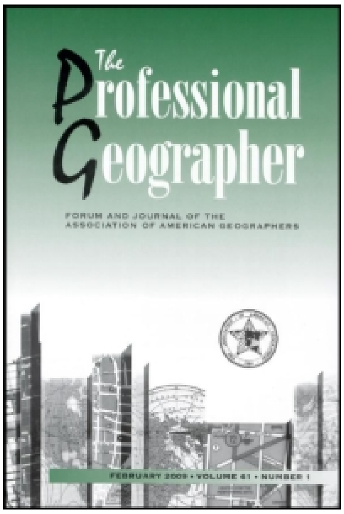

The Professional Geographer
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t788352615

SIMPLE MARKOV CHAIN MODEL OF SMOG PROBABILITY IN THE SOUTH COAST AIR BASIN OF CALIFORNIA

Gong-Yuh Lin[a]
[a] Associate Professor of Geography, California State University, Northridge

Online publication date: 15 March 2010

To cite this Article Lin, Gong-Yuh(1981) 'SIMPLE MARKOV CHAIN MODEL OF SMOG PROBABILITY IN THE SOUTH COAST AIR BASIN OF CALIFORNIA', The Professional Geographer, 33: 2, 228 — 236
To link to this Article: DOI: 10.1111/j.0033-0124.1981.00228.x
URL: http://dx.doi.org/10.1111/j.0033-0124.1981.00228.x

Professional Geographer, 33(2), 1981, pp. 228–236

SIMPLE MARKOV CHAIN MODEL OF SMOG PROBABILITY IN THE SOUTH COAST AIR BASIN OF CALIFORNIA

Gong-Yuh Lin

California State University, Northridge

The simple Markov chain model approximates sequences of episode days for those air-monitoring stations with sufficient data for analysis. The conditional probability for episode days shows a distinct regional pattern, with highest values over the inland valleys and decreasing values toward the coast.

OXIDANT concentration for any given place and time is primarily the result of three factors: pollution emission rates, weather conditions, and topography. Because emission rates do not vary greatly from day to day and topography is constant, daily variations in oxidant concentration are most likely caused by fluctuations in weather conditions. Even though weather conditions as a whole vary from day to day, certain conditions persist. The persistence of these weather events has been a topic of great interest to meteorologists.

The first-order Markov chain model has been employed frequently to study the conditional probability and persistence of wet and dry days [3] and visibility categories [5]. Likewise, smog levels, which are to a large extent controlled by weather conditions, may show a similar persistence. Previous research indicates that the concentration of oxidant on a given day is very useful for predicting the oxidant concentration for the following day [*1*]. The one-day-lag oxidant correlation coefficient for inland valleys of the South Coast Air Basin is approximately 0.65, implying a high persistence of oxidant levels. This article reports on a study of the conditional probability and sequences of episode and nonepisode days derived from the simple Markov chain models for the forty air-monitoring stations in the South Coast Air Basin (Figure 1).

Background

Since 1974, the California Air Resources Board has mandated an emergency warning system for high pollution concentrations. The episode system has been in force in the South Coast Air Basin since 1976 [*8*]. Episode criteria for oxidant concentrations are one-hour average concentrations of 0.20, 0.35, and 0.50 ppm for Stages 1, 2, and 3, respectively. An episode stage must be called when an oxidant concentration equaling or exceeding the criterion is either predicted or reached. The major purpose of the episode declaration is to prevent excessive build-ups of pollution in order to protect the public health from pollution effects. In this study, an episode day is defined as a day with one-hour average concentration of oxidant equaling or exceeding 0.20 ppm. Thus it includes all three episode stages. However, Stage 2 and Stage 3 episode days have rarely occurred at most stations. During the study period, almost all the episode days were called for the Stage 1 criterion.

Oxidant is a secondary pollutant that is formed in the air through complicated photochemical reactions among various primary pollutants, mainly oxides of nitrogen, hydrocarbon, and carbon monoxide, in the presence of ultraviolet radiation [6, *7*]. Hence oxidant tends to build up in the air during warm periods. In the South Coast Air Basin, oxidant concentrations peak in summer seasonally and at noon or in the afternoon diur-

