A Method for Determining Change in Precipitation Data

MARIE R. KLUGMAN

Department of Geography, University of Iowa, Iowa City 52242

STUART A. KLUGMAN

Department of Statistics and Actuarial Science, University of Iowa, Iowa City 52242

(Manuscript received 13 October 1980, in final form 4 September 1981)

ABSTRACT

A method is proposed for determining the occurrence of climatic change with respect to precipitation. The method is an analysis of variance procedure with a Markov chain model used to estimate the within-group variance. An illustration using data from Medford, Oregon, for the fall seasons from 1948 to 1976 is provided.

1. Introduction

Temporal and spatial precipitation variations have interested many climatologists, but most investigations into these topics have been primarily concerned with delineating specific areas and time periods affected by abnormal precipitation amounts. In many of these studies mean precipitation amounts for months, seasons or years have been compared to determine whether significant precipitation variations occurred during the period of record. Trends, periodicities and abrupt changes in mean precipitation amounts were sought which would conform to respective changes in hypothesized causal factors such as sunspots and circulation regimes.

Two properties of precipitation data which may be of greater importance in determining whether significant climatic variations have occurred are the variance and the persistence. Knowledge of these two properties of past precipitation amounts would help to set bounds on predictions of future precipitation. Little work in precipitation climatology has been concerned with these properties of precipitation amounts; two notable exceptions are Skaggs (1978) and Granger (1979). In both of these studies precipitation means were investigated but no significant changes were found over time. However, in western Kansas for the period 1887–1968, Skaggs (1978) found that the autocorrelation or persistence in the precipitation data had changed through time. Granger (1979) found that the probability of extreme amounts of precipitation (very high or very low) had increased in California between the periods 1921–40 and 1961–77. In both cases it can be argued that the precipitation factors which changed over time are as important, if not more important, for agricultural and public water needs as changes in the simple precipitation mean. Considering these implications, a more meaningful determination of climatic change may result from the investigation of the variance of precipitation.

Jones (1975) suggested an analysis of variance approach to test for changes in climatic variables over time. He noted that autocorrelation in climatic variables is a complicating feature of this problem. If dependence is not taken into account, the within-groups sum of squares underestimates the group to group variability under the null hypothesis of equal group means. The result is an inflated F-ratio and a higher probability of concluding that the climate has changed when in fact it has not. Madden (1976) and Madden and Shea (1978) have applied Jones' suggestions to sea level pressure data and temperature data, respectively.

To determine a more accurate estimate of the within-group (daily) variability of precipitation data, a model which allows for dependence must first be fitted to the daily data. A first-order autoregressive [AR(1)] model has been fitted to climatic data (e.g., Jenkinson, 1957; Leith, 1973). Using precipitation data the AR(1) model assumes that the amount of precipitation falling on day n is a function of the amount on day $n - 1$ plus a random component. While this may describe the dependence in most climatic variables, it does not in general describe the dependence in precipitation. A more accurate description of the dependence in precipitation amounts is that the amount of precipitation falling on day n depends only on whether the previous k days were wet or dry. The generalization of the Markov chain as proposed by Katz (1977a) fits this description.

0021-8952/81/121506-04$05.00

From Katz's model an estimate of the group to group variability under the null hypothesis which allows for dependence can be obtained.

2. The model for k-day dependence

The model for daily precipitation developed by Katz (1977b) considers the process $\{(X_n, J_n): n = \cdots, -1, 0, 1, \cdots\}$ where X_n is the amount of rainfall on day n and J_n indicates the rainfall pattern (wet or dry) for days $n, n-1, \ldots, n-k+1$. Let $H_i = 1$ if it rained more than a trace in day i and $H_i = 0$ otherwise. Then $J_n = \sum_{i=0}^{k-1} H_{n-i}2^i$. The $\{J_n\}$ process forms a Markov chain with states $\{0, 1, \ldots, 2^k - 1\}$. For example, if $k = 2$, then J_n can assume the values 0, 1, 2 and 3. The rainfall pattern can best be discerned by writing the value in base two. Thus, $2 = 10_2$ and so $J_n = 2$ indicates that there was rain on day $n - 1$ and no rain on day n. Similarly, if $k = 3$ and $J_n = 6 = 110_2$ there was rain on days $n - 2$ and $n - 1$ but not on day n.

Let $\mathsf{P} = (P_{ij})$ be the matrix of transition probabilities, i.e., $P_{ij} = P[J_n = j|J_{n-1} = i]$. Each row and column of P contains at most two nonzero elements and the rows each sum to 1. For example, if $k = 2$ and $i = 2$, then only P_{20} and P_{21} can be nonzero, as the other transitions are not possible. $P_{23} > 0$ implies the possibility of a transition from rain on day $n - 2$ and no rain on day $n - 1$ to rain on days $n - 1$ and n and $P_{22} > 0$ also requires both no rain and rain on day $n - 1$.

