also leads to their persistence in the environment. PCBs are formulated as liquid mixtures of congeners and were manufactured in the United States under the trade name Aroclor. These mixtures have been used extensively in this country as coolants in high-voltage electrical components and may be introduced into the environment when these components are serviced, repaired, or discarded. Since the PCB fluid physically resembles lubricating oil, there have been instances in which PCBs have been added mistakenly to motor oils being collected for recycling purposes.

For these reasons, it is important to be able to measure small concentrations of PCBs. These measurements are now aided by NBS Standard Reference Material 1581 which provides certified concentrations of two congeneric PCB mixtures in two forms, known as Aroclor 1260 and Aroclor 1242. Four materials constitute this SRM, identified as Aroclor 1260 in motor oil, Aroclor 1260 in transformer oil, Aroclor 1242 in motor oil, and Aroclor

1242 in transformer oil. Large lots of these materials were prepared, carefully blended and dispensed into 5-mL amber ampoules, yielding about 4000 ampoules of each material. Six sample vials of each material were selected at random for chemical analysis.

This paper describes the statistical analysis of the measurements on the selected sample vials. The method of analysis was the same for all four types of material; hence, we will limit our discussion to its application on Aroclor 1260 in motor oil, for the most part.

2.2 Measurement Techniques

A method for chemical analysis of PCBs in oil was developed at the National Bureau of Standards and employs both liquid and gas chromatography for analysis. Quantitation is performed by a calibrated internal standard (IS) technique which is described in detail by Chesler et al. [2]. Basically, this technique works in the following manner. First, a sample containing a known amount of analyte (in this case, PCB) and known amounts of one or more internal standards is analyzed. The instrumental responses, measured as individual peak areas, are recorded, and the relative responses between the analyte and the internal standards (ISs) are ascertained. Second, known amounts of the internal standards are added to a sample for which the concentration of the analyte is desired, and the sample is subsequently analyzed. Then, using the relative response factors previously ascertained, the concentration of the analyte is determined from the ratio of the responses between the analyte and internal standard. In the particular experiments discussed below, each measurement of the analyte gave 10 separate measurable responses (i.e. 10 distinct peaks) and each of the three internal standards gave one measurable response. Figure 1 shows a sample chromatogram from one of the analyses. Thus, since analyte concentration can be calculated from any combination of one internal standard plus one analyte peak, the analyte concentration can be calculated in many ways for each single analysis.

2.3 A Formula for Determining Aroclor Concentration

The response factors, i.e., the ratios (peak area)/(concentration), are not necessarily constant from run to run. In fact, the response factor for a given peak is subject to many kinds of instrumental variation, such as the electronic detector and heat source differences inside the instrument and variations in the flow rate of material through the gas chromatographic column. But, to a large extent, these variations affect all peaks (including those that arise from the internal standard) in the same way.

Therefore, on a given "occasion" when several runs are all reasonably close together in time, the ratio

$$\frac{(\text{peak area for Aroclor})/(\text{concentration of Aroclor})}{(\text{peak area for IS})/(\text{concentration of IS})} \quad (1)$$

should be constant for each run and each solution, whether it is a sample solution or a specially-prepared calibration solution. In principle, then, (1) should be the same for the sample and the calibration material; i.e.,

$$\frac{P(I,s)/C(I,s)}{P(a,s)/C(a,s)} = \frac{P(I,c)/C(I,c)}{P(a,c)/C(a,c)}$$

where
P denotes peak area
C denotes concentration
I indexes internal standard peak
a indexes Aroclor peak
s denotes sample solution
c denotes calibration solution.

Therefore, a calibration solution is passed through the chromatograph at approximately the same time a sample solution is run, and the concentration of the Aroclor in a sample, based on a given peak area, may be determined as

$$C(a,s) = \frac{P(I,c)}{P(a,c)} \times \frac{P(a,s)}{P(I,s)} \times \frac{C(I,s)}{C(I,c)} \times C(a,c). \quad (2)$$

Note that different peaks will yield different values for concentration.

The concentration of a substance in a solution is very nearly the mass of the analyte divided by the mass of the oil in that solution. In the actual experiment, the weight of the internal standard was measured and held constant for both the calibration solution and the sample solution. So in this experiment,

$$\frac{C(I,s)}{C(I,c)} = \frac{W(I,s)/W(oil,s)}{W(I,c)/W(oil,c)} = \frac{W(oil,c)}{W(oil,s)}$$

where W denotes weight. Therefore, the formula

$$C(a,s) = \frac{P(I,c)}{P(I,s)} \times \frac{P(a,s)}{P(a,c)} \times \frac{W(oil,c)}{W(oil,s)} \times C(a,c), \quad (2')$$

