

Stochastic Simulation of Atmospheric Trajectories

MITCHELL J. SMALL

Civil Engineering/Engineering and Public Policy, Carnegie-Mellon University, Pittsburgh, PA 15213

PERRY J. SAMSON

Atmospheric and Oceanic Sciences, University of Michigan, Ann Arbor, 48109

(Manuscript received 18 February 1982, in final form 2 November 1982)

ABSTRACT

Methods are presented for generating an ensemble of synthetic atmospheric trajectories. These include methods for a set of independent trajectories, and methods for a correlated set of sequential trajectories. The models incorporate first-order autocorrelation of successive zonal and meridional displacements, and in the case of the sequential set, correlations between trajectories due to spatial patterns of air flow.

The model for independent trajectories is applied and validated using three years (1976–78) of observed forward trajectories originating from southeast Michigan. The synthetic trajectories are statistically consistent with observed wind fields, and result in large-scale diffusion which is comparable to existing estimates. Limitations in the assumption of spatial homogeneity of wind statistics are explored using backward trajectories arriving at six locations in eastern North America during July 1978. The applications highlight the problems which may arise when observed trajectories are used in atmospheric transport modeling, and support the use of synthetic trajectories, at least in a complementary analysis, to obtain a more accurate long-term representation. The procedure for correlated trajectories is not implemented; however, it may prove useful for transport problems of shorter time scales where wind persistence is important.

1. Introduction

Stochastic simulation has been applied in a wide range of meteorological analyses involving geopotential, temperature, precipitation and wind field synthesis. In this paper, methods for generating synthetic mixed-layer air trajectories, primarily for use in long-range atmospheric transport modeling, are presented. Separate techniques are developed for simulating a series of independent trajectories and a sequential series of correlated trajectories. The generation of synthetic atmospheric trajectories is particularly motivated by difficulties with observed trajectories such as gaps and bias introduced by missing data, sparsely located upper air stations, and trajectory termination over the sea.

A number of statistical representations of large-scale air motion have been developed to provide a basis for stochastic simulation. Pratte and Lee (1979) present a digital filtering procedure to simulate two-dimensional wind fields with a specified level of spatial correlation. Bacon *et al.* (1981) present a Markov chain model for representing the downwind air motion from an emissions source. The probability of a given wind speed and direction is dependent on the wind speed and direction at the previous time step. The model presented in this study is similar in purpose, but is based on a two-dimensional, first-order

autoregressive representation of trajectory displacements downwind from a source, or upwind from a receptor location.

Beginning with the early work of Taylor (1921), interest has focused on the correlative structure of wind fields and its relationship to diffusion processes (see, e.g., Gifford, 1968; Kao, 1974; and Pasquill, 1974). Durst *et al.* (1959) applied these concepts to large-scale, long-term diffusion; that is, the horizontal diffusion which occurs as a result of variations in trajectory paths over months, seasons and years. Later work by Murgatroyd (1969) and applications of this approach in long-range air pollution models by Bolin and Persson (1975), Sheih (1977), Fay and Rosenzweig (1980), and Shannon (1981) have provided additional insight into the nature of long-term diffusion.

A fundamental characterization used in the analysis of atmospheric diffusion is the autocorrelation coefficient of component wind speeds along a trajectory, $R(\tau)$. While many formulations for $R(\tau)$ have been analyzed both theoretically and empirically, among the simplest and most basic is the exponential form:

$$R(\tau) = e^{-\alpha\tau}, \quad (1)$$

where the term $(1/\alpha)$ is referred to as the time scale of characteristic eddies. The trajectory generating scheme proposed in this paper is consistent with Eq.

(1) and a simple relationship between α and model parameters is given.

The mean and standard deviation of trajectory locations as a function of downwind travel time are commonly used to characterize transport climatology and long-term diffusion. These are computed for the generating models and shown to be consistent with existing estimates.

The generating method presented in the first part of the paper produces a set of independent trajectories, with no sequencing or temporal interrelationships. A method for generating a correlated series of trajectories, leaving or arriving at a source location at fixed time intervals, is presented in the final sections of the paper.

2. Time series model for independent trajectories

The forward movement of an air parcel may be described by a series of (x, y) coordinate displacements away from an initial generation point. Backward trajectories are similarly described, except the displacements move backward in time. Defining the time history of the location of an air trajectory as the vector $P_t = (X_t, Y_t)$, the successive displacements (for forward trajectories) are

$$\Delta X_t = X_{t+1} - X_t, \tag{2}$$

$$\Delta Y_t = Y_{t+1} - Y_t. \tag{3}$$

These displacements may be standardized to define

$$x_t = (\Delta X_t - \overline{\Delta X}) / \sigma_{\Delta X}, \tag{4}$$

$$y_t = (\Delta Y_t - \overline{\Delta Y}) / \sigma_{\Delta Y}. \tag{5}$$

where $\overline{\Delta X}$, $\sigma_{\Delta X}$, $\overline{\Delta Y}$ and $\sigma_{\Delta Y}$ are the respective means and standard deviations.

The proposed generation scheme uses a simple first-order autoregressive equation. Assuming the x and y displacements are uncorrelated, they are generated by

$$x_t = \alpha_x x_{t-1} + \epsilon_{x,t}, \tag{6}$$

$$y_t = \alpha_y y_{t-1} + \epsilon_{y,t}, \tag{7}$$

where α_x and α_y are the first-order autoregression coefficients and $\epsilon_{x,t}$ and $\epsilon_{y,t}$ are serially uncorrelated normal random variates with mean zero and variances equal to $1 - \alpha_x^2$ and $1 - \alpha_y^2$ respectively. If significant correlation exists between the x and y displacements, this may be approximated by generating $\epsilon_{x,t}$ and $\epsilon_{y,t}$ simultaneously with

$$E[\epsilon_{x,t}\epsilon_{y,t}] = \rho(1 - \alpha_x\alpha_y), \tag{8}$$

where E denotes the expected value and ρ is the correlation coefficient between x_t and y_t . In most cases ρ should be small enough so that the use of this procedure is not warranted.

The generating equations (2)–(7) are formally described as a two-dimensional, first-order, integrated autoregression model, or an ARIMA (1, 1, 0) model (Box and Jenkins, 1976) for trajectory location. The first argument, corresponding to the AR component of the ARIMA, indicates a first-order autoregression. The second argument, corresponding to the I component, indicates one level of integration (i.e., differences in location are modeled rather than the locations themselves). The third argument, corresponding to the MA component, indicates that there is no moving average term in the model.

The model formulation assumes stationarity, that is, for a given location (X, Y) ; $\overline{\Delta X}$, $\sigma_{\Delta X}$, $\overline{\Delta Y}$, and $\sigma_{\Delta Y}$ are constant over time. Spatial inhomogeneity, however, may still be incorporated with the means and standard deviations a function of P_t . Temporal stationarity may be more rigorously ensured in practice by separately analyzing particular months or seasons.

In the proposed application P_t is defined at three-hour time intervals, so x_t and y_t are the standardized three-h displacements. The initial displacements, x_0 and y_0 , are sampled from a normal distribution with mean zero and unit variance, and the resulting x_t and y_t are also $N(0, 1)$. Successive trajectories are independent.

Eqs. (6) and (7) are in a form consistent with the generation of forward trajectories. The fact that the same α_x and α_y may be used in the generation of backward trajectories is made evident by noting that α_x and α_y are estimated by the correlation coefficients between x_t and x_{t-1} and y_t and y_{t-1} respectively.