Let $\boldsymbol{\pi}' = (\pi_0, \pi_1, \ldots, \pi_{2^k-1})$ be the vector of stationary probabilities ($\boldsymbol{\pi}'\mathsf{P} = \boldsymbol{\pi}'$ and its elements sum to 1). The $\{X_n\}$ process is conditionally independent given the $\{J_n\}$ process. In particular, the probability distribution of rainfall on day n depends only on J_n and J_{n-1}. That is,

$$F_{ij}(x) = P[X_n \leqslant x| \cdots, X_{n-2}, X_{n-1} \text{ and } \cdots, J_{n-2}, J_{n-1} = i, J_n = j] = P[X_n \leqslant x|J_{n-1} = i, J_n = j].$$

It is assumed that observations begin with X_1 and that the process was in its stationary condition on day 0 (i.e., $P[J_0 = i] = \pi_i$).

Katz (1977b) obtained

$$\mu = E[X_n] = \sum_{i,j} \pi_i P_{ij} \mu_{ij}$$

and

$$\sigma^2 = \mathrm{Var}(X_n) = \sum_{i,j} \pi_i P_{ij} m_{ij} - \mu^2,$$

where

$$\mu_{ij} = E[X_n|J_{n-1} = i, J_n = j]$$

and

$$m_{ij} = E[X_n^2|J_{n-1} = i, J_n = j].$$

Let $S_r = X_1 + \cdots + X_r$. Then as r goes to infinity the distribution of $(S_r - r\mu)/\gamma\sqrt{r}$ converges to the standard normal distribution where $\gamma^2 = \sigma^2 + 2\boldsymbol{\pi}'\mathsf{Q}(\mathsf{Z} - \mathsf{F})\mathsf{Q}\mathbf{e}$. The matrix $\mathsf{Q} = (P_{ij}\mu_{ij})$, $\mathbf{e}$ is a column vector of ones, $\mathsf{F} = \mathbf{e}\boldsymbol{\pi}'$, and $\mathsf{Z} = [\mathsf{I} - (\mathsf{P} - \mathsf{F})]^{-1}$.

3. Using the model to determine variability and climatic change

To investigate climatic change with regard to precipitation it is necessary to know how much of the year-to-year variation is due to day-to-day variability. This may be expressed by the variance components model (Graybill, 1961) $X_{st} = \mu + a_s + b_{st}$, where $s = 1, \ldots, n$ is the year and $t = 1, \ldots, m$ is the day of the year. Here the subscripts s and t are replacing the single subscript n used in the previous section so that days can be identified with the appropriate year. The a_s and b_{st} are random variables with zero mean and variances σ_a^2 and σ_b^2. The null hypothesis of no change from year to year is equivalent to $\sigma_a^2 = 0$. To test this hypothesis, the variance of the yearly averages as estimated by $\sum_{s=1}^{n} (X_{s\cdot} - X_{\cdot\cdot})^2/(n-1)$, where $X_{s\cdot} = (\sum_{t=1}^{m} X_{st})/m$ and $X_{\cdot\cdot} = (\sum_{s=1}^{n} X_{s\cdot})/n$ is compared to a similar estimate based on within-year observations assuming the null hypothesis is true. If the X_{st} are independent and a_s and b_{st} have normal distributions, the denominator is $\sum_{s,t} (X_{st} - X_{s\cdot})^2/mn(m-1)$ and under the null hypothesis the ratio has an F distribution with $n - 1$ and $n(m-1)$ degrees of freedom.

For the case of dependence, the analysis can still be performed. The denominator must be replaced by an estimate of

$$\mathrm{Var}(X_{s\cdot}) = \mathrm{Var}(\sum_{t=1}^{m} X_{st}/m) = \frac{1}{m^2}\mathrm{Var}(\sum_{t=1}^{m} X_{st}).$$

Under the null hypothesis the value of X_{st} does not depend on s and therefore an estimate of $\mathrm{Var}(\sum_{t=1}^{m} X_{st})$ using data from all years is valid. Using the model of Section 2, an estimate of $\mathrm{Var}(\sum_{t=1}^{m} X_{st})$ is given by $\gamma^2 m$ and so $\mathrm{Var}(X_{s\cdot})$ may be estimated by γ^2/m. The appropriate test statistic is then $\chi^2 = m \sum_{s=1}^{n} (X_{s\cdot} - X_{\cdot\cdot})^2/\gamma^2$ and under the null hypothesis it will (as $m \to \infty$) have a chi-square distribution with $n - 1$ degrees of freedom. The large sample sizes encountered in these investigations make it reasonable to assume that γ^2 is known, not random.

TABLE 1. Analysis of Medford, Oregon data.

