SESSION II: ESTIMATION OF TRENDS

William Hill, a statistical scientist, began his presentation with an illustration of ozone depletion curves predicted by the findings of the NAS (Fig. 27). In curve A, CFMs are assumed to be released at 1973 rates until some point in time where it is assumed that the releases are suddenly halted. The theory underlying curve A suggests that even after the release of CFMs is ended, a reduction in ozone will continue for approximately 10 additional years before the ozone gradually begins to return to its previous level. Curve B illustrates the predicted depletion where it is assumed that CFMs are released at 1973 rates without interruption. By varying the rate constraints underlying the chemical reactions involved in the ozone destruction mechanism, a family of curves similar to A and B is produced.

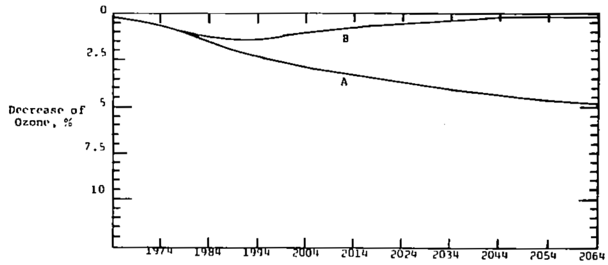

Figure 27. Ozone Depletion Curves

The application of statistical methods to recorded ozone measurements has an important role in the evaluation of the effect of human-related activities on the environment. Since the effects of a long-term depletion of ozone at magnitudes predicted by the NAS would probably be harmful to most forms of life, it is important to determine whether the leading edge of the hypothesized decline has occurred. Seeking to let the data speak for themselves, Hill created empirical pre-whitening filters the derivation of which was independent of the underlying physical mechanisms. When the data themselves are in question, statistical analysis can perform a "checks and balances" effort. Hill noted that time series modeling has some distinct advantages. It filters variations into systematic and random parts, errors are uncorrelated, and significant phase lag dependencies are identified. Hill discussed using time series modeling to enhance the capability of detecting trends.

Hill presented an analysis of ozone data using time series intervention analysis to determine whether the predicted decline has occurred in ozone. He first examined existing ozone data to determine whether a significant global abnormal trend--any positive or negative trend, man-made or natural, which cannot be explained by past ozone data records--has occurred as predicted in the ozone level in the 1970s. The second objective of Hill's analysis was to quantify the potential detectability that could be provided by future monitoring of ozone concentrations through a global network of recording stations. Detectability refers to the smallest abnormal trend that would have to occur in the ozone measurements to be judged significantly different from zero trend. Early warning of a trend followed by correction of the cause would lead to the return to normal ozone levels (Fig. 28).

Hill presented plots of monthly total ozone values recorded at three sites: Tateno, Japan (36N, 140E), Mauna Loa, Hawaii (20N, 156W), and Aspendale, Australia (38S, 145E) (Fig. 29). Many characteristics of total ozone data are illustrated in these plots. The mean ozone levels increase as the distance from the equator increases. The amplitude of the seasonal variation exhibits a similar latitudinal dependency. Figure 29 also illustrates the phase difference in the ozone peaks between North Temperate and South Temperate Zone stations. One predominant characteristic of ozone data which is not obvious from this illustration is the strong seasonal and latitudinal dependency of the month-to-month variance of ozone concentrations.

Since ozone recording stations are not uniformly distributed around the globe, the close proximity of many of the stations casts doubt on the independence of the data records. Thus Hill selected a representative global sample of stations for analysis, a sample in which no particular region has a larger influence than any other region, by dividing the globe into nine equal areas (dark lines in Fig. 30) such that each area contains at least one active recording site with at least 10 years of continuous data. One station with no more than two missing values was chosen for analysis in each area. All data were recorded using the same type of instrument, and missing values were estimated by a graphical linear interpolation procedure.

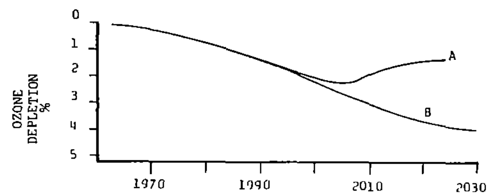

Figure 28. Hypothesized Ozone Depletion Profiles. Profile A: CFMs released at constant rate until some point in time at which all emissions are assumed to be curtailed. Profile B; CFMs released at constant rate without interruption.

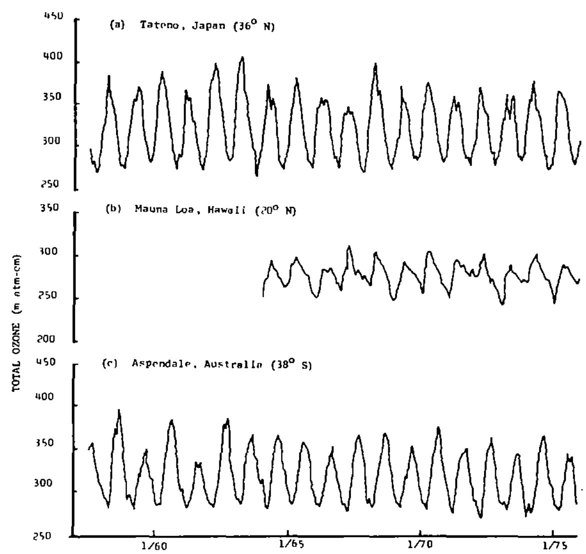

Figure 29. Mean Monthly Total Ozone Measurements Representative of the North Temperate (a), Tropical (b), and South Temperate (c) Data

The stations chosen for analysis using the above criteria are listed in Table 1 and are indicated by the large dots in Figure 30. Since ozone measurements were not made at Kodaikanal in May and June 1975, Hill truncated the series at April 1975. Other missing values occur prior to the period of hypothesized trends, and estimates of these missing values would be expected to have a small effect, if any, on the results.

Hill noted that while the global sample of stations was not truly a random sample of ozone recording sites, the restrictions did not compromise the results of the analysis.

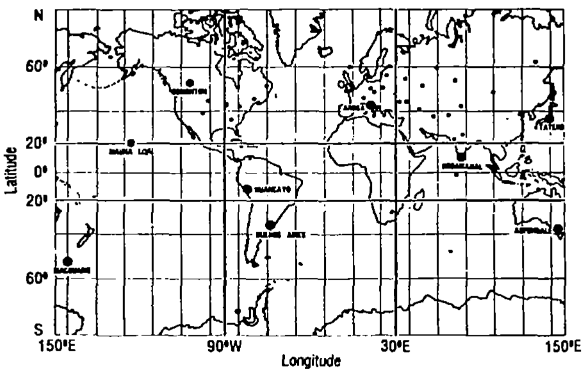

Figure 30. Stations Selected for Global Analysis of Total Ozone Data

Table 1. Stations selected for global analysis of total ozone data.

ZONE	STATION	LOCATION	PERIOD	MEAN LEVEL	# OF MISSING VALUES
North Temp.	Edmonton	54N, 114W	7/57-12/75	357	0
	Arosa	47N, 10E	1/57-12/75	333	2
	Tateno	36N, 140E	7/57-12/75	323	0
Tropics	Mauna Loa	20N, 156W	1/64-12/75	277	0
	Huancayo	12S, 75W	2/64-12/75	264	1
	Kodaikanal	10N, 77E	1/61-4/75	261	0
South Temp.	MacQuarie Isl.	54S, 159E	3/63-12/75	340	0
	Buenos Aires	35S, 58W	10/65-12/75	288	0
	Aspendale	38S, 145E	7/57-12/75	320	0

The ozone change, or trend, analysis is an application of the intervention analysis technique described by G.E.P. Box and G. C. Tiao in the Journal of the American Statistical Association in 1975. Intervention refers to the occurrence of a phenomenon (man-related or natural) which could possibly affect the level of a time series of data.

Hill's intervention analysis of ozone data attempts to determine whether a change exists in each of nine univariate series that would support the theory of a hypothesized depletion in ozone due to CFMs and other ozone depletion sources. Although the analysis can be completed in one step, Hill broke it into two steps so that the changing month-to-month variance of the ozone data can be more easily incorporated into the analysis.

In this analysis, time series models are first identified. One of the main reasons small trends can be detected is that there is a variance reduction capability in time series modeling. Tukey noted that Hill's "major output is standard errors because that will be most useful in trend detection." This is graphically illustrated (Fig. 31) using the monthly ozone data from Tateno, Japan.