In the application presented in this paper, $\overline{\Delta X}$, $\sigma_{\Delta X}$, $\overline{\Delta Y}$ and $\sigma_{\Delta Y}$ are assumed to be spatially homogeneous. The generating equations then simplify to

$$\Delta X_t = \alpha_x \Delta X_{t-1} + \overline{\Delta X}(1 - \alpha_x) + \sigma_{\Delta X} \epsilon_{x,t}, \tag{9}$$

$$\Delta Y_t = \alpha_y \Delta Y_{t-1} + \overline{\Delta Y}(1 - \alpha_y) + \sigma_{\Delta Y} \epsilon_{y,t}, \tag{10}$$

and the initial displacements are calculated by

$$\Delta X_0 = \overline{\Delta X} + \eta_x, \tag{11}$$

$$\Delta Y_0 = \overline{\Delta Y} + \eta_y, \tag{12}$$

where η_x is sampled from a $N(0, \sigma_{\Delta X}^2)$ distribution and η_y is $N(0, \sigma_{\Delta Y}^2)$.

a. Implications for long-term diffusion

The long-term diffusion characteristics of the spatially homogeneous model [Eqs. (9)–(12)] are now derived. The autocorrelation functions of component displacements (equivalent to those for the components of wind speed) are

$$R_x(t) = \alpha_x^t, \tag{13}$$

$$R_y(t) = \alpha_y^t \tag{14}$$

Note that t measures discrete time in units of three hours, so for example, the correlation between a zonal displacement along a trajectory and another 12 h downwind is α_x^4 . Comparing Eqs. (13) and (14) to Eq. (1) (note that τ in Eq. (1) measures continuous time), α_x and α_y may be estimated from the characteristic eddy correlation time by

$$\alpha_x \text{ or } \alpha_y = e^{-3\alpha} \tag{15}$$

This relationship is shown in Fig. 1. The characteristic eddy time is estimated by Durst *et al.* (1959) to be ~24 h, by Murgatroyd (1969) as ~33 h, and by Bolin and Persson (1975) to be on the order of 40–60 h. Fig. 1 indicates that α_x and α_y should thus lie in the range of 0.85–0.95. Results presented in the next section confirm this expectation.

Long-term horizontal diffusion is characterized by the standard deviation of the location of trajectory endpoints (σ) as a function of downwind travel time (τ). The component standard deviations (σ_x and σ_y) for the time series model (Eqs. 9–12) are computed by noting that the zonal location of the trajectory after n displacements ($3n$ h downwind) is

$$X_n = \eta_x \sum_{j=0}^{n-1} (\alpha_x^j) + \sigma_{\Delta X} \sum_{j=0}^{n-2} \epsilon_{n-j} \left(\sum_{i=0}^j \alpha_x^i \right) + n\overline{\Delta X}, \tag{16}$$

with expected value

$$\overline{X_n} = n\overline{\Delta X}, \tag{17}$$

and standard deviation

$$\sigma_X = \sigma_{\Delta X} \left[\left(\sum_{j=0}^{n-1} \alpha_x^j \right)^2 + (1 - \alpha_x^2) \sum_{j=0}^{n-2} \left(\sum_{i=0}^j \alpha_x^i \right)^2 \right]^{1/2}. \tag{18}$$

Equivalent expressions apply to Y_n , $\overline{Y_n}$ and σ_y . As X_n and Y_n are sums of normal random variables, they are normally distributed. Eq. (18) is plotted in Fig. 2 for a representative range of α_x and α_y . A nearly linear relationship between the standard deviation of

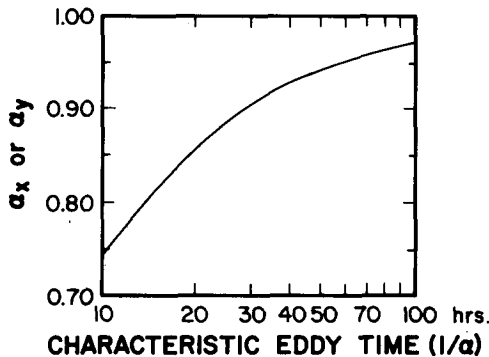


FIG. 1. Relationship between characteristic eddy time and first-order (three-h) autoregression coefficients.

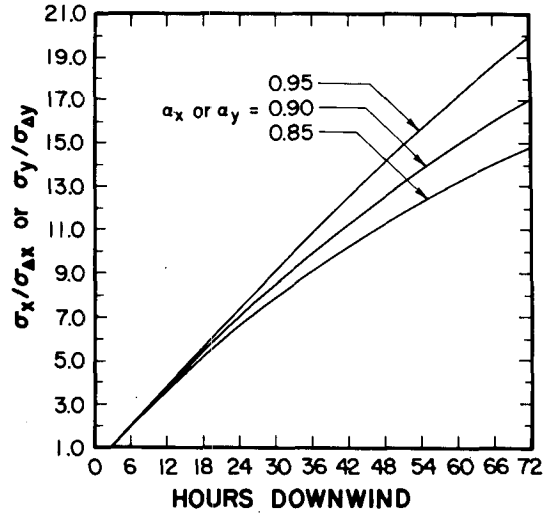


FIG. 2. Determination of long-term diffusion.

trajectory endpoints and travel time is obtained initially, but the rate of increase is less than linear further downwind. Eq. 18 approaches complete linearity as α_x approaches one, and approaches a quadratic form for low values of α_x (i.e., a random walk).

3. Test of model for independent trajectories

To demonstrate the application of the proposed generating method, the likely range of parameters, and its ability to produce a representative set of trajectories, the model was used to generate synthetic forward trajectories from a location in southeast Michigan. Model parameters were estimated from a three-year set (1976–78) of observed mixed-layer trajectories calculated using the method of Heffter (1980). The Heffter method applies spatial and temporal interpolation to rawinsonde and pibal data collected every 12-h at upper air stations in the study area. The interpolated trajectories leave the origin every six-h and the location is computed every three-h for 72-h downwind, or until the trajectory is terminated due to unavailable or missing rawinsonde and pibal data. A number of trajectories in the data set are missing or incomplete, but in terms of the number of individual trajectory displacements included (more than 65 000)¹ the calculated trajectories still provide a substantial basis for analysis.

a. Estimation of model parameters

To estimate the model parameters individual 3-h displacements were calculated in the zonal and me-

¹ Note that due to the interpolation used in the Heffter method, many of the displacements used for parameter estimation are not independent. The effective number of independent estimates used in the analysis is thus actually less than 65 000.

TABLE 1. Summary of forward trajectory statistics and correlations (originating from southeast Michigan, 1976–78).

| | Number of observations | $\overline{\Delta X}$ (km) | $\sigma_{\Delta X}$ (km) | $\overline{\Delta Y}$ (km) | $\sigma_{\Delta Y}$ (km) | α_x | α_y | ρ |
|---------------------------------|------------------------|----------------------------|--------------------------|----------------------------|--------------------------|------------|------------|--------|
| Total period | 65 521 | 44.1 | 60.7 | 3.5 | 52.6 | 0.92 | 0.89 | 0.019 |
| Winter (21 December–20 March) | 14 680 | 57.4 | 65.5 | -1.4 | 57.5 | 0.90 | 0.88 | -0.037 |
| Spring (21 March–20 June) | 15 998 | 35.7 | 63.3 | -0.83 | 54.5 | 0.92 | 0.89 | 0.020 |
| Summer (21 June–20 September) | 18 977 | 43.0 | 47.5 | 7.5 | 41.3 | 0.91 | 0.89 | 0.160 |
| Fall (21 September–20 December) | 15 866 | 41.7 | 65.2 | 7.4 | 57.0 | 0.92 | 0.91 | -0.014 |

ridional directions (ΔX and ΔY , positive toward the east and north, respectively). Their means and standard deviations were determined, and α_x and α_y estimated as the correlation coefficients between ΔX_i and ΔX_{i-1} and ΔY_i and ΔY_{i-1} , respectively. The results are summarized in Table 1 along with ρ , the correlation coefficient between zonal and meridional displacements.

