519.246.87:551.578.1(492)

MEDEDELINGEN LANDBOUWHOGESCHOOL
WAGENINGEN • NEDERLAND • 78-8 (1978)

THE SPECTRAL ANALYSIS OF DAILY RAINFALL SEQUENCES

T. A. BUISHAND
Department of Mathematics, Agricultural University, Wageningen, The Netherlands

(Received 24–X–1977)

H. VEENMAN & ZONEN B.V. – WAGENINGEN – 1978

CONTENTS

1. INTRODUCTION

Spectral analysis is known as an important tool in studying time series. This paper discusses the application of this technique to daily rainfall sequences in the Netherlands. Univariate and bivariate spectral analysis are studied successively. The former deals with a rainfall series at one point: the latter with rainfall sequences at two points.

2. UNIVARIATE SPECTRAL ANALYSIS

2.1. THE SPECTRUM

2.1.1. *Definition of the spectrum of a stochastic process*

Since the rainfall process is observed at discrete time intervals the definition of the spectrum will be based on stochastic processes in discrete time. Further for the definition given here it will be required that the process is stationary or at least second-order stationary, that is the first and second-order moments (mean, variance, autocorrelation coefficients) may not change in time.

Let $\{X_t\}$ be a stationary process with mean μ_x and variance σ_x^2, then its autocovariance function is defined by

$$\gamma_{xx}(k) = E(X_t - \mu_x)(X_{t+k} - \mu_x). \tag{1}$$

For $k = 0$ one gets the variance. The function $\gamma_{xx}(k)$ is an even function in k (and for rainfall sequences $\gamma_{xx}(k)$ is positive and monotonically decreasing in $|k|$). Division of $\gamma_{xx}(k)$ by σ_x^2 gives the autocorrelation function $\rho_{xx}(k)$.

The Fourier transform of the autocovariance function is called the power spectrum:

$$\Gamma_{xx}(f) = \sum_{k=-\infty}^{\infty} \gamma_{xx}(k) e^{-2\pi i f k}\ , -{}^1/_2 < f \leqslant {}^1/_2. \tag{2}$$

But since $\gamma_{xx}(k) = \gamma_{xx}(-k)$ Eq. (2) reduces to

$$\Gamma_{xx}(f) = \gamma_{xx}(0) + 2\sum_{k=1}^{\infty} \gamma_{xx}(k) \cos 2\pi f k\ , -{}^1/_2 < f \leqslant {}^1/_2. \tag{3}$$

Notice further that

$$\Gamma_{xx}(f) = \Gamma_{xx}(-f) \tag{4}$$

$$\int_{-1/2}^{1/2} \Gamma_{xx}(f)\, df = \gamma_{xx}(0) = \sigma_x^2. \tag{5}$$

Thus the power spectrum shows how the variance is decomposed over the various frequencies. Since $\Gamma_{xx}(f)$ is an even function in f it is usual to define a power spectrum for non-negative f

$$\Gamma_{xx}^{+}(f) = 2\Gamma_{xx}(f) = 2\left\{\gamma_{xx}(0) + 2\sum_{k=1}^{\infty} \gamma_{xx}(k)\cos 2\pi fk\right\}, 0 \leqslant f \leqslant {}^{1}/_{2}. \quad (6)$$

When the power spectrum is divided by the variance one obtains the spectral density (the Fourier transform of the autocorrelation function). This function has the same properties as a probability density.

2.1.2. *Examples*

The first example given here is a stationary first-order autoregressive process $\{X_t\}$ with parameter α_1. This process is defined by

$$X_t = \alpha_1 X_{t-1} + \sqrt{1-\alpha_1^2}\, Z_t \quad (7)$$

where $\{Z_t\}$ is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance. The process $\{X_t\}$ has zero mean and unit variance: its autocorrelation function is given by

$$\rho_{xx}(k) = \alpha_1^{|k|} \quad (8)$$

and its spectral density is

$$\Gamma_{xx}^{+}(f) = \frac{2(1-\alpha_1^2)}{1+\alpha_1^2 - 2\alpha_1 \cos 2\pi f}, 0 \leqslant f \leqslant {}^{1}/_{2}. \quad (9)$$

Fig. 1. shows the spectral density for some values of α_1. For $\alpha_1 = 0$ one has an uncorrelated process with a constant spectral density at all frequencies (so-called white noise process). For positive values of α_1 the spectral density is large for the low frequencies and small for the high frequencies. Stochastic processes with this property are called red noise processes. The slope of the curve in Fig. 1 becomes steeper when α_1 increases.