RESULTING TATENO TIME SERIES MODEL

$$Y_t = \frac{(1 - \theta_{12} B^{12}) A_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})} = \text{[FILTER]} \quad \times \quad \text{[ERROR]}$$

where

Y_t = total ozone observed in month t

A_t = random uncorrelated noise (error) in month t

B = backshift operator such that $B^{12}Y_t = Y_{t-12}$

ϕ_1, ϕ_2 = autoregressive parameters representing dependencies between ozone values 1 and 2 months apart, respectively

θ_{12} = seasonal moving average parameter

rewritten

$$Y_t = Y_{t-12} + \phi_1 (Y_{t-1} - Y_{t-13}) + \phi_2 (Y_{t-2} - Y_{t-14})$$

$$-\theta_{12} A_{t-12} + A_t$$

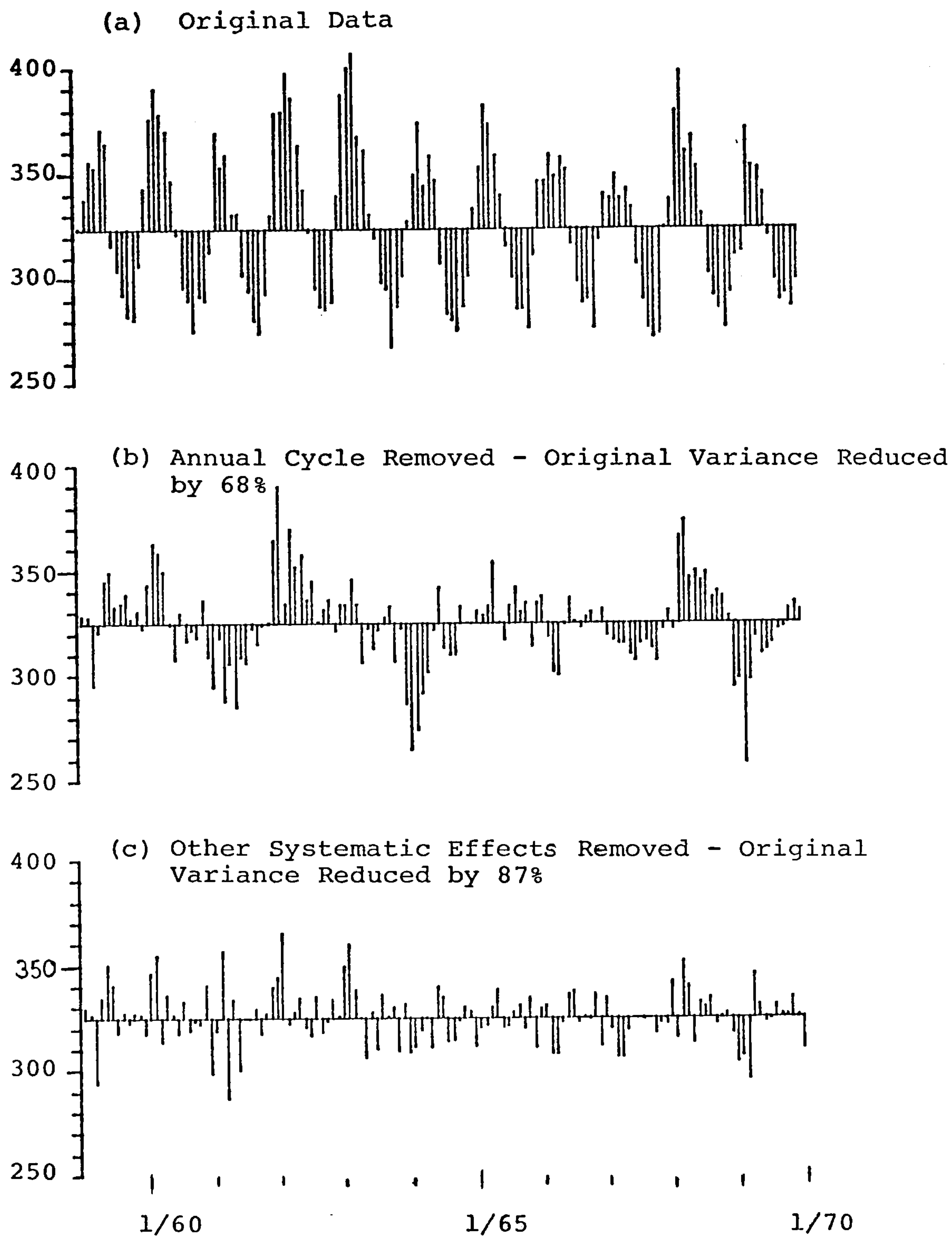

Figure 31. Removing the Systematic Variation at Tateno by Time Series Analysis

Removing the seasonal or annual cycle by using 12-month differences, the variance is reduced by 68% (Fig. 31b). By further identifying and removing the significant dependencies that are still remaining (Fig. 31c), the original variance is reduced by a total of 87%. The eventual residual variation is characteristic of random error and has been checked for randomness by tests of significance.

To identify models for Tateno and the other stations such that the data are reduced to random error (a_t), the autocorrelation function which represents the correlations between data (e.g., deseasonalized data) separated by 1, 2, ..., k months is constructed and is examined for meaningful patterns. For Tateno, the autocorrelation function for the deseasonalized data (Fig. 32) is typical of a second order autoregressive model with a seasonal moving average term. When such a model is postulated and the corresponding coefficients estimated (see model in Table 2), Hill obtains the estimated residuals or errors (a_t) shown in Figure 33. Each model was arrived at independently. Discussion at this point included a comment by John Tukey that "nobody can look at an autocorrelation function and tell what's happening."

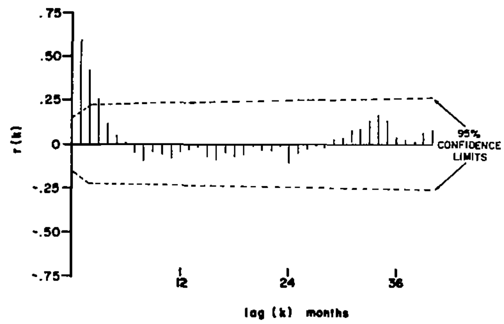

Figure 32. Autocorrelation Function of "Deseasonalized" Tateno Data 7/57-12/69

Hill reiterated that he is letting the data decide what is significant. Elmar Reiter countered that the "periodicity of the atmosphere varies too much to do this" and further proposed that eigenvalues be calculated for as many stations as possible.

As a check of the independency of the residuals, the residual autocorrelation function which shows no unusual correlations or patterns is generated (Fig. 34). This supports the adequacy of the model and reaffirms the result that the data have had their systematic variation removed, leaving only the random part for estimating the background variance (σ^2) in trend detection calculations.

Hill identified the pre-intervention time series models and estimated parameters for each station using the Box-Jenkins Univariate Time Series computer package developed by D. J. Pack at Ohio State University. This package uses an unweighted nonlinear least squares algorithm to estimate the ϕs and θs.

Table 2. Fitted time series models.
Case 1: Identification and fit using data through 12/69
Case 2: Identification and fit using data through 12/71
Case 3: Identification and fit using data through end of series

STATION	CASE	
Edmonton	1	$(1-.20B^1-.24B^2-.08B^3)\ (1-B^{12})y_t = (1-.65B^{12})a_t$
	2	$(1-.22B^1-.21B^2-.08B^3)\ (1-B^{12})y_t = (1-.66B^{12})a_t$
	3	$(1-.19B^1-.20B^2-.06B^3)\ (1-B^{12})y_t = (1-.69B^{12})a_t$
Arosa	1	$(1-.82B^1)\ (1-B^{12})y_t = (1-.66B^1)\ (1-.77B^{12})\ (1+.17B^{25})a_t$
	2	$(1-.82B^1)\ (1-B^{12})y_t = (1-.65B^1)\ (1-.79B^{12})\ (1+.24B^{25})a_t$
	3	$(1-.81B^1)\ (1-B^{12})y_t = (1-.65B^1)\ (1-.80B^{12})\ (1+.26B^{25})a_t$
Tateno	1	$(1-.50B^1-.13B^2)\ (1-B^{12})y_t = (1-.76B^{12})a_t$
	2	$(1-.48B^1-.14B^2)\ (1-B^{12})y_t = (1-.77B^{12})a_t$
	3	$(1-.45B^1-.13B^2)\ (1-B^{12})y_t = (1-.81B^{12})a_t$
Mauna Loa	1	$(1-.65B^1)\ (1-B^{12})y_t = (1-.79B^{12})a_t$
	2	$(1-.62B^1)\ (1-B^{12})y_t = (1-.74B^{12})a_t$
	3	$(1-.64B^1)\ (1-B^{12})y_t = (1-.82B^{12})a_t$
Huancayo	1	$(1-.73B^1+.22\ B^2-.27B^3+.17B^4-.34B^5+.18\ \ B^6)\ (1-B^{12})y_t = (1-.73B^{12})a_t$
	2	$(1-.57B^1+.003B^2-.04B^3-.08B^4-.16B^5+.10\ \ B^6)\ (1-B^{12})y_t = (1-.71B^{12})a_t$
	3	$(1-.49B^1-.02\ B^2-.09B^3-.17B^4-.03B^5+.0003B^6)\ (1-B^{12})y_t = (1-.85B^{12})a_t$
Kodaikanal	1	$(1-.72B^1-.17B^2)\ (1-B^{12})y_t = (1-.62B^{12})a_t$
	2	$(1-.64B^1-.24B^2)\ (1-B^{12})y_t = (1-.67B^{12})a_t$
	3	$(1-.63B^1-.25B^2)\ (1-B^{12})y_t = (1-.70B^{12})a_t$
Buenos Aires	1	$(1-.56B^1+.16B^2-.17B^3)\ (1-B^{12})y_t = (1-.66B^{12})a_t$
	2	$(1-.48B^1+.13B^2-.24B^3)\ (1-B^{12})y_t = (1-.60B^{12})a_t$
	3	$(1-.40B^1+.03B^2-.19B^3)\ (1-B^{12})y_t = (1-.65B^{12})a_t$
MacQuarie Isles	1	$(1-.55B^1)\ (1-B^{12})y_t = (1-.73B^{12})a_t$
	2	$(1-.53B^1)\ (1-B^{12})y_t = (1-.68B^{12})a_t$
	3	$(1-.46B^1)\ (1-B^{12})y_t = (1-.75B^{12})a_t$
Aspendale	1	$(1-.47B^1-.13B^2)\ (1+.17B^{14})\ (1-B^{12})y_t = (1-.70B^{12})a_t$
	2	$(1-.47B^1-.13B^2)\ (1+.17B^{14})\ (1-B^{12})y_t = (1-.72B^{12})a_t$
	3	$(1-.45B^1-.15B^2)\ (1+.17B^{14})\ (1-B^{12})y_t = (1-.74B^{12})a_t$