The results indicate a mean wind vector directed toward the east with an average mixed-layer transport velocity on the order of 15 km h^{-1} (4.2 m s^{-1}). Small seasonal differences are apparent as the mean wind during the summer and fall are directed east by northeast, mean winter winds are displaced more directly toward the east, and mean spring winds are less displaced toward the east. Note that the individual downwind displacements occur throughout the study area, and the statistics in Table 1 are thus partially indicative of a spatial average over the eastern United States and Canada. A bias is introduced, however, by the fact that the trajectories originate from a single location (southeast Michigan). The issue of the spatial range of applicability of estimated parameters is addressed later in the section on spatial inhomogeneity.

The standard deviation of 3-h zonal displacements is $\sim 60 \text{ km}$, while the meridional standard deviation is $\sim 55 \text{ km}$. These values are each $\sim 20\text{--}25\%$ lower during the summer than during the other seasons, indicating a correspondingly lower amount of long-term horizontal dispersion in the summer. The estimated values of α_x and α_y are 0.92 and 0.89, respectively, with negligible seasonal variation. The values of ρ indicate that correlations between ΔX and ΔY are negligible, with the exception of a small, positive correlation during the summer months ($\rho = 0.16$). This suggests that the assumption $\rho = 0$ is reasonable for the simple, first-order generation method used in this analysis.

b. Model validation

The statistical accuracy of the model applied to southeastern Michigan is evaluated by comparing predicted and observed trajectory distributions in terms of their mean locations and standard deviations. Prior to this, it is useful to check an underlying assumption of the model; that individual trajectory

displacements are Gaussian. The histograms of zonal and meridional displacements during the total three year period are shown in Fig. 3. The normal distributions, fitted using the observed means and standard deviations, appear to provide a good approximation to the observed histograms.

Comparisons between the means and standard deviations of observed and predicted zonal and meridional trajectory locations are shown in Figs. 4 and 5. The limiting mean and standard deviation of model generated trajectories (solid line in Figs. 4 and 5) are calculated from Eqs. (17) and (18) (or Fig. 2) using

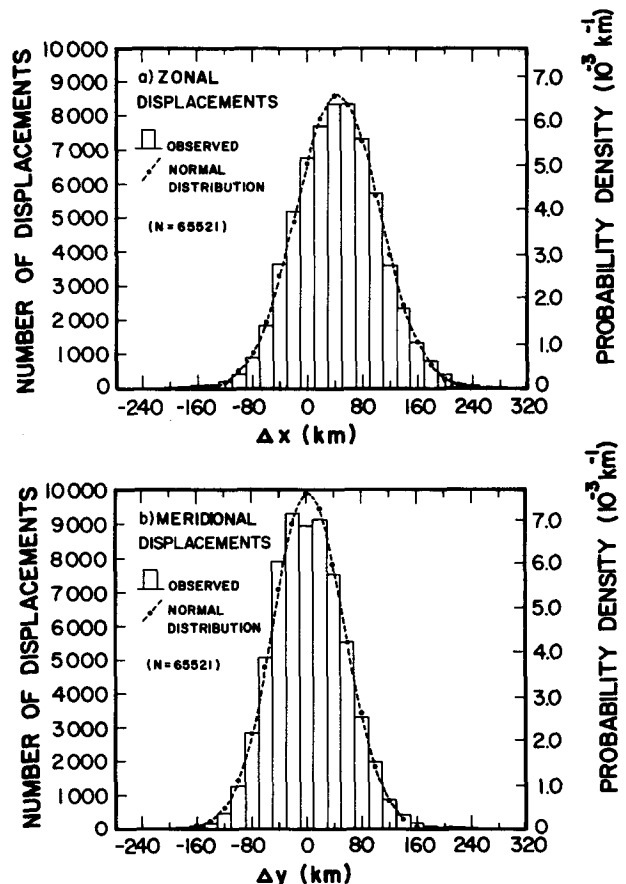


FIG. 3. Frequency distribution of 3-h displacements along trajectories originating from southeast Michigan (1976–78).

the parameters in the first row of Table 1. The limiting prediction is shown to be nearly identical to the observed distribution (dashed line) for the first 30 h downwind, however, after 30 h there is a significant divergence, as the observed σ_x and σ_y level off, while the predicted σ_x and σ_y continue to increase. This is accompanied by a similar divergence in the observed and predicted values of \bar{X} . No clear pattern is apparent in the estimates of \bar{Y} (note that the scale in Fig. 5a is blown up by a factor of 4 relative to Fig. 4a, and that the differences in the observed and predicted \bar{Y} are much smaller than the differences in \bar{X}).

The differences between the observed and predicted, limiting behavior of trajectory distributions in Figs. 4 and 5 are for the most part artificial, due primarily to the premature termination (over the sea) of trajectories computed from observed winds. Observed trajectories with a strong component toward the east begin passing over the Atlantic Ocean in significant numbers 30 h downwind and thereafter. These trajectories are terminated due to the absence of upper air stations over the sea, and their large, positive contributions to \bar{X} , σ_x and σ_y are lost. To

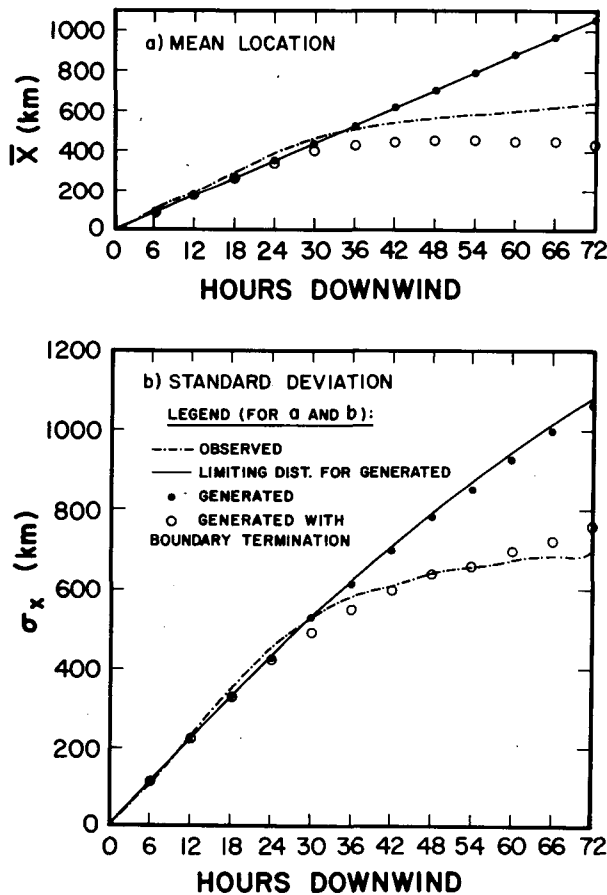


FIG. 4. Zonal endpoint distribution of trajectories originating from southeast Michigan (1976-78).