	Independence	One-day dependence	Two-day dependence
Transition probabilities (those not given are 0)		$P_{00} = 1 - P_{01} = 0.82871$ $P_{10} = 1 - P_{11} = 0.44272$	$P_{00} = 1 - P_{01} = 0.85383$ $P_{12} = 1 - P_{13} = 0.47352$ $P_{20} = 1 - P_{21} = 0.67732$ $P_{32} = 1 - P_{33} = 0.40609$
Stationary probabilities	$P(\text{rain}) = 0.27392$	$\pi_0 = 0.72103$ $\pi_1 = 0.27897$	$\pi_0 = 0.58432$ $\pi_1 = \pi_2 = 0.12610$ $\pi_3 = 0.16348$
Conditional means (mm) (those not given are 0)	$\mu_1 = 5.38861$	$\mu_{01} = 3.90017$ $\mu_{11} = 6.59994$	$\mu_{01} = 3.66700$ $\mu_{13} = 5.66776$ $\mu_{21} = 4.42417$ $\mu_{33} = 7.27304$
Conditional second moments (mm^2) (those not given are 0)	$m_1 = 100.85708$	$m_{01} = 49.05151$ $m_{11} = 137.98036$	$m_{01} = 43.43862$ $m_{13} = 89.65789$ $m_{21} = 61.66439$ $m_{33} = 172.87707$
Unconditional mean (μ)	1.47605	1.50774	1.57556
Unconditional variance (σ^2)	25.06376	25.23866	26.47091
Variance of seasonal averages (γ^2)	25.06376	35.89670	44.49669
Chi-square (28 df)*	68.61	47.70	38.65
p-value	0.00003	0.011	0.087

* Chi-square $= 29 \sum_{s=1}^{29} (X_{s\cdot} - X_{\cdot\cdot})^2/\gamma^2 = 1719.59/\gamma^2$.

4. An example

Daily rainfall data (in mm) for Medford, Oregon, were collected for the months of September, October and November from 1948 through 1976. The transition probabilities were assumed to be stable from day to day and from year to year. These assumptions were tested as outlined by Anderson and Goodman (1957) and both stationarity assumptions were upheld. It also was assumed that the amount of precipitation on a given day does not depend on the amount of precipitation on the previous day (or days, if $k > 1$). Following Katz (1977a), the correlation coefficients of the natural logarithms of precipitation amounts were calculated. All were found to be near zero (0.19 for first order and 0.14, 0.12, and 0.18 for the three second-order cases 001, 101 and 111) which indicates no significant dependence.

TABLE 2. Estimating the order of dependence.

Order(k)	SBC(k)
0	324.67
1	−20.30
2	−54.51*
3	−41.68

* The Schwarz Bayesian criterion selects the order for which $\text{SBC}(k) = \eta_k - (16 - 2^k)\ln(2523)$ is smallest. η_k is $-2\ln\lambda_k$, where λ_k is the likelihood ratio test statistic for order k vs order 4.

Table 1 gives the relevant estimated quantities for the model with independence and for the models with one- and two-day dependence. The first two models yield chi-square values that are clearly significant. However, the Schwarz Bayesian Criterion (Schwarz, 1978; Katz, 1979) indicates the appropriate model involves 2-day dependence (Table 2). With this model the significance vanishes because the extra variance is accounted for in the model.

5. Further directions

This analysis of variance approach will be used to determine whether the seasonal precipitation climates of 104 United States stations have changed during the 29-year period, December 1947–November 1976. For those stations where no change is found, the variance estimated using the Katz model will be interpreted as the amount of variability that can be expected within an unchanging climate. For those stations where change is found, upper air circulation data will be investigated to determine relationships between precipitation variability and circulation variability.

REFERENCES

Anderson, T. W., and L. A. Goodman, 1957: Statistical inference about Markov chains. *Ann. Math. Stat.*, **28**, 89–110.

Granger, O., 1979: Increasing variability in California precipitation. *Ann. Assoc. Amer. Geogr.*, **69**, 533–543.

Graybill, F. A., 1961: *An Introduction to Linear Statistical Models,* Vol. 1. McGraw-Hill, 463 pp.

Jenkinson, A. F., 1957: Relations between standard deviations of daily, 5-day, 10-day and 30-day mean temperatures. *Meteor. Mag.*, **86**, 169–176.

Jones, R. H., 1975: Estimating the variance of time averages. *J. Appl. Meteor.*, **14**, 159–163.

Katz, R. W., 1977a: Precipitation as a chain-dependent process. *J. Appl. Meteor.*, **16**, 671–676.

——, 1977b: An application of chain-dependent processes to meteorology. *J. Appl. Prob.*, **14**, 598–603.

——, 1979: Estimating the order of a Markov chain: Another look at the Tel Aviv rainfall data. *Preprints Sixth Conference on Probability and Statistics in Atmospheric Sciences,* Banff, Amer. Meteor. Soc., 217–221.

Leith, C. E., 1973: The standard error of time-averaged estimates of climatic means. *J. Appl. Meteor.*, **12**, 1066–1069.

Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.*, **104**, 942–952.

——, and D. J. Shea, 1978: Estimates of the natural variability of time-averaged temperatures over the United States. *Mon. Wea. Rev.*, **106**, 1695–1703.

Schwarz, G., 1978: Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.

Skaggs, R. H., 1978: Climatic change and persistence in western Kansas. *Ann. Assoc. Amer. Geogr.*, **68**, 73–80.