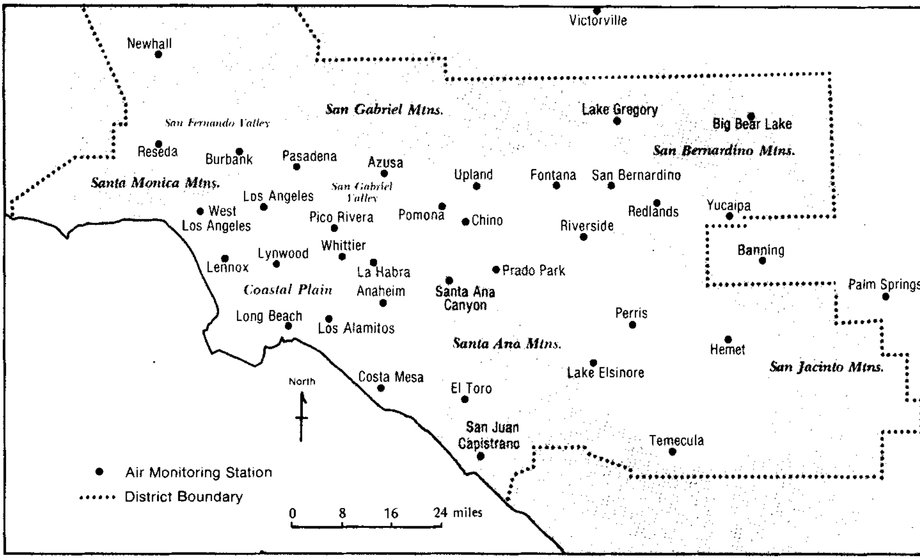

Figure 1. Locations of air-monitoring stations in the South Coast Air Basin.

nally. Once formed, the vertical dispersion of oxidant is hampered by the presence of an inversion layer caused by the compressional heating of the subsiding air current associated with the North Pacific subtropical anticyclone, which persists off the California coast in summer. Day-to-day variations in oxidant levels in summer are rarely reflected by drastic changes in the synoptic circulation pattern, i.e., the presence or absence of the anticyclone, but rather are revealed by variations in the intensity of ultraviolet radiation and the subsiding air motion associated with the anticyclone. However, data on ultraviolet radiation are not routinely available. The South Coast Air Quality Management District has employed radiosonde data taken at Los Angeles International Airport as predictor variables to develop statistical models for forecasting oxidant concentrations in the South Coast Air Basin [9]. It was found that episodes occurred on the days with higher temperatures, particularly at the upper levels approximately from 2,000- to 5,000-foot (850 mb) heights (Figure 2). The morning (0600 PST) 850-mb temperature shows a significantly high correlation coefficient (0.60) with the oxidant concentration for inland valley stations where episodes occur most frequently. At Upland, for instance, the morning 850-mb temperature is on average 5°C higher on episode than on nonepisode days. The onset of episode spells is frequently accompanied by increasing air temperatures aloft. It was found that 71.3 percent of the episode spells at Upland began with an average 3.2°C 850-mb temperature increase from the previous nonepisode day, whereas 70.1 percent of episode spells ended with an average 2.6°C temperature decrease.

The morning 850-mb temperature, an index of the intensity of both radiation and subsiding air motion associated with the anticyclone, shows strong persistence as revealed by a significantly high one-day-lag correlation coefficient of 0.82.

Method

A finite (finite number of states) Markov chain process is a finite stochastic process such that the outcome of any future trial is determined by the outcome of the last experiment

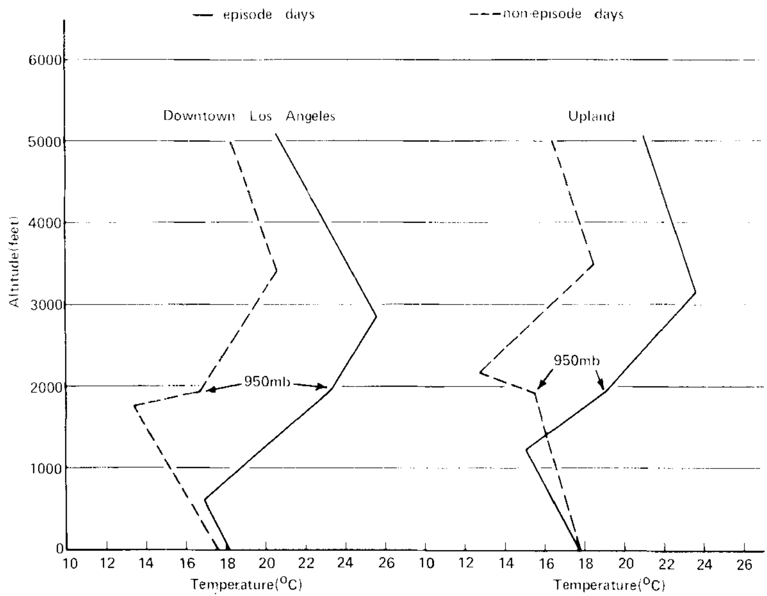

Figure 2. Mean morning temperature profiles observed at Los Angeles International Airport on episode and nonepisode days at Los Angeles and Upland.