In the following example the influence of the tail of the autocorrelation function is studied. The stochastic process being considered is defined as follows

$$X_t = a_1 Y_{1t} + a_2 Y_{2t} \quad (10)$$

where $\{Y_{1t}\}$ and $\{Y_{2t}\}$ are two independent first-order autoregressive processes with parameter α_1 and α_2, respectively. For simplicity it will be assumed that $\{Y_{1t}\}$ and $\{Y_{2t}\}$ have zero mean and unit variance. When the parameter α_2 of the process $\{Y_{2t}\}$ is chosen close to 1 and when a_2 is taken small with respect to a_1, one gets a stochastic process whose autocorrelation function at small lags hardly differs from that of a first-order autoregressive process (with parameter α_1), but the tail of its correlogram is much longer. The process $\{X_t\}$ has zero mean, its variance is $a_1^2 + a_2^2$, and its autocorrelation function is given by

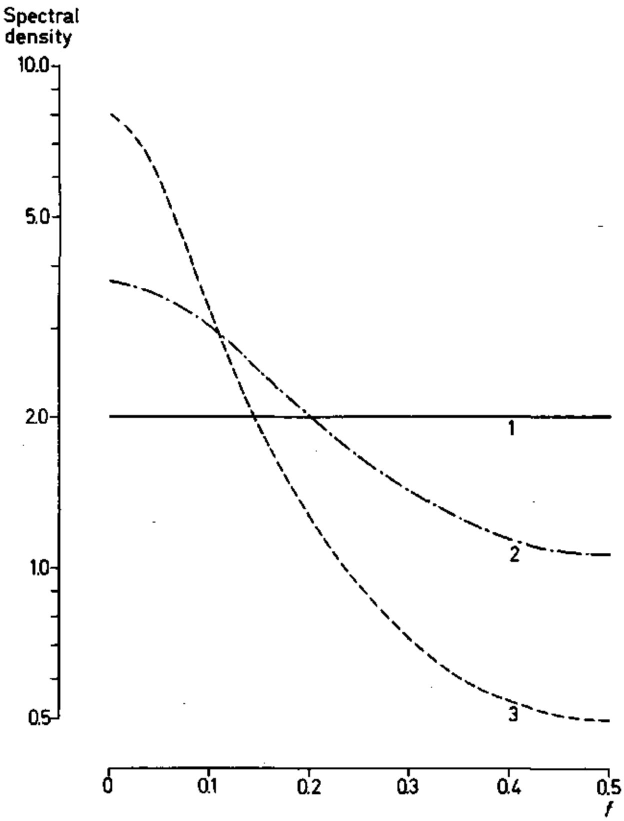

FIG. 1. The spectral density of an autoregressive process. The parameter α_1 is 0 for Process 1 (white noise), 0.3 for Process 2 and 0.6 for Process 3.

$$\rho_{xx}(k) = p\alpha_1{}^{|k|} + (1-p)\,\alpha_2{}^{|k|} \tag{11}$$

with $p = a_1^2/(a_1^2 + a_2^2)$. So the shape of the autocorrelation function depends on 3 parameters, namely α_1, α_2 and the ratio b of a_1 to a_2. The spectral density of $\{X_t\}$ is

$$\Gamma^+_{xx}(f)/\sigma_x^2 = \frac{2p(1-\alpha_1^2)}{1+\alpha_1^2 - 2\alpha_1\cos 2\pi f} + \frac{2(1-p)\,(1-\alpha_2^2)}{1+\alpha_2^2 - 2\alpha_2\cos 2\pi f}\,,$$
$$0 \leqslant f \leqslant {}^1/_2. \tag{12}$$

Fig. 2 gives the correlogram and the spectral density of some processes with approximately the same first-order autocorrelation coefficient (0.3), but with different tails of the autocorrelation function. From this figure it is seen that the spectra stongly differ at the low frequencies: the spectra of Process 1 and Process 3 differ only visibly near the initial point. The peak of the spectral

