

Disaggregative Invariance of Daily Precipitation

ROMAN KRZYSZTOFOWICZ

Department of Systems Engineering and Division of Statistics, University of Virginia, Charlottesville, Virginia

THOMAS A. POMROY

Department of Systems Engineering, University of Virginia, Charlottesville, Virginia

(Manuscript received 28 February 1996, in final form 18 October 1996)

ABSTRACT

Disaggregative invariance refers to stochastic independence between the total precipitation amount and its temporal disaggregation. This property is investigated herein for areal average and point precipitation amounts accumulated over a 24-h period and disaggregated into four 6-h subperiods. Statistical analyses of precipitation records from 1948 to 1993 offer convincing empirical evidence against the disaggregative invariance and in favor of the conditional disaggregative invariance, which arises when the total amount and its temporal disaggregation are conditioned on the timing of precipitation within the diurnal cycle.

The property of conditional disaggregative invariance allows the modeler or the forecaster to decompose the problem of quantitative precipitation forecasting into three tasks: (i) forecasting the precipitation timing; (ii) forecasting the total amount, conditional on timing; and (iii) forecasting the temporal disaggregation, conditional on timing. Tasks (ii) and (iii) can be performed independently of one another, and this offers a formidable advantage for applications.

1. Introduction

In broad terms, disaggregative invariance refers to some form of stochastic independence between the total precipitation amount and its temporal disaggregation. This article investigates the existence and the nature of disaggregative invariance for the 24-h precipitation amount disaggregated into four 6-h subperiods. This particular precipitation process is of interest because it is the predictand of probabilistic quantitative precipitation forecasts (QPFs) for river basins. A methodology for preparing QPFs has been tested operationally in the Weather Service Forecast Office in Pittsburgh, Pennsylvania, since August 1990 and is currently being developed into an operational prototype for dissemination to other offices of the National Weather Service (NWS).

The format of the predictand is dictated by (i) feasibility of preparing QPFs operationally and (ii) requirements of a hydrometeorological forecasting system that will transform probabilistic forecasts of precipitation into probabilistic forecasts of river stages (Krzysztofowicz 1993). The *conditional disaggregative invariance* of the 24-h precipitation, which this study documents through extensive statistical analyses, plays a

pivotal role in designing the forecasting methodology (Krzysztofowicz et al. 1993) and the local climatic guidance (Krzysztofowicz and Sigrest 1997). It may also be of interest to hydrometeorologists who model precipitation processes for purposes other than operational forecasting.

The next section lays down the methodological background. Section 3 reviews past studies related to disaggregative invariance of precipitation. Section 4 analyzes the marginal disaggregative invariance that facilitates the understanding of statistical disaggregation. Analyses that lead to the main finding, the conditional disaggregative invariance, are reported in section 5. The last section summarizes all findings.

2. Methodological background

a. Predictand

Consider a fixed period beginning at a designated hour of the day and divided into n subperiods of equal or unequal length. Precipitation amounts to be defined can be either observed at a rain gauge or calculated as areal averages based on rain gauge observations.

Let W denote the precipitation amount accumulated during a wet period. Let W_i denote the precipitation amount accumulated during the i th subperiod, $i \in \{1, \dots, n\}$. Thus,

Corresponding author address: Prof. Roman Krzysztofowicz, University of Virginia, Thornton Hall, SE, Charlottesville, VA 22903.

$$W > 0; \quad W_i \geq 0, \quad i = 1, \dots, n;$$

$$W_1 + \dots + W_n = W. \quad (1)$$

For each i define a variate, $\Theta_i = W_i/W$, representing a fraction of the total amount accumulated during subperiod i . Thus,

$$0 \leq \Theta_i \leq 1, \quad i = 1, \dots, n;$$

$$\Theta_1 + \dots + \Theta_n = 1. \quad (2)$$

The vector of fractions $(\Theta_1, \dots, \Theta_n)$ defines the temporal disaggregation of the total precipitation amount accumulated during the period into n subperiods. Because one of the fractions can always be expressed in terms of the remaining fractions through the unit sum constraint, only $n - 1$ fractions must be forecast.

The predictand is the vector $(W; \Theta_1, \dots, \Theta_n)$. The property to be investigated is the stochastic dependence between the vector of fractions $(\Theta_1, \dots, \Theta_n)$ and the amount W . The structure of this dependence is of paramount import to formulating probability models of the predictand and designing methodologies for PQPF. In the extreme case, if $(\Theta_1, \dots, \Theta_n)$ turns out to be stochastically independent of W , the problem could be decomposed into two independent tasks: (i) forecasting the total precipitation amount W , and (ii) forecasting the temporal disaggregation $(\Theta_1, \dots, \Theta_n)$.

b. Disaggregative invariance

The stochastic dependence between the vector of fractions $(\Theta_1, \dots, \Theta_n)$ and the amount W is characterized by a family of conditional generalized probability density functions that, when evaluated at an observation $(\Theta_1 = \theta_1, \dots, \Theta_n = \theta_n)$ and conditioned on the hypothesis $W = \omega$, are denoted $\eta(\theta_1, \dots, \theta_n | \omega)$. The term generalized density stems from the fact that each Θ_i , $i = 1, \dots, n$, is a discrete-continuous variate, as will become apparent soon. Because one fraction can be eliminated through the unit sum constraint (2), and the sample space of each fraction is the unit interval $[0, 1]$, the density $\eta(\cdot, \dots, \cdot | \omega)$ is $(n - 1)$ -variate, and its support is the $(n - 1)$ -dimensional simplex $[0, 1]^{n-1}$.

DEFINITION 1. *The predictand $(W; \Theta_1, \dots, \Theta_n)$ possesses the property of disaggregative invariance if and only if the vector of fractions $(\Theta_1, \dots, \Theta_n)$ is stochastically independent of the amount W ; that is, at every point $(\omega; \theta_1, \dots, \theta_n)$ one can write*

$$\eta(\theta_1, \dots, \theta_n | \omega) = \eta(\theta_1, \dots, \theta_n). \quad (3)$$

Various procedures for testing the disaggregative invariance can be found in the literature (Aitchison 1986) for the case when each Θ_i is a strictly continuous variate, constrained by $0 < \Theta_i < 1$. Because in the case of precipitation $0 \leq \Theta_i \leq 1$ and values 0 and 1 can be observed with nonnegligible probabilities, different testing procedures are required. To develop a comprehensive understanding of the statistical behavior of frac-

tions, testing progresses through two stages. First, analyses are performed to test marginal independence of Θ_i from W for each $i = 1, \dots, n$ (section 4). Next, an analysis according to (3) is performed to test joint independence of $(\Theta_1, \dots, \Theta_n)$ from W (section 5).

All testing procedures employ empirical distributions and nonparametric statistics. Hence the conclusions are mostly free from assumptions about probability models and estimators of their parameters.

c. Test cases and data

The analyses reported herein were performed in support of the PQPF methodology that has been tested operationally by the Weather Service Forecast Office in Pittsburgh since August 1990. Forecasts are made for a 24-h period divided into four 6-h subperiods. The beginning hours are 0000 and 1200 UTC. For each beginning hour, the following predictands were analyzed. Areal average precipitation amounts were analyzed for (i) the Upper Allegheny River basin above the Kinzua Dam, which covers 5853 km² (2260 square miles) in Pennsylvania and New York, and (ii) the Lower Monongahela River basin above Connellsville, which covers 3429 km² (1324 square miles) in Pennsylvania and Maryland. Point precipitation amounts were analyzed for rain gauges at (i) Bolivar, New York, located within the Upper Allegheny River basin, and (ii) Confluence, Pennsylvania, located within the Lower Monongahela River basin.

The source data consist of hourly precipitation amounts recorded by rain gauges within or near the basins of interest from 1948 to 1993. For each hour, the basin average precipitation amount is estimated as a weighted sum of rain gauge observations with weights determined via the Thiessen method (Linsley et al. 1975). A realization of the random vector (W_1, \dots, W_n) is obtained by summing up the hourly basin average precipitation amounts over appropriate subperiods. The number of realizations is thus equal to the number of periods with complete hourly observations. Realizations are assumed to form a random sample. To obtain samples of adequate length, analyses are performed for six seasons, each covering two months: January–February, March–April, May–June, July–August, September–October, and November–December.

Whereas general conclusions drawn in this paper are based on all cases, data analyses and test results are illustrated for the Lower Monongahela River basin, season March–April, 24-h period beginning at 1200 UTC, and disaggregation into four 6-h subperiods: 1200–1800, 1800–0000, 0000–0600, 0600–1200. Hourly basin average precipitation amounts were estimated from observations at six stations with the following weights: Sines Deep Creek (0.24), New Germany (0.11), Confluence (0.26), Glencoe (0.10), Connellsville (0.14), and Boswell (0.15). All amounts are reported herein in inches.