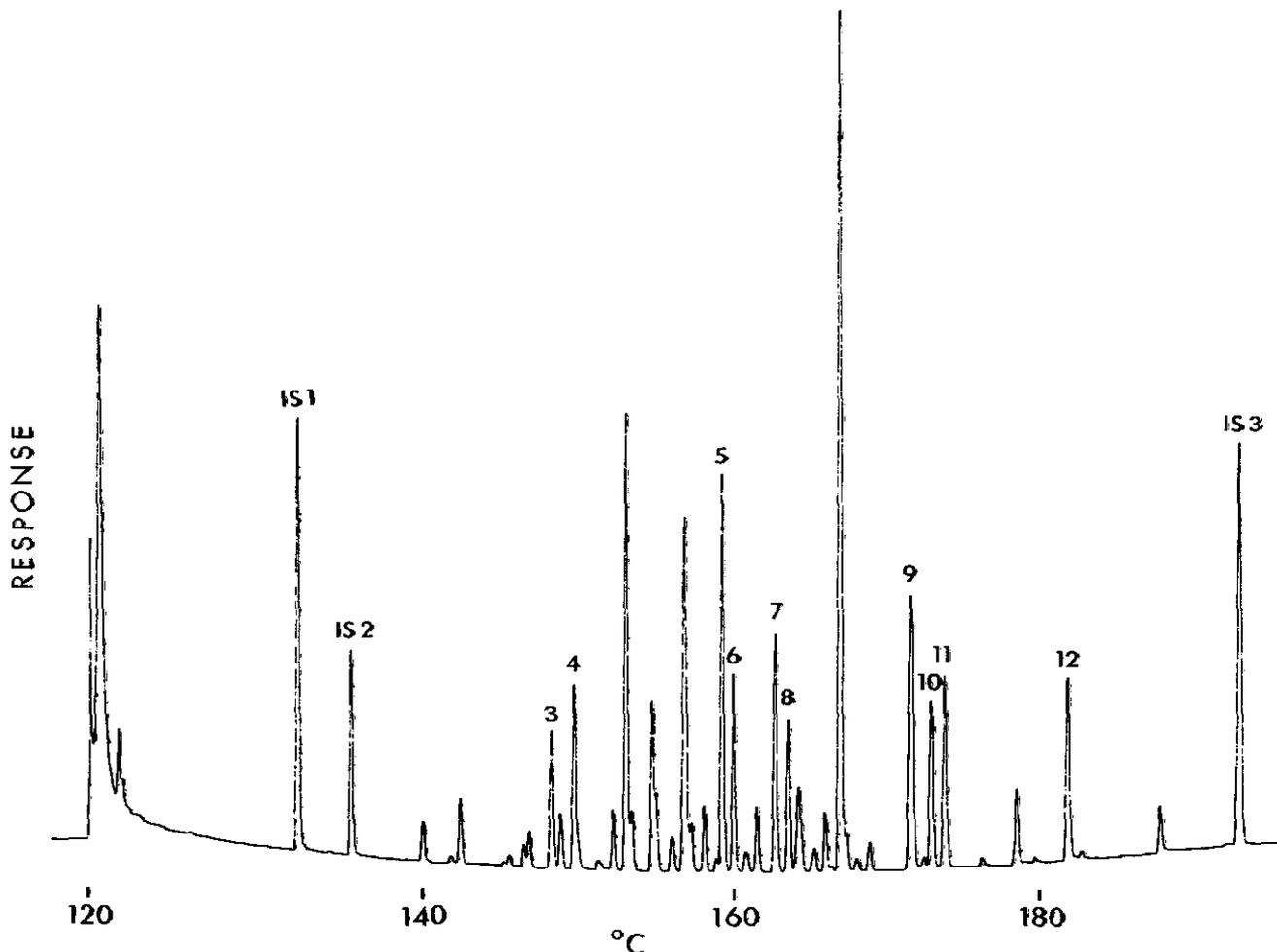is actually sufficient for determining Aroclor concentration.

FIGURE 1. A gas chromatogram from one subsample of Aroclor 1260 in motor oil. IS 1, IS 2, IS 3 denote peaks from three internal standards; 3 through 12 denote Aroclor peaks. The horizontal scale indicates the temperature levels in the gas chromatographic column at which the compounds are eluted.

2.4 Experiment Design

For each material, three subsamples were analyzed from each of the six selected ampoules. On a given occasion (2-3 hour period), two subsamples, one from each of two ampoules, were run along with one calibration solution. Display 1 presents the experimental design. Given the method of preparation, a high degree of homogeneity among the ampoules was expected. With these limited data, tests on the variance among the six ampoules showed no evidence of inhomogeneity compared to the overall variability of the measurements.

As a result of the design in display 1, the calculated values of concentration for two subsamples analyzed on the same occasion are not independent. In eq (2), these two calculated concentrations would share the same value of the ratio $P(I,c)/P(a,c)$. Thus, for a given (internal standard peak, Aroclor peak) pair, the 18 concentration

values obtained from six ampoules each with three replications are not independent. Furthermore, since there are many pairs of internal standards and Aroclor peaks for each sample, this design leads to a large number of calculated concentration values having strong and complex interdependencies. The challenge is to take proper account of these dependencies in obtaining an overall certified value and uncertainty statement for the SRM certificate.

3. A Linear Model for the Measurements

The basic properties of the gas chromatography procedure described by eq (1), plus the rules and provisos mentioned in the surrounding text, can be represented by a linear statistical model. Denoting $\log_{10}(\text{peak area})$ by Y, these considerations lead to the model in display 2. This model is an unbalanced partially nested and partially crossed analysis of variance model with a total of 211

39

Allocation of Subsamples and Calibrations to
Occasions on Which GC Analysis Runs Were Made*

Occasion	Sample ID	Calibration Solution
1	S1, S2	A
2	S1, S2	A
3	S1, S2	A
4	S3, S4	B
5	S3, S4	B
6	S3, S4	B
7	S5, S6	C
8	S5, S6	C
9	S5, S6	C

*Note: The order in which the two samples and one calibration
were run within each occasion was varied.

