Tellus (1991), *43B*, 171–187

An exploratory analysis of long-term trends in atmospheric CO_2 concentrations

By M. YA. ANTONOVSKY, *International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria and* V. M. BUCHSTABER, *All-Union Research Institute of Physicotechnical and Radiotechnical Measurements, Moscow, USSR*

(Manuscript received 17 October 1989; in final form 27 December 1990)

ABSTRACT

A new methodological approach for the analysis of monitoring data is discussed. The main ideas are illustrated for the example of the CO_2 problem. The analysis of CO_2 concentrations obtained from a global network of monitoring stations permitted us to construct a nonparametric evaluation of the spatial-temporal distribution of this field. We propose a parabolic parameterization of the long-term tendency of this field as a function of time (in one-year time steps). A function of the predictive ability of a model is defined on the basis of the technique of "supervised training." This function is computed for a parabolic model and it is shown that this model constructed for the first 15 years of observations evaluates the tendency for the next 15 years quite well. The main problem that we solve in this paper is how to correlate the projections of different models for the carbon cycle and different scenarios of the annual release of carbon into the atmosphere with the projections that reflect parameterization of the trends of CO_2-monitoring data.

1. Introduction

It is shown that the projection of a parabolic parameterization agrees well with the series of projections obtained on the base of models that used the so-called "Reference Scenarios." We introduce the criteria of the risk of a projection using the functional of connection between the parameterization of the observed trend and the analytical expression for the future concentrations. This functional plays the role of functional of least action in the problems of variational calculus and optimal control. It gives a comparison of the different analytical expressions of future trends in the concentrations of atmospheric CO_2 discussed in the literature. Described in the Appendix are the relationships between different methods of decomposition of seasonal time series on components (the Tukey method, factor analysis, the SABL method).

2. Statement of the problem and discussion of the results

Using the data base provided by CDIAC (the Carbon Dioxide Information Analysis Center), Oak Ridge National Laboratory, it is possible to construct a realization (picture) of a spatially and temporally distributed field of monthly mean concentrations of atmospheric CO_2 on the globe. In this connection, it is important to stress the pioneering work of C. D. Keeling that is the benchmark (from 1957) of regular gathering and data analysis of atmospheric CO_2 (see Keeling, 1987 and Keeling et al., 1989). In Trabalka (1985) a three-dimensional perspective of the latitude and time variation of global atmospheric CO_2 concentrations ("the pulse-of-the-planet") was constructed based on flask measurements for 1979–1982.

In the paper of Tans et al. (1990) are given annual average concentrations of CO_2 since 1981

till 1987 obtained from the Geophysical Monitoring for Climatic Change (GMCC) division of the National Oceanic and Atmospheric Administration (NOAA), which has been collecting air samples in flasks for CO_2 analysis from more than 20 sites.

After constructing such realizations in the form of a spatial and temporal table, the question of analysis arises. The character of this analysis is defined by the problem under consideration. We are speaking about the following: there exists a set of factors that define the global carbon cycle. The role of an exploratory analysis of CO_2 monitoring data is to determine the regularities in the structure of the data and in the context of explanatory notions to evaluate the sensitivity of monitoring to changes in these factors.

Antonovsky et al. (1988) investigated the presentation of the series of mean monthly concentrations at a monitoring station in the following form:

$$C(Y, M, r) = C_0(r) + C_1(Y, r) + C_2(M, r) + E(Y, M, r) \quad (1)$$

where Y is a year of measurement, M is a month of measurement, r are the coordinates of a station, C_0 is the characteristic value, $C_1(Y_1, r)$ are yearly effects, $C_2(M, r)$ are the monthly effects, and $E(Y, M, r)$ is a table of residuals. The values of C_0, C_1, C_2, and E are produced by the Tukey method of median analysis of the two-way tables (Emerson and Hoaglin, 1983) (for details, see Section 2.) This method will be applied to several geophysical applications. Further, we will show the validity of Tukey's method for the description of a realization

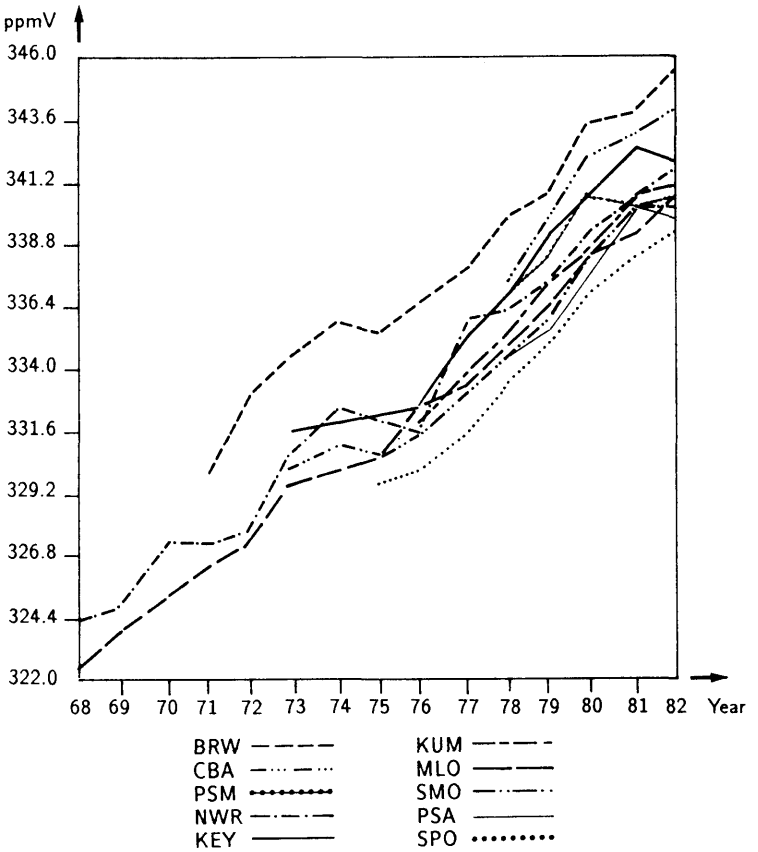

Fig. 1. Characteristic value + yearly effects (Tukey method of median analysis).

of the field of concentrations of CO_2 and we will discuss its variants connected with different criteria of the smallness of residuals $\{E(Y, M, r)\}$ and its linkages with other methods of decomposition of seasonal time series on components including the SABL method using the problem under consideration (Cleveland et al., 1983).

Fig. 1 shows a graph of the set of times series $\{C_0(r) + C_1(Y, r)\}$, where r is a running coordinate of the monitoring stations, whose locations are given in Table 1. Fig. 2 shows a graph of the set of the time series $\{C_2(M, r)\}$. The set of the series $\{C_0(r) + C_1(Y, r)\}$ (see Fig. 1) as is shown in Section 2, can be considered as a realization of some process and as an estimation of the trend of the time series of concentration of atmospheric CO_2, which describes the whole atmosphere. Then the series $C_0 + C_1(Y)$ for the Mauna Loa station can be considered as a nonparametric estimation of the trend of the process. Use of the data from Mauna Loa (the longest series of observations) for constructing nonparametric and parametric estimations of trends in global atmospheric CO_2 is discussed very widely in the literature. At the same time, Keeling et al. (1989) have proposed as a nonparametric estimation of the trend, the mean of the series of the mean yearly concentrations of CO_2 from Mauna Loa and from the South Pole. Figs. 1 and 2 show the high information content of the method we chose for describing the initial realization of the field of concentration. In Fig. 1, one can see a decrease in the characteristic values of concentration in moving from the North Pole to the South Pole in each year of the observation.

From Fig. 2 it is seen that the shape of the curve's seasonal oscillations depends essentially on

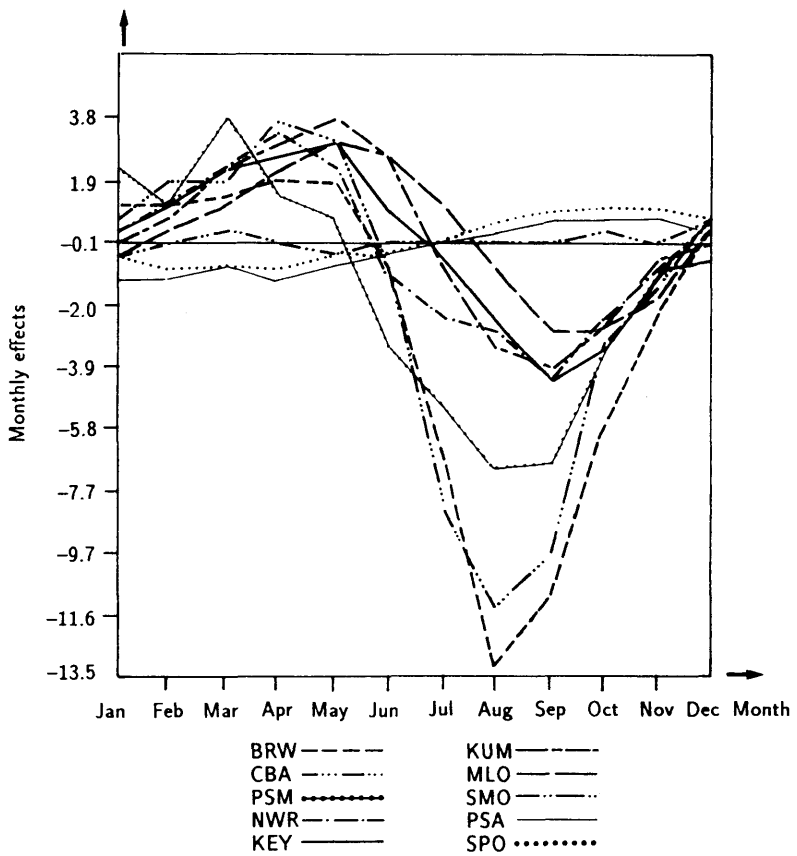

Fig. 2. Monthly effects (Tukey method of median analysis).