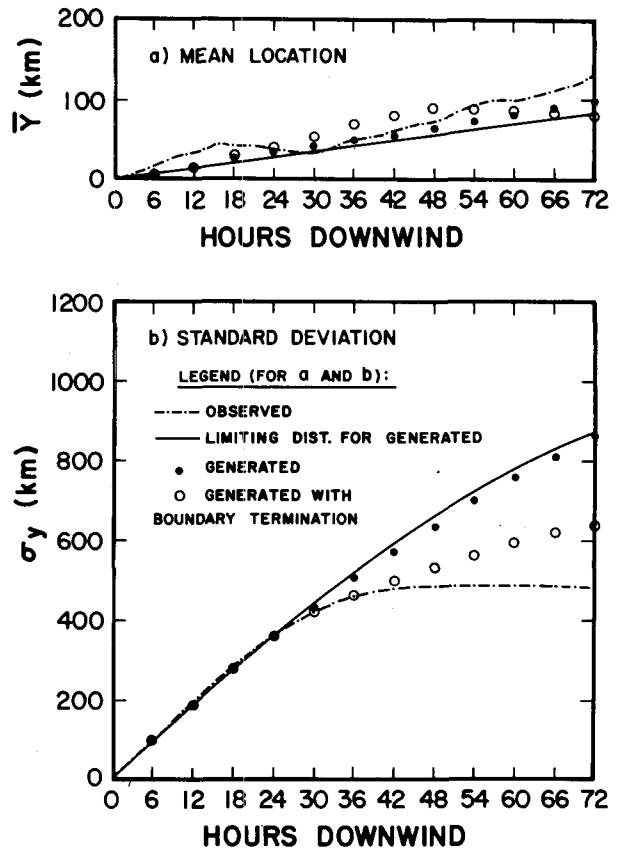


FIG. 5. Meridional endpoint distribution of trajectories originating from southeast Michigan (1976-78).

highlight this effect, Figs. 6-8 show the annual limiting, bivariate normal distribution predicted by the model 24, 48 and 72 h downwind. The contours indicate the fraction of trajectory endpoints during the year predicted to arrive in the given area per square kilometer ($\times 10^8$). A large portion of the probability is predicted to be over the Atlantic ocean at 48 and 72 h.

To test the assertion that the differences in the modeled and observed behavior in Figs. 4 and 5 are due to trajectory termination, 5000 synthetic trajectories were generated for southeastern Michigan using Eqs. (9)-(12). The mean and standard deviation of these generated trajectory locations (solid circles in Figs. 4 and 5) are very close to the predicted, limiting behavior, with some small differences due to the statistical variability associated with a finite sample size. The generated trajectories were then modified, terminating all trajectories which pass over the sea or into northern Canada. The boundaries for termination of generated trajectories are shown in Fig. 9. The modified distribution (Figs. 4 and 5, open circles) is now very similar to the observed distribution. Differences which remain may be due to the imprecise selection of termination boundaries or the premature

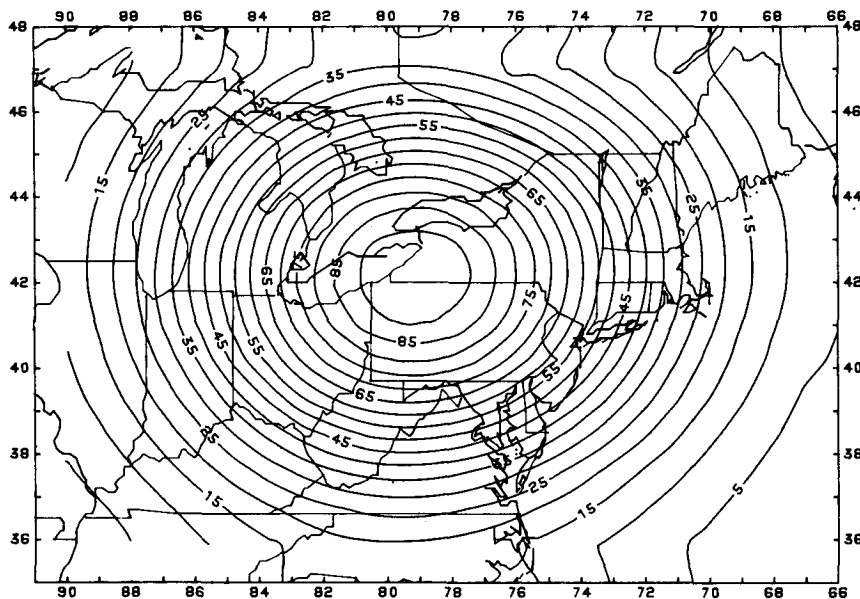


FIG. 6. Limiting endpoint distribution of modeled trajectories 24 h downwind from southeast Michigan during 1976-78 ($10^8 \times$ fraction km^{-2}).

termination of observed trajectories within the boundary area due to missing data. This is supported by Fig. 10 which compares the observed and simulated fractions of trajectories terminated. Termination due to incomplete data in overland areas does in fact occur in the trajectory data set.

Some of the differences noted in Figs. 4 and 5 may also be due to an incomplete representation of the statistical properties of atmospheric trajectories by the

simplified generating model. A possible modification could involve the use of a higher order ARIMA equation. Preliminary estimation with ARIMA (2, 1, 0) model indicates small, but significant coefficients for the lag-two (six-h lag) term. Given the overall uncertainties in model analysis and estimation, however, it is not felt that the small improvement in accuracy achieved with a lag-two model warrants the additional complexity. A better representation may also

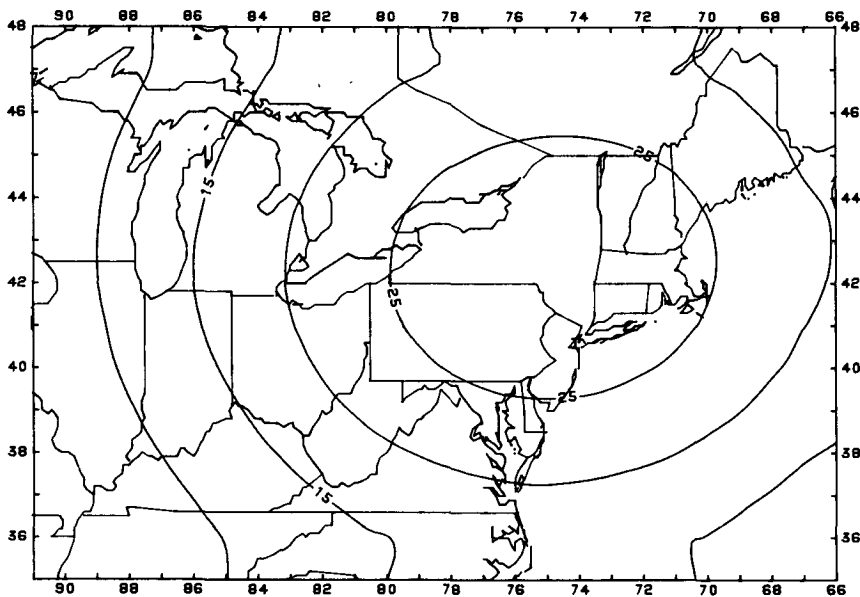


FIG. 7. Limiting endpoint distribution of modeled trajectories 48 h downwind from southeast Michigan during 1976-78 ($10^8 \times$ fraction km^{-2}).