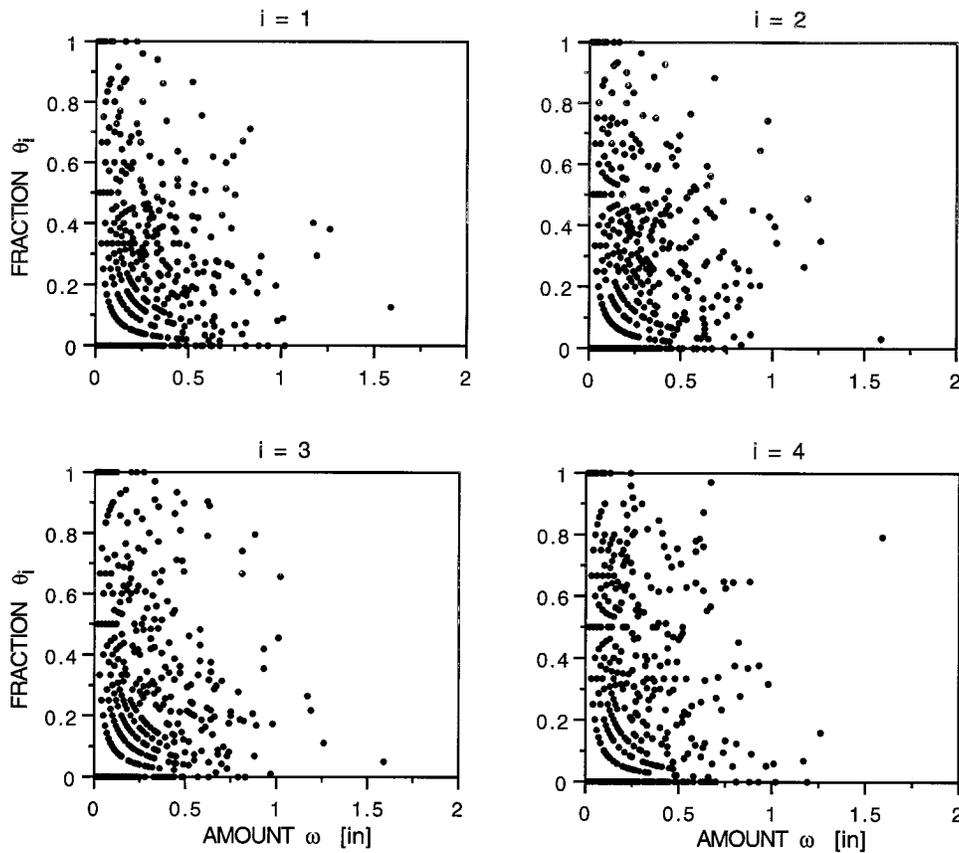


FIG. 1. Scatterplots of fraction θ_i versus precipitation amount ω ; Monongahela Basin, March–April, from 1200 UTC.

3. Prior studies

a. Disaggregation within storm

Several authors investigated characteristics of precipitation related to the concept of temporal disaggregation. Huff (1967) analyzed 261 heavy storms having duration from 3 to 48 h and producing precipitation amounts in excess of 0.5" over an area of 400 square miles, or 1" at one or more of 49 rain gauges in east-central Illinois. No stratification of storms according to season or time of the day was considered. Of central interest was the *dimensionless mass curve*—a plot of the cumulative percent of precipitation amount versus the cumulative percent of storm duration. In essence, such a plot depicts the temporal disaggregation of the total amount within the unitized storm duration. The plots were classified into four patterns, depending upon whether the largest fraction of precipitation occurred in the first, second, third, or fourth quarter of the storm duration. Frequencies of the four patterns were estimated: unconditional, conditional on precipitation type (continuous rain, snow, rainshower, thunderstorm, hail, and combinations thereof), and conditional on storm duration (<12, 12–24, >24 h). Stochastic dependence was revealed in both cases. In particular, disaggregation patterns of type 1

and 2 occurred most frequently with durations less than 12 h, type 3 occurred most frequently with durations 12–24 h, whereas type 4 occurred most frequently with durations greater than 24 h. Thus according to Huff's study, disaggregation of the amount within the unitized storm duration depends stochastically on the duration. And because duration is known to be dependent on the total amount, the study supports the hypothesis that disaggregation of the amount within the unitized storm duration depends on the amount. In fact, such dependence glares from a display of five distinct dimensionless mass curves, each for a different total amount, shown by Hjelmfelt (1981, Fig. 1) and originally published by the United States Weather Bureau (1947).

The concept of the dimensionless mass curve was recently revisited by Koutsoyiannis and Foufoula-Georgiou (1993), who proposed a stochastic model to represent the ensemble of mass curves for storms occurring over a point or an area within a month or season. The dimensionless mass curve was defined to be independent of both storm duration and precipitation amount, and this independence was considered to be the normative requirement for precipitation models. This requirement is contrary to empirical evidence: it overlooks the fact that Huff's classification of the dimensionless mass

curves into four patterns amounted to a somewhat crude but effective conditioning of the curves on storm duration (Huff 1967, 1008, 1015).

The conclusions from the aforementioned studies do not extend directly to our definition of the temporal disaggregation, however. A direct comparison is impeded by two features of the dimensionless mass curve. First, the length of subperiods on the real-time scale depends on storm duration. Second, stochastic properties of precipitation are assumed to be independent of the timing of the storm within the day. Both features are unlike those of our model wherein the subperiods are defined by the hours of the day and thus their length and timing are fixed.

The advantage of fixed subperiods is that the timing of precipitation can be captured. This is critical for the purpose of forecasting in light of the overwhelming empirical evidence that the diurnal cycle is a significant predictor of precipitation variability, especially of convective origins (Wallace 1975). The disadvantage of fixed subperiods as long as 6 h is that characteristics of individual storms, especially those of short duration, cannot be captured. This disadvantage could be alleviated, of course, by resorting to shorter, say hourly, subperiods. However, the current state of the art precludes skillful forecasting of hourly precipitation amounts over periods of 24 h. Hence a daily precipitation model with hourly subperiods, while technically feasible to develop and estimate, would not provide a useful guidance for the forecaster who must prepare a 24-h PQPF.

b. Disaggregation among storms

Another stochastic model, which incorporates some form of the disaggregative invariance, was developed by Hershenson and Woolhiser (1987) for the purpose of simulating point precipitation from summer thunderstorms occurring in the Southwest. The model disaggregates daily precipitation amount into amounts produced by individual storms. The number of storms per day and the duration of each storm are generated from distributions. In effect, the daily amount is disaggregated into a random number of wet subperiods having random duration. Most significantly, all fractions are strictly positive and are assumed to have the following properties: (i) the number of positive fractions (equivalently, the number of storms per day) and the daily amount are stochastically dependent, and (ii) the vector of positive fractions, given their number, and the daily amount are stochastically independent. The first assumption received some limited testing on data; the second assumption received no empirical justification. Most interesting to us, however, is the fact that these two assumptions are, in a sense, compatible with the notion of conditional disaggregative invariance, which is studied in section 5. Thus our findings are not totally without precedence.

4. Marginal analyses

a. Marginal generalized density

Subperiod i may be dry, $\Theta_i = 0$, may receive the total amount for the period, $\Theta_i = 1$, or a fraction thereof, $0 < \Theta_i < 1$. The probability function defined over these events, conditional on the amount $W = \omega$, is specified by probabilities

$$p_{i0}(\omega) = P(\Theta_i = 0 | W = \omega),$$

$$p_{i1}(\omega) = P(\Theta_i = 1 | W = \omega), \tag{4}$$

which imply $1 - p_{i0}(\omega) - p_{i1}(\omega) = P(0 < \Theta_i < 1 | W = \omega)$. Next define the distribution of Θ_i restricted to $0 < \Theta_i < 1$ and conditional on $W = \omega$: $h_i(\theta_i | \omega) = P(\Theta_i \leq \theta_i | 0 < \Theta_i < 1, W = \omega)$. The corresponding density takes on values $h_i(\theta_i | \omega) > 0$ if $0 < \theta_i < 1$ and $h_i(\theta_i | \omega) = 0$ otherwise.