Table 1. *GMCC (NOAA) Stations Network*

Name	Symbol	Longitude	Latitude	Region
Barrow	BRW	156°W	71°N	Alaska
Cold Bay	CBA	162°W	55°N	Alaska
Point Six	PSM	110°W	47°N	State of Montana
Niwot Ridge	NWR	105°W	40°N	State of Colorado
Key Biscayne	KEY	80°W	25°N	State of Florida
Kumukahi	KUM	158°W	22°N	Hawaii
Mauna Loa	MLO	155°W	19°N	Hawaii
American Samoa	SMO	170°W	14°S	South Pacific
Palmer	PSA	64°W	64°S	Antarctica
Amundsen Scott	SPO	24°W	89°S	Antarctica

the station (r). If we choose as characteristics of these curves the following set of criteria:

1. $T_+ = \{M: C_2(M) > 0\}$,
   $T_- = \{M: C_2(M) < 0\}$,
2. $A_+ = \max_{M \in T_+} C_2(M)$,
   $A_- = \max_{M \in T_-} |C_2(M)|$
3. $M_{*+} = \arg A_+$,
   $M_{*-} = \arg A_-$
4. Histogram of the sets $\{C_2(M); M \in T_+\}$,
   $\{C_2(M); M \in T_-\}$,

then a simple analysis shows that the values of these characteristics for each of the stations are in accordance with the behavior of the biota in a latitude belt presented by the stations. One of the main problems of this paper is to conduct a comparison between the main projections of future CO_2 concentrations. The future concentrations of CO_2 that are noted by B. Bolin (1986) could be presented as a function of scenarios of emissions of CO_2 in the atmosphere based on one of the two methods most accepted by the CO_2 community: the concept of airborne fraction and globally-averaged models of the carbon cycle, that due to rapid turbulent transfer ensure that background CO_2 concentration variations over different parts of the world remain near 4 ppmv. For this purpose, in Section 3 we construct and research the parameterizations of the series $C_0 + C_1(Y)$ for the Mauna Loa station in the classes of polynomial and exponential functions. For comparison of parameterizations that are equivalent by the criterion R^2 it is proposed to use the function of projection ability that exploits the well-known procedure in pattern recognition known as "supervised training." It is shown that by this criterion from two of the three parameters of the parabolic $\{at^2 + bt + c\}$ and exponential $\{ae^{bt} + c\}$ family for the series $C_0 + C_1(Y)$ under consideration, the preferred one is a parabolic family. Let us remark that for this series the parabolic model constructed on the data for the first 15 years of observations projects the next 15 years quite well, i.e., is in good accordance with the measured data while the exponential models yield poor results. For the parabolic family the function of projection ability is determined.

Finally, the following parameterization of the trend of the time series of atmospheric concentrations of CO_2 is proposed:

$$C(t) = 0.0186(t - 1957)^2 + 0.58(t - 1957) + 314.6. \quad (2)$$

In Section 4, a comparative analysis of this parameterization is given with several parameterizations taken from the literature on the CO_2 problem. Supposing that the function $C(t)$ in (2) describes the trend up to the present time of the concentrations of atmospheric CO_2, we make a comparison of the values for the years 2000–2100 with the projections of future concentrations of CO_2 discussed in the works of Bolin (1986), Budyko and Izrael (1987), and others, and in the most recent IPCC (WMO/UNEP Intergovernmental Panel of Climatic Change) publications.

In estimating the dynamics of future inputs of CO_2 on the greenhouse effect (including all radiative gases), Wigley (1987) described a func-

tion $W(t)$ expressing the concentration of CO_2 in the atmosphere after the year 1985. (Let us recall that at the present time the contribution of this input is estimated at 50%). For this purpose, Wigley constructed a parabolic smoothing

$$C_W(t) = 311 + 0.0208(t - 1944)^2$$

of the data for the period 1958–1985. Then he constructed the function $W(t)$ as a parabola with the following conditions: $W(1985) = C_W(1985) =$ 346 ppmv, $W'(1985) = C'_W(1985)$ and $W(2030)$ is the best 2030 estimate given by Ramanathan et al. (1985), viz., 450 ppmv. This projected value is based on earlier work by Wuebbles (1981) but it agrees well with other, more recent estimates. In their US Department of Energy study, for example, Trabalka et al. (1985) give a range of 450–530 ppmv for 2025, equivalent to 430–578 ppmv for 2030. From formula (2) it follows that $C(1985) = 345$ ppmv, with a confidence interval of 343–347 ppmv, $C(2030) = 456$ with a confidence interval of 448–464 ppmv.

Thus, we have in the interval from 1958 to 2030, two functions $C(t)$ and $W(t)$, where

$$W(t) = \{C_W(t): 1958 \leqslant t \leqslant 1985;\ W(t): t \geqslant 1985\}.$$

In Section 4, we introduce and discuss a method of constructing functionals (statistical criteria of the type "Risk of Projections") for comparing the functions that were obtained from the monitoring data and the model projections. Also constructed in Section 4 are the functionals that show preference for the function $C(t)$ in comparison with the function $W(t)$ in the interval 1958–2100. Also, the prediction on the basis of the model of the carbon cycle (see Siegenthaler, 1983) for the upper boundary of emissions of CO_2 by the year 2050 gives the value 531 ppm, compared to a value $C(2050) = 529$ ppm. Finally, many models predict a doubling of CO_2 concentrations with respect to preindustrial levels (275–290 ppm) in the interval 2050–2060. These predictions are also confirmed by our parabola value: $C(2055) = 550$ ppm, $C(2060) = 572$ ppm. Moreover, Rotty and Reister (1986) describe reference energy scenarios in the framework of the model of the carbon cycle of Björkström (1979), who for 2100 predicts a level of concentration of CO_2 equal to 775 ppm, while $C(2100) = 778$ ppm. Our computer experiments with models of Emanuel et al., 1984; Goudariaan and Ketner, 1984, are also confirmed by our parabola.

3. Data analysis of observations of a spatial-temporal field of CO_2 concentrations

The initial point of analysis, as we stressed above, is the $(N \times K)$-table, where N is the number of months of observations and K is the number of stations. Each of the rows of the table corresponds to one of the stations of the monitoring network and is the time series of mean monthly concentrations of atmospheric CO_2. Classical spectral analysis (discrete Fourier transform) has shown that each of the time series under consideration has a maximal value of the amplitude spectrum clearly expressed on a frequency $\frac{1}{12}$ corresponding to a yearly cycle. A typical amplitude spectrum is given in several papers (see, for example, Antonovsky et al., 1988, where we applied Vinograd's algorithm to determine the time series obtained for different stations, having a different length non-equal to the power of 2). This result is in accordance with the idea that the biota play a leading role in forming the cycles of the component of time series of mean monthly concentrations of atmospheric CO_2. So at this stage of exploratory analysis we have a reason to apply the methods of seasonal times series that could be subdivided in two groups of methods: (1) on the basis of the "autoregressive-integrated moving average" (ARIMA) model (Box and Jenkins, 1976) and (2) on the methods of decomposition into components (Cleveland et al., 1983 and others).

We apply Tukey's method for analyzing two-way tables. In the appendix we will discuss general approaches to decomposition of seasonal time series into components. On this basis we will discuss the connection between Tukey's method, classical factor analysis (Aivazyan et al., 1989), and the SABL method (Cleveland et al., 1983). In Tukey's method (Emerson and Hoaglin, 1983), the table $C(Y, M)$ in the form

$$C(Y, M) = C_0 + C_1(Y) + C_2(M) + E(Y, M) \quad (1)'$$

is constructed by the use of a function $G(E(Y, M))$ that assesses the level of smallness of the table of

residuals using statistical criteria. Usually the following criteria are used:

$$G_1(E(Y, M)) = \sum_Y \sum_M |E(Y, M)|;$$

$$G_2(E(Y, M)) = \sum_Y \sum_M E(Y, M)^2,$$

when G is constructed on the basis of maximal likelihood principles under the hypothesis of normality of the distribution for residuals (in the case of G_2) and the hypothesis of a Laplace distribution for residuals (in the case of G_1). In these two cases we can get the table (1)′ for $C(Y, M)$ with the help of Tukey's algorithm of smoothing of a table by the method of means in the case of G_2 and by the method of medians in the case of G_1. A detailed study of the relationship between the method of means and medians (practical recommendations for their applications) is contained in Emerson and Hoaglin (1983). As a result of a comparative analysis of the different methods of decomposition of the seasonal series of components (see appendix) we chose Tukey's method of medians. The algorithm of this method uses the procedure of minimization of the functional $G_1(E(Y, M))$ and is an iterational process of constructing the decompositions: $C(Y, M) \overset{(i)}{=} C_{0,i} + C_{1,i}(Y) + C_{2,i}(M) + E_i(Y, M)$, $i = 0, 1, \ldots$ where $C_{k,0} \equiv 0$, $k = 0, 1, 2$, $E_0(Y, M) = C(Y, M)$.