Let y_t, t = 1, . . ., N be a set of N observations collected at equal time intervals. Using all data obtained prior to the (hypothesized) intervention, the first step of the analysis is to identify a time series model of the form

$$\phi(B)\ (1-B^{12})\ y_t = \theta(B)a_t \quad t=1,2,\ldots,T-1 \tag{7}$$

for each station, where

y_t is the mean monthly total ozone measurements,

a_t is independently distributed $N(o,\sigma_i^2)$ random errors, i=1,...,12 referring to the 12 months

B is the backshift operator (i.e., $B^k y_t = y_{t-k}$)

$\theta(B)$ is the moving average transfer function,

$\phi(B)$ is the autoregressive transfer function,

T is the time of hypothesized intervention, and

$(1-B^{12})$ is used to remove the seasonal variation of the monthly observations.

After obtaining estimates $\hat{\theta}$ (B) and $\hat{\phi}$ (B) of $\theta(B)$ and $\phi(B)$ which account for the phase lag dependencies in the data, a linear ramp function is introduced into the model at the point of intervention as the second step in the analysis. The model is now expressed as

$$y_t = \{\omega/(1-B^{12})\}\ \xi_t + \frac{\hat{\theta}(B)}{\hat{\phi}(B)(1-B^{12})}\cdot a_t \quad t=1,2,\ldots,N \tag{8}$$

where
$$\xi_t = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases}$$

and ω represents the yearly rate of abnormal change in ozone measured in (m atm cm) per year. Rewriting equation (7) as

$$z_{t'} = \omega\ x_{t'} + a_{t'}\ ,\quad t' = -T+1,\ -T+2,\ldots,-1,0,1,\ldots,n \tag{9}$$

where
$$t' = t-T$$
$$n = N - T$$
$$z_{t'} = [\hat{\phi}(B)\ (1-B^{12})/\hat{\theta}(B)]\ y_{t'}$$
$$x_{t'} = [\hat{\phi}(B)/\hat{\theta}(B)]\ \xi_{t'}$$

ω can be easily estimated by linear least squares.

Figure 33. General Methodology

In these series where the variance is not constant from month to month, approximately unbiased but not necessarily minimum variance estimators should be gotten for the ϕs and θs. (The transformation procedure of Box and Cox was applied to the original data [y_t] to see if some power or logarithm transformation of y_t led to constant variance in the transformed variable. No variance stabilizing transformation was found. However, this posed no real problem since the main objective was to find nearly unbiased estimators for the ϕs and θs which could be fixed when estimating ω in the next step.)

The results of the model identification and estimation are summarized in Table 2 for Case 1, Case 2 and Case 3. The latter

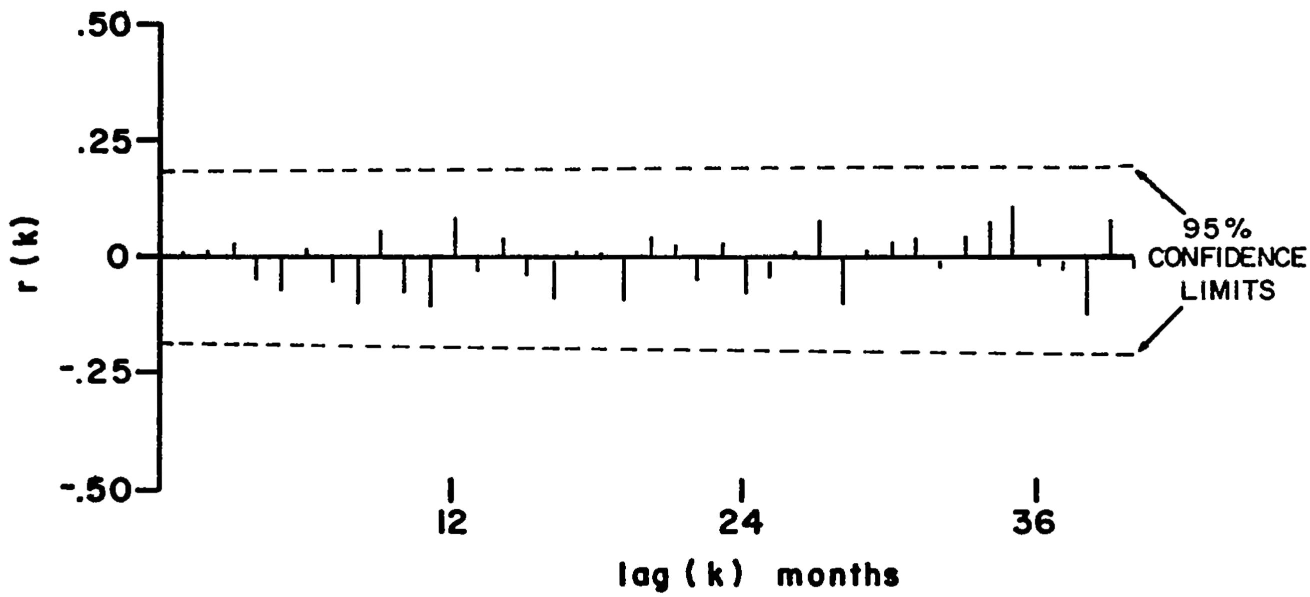

Figure 34. Autocorrelation Function of the Residuals (Tateno Data 7/57-12/69)

is the fit for the complete series through 1975 which is needed for later calculations. For each station the identification program suggests the same model for both the shorter and longer pre-intervention series (Case 1 vs. Case 2).

Once the time series models are thus identified and the parameters are estimated using nonlinear least squares and with data first through 12/69 (Case 1) and then through 12/71 (Case 2), then the ramp parameter ω is estimated from data beginning 1/70 to the end of the series or from 1/72 to the end of the series. By proceeding in this fashion the interval 1970-75 is examined for a possible abnormal change due to intervention (as measured by ω) since it is a period often associated with the predicted onset of man-made ozone depletion. Each model is verified by applying tests of significance to the residual autocorrelations. With the exception of Huancayo, parameter estimates for Case 1 and Case 2 exhibit only slight differences. (Negligible terms are left in the model for Case 2 at Huancayo for comparison purposes only.) The results, in general, suggest that the pre-intervention series are long enough to allow for consistent model identification and estimation. With regard to Huancayo, the relatively large change in parameter estimates may be due to the near nonstationarity of the data series as suggested by the large number of autoregressive terms required to reduce the series to white noise. An instrument drift is one possible explanation of the near nonstationary behavior of the Huancayo series. Inspection of the identified models gives some support for a suspected quasi-biennial cycle. (See, for example, Arosa's moving average term of order 25.)

The results of the first step are the input to the second step which involves estimating the abnormal trend parameter (ω) for each series over the period of hypothesized change or intervention. Estimates $\hat{\omega}$ of ω are obtained as the weighted least squares solution to equation (9). Here the emphasis is on obtaining not only an accurate or unbiased estimate for each ω but also a precise estimate leading to improved sensitivity in trend detection. Theoretically, weighted linear least squares will give minimum variance unbiased estimators when there is non-homogeneity of variance.