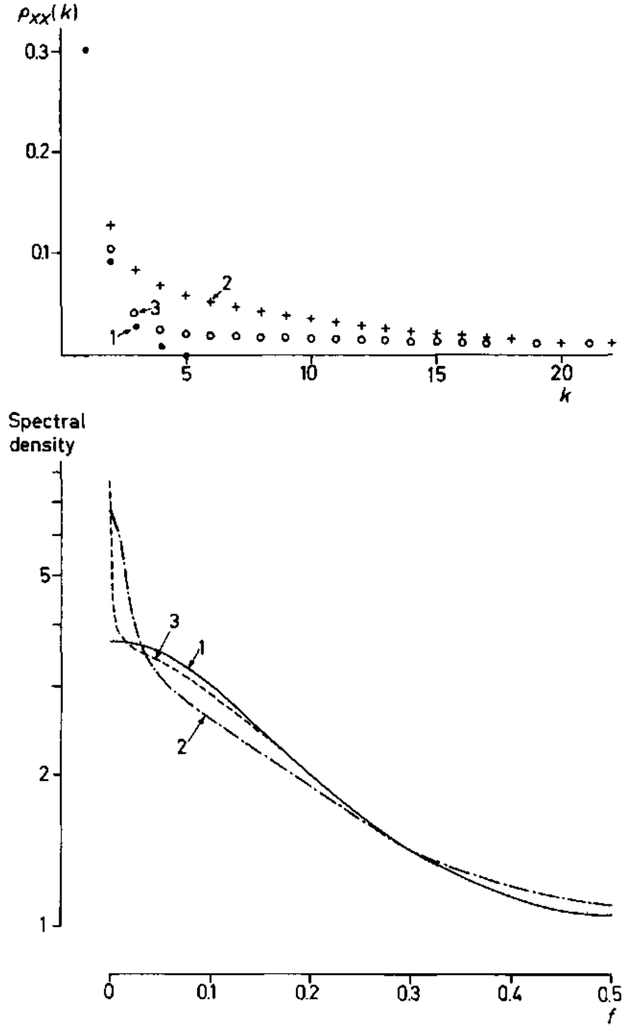

FIG. 2. The autocorrelation function and the spectral density for an autoregressive process (Process 1) and for mixtures of autoregressive processes (Processes 2 and 3). The parameter α_1 of Process 1 is 0.3; the parameters of Process 2 are $\alpha_1 = 0.23$, $\alpha_2 = 0.90$ and $b = 3$, and the parameters of Process 3 are $\alpha_1 = 0.29$, $\alpha_2 = 0.98$ and $b = 7$.

density at the zero frequency becomes more pronounced when the correlogram falls off more slowly.

From the examples given here it can be concluded:

a. Differences in the first-order autocorrelation coefficients give rise to differences in the slope of the spectral density.
b. Differences in the tail of the correlogram give rise to differences at the low frequencies.

2.1.3. *Relation to the variance-time diagram*

Since the power spectrum is the Fourier transform of the autocovariance function it gives the same information about the stochastic process. Another measure equivalent to the autocovariance function is the variance-time diagram. Let $V_x(n)$ be the variance of $X_1 + X_2 + \ldots + X_n$, then the variance-time diagram gives $V_x(n)$ as a function of n. Clearly

$$V_x(n) = \sum_{k=-(n-1)}^{n-1} (n - |k|)\, \gamma_{xx}(k). \tag{13}$$

From this relation it follows that

$$S_x(n) = V_x(n+1) - V_x(n) = \sum_{k=-n}^{n} \gamma_{xx}(k). \tag{14}$$

When n tends to infinity one gets

$$\lim_{n\to\infty} S_x(n) = \sum_{k=-\infty}^{\infty} \gamma_{xx}(k) = \Gamma_{xx}(0) \tag{15}$$

using Eq. (2). So the asymptotic slope of the variance-time diagram is equal to the power spectrum at $f = 0$.