As the ith decomposition is constructed, the components of the $(i+1)$th are defined by the following formula:

$$C_{0,i+1} = C_{0,i} + m_{1,i} + m_{2,i};$$

$$C_{1,i+1} = C_{1,i} + m_i(Y) - m_{1,i};$$

$$C_{2,i+1} = C_{2,i} + m_i(M) - m_{2,i};$$

$$E_{i+1}(Y, M) = E_i(Y, M) - m_i(Y) - m_i(M);$$

$$m_i(Y) = \operatorname{med}\{E_i(Y, M),\ M = 1, \ldots, 12\};$$

$$m_{1,i} = \operatorname{med}\{C_{1,i}(Y) + m_i(Y),\ Y = 1, \ldots, \tilde{Y}\};$$

$$m_i(M) = \operatorname{med}\{E_i(Y, M) - m_i(Y),\ Y = 1, \ldots, \tilde{Y}\};$$

$$m_{2,i} = \operatorname{med}\{C_{2,i}(M) + m_i(M),\ M = i, \ldots, 12\}.$$

To stop the algorithm on the $(i+1)$th step, one may use, for example, the condition

$$\max_{Y,M} \frac{E_i(Y, M) - E_{i+1}(Y, M)}{E_i(Y, M)} \leqslant 0.01.$$

The algorithm described above was applied to a table of data of a spatially and temporally distributed field of concentrations of atmospheric CO_2 provided by CDIAC. As we stressed above, we have a description of this table in the form of three sets: the set of characteristic values $\{C_0(r)\}$, the set of the series of yearly effects $\{C_1(Y, r)\}$, and the set of the series of the monthly effects $\{C_2(M, r)\}$.

The next task is to estimate the statistical validity of the decomposition we have obtained and to determine the geophysical sense of these components. We analyze the data for the monitoring stations given in Table 1.

The stations are numbered in order from the North Pole to the South Pole. Let $f_k(t)$ be the analyzing series of observations, $t = 1, \ldots, 12Y_k$, where Y_k is a number of years of observations for the kth station, and t is the number of month. To the series $f_k(t)$ corresponds the $12 \times Y_k$th table $C(Y, M, r_k)$, where r_k are the coordinates of the kth station. Applying the method of median smoothing to the table $C(Y, M, r_k)$, we obtain the characteristic value C_0^k, yearly effects $C_1^k(Y)$ and monthly effects $C_2^k(M)$. Then we form the series $\tilde{f}_k(t)$, the value of which in the tth month of observation equals $f_k(t) - C_0^k - C_1^k(Y)$, where $Y = [t/12] + 1$ is a year in which t observations have been made. A spectral analysis of the series $\tilde{f}_k(t)$ for each k shows that in the amplitude spectrum of this series the first statistically significant value is determined on a frequency that corresponds to one year. So it is shown that in the subtractions from the series $f_k(t)$ the component $C_0^k + C_1^k(Y)$ is equivalent to a pass through a low-frequency filter. It gives assurance that the series $C_0^k + C_1^k(Y)$ is a non-parametric estimation of the trend of the series $f_k(t)$.

Fig. 1 shows graphs of the set of the time series $\{C_0^k + C_1^k(Y)\}$, $k = 1, \ldots, 10$. The set of these series can be considered as a realization of some process and is an estimation of the trend of the time series of CO_2 concentrations that describes the whole compartment of the atmosphere. From an analysis of this picture, it follows that the series $C_0 + C_1(Y)$ for the Mauna Loa station can be considered as a nonparametric estimation of the trend of the global process.

Section 3 is devoted to parameterizations of the trend and comparisons among them. For each station, let us consider the series $\tilde{\tilde{f}}_k(t)$, the value

of which in the tth month of observations is equal to $\tilde{f}_k(t) - C_2^k(M)$, where $M = t - 12[(t-1)/12]$, $1 \leqslant t \leqslant 12Y_k$, $M = 1, 2, ..., 12$, and $[p/q]$ represents the integer part of the fraction p/q.

There is no statistically significant discrete value in the amplitude spectrum of the series $\tilde{\tilde{f}}_k$ for each k as it is shown by the spectral analysis of this series. This means that subtracting the component $C_2^k(M)$ from the series $\tilde{f}_k(t)$ leaves no single separate periodic component. It gives assurance that the component $C_2^k(M)$ is a nonparametric estimation of seasonal oscillations in the initial series. Fig. 2 shows graphs of the monthly effects of each for the 10 monitoring stations. Fig. 2 was discussed in Section 1.

For an analysis of the importance of the biota in the global carbon cycle, it is important to investigate the variation of the amplitude of interannual oscillations. In the appendix it is shown that Tukey's method is a particular case of a more general method that, for example, gives the algorithm of computation of variations of such amplitudes.

Concluding this section, let us remark that a statistical validation of the presentation of the series $f_k(t)(\approx C(Y, M, r_k))$ as a sum of components $C_0^k + C_1^k(Y) + C_2^k(M)$ is confirmed by the fact that the matrix of remainders $E(Y, M)$ is small relative to many important criteria, independent of the criterion that has been used for constructing the expansion in equation (1). For example, the randomness of the series $E(Y, M, r_k)$ is characterized by the fact that its amplitude spectrum of the series $\tilde{\tilde{f}}_k(t)(\approx E(Y, M, r_k))$ has no statistically significant value.

4. A parabolic parameterization of the trend of concentrations of CO_2

In the literature analytical expressions (parameterizations) are discussed for the trend of concentrations of CO_2 in the atmosphere in the class of exponential functions (Wuebbles et al., 1984; Baes and Killough, 1985; Keeling et al., 1989, and others) and in the class of parabolas (T. Wigley, 1987).

As we stressed above for parameterization of a trend of CO_2 we use the series $C_0 + C_1(Y)$ constructed on the data of Mauna Loa (see Section 2).

Using the least squares method, Antonovsky et al. (1988) constructed the parameterization in the class of exponential functions in the form

$$C_e(t) = 292 + 22.2 \exp[0.03(t - 1957)] \qquad (3)$$

$t = 1958, 1959, ...$ with the mean square of the error $\sigma^2 = 0.28$ and in the class of polynomial function in the form (2) with the mean square of the error $\sigma^2 = 0.22$.

The task of computation of polynomials by the least squares method requires multiple linear regression. The standard program determines the coefficients of polynomials and approximate confidence intervals for the parameters.

Thus a low parabola $LC(t)$ and an upper parabola $UC(t)$ can be constructed simultaneously with parabola $C(t)$ (see eq. (2)).

$$LC(t) = 0.0176(t - 1957)^2 + 0.580(t - 1957) + 314.37,$$

$$UC(t) = 0.0196(t - 1957)^2 + 0.614(t - 1957) + 314.83.$$

It appears that the graph of the exponential function $C_e(t)$ (see equation (3)) for t in the interval 1958–2000 is in the domain, restricted by the parabolas $LC(t)$ and $UC(t)$. In the same domain are the graphs of the other parameterizations of the trend of atmospheric CO_2 taken from the literature. From this fact and from the general statistical principles it is concluded that the parabolic model is preferable to the exponential, which is the solution of the problem of a nonlinear regression.

During the interval 2000–2100, the values of the functions $C(t)$ and $C_e(t)$ essentially differ. For example $C(2050) = 529$ ppmv, $C_e(2050) = 700$ ppmv.

As noted in Section 1, the value of function $C(t)$ in this interval, unlike the function of $C_e(t)$, is in accordance with the series of other published projections. The question remains as to which of the models, parabolic or exponential, is in better accordance with the monitoring data by the criteria "supervised training." Within these criteria, the data are subdivided into two groups: training and controlling. From the training group of data we estimate the parameters of the model. Then we statistically compare the projection from

Tellus B/43/2-7

the model constructed with the data from the control group.

In this connection, let us describe the realization of this approach as a function of projection ability of the model. Let $\Phi(t; \alpha_1, ..., \alpha_p)$ be a chosen model for data $y_1, ..., y_n$, where $\alpha_1, ..., \alpha_p$ are estimated parameters and $\Phi(\ )$ is a chosen analyzing shape of functional dependence. Then the projection ability of the model is a function $F(\Phi)$ of two natural (numbers) arguments k and l, $k + l \leqslant n$, where k is the number of years of observations on which we are calculating the parameters of the models, l is the number of years during which we use these models for prediction, n is the number of years of observation, and a value of function $F(k, l)$ is a mean square of the error of the prediction σ^2.