With δ denoting the Dirac delta function [whose defining property is $\delta(0) = \infty$ and $\delta(\tau) = 0$ for $\tau \neq 0$], the marginal generalized density of Θ_i , conditional on $W = \omega$, can be expressed as follows:

$$\eta_i(\theta_i | \omega) = \delta(\theta_i)p_{i0}(\omega) + \delta(\theta_i - 1)p_{i1}(\omega) + h_i(\theta_i | \omega) \times [1 - p_{i0}(\omega) - p_{i1}(\omega)]. \tag{5}$$

DEFINITION 2. *The predictand ($W; \Theta_1, \dots, \Theta_n$) possesses the property of marginal disaggregative invariance if and only if each fraction Θ_i is stochastically independent of the amount W ; that is, for every $i = 1, \dots, n$ and each point (ω, θ_i) one can write $\eta_i(\theta_i | \omega) = \eta_i(\theta_i)$.*

Scatterplots of θ_i versus ω , such as those shown in Fig. 1, have generally the following pattern: (i) points on the lower and upper boundaries have different scatter, suggesting that $p_{i0}(\omega)$ and $p_{i1}(\omega)$ are conditioned on ω ; (ii) points in the interior form a cloud, having a non-uniform density and a sloping right fringe, which suggests that $h_i(\theta_i | \omega)$ is conditioned on ω . Further examination of these hypotheses is described next. (The curvy layers of points in the left corners are the artifact of the coarse precision of hourly rainfall measurements; this artifact is explained in appendix A.)

b. Discrete part of fraction

To investigate the dependence of the discrete part of Θ_i on the continuous variate W , it is advantageous to invert the conditioning. Toward this end, define the marginal probability function of Θ_i : $\xi_{i0} = P(\Theta_i = 0)$, $\xi_{i1} = P(\Theta_i = 1)$, and $1 - \xi_{i0} - \xi_{i1} = P(0 < \Theta_i < 1)$. Next, define three conditional distributions of W :

$$G_{i0}(\omega) = P(W \leq \omega | \Theta_i = 0),$$

$$G_{i1}(\omega) = P(W \leq \omega | \Theta_i = 1),$$

$$G_i(\omega) = P(W \leq \omega | 0 < \Theta_i < 1). \tag{6}$$

With the densities corresponding to these distributions

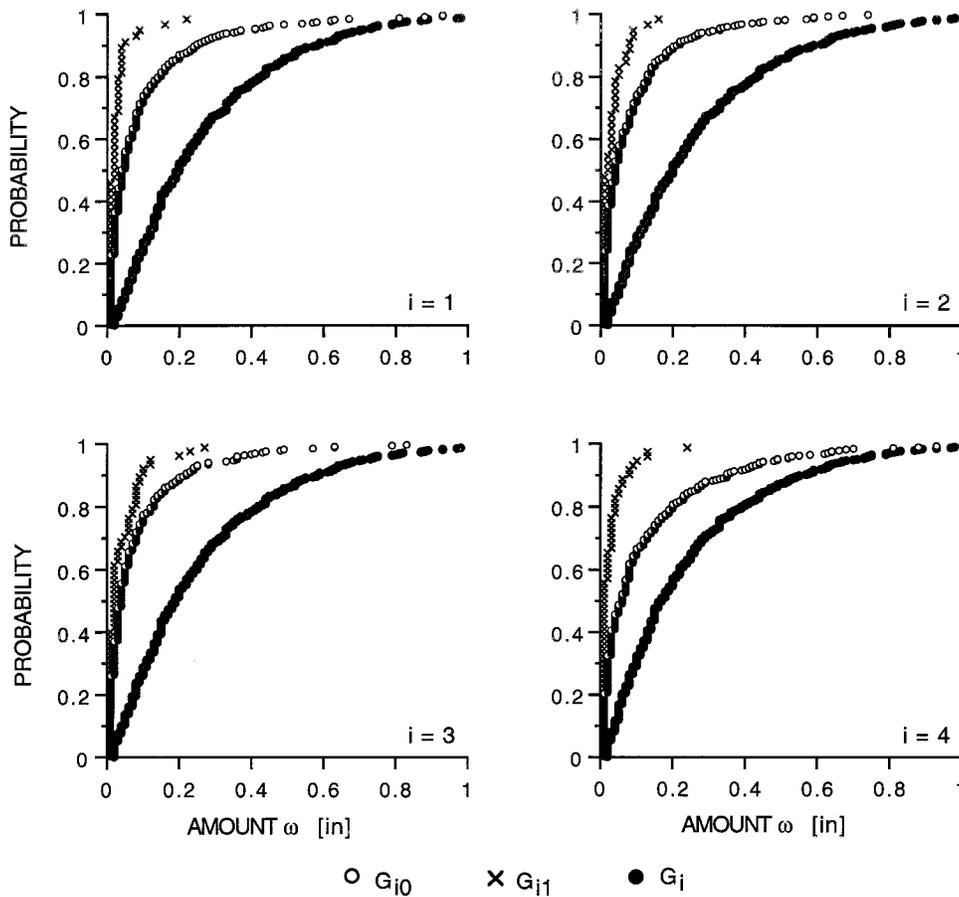


FIG. 2. Empirical distributions of the precipitation amount W , conditional on values of the fraction: $\Theta_i = 0$, $\Theta_i = 1$, and $0 < \Theta_i < 1$; Monongahela Basin, March–April, from 1200 UTC.

denoted g_{i0} , g_{i1} , and g_i , the marginal density of W follows from the total probability law:

$$g(\omega) = g_{i0}(\omega)\xi_{i0} + g_{i1}(\omega)\xi_{i1} + g_i(\omega) \times (1 - \xi_{i0} - \xi_{i1}). \tag{7}$$

Finally, Bayes' theorem yields

$$p_{ij}(\omega) = \frac{g_{ij}(\omega)\xi_{ij}}{g(\omega)}, \quad j = 0, 1. \tag{8}$$

It is now apparent that the probability function of Θ_i is not conditioned on ω , that is, $p_{ij}(\omega) = \xi_{ij}$ for $j = 0, 1$, if and only if $g_{i0} = g_{i1} = g_i$.

To test this hypothesis, empirical estimates of the distributions G_{i0} , G_{i1} , and G_i are compared. Figure 2 shows these estimates for $i = 1, 2, 3, 4$. Although a formal statistical test could be administered, the conclusion is apparent from the plots. For every i , the distributions are distinct and form a dominating sequence wherein $G_{i1}(\omega) > G_{i0}(\omega) > G_i(\omega)$ for all $\omega > 0$. Thus, the discrete part of fraction Θ_i is stochastically dependent on the amount W .

c. Continuous part of fraction

The independence of the continuous part of Θ_i from W is ascertained if at each point (ω, θ_i) , such that $0 < \theta_i < 1$, one can write $h_i(\theta_i|\omega) = h_i(\theta_i)$. To test this hypothesis, the original variates are transformed into standard normal variates. Toward this end, construct the empirical marginal distributions, restricted to the continuous part of the fraction:

$$K_i(\theta_i) = P(\Theta_i \leq \theta_i | 0 < \Theta_i < 1),$$

$$G_i(\omega) = P(W \leq \omega | 0 < \Theta_i < 1). \tag{9}$$

Next process each observation (ω, θ_i) through the normal quantile transform (NQT):

$$z_i = Q^{-1}(K_i(\theta_i)),$$

$$v = Q^{-1}(G_i(\omega)), \tag{10}$$

where Q^{-1} is the inverse of the standard normal distribution Q . The resultant pair (v, z_i) is an observation of variates (V, Z_i) whose marginal distributions are standard normal.

The scatterplots of z_i versus v , for $i = 1, 2, 3, 4$, shown in Fig. 3, take on remarkable, diamond-like shapes. They can be traced to two unrelated effects. The left triangle of the diamond and its sloping layers of points are caused by the crude precision of hourly rainfall measurements, as explained in appendix A. This is a spurious effect that distorts the dependence structure between fraction Θ_i and amount W when observations of W are low. The right triangle of the diamond reflects a property that will be called the *temporal smoothing of high precipitation*: as $V = v$ becomes high, observations of Z_i become less dispersed and tend toward zero. Inasmuch as NQT is strictly increasing, and zero of Z_i corresponds to the median of Θ_i , this effect can be restated as follows. As $W = \omega$ becomes high, observations of Θ_i become less dispersed and tend toward the median. This effect is seen for every $i = 1, 2, 3, 4$. Hence one may conclude that as the amount $W = \omega$ becomes high, the temporal disaggregation of ω becomes less variable and tends toward the median of fractions $(\Theta_1, \Theta_2, \Theta_3, \Theta_4)$. Still another way of describing the temporal smoothing of high precipitation is this: when the amount becomes high, it is more likely that the precipitation will be evenly spread out over all four subperiods; it is therefore less likely that a single subperiod will receive little or most of this high amount. Mathematically, the right triangle of the diamond in Fig. 3 implies that $h_i(\theta_i|\omega) \neq h_i(\theta_i)$ and that, therefore, the continuous part of fraction Θ_i is stochastically dependent on the amount W .

In summary, the marginal analyses offer a convincing empirical evidence against the disaggregative invariance of the 24-h precipitation. They also uncover an obstacle to modeling the dependence between Θ_i and W based on standard measurements of hourly precipitation amounts whose crude precision produces a spurious dependence structure when the total amounts are low.

5. Multivariate analyses

a. Precipitation timing

When the precipitation occurs during a period, it occurs in one or more of the n subperiods. The occurrence or nonoccurrence of precipitation in subperiods defines the timing T . There are $2^n - 1$ possible realizations of T , denoted t and called timing patterns. They can be formally defined in terms of fractions $(\Theta_1, \dots, \Theta_n)$, as shown in Table 1 for $n = 4$. A particular pattern is identified by the concatenation of indices of subperiods that are wet. For example, if all precipitation for the period occurs in subperiods 1 and 3, then $\Theta_1 > 0, \Theta_3 > 0, \Theta_1 + \Theta_3 = 1$, and $T = 13$.