The weight assigned to each observation in the analysis is the reciprocal of the standard deviation of all data for that month prior to the hypothesized intervention. For example, in Case 1, the weight for Tateno in May 1972 is the reciprocal of the standard deviation for all May observations for Tateno prior to 1970. By assigning weights in this manner, the weights are not "contaminated" by observations which are potentially depleted. Thus, defining

$$m = 1 + (\text{remainder } t'/12),\ t' \geq 0$$

and w_1 = January "weight"

w_2 = February "weight"

etc.,

the $\hat{\omega}$ is obtained for each series and case as the least square solution of

$$w_m z_{t'} = \omega\, w_m x_{t'} + w_m a_{t'},\ t' = 0,1,\ldots,n \tag{10}$$

where $z_{t'}$, $x_{t'}$ and t are as defined in equation (9). The standard error of $\hat{\omega}$ is calculated for each station as

$$SE\ (\hat{\omega}) = \hat{\sigma}^2(\underset{\sim}{X}'\underset{\sim}{W}^2\underset{\sim}{X})^{-1} \tag{11}$$

where the elements of the vector $\underset{\sim}{X}$,

$$x_{t'} = \{\hat{\phi}(B)/\hat{\theta}(B)\}\xi_t,\ t' = 0,1,\ldots,n$$

(Note $\underset{\sim}{X}'$ is the transpose of the vector $\underset{\sim}{X}$.)

W is a diagonal matrix with w_m on the diagonal

and $\hat{\sigma}^2$ is an estimate of the weighted residual variance.

The estimates of ω and the standard errors are presented in Table 3. For both cases, there are four positive estimates and five negative values for ω covering the nine stations. In only one instance, Huancayo (Case 2), is the estimate of ω different from 0 at the 5% level of significance. The large difference between $\hat{\omega}$ (Case 1) and $\hat{\omega}$ (Case 2) for Huancayo suggests that the increase in the ozone level is a recent phenomenon and may be due to nonenvironmental factors such as an instrument drift. Overall, the results summarized in Table 3 suggest that, in the nine stations analyzed, there has been neither a significant change in the ozone level during the 1970s nor a positive or negative tendency.

A global estimate of change in the ozone, $\hat{\omega}_G$, is obtained by averaging the individual estimates of ω. To simplify the calculation of the standard error of $\hat{\omega}_G$, the nine station residuals were assumed to be independent of one another.

Table 3. Estimated values of ω and standard errors measured in (m atm-cm) per year.

STATION	CASE 1 $\hat{\omega}$	CASE 1 $SE(\hat{\omega})$	CASE 2 $\hat{\omega}$	CASE 2 $SE(\hat{\omega})$
Edmonton	+0.582	1.96	+0.727	2.56
Arosa	-0.407	1.10	-0.638	1.64
Tateno	+0.471	1.10	+0.185	1.56
Mauna Loa	-0.170	0.70	-0.400	0.99
Huancayo	+0.886	0.92	+2.330(1)	1.18
Kodaikanal	-2.220	2.10	-1.895	2.30
MacQuarie Isl.	+1.610	1.84	+3.710	2.70
Buenos Aires	-0.277	1.59	-0.434	2.45
Aspendale	-1.180	0.90	-1.167	1.25
Global Avg.	-0.078	0.48(2)	+0.269	0.65(2)

(1) Significantly different from 0 at 5% level of significance (2-sided).

(2) Pooled estimate.

Hill checked this assumption by studying the cross-correlation coefficients between the residuals for all 36 pairings of the nine series at different lead/lag values. If two stations are independent, the cross-correlation coefficients should have zero mean and show no pattern that clearly denotes a relationship. Hill detailed his tests of the data for independency.

Since not all the series are variance stationary and hence not likely jointly covariance stationary, the cross-correlation analysis is applied to the weighted residuals. It can be expected that the weighted residuals will be approximately white noise. For two independent white noise series, the 95% confidence limits for the estimated cross-correlation coefficient for a lag of k months are approximately $\pm 2 \times (N-|k|)^{-\frac{1}{2}}$. Figure 35 illustrates a typical cross-correlation function which was observed in the analysis.

A summary of the significant cross-correlations for the weighted residuals is given in Table 4 for up to lead/lag 12 months, a period Hill said is more likely to show a relationship between stations, if one exists.

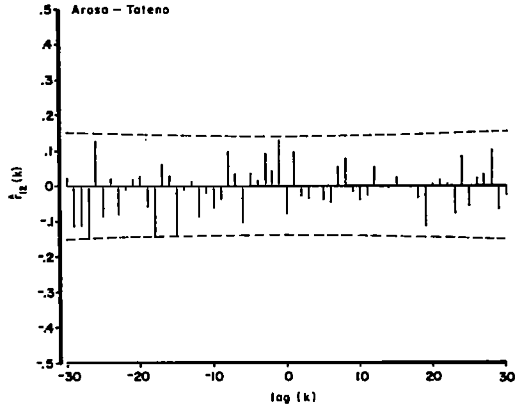

Figure 35. Estimated Cross-correlation Coefficients $\hat{r}_{12}(k)$ of Weighted Residuals from Arosa and Tateno Models (fit through 1975). A Positive Lag (k) Represents Tateno Lagging Arosa by k Months. The Dashed Lines are the Approximate 95% Confidence Limits.

There are 35 significant cross-correlations out of a total of 900 values, 25 lead/lag cross-correlation coefficients calculated for each of 36 pairings. The observed percentage of significant cross-correlations is therefore 4% as compared with the theoretical 5%, if each series is white noise. Although there are no obvious patterns in Table 4, certain of the significant cross-correlations might indicate either a chemical or physical transport phenomenon. For example, two pairings of tropical stations--Huancayo-Mauna Loa and Kodaikanal-Huancayo--show a positive cross-correlation between residuals of the same month (or lag 0). One of these, the largest cross-correlation coefficient to be estimated in this analysis, is 0.35 between Huancayo and Mauna Loa. Despite the fact that the significant cross-correlations are small in magnitude, these two pairings might be suggesting some relationship between tropical stations where the chemical effects related to ozone production dominate. There is a possibility that both chemical production and physical transport factors may explain these and some of the other significant lead/lag cross-correlations. Regardless, neither the pattern of the cross-correlations nor the proportion of significant values seems to contradict the general assumption of independency.

A further test of independency is obtained by applying the asymptotic approximation formula of Haugh

$$S_M^* = N^2 \sum_{k=-M}^{M} (N-|k|)^{-1} \hat{r}_{12}(k)^2 \qquad (12)$$

Table 4. Significant cross-correlation coefficients for weighted residuals. ($A \to B^+$ means B lags A by k months with a significant positive (+) correlation.)

Lead/Lag (k)	Significant Cross-Correlations
0	Kod $\to$ Hua$^+$, Hua $\to$ Mau$^+$, Asp $\to$ Mac$^-$
1	Edm $\to$ Mau$^-$, Tat $\to$ Asp$^-$, Bue $\to$ Kod$^-$, Kod $\to$ Edm$^-$
2	
3	Mac $\to$ Edm$^+$, Tat $\to$ Asp$^+$, Kod $\to$ Asp$^-$, Mac $\to$ Kod$^+$, Mau $\to$ Kod$^-$
4	Mac $\to$ Mau$^-$
5	Mau $\to$ Tat$^-$, Mau $\to$ Mac$^-$, Asp $\to$ Bue$^+$
6	Bue $\to$ Hua$^-$, Tat $\to$ Asp$^+$, Mau $\to$ Asp$^+$, Edm $\to$ Aro$^-$
7	Hua $\to$ Aro$^-$, Kod $\to$ Aro$^+$, Kod $\to$ Tat$^-$
8	Tat $\to$ Kod$^-$, Bue $\to$ Hua$^+$, Tat $\to$ Asp$^-$, Asp $\to$ Tat$^-$
9	Tat $\to$ Mau$^-$, Tat $\to$ Asp$^+$
10	
11	Bue $\to$ Mau$^+$, Hua $\to$ Mau$^-$, Aro $\to$ Asp$^-$
12	Bue $\to$ Mau$^+$, Kod $\to$ Asp$^-$, Bue $\to$ Kod$^+$

where $\hat{r}_{12}$ is the estimated cross-correlation coefficient between series 1 and series 2 at lag(k), and M is set equal to 12. The test statistic S_M* is compared to the χ^2 distribution with $2M+1 = 25$ degrees of freedom. We would not reject series 1 and 2 as being independent if S_M* is less than the $\chi^2 = 37.7$ at the 5% significance level. Only four of the 36 pairings have a significant $S_M^* \geq 37.7$. These are Aspendale-Tateno, Buenos Aires-Mauna Loa, Huancayo-Mauna Loa, and Mauna Loa-Tateno. In the two latter pairings, a single cross-correlation dominates the estimate of S_M*. There is the lag (0) positive cross-correlation between Huancayo and Mauna Loa, and the negative cross-correlation for Tateno lagging Mauna Loa by 5 months. The high S_M* between Aspendale and Tateno is reflecting the significant correlations at k = -9, -8, -6, -3, -1, 8 in Figure 36 and Table 4. (The negative k means Aspendale lags Tateno.) This may be reflecting some transport pattern of ozone between two stations which have nearly the same longitude and are approximately equal distance but opposite in direction from the equator. The Buenos Aires-Mauna Loa value for S_M* is largely affected by the cross-correlations at lags 11 and 12 months (Table 4).