Thus

$$F(k, l) = \frac{1}{l} \sum_{j=1}^{l} (y_{k+j} - y^*_{k+j})^2,$$

where

$$y^*_{k+j} = \Phi(t_{k+j}, \alpha_1^*(k), ..., \alpha_p^*(k)),$$

$$j = 1, ..., l$$

and $(\alpha_1^*(k), ..., \alpha_p^*(k))$, are the parameters of model for the data $y_1, ..., y_k$. That is,

$$(\alpha_1^*(k), ..., \alpha_p^*(k)) = \arg\min_{\alpha_1, ..., \alpha_p} \sigma^2_{tr}(k)$$

and

$$\frac{1}{k} \sum_{s=1}^{k} (y_s - \Phi(t_s, \alpha_1, ..., \alpha_p))^2 = \sigma^2_{tr}(k)$$

is the mean square of error on the $k$step (year) of training.

Let us take as $y_1, ..., y_n$ $(n = 30)$ the series of data $C_0 + C_1(Y)$ for the Mauna Loa station; then for $\Phi_e(t, \alpha_1, \alpha_2, \alpha_3) = \alpha_1 e^{\alpha_2(t-1957)} + \alpha_3$ we get $\sigma^2_{e,tr}(15) = 0.07$, and for $\Phi_p(t, \alpha_1, \alpha_2, \alpha_p) = \alpha_1(t-1957)^2 + \alpha_2(t-1957) + \alpha_3$ we get $\sigma^2_{p,tr}(15) = 0.08$. At the same time, $F(\Phi_e)(15, 15) = 9.50$ and $F(\Phi_p)(15, 15) = 0.47$.

Thus the parabolic model, even for the first 15 years of observations, estimates the tendency exactly for the next 15 years, while the exponential yields no such results.

Fig. 3 shows a graph of the function $F(k, l)$ for a parabolic family. Here we see that the error of the prediction $F(k, l) = \sigma^2_{pr}(k, l)$ for fixing the number of the year of the training k as a function of the number of the year of prediction l increases slowly. At the same time, fixing the number of the predictions forward, l, $\sigma^2_{pr}(k, l)$ as a function of the number of years of training (k) decreases rapidly. This fact show a good projection ability of the parabolic family for 30 years of observations at the Mauna Loa station.

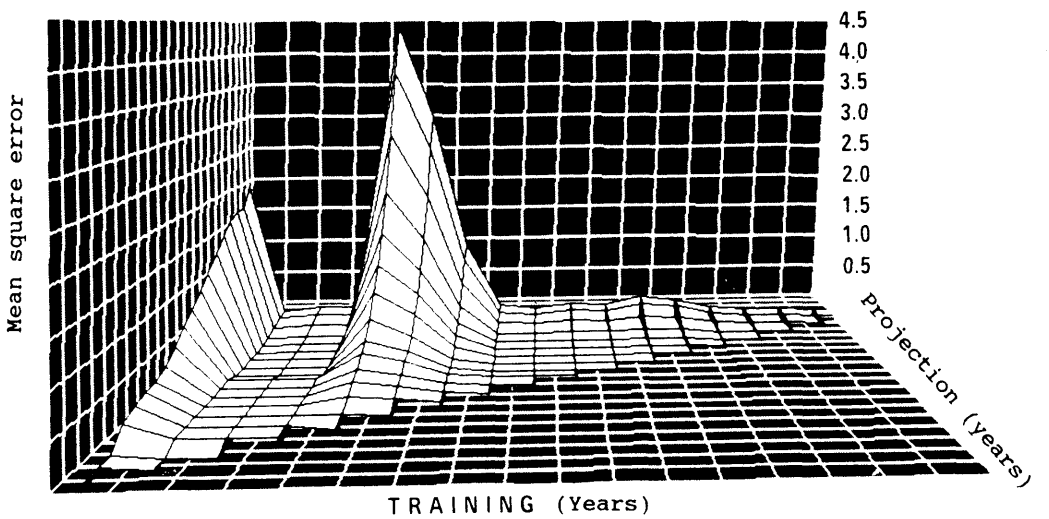

Fig. 3. Function of projection ability of the parabolic models.

5. Analytical expressions of the trends of atmospheric CO_2

The interval on which the different authors construct analytical expressions for the trend of the concentrations of CO_2 in the atmosphere at the present time includes the period from 1750 to 2100. For example in the paper by Crutzen and Brühl (1988) there is a formula

$$C_{CB}(t) = 274 + 6 \exp(0.01965(t - 1860)). \quad (4)$$

We have: $C_{CB}(1860) = 280$, $C_{CB}(1900) = 287$; $C_{CB}(1958) = 315$, $C_{CB}(1980) = 337$; $C_{CB}(2030) = 443$, $C_{CB}(2050) = 525$.

Thus, the value of function (4) is in accordance with evaluations of past concentrations of CO_2 obtained from measurements of the air trapped in bubbles in ice cores, isotopic C_{13} and C_{14} data from analysis of tree rings, data from monitoring stations, and also with some model predictions. Fig. 4 shows the behavior of function (4) relative to parabola (2). Here, it is convenient to introduce the following three intervals. The first interval $[T_0, T_1]$ is from preindustrial times to the beginning of active monitoring. For this interval we have no continuous series of observations, but only evaluations at some points. In the second interval $[T_1, T_2]$, we have a "continuous" series of measurements, where $T_1 \geqslant 1958$, $T_2 \leqslant 1989$. In the interval $[T_1, T_2]$ the precision of the data is high. The third interval $[T_2, T_3]$ is that for which we have the model evaluations of future values. The precision of these evaluations naturally is not high and depends on uncertainties used during the calculations of models of the carbon cycle, on scenarios of "fossil fuel emission," and on land use.

The paper Wuebbles et al. (1984) considers the analytical expressions for concentrations of CO_2 in all three intervals. Moreover, in the interval $[T_2, T_3]$ three expressions are given corresponding to different assumptions for future energy use and economic developments extracted from Edmonds et al. (1985). If in the first two intervals $[T_0, T_1]$ and $[T_1, T_2]$ the exponential functions are used, then in the interval $[T_2, T_3]$ we use the functions from the family of

$$\alpha_1 + \alpha_2(t - T_2) \exp \alpha_3(t - T_2). \quad (5)$$

Wigley (1987) proposed an expression in the interval [1750, 2030] as a parabolic spline. As noted in Section 1 for the construction of the parabola in the interval $[T_2, T_3]$ as a control an estimation of the concentration of CO_2 in the year 2030 of 450 ppmv was used.

Also discussed in Section 1 is the method of construction of the function $W(t)$ in the inter-

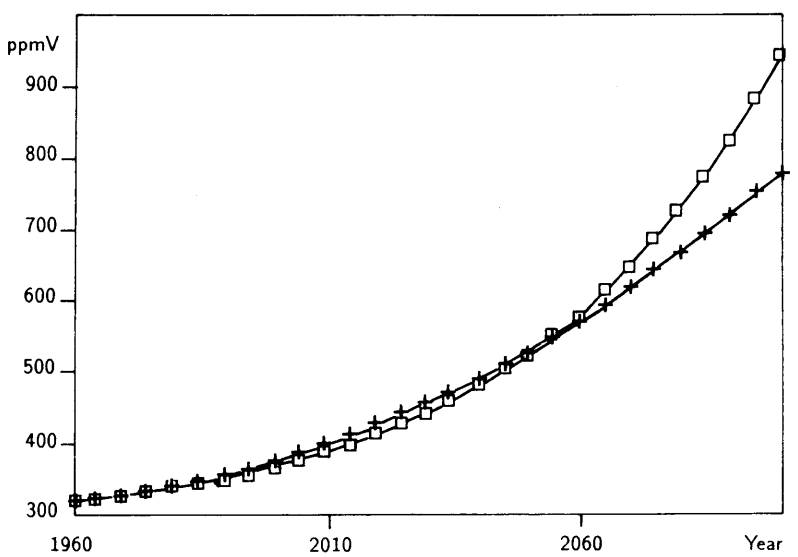

Fig. 4. Behavior of function (4) (□) relative to parabola (2) (+).

val [1958, 2030]. Using this example we shall demonstrate our approach.

Let us denote as P_2^1 the set of all continuous one-time differentiable functions of the form

$$\varphi(t) = \{P_{\varphi,1}(t - T_1),\ T_1 \leqslant t \leqslant T_2;$$
$$P_{\varphi,2}(t - T_2),\ T_2 \leqslant t \leqslant T_3\},$$

where $P_{\varphi,1}(t)$, $P_{\varphi,2}(t)$ are the parabolas for which $P_{\varphi,1}(T_2 - T_1) = P_{\varphi,2}(0)$ and $P'_{\varphi,1}(T_2 - T_1) = P'_{\varphi,2}(0)$.