and is independent of any previous outcome [4, 2]. Because this article deals with the two-state (episode and nonepisode) Markov chain probability, it requires the computation of two parameters, p_0 and p_1. P_0 is the conditional probability of an episode day (Y) given that the previous day had no episode (N), and p_1 is the conditional probability of an episode day given that the previous day had an episode. According to this probability model, any given day should fall into one of the following categories: $Y|N$, $Y|Y$, $N|Y$, and $N|N$. The first letter represents today and the second letter yesterday. Thus,

$$p_1 = \frac{n(Y|Y)}{n(Y|Y) + n(N|Y)}$$

and

$$p_0 = \frac{n(Y|N)}{n(Y|N) + n(N|N)},$$

where n denotes the total number of days for a given category. The probability of an episode sequence of length k is

$$(1 - p_1)\, p_1^{k-1},$$

and the probability for episode sequence greater than k is

$$p_1^{k}.$$

Similarly, the probability of a nonepisode sequence of length k is

$$p_0\,(1 - p_0)^{k-1},$$

and the probability for nonepisode sequence greater than k is

$$(1 - p_0)^{k}.$$

A state of a Markov chain is called absorbing if the system remains in the state once it enters there. A quantity of interest is the proportion of a long period of time spent in each of the two states. If the episode days are set to be the absorbing state, the mean first passage time is computed as

$$\frac{1}{(1-(1-p_0))} = p_0^{-1}.$$

This can be interpreted as the average length of nonepisode spells or the mean time before reaching the absorbing state. Likewise, the mean length of episode spells is

$$(1-p_1)^{-1}.$$

The estimated transition probability matrix is

$$P = \begin{array}{c c} & \begin{array}{cc} N & Y \end{array} \\ \begin{array}{c} N \\ Y \end{array} & \begin{pmatrix} 1-p_0 & p_0 \\ 1-p_1 & p_1 \end{pmatrix} \end{array},$$

and the kth-step transition probabilities can be obtained from the matrix P^k. After a sufficiently long period of time, the system is expected to settle down to a condition of statistical equilibrium in which the state occupation probabilities are independent of the initial conditions. The equilibrium probability vector can be written as

$$P^{(k)} = (\pi_0, \pi_1),$$

where π_0 and π_1 are the equilibrium probabilities of nonepisodes and episodes, respectively. These two quantities can be computed as relative frequencies of nonepisode and episode days or can be derived by computing successively the kth power of the transition matrix P until the equilibrium probabilities are obtained, i.e.,

$$P^k = \begin{pmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{pmatrix}.$$

Discussion

The conditional probabilities of episode days and frequencies of episode sequences of various lengths for the four summer smog seasons, May 1 through October 31 of 1974 through 1977, were calculated according to the above equations for the forty air-monitoring stations in the South Coast Air Basin. Table 1 shows an example of the procedure for estimating the conditional probability of episode occurrences at Upland. The probabilities for the other stations were calculated similarly. The conditional probability of episode days shows distinct spatial variation (Figures 3 and 4). Values of both p_1 and p_0 increase from the coast to the inland valleys. The occurrence of episode days is rather persistent at the inland valleys, as revealed by high p_1 values. For instance, p_1 at Upland is 0.76, the highest value in the basin, whereas p_0 is only 0.23. At the coastal stations, such as Lennox and Long Beach, the oxidant concentration was too low to have any episode days during the study period. This is reflected in zero probability values for these stations.

Table 2 shows the predicted and observed frequencies of episode sequences of various lengths for selected stations. Chi-square values were calculated to test goodness of fit of actual occurrences of episode sequences to the simple Markov chain model. Sequences with predicted frequencies of less than 5 were combined together to form a new group. Yate's correction factor of 0.5 was applied to calculate the chi-square value in case only two sequences (1 degree of freedom) had sufficient frequencies to perform the chi-square test. Most stations had small chi-square values, with the highest value of 13.20 for Upland,

TABLE 1

ESTIMATES OF p_1 AND p_0 VALUES AT UPLAND

Yesterday	Today: Yes	Today: Yes & No	p_1 (Y|Y)	p_0 (Y|N)
Yes	271	358	0.76	
No	86	374		0.23