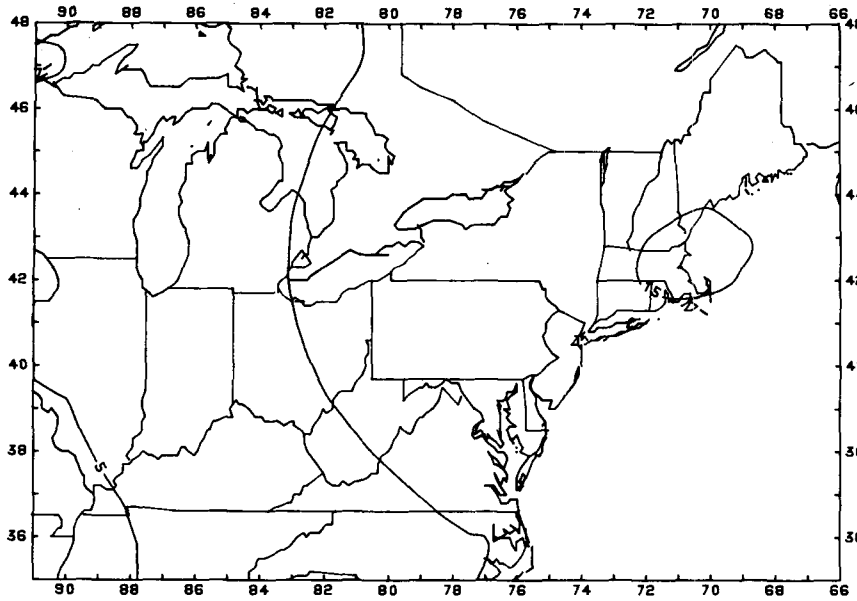


FIG. 8. Limiting endpoint distribution of modeled trajectories 72 h downwind from southeast Michigan during 1976-78 ($10^8 \times$ fraction km^{-2}).

be obtained by estimating α_x and α_y to fit the entire observed autocorrelation functions in Eqs. 13 and 14: $R_x(t)$ and $R_y(t)$; rather than only the lag-one correlation coefficients, $R_x(1)$ and $R_y(1)$ (Kao, 1965). Finally, the assumption of spatial homogeneity may introduce error. The issue of spatial homogeneity is explored further in the next section.

It is useful to compare the long-term diffusion characteristics of the model to existing estimates. One

such estimate, presented by Fay and Rosenzweig (1980) is based on the equation

$$\sigma^2 = 2D_h\tau, \tag{19}$$

in which the standard deviation of the location of trajectory endpoints (σ) is related to the travel time (τ) and the horizontal diffusivity (D_h). Eq. (19) is formally based on a random walk assumption, and is

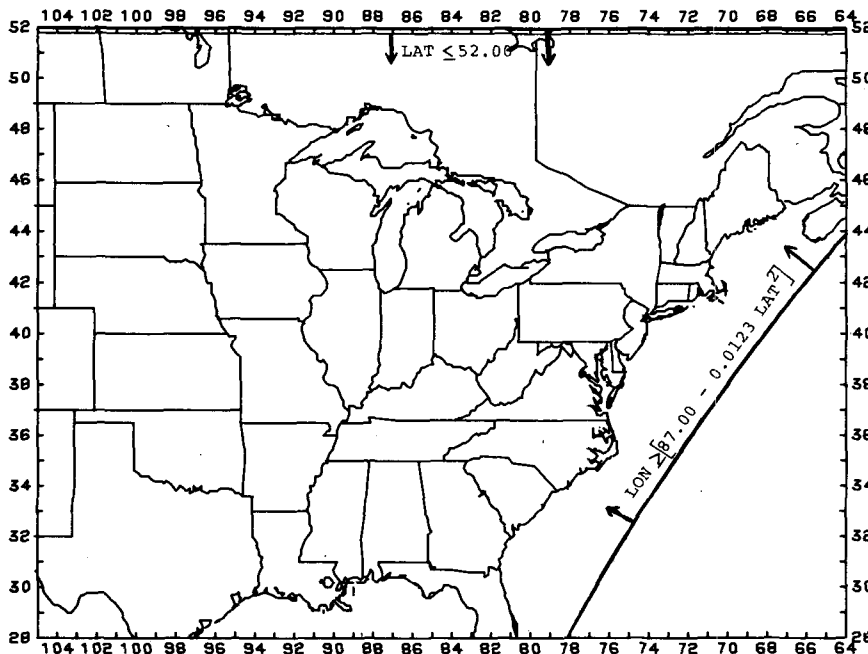


FIG. 9. Assumed termination boundaries for synthetic trajectories.

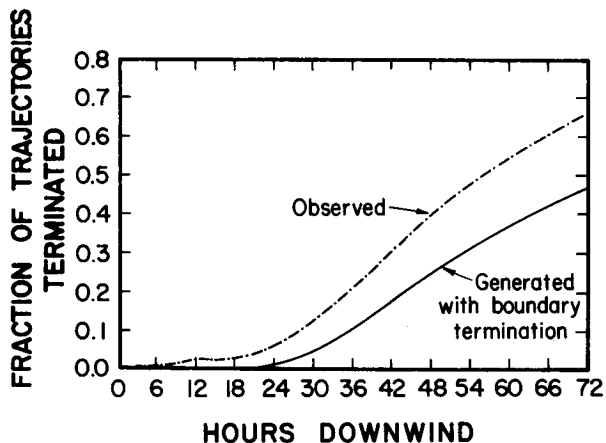


FIG. 10. Termination of trajectories originating from southeast Michigan (1976-78).

considered applicable at long travel times. The standard deviation predicted by the time series model for trajectories originating in southeast Michigan is compared to that predicted by Eq. (19) in Fig. 11. Fay and Rosenzweig estimate values of D_h appropriate for the eastern United States to be $\sim 1.0 \times 10^6$ – $2.0 \times 10^6 \text{ m}^2 \text{ s}^{-1}$. Fig. 11 indicates that the generation model produces approximately the same magnitude of horizontal diffusion as that obtained by this previous estimate, but that the generation model results in a more linear relationship at short travel times, rather than the quadratic relationship of Eq. (19). The linear result at short travel times is consistent with other theories of large-scale diffusion (Gifford, 1968; Kao, 1974; Pasquill, 1974 and the fact that Eq. (19) is generally qualified to be applicable only to long travel times.

c. Spatial inhomogeneity

A major question in the use of the simple generating method is the effect of inaccuracies introduced by assuming that model parameters are spatially homogeneous during a given study period. To provide additional insight into this issue, one month of backward trajectories (July 1978) were computed for six locations in the eastern United States and Canada, and used to estimate the parameters in the simple (spatially homogeneous) model. These results are summarized in Table 2.

Table 2 indicates that trajectories generated from the first five locations in the northeastern United States and Canada produce reasonably similar model parameters. In particular, $\sigma_{\Delta x}$, $\sigma_{\Delta y}$, α_x and σ_y are fairly uniform throughout, and a small, but significantly positive value of ρ is determined in each case. The mean wind vectors are also similar, but reductions in the zonal component begin to occur with the west-

ern Pennsylvania trajectories and become pronounced in the southern Appalachia analysis. The relatively large negative value of ΔY determined for Quebec City (-8.6 km) may be somewhat biased by the tendency for trajectories traveling north (positive ΔY) to terminate prematurely in northern Canada due to missing upper air data. The use of observed trajectories for parameter estimation may indeed be questioned throughout since the bias in the raw data will also result in bias in the estimated parameters of the stochastic model. The contention, however, is that the bias in trajectory displacements (used for parameter estimation) is much smaller than the bias in trajectory locations. This contention is supported by the successful model validation, nevertheless, the problem suggests the need for more careful screening and analysis in determining model parameters in certain cases. For example, in fringe areas such as Quebec, it may be appropriate to use only trajectory displacements within 9 or 12 h of the receptor for parameter estimation.