Display 2—Linear model for $Y = \log_{10}(\text{peak area})$

	Notation	Meaning/levels	Number of Parameters
$Y_{ijklm} =$	Q_i	occasion $i = 1,2,...,9$	9
$+$	$R_{(i)j}$	run within occasion $j = 1,2,3$	27
$+$	M_k	material: $k = 1,2,...,9$ (samples S1,...,S6 and calibration solutions A,B,C)	9
$+$	$C_{(k)l}$	concentration within material $l = 1$ (IS1), 2 (IS2), 3 (Aroclor), 4 (IS3)	36
$+$	P_m	peak: $m = 1 \quad 2 \quad 3...12 \quad 13$ IS1 \quad IS2 \quad Aroclor \quad IS3	13
$+$	$(QP)_{im}$	interaction	117
$+$	ε_{ijklm}	measurement error	-
		Total	211

parameters and 351 observations. The 351 measured peak
areas come from 9 occasions with 3 runs each (one
calibration and two SRM subsamples—see display 1), for
a total of 27 runs, times 13 peaks per run.

While we feel that this linear model provides a
reasonably complete and realistic representation of the
measurements, there were several reasons which prevented
its use as a sufficient guide for the data analysis.

First, some implicit constraints in the model are difficult
to incorporate in an analysis with available statistical
computer software. These constraints arise from the fact
that Aroclor concentrations in the calibration solutions are

actually known before any peak areas are measured.
Similarly, the ratios of concentrations for the internal
standards are controlled, as described in section 2.3 above.
Taken together, these considerations imply that there are
actually only six free parameters (corresponding to the
concentrations of Aroclor in the six selected ampoules)
among the 36 model terms denoted $C_{(k)l}$. In contrast, the
statistical computer packages which can handle a problem
of this size and of incomplete rank do not have any
provision for imposing the necessary constraints, so that
the $C_{(k)l}$ terms in the model are treated (inappropriately)
as 36 free parameters. A future manuscript will give a
detailed look at this and other aspects of the least squares
treatment of these data under the linear model.

Another reason for not using the model directly is the
probable lack of homogeneity in the variances of the
logarithms of the measured peak areas, which is implicitly
assumed in computer packages. Evidence of variance
inhomogeneity arose from the concerns of the chemists
and was exhibited in the data (see section 4.3).

Although a least squares analysis of the model was not
used to obtain the certified value and uncertainty for the
SRM certificate, it is instructive to compare the data
summaries suggested by least squares with the more
robust alternative actually used. For example, for each
replication on each sample, it is possible to calculate the
Aroclor concentration in 30 different ways by using one
of the three internal standards with any one of the 10
available Aroclor peaks. It can be shown that a least
squares fit of the model implies that one should summarize
those 30 possible concentration values by taking their
arithmetic mean value (in log scale). In the following
sections, some useful alternatives to simply averaging
log(concentration) values will be described.

4. An Exploratory Analysis

4.1 Concentrations Revisited

As illustrated in display 1, samples S1 and S2 were run
on the same occasion as a calibration solution denoted by
A; likewise, samples S3 and S4 were run with calibration
B, and samples S5 and S6 were run with calibration C.
Three replications were performed on each occasion.
Transforming eq (2′) via logarithms, a calculation of
log(concentration) may be described as

$$Z = \underbrace{\log(\text{weight of oil} \times \text{one IS peak/one Aroclor peak})}_{\text{calibration}}$$

$$- \underbrace{\log(\text{weight of oil} \times \text{same IS peak/same Aroclor peak})}_{\text{sample}} \quad (3)$$

$$+ \log(\text{concentration of Aroclor in calibration solution}).$$

Notice that there are (3 calibrations) × (3 replications) × (3 IS peaks) × (10 Aroclor peaks) = 270 values for the first term in eq (3), and (6 samples) × 3 × 3 × 10 = 540 values for the second term. The concentration in each calibration solution is determined gravimetrically once for all three replications, so there are only three values for the third term in (3) (typically, about $\log_{10}$ (100.01 $\mu g/g) \approx 2$).

4.2 Summarizing the Data

The quantity $U = \log_{10}[(\text{weight of oil} \times (\text{IS peak}) / (\text{Aroclor peak})]$, for either the calibration or the sample, is of interest, for its value can be determined entirely from the measurements made during a single run. Corresponding values of U must be combined for the appropriate sample and calibration within a single occasion. If we consider the measurements based on a particular internal standard and peak area, we can reduce the data for two samples by averaging the U-values from the independent replications of two samples, and combining this average with the value from the calibration solution on the same occasion. Thus, for each peak and internal standard, we can obtain nine summary values, one for each replication on each of three calibration solutions. Let us denote these "summary values" by $Z_{\alpha j l n}$, where $\alpha = $ A,B,C identifies the calibration solution, $j = $ 1,2,3 indexes replications, $l = $ 1,2,3 indexes the internal standard, and $n = $ 3,4,...,12 indexes the Aroclor peak. Thus, $Z_{\alpha j l n}$ is the average log(Aroclor concentration) for the two samples which were run on the jth replication with calibration α, using internal standard l and Aroclor peak n. The values $Z_{\alpha j l n}$ are displayed schematically in figure 2.