For $\varphi(t) \in P_2^1$, the parabola $P_{\varphi,1}(t - T_1)$ is the analytical expression for the trend of concentration of atmospheric CO_2 in the interval $[T_1, T_2]$ and $P_{\varphi,2}(t - T_2)$ is the analytical expression for its projection on the interval $[T_2, T_3]$.

Additional conditions, linking $P_{\varphi,1}$ and $P_{\varphi,2}$ could be considered as an example of formalization of an a priori image of extention of the function that describes the monitoring data and the function that describes the forecast.

It is clear that $W(t) \in P_2^1$ is a solution of the following problem:

$$W(t) = \arg\min_{\varphi \in P_2^1} \frac{1}{n} \sum_{q=1}^{n} (y_q - \varphi(t_q))^2 + \lambda(y_* - \varphi(T_3))^2, \quad (6)$$

where $y_1, \ldots, y_n$ are values with a one-year time step used by Wigley for obtaining the analytical expression of the trend in the interval [1958, 1985], $n = 27$, $T_3 = 2030$, and $y_* = 450$ ppmv.

Our main idea is, if the hypotheses about the connection between the observed trend and its projection in the future (considering also the inertia of the global carbon cycle) can be expressed in terms of the functionals of connections $S(P_{\varphi,1}, P_{\varphi,2})$ then it is possible to introduce a general functional, called the functional of risk of projection:

$$F(\varphi(t)) = \frac{\lambda_1}{n_1} \sum_{q=1}^{n_1} (y_q - P_{\varphi,1}(t_q - T_1))^2 + \lambda_2 \frac{n_1}{n_1 + n_2} (y_* - P_{\varphi,2}(T_3 - T_2))^2 + \lambda_3 S(P_{\varphi,1}, P_{\varphi,2}) \cdots, \quad (7)$$

where $n_1 = T_2 - T_1$, $n_2 = T_3 - T_2$, $y_1, \ldots, y_n$ and y_* are the monitoring data and the value of projection in the year T_3. Here λ_1, λ_2, and λ_3 are Lagrange multipliers, penalty coefficients, $\lambda_i \geqslant 0$, $i = 1, 2, 3$.

The functional of connection S reflects a local condition in the transition from monitoring to prediction. For example

$$S(P_{\varphi,1}, P_{\varphi,2}) = (P''_{\varphi,1}(T_2 - T_1) - P''_{\varphi,2}(0))^2 \cdots, \quad (8)$$

and

$$S(P_{\varphi,1}, P_{\varphi,2}) = \frac{1}{T_3 - T_2} \times \int_{T_2}^{T_3} (P_{\varphi,1}(t - T_1) - P_{\varphi,2}(t - T_2))^2 \, dt. \quad (9)$$

Let us remark that in our case when $P_{\varphi,i}$, $(i = 1, 2)$ are parabolas, the difference between the integral characteristic (9) and the local one (8) consists of the integral characteristic taking into account the length of the interval of the projection. Local characteristics are not taken into account.

For functionals of the connection of the form (8) and (9), determination of the function

$$\varphi^* = \arg\min_{\varphi \in P_2^1} F(\varphi(t))$$

is reduced to the problem of linear programming for 4 unknown parameters. It allows one the possibility to determine the function φ^* as a result of an effective computational experiment. The task of the computational experiment is to discover the dependence of the solution from penalty coefficients $\lambda_1, \lambda_2, \lambda_3$. As it directly follows from the shape of the functional (7), these factors have the following interpretation: λ_1 is inversely proportional to the estimation of the preciseness of monitoring data, λ_2 is inversely proportional to the estimation of preciseness of projection, and λ_3 has the sense of the coefficient of confidence to the functional of connectedness S.

Returning to the function $W(t)$, we see that there is not an extremal for functional (8) in the case $\lambda_3 \neq 0$. Given the condition for achieving a projected value 450 ppmv, the projecting parabola is

$$P_{W,2}(t) = 291.7 + 0.0135(t - 1921.6)^2.$$

That essentially differs from

$$P_{W,1}(t - 1958) = C_W(t)$$

by the criteria of connectedness of eqs. (8) and (9).

As noted, the projection of 450 ppmv in 2030 has an interval of uncertainty of [430, 578]. So the value $C(2030) = 456$ given by the parabola $C(t)$ (see eq. (2)) differs from $W(2030) = 450$ by 4% relative to the interval of uncertainty of this projection.

Let us consider the functional of risk of the projection $F(\varphi(t))$ as a function $F(\varphi(t); \lambda_3)$ of the parameter λ_3. Let us stress that the parameter λ_3 expresses *the level of confidence to conserving of the tendency.*

From the discussion above, it follows $F(W(t), 0) \leqslant F(C(t), 0)$ and $F(W(t), \lambda_3) > F(C(t), \lambda_3)$ for $\lambda_3 > \lambda_{3*}$, where λ_{3*} appears small enough. In the section above we showed the following possibilities of functional F for comparison of different projections. Fixing λ_3, the functional F introduces a partial order on the set P_2^1: $\varphi(t) \leqslant \Psi(t)$ if

$$F(\varphi(t); \lambda_3) \geqslant F(\Psi(t), \lambda_3).$$

And as seen from the example of the function $W(t)$ and $C(t)$ the partial order depends on λ_3. In Fig. 5 the parabola is pictured in the background of the 5 curves of the projection corresponding to different scenarios (see Table 2).

For comparison let us give the values of the analytical expressions (see Table 3).

Table 2. *Scenarios of annual CO_2 emissions and atmospheric concentrations (Report of the Expert Group on Emissions Scenarios, IPCC WG-III.): (a) 2030 High Emissions Scenario; (b) 2060 Low Emissions Scenario; (c) Control Policies Scenarios; (d) Accelerated Policies Scenario; (e) Alternative Accelerated Policies Scenario*

Year	Input-output	(a)	(b)	(c)	(d)	(e)
2025	emissions (PgC)	11.5	6.4	6.3	5.1	3.8
	concentrations (ppmv)	437	398	398	393	384
2075	emissions (PgC)	18.7	8.8	5.1	3.0	3.5
	concentrations (ppmv)	679	492	469	413	407

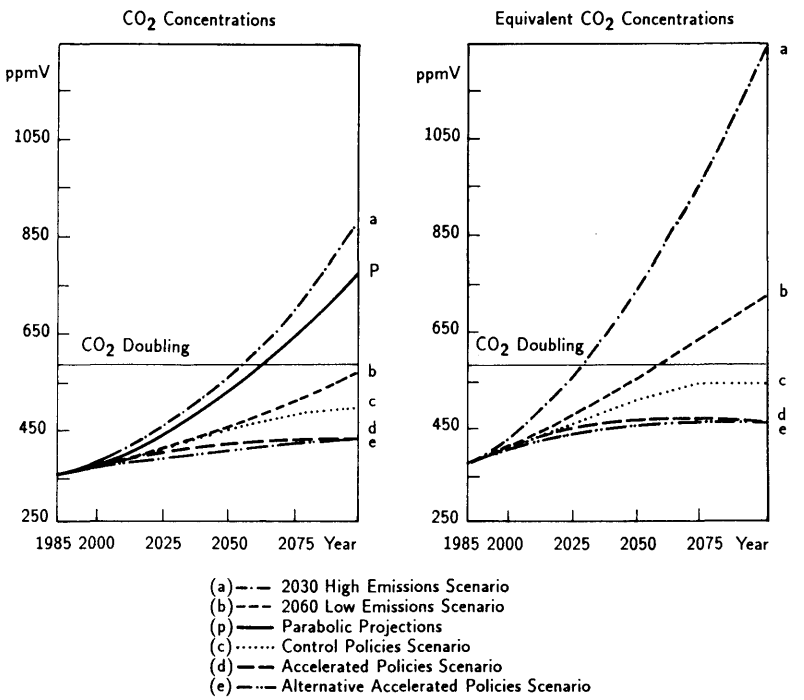

Fig. 5. CO_2 and equivalent concentrations from Report of the Expert IPCC WG-III.

Table 3. *Analytical expression for the trend of atmospheric CO_2 on the interval $[T_2, T_3]$*

P_2	$: 314.596 + 0.5803(t-1957) + 0.0186(t-1957)^2$;	$T_2 = 1989$
W_2	$: 291.7 + 0.0135(t-1921.6)^2$;	$T_2 = 1985$
$C_{2,\mathrm{CB}}$	$: 274 + 6 \exp 0.01965(t-1860)$;	$T_2 = 1988$
$C_{2,\mathrm{DOE}}$	$: 290 + 24.63 \exp 0.02967(t-1958)$;	$T_2 = 1985$
$C^{\mathrm{H}}_{2,\mathrm{WML}}$	$: 341.4 + 1.081(t-1983) \exp 0.02581(t-1983)$;	$T_2 = 1983$
$C_{2,\mathrm{WML}}$	$: 341.4 + 1.539(t-1983) \exp 0.009173(t-1983)$;	$T_2 = 1983$
$C^{\mathrm{L}}_{2,\mathrm{WML}}$	$: 341.4 + 1.82(t-1983) \exp 0.0000834(t-1983)$;	$T_2 = 1983$

Year	Projection	P_2	W_2	$C_{2,\mathrm{CB}}$	$C_{2,\mathrm{DOE}}$	$C^{\mathrm{H}}_{2,\mathrm{WML}}$	$C_{2,\mathrm{WML}}$	$C^{\mathrm{L}}_{2,\mathrm{WML}}$
2025	concentrations (ppmv)	440	436	428	470	476	436	418
2075	concentrations (ppmv)	642	609	684	1083	1410	671	510

Each of these projections can be described by analytical expressions from the class P_2^1 with $T_1 = 1958$, $T_2 = 1985$ and $T_3 = 2100$. It is easy to show that there exists λ_3 such that $F(\varphi, \lambda_3)$ orders these expressions in correspondence with their locations (see Fig. 5). There exists λ_3^* for which functional $F(\varphi, \lambda_3^*)$ takes the smallest value on the parabola $C(t)$.