General conclusions may be drawn from Table 2 as to the applicability and limitations of the spatially homogeneous model applied to July 1978. If the simple model is used, parameters estimated for one location appear to be applicable to other locations over a fairly large, but still limited portion of the northeastern United States and Canada. In all cases, however, backward trajectories generated toward the southwest may be somewhat less representative as the model fails to reproduce the reduction in the mean zonal wind component in this region. The benefits of including a spatially varying mean wind vector in the generating model are implicit, but must be weighed against the considerable increase in analysis which must be performed to estimate and incorporate the spatial variation.

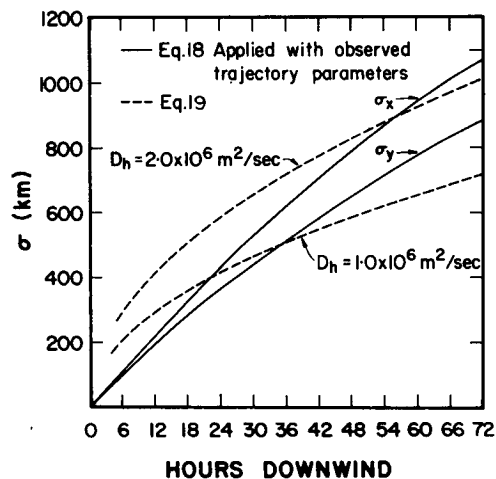


FIG. 11. Comparison of modeled long-term diffusion with previous estimate for eastern United States.

TABLE 2. Summary of backward trajectory statistics and correlations (Originating from indicated locations during July 1978).

| Location | Latitude | Longitude | Number of observations | $\overline{\Delta X}$ (km) | $\sigma_{\Delta X}$ (km) | $\overline{\Delta Y}$ (km) | $\sigma_{\Delta Y}$ (km) | α_x | α_y | ρ |
|---------------|----------|-----------|------------------------|----------------------------|--------------------------|----------------------------|--------------------------|------------|------------|--------|
| Southern | | | | | | | | | | |
| Nova Scotia | 44.00 | 66.00 | 2305 | -36.9 | 45.6 | 2.5 | 47.3 | 0.93 | 0.92 | 0.20 |
| Quebec City | 47.00 | 72.50 | 1917 | -38.9 | 37.4 | -8.6 | 42.6 | 0.88 | 0.91 | 0.28 |
| New Hampshire | 45.00 | 71.00 | 2298 | -36.3 | 39.1 | 0.18 | 48.7 | 0.90 | 0.92 | 0.10 |
| Adirondacks | 44.00 | 74.00 | 2364 | -32.3 | 39.0 | 3.2 | 48.0 | 0.91 | 0.93 | 0.10 |
| Western | | | | | | | | | | |
| Pennsylvania | 40.50 | 78.50 | 2107 | -26.2 | 40.9 | -3.9 | 38.6 | 0.89 | 0.92 | 0.13 |
| Southern | | | | | | | | | | |
| Appalachia | 35.00 | 84.00 | 2651 | -4.9 | 38.6 | 1.7 | 30.8 | 0.93 | 0.93 | -0.11 |

4. Correlated series of atmospheric trajectories

In the previous sections, a method is presented for generating a set of independent, synthetic air trajectories. With this method, each trajectory represents an independent random sample from the population of possible trajectories leaving a location at any time during a period when the statistical properties of air flow are assumed to be stationary. In contrast to this, consider a series of trajectories leaving as an ordered sequence. The spatial correlation structure of atmospheric circulation patterns dictates that there be correlation between displacements occurring on different trajectories at the same time. The amount of correlation is dependent upon the distance between the trajectory locations. A technique for generating a correlated series of trajectories is presented, and is based on the method for generating multivariate synthetic sequences of stream-flows developed by Matalas (1967). The procedure has been applied to meteorological variables by Richardson (1981).

a. Generating method

Generalizing the notation in Eqs. 4 and 5, standardized trajectory displacements may be defined as

$$d_i^j(j) = [\Delta D_i^j(j) - \overline{\Delta D(j)}] / \sigma_{\Delta D(j)}, \quad (20)$$

where the superscript i denotes the particular trajectory ($i = 1, m$), j indicates the direction of displacement ($j = 1$ for zonal, $j = 2$ for meridional), t measures discrete time, usually in units of three hours, and $\overline{\Delta D(j)}$ and $\sigma_{\Delta D(j)}$ are the respective means and standard deviations of (three-h) displacements. Comparing Eq. 20 to Eqs. 4 and 5,

$$d_i^j(1) = x_t^i, \quad (21)$$

$$d_i^j(2) = y_t^i. \quad (22)$$

Temporal stationarity is assumed in the model, but spatial inhomogeneity may be incorporated by defining $\overline{\Delta D(j)}$ and $\sigma_{\Delta D(j)}$ as a function of location.

The generating equation is

$$d_t = \mathbf{A}_i d_{t-1} + \mathbf{B}_i \gamma_t, \quad (23)$$

where d_t is a $(2m \times 1)$ matrix of the $d_i^j(j)$, γ_t is a $(2m \times 1)$ matrix of normally distributed, independent random variables with zero mean and unit variance, and \mathbf{A}_i and \mathbf{B}_i are $(2m \times 2m)$ matrices calculated to preserve the lag-zero and lag-one auto- and cross-correlations among the $d_i^j(j)$. The \mathbf{A} and \mathbf{B} matrices are computed from

$$\mathbf{A} = \mathbf{M}_1 \mathbf{M}_0^{-1}, \quad (24)$$

$$\mathbf{B} \mathbf{B}^T = \mathbf{M}_0 - \mathbf{M}_1 \mathbf{M}_0^{-1} \mathbf{M}_1^T, \quad (25)$$

where \mathbf{M}_0 and \mathbf{M}_1 are the respective $(2m \times 2m)$ lag-zero and lag-one covariance matrices for the $d_i^j(j)$ (equivalent to correlation coefficient matrices as the $d_i^j(j)$ have unit variance) and the superscripts -1 and T denote matrix inversion and transposition. Additional assumptions are required to generate the initial displacement for each trajectory. A reasonable approximation can be made with a first-order autocorrelation equation representing Eulerian wind speed and direction at the origin.

The correlation matrices \mathbf{M}_0 and \mathbf{M}_1 depend on the relative locations of trajectories 1 through m and must therefore be recomputed at each time step. Similarly, Eqs. (24) and (25) must be solved at each step. If trajectories are computed to leave a source or receptor location every six h and include 24, three-h displacements (for a total of 72 h downwind or upwind), then displacements are generated for 12 trajectories ($m = 12$) at each time step (except during the first and last 72 h of the simulation when $m < 12$). The procedure thus requires the inversion of a (24×24) matrix at each time step, resulting in large computational requirements. A simplified approximation presented by Matalas (1967) eliminates the need for matrix inversion and is described in the following section.