Figure 2 suggests that further summarization is possible across those values that are estimating the same quantity. We can do this in two stages, corresponding to the two two-way layouts that are shown. If we analyze the data via the least-squares model of section 3, the order of summarization will not matter, and we can simply average all $9 \times 30 = 270$ values together. However, these values are not independent, and obtaining a standard error for such a grand average is not simple.

Clearly, the 30 values in the secondary layout (b) of figure 2 are highly correlated. However, a "typical value" from the 30 values $Z_{\alpha j l n}$, $l = $ 1,2,3, $n = $ 3,4,...,12, will be strictly independent of any $Z_{\alpha' j' l n}$ ($\alpha \neq \alpha'$, $j \neq j'$). It will simplify our uncertainty assessment from this four-way table if we reduce each secondary table containing the correlated values to a single "typical value" representing these numbers. Having found such a "typical value," it may be placed in the appropriate cell of (a) of figure 2, where we are assured that the resulting nine values will be independent. A one-way analysis of variance
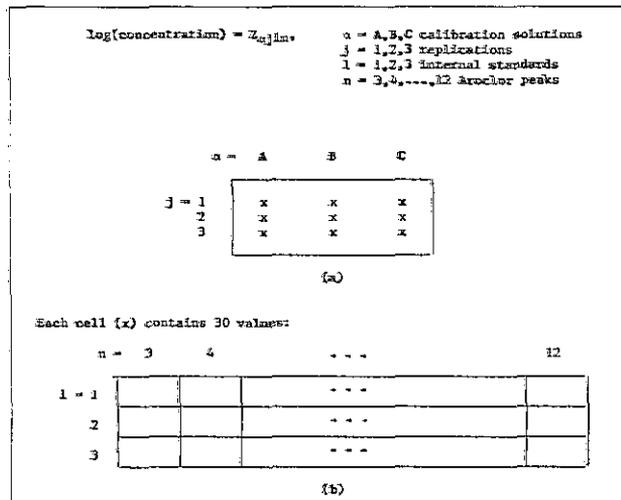

FIGURE 2. Schematic layout for summary values of $\log_{10}$(concentration).

on these nine values may then be performed, and an appropriate standard error obtained herein.

Thus, the analysis proceeds in two stages in which we:

1) Obtain a "typical value" from the entries $Z_{\alpha j l n}$ for each $\alpha = $ A,B,C and $j = $ 1,2,3 (replications), and,

2) Analyze the nine "typical values" $Z_{\alpha j..}$ by a standard one-way analysis of variance.

Since chemical reasoning leads us to expect the data in some cells of the secondary tables to be much less reliable than those in other cells, we chose to use two-way median polish to obtain a summary value for the table. This procedure is analogous to a two-way analysis of variance, but is insensitive to large deviations from the typical value which might occur in the data. (An example illustrating the procedure is given in the Appendix.) Furthermore, the correlation among the 30 values in the table will not worry us at this stage since we are interested only in reducing the table to a single typical value.

4.3 Summarizing the Dependent Pieces: Median Polish

It is interesting to compare the summary value obtained from median polish with the corresponding summary (the grand mean) obtained from a two-way analysis of variance of the same data.

The differences between the median polish typical value and the grand mean of the 30 values, for each of the nine two-way tables, are illustrated in figure 3. Notice that the median polish typical value differs from the grand mean whenever one group of values (corresponding to a given internal standard) does not agree with the other two. The median polish typical value gravitates towards the heaviest concentration of values, downplaying the influence of a

group which is some distance away (e.g., internal standard 3 in the third replication using calibration A). In contrast, the grand mean gives all values equal weight, regardless of their relative positions among the others. The grand mean may not be appropriate, given the possibility of chemical reasons for all values from a given internal standard to be high or low. Notice also from figure 3 that the data exhibit evidence of variance inhomogeneity, as mentioned in section 3. For these reasons, we adopt the median polish typical values as the summary values for this problem.

4.4 Summarizing the Independent Pieces: One-way Analysis of Variance

The nine median polish typical values for Aroclor 1260 in motor oil are shown in the top portion of display 3.

Since different replications involved preparation of distinct subsamples of material, these nine entries are independent. Furthermore, they are already summaries of 30 values from peak × IS combinations. By virtue of central limit theorem considerations, we feel safe in analyzing this table using a standard one-way analysis of variance model with normal errors.