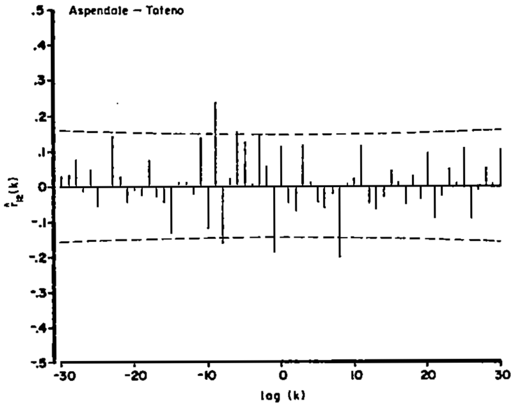

Figure 36. Estimated Cross-correlation Coefficients of Weighted Residuals from Aspendale and Tateno Models (fit through 1975). A Positive Lag (k) Represents Tateno Lagging Aspendale by k Months. The Dashed Lines are the Approximate 95% Confidence Limits.

In summary, two types of statistical tests have been performed on the cross-correlations of the residuals from all 36 pairings of stations. The proportion of significant results does not appear unusual, nor does there appear to be a dominant pattern that would lead one to reject the net or general assumption of independency. There are, however, certain significant cross-correlation coefficients that could be reflecting ozone production characteristics in the tropics and ozone transport between regions. These cross-correlation coefficients are relatively small, and since they represent a reasonably balanced mix of positive and negative covariances, their additive effect on $SE(\hat{\omega}_G)$ is likely to be slight with $SE(\hat{\omega}_G)$ either being slightly larger or slightly smaller than already estimated.

Thus, an analysis of the cross-correlations of the residual series does not lead to a contradiction of the assumption that the nine station residuals are independent of one another. The individual estimates of the standard error of $\hat{\omega}_i$, i = 1,...9, are therefore combined to provide an estimate, $SE(\hat{\omega}_G)$, of the standard deviation of $\hat{\omega}_G$. That is:

$$SE(\hat{\omega}_G) = \left[(1/9)^2 \sum_{i=1}^{9} SE(\hat{\omega}_i)^2\right]^{\frac{1}{2}} \qquad (13)$$

By dividing $\hat{\omega}_G$ and $SE(\hat{\omega}_G)$ by 307, the overall ozone average can be obtained based on the sample of nine stations. To express this as a percent, the estimated abnormal global rate of change per year for Case 1 is -0.03% ± 0.31% (95% confidence limits). For Case 2, the estimate is 0.09% ± 0.42%. Both results suggest there has been no statistically significant change in global ozone persisting in the 1970's.

Setting out to check his linear ramp function with a simulation, Hill determined how well the methodology estimates a predicted decline if the decline were moderately exponential (Fig. 28) instead of linear. All ozone data are artificially reduced according to the ozone depletion model proposed by Jesson (Fig. 37). Using the pre-intervention models in Table 2, a new trend estimate, $\hat{\omega}'$ is calculated for each station after the data are artificially depleted and compared to the original. If the methodology is to be appropriate for ozone trend estimation, the differences $\hat{\omega}_i - \hat{\omega}_i'$, i=1, ...9, when expressed as a percentage of the mean level for station i, should be close to 0.11% for Case 1, where 0.11% is the average amount each data series is depleted per year in the intervention interval. For Case 2, the percent difference should be close to 0.13%. The results of the simulation, summarized in Table 5, indicate close agreement between the artificial exponential depletion and the estimate of depletion from the intervention analysis. These results indicate that the use of the linear ramp function of equation (11) will serve as a good approximation to typical ozone depletion profiles in the 1970s. As a further check on the analysis, each data series was artificially depleted using a linear depletion model. The trend analysis estimated the reduction exactly, as would be expected from the underlying theory.

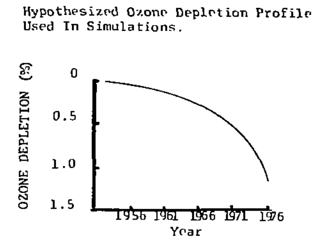

Figure 37. This Profile Represents an Earlier Estimate of Depletion Where the Effect of the Chemistry of Chlorine Nitrate is to Reduce the Depletion Predictions. The Predictions of Figure 37 should not be Compared with Those in Figure 28.

Pursuing the issue of global detectability afforded by the monitoring of ozone levels beyond 1975, Hill recalled that detectability is defined as the smallest abnormal change that would have to occur in the ozone data to be considered significantly different from zero change. Quantitatively, at the 95% confidence level, this is simply expressed as 1.96 x SE($\hat{\omega}_G$). This is converted to a percentage by dividing by 307, the global average of the nine stations and multiplying by 100%.

Since no abnormal trend is found in the period prior to 1975 (Figs. 38 and 39), the models are refitted over the complete data set (Case 3, Table 2). These show no inadequacies such that the identification step had to be redone. Special attention is paid to the ratio: (mean residual)/(standard error) at Huancayo. Since this is not significant, a trend term did not need to be included in the model.

Table 5. Simulation results for artificial depletion shown in Figure 37, where $\hat{\omega}$ is the estimated trend parameter for the original data, and $\hat{\omega}'$ is the estimated trend parameter for the artificially depleted data.

$\Delta\hat{\omega}(\%) = 100\% \times (\hat{\omega}' - \hat{\omega})/(\text{average ozone level for the station})$

	CASE 1			CASE 2		
STATION	$\hat{\omega}$	$\hat{\omega}'$	$\Delta\hat{\omega}(\%)$[1]	$\hat{\omega}$	$\hat{\omega}'$	$\Delta\hat{\omega}(\%)$[2]
Edmonton	+0.582	+0.108	-.13%	+0.727	+0.050	-.19%
Arosa	-0.407	-0.870	-.14	-0.638	-1.200	-.17
Tateno	+0.471	-0.054	-.16	+0.185	-0.400	-.18
Mauna Loa	-0.170	-0.539	-.13	-0.400	-0.914	-.19
Huancayo	+0.886	+0.578	-.18	+2.330	+1.950	-.14
Kodaikanal	-2.220	-2.420	-.08	-1.895	-2.150	-.10
MacQuarie Isl	+1.610	+1.230	-.11	+3.710	+3.080	-.16
Buenos Aires	-0.277	-0.627	-.12	-0.434	-0.812	-.13
Aspendale	-1.180	-1.510	-.10	-1.167	-1.660	-.15
Global Avg.	-0.078	-0.456	-.12%	+0.269	-0.228	-.16%

1 Compare with -.11%

2 Compare with -.13%

EVALUATING FOR TREND 1970 - 1975 AT TATENO

PRE 1970, MODEL IS

$$Y_t = \frac{(1 - \theta_{12} B^{12}) A_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})}$$

IF TREND 1970 - 75, THEN

$$Y_t = \frac{\omega}{(1 - B^{12})} \xi_t + \frac{(1 - \theta_{12} B^{12}) A_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})}$$

WHERE

$$\xi_t = \begin{cases} 0 & \text{BEFORE } 1/70 \\ 1 & \text{FROM } 1/70 \end{cases}$$

QUESTION: IS ω SIGNIFICANTLY DIFFERENT FROM ZERO?