2.2. Estimation of spectra

2.2.1. *The sample spectrum*

Let $x_1, \ldots, x_N$ be a realization of a stationary stochastic process. Then the sample autocovariance function is defined by

$$c_{xx}(k) = \frac{1}{N} \sum_{j=1}^{N-k} (x_j - \bar{x})(x_{j+k} - \bar{x}),\ k = 0, \ldots, N\text{–}1 \tag{16}$$

where $\bar{x}$ is the mean of the N x_js. Dividing $c_{xx}(k)$ by the sample variance s_x^2 gives the lag k autocorrelation coefficient $r_{xx}(k)$, shortly denoted by r_k. The sample (power) spectrum is obtained by taking the Fourier transform of (16)

$$C_{xx}(f) = c_{xx}(0) + 2 \sum_{k=1}^{N-1} c_{xx}(k) \cos 2\pi fk\ ,\ -{}^1/_2 < f \leqslant {}^1/_2 \tag{17}$$

or, when only non-negative frequencies are considered

$$C_{xx}^{+}(f) = 2 \left\{ c_{xx}(0) + 2 \sum_{k=1}^{N-1} c_{xx}(k) \cos 2\pi fk \right\},\ 0 \leqslant f \leqslant {}^1/_2. \tag{18}$$

The sample spectrum has the following properties

$$C_{xx}(0) = 0, \text{ see Appendix} \tag{19}$$

$$\int_{-1/2}^{1/2} C_{xx}(f)\, df = c_{xx}(0) = s_x^2 \tag{20}$$

that is the sample spectrum shows how the sample variance is distributed over the various frequencies.

For $f \neq 0$ the sample spectrum $C_{xx}(f)$ is an asymptotically unbiased estimator of $\Gamma_{xx}(f)$: its asymptotic variance is $\Gamma^2_{xx}(f)$ when $f \neq {}^1/_2$ and $2\Gamma^2_{xx}(f)$ when $f = {}^1/_2$. So $C_{xx}(f)$ is an inconsistent estimator since its variance does not tend to zero when N tends to infinity.

2.2.2. *Smoothing of spectral estimates*

The variance of spectral estimates can be reduced by smoothing. Smoothed estimates can be obtained by giving weights to the estimated autocovariances at different lags

$$\tilde{C}_{xx}(f) = w(0)\, c_{xx}(0) + 2 \sum_{k=1}^{N-1} w(k)\, c_{xx}(k) \cos 2\pi f k, \; -{}^1/_2 < f \leqslant {}^1/_2. \tag{21}$$

The weight function $w(k)$ is called the lag window. To preserve the relation (20) $w(0)$ is taken to be 1.

There are many different types of lag windows known in the literature (JENKINS and WATTS, 1969: NEAVE, 1972). In this study use was made of the Tukey window

$$w_T(k) = \begin{cases} {}^1/_2 \left(1 + \cos \dfrac{\pi k}{M}\right) &, k \leqslant M \\ 0 &, k > M. \end{cases} \tag{22}$$

When this window is used Eq. (21) becomes

$$\tilde{C}_{xx}(f) = c_{xx}(0) + \sum_{k=1}^{M-1} \left(1 + \cos \frac{\pi k}{M}\right) c_{xx}(k) \cos 2\pi f k, \quad -{}^1/_2 < f \leqslant {}^1/_2 \tag{23}$$

or, when only non-negative frequencies are considered

$$\tilde{C}^+_{xx}(f) = 2\left\{c_{xx}(0) + \sum_{k=1}^{M-1} \left(1 + \cos \frac{\pi k}{M}\right) c_{xx}(k) \cos 2\pi f k\right\}, \quad 0 \leqslant f \leqslant {}^1/_2. \tag{24}$$

A crude approximation of var $\{\tilde{C}_{xx}(f)\}$ is

$$\text{var}\,\{\tilde{C}_{xx}(f)\} \approx 0.75 \frac{M}{N} \Gamma^2_{xx}(f). \tag{25}$$

It is not permitted to apply this approximation in the neighbourhood of $f = 0$ and $f = {}^1/_2$. Caution is also needed in the neighbourhood of peaks and troughs.

From (25) it is seen that the variance is considerably reduced for small values of M (that is the number of lags for which the autocovariances have to be estimated is small). On the other hand a small value of M may give rise to considerable bias. In practice different values of M are tried in estimating the spectrum.

Notice that (23) may give a reasonable estimate of $\Gamma_{xx}(f)$ for $f = 0$. Therefore the parameter M should be small with respect to N, but large enough that $\gamma_{xx}(k)$ is negligible for $k \geqslant M$.