TABLE 2

PREDICTED AND OBSERVED FREQUENCIES OF SEQUENCES OF EPISODE DAYS OF VARIOUS LENGTHS

	Los Angeles		Azusa		Upland	
Lengths (Days)	PR	OB	PR	OB	PR	OB
1	22.73	23.00	26.79	27.00	20.64	24.00
2	6.27*	6.00*	17.59	18.00	15.48	23.00
3			11.55	8.00	12.04	8.00
4			7.58	8.00	9.46	7.00
5			14.45*	17.00*	6.88	2.00
6					5.16	9.00
7					16.34*	13.00*
chi-square	0.01		1.58		13.20	
degree of freedom	1		4		6	

* Indicates frequencies of sequences of a given length and longer.

TABLE 3

PREDICTED AND OBSERVED FREQUENCIES OF SEQUENCES OF NONEPISODE DAYS OF VARIOUS LENGTHS

	Los Angeles		Azusa		Upland	
Lengths (Days)	PR	OB	PR	OB	PR	OB
1	1.39	2.00	12.59	16.00	20.47	22.00
2	1.32	1.00	10.64	8.00	15.75	25.00
3	1.25	1.00	8.98	13.00	12.10	7.00
4	1.22	0.00	7.58	9.00	9.35	6.00
5	1.16	3.00	6.41	9.00	7.21	7.00
6	1.12	2.00	5.41	2.00	5.52	5.00
7	1.09	0.00	29.41*	24.00*	18.51*	17.00*
8	1.02	3.00				
9	0.99	1.00				
10	22.59	20.00				
chi-square	0.55**		7.84		9.07	
degree of freedom	1**		6		6	

* Indicates frequencies of sequences of a given length and longer.
** Indicates chi-square value is calculated by combining sequences of lengths 1 through 9 days into one category.

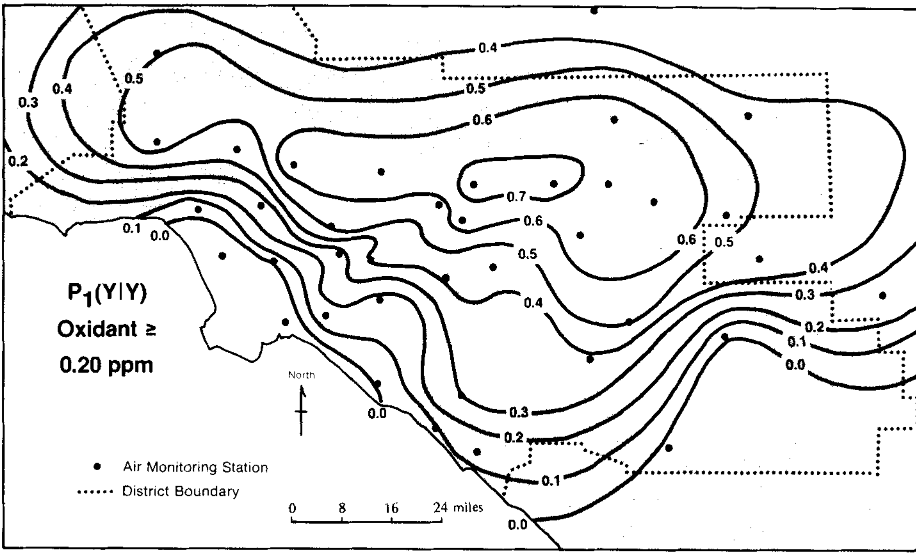

Figure 3. Conditional probabilities of an episode day, given that the previous day had an episode.