Thus, it appears possible to correlate the numerical expression to each of the scenarios, and see how the projections differ from that of a corresponding scenario.

In the conclusion, we briefly summarize the main aspects of our approach.

Introductory informations:

(a) interval $[T_1, T_3] = [T_1, T_2] \cup [T_2, T_3]$;

(b) a set $y, ..., y_{n_1}$, $n_1 = T_2 - T_1$ for estimation of the trend from the monitoring data

(c) the set of pairs $\{(t_{*1}, y_{*1}), ..., (t_{*r}, y_{*r})\}$, where $(t_{*l}$ is the year of lth projection and y_{*l} is a model evaluation of concentration in this year).

Chosen assumptions:

(a) a class of functions $\varphi_1(t; \alpha_1, ..., \alpha_p)$ for the construction of the analytical expression of the trend in the interval $[T_1, T_2]$,

(b) a class of the functions $\varphi_2(t; \beta_1, ..., \beta_q)$ for analytical expression of projection.

For example, in the paper by Wuebbles et al., 1984, φ_1 is a selected class of exponential functions, and as φ_2 is a class of functions of the form in eq. (5).

From S, it is possible to choose a functional of the shape of eq. (9), where $P_{\varphi,k}$ is the change by φ_k.

Construction: the class of function Ψ is introduced in the form $\varphi(t) = \varphi(t; \alpha_1, ..., \alpha_p; \beta_1, ..., \beta_q) = \{\varphi_1, (t), T_1 \leqslant t \leqslant T_2; \varphi_2(t), T_2 \leqslant t \leqslant T_3\}$, where $\varphi_1(t)$ and $\varphi_2(t)$ are linked by the conditions in the model of the inertia of the global carbon cycle. For example, the continuity is $\varphi_1(T_2) = \varphi_2(T_2)$, q-times differentiations, $q \geqslant 1$ is $\varphi_1^{(q)}(T_2) = \varphi_2^{(q)}(T_2)$.

The class of functionals of connections $S(\varphi_1, \varphi_2)$ is chosen for modelling the different variants of changing of tendency.

The search for unknown analytical expressions in the interval $[T_1, T_3]$ is realized on the basis of the functional F:

$$F(\varphi(t)) = \frac{\lambda_1}{n_1} \sum_{i=1}^{n_1} (y_i - \varphi_1(t_1))^2 + \left(\frac{n_1}{n_1 + n_2}\right) \frac{\lambda_2}{r} \sum_{j=i}^{r} (y_{*,j} - \varphi_2(t_j))^2 + \lambda_3 S(\varphi_1, \varphi_2).$$

We applied the method of computational experiment, using the interpretation of coefficient of $\lambda_1, \lambda_2, \lambda_3$, described above. Let us stress that the functional $S(\varphi_1, \varphi_2)$ plays the same role as the functional of the type of least action in the problem of the calculus of variation and optimal control.

6. Acknowledgments

The authors wish to express their gratitude to Professors B. R. Döös, Yu. A. Izrael, R. E. Munn, and M. C. MacCracken for their advice and support. Special recognition is due to Dr. W. M. Stigliani for useful discussions and editing. The authors would also like to thank Ms. C. Fuhrmann for the organization and final preparation of this paper.

7. Appendix

Factor decomposition in data tables.
Let $C=(c(i, j), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m)$ be a data table. In the problem of the analysis of the series of mean monthly concentration we have: $m=12$, n, the number of the year of observation, $c(i, j)$, a value of mean monthly concentration of the jth month of the ith year of observation. Without any loss in generality, we can accept that $m \leqslant n$.

In factor analysis there is a hypothesis that nm-matrix C is presented (factorized) in the form:

$$C=VW'+E, \tag{10}$$

where V, nk-matrix, $k<n$, the vector column of which is called the main factors, W, mk-matrix, and E, nm-matrix, of the remainders (Aivazyan et al. 1989), and $'$ is a symbol of transposition.

Let $v_1, ..., v_k$ and $w_1, ..., w_k$ be vector columns of matrices V and W, respectively. Then, eq. (10) can be presented in the form:

$$C=\sum_{l=1}^{k} v_l w_l' + E, \tag{11}$$

where vw' is matrix equal to product of vector column by vector row. It is easy to see that in case $k=2$ and $v_1=(v_1(i))$, $v_1(i)=1$, $1 \leqslant i \leqslant n$; $w_1=(w_1(j))$, $(w_1(j)=1, 1 \leqslant j \leqslant n)$, decomposition (11) for unknown v_1 and w_2 turns into decomposition (1)$'$. So decomposition of the matrix of data by the Tukey method and by the method of factor analysis represents particular cases of a more general method. This method permits decomposition (10) as in the case when some of v_l and w_l have a given shape. This is true also in the case when each of the vectors v_l and w_l is found only from the condition of smallness of the value of the matrix of remainders.

Let us consider a variant of this method which uses the common notion of two-factor decomposition.

Definition. Two-factor decomposition of a table C is its presentation in the form:

$$C=v_0 w_1' + v_1 w_0' + C_1, \tag{12}$$

where $v_k \in R^n$, $w_k \in R^m$, $k=0, 1$ are the vector columns with the coordinates $(v_k(i), 1 \leqslant i \leqslant n; w_k(j), 1 \leqslant j \leqslant m)$, respectively; $C_1=(c_1(i, j); 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m)$ is a table of remainders. Every two-factor decomposition of a given matrix C is defined by vectors v_k, w_k, $k=0, 1$. Different decompositions of matrix C can be used depending on the goals of the analysis. Usually, these goals are formalized as some functional from vectors v_k, w_k, $k=0, 1$. Unknown decomposition corresponds to the vectors v_k^*, w_k^*, $k=0, 1$, that give an extremum to this functional. In some cases, the functional depends only on the table of remainders C_1 and expresses its "smallness" in some sense.

The matrix $C_1=(c_1(i, j))$ may be considered as an mn-dimensional vector with coordinates $c_1(i, j)$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, the indexes of which $\{(i, j)\}$ are ordered lexicographically. A criterion of smallness usually is chosen as a function of a metric in nm-dimensional space. The Minkovskii metric,

$$\rho_\alpha(x, y)=\left(\sum_{l=1}^{L} |x(l)-y(l)|^\alpha\right)^{1/\alpha},$$

is often used, where $\alpha>0$ and $x=(x(l), 1 \leqslant l \leqslant L)$; $y=(y(l), 1 \leqslant l \leqslant L)$.

Let $\|x\|_\alpha=\rho_\alpha(x, 0)$. Then the table of remainders C_1 is estimated by value

$$\|C_1\|_\alpha^\alpha=\sum_{i=1}^{n}\sum_{j=1}^{m} |c_1(i, j)|^\alpha. \tag{13}$$

Let us remark that if $\alpha=2$ (Euclidian metric) for the value of the norm, then there is a useful matrix presentation for computation:

$$\|C_1\|_2^2=\mathrm{Sp}(C_1' C_1),$$

where $\mathrm{Sp}(A)$ is a trace of matrix A, i.e., the sum of its diagonal elements.

Table C_1 could be considered as a sample of random numbers of volume nm.

The criterion of smallness has to be chosen as a function of sample distribution (histogram). Usually an entropy functional of a sample distribution is used (Aivazyan et al. 1989).

Let us consider, first of all, more carefully the two-factor decomposition, connected with a Euclidian metric. In this case we omit the lower index of norm (the 2) and write $\|C_1\|^2$. We have

$$\|C_1\|^2 = \|C - (v_0 w_1' + v_1 w')\|^2$$
$$= F(v_0, v_1, w_0, w_1).$$

Problem 1. Let $v_0 \in R^n$, $w_0 \in R^m$ be given vectors, $\|v_0\| \neq 0$, $\|w_0\| \neq 0$; we have to construct the two-factor decomposition

$$C = v_0 w_1' + v_1 w_0' + C_1,$$

where $v_1 \in R^n$, $w_1 \in R^m$ are vectors giving the minimum to the functional

$$F(v_1, w_1) = \|C_1\|^2.$$

Solving the gradient equation $\nabla_{v_1, w_1} F = 0$, we obtain a solution of *Problem* 1 in the form

$$C = \frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2} v_0 w_0'$$
$$+ v_0 \tilde{w}_1' + \tilde{v}_1 w_0' + C_1, \qquad (14)$$

where

$$\tilde{w}_1 = \frac{C' v_0}{\|v_0\|^2} - \frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2} w_0,$$

$$\tilde{v}_1 = \frac{C w_0}{\|w_0\|^2} - \frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2}.$$

It is easy to show that decomposition (14) represents table C as a sum of four mutual orthogonal vectors.