WHERE ω = ABNORMAL YEARLY RATE OF CHANGE IN TOTAL OZONE

Figure 38. Evaluating Trend at Tateno

TREND DETECTABILITY THRESHOLDS FOUND BY

(1) REFITTING MODELS THRU 1975 (SINCE NO PRIOR TREND)

(2) CALCULATE STANDARD ERROR ($SE(\hat{\omega})$) OF FUTURE $\hat{\omega}$

(3) CALCULATE STANDARD ERROR OF GLOBAL AVERAGE $\hat{\omega}_G$

$$SE(\hat{\omega}_G) = \left[(1/9)^2 \sum_{i=1}^{9} SE(\hat{\omega}_i)^2 \right]^{\frac{1}{2}}$$

IF 9 STATIONS INDEPENDENT

(4) CALCULATE THRESHOLD AT 95% CONFIDENCE

$$1.96 \times SE(\hat{\omega}_G)$$

CONVERT TO %

Figure 39. Finding Trend Detectability Thresholds

Prior to calculating $SE(\hat{\omega}_i)$ and hence $SE(\hat{\omega}_G)$ corresponding to an intervention starting at 1/76 and going into the future, consider each term of equation (11). The vector X is a function of the pre-1/76 data and the length of the intervention interval; W^2, the diagonal matrix of weights, is a function only of the preintervention data, and $\hat{\sigma}^2$ is the only term which depends on the post-intervention data. Assuming the residual variation prior to 1/76 has the same variance structure as after 1/76, then $\hat{\sigma}^2$ can be calculated as

$$\hat{\sigma}^2 = (T-1-(p+q))^{-1} \sum_{s=L+1}^{T-1} W_m^{\ 2} \, (y_s-\hat{y}_s)^2 \tag{14}$$

where T corresponds to 1/76, the point of intervention

p is the number of autoregressive terms in the model

q is the number of moving average terms in the model

L is the maximum back order

and $\hat{y}_s$ is the one step ahead forecast made at time s-1 using models of the form in equation (7)

Estimates of detectability for future monitoring periods of 3 to 8 years are presented in Table 6. Column 2 of Table 6 presents detectability estimates based on the sample of the nine stations. The results indicate that an abnormal change of 0.26% per year, persisting for six years (1.56% total), would represent a significant change in the ozone level, if it were to occur. If the monitoring period extended for eight years, a persistent yearly rate of change of 0.21% per year (1.68% total) would be considered significant. Column 3 gives the detectability estimates based on a global network of recording locations equivalent to 18 independent uniformly-distributed sites with residual variation similar to the nine stations analyzed. This "18-station network" can be constructed by including more of the existing ground-based stations in the analysis and/or using satellite data which should be available shortly. Calculations indicate that an abnormal change close to 1% is detectable from the total ground-based network, if such a change were to occur. A combination of data prior to and after January 1976 (e.g., January 1974 - 78) should provide detectability close to the tabulated estimates.

Table 6. Yearly global ozone changes that must persist for p years to be judged statistically significant.

NUMBER OF YEARS	9-STATION GLOBAL NETWORK	18-STATION GLOBAL NETWORK
3	.48%	.34%
4	.37	.26
5	.31	.22
6	.26	.19
7	.23	.16
8	.21	.15

One apparent characteristic of the intervention analysis is that the total detectability lessens as the monitoring interval lengthens. For example, based on the nine stations analyzed, a total change of 1.44% corresponding to 0.48%/year for three years would be significant, while the total change in eight years at 0.21%/year would have to be 1.68% before it could be judged significant (see Table 6). Hill noted that, "intuitively, this is what one might expect. The faster the yearly rate of change, the smaller the total effect needs to be to be judged significant. Very gradual rates of change are more difficult to detect leading to longer elapsed times and greater total changes. A rigorous interpretation lies in the error propagation characteristics of the estimated step function $\{\hat{\omega}/(1-B^{12})\}\xi_t$ with increasing time."

Assuming that the predicted ozone depletion effects for the various compounds are additive, the predicted net global effect is in the range of 1-2% and should by now be large enough to have produced a detectable change in the ozone level. The fact that the trend analysis shows no significant abnormal change in ozone suggests that, although the depletion theories may be correct, the depletion predictions when treated cumulatively yield a result that appears to be too large.

Hill concluded that, "The detectability analysis indicates that the ozone data provide an excellent basis for future monitoring of ozone concentrations. The effect of the early warning provided by the data is to minimize the impact on the environment of a change in the ozone level due to man-related activity, if such a change were to occur. For example, if FC-11 and FC-12 were to cause a 1.56% depletion in the ozone in the next six years, an estimated maximum depletion 1.5 times greater (factor based on NAS calculations), or 2.3%, would occur and be followed by a gradual reversal to normal, assuming that the cause is identified and controlled. (See curve A, Fig. 28.) Thus, attention could center upon climatic and biological impacts resulting from potential maximum reversible changes of 2.3%. Further calculations indicate that the detection capability can be increased by incorporating additional ground station data and/or satellite data into the monitoring scheme (Table 6, column 3)."

Hill noted his assumptions that the cause or causes of an ozone depletion can be identified and controlled. If future monitoring should reveal a significant change in the ozone level, careful investigation of all potential depletion sources, human-related and natural, would be necessary before a cause could be identified. For example, natural trends could be mistaken for man-made effects if the periodicity of the natural trend is greater than the ozone record. This would be true of some shorter data series where cycles, such as a suspected 11-year cycle, may not be fully identified and accounted for in the time series model. Trends which might have been caused by instrument drift or local phenomena can be verified by comparing the suspicious results with those of neighboring stations for consistency. Thus, knowledge of both chemical and physical processes associated with ozone activity will be necessary to complete a cause-and-effect evaluation if statistical analysis of ozone data reveals a significant change in ozone concentration.

Next, Marcello Pagano, from the State University of New York at Buffalo, presented his methodology for analyzing the data by using the time series of ozone monthly means from the same nine-station network (Table 7) that Hill used. Pagano re-iterated that this network serves as a globally-balanced sample of ozone monitoring stations whose time series had no missing

Table 7. Time series of ozone monthly means.

Station and Dates of Observations	Model Method	Ratio of Before and After Mean Square Prediction Errors PRER			Proportion Negative Forecast Errors NEGER		
		24 mo.	48 mo.	72 mo.	24 mo.	48 mo.	72 mo.
AROSA Jan 58-Dec 75	2	1.47	1.26	1.28	.67	.65	.61
ASPENDALE Jan 58-Dec 75	2	.92	1.09	.80	.63	.69	.54
BUENOS AIRES Jan 66-Dec 75	1	1.14	1.46	--	.54	.54	
EDMONTON Jan 58-Dec 75	2	.96	.89	.88	.38	.48	.43
HUANCAYO Jan 65-Dec 75	1	2.05	1.66	1.73	.29	.42	.36
KODAIKANAL Jan 61-Apr 75	3	1.23	1.08	1.19	.56	.57	.48
MACQUARIE ISLES Jan 64-Dec 75	2	1.5	1.76	1.80	.58	.52	.54
MAUNA LOA Jan 64-Dec 75	4	.84	1.23	1.47	.54	.56	.50
TATENO Jan 58-Dec 75	2	.82	1.23	.88	.50	.56	.53
95% Significance Level						.36 .64	.38 .62
PRER., 60		1.70	1.57	1.52			
PRER., 120		1.60	1.47	1.42			

values. The series is also long enough for statistically significant data modeling and parameter estimation.

Analyzing the data consists of dividing each time series into two parts, the earlier part to fit the model and the later part to generate predictors which can be used to judge the difference between the later observations and the earlier. Because of the short length of the ozone series available, Pagano considered three cases of dividing each ozone series into two parts:
(i) data through 1973 for modeling, 1974-75 data for predicting;
(ii) data through 1971 for modeling, 1972-75 data for predicting;
(iii) data through 1969 for modeling, 1970-75 data for predicting.
These three cases are referred to as data sets 2, 4, and 6, respectively. Data set 2 yields the longest record for fitting the model, and data set 6 yields the longest record for judging the predictors.

The following is taken directly from Pagano's paper , as submitted to the proceedings, with the exception of italicized comments.

Tests for detecting changes in probability distribution and downward trends in time series

When the state of a system is describable by a time series $Y(t)$ of measurements over time, a natural question that arises is to test a hypothesis H_0 that there have been no changes in the probability distribution of the state of that system starting at a specified time t_0. One approach to testing H_0, whose rationale has been discussed by Box and Tiao (1976) is as follows: (1) form a data base of values $Y(t)$ at times denoted $t = 1, \ldots, T$; (2) fit a statistical model to the time series $Y(\cdot)$, using its values only up to time t_0 where $t_0 < T$; (3) at each $t = 1,2,\ldots,T$, form the one-step ahead forecasts $Y^{\mu}(t)$ of the value $Y(t)$ at time based on the values $Y(t-1), Y(t-2), \ldots$ at immediately preceding times; (4) comparison of forecasts $Y^{\mu}(t)$ with actuality $Y(t)$ for $t > t_0$ can be used to determine (qualitatively and quantitatively) whether the model for the time series $Y(\cdot)$ fitted to the values before time t_0 describes the probability distribution of the values $Y(t)$ at times after t_0.