The timing T plays a key role in further analysis. First, it allows us to capture an effect of the diurnal cycle on the distribution of the predictand. Second, it allows us to decompose the distribution of the predictand into discrete and continuous parts, thereby faci-

tating the test of the disaggregative invariance hypothesis (3). To wit, the conditional generalized density of fractions can be decomposed as follows:

$$\eta(\theta_1, \dots, \theta_n|\omega) = \sum_t \phi(\theta_1, \dots, \theta_n|T = t, W = \omega) \times P(T = t|W = \omega), \tag{11}$$

where $\phi(\cdot, \dots, \cdot|T = t, W = \omega)$ is a joint density of fractions $(\Theta_1, \dots, \Theta_n)$, conditional on timing $T = t$ and amount $W = \omega$, and $P(T = t|W = \omega)$ is the probability of timing $T = t$, conditional on amount $W = \omega$.

b. Conditional disaggregative invariance

The decomposition of η according to (11) implies that the hypothesis of disaggregative invariance can also be decomposed into two parts: (i) independence of timing T from amount W , and (ii) independence of the vector of fractions $(\Theta_1, \dots, \Theta_n)$ from amount W , conditional on timing $T = t$. Each part of the hypothesis can hold independently of the other part. Because the second part involves fractions, it is of particular interest herein.

DEFINITION 3. *The predictand $(W; \Theta_1, \dots, \Theta_n)$ possesses the property of conditional disaggregative invariance if and only if the vector of fractions $(\Theta_1, \dots, \Theta_n)$ is stochastically independent of the amount W , conditional on timing $T = t$, for all realizations t ; that is, at every point $(\omega; \theta_1, \dots, \theta_n)$ and for every t one can write*

$$\phi(\theta_1, \dots, \theta_n|T = t, W = \omega) = \phi(\theta_1, \dots, \theta_n|T = t). \tag{12}$$

In the subsequent sections, each element of (11) is structured and then tested for independence from ω .

c. Probability function of timing

To investigate the dependence of timing T on amount W , it is advantageous to invert the conditioning. For this purpose, define a conditional distribution of the amount, $G_t(\omega) = P(W \leq \omega|T = t)$, with the corresponding density denoted g_t . The marginal density of W is thus specified by

$$g(\omega) = \sum_t g_t(\omega)P(T = t), \tag{13}$$

and the conditional probability function of T is specified by

$$P(T = t|W = \omega) = \frac{g_t(\omega)P(T = t)}{g(\omega)}. \tag{14}$$

Thus, T is stochastically independent of W if and only if $g_t = g$ for every t .

For the purpose of testing, the null hypothesis is restated in terms of the distributions: $G_t = G$ for every t . Empirical estimates of G_t and G are then used in the two-sample Kolmogorov–Smirnov test (Lindgren 1976, 494). Table 2 presents test results as well as statistics

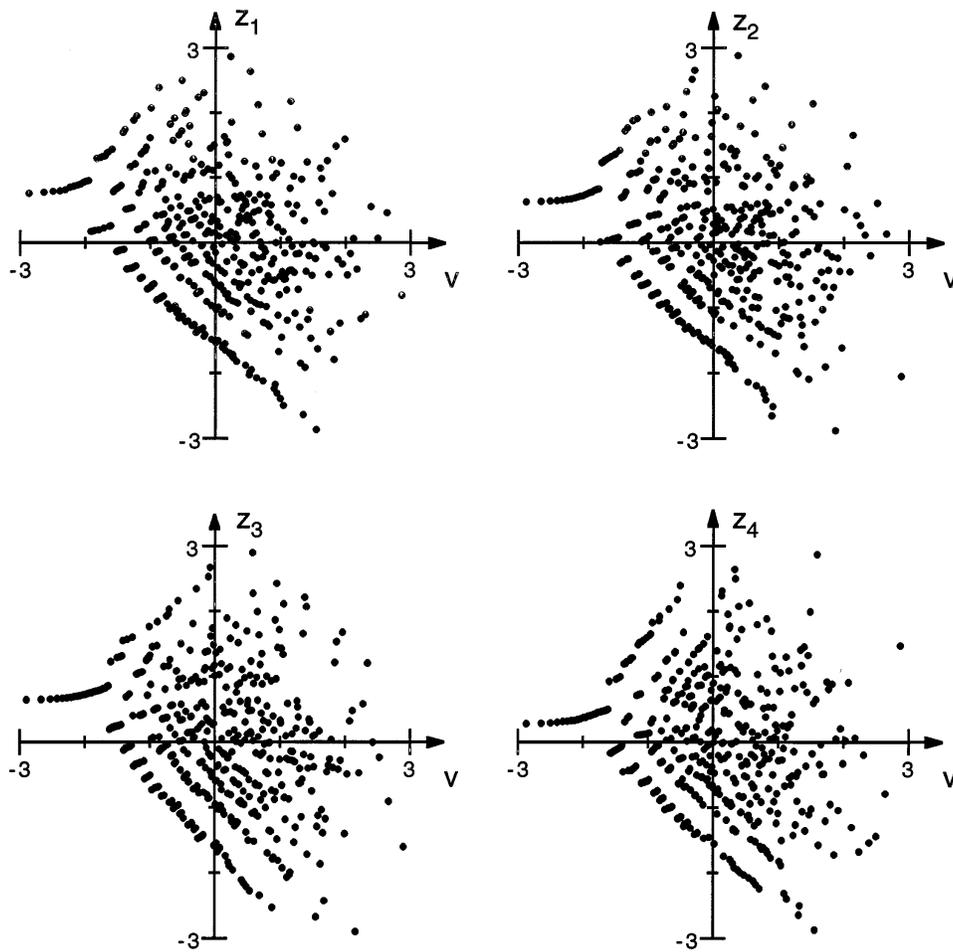


FIG. 3. Scatterplots of transformed fraction z_i versus transformed precipitation amount v , when the original fraction θ_i is restricted to the continuous part, $0 < \theta_i < 1$; Monongahela Basin, March–April, from 1200 UTC.

TABLE 1. Definition of precipitation duration, timing pattern, and conditional restricted densities of fractions for four subperiods.

Duration index* d	Timing pattern t	Definition in terms of fractions		Conditional restricted density of fractions
1	1	$\Theta_1 = 1$		$\delta(\theta_1 - 1)$
1	2	$\Theta_2 = 1$		$\delta(\theta_2 - 1)$
1	3	$\Theta_3 = 1$		$\delta(\theta_3 - 1)$
1	4	$\Theta_4 = 1$		$\delta(\theta_4 - 1)$
2C	12	$\Theta_1, \Theta_2 > 0$	$\Theta_1 + \Theta_2 = 1$	$f_{12}(\theta_1 \omega)$
2C	23	$\Theta_2, \Theta_3 > 0$	$\Theta_2 + \Theta_3 = 1$	$f_{23}(\theta_2 \omega)$
2C	34	$\Theta_3, \Theta_4 > 0$	$\Theta_3 + \Theta_4 = 1$	$f_{34}(\theta_3 \omega)$
2N	13	$\Theta_1, \Theta_3 > 0$	$\Theta_1 + \Theta_3 = 1$	$f_{13}(\theta_1 \omega)$
2N	14	$\Theta_1, \Theta_4 > 0$	$\Theta_1 + \Theta_4 = 1$	$f_{14}(\theta_1 \omega)$
2N	24	$\Theta_2, \Theta_4 > 0$	$\Theta_2 + \Theta_4 = 1$	$f_{24}(\theta_2 \omega)$
3C	123	$\Theta_1, \Theta_2, \Theta_3 > 0$	$\Theta_1 + \Theta_2 + \Theta_3 = 1$	$f_{123}(\theta_1, \theta_2 \omega)$
3C	234	$\Theta_2, \Theta_3, \Theta_4 > 0$	$\Theta_2 + \Theta_3 + \Theta_4 = 1$	$f_{234}(\theta_2, \theta_3 \omega)$
3N	124	$\Theta_1, \Theta_2, \Theta_4 > 0$	$\Theta_1 + \Theta_2 + \Theta_4 = 1$	$f_{124}(\theta_1, \theta_2 \omega)$
3N	134	$\Theta_1, \Theta_3, \Theta_4 > 0$	$\Theta_1 + \Theta_3 + \Theta_4 = 1$	$f_{134}(\theta_1, \theta_3 \omega)$
4	1234	$\Theta_1, \Theta_2, \Theta_3, \Theta_4 > 0$		$f_{1234}(\theta_1, \theta_2, \theta_3 \omega)$

*C—consecutive subperiods; N—nonconsecutive subperiods.

TABLE 2. Test of stochastic independence between precipitation amount W and timing T ; Monongahela Basin, March–April, from 1200 UTC.