We have seen that Problem 1 is equivalent to the following important problem.

Problem 1'. Let $v_0 \in R^n$ and $w_0 \in R^m$ be given vectors, $\|v_0\| \neq 0$, $\|w_0\| \neq 0$. The problem is to construct a two-factor expansion

$$C = v_0 w_1' + v_1 w_0' + C_1,$$

where $v_0' C_1 = 0$ and $C_1 w_0 = 0$, i.e., we have to obtain a matrix of remainders C_1, each column of which is orthogonal (noncorrelated) to vector v_0 and simultaneously each row of C_1 is orthogonal to vector w_0.

Application of Tukey's method for smoothing two-way tables (by means) (Emerson and Hoaglin, 1983)

Let us consider Problem 1 in the particular case, when v_0, w_0 are such vectors, that $v_0(i) = 1$ for each $i \leqslant n$ and $w_0(j) = 1$ for each $j \leqslant n$. In this case we have for arbitrary $v \in R^n$ and $w \in R^m$

$$v_0' v = (v_0, v) = n\bar{v}, \qquad w_0' w = m\bar{w},$$

where

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v(i), \qquad \bar{w} = \frac{1}{m} \sum_{j=1}^{m} w(j);$$

mean values of coordinates.

On the basis of this remark, let us integrate the components of expansion (14):

$$\frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} c(i, j);$$

mean value of coordinates of table C. We have:

$$\tilde{w}_1 = \frac{C' v_0}{\|v_0\|^2} - \frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2} w_0.$$

Let us remark that $C' v_0 / \|v_0\|^2$ is a vector column whose ith coordinate is equal to the mean value of the ith column of table C. Further, $(v_0' C w_0)/(\|v_0\|^2 \|w_0\|^2) w_0$ is a vector column whose ith coordinate is equal to the mean value of the elements of matrix C, i.e., to the mean value of coordinates of vector $C' v_0 / \|v_0\|^2$.

Thus, $\tilde{w}_1$ is a vector column, the ith coordinate of which is equal to the mean value of coordinates of the ith "column" of matrix C minus the mean value of means of columns of matrix C. By analogy, v_1' is a vector row, the jth coordinate of which is equal to the mean value of means of rows of matrix C. Hence, expansion (12) in the case of concrete vectors $v_0 = (v_0(i) = 1, 1 \leqslant i \leqslant n)$, $w_0 = (w_0(j) = 1, 1 \leqslant j \leqslant n)$ describes Tukey's algorithm of smoothing of two-way tables (by means).

Geophysical digression. The table year-month is presented in the form (14), where $\tilde{v}_1 = (\tilde{v}_1(i), i = 1, ..., n)$, $\tilde{v}_1(i)$ is a centered mean year observation for the ith year of observation, $\tilde{w}_1 = (\tilde{w}_1(j), j = 1, ..., 12)$ and $\tilde{w}_1(j)$ is a centered mean by "n" mean monthly observation for the jth month of

year. (Let us remember that in statistics the operation of centering is a subtraction from each of the coordinates of a given vector v, the mean value $\bar{v}_m$ of these coordinates: let $v=(v_1, v_2, ..., v_n)$, let $\bar{v}_m$ be the mean value of coordinate $v_1, v_2, ..., v_n$, then $v=(v_1-\bar{v}, v_2-\bar{v}, ..., v_n-\bar{v})$ is the operation of centering of vector v.)

As noted above, we consider a vector $\tilde{v}_1$ to be a nonparametric estimation of a centered trend, and a vector $\tilde{w}_1$, an estimation of annual behavior of the series of observations.

Let us remark that the coordinates of vector $C_1\tilde{w}_1$ are exactly the coefficients of covariation of the rows of the matrix of remainders C_1 with a vector of annual variation. Hence, if $C_1\tilde{w}_1 \neq 0$, then the matrix of remainders C_1 continues to contain information about the annual variation.

By analogy, if $\tilde{v}_1' C_1 \neq 0$, then in the matrix of remainders there is information about the trend. Thus, it is natural to use two-factor decomposition of the matrix of remainders of C_1 by the vectors $\tilde{v}_1$ and $\tilde{w}_1$, i.e., to present the matrix of remainders in the form of (14)

$$C_1 = \frac{\tilde{v}_1' C_1 \tilde{w}_1}{\|v_1\|^2 \|w_1\|^2} \tilde{v}_1 \tilde{w}_1' + \tilde{v}_1 \tilde{w}_2' + \tilde{v}_2 \tilde{w}_1' + C_2.$$

From the properties of expansion (14) of matrix C_1 by vectors $\tilde{v}_1$ and $\tilde{w}_1$, it follows that $v_1' C_1 \tilde{w}_1 = v_1' C \tilde{w}_1$.

In the end, we obtain

$$C = \frac{v_0' C w_0}{\|v_0\|^2 \|w_0\|^2} v_0 w_0' + v_0 \tilde{w}_1' + \tilde{v}_1 w_0' + \frac{\tilde{v}_1' C \tilde{w}_1}{\|\tilde{v}_1\|^2 \|\tilde{w}_1\|^2} \tilde{v}_1 \tilde{w}_1' + \tilde{v}_1 \tilde{w}_2' + \tilde{v}_2 \tilde{w}_1 + C_2. \quad (15)$$

Thus, for a more comprehensive description of the initial matrix C, we introduced the new vectors v_2 and w_2. As a result we transferred from the matrix of remainders C_1 to the matrix of remainders C_2. In some sense matrix C_2 is smaller than matrix C_1, namely, the rows of matrix C_1 are not correlated to a vector of annual variation $\tilde{w}_1$, and the columns are not correlated to a vector of trend $\tilde{v}_1$.

It appears that the vectors v_2 and w_2 have a natural interpretation. Indeed, the ith coordinate $v_2(i)$ of vector v_2 is an amplitude correction to a description of an annual variation by the data of the ith year of observation. Further, a j-coordinate $w_2(j)$ of vector w_2 gives scale correction to the description of the trend by the mean monthly observation in the jth month during n-year of observation. More precisely, let us present a centered vector of mean monthly observation for the ith year (centered i-row $r'(i)-\overline{r(i)}\, w_0'$ of matrix C in form

$$r'(i) - \overline{r(i)}\, w_0' = \tilde{w}_1' + \alpha(i)\, \tilde{w}_1' + D_r'(i),$$

where $D_r(i)$ is a vector of remainders such that $\mathrm{cov}(D_r(i), \tilde{w}_1) = 0$ and $\alpha(i)$ is a correction to the amplitude of interannual variation, which is described by the vector $\tilde{w}_1$. Then a relation is fulfilled

$$\alpha = (\alpha(i), 1 \leqslant i \leqslant n) = \tilde{v}_2 + \frac{\tilde{v}_1' C \tilde{w}_1}{\|\tilde{v}_1\|^2 \|\tilde{w}_1\|^2}$$

(covariation cov of the vectors $D_r(i)$ and w_1 coincides with their scalar product because they are centered).

By analogy, let us present a centered vector of mean monthly observation in the jth month (centered column $C(j)-\overline{C(j)}\, v_0$ of matrix C) in the form:

$$c(j) - \overline{c(j)}\, v_0 = \tilde{v}_1 + \beta(j)\, \tilde{v}_1 + D_c(j)$$

where $\mathrm{cov}(D_c(j), \tilde{v}_1) = 0$ and $\beta(j)$ is a scale correction to a description of a trend by the vector $\tilde{v}_1$. We have

$$\beta = (\beta(j), 1 \leqslant j \leqslant 12) = \tilde{w}_2 + \frac{\tilde{v}_1' C \tilde{w}_1}{\|\tilde{v}\|^2 \|\tilde{w}_1\|^2}.$$

Two-factor expansion of the matrices, connected with the Minkovskii metrics. Among the Minkovskii metrics $\rho_\alpha(x, y)$, together with the Euclidian metric $\rho_2(x, y)$, a special case is the metric $\rho_1(x, y)$. The step of Tukey's algorithm of smoothing of matrix by medians (see Section 2) is based on the fact that for any number sequence $a_1, ..., a_N$ a minimum of the function

$$\varphi(a) = \frac{1}{N} \sum_{l=1}^{N} |a_l - a|$$

is reached when a is equal to the median of this sequence.

In conclusion, we give a short scheme of algorithm for the metrics $\rho_\alpha(x, y)$, $\alpha > 0$.