It is possible that differences among the groups identified by calibration solutions could be introduced into the data at the stage when the chemists add the internal standard "spikes" to the sample, or when the calibration solutions are prepared. Since the results of these operations are treated as values known without error, errors introduced at this stage would persist as systematic errors between the calibration solution groups. However, the F-tests for these group effects were all clearly nonsignificant, with computed α-levels ranging from 35
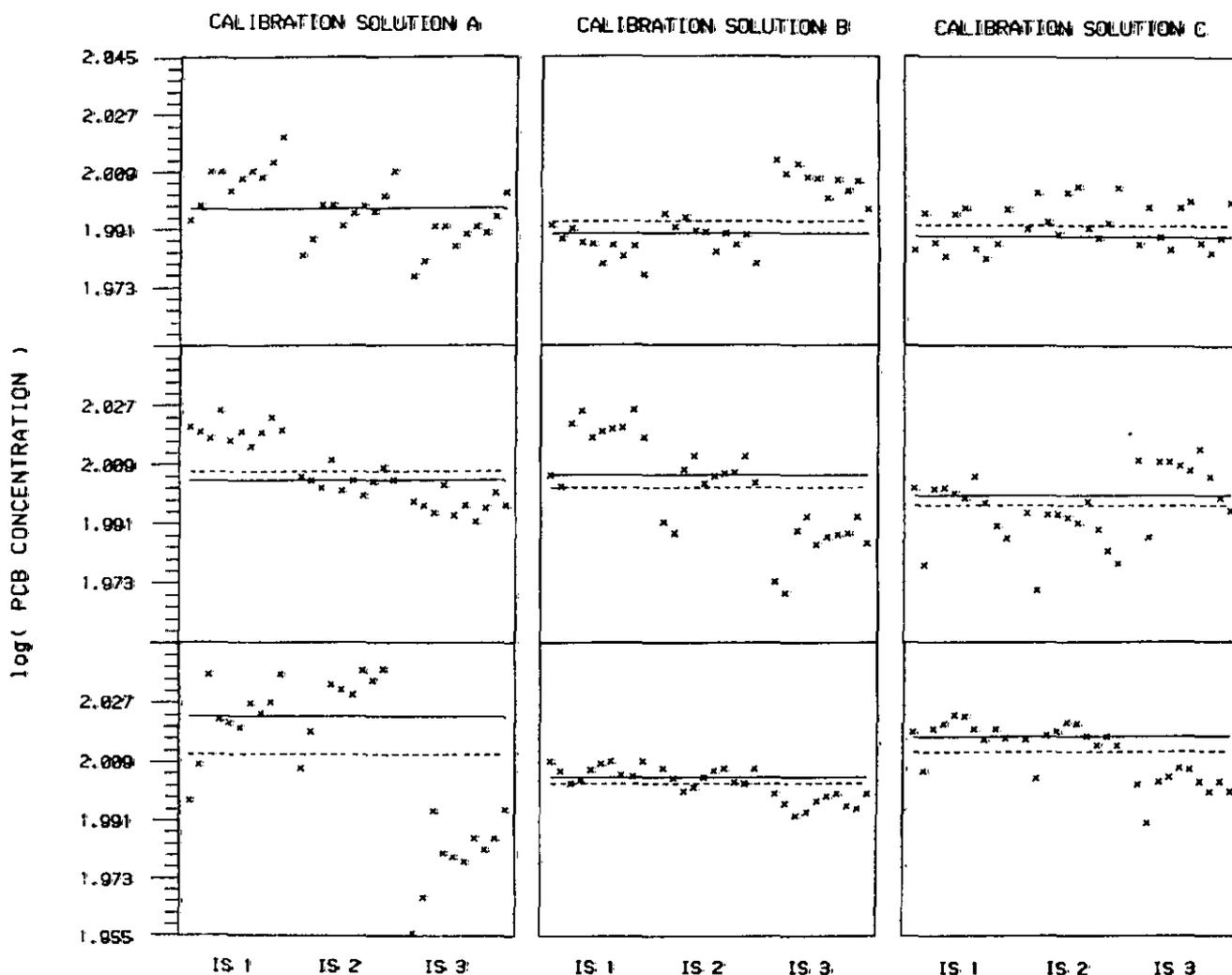

FIGURE 3. Median polish typical values (solid) and grand means (dot) of 30 values of log(concentration). Each set of 10 points refers to a different internal standard. The leftmost graphs are from the replications using calibration A; the middle are from those using calibration B; and the rightmost are from those using calibration C.

Display 3—One-way analysis of variance for nine "typical values" of log(concentration) for Aroclor 1260 (motor)

		Calibration Solution:		
		A	B	C
Replication				
1		1.99723	1.98942	1.98833
2		2.00399	2.00542	1.99936
3		2.02271	2.00415	2.01660

Source	d.f.	Sums of Squares	Mean Squares	F-ratio
Between Calibrations	2	.0001151	.0000576	.378
Within Calibrations	6	.0009127	.0001521	
Total	8	.0010278		

Student's t confidence interval for grand mean (8 d.f.): (1.99431, 2.01174)

percent to 87 percent for the four materials studied. The one-way analysis of variance for Aroclor 1260 (motor) is shown in display 3.

In the absence of between-group errors, it is reasonable to treat the nine values in our one-way table as independent and identically distributed observations on the Aroclor concentration. Taking this approach, the standard error of the mean of the nine values is estimated in the usual way as $s/\sqrt{9}$. The results of these calculations for all four materials are summarized in table 1.