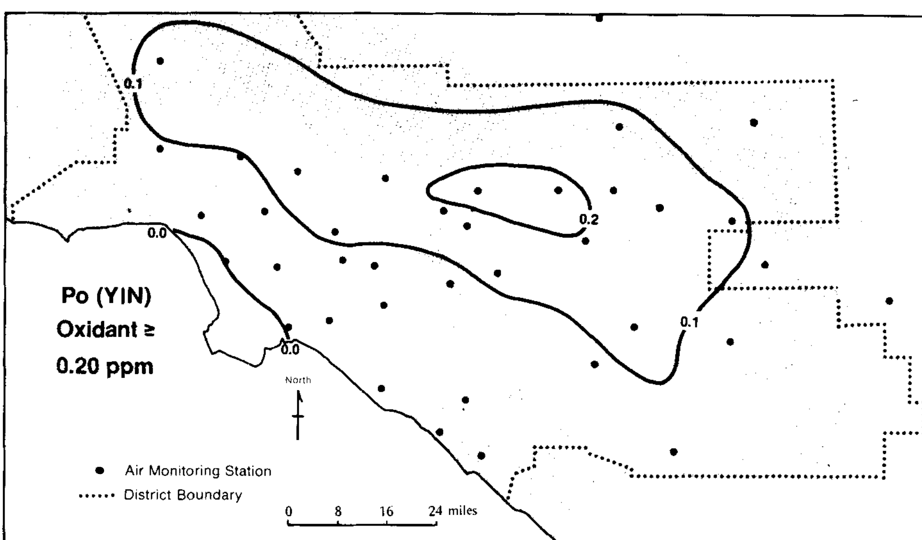

Figure 4. Conditional probabilities of an episode day, given that the previous day had no episode.

which is less than the 1 percent critical value of 16.81 with 6 degrees of freedom. It appears that the predicted frequencies derived from the Markov chain model agree well with the observed frequencies, particularly for the shorter sequences.

Table 3 shows predicted and observed frequencies of nonepisode sequences of various

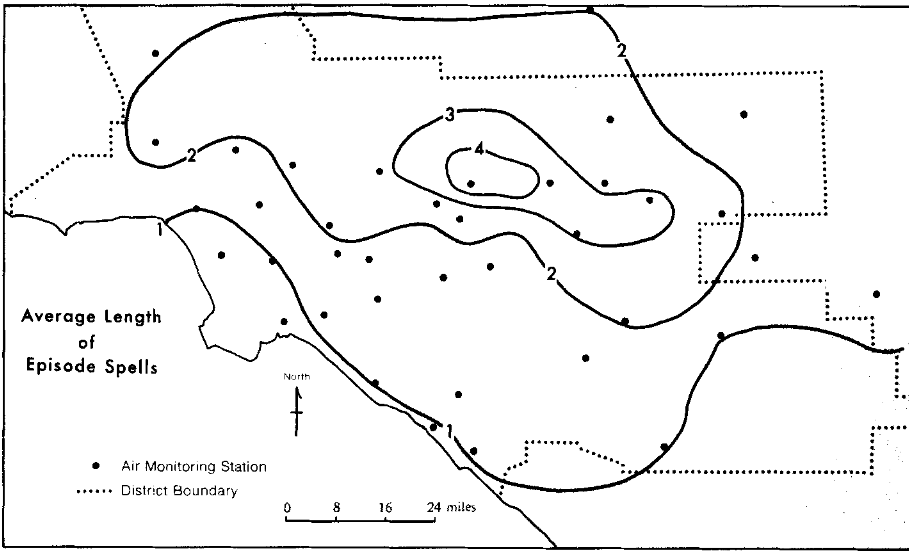

Figure 5. Average lengths of episode spells.

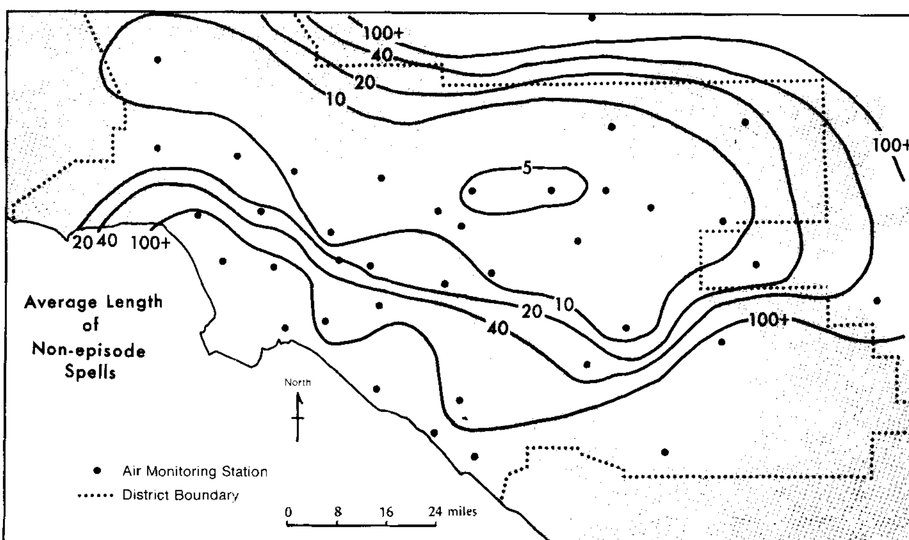

Figure 6. Average lengths of nonepisode spells.

lengths for Azusa, Los Angeles, and Upland. Chi-square tests indicate that there are agreements between the predicted and observed frequencies of nonepisode spells for Azusa and Upland. At Los Angeles, where episode days occur less frequently, the nonepisode spell is long. The Markov chain model predicts that approximately 68 percent of nonepisode spells would last for 10 days or longer. If spells of lengths 1 through 9 days are combined into one category, as opposed to a second category for nonepisode spells of 10 days and longer, the chi-square value is 0.55, which reflects a good fit between observed and predicted frequencies.