Let us remark first of all that the solution of the problem in the form of (14) is easily extended to a case of general Euclidian metric with a scalar product

$$(x, y) = x'Ay,$$

where A is a fixed, positively defined symmetrical matrix. This permits, with the aid of standard techniques, construction of an iterational algorithm for solving the problem of two-factor expansion connected with the Minkovskii metric $\rho_\alpha(x, y)$, $\alpha > 0$. Indeed, let us remark that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (\delta^2 + |c_1(i, j)|)^{\alpha-2} (c_1(i, j))^2 = \mathrm{Sp}(C_1' A C_1) \xrightarrow{\delta \to 0} \|C_1\|_\alpha^\alpha,$$

where $A = A(C_1)$ is a diagonal nm-symmetrical matrix with diagonal elements $(\delta^2 + |c_1(i, j)|)^{\alpha-2}$.

Algorithm

1. As a first approach we obtain the solution of Problem 1 in the shape of (12). We obtain a matrix $C_1 = C_{11}$ and corresponding to it the matrix $A = A(C_{11})$.
2. As a second approach we obtain the solution of the problem of two-factor decomposition for Euclidian metric with a matrix $A(C_{11})$. We obtain a matrix of remainders $C_1(A(C_{11})) = C_{12}$ and corresponding to it a matrix $A(C_{12})$.

By iteration we obtain the next approximation. To stop the algorithm, one may use the same criteria as for solution of minimization function problems.

The SABL Method. As we stressed (see also Box and Jenkins (1976) chapter 9), the condition of seasonality of time series permits one to roll it up into a $m \times n$ table, where m is the number of the value of seasons and n is the number of seasons (in the case of monthly average concentrations, $m = 12$). By this it is based on hypothesis: the rows of the table, or the columns of the table or simultaneously both have the characteristics of a time series, i.e., the methods of analysis of such series are applied to them.

The methods of numerical filtration of the time series, or smoothers, are based on the SABL procedure, Cleveland et al. (1983).

In the mentioned paper by Cleveland et al. (1983), the SABL method is applied to the analysis of monthly average of concentration of series of CO_2. For the convenience of accounting let us consider a summary of this method. Let $x(1), ..., x(12n)$ be a series of monthly average concentrations. The problem is to present each value of series in the form $x(m) = t(m) + s(m) + i(m)$, $m = 1, ..., 12n$ where $t(m)$ is a trend of series, $s(m)$ is a seasonal component, $i(m)$ is a series of remainders.

The solution of this problem as in Tukey's method are given as an iterative procedure. The initial estimation of $t_0(m)$ of the trend $t(m)$ is the result of smoothing of the whole series with the goal of picking out the low frequency component; the series of remainders $x(m) - t_0(m)$ is rolled up into the $12 \times n$-table, and to each of the 12 vector columns of the length n the smoother is used. The vector columns obtained are $12 \times n$ table, that unrolls into the series of the length $12n$. This series is called the initial estimation of the seasonal component of the initial series. Let on the lth step the lth estimations $s_l(m)$ of the season component be constructed. Then the $l+1$th estimation $t_{l+1}(m)$ of the trend $t(m)$ is a result of smoothing of the series $x(m) - s_l(m)$. The $l+1$th estimation $S_{e+1}(m)$ of a seasonal component $s(m)$ is a $12 \times n$ table composed from the columns, each of which is obtained by smoothing of the corresponding column of the table for seasonal series $x(m) - t_{l+1}(m)$. The iterational procedure is stopped when the estimations $t_l(m)$ and $t_{l+1}(m)$ and estimations $s_l(m)$ and $s_{l+1}(m)$, respectively, are indistinguishable by some criteria.

From the given description of the SABL method, it follows that Tukey's method would be considered as some variant of the SABL method for a special choice of smoothers.

In summary, it is shown that Tukey's well-known method could be included in the family of methods of factor decomposition of data tables. This method, applied to the special tables, corresponding to the seasonal time series, is included in the family of SABL methods. This fact allows for a method of exploratory analysis that unifies the results of factor analysis and SABL. Let us stress that the essential advantage of factor analysis in interpretation of the results is that they are realized via an optimization procedure. The advantage of the SABL method is the possibility of its realization in computational experiments.

REFERENCES

Aivazyan, S. A., Buchstaber, V. M., Yenyukov, I. S. and Meshalkin, L. D. 1989. *Classification and reduction of dimensionality.* Finansy i statistika, Moscow.

Antonovsky, M. Ya., Buchstaber, V. M. and Zubenko, A. A. 1988. *Statistical analysis of long-term trends in atmospheric CO_2 concentrations at baseline stations.* WP-88-122. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Antonovsky, M. Ya., Buchstaber, V. M. and Zubenko, A. A. 1989. Exploratory analysis of trends in the series of observations of concentrations of CO_2 in atmosphere. In: *Problems of ecological monitoring and ecosystem modeling*, Volume XII, Hidrometeoizdat, Leningrad.

Baes, C. F., Jr. and Killough, G. G. 1985. *A two-dimensional CO_2-ocean model including the biological processes*, TR021, DOE/NBB-0070.

Björkström, A. 1979. A model of CO_2 interaction between atmosphere, oceans, and land biota. In: *The global carbon cycle* (eds. B. Bolin, E. T. Degens, S. Kempe, and P. Ketner). SCOPE 13. John Wiley & Sons, Chichester, New York, 403–457.

Bolin, B. 1986. How much CO_2 will remain in the atmosphere? The carbon cycle and projections for the future. In: *The greenhouse effect, climatic change, and ecosystems* (eds. B. Bolin, B. R. Döös, J. Jäger, and R. A. Warrick). SCOPE 29. John Wiley & Sons, Chichester, New York, 93–155.

Box, G. E. P. and Jenkins, G. M. 1976. *Times series analysis: forecasting and control.* Holden-Day, San Francisco.

Budiko, M. I. and Izrael, Yu. A. (eds.). 1987. *Anthropogenic changes of climate.* Hidrometizdat, Leningrad.

Cleveland, W. S., Freeny, A. E. and Graedl, T. E. 1983. The seasonal component of atmospheric CO_2: information from new approaches to the decomposition of seasonal time series. *J. Geophysical Res. 88, C15,* 10934–10946.

Crutzen, P. and Brühl, C. 1988. *Carbon dioxide and other greenhouse gases: climatic and associated impacts* (eds. R. Fantechi and Ghazi). 159–167.

Edmonds, J. A., Reilly, J., Trabalka, J. R. and Reichle, D. E. 1985. *An analysis of possible future atmospheric retention of fossil fuel CO_2.* DOE/OR/21400-1. US Department of Energy, Washington, D.C.

Emanuel, W. R., Killough, G. G., Post, W. M., Shugart, H. H. and Stevenson, M. P. 1984. *Computer implementation of a globally averaged model of the world carbon cycle.* TR010. Carbon Dioxide Research Division, US Department of Energy, Washington, D.C.

Emerson, J. D. and Hoaglin, D. C. 1983. Analysis of two-way tables by medians. *Understanding robust and exploratory data analysis* (eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey), John Wiley & Sons, New York, 166–209.

Goudriaan, J. and Ketner, P. 1984. A simulation study for the global carbon cycle including man's impact on the biosphere. *Climatic Change 6,* 167–192.

Keeling, C. D. 1987. *Hourly calibration atmospheric CO_2 concentration 1958–1986. Mauna Loa observatory.* CDIAC-NDP-043. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heiman, M., Mook, W. G. and Roeloffzen, H. 1989. A three-dimensional model of atmospheric CO_2 transport based on observed winds: 1. Analysis of observational data. Pages 165–236. In: *Aspects of climate variability in the Pacific and the western Americas* (ed. D. H. Peterson). Monograph No. 55 (December 1989). American Geophysical Union, Washington, DC.

Ramanathan, V., Cicerone, R. J., Singh, H. B. and Kiehl, J. T. 1985. Trace gas trends and their potential role in climate change. *J. Geophys. Res. 90,* 5547–5566.

Rotty, R. M. and Reister, D. B. 1986. Use of energy scenarios in addressing the CO_2 question. *J. Air. Pollut. Control. Assoc. 36(10),* 1111–1115.

Siegenthaler, U. 1983. Uptake of excess CO_2 by an outcrop-diffusion model of the ocean. *J. Geophysical Res. 88(C6),* 3599–3608.

Tans, Pioter P., Fung, I. Y. and Takahashi, T. 1990. Observational Constraints on the Global Atmospheric CO_2 Budget. *Science 247,* 1431–1438.

Trabalka, J. R. (ed.). 1985. *Atmospheric carbon dioxide and the global carbon cycle.* DOE/ER-0239. US Department of Energy, Washington, DC.

Wigley, T. M. L. 1987. Relative contributions of different trace gases to the greenhouse effect. *Climate Monitoring 16,* 14–28.

Wuebbles, D. J. 1981. *Scenarios for future anthropogenic emissions of trace gases in the atmosphere.* UCID-18997. Lawrence Livermore National Lab., Berkeley, CA, USA.

Wuebbles, D. J., MacCracken, M. C. and Luther, F. M. 1984. *A proposed reference set of scenarios for radiatively active atmospheric constituents.* TR015. Carbon Dioxide Research Division, US Department of Energy, Washington, DC.