TABLE 1. *Summary Values for Concentrations of PCBs in Oils*

	Aroclor 1260 (3 internal standards)		Aroclor 1242 (2 internal standards)	
	Motor Oil	Transformer Oil	Motor Oil	Transformer Oil
log(Concentration), mean of typical values from median polish	2.0030	2.00027	2.003	2.0017
Standard error for log(concentration)	.0038	.0050	.0027	.0023
95% Confidence limits (log scale)	(1.9943, 2.0117)	(1.9912, 2.0142)	(1.9941, 2.0064)	(1.9964, 2.0070)
Concentration	100.69	100.62	100.07	100.39
95% Confidence limits (original scale): (µg/g)	(98.70, 102.74)	(97.99, 103.32)	(98.65, 101.48)	(99.18, 101.62)

4.5 Peak Effects

As in a classical analysis of variance, median polish yields estimates of the effects from the data. It is instructive to examine these estimates, particularly in cases presenting anomalies among them. Nine median polishes were performed to obtain the entries for the layout in (a) of figure 2. The resulting peak effects are plotted in figure 4. Notice that in all cases the median peak effect is zero (by the nature of the median polish analyses) and that in several cases peaks 3 and 4 (first and second from the left) are substantially lower than the others. The average across all nine independent replications is plotted in figure 5, shown with limits of one standard error. Graphs such as this one suggest looking for chemical explanations for the low values on peaks 3 and 4.

5. Conclusions

Four data sets consisting of gas chromatography measurements were analyzed for purposes of SRM certification and for eliciting unusual features in the data. A linear model was proposed, but practical difficulties limit its usefulness for the problem at hand. Exploratory techniques led not only to a certified value and uncertainty but also to the investigation of chemical causes for unusual patterns in the effects.

The proposed analysis for these data is not specific to an experiment of this form—i.e., GC measurements using multiple internal standards and multiple Aroclor peaks, with one calibration solution for every two sample solutions. Rather, it illustrates how the data from independent runs yielding several, but dependent, answers, can be analyzed without resorting to complicated linear models that can sometimes be unwieldly. Basically, our general approach proceeds in two stages:

1) The structure for the dependent pieces is determined, and the pieces are reduced to an appropriate summary value (e.g., median, median polish typical value, or another location estimator), and

2) The structure for the independent pieces is identified, and the data analyzed accordingly.

Since the data at the second stage are independent, standard techniques for obtaining uncertainty statements apply.

As a summarizing technique for the first stage, we used median polish to provide a robust summary for each data set, leading to the values reported in table 1. Without exploratory analysis, classical techniques may well have obscured some of the interesting features in the data.
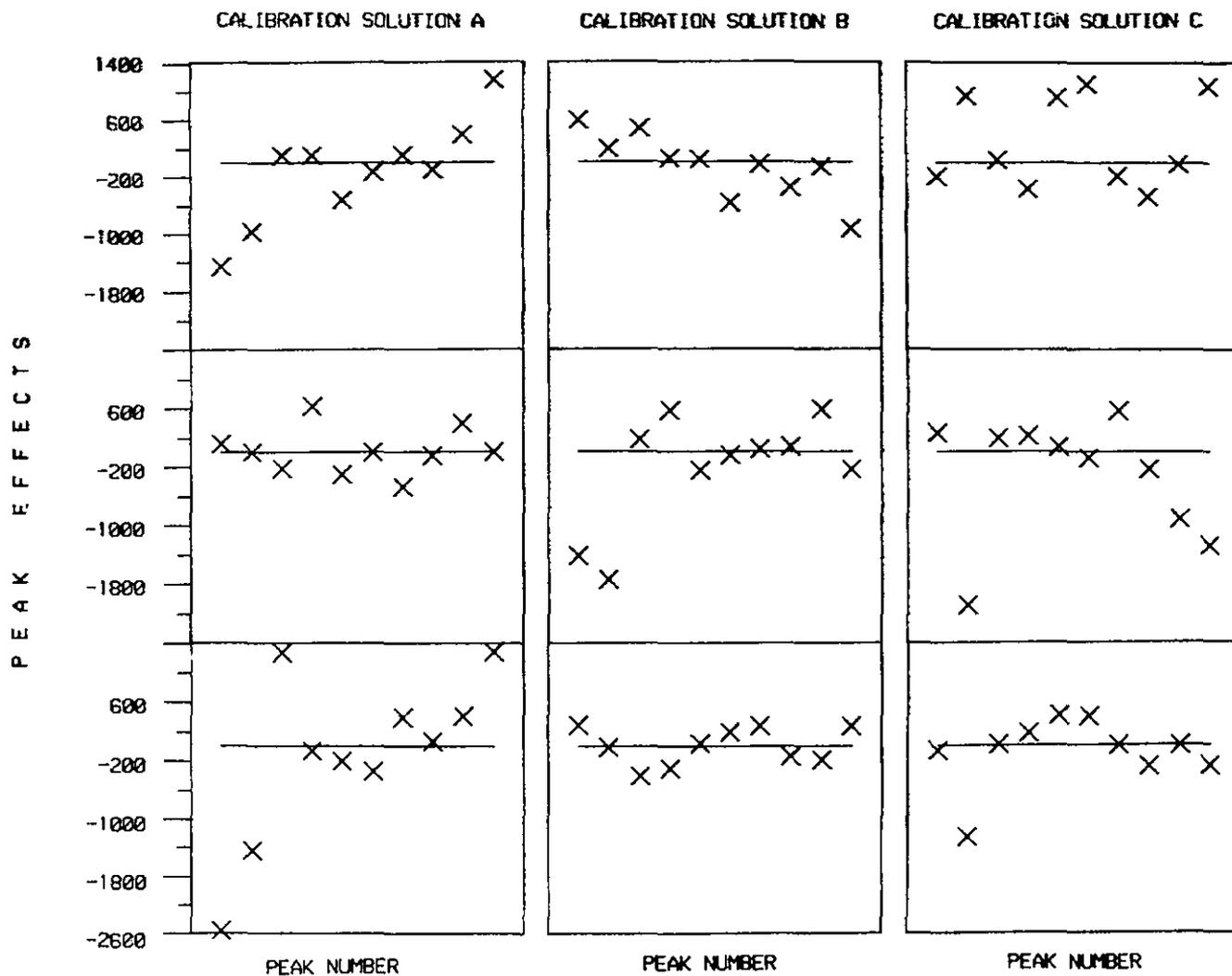
FIGURE 4. *Peak effects from the median polishes corresponding to the data in figure 3.*