From Table 1, the two-state transition probability matrix for Upland is

$$P = \begin{pmatrix} 0.77 & 0.23 \\ 0.24 & 0.76 \end{pmatrix},$$

and the equilibrium probability vector is

$$p^{(k)} = (0.51, 0.49).$$

This implies that the probability that any day will be a nonepisode day is 0.51, and 0.49 that it will be an episode day. The equilibrium condition would be reached in 12 days, as the 12th ($k = 12$) power of matrix P yields these equilibrium probabilities. This reflects the fact that the probability of episode or nonepisode occurrence on a given day is independent of the probability of episode or nonepisode occurrence 12 days earlier. The equilibrium probability vector is (0.95, 0.05) for Los Angeles and (0.69, 0.31) for Azusa. It thus requires 4 days for Los Angeles and 9 days for Azusa to reach the equilibrium condition.

The average length of episode spells varies from approximately 4 days at inland valley stations to less than 1 day at coastal stations (Figure 5). By contrast, the nonepisode spell lasts much longer at coast stations than at inland valley stations (Figure 6). It varies from approximately 5 days at inland valley stations to more than 100 days at coastal stations. At stations where there are no episode days, the nonepisode spell lasts indefinitely.

A smog cycle may be defined as an episode spell followed by a nonepisode spell. At Upland, the average lengths of episode and nonepisode spells are 4.17 and 4.35 days, respectively. Hence the smog cycle is 8.52 days. The cycles for Azusa and Los Angeles are 9.35 and 25.09 days, respectively.

Conclusion

The declaration of episode days is of great concern to the public because abatement actions, ranging from stopping all unnecessary driving to curtailing industrial and commercial operations, could result in serious economic impacts and personal inconvenience. The occurrence of episode days thus deserves attention.

This study reveals that the first-order Markov chain models fit sequences of occurrence of episode days in areas where episode days occur frequently. The aim of the Markov chain model is to provide statistical description of the persistence of episode days. The good fit of the model suggests that smog concentration on one day could be enhanced from the addition of residual smog from the preceding day or, alternatively, it could imply that the weather conditions that favor high smog concentrations show persistence.

Literature Cited

1. Aron, R. H., and I. M. Aron. "Statistical Forecasting Models: II. Oxidant Concentration in the Los Angeles Basin." *Journal of the Air Pollution Control Association,* 28 (1978), 684–88.
2. Cox, D. R., and H. D. Miller. *The Theory of Stochastic Processes.* New York: John Wiley & Sons, 1965.
3. Gabriel, K. R., and J. Neumann. "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv." *Quarterly Journal of the Royal Meteorological Society,* 88 (1962), 90–95.
4. Kemeny, J. G., and J. L. Snell. *Finite Markov Chains.* New York: Van Nostrand, 1960.
5. Lund, I. A., and D. D. Grantham. *Persistence, Runs, and Recurrence of Visibility.* Hanscom AFB, Mass: Air Force Geophysics Laboratory Enviromental Research Papers, No. 624, 1978.
6. Perkins, H. C. *Air Pollution.* New York: McGraw-Hill, 1974.

7. Peterson, J. T., and K. L. Demerijian. "The Sensitivity of computed Ozone Concentrations to U.V. Radiation in the Los Angeles Area." *Atmospheric Environment,* 10 (1976), 459–68.
8. South Coast Air Quality Management District. *Air Quality & Meteorology*. El Monte, Calif. SCAQMD Annual Report, 1976.
9. Zeldin, M. D., and J. C. Cassmassi. *Development of Improved Methods for Predicting Air Quality Levels in the South Coast Air Basin.* Santa Monica, Calif. Technology Service Corporation Report, 1979.

GONG-YUH LIN, Associate Professor of Geography at California State University, Northridge, is interested in air pollution climatology. He has a Ph.D. from the University of Hawaii.