One important diagnostic tool is the prediction error ratio, abbreviated PRER. The mean square prediction errors before and after t_0 are denoted

$$\text{PREDERRBEF}\ (t_0) = \frac{1}{t_0} \sum_{t=1}^{t_0} \{Y(t) - Y^{\mu}(t)\}^2$$

$$\text{PREDERRAFT}\ (t_0) = \frac{1}{T-t_0} \sum_{t=t_0+1}^{T} \{Y(t) - Y^{\mu}(t)\}^2$$

in terms of which we define

$$\text{PRER } (t_0) = \frac{\text{PREDERRAFT } (t_0)}{\text{PREDERRBEF } (t_0)}$$

Under the hypothesis that there has been no change in the model, the probability distribution of the statistic PRER (t_0) is approximately the F distribution with ($T-t_0$) and (t_0-p) degrees of freedom, where p is the number of parameters used in fitting the time series model.

The statistic PRER is a test statistic for the hypothesis of no model change at time t_0 which is an "omnibus" or "overall" criterion, in the sense that the test does not specify the nature of the change against which one is testing. One should also employ a "specific" test statistic which specifically tests for the kind of change one is concerned about detecting.

To test the hypothesis that there is a (downward) trend in the measurements, one would use the sign-test statistic

$$\text{NEGER } (t_0) = \text{proportion of prediction errors } Y(t) - Y^{\mu}(t),\ t > t_0, \text{ which are negative}$$

If the process generating the data is stable, then the proportion of negative residuals (actual value Y(t) minus predicted value $Y^{\mu}(t)$) should be about 50%. *That is, Pagano commented, "We are just as likely to underpredict as to overpredict."* If the process measurements have a downward trend, then NEGER (the proportion of negative residuals) should be significantly greater than 50%. (If there is an upward trend, NEGER should be significantly less than 50%.) The expected variability of about 50% NEGER (t_0) when the hypothesis of no model change is true is described by the binomial distribution (with parameters t_0 and 0.5). Under the hypothesis of no model change, a 95% two-sided confidence region for NEGER (48) is 36% to 64%, and for NEGER (72) is 38% to 62% (see table 7).

Ninety-five percent significance levels for the value of PRER are approximately 1.70, 1.57, or 1.52, depending on whether the time span being predicted is the last two, four, or six years, and assuming that the degrees of freedom used in estimating the mean square prediction error over the fitted period is 60. For 120 degrees of freedom these thresholds are approximately 1.6, 1.47 and 1.42.

A technical note: inadvertently, instead of PREDERRBEF (t_0) we computed

$$\text{PREDERRTOT } (t_0) = \frac{1}{T} \sum_{t=1}^{T} \{Y(t) - Y^{\mu}(t)\}^2$$

using the model fitted to the data up to time t_0. One then computes PRER (t_0) using the relation

$$1 - \{\text{PRER } (t_0)\}^{-1} = \frac{T}{t_0} \left(1 - \left\{\frac{\text{PREDERRTOT } (t_0)}{\text{PREDERRAFT } (t_0)}\right\}\right)$$

Methods of time series model fitting

The first step in modeling a time series $Y(t)$ is to consider its level, or means. Since each station clearly exhibits a seasonal pattern (a 12-month periodicity), the monthly means (means of January, February, ..., December, respectively) are first calculated (Fig. 40, Fig. 41). A test is then performed to see if the monthly means can be represented as the sum of a small number of fundamental harmonics; this would achieve a reduction in the number of parameters required to model the mean. Usually the first two harmonics of the period 12 (frequency $2\pi/12$) suffice to model the monthly means by values called the fitted monthly means. The time series is then demeaned by subtracting from each monthly value the fitted mean for that month; the demeaned series is denoted $Z(t)$.

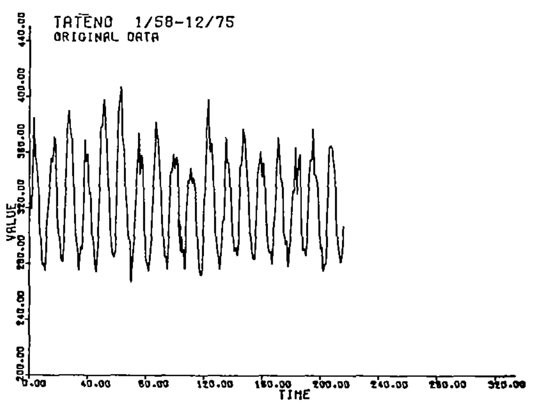

Figure 40. Monthly Means, Original Data

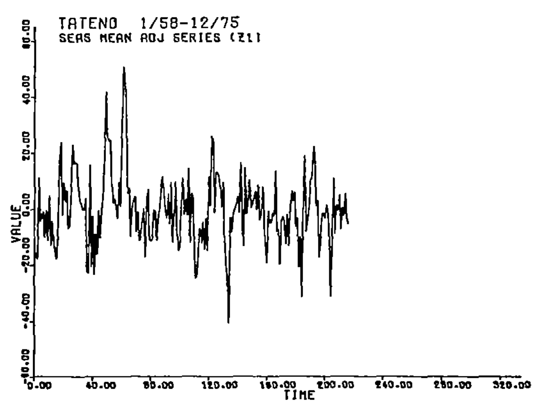

Figure 41. Monthly Means, Seasonal Means Adjusted Series

The first step in modeling $Z(t)$, representing the fluctuations of a monthly time series $Y(t)$ about its fitted monthly means, is to examine the monthly variances; that is the variance of all January values about the fitted mean of January values, ..., the variance of all December values about the fitted mean of December values. Having calculated the monthly variances one would like to test the hypothesis that the variance is constant over the year. Tests of this hypothesis are available only under the simplifying assumption that time series is Gaussian white noise; it is felt that these tests can be used to provide a vague indication, on the basis of which most stations are regarded as having monthly variances which are not constant but vary. *"This correlation," Pagano added, "is exactly what we want--[we want to know] how dependent the future is on the past."* The only stations which we considered whose variances would be regarded as constant are Buenos Aires, Huancayo, and Kodaikanal.

When the monthly variances are regarded as constant we denote $Z(t)$ by $Z1(t)$. When the monthly variances are regarded as varying, we form a de-varianced time series $Z2(t)$ whose value for a given time t is $Z(t)$ divided by the monthly standard deviation for the month corresponding to time t.

For each series $Z1(\cdot)$ and $Z2(\cdot)$, we have two cases: the series is either stationary or periodic-stationary. To intuitively define these concepts, denote the series for expository purposes as $Z(t)$; we will model it as an autoregressive scheme (stochastic difference equation whose right-hand side $\varepsilon(t)$ is white noise or independent random variables):

$$Z(t) + \alpha_t(1)\, Z(t-1) + \ldots + \alpha_t(m)\, Z(t-m) = \varepsilon(t) .$$

Using a periodically varying filter rather than a static one, it is necessary to determine the filter length. Pagano pointed out that "statistical theory argues for a shorter filter to have fewer parameters, while reality argues for a long filter length."

$Z(\cdot)$ is stationary is equivalent to: the autoregressive coefficients $\alpha_t(j)$ do not depend on t and the variance of $\varepsilon(t)$ is constant in t. How many autoregressive coefficients to use is determined by a statistical testing criterion; we consider two criteria which we call CAT and SELECT. $Z(t)$ is periodic-stationary is equivalent to: the coefficients $\alpha_t(j)$ depend only on the month of t, and the variance of $\varepsilon(t)$ also depends only on the month of t. In modeling period-stationary time series we consider three criteria for determining how many coefficients to use for a given month (described in methods 6, 7, 8 below).

The foregoing considerations yield eight possible models for the fluctuations $Z(\cdot)$ of a time series $Y(\cdot)$ about its monthly means.

Method 1: Treat monthly variances as constant, model Z1 as stationary time series, fit autoregressive scheme by CAT.

Method 2: Treat monthly variances as varying, model Z2 as stationary time series, fit autoregressive scheme by CAT.

Method 3: Same as method 1, but fit autoregressive scheme by SELECT.

Method 4: Same as method 2, but fit autoregressive scheme by SELECT.

Method 5: Treat monthly variances as constant, model Z1 as periodic-stationary, fit autoregressive schemes using order determined in method 1.

Method 6: Treat monthly variances as varying, model Z2 as periodic-stationary, fit autoregressive schemes using order determined in method 2.

Method 7: Same as method 6, but fit autoregressive schemes by PCAT for each month.

Method 8: Same as method 6, but fit autoregressive schemes by SELECT for each month.

The length of ozone time series does not seem long enough to use the model of periodic-stationary time series (methods 5, 6, 7 and 8) because of the number of parameters that need to be estimated. In our detailed data summaries, we report the model fitting results using these methods, but we explicitly consider interpretable only the model fitting results using methods 1 through 4.

To choose the most representative model for an ozone time series, the choice will be made from either methods 1, 3 or from methods 2,4 depending on whether one accepts or rejects the hypothesis that monthly variances are constant.

If one would like to select one of the models fitted as being "best fitting," a principle for choosing a modeling method is the following: choose the method which yields smallest overall mean square prediction error using PREDERRTOT on data set 2, and smallest mean square prediction error over the data set not used to fit the model using PREDERRAFT on data set 6. We believe that the conclusions are essentially similar for all models fitted by methods 1-4, but it seems worthwhile to choose one method as being most representative. The test statistics for this method are reported in Table 7.