Timing pattern t	Sample size	Probability $P(T = t)$	Sample moments*		Test** of $G_t = G$		
			$E(W T = t)$	$SD(W T = t)$	K–S stat.	p value	
1	49	0.056	2.88	3.82	0.588	<0.01	R
2	47	0.054	3.28	3.39	0.496	<0.01	R
3	68	0.078	4.24	5.21	0.428	<0.01	R
4	64	0.073	3.00	3.90	0.534	<0.01	R
12	45	0.051	11.00	9.95	0.164	0.19	
23	63	0.072	19.46	23.56	0.126	>0.20	
34	61	0.070	9.20	8.24	0.213	<0.01	R
13	13	0.015	10.85	10.97	0.243	>0.20	
14	45	0.051	10.09	11.14	0.261	<0.01	R
24	15	0.017	4.87	5.45	0.562	<0.01	R
123	72	0.082	28.35	21.24	0.321	<0.01	R
234	59	0.067	21.34	19.10	0.257	<0.01	R
124	45	0.051	23.42	19.77	0.311	<0.01	R
134	57	0.065	22.60	17.25	0.331	<0.01	R
1234	174	0.198	36.90	25.22	0.455	<0.01	R
Unconditional	877		18.04	20.98			

* E –expectation; SD –standard deviation.
 ** K–S stat.–Kolmogorov–Smirnov statistic; R–rejection of the null hypothesis.

of T and W for March–April. The null hypothesis is rejected in 12 out of 15 tests at the significance level $\alpha < 0.01$. In all 90 tests (6 seasons \times 15 timing patterns), the null hypothesis is rejected in 62 cases (69%) at $\alpha < 0.01$ and in 74 cases (82%) at $\alpha \leq 0.05$. Thus the data strongly support the stochastic dependence between timing T and amount W .

The 16 cases out of 90 tests, in which the identity $G_t = G$ cannot be rejected at $\alpha > 0.05$, occur in no more than three instances per season. Inasmuch as $G(\omega)$ is an expectation of $G_t(\omega)$ for every ω , as can be seen from (13), it should not be surprising that distribution G is close to some distributions G_t that form the expectation. In other words, a uniform rejection of the null hypothesis for all t need not occur, no matter how strongly W depends on T .

The stochastic dependence between W and T could be anticipated because the 15 realizations of T comprise precipitation events of duration D equal to one, two, three, and four subperiods (equivalently, 1–6, 7–12, 13–18, and 19–24 h). Hence, the dependence between W and T could be explained indirectly in terms of the dependence between the amount W and duration D , which is well known. However, the sample estimates of conditional moments shown in Table 2 reveal a direct dependence of the amount W on timing T . We submit that the source of this dependence is the diurnal cycle of precipitation (Wallace 1975).

d. Conditional density of fractions

A realization t of timing T specifies a subset of m_t fractions ($m_t \leq n$) that take on positive values and whose sum is one; the remaining $n - m_t$ fractions take on value zero with probability one. Consequently, the conditional joint density of n fractions, $\phi(\cdot, \dots, \cdot | T = t, W = \omega)$,

is uniquely determined by an $(m_t - 1)$ -variate density with support restricted to the $(m_t - 1)$ -dimensional simplex. For $n = 4$, these restricted densities are listed in the last column of Table 1.

When $t \in \{1, 2, 3, 4\}$, all precipitation falls in one subperiod so that $\Theta_t = 1$. Thus, the restricted density of fractions degenerates to an impulse $\delta(\theta_t - 1)$, and the conditional joint density of fractions is specified as follows:

$$\phi(\theta_1, \theta_2, \theta_3, \theta_4 | T = t, W = \omega) = \begin{cases} \delta(\theta_t - 1), & \theta_t = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

When $t > 4$, the restricted density of fractions, conditional on $T = t$ and $W = \omega$, is denoted $f_t(\cdot | \omega)$. For example, when $T = 13$, positive fractions are Θ_1 and Θ_3 , the restricted density $f_{13}(\cdot | \omega)$ is univariate and applies to Θ_1 (whereas $\Theta_3 = 1 - \Theta_1$), and the conditional joint density of fractions is specified as follows:

$$\begin{aligned} \phi(\theta_1, \theta_2, \theta_3, \theta_4 | T = 13, W = \omega) \\ = \begin{cases} f_{13}(\theta_1 | \omega), & \theta_1, \theta_3 > 0, \theta_1 + \theta_3 = 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{16}$$

With the aid of Table 1, the specification of ϕ for the remaining realizations of T should be apparent.

In summary, for $n = 4$, the conditional joint density of fractions can be decomposed into four impulse functions, six univariate densities, four bivariate densities, and one trivariate density. Thus, to ascertain whether or not the conditional disaggregative invariance holds, as defined by (12), one must test whether or not the conditioning on $W = \omega$ in each of the 11 restricted densities of fractions can be eliminated so that $f_t(\cdot | \omega) = f_t(\cdot)$ for all $t > 4$.

e. Testing procedure

The test procedure employs a multivariate extension of the NQT described in section 4c. The procedure must ensure that the transformation of fractions into standard normal variates has a unique inverse that preserves the unit sum constraint (2). Toward this end, fractions are first transformed into ratios (Aitchison 1986). For a univariate restricted density, conditional on timing pattern from the set $\{t = ij: i = 1, 2, 3; j = 2, 3, 4; i < j\}$, the transform is trivial:

$$y_1 = \theta_i/\theta_j. \tag{17}$$

For a bivariate restricted density conditional on timing pattern from the set $\{t = ijk: i = 1, 2; j = 2, 3; k = 3, 4; i < j < k\}$, the transforms are

$$y_1 = \theta_{(1)}/\theta_{(3)}, y_2 = \theta_{(2)}/\theta_{(3)}, \tag{18}$$

where $(\theta_{(1)}, \theta_{(2)}, \theta_{(3)})$ is one of the three possible permutations of fractions: $(\theta_i, \theta_j, \theta_k), (\theta_k, \theta_i, \theta_j), (\theta_j, \theta_k, \theta_i)$. For the trivariate restricted density conditional on timing pattern $t = 1234$, the transforms are

$$y_1 = \theta_{(1)}/\theta_{(4)}, y_2 = \theta_{(2)}/\theta_{(4)}, y_3 = \theta_{(3)}/\theta_{(4)}, \tag{19}$$

where $(\theta_{(1)}, \theta_{(2)}, \theta_{(3)}, \theta_{(4)})$ is one of the four possible permutations of fractions: $(\theta_1, \theta_2, \theta_3, \theta_4), (\theta_2, \theta_1, \theta_4, \theta_3), (\theta_3, \theta_4, \theta_1, \theta_2), (\theta_4, \theta_3, \theta_2, \theta_1)$. The distinction between permutations is that different fraction serves as the denominator of the ratios. Each ratio is bounded below, $0 < y_s < \infty, s \in \{1, 2, 3\}$, but there are no other constraints.

For a fixed t and a fixed permutation of fractions, the available sample of fractions is transformed into a sample of ratios according to (17), (18), or (19). Then empirical marginal distributions are constructed:

$$K_{s|t}(y_s) = P(Y_s \leq y_s | T = t), \quad s \in \{1, 2, 3\},$$

$$G_t(\omega) = P(W \leq \omega | T = t). \tag{20}$$

Each joint observation of ratio and amount is processed through NQT:

$$z_s = Q^{-1}(K_{s|t}(y_s)), \quad s \in \{1, 2, 3\},$$

$$v = Q^{-1}(G_t(\omega)). \tag{21}$$

The transformed observations are used to estimate the Pearson's product-moment correlation coefficient $\gamma_s = \text{cor}(V, Z_s)$.

Under the assumption that the joint distribution of (V, Z_s) is bivariate standard normal, γ_s provides a fully efficient measure of dependence between the original variates Y_s and W (Farlie 1960). Consequently, Y_s is stochastically independent of W , conditional on $T = t$, if and only if $\gamma_s = 0$. One may also note that Spearman's rank correlation coefficient between Z_s and V , as well as between Y_s and W , is (Kruskal 1958) $\rho_s = (6/\pi) \arcsin(\gamma_s/2)$. Thus, in fact, γ_s is an ordinal measure of association.

Applying this model to the three cases yields the fol-

lowing inference procedures. In the univariate case, Y_1 is conditionally independent of W if $\gamma_1 = 0$. In the bivariate case, (Y_1, Y_2) is conditionally independent of W if $\gamma_1 = \gamma_2 = 0$. And in the trivariate case, (Y_1, Y_2, Y_3) is conditionally independent of W if $\gamma_1 = \gamma_2 = \gamma_3 = 0$. From the uniqueness of transformations (17)–(19), it follows that if a vector of ratios is conditionally independent of W , then the corresponding vector of fractions is also conditionally independent of W . Thus, we have obtained a means of testing the conditional disaggregate invariance.

Formally, the null hypothesis $\gamma_s = 0$ for $s \in \{1, 2, 3\}$ is subjected to a two-sided test (Lindgren 1976, 478) based on the sample estimate of γ_s and the sample size. The result is the p_s value, the largest significance level α at which the null hypothesis cannot be rejected.

In the univariate case, transforms (17) and (21) define a unique model of (W, Y_1) and thus there is only one p_1 value that is reported. In the bivariate case, transforms (18) and (21) define a family of three models of (W, Y_1, Y_2) , each resulting from a different permutation of fractions. The test is performed for each permutation and thus three pairs of (p_1, p_2) values are obtained. To make the acceptance of the null hypothesis convincing, p_1 and p_2 should be as large as possible. For this reason, we select the permutation for which the metric $(\sqrt{p_1} + \sqrt{p_2})$ is the highest and report the corresponding (p_1, p_2) values. In a sense, we select the model under which the dependence between (Y_1, Y_2) and W is the weakest. The rationale is that none of the models is a perfect representation of (W, Y_1, Y_2) and thus the consideration of alternative models is appropriate.