b. Simplified procedure

The simplified procedure allows for simulation along m trajectories with $2m$ equations:

$$d_i^j(j) = \alpha_j d_{i-1}^j(j) + \epsilon_i^j(j), \quad (26)$$

where the α_j : $\alpha_1 (= \alpha_x)$ and $\alpha_2 (= \alpha_y)$, are the lag-one

autocorrelation coefficients of zonal and meridional displacements, respectively, and the $\epsilon_i^j(j)$ are a correlated set of normal random variables with mean zero and variance equal to $1 - \alpha_j^2$. The $\epsilon_i^j(j)$ are generated simultaneously with an appropriate covariance matrix to preserve the lag-zero cross-correlations between the $d_i^j(j)$. These include the following terms:

$$E[d_i^j(1)d_i^j(2)] = \rho_i^{j,j}(1, 2), \tag{27}$$

$$E[d_i^j(1)d_i^k(2)] = \rho_i^{j,k}(1, 1), \tag{28}$$

$$E[d_i^j(2)d_i^k(2)] = \rho_i^{j,k}(2, 2), \tag{29}$$

$$E[d_i^j(1)d_i^k(2)] = \rho_i^{j,k}(1, 2), \tag{30}$$

where, for example, $\rho_i^{j,k}(1, 2)$ is the correlation coefficient between the zonal displacement of trajectory i and the meridional displacement of trajectory k at time t . It is further assumed that

$$\rho_i^{j,k}(1, 2) = \rho_i^{j,k}(2, 1). \tag{31}$$

To meet the conditions in Eqs. (26)–(31) the vector $\epsilon_i = [\epsilon_i^1(1), \epsilon_i^2(1), \dots, \epsilon_i^m(1),$

$$\epsilon_i^1(2), \epsilon_i^2(2), \dots, \epsilon_i^m(2)] \tag{32}$$

is generated at each time step with a covariance matrix of the form (for $m = 6$, for illustrative purposes):

$$\text{Cov}(\epsilon_i) = \begin{bmatrix} A & C & C & C & C & C & E & F & F & F & F & F \\ C & A & C & C & C & C & F & E & F & F & F & F \\ C & C & A & C & C & C & F & F & E & F & F & F \\ C & C & C & A & C & C & F & F & F & E & F & F \\ C & C & C & C & A & C & F & F & F & F & E & F \\ C & C & C & C & C & A & F & F & F & F & F & E \\ E & F & F & F & F & F & B & D & D & D & D & D \\ F & E & F & F & F & F & D & B & D & D & D & D \\ F & F & E & F & F & F & D & D & B & D & D & D \\ F & F & F & E & F & F & D & D & D & B & D & D \\ F & F & F & F & E & F & D & D & D & D & B & D \\ F & F & F & F & F & E & D & D & D & D & D & B \\ F & F & F & F & F & E & D & D & D & D & D & B \end{bmatrix}, \tag{33}$$

where

$$A = 1 - \alpha_1^2, \tag{34}$$

$$B = 1 - \alpha_2^2, \tag{35}$$

$$C = \rho_i^{j,k}(1, 1) - \alpha_1^2 \rho_{i-1}^{j,k}(1, 1), \tag{36}$$

$$D = \rho_i^{j,k}(2, 2) - \alpha_2^2 \rho_{i-1}^{j,k}(2, 2), \tag{37}$$

$$E = \rho_i^{j,j}(1, 2)(1 - \alpha_1 \alpha_2), \tag{38}$$

$$F = \rho_i^{j,k}(1, 2) - \alpha_1 \alpha_2 \rho_{i-1}^{j,k}(1, 2). \tag{39}$$

If zonal and meridional displacements are assumed to be independent, $E = F = 0$ and the $\epsilon_i^j(1)$ may be generated separately from the $\epsilon_i^j(2)$. The covariance matrix must be redetermined at each time step and ϵ_i generated in a manner consistent with $\text{Cov}(\epsilon_i)$. This may be accomplished using a number of common procedures for generating a set of correlated normal random variates (Sheuer and Stoller, 1962; Yakowitz, 1977). Note that lag-one cross-correlations are not preserved by the simplified model, but are assumed equal to the product of the lag-zero cross-correlation and the lag-one autocorrelation. Assuming this approximation is satisfactory, the simplified procedure has the added benefit that lag-one cross-correlations need not be estimated from historical data. A spatial-lag-zero relationship is still required though to estimate $\rho_i^{j,k}(1, 1)$ and $\rho_i^{j,k}(2, 2)$.

Spatial correlation functions for winds and other

meteorological variables are presented by Buell (1972), Julian and Thiebaut (1975), Thiebaut (1976) and Pratte and Lee (1979). To illustrate this type of relationship, Figs. 12a and b show $\rho_i^{j,k}(1, 1)$ and $\rho_i^{j,k}(2, 2)$, respectively, as a function of the separation distance between trajectories i and k at time t . These were computed using observed, mixed-layer trajectories computed with the method of Heffter (1980) for 20 cities in the eastern United States for the period 1 July to 15 September 1978. These data may be represented by equations of the following form

$$\rho_i^{j,k}(1, 1) = [1 - (r_x/L)^2] \exp(-r^2/2L^2), \tag{40}$$

$$\rho_i^{j,k}(2, 2) = [1 - (r_y/L)^2] \exp(-r^2/2L^2), \tag{41}$$

where r is the separation distance, r_x and r_y are the zonal and meridional components of the separation distance, respectively (i.e., $r^2 = r_x^2 + r_y^2$), and L is a characteristic distance of correlation. In each case a value of L on the order of 700 km would be appropriate.

5. Application of stochastic trajectories to long-range atmospheric transport models

Synthetic air trajectories may be used as inputs in a variety of Lagrangian, long-range transport models. The most direct application is to receptor-oriented models where a single set of backward trajectories is used to compute pollutant concentrations at the re-

ceptor location (Eliassen and Saltbones, 1975; Eliassen, 1978; Voldner, *et al.* 1981; and Samson and Small, 1983). Stochastic forward trajectories may be used for source-oriented models, however, they are appropriate only when estimating the distribution of effects from a single source, but not the combined effects of many sources. This is due to the correlations which occur between forward trajectories leaving each of the separate sources. The generation of an entire, spatially correlated wind field using methods such as those described by Pratte and Lee (1979) would be required for source-oriented models.

The output from a stochastic simulation of atmospheric pollution transport allows a probabilistic characterization of pollutant concentrations. The results may be compared to observed concentrations in a statistical sense, but nothing can be said concerning air pollution episodes occurring on a particular day or the pH of a particular rainstorm. It is useful to contrast results which may be obtained from a (receptor-oriented) long-range transport model using independent trajectories to results obtainable using a correlated series of atmospheric trajectories. Air pollution concentrations, calculated with the arrival of independent trajectories, represent instantaneous, independent estimates. These may be analyzed to determine a mean, variance and frequency distribution of pollutant concentrations. Nothing can be said, however, concerning the temporal structure and pattern of air pollution levels. To obtain this capability, simulations using a correlated series of synthetic trajectories are necessary. In addition to the mean, variance and frequency distribution, these results may be analyzed to determine temporal features such as the autocorrelation function, spectral density, and statistical properties of time averaged (e.g., 24 h) concentrations.

The procedure for generating an independent series of trajectories has been incorporated in a stochastic model of atmospheric regional transport (Small, 1982). In this application, synthetic Lagrangian precipitation is also simulated in a cogeneration method which accounts for relationships between wind fields and precipitation. The method was found to produce representative meteorological inputs for long-term (monthly, seasonal and yearly) simulations. Furthermore, considerable savings in both data preparation effort and computation time were obtained when compared to the calculation and use of observed trajectory and precipitation fields.