A smoothed estimate of the spectral density is obtained by dividing $\tilde{C}_{xx}(f)$ by the sample variance.

Usually the logarithms of spectral estimates are plotted, since their variances are approximately independent of the theoretical value $\Gamma_{xx}(f)$. From (25) it follows that

$$\text{var}\,\{\ln \tilde{C}_{xx}(f)\} \approx 0.75\,\frac{M}{N} \tag{26}$$

and also when only non-negative frequencies are considered, Eq. (24), one gets

$$\text{var}\,\{\ln \tilde{C}^{+}_{xx}(f)\} \approx 0.75\,\frac{M}{N}. \tag{27}$$

2.3. Spectral analysis of Winterswijk data

2.3.1. *Description of the data*

The meteorological station of Winterswijk is situated in the east of the Netherlands (51°58′N, 6°49′E). The data used were daily values of the period Dec. 1907 – Nov. 1973.

The rainfall data show seasonal variation in the mean, the standard deviation and the autocorrelation coefficients. The largest values for the mean and standard deviation are found in summer: the largest autocorrelation coefficients are found in winter (see Fig. 3). To reduce this seasonal variation the rainfall sequence was split into 4 subseries: each subseries contained the data for a particular season. The construction of the subseries was done in such a way that wet or dry spells having days in two seasons were assigned to only one of these seasons (Buishand, 1977a, III. 2 and III. 6). Since features of the rainfall process show little variation within a season the subseries are approximately stationary. In this paper special attention is given to the winter (December-February) and summer season (June-August).

Spectra were also estimated for the sequences of wet and dry days (wet-dry sequence). A day was taken to be wet when its rainfall amount was at least 0.3 mm. The value 1 was assigned to a wet day and the value 0 to a dry day.

2.3.2. *The spectral density of the wet-dry sequence*

In this subsection smoothed estimates are compared with theoretical spectra of fitted processes. The processes considered are:

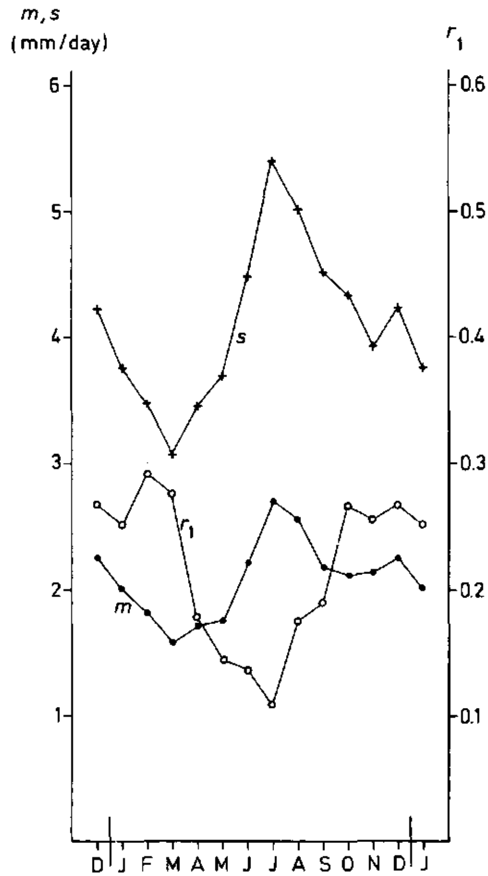

FIG. 3. Mean (m), standard deviation (s) and first-order serial correlation coefficient (r_1) of daily rainfall at Winterswijk.

a. A first-order Markov chain. For this process the probability of a wet day depends on the state (wet or dry) of the previous day. Let $P(W \mid D)$ denote the probability of a wet day after a dry day and $P(W \mid W)$ the probability of a wet day after a wet day. Most rainfall sequences have the property that $P(W \mid W) > P(W \mid D)$. For example. for Winterswijk $P(W \mid W)$ ranges from 0.59 to 0.70 and $P(W \mid D)$ ranges from 0.27 to 0.38. For Markov chains the probabilities $P(W \mid D)$ and $P(W \mid W)$ are called transition probabilities.

b. An alternating renewal process. For this process the lengths of successive wet and dry spells are independent. In this study truncated negative binomial distributions were fitted to lengths of wet and dry spells. Therefore. the process is abbreviated as TNBD-TNBD process.