In the trivariate case, transforms (19) and (21) define a family of four models of (W, Y_1, Y_2, Y_3) , each resulting from a different permutation of fractions. Thus, there is a similar problem of choice. We select the permutation for which the metric $(\sqrt{p_1} + \sqrt{p_2} + \sqrt{p_3})$ is the highest and report the corresponding (p_1, p_2, p_3) values.

f. Effect of rainfall measurements

Prior to applying the test procedure, it is expedient to examine the scatterplots of θ_i versus ω , and z_s versus v , conditional on $T = t$. The objective is to judge the validity of assuming a bivariate normal distribution for the transformed variates. In the present case, an overriding concern is the effect of the coarse rainfall measurements, which is described in appendix A. From the examples in Fig. 4, it can be seen that the effect is still present in the form of a triangle of points for $v < 0$. Thus, one faces a dilemma. One can either eliminate observations ω smaller than a threshold, say $\omega < 0.20''$, and conduct a restricted test, or one can use all observations in an unrestricted test. In either case, the test will be approximate.

We chose the second approach, which has the following property. Under the assumptions of the test, sto-

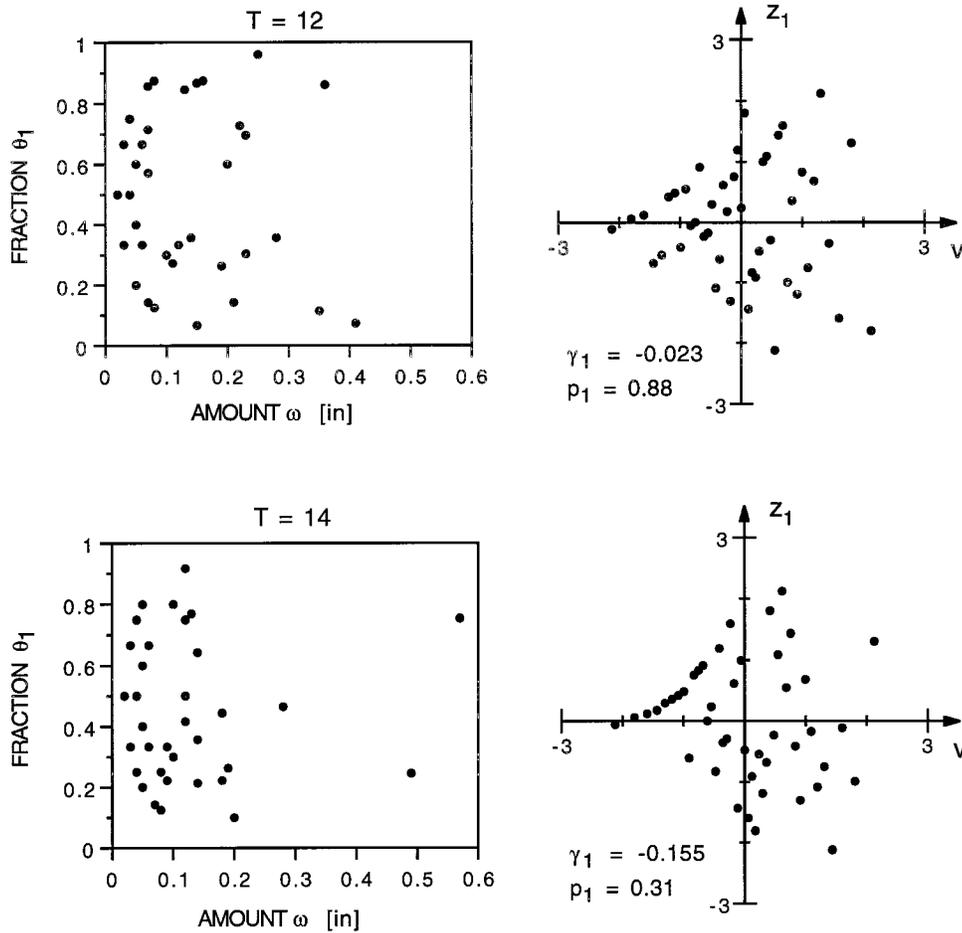


FIG. 4. Scatterplots of fraction θ_1 versus precipitation amount ω and transformed fraction z_1 versus transformed precipitation amount v , conditional on timing pattern $T = t$; Monongahela Basin, March–April, from 1200 UTC.

chastic independence of Θ_i from W is manifested by a scatterplot of z_s versus v , which is symmetric around the line $z_s = 0$. Hence, a symmetric triangle of points on the left will not affect the value of the correlation coefficient $\gamma_s = 0$. On the other hand, stochastic dependence of Θ_i on W is manifested by an asymmetric scatterplot. The asymmetry will result in a value $\gamma_s \neq 0$ and thus will be detected regardless of the triangle constraining the points on the left.

In summary, the spurious effect of coarse rainfall measurements undermines the assumptions of the test. Nonetheless, the test based on γ_s should be adequate to discriminate between cases of independence and cases of significant dependence. Scatterplots suggest that the dependence is mostly none or weak.

g. Test results

Table 3 reports complete test results for March–April. Among 17 correlation coefficients, only two have p val-

ue less than 0.10. Thus, the null hypothesis cannot be rejected in 15 cases (88%) at the significance level $\alpha = 0.10$. In fact, most p values are considerably higher than 0.10.

Table 4 reports the overall frequency of p values. In all 102 tests (6 seasons \times 17 correlation coefficients), the null hypothesis cannot be rejected in 83 cases (81%) at the significance level $\alpha = 0.10$. Thus, overall, the test supports the conditional stochastic independence between fractions and the amount.

In addition to the overall frequency, it is important to examine the actual p values and their structure. For this purpose, Fig. 5 displays a frequency function of p_1 values for the univariate distributions, a scatterplot of (p_1, p_2) values for the bivariate distributions, and profiles of (p_1, p_2, p_3) values for the trivariate distributions. Two observations are of import. First, most of the p values are considerably higher than (the already high) significance level $\alpha = 0.10$ for rejecting the null hypothesis. Second, there are no bivariate distributions with two

TABLE 3. Test of conditional disaggregative invariance; Monongahela Basin, March–April, from 1200 UTC.

Timing pattern <i>t</i>	Sample size	Permutation of fractions	Sample correlation coefficient			<i>p</i> value of two-sided test		
			γ_1	γ_2	γ_3	p_1	p_2	p_3
12	45		-0.023			0.88		
23	63		-0.236			0.06		
34	61		-0.127			0.33		
13	13		-0.353			0.24		
14	45		-0.155			0.31		
24	15		0.083			0.77		
123	72	123	-0.070	-0.012		0.56	0.92	
234	59	234	-0.175	0.069		0.18	0.60	
124	45	124	0.189	-0.028		0.21	0.85	
134	57	341	-0.166	0.320		0.22	0.02	
1234	174	1234	0.052	0.080	-0.026	0.49	0.30	0.74

significant correlation coefficients and there are no tri-variate distributions with three significant correlation coefficients. Thus, any stochastic dependence between fractions and the amount, which cannot be rejected, occurs sporadically.

Inasmuch as the test results may be sensitive to the criterion that selects the permutation of fractions, in addition to the sum of square roots of *p* values, two other criteria were used: a maximin criterion and a lexicographic criterion. Under the three criteria, the overall number of *p* values greater than 0.10 was, respectively, 83, 81, and 83.

Another potential effect on the test results may come from the choice of the transforms (17)–(19). To examine this possibility, the entire test was repeated using ratio definitions of Krzysztofowicz and Reese (1993). Again three criteria were used to select the permutation of fractions. The overall number of *p* values greater than 0.10 was, respectively, 81, 82, and 84. Thus, the results are robust with respect to the choice of the transform and the criterion for selecting the permutation.

The entire testing procedure was next applied to point precipitation data. The test results for a rain gauge in Confluence, Pennsylvania, which are summarized in appendix B, make even a stronger case for conditional stochastic independence between the fractions and the amount.

TABLE 4. Frequency of *p* values in the two-sided test of the null hypothesis $\gamma_s = 0$, $s \in \{1, 2, 3\}$; Monongahela Basin, from 1200 UTC.

<i>p</i> value of two-sided test	Number of cases			Total*
	Univariate	Bivariate	Trivariate	
$p_s \leq 0.5$	6	7	1	14
$p_s \leq 0.10$	8	9	2	19
$p_s > 0.10$	28	39	16	83
Total	36	48	18	102

*6 seasons \times 17 correlations = 102.

6. Conclusions

The process of temporal disaggregation of the total precipitation amount accumulated over a period is of considerable import to modeling and forecasting activities. The process is statistically characterized by a family of distributions of the vector of fractions $(\Theta_1, \dots, \Theta_n)$, conditional on the total amount *W*. The absence of statistically significant conditioning is termed the *disaggregative invariance*. This property has been investigated herein for areal average and point precipitation amounts accumulated over a 24-h period and disaggregated into four 6-h subperiods.