The generation method for a correlated sequence of trajectories has not, as yet, been implemented in an atmospheric transport model. The greater computational complexity relative to the procedure for independent trajectories may discourage its use, however, it remains a scientifically sound method for generating more representative trajectory fields. The method is particularly appropriate for the evaluation

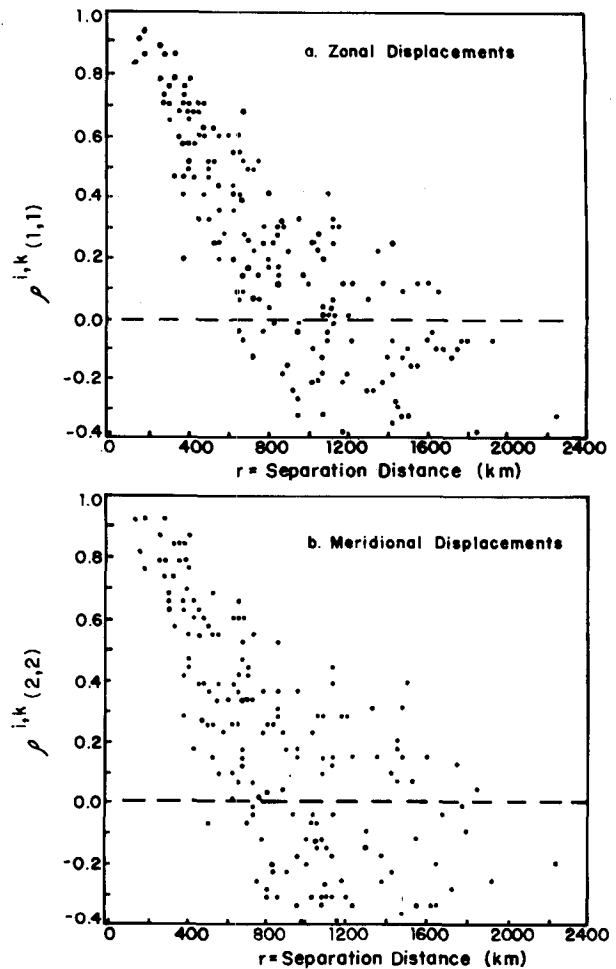


FIG. 12. Observed spatial correlation of trajectory displacements (based on forward trajectories computed for 20 cities in the eastern United States, 1 July–15 September 1978).

of the statistical properties of episodes and exposures with time scales on the order of one to a few days.

6. Summary

Methods for generating stochastic atmospheric trajectories have been developed, including a simple time series/autoregression model for simulating independent trajectories, and a somewhat more involved procedure for simulating a correlated ensemble of sequential trajectories. The application of the model for independent trajectories to three years of data from southeastern Michigan demonstrates that the method produces a set of trajectories which is statistically consistent with observed atmospheric transport and existing estimates of long-term horizontal diffusion. It also highlights the problems which may arise when observed trajectories are used in atmospheric transport modeling, and supports the use of synthetic trajectories, at least in a complementary

analysis, to obtain a more accurate long-term representation. The procedure for correlated trajectories is not implemented, though it may prove useful for transport problems of shorter time scales where wind persistence is important.

Acknowledgments. The work presented in this paper was supported in part by grants from the Environmental Protection Agency and a Rackham predoctoral fellowship at the University of Michigan. At the time of the study, Mitchell J. Small was a graduate student in the Environmental and Water Resources Engineering Program at the University of Michigan. The help and guidance of Professor Walter J. Weber, Jr. is gratefully acknowledged.

REFERENCES

- Bacon, J. W., B. Henderson-Sellers and A. Henderson-Sellers, 1981: The application of a stochastic wind model to the meteorology of north-west England. *Proc. 11th NATO-CCMS Int. Tech. Meeting on Air Pollution Modeling*. Amsterdam. *Air Pollution Modeling and its Application I*. C. D. Wispelaere Ed., Plenum Press, 223-235.
- Bolin, B., and C. Persson, 1975: Regional dispersion and deposition of atmospheric pollutants with particular application to sulfur pollution of western Europe. *Tellus*, **27**, 281-310.
- Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis: Forecasting and Control*. Rev. ed., Holden Day, 575 pp.
- Buell, E. C., 1972: Correlation functions for wind and geopotential on isobaric surfaces. *J. Appl. Meteor.* **11**, 51-59.
- Durst, C. S., A. F. Crossley and N. E. Davies, 1959: Horizontal diffusion in the atmosphere as determined by geostrophic trajectories. *J. Fluid Mech.*, **6**, 401-422.
- Eliassen, A., 1978: The OECD study of long-range transport of air pollutants: long-range transport modeling. *Atmos. Environ.*, **12**, 479-488.
- , and J. Saltbones, 1975: Decay and transformation rates of sulphur dioxide as estimated from emission data, trajectories, and measured air concentrations. *Atmos. Environ.*, **9**, 425-429.
- Fay, J. A., and J. J. Rosenzweig, 1980: An analytical diffusion model for long distance transport of air pollutants. *Atmos. Environ.*, **14**, 355-365.
- Gifford, F. A., Jr., 1968: An outline of theories of diffusion in the lower layers of the atmosphere. *Meteorology and Atomic Energy*, 1968, D. H. Slade, Ed., U.S. Atomic Energy Commission, 65-116.
- Heffter, J. L., 1980: Air resources laboratories atmospheric transport and dispersion model (ARL-ATAD)/NOAA. Tech. Memo. ERL ARL-81. Air Resources Laboratories. Silver Spring, Maryland, 17 pp.
- Julian, P. R., and H. J. Thiebaut, 1975: On some properties of correlation functions used in optimum interpolation schemes. *Mon. Wea. Rev.*, **103**, 605-616.
- Kao, S. K., 1965: Some aspects of the large-scale turbulence and diffusion in the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **91**, 10-17.
- , 1974: Basic characteristics of global scale diffusion in the troposphere. *Advances in Geophysics*, Vol. 18B, Academic Press, 15-32.
- Matalas, N. C., 1967: Mathematical assessment of synthetic hydrology. *Water Resour. Res.*, **3**, 937-945.
- Murgatroyd, R. J., 1969: Estimations from geostrophic trajectories of horizontal diffusivity in the mid-latitude troposphere and lower stratosphere. *Quart. J. Roy. Meteor. Soc.*, **95**, 40-62.
- Pasquill, J., 1974: *Atmospheric Diffusion*, 2nd ed. Halsted Press, Wiley, 429 pp.
- Pratte, J. F., and R. W. Lee, 1979: A short method of generating meteorological fields for simulation studies. *J. Appl. Meteor.*, **18**, 1670-1673.
- Richardson, C. W., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.*, **17**, 182-190.
- Samson, P. J., and M. J. Small, 1983: The use of atmospheric trajectory models for diagnosing the sources of acid precipitation. *Acid Precipitation*, Vol. 9, *Modeling of Total Acid Precipitation Impacts*, J. L. Schnoor, Ed., Ann Arbor Science.
- Shannon, J. D., 1981: A model of regional long-term average sulfur atmospheric pollution, surface removal, and net horizontal flux. *Atmos. Environ.* **15**, 689-701.
- Sheih, C. M., 1977: Application for a statistical trajectory model to the simulation of sulfur pollution over northeastern United States. *Atmos. Environ.*, **11**, 173-178.
- Sheuer, E. M., and D. S. Stoller, 1962: On the generation of normal random vectors. *Technometrics*, **4**, 278-281.
- Small, M. J., 1982: Stochastic model of atmospheric regional transport. Ph.D. thesis, Environmental and Water Resources Engineering Program, University of Michigan, 237 pp.
- Taylor, G. I., 1921: Diffusion by continuous movements. *Proc. London Math. Soc. (2)*, **20**, 196-202.
- Thiebaut, H. J., 1976: Anisotropic correlation functions for objective analysis. *Mon. Wea. Rev.*, **104**, 994-1004.
- Yakowitz, S., 1977. *Computational Probability and Simulation*. Addison-Wesley, 240 pp.