Fig. 4 compares smoothed estimates of the spectral density with theoretical values for the winter and summer season. From the behaviour at the low frequencies it is seen that the TNBD-TNBD process has a longer memory than the first-order Markov-chain. Further for small values of $M (M = 4)$ the

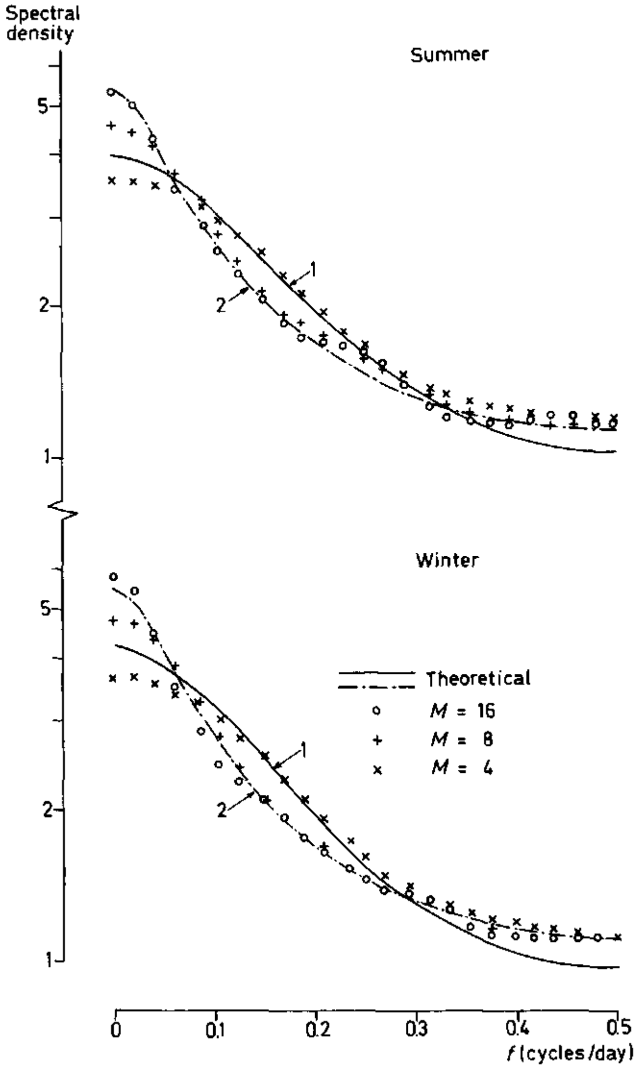

FIG. 4. Smoothed spectral density estimates and spectral densities of fitted processes for the wet-dry sequence of Winterswijk. Process 1 is a first-order Markov chain and Process 2 is an alternating renewal process with truncated negative binomial distributions fitted to lengths of wet and dry spells.

estimates are heavily biased. When only a small value of M had been taken into account, one should have come to the wrong conclusion that the first-order Markov chain is a good model! For $M = 16$ there is hardly any bias and the spectrum of the TNBD-TBBD process fits the estimated values well. There are hardly any differences between the spectra of the winter and summer season.

indicating that there is little seasonal variation in the dependence of successive wet and dry days.

The following remarks can be made on the calculation of the theoretical spectra. For the first-order Markov chain the autocorrelation function and the spectral density have the same form as those of a first-order autoregressive process, see Eqs. (8) and (9). The parameter α_1 can be expressed in the transition probabilities. An expression for the autocovariances of an alternating renewal process was given by BUISHAND (1977a, IV.4.1). The power spectrum can be obtained by applying Eq. (6) to the calculated autocovariances. It is, however, also possible to express the power spectrum or the spectral density in the probability generating functions of the lengths of wet and dry spells (see Appendix).

2.3.3. *The spectral density of the rainfall sequence*

In this subsection the fit of a rainfall model is tested by comparing its spectral density with estimated values. The most general form of the rainfal model is:

a. The wet-dry process is a TNBD-TNBD process.
b. The distribution of the amount of rainfall on a particular wet day depends on the number of adjacent wet days: the mean and variance increase with this number.
c. Rainfall amounts within a wet spell