Table 8. Autoregressive filter of model fitted to fluctuations Z(t)

$$Z(t) + \alpha_1 Z(t-1) + \ldots + \alpha_m Z(t-m) = \varepsilon(t)$$

STATION	DATA SET	α_1	α_2	α_3	α_4	α_5	α_6	α_7
AROSA	2	-.061	-.130	-.048	-.102	-.048	-.153	
	4	-.126	-.135	-.041	-.135	-.057	-.158	.112
	6	-.190	-.141					
ASPENDALE	2	-.283	-.163	-.193				
	4	-.189	-.097	-.214	-.050	-.178	.035	-.025
		-.001	.090	.209	(coefficients α_8, α_9, α_{10})			
	6	-.203	-.165	-.229				
BUENOS AIRES	2	-.257						
	4	-.371						
EDMONTON	2	-.097	-.118	-.028	-.067	-.146		
	4	-.148	-.048	-.073	-.073	-.159		
	6	-.140	-.059	-.070	-.085	-.202		
HUANCAYO	2	-.476	-.195					
	4	-.652						
	6	-.637						
KODAIKANAL	2	-.713	0	0	-.222			
	4	-.730	0	0	-.200			
	6	-.875						
MACQUARIE ISLES	2	-.323	-.068	-.091	.217			
	4	-.382						
	6	-.434	.174					
MAUNA LOA	2	-.576						
	4	-.470						
	6	-.457						
TATENO	2	-.247	-.285					
	4	-.312	-.253					
	6	-.384						

Table 8 summarizes the coefficients of the stationary autoregressive models fitted to the fluctuation series Z(t) at each station.

Since this methodology should work with any parameter that varies seasonally, London proposed applying the same technique to temperature data to see if the methodology successfully predicts the world-wide cooling that has occurred since the 1940s. If the technique does forecast the temperature change, it would clearly strengthen the methodology and lend greater evidence to the conclusions about other seasonal variations such as ozone.

Conclusions

The values of the test statistics summarized in Table 7 do not reject the hypothesis that there has been no downward trend in the measurements of ozone levels in the period through 1975.

By the test statistic NEGER (proportion of negative forecast errors) Arosa and Aspendale could be considered to have a significantly high proportion in their forecasts over 1971-75, but not over 1969-75. Their values of the test statistic PRER is not significantly high.

The values of PRER for Huancayo are significantly high which indicates a change in the probability distributions of ozone levels; to interpret this one uses the values of NEGER which are just barely significantly low for Huancayo. Therefore, if there is any statistical evidence of trend in ozone measurements at Huancayo, it is an upward trend.

On the other hand, the values of PRER for Macquarie Islands are significantly high, but NEGER is non-significant. Therefore, the ozone measurements at Macquarie Isles might provide statistical evidence of a downward trend. It is the only station with this property. It is also the station for which our time series model fits the worst when one compares the mean square forecast error with the overall variance of the time series (summarized in Table 9).

Table 9. Comparison of mean square forecast errors with overall variance of time series

STATION	MEAN	VARIANCE	MEAN SQUARED FORECAST ERROR PREDERRAFT	
			Last 4 Years	Last 6 Years
AROSA	334.3	245.5	283.1	276.1
ASPENDALE	320.2	138.2	94.7	79.4
BUENOS AIRES	287.9	152.9	168.6	
EDMONTON	358.0	324.0	250.1	252.8
HUANCAYO	263.5	22.8	21.4	22.0
KODAIKANAL	261.2	103.6	19.1	20.0
MACQUARIE ISLES	340.5	374.3	462.1	455.1
MAUNA LOA	277.1	78.4	59.7	66.2
TATENO	324.6	179.4	123.9	113.3

Janet Campbell of NASA Langley reviewed the "imperfect data question." She defined the following terms:

$\hat{O}_3(t,x)$ = Dobson measurement

$O_3(t,x)$ = Actual total ozone

where both are associated with a time t and position x. The error associated with this measurement is:

$$\varepsilon(t,x) = \hat{O}_3(t,x) - O_3(t,x)$$

In order to determine data quality, one must know something about the properties of $\varepsilon(t,x)$.

Campbell showed two data records which were made simultaneously by side-by-side Dobson instruments at Arosa, Switzerland. Since both instruments are attempting to measure the same $O_3(t,x)$, then differences in simultaneous measurements are, essentially, differences in errors. Thus, one can gain some insight into the magnitude of errors at this station by examining these differences.

Writing:

$$\begin{array}{ccc} \text{known} \leftarrow & \vdots & \rightarrow \text{unknown} \\ \hat{O}_3(t,x) & = & O_3(t,x) + \varepsilon(t,x) \end{array}$$

and noting that the left-hand side of the equation is the known (observable) information and the right-hand side represents an unknown partitioning, then the known average of a set of Dobson measurements is an estimate of the average true ozone plus the average error (bias). That is:

$$\begin{array}{ccc} \text{known} \leftarrow & \vdots & \rightarrow \text{unknown} \\ E(\hat{O}_3(t,x)) & = & E(O_3(t,x)) + E(\varepsilon(t,x)) \end{array}$$

If $\varepsilon(t,x)$ is unbiased, then $E(\varepsilon(t,x))$ tends to zero for a "long enough" averaging period. The assumption of no bias may not be reasonable, however.

Trend estimates are limited by the variance of the data, that is, by:

$$\begin{array}{ccc} \text{known} \leftarrow & \vdots & \rightarrow \text{unknown} \\ \mathrm{Var}(\hat{O}_3(t,x)) & = & \mathrm{Var}(O_3(t,x)) + \mathrm{Var}(\varepsilon(t,x)) + 2\,\mathrm{Cov}(O_3,\varepsilon) \end{array}$$

It is desirable for the errors to be independent of the actual total ozone (i.e., $\mathrm{Cov}(O_3,\varepsilon) = 0$). If this is the case, then

$$\mathrm{Var}(\hat{O}_3(t,x)) \geq \mathrm{Var}(O_3(t,x))$$

and

$$\mathrm{Var}(\hat{O}_3(t,x)) \geq \mathrm{Var}(\varepsilon(t,x))$$

so that the known data variance provides an upper bound on the variances of O_3 and ε.

To decide about the existence of a bias or whether or not errors are correlated to O_3, one should "pull the errors apart" and look at potential error sources. Three major causes of error are:

1. incorrect instrument calibration, poor maintenance, etc.
2. algorithms used to convert measured radiances to total ozone estimates
3. meteorological/geophysical variables.

A calibration error, for example, could produce either a constant bias or a time-varying bias (drift) in the data. Correlations between ε and O_3 can result from the correlation of both with a third variable such as another atmospheric constituent.

There are some types of errors which can seriously affect trend estimation techniques whereas others are not so serious. An unknown but constant bias will not affect trend estimates, whereas a bias which changes over time can either be mistaken for an ozone trend or cancel a real ozone trend of opposite sign. The actual magnitude of errors is not necessarily a problem because this is accounted for in the trend estimation techniques, provided that the data variance properly reflects these magnitudes. This condition will be met, as discussed earlier, if $Cov(O_3,\varepsilon) = 0$. It is important to examine error sources and attempt to identify or remove the serious errors.

There are two possible mistakes which can be made in our conclusions. The "Type I" mistake would occur if we were to detect a trend which doesn't exist, and the "Type II" mistake would result if we were to fail to detect a trend which does exist. As previously mentioned, errors which contain a trend in themselves could result in either of these mistakes. A Type I error could also result from too short a data record when a natural low frequency oscillation is mistaken for a monotonic trend. A Type II error can result from an inadequate model in which residual variances are too high. The models of Hill, Sheldon and Tiede, with their low trend detectability thresholds, do not suffer from this problem. The major type of data inadequacy which can invalidate their results would be trending errors.

(Campbell noted: "This discussion of errors applies only to situations where one is analyzing time series at one or more stations and making inferences about those stations. Where inferences are 'extrapolated' beyond the stations for which data are available, as for example, a global mean estimated using data from 9 stations, other errors can occur and these are not addressed here.")

Komhyr emphasized the importance of Type I errors where the "net effect could be no trend" and suggested that it might be useful to look at variations in different levels of the atmosphere. He added, "Statistical analysis can tell you if a trend is going on or not, but physical and chemical analysis must explain the data."

Gille observed that the ozone concentration in the 40-km region reflects the first effects of photochemistry. Since the natural variance of ozone concentration is thought to be low at this altitude, it is a good place to look for the first evidence of changes in ozone photochemistry. In addition, the variance in limb scanning data is low at this altitude, giving two reasons for an improved signal-to-noise ratio.