We are grateful to Dr. Stephen N. Chesler for his advice on the chemical aspects of this problem and for providing us with the data, to Dr. Walter S. Liggett for his comments on this manuscript and to Dr. Churchill Eisenhart for suggesting the addition of the Appendix.

6. References

[1] Tukey, John W. (1977) *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.
[2] Chesler, S.N., Guenther, F.R., May, W.E., Parris, R.M. (1982) Standard reference materials for accurate analyses of PCBs in oil. ASTM Standardization News, 10, (6); 15-20.

7. Appendix: Median Polish

An analysis of a two-way table via median polish may be unfamiliar to many readers, so we present an example using a subset of the PCB data. A typical additive model for a two-way table is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

where $i=1,2,...,$(number of rows) and $j=1,2,...,$(number of columns). An analysis of variance estimates μ by $\overline{Y}_{i..}$, α_i by $(\overline{Y}_{i.}-\overline{Y}..)$ and β_j by $(\overline{Y}._{j}-\overline{Y}..)$, where dots indicate averaging over the missing subscripts. Median polish offers more robust estimates of these parameters as illustrated below.
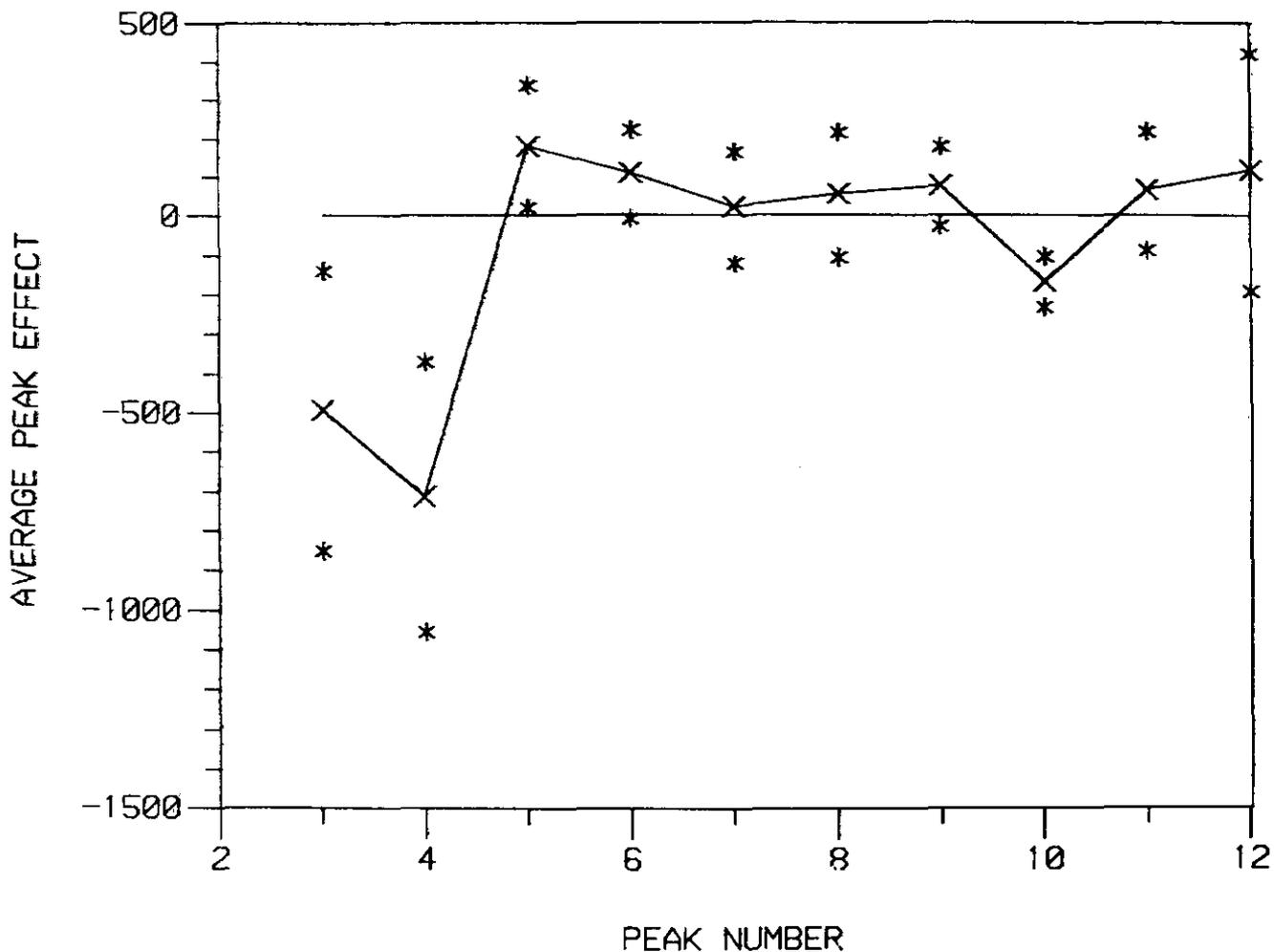
44