The scatterplots of observations, statistical tests, and sensitivity analyses together offer convincing empirical evidence in support of two hypotheses. First, in concordance with some earlier related studies (United States Weather Bureau 1947; Huff 1967), the vector of fractions is stochastically dependent on the amount; that is, the disaggregative invariance is not a property of precipitation. Second, when inference is conditioned on the timing of precipitation within the diurnal cycle, two properties manifest themselves: (i) the amount *W* is stochastically dependent on timing *T*—a property that is a refinement of the known dependence between the amount and duration, and (ii) the vector of fractions $(\Theta_1, \dots, \Theta_n)$, conditional on timing $T = t$, is stochastically independent of the amount *W*—a property called the *conditional disaggregative invariance*.

The practical implication of the conditional disaggregative invariance is formidable. It allows the modeler or the forecaster to decompose the problem into three tasks: (i) forecasting the precipitation timing *T*, (ii) forecasting the total amount *W*, conditional on timing $T = t$, and (iii) forecasting the temporal disaggregation $(\Theta_1, \dots, \Theta_n)$, conditional on timing $T = t$. Tasks (ii) and (iii) can be performed independently of one another, and this reduces the complexity of models or judgments.

Future research should aim at testing the conditional disaggregative invariance property for precipitation on different timescales, and possibly different spatial averaging scales, and exploiting practical advantages of

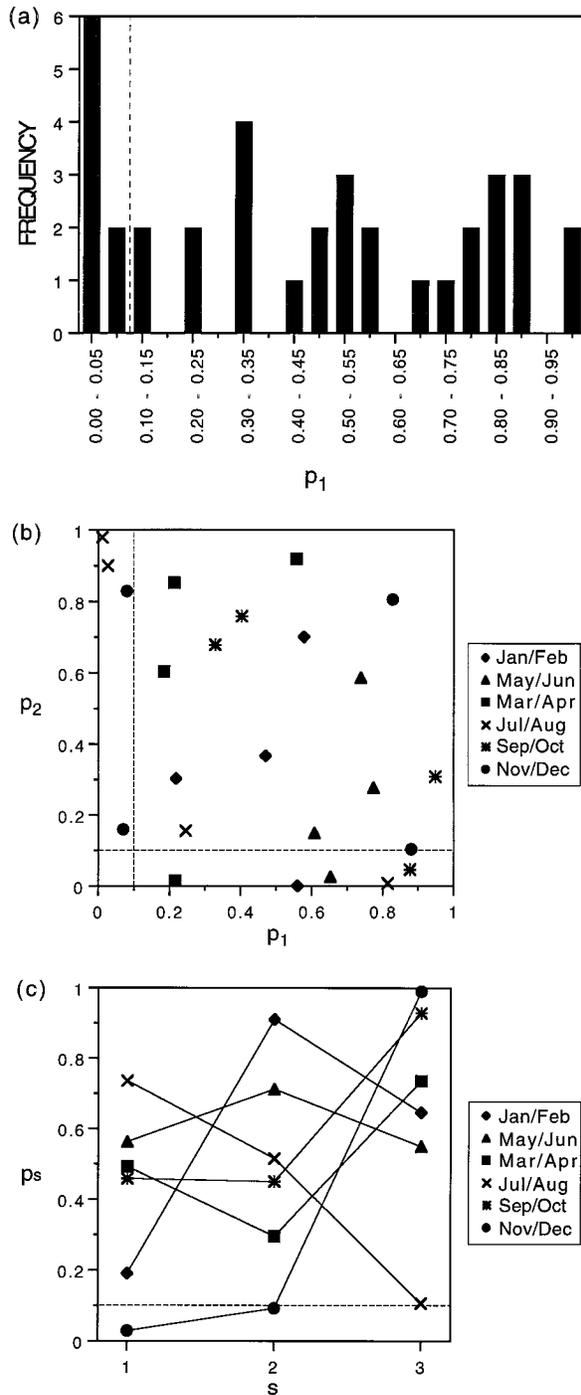


FIG. 5. Plots of p values from the two-sided test of the null hypothesis $\gamma_i = 0$, $s \in \{1, 2, 3\}$, for (a) univariate, (b) bivariate, and (c) trivariate distributions of fractions conditional on $T = t$ and $W = \omega$; Monongahela Basin, from 1200 UTC.

the property in the realms of modeling and forecasting. While the immediate need of operational forecasters is for models on the space and time scales at which precipitation amounts are predictable, scientific insight may

be gained by developing models that hold across a hierarchy of scales (Rodriguez-Iturbe et al. 1984; Rodriguez-Iturbe et al. 1987). To arrive at a general model for temporal disaggregation of a total precipitation amount, one possible approach is to proceed along the theory of stochastic bifurcation processes (Krzysztofowicz and Reese 1993). While its geometric interpretation encompasses a random cascade, which has been explored for modeling spatial rainfall (Gupta and Waymire 1993), the stochastic bifurcation process allows for different topological structures and nonidentical bifurcations that result in a nonstationary sequence of fractions—a feature indispensable for capturing the diurnal cycle of precipitation.

Acknowledgments. The hourly precipitation data were obtained through kind cooperation of John Vogel and Michael Yekta, Hydrometeorological Branch, Office of Hydrology, National Weather Service. This article was completed while Roman Krzysztofowicz was on assignment with the National Weather Service, Eastern Region, under an Intergovernmental Personnel Act agreement. Research leading to this article was supported by the National Weather Service, under the project “Development of a Prototype Probabilistic Forecasting System.” Leadership of Gary Carter in promoting this project and fostering a collaborative research environment within the Eastern Region is gratefully acknowledged.

APPENDIX A

Spurious Effects of Rainfall Measurements

a. Discrete nature of measurements

For every subperiod $i = 1, 2, 3, 4$, the scatterplot of θ_i versus ω in Fig. 1 shows curvy layers of points in the left corners and horizontal layers at ordinates $\theta_i = 0.33, 0.50, 0.67$ for ω near zero. The cause of this structure is the measurements ω and ω_i of precipitation amounts, which define the value $\theta_i = \omega_i/\omega$ of the fraction.

Theoretically, amounts W and W_i are continuous variates with the domain $[0, \infty)$ and fraction Θ_i is a continuous variate with the domain $[0, 1]$. Practically, the domain of measurements ω and ω_i is discretized by the precision of the rain gauge. Consequently, the domain of θ_i is also discretized. For example, when the rain gauge precision is $0.01''$ and a measurement of the amount for the period is $\omega = 0.02$, there are only three possible ω_i values, $\{0, 0.01, 0.02\}$, and three possible θ_i values, $\{0, 0.5, 1\}$. Table A1 illustrates the relationship: as ω increases, the number of possible θ_i values increases correspondingly. When projected into a scatterplot, this relationship implies that for a low ω value, there are relatively few possible θ_i values, and thus distinct layers of points can be discerned. As ω increases, the outer possible θ_i values tend toward either zero or one, and this gives the layers their curvy shapes.

TABLE A1. Discrete measurements of amounts and fraction.

Measured total amount ω (0.01")	Possible values of subamount ω_i (0.01")	Possible values of fraction θ_i
1	0, 1	0, 1
2	0, 1, 2	0, 0.5, 1
3	0, 1, 2, 3	0, 0.33, 0.67, 1
4	0, 1, 2, 3, 4	0, 0.25, 0.5, 0.75, 1

b. Implication on normalized scatterplots

The ramifications of the discretization are even more pronounced in the scatterplots of transformed observations, z_i versus v , shown in Fig. 3. An annotated enlargement of the left side of the first scatterplot (Fig. A1) highlights the structure of points. The first cluster of points on the left results from observations ($\omega = 0.02, \theta_1 = 0.5$). Because they have identical coordinates, the original observations plot as one point (Fig. 1), but after NQT each observation plots as a unique point (Fig. 3). There are no clusters resulting from observations ($\omega = 0.02, \theta_1 = 0$) and ($\omega = 0.02, \theta_1 = 1$) because the plot is restricted to values $0 < \theta_1 < 1$. Further to the right there are clusters resulting from observation $\omega = 0.03$ and two possible fraction values, $\theta_1 = 0.33, 0.67$. The next clusters to the right result from observation $\omega = 0.04$ and three possible fraction values, $\theta_1 = 0.25, 0.5, 0.75$. As ω increases, the clusters fan out, creating layers of points and demarcating the lower and upper bounds of the scatterplot.

In summary, the coarse precision of rainfall measurements affects the image of stochastic dependence between fraction Θ_i and amount W_i , as portrayed by a scatterplot of observations. In the space of transformed variates, the scatterplot takes the shape of a diamond. The left triangle of this diamond is an artifact of the coarse measurements. This artifact should be accounted for in inference.

c. Mixed rain gauge precision

Prior to 1971, the universal rain gauge was the standard device for measuring rainfall. The precision of this gauge was 0.01" and observations were recorded on strip charts. Sometime between 1971 and 1972, the Fischer-Porter rain gauge began to be used. The paper tape record produced by this gauge is easier to process, but the precision is coarser: 0.1". The exact timing of the change from the universal gauge to the Fischer-Porter gauge varies from station to station, and, in some cases, the switch was never made.

The change in rain gauge precision makes the domain of measurements ω and ω_i coarser and this, in turn, may impact the scatterplot of θ_i versus ω , especially for small amounts ω . Moreover, a long climatic sample (such as the one used here from 1948 to 1993) will contain measurements whose precision varies by the order of magnitude. The effect of this nonhomogeneity depends upon

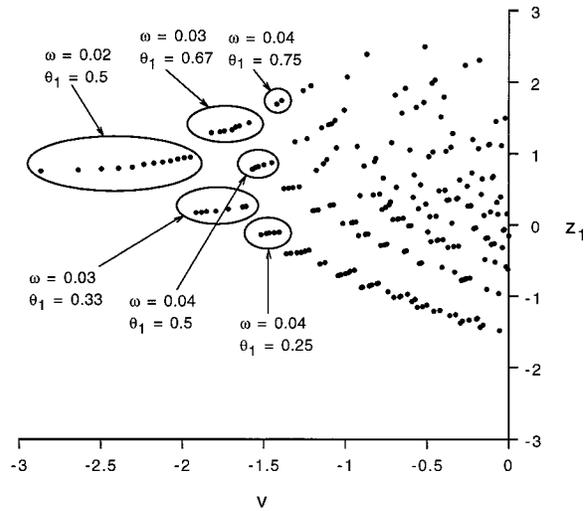


FIG. A1. Enlarged left side of the scatterplot of transformed observations, z_1 versus v , from Fig. 3.

the sizes of the two subsamples and the purpose of inference. We examined scatterplots based on two subsamples and judged that the nonhomogeneity of the combined sample should not affect conclusions from our analyses.

APPENDIX B

Test Results for a Station

The two-sided test of the null hypothesis $\gamma_s = 0, s \in \{1, 2, 3\}$, detailed in section 5e, was applied to precipitation data from a rain gauge in Confluence, Pennsylvania. Table B1 reports the overall frequency of p values in the same format as Table 4. In all 102 tests, the null hypothesis cannot be rejected in 97 cases (95%) at the significance level $\alpha = 0.10$. Figure B1 displays the p values in the same format as Fig. 5.

Whereas the case for conditional stochastic independence between the vector of fractions and the amount is strong for the Monongahela Basin (Table 4, Fig. 5), it is uniformly stronger for the Confluence station (Table B1, Fig. B1). An explanation can be deduced from the nature of the spatial averaging of precipitation. Relative to the point precipitation process, the areal average precipitation process has (i) a higher frequency of longer

TABLE B1. Frequency of p values in the two-sided test of the null hypothesis $\gamma_s = 0, s \in \{1, 2, 3\}$; Confluence station, from 1200 UTC.

p value of two-sided test	Number of cases			Total*
	Univariate	Bivariate	Trivariate	
$p_s \leq 0.5$	2	1	0	3
$p_s \leq 0.10$	2	2	1	5
$p_s > 0.10$	34	46	17	97
Total	36	48	18	102

*6 seasons \times 17 correlations = 102.

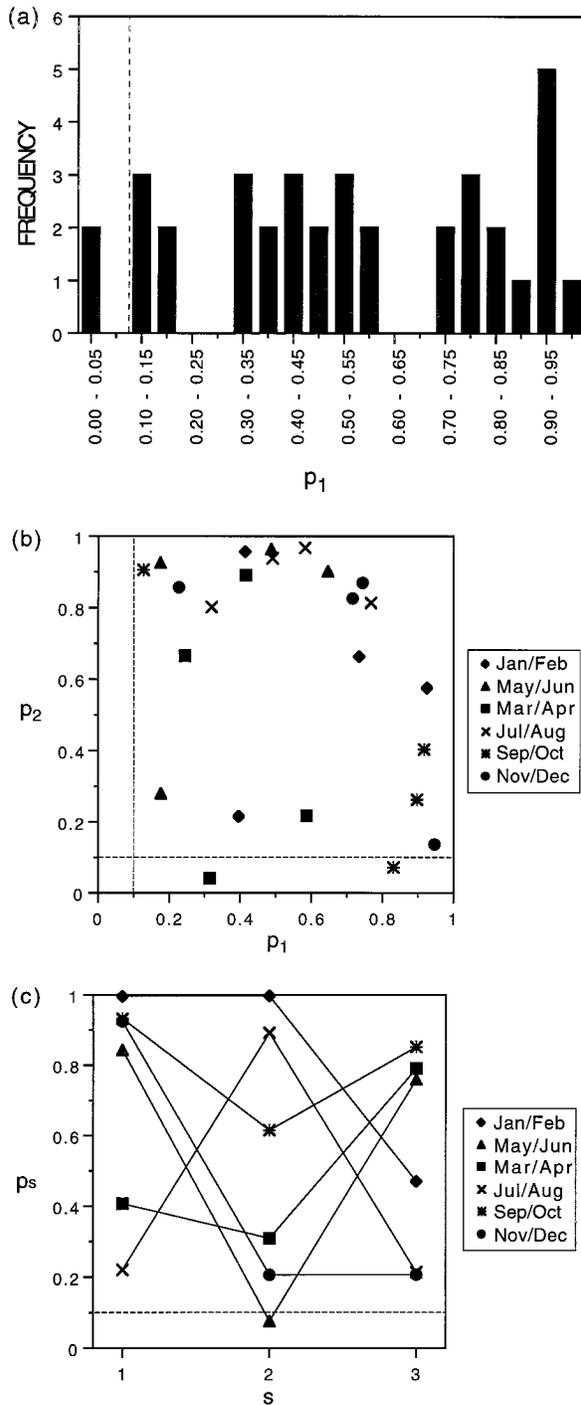


FIG. B1. Plots of p values from the two-sided test of the null hypothesis $\gamma_s = 0$, $s \in \{1, 2, 3\}$, for (a) univariate, (b) bivariate, and (c) trivariate distributions of fractions conditional on $T = t$ and $W = \omega$; Confluence station, from 1200 UTC.

durations and (ii) a more uniform temporal disaggregation of the amount. The combined effect is the *temporal smoothing of high precipitation*— the phenomenon, which is discussed in section 4c and which is responsible for a negative correlation between fraction Θ_i and amount W . The effect is subtle though quantifiable: the proportion of significant correlation coefficients (at $\alpha \leq 0.10$) is 5/102 for the Confluence station and increases to 19/102 for the Monongahela Basin.

REFERENCES

Aitchison, J., 1986: *The Statistical Analysis of Compositional Data*. Chapman and Hall.

Farlie, D. J. G., 1960: The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**(3), 307–323.

Gupta, V. K., and E. C. Waymire, 1993: A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteor.*, **32**, 251–267.

Hershendorfer, J., and D. A. Woolhiser, 1987: Disaggregation of daily rainfall. *J. Hydrol.*, **95**, 299–322.

Hjelmfelt, A. T., Jr., 1981: Overland flow from time-distributed rainfall. *J. Hydraul. Div.*, **107**(HY2), 227–238.

Huff, F. A., 1967: Time distribution of rainfall in heavy storms. *Water Resour. Res.*, **3**(4), 1007–1019.

Koutsoyiannis, D., and E. Foufoula-Georgiou, 1993: A scaling model of a storm hyetograph. *Water Resour. Res.*, **29**(7), 2345–2361.

Kruskal, W. H., 1958: Ordinal measures of association. *J. Amer. Stat. Assoc.*, **53**, 814–861.

Krzysztofowicz, R., 1993: Probabilistic hydrometeorological forecasting system: A conceptual design. Post-Print Volume: Third National Heavy Precipitation Workshop, NOAA Tech. Memo. NWS ER-87, 29–42. [Available from the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161.]

—, and S. Reese, 1993: Stochastic bifurcation processes and distributions of fractions. *J. Amer. Stat. Assoc.*, **88**(421), 345–354.

—, and A. A. Sigrest, 1997: Local climatic guidance for probabilistic quantitative precipitation forecasting. *Mon. Wea. Rev.*, **125**, 305–316.

—, W. J. Drzal, T. R. Drake, J. C. Weyman, and L. A. Giordano, 1993: Probabilistic quantitative precipitation forecasts for river basins. *Wea. Forecasting*, **8**, 424–439.

Lindgren, B. W., 1976: *Statistical Theory*. Macmillan.

Linsley, R. K., Jr., M. A. Kohler, and J. L. H. Paulhus, 1975: *Hydrology for Engineers*. McGraw-Hill, 82–84.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire, 1984: Scale considerations in the modeling of temporal rainfall. *Water Resour. Res.*, **20**(11), 1611–1619.

—, B. Febres de Power, and J. B. Valdés, 1987: Rectangular pulses point process models for rainfall: Analysis of empirical data. *J. Geophys. Res.*, **92**(D8), 9645–9656.

United States Weather Bureau, 1947: Thunderstorm rainfall. Hydro-meteorological Rep. 5.

Wallace, J. M., 1975: Diurnal variations in precipitation and thunderstorm frequency over the conterminous United States. *Mon. Wea. Rev.*, **103**, 406–419.