FIGURE 5. Average peak effect (averaged across all nine replications shown in figure 4). Limits of one standard error of the average shown by *.

Recall from section 4.3 that for each replication ($j=1,2,3$) of each calibration run ($\alpha=$A,B,C), we have thirty values, $Z_{\alpha j l n}$, corresponding to the three internal standards ($l=1,2,3$), and 10 Aroclor peaks ($n=3,\ldots,12$). These are the thirty values Y_{ij} in the two-way table.

For the data set of Aroclor 1260 in Motor Oil, $j=1$ (first replication) and $\alpha=$A (calibration solution), of log(Aroclor concentration), part of the two-way table is shown below:

| | Aroclor Peak | | | | |
	3	4	5	6	7
IS 1	1.9937	1.9985	2.0093	2.0093	2.0030
IS 2	1.9827	1.9873	1.9983	1.9983	1.9920
IS 3	1.9761	1.9809	1.9917	1.9917	1.9892

To simplify the arithmetic, let us subtract 2 from each of the numbers and multiply by 10^4:

| | Aroclor Peak | | | | |
	3	4	5	6	7
IS 1	−63	−15	93	93	30
IS 2	−173	−127	−17	−17	−80
IS 3	−239	−191	−83	−83	−108

As a first step in estimating the column effects, we write down the median of each column below a single line:

| | Aroclor Peak | | | | |
	3	4	5	6	7
IS 1	−63	−15	93	93	30
IS 2	−173	−127	−17	−17	−80
IS 3	−239	−191	−83	−83	−108
	−173	−127	−17	−17	−80

45

We subtract each median from the value in the cell, and put a double line between the table and the medians to indicate that a number has been subtracted:

	Aroclor Peak				
	3	4	5	6	7
IS 1	110	112	110	110	110
IS 2	0	0	0	0	0
IS 3	-66	-64	-66	-66	-28
Column effect	-173	-127	-17	-17	-80

Next we operate on the rows in the same fashion. First, write down the median of each row (including the row corresponding to the column effects):

	Aroclor Peak					
	3	4	5	6	7	
IS 1	110	112	110	110	110	110
IS 2	0	0	0	0	0	0
IS 3	-66	-64	-66	-66	-28	-66
Column effect	-173	-127	-17	-17	-80	-80

and then subtract the row medians from the values in the table (again indicated by the double line):

	Aroclor Peak					Row effect
	3	4	5	6	7	
IS 1	0	2	0	0	0	110
IS 2	0	0	0	0	0	0
IS 3	0	2	0	0	38	-66
Column effect	-93	-47	63	63	0	-80

Now we return to columns again: write down the median of each column below the table (single line):

	Aroclor Peak					Row effect
	3	4	5	6	7	
IS 1	0	2	0	0	0	110
IS 2	0	0	0	0	0	0
IS 3	0	2	0	0	38	-66
Column effect	-93	-47	63	63	0	-80
	0	2	0	0	0	0

Subtract these medians from the values in the corresponding column, and add the two sets of column effects together:

	Aroclor Peak					Row effect
	3	4	5	6	7	
IS 1	0	0	0	0	0	110
IS 2	0	-2	0	0	0	0
IS 3	0	0	0	0	38	-66
Column effect	-93	-45	63	63	0	-80

Now we return to the rows, the medians of which are easily seen to be all 0's. So there is no further "polishing" of this table. Returning to the original scale, the row effects $\hat{\alpha}_i$ (corresponding to internal standards) are (0.0110, 0.0000, -0.0066), the column effects $\hat{\beta}_j$ (corresponding to Aroclor peaks) are (-.0093, -.0045, .0063, .0063, .0000), and the overall typical value $(\hat{\mu})$ is 1.9920.

Notice that:
- Median polish essentially "converges," in that, beyond a certain point, the table changes little;
- At each stage, the values in the table ("residuals") are further polished, so that, at any stage of the process, row effect + column effect + overall typical value + tabled entry = original data value;
- Whereas the usual analysis of variance summarizes rows and columns by means, median polish summarizes by medians;
- A plot of the peak effects versus peak number would reveal that the effect from peak 3 is considerably lower than the others; and
- An extreme value in the table is flagged by having an unusually large or small residual in the cell of the final table (e.g., IS 3 and Aroclor Peak 7). Most of the entries in the final tableau, however, are nearly zero.

Three- and four-way median polish are also possible; see [1] (pp. 452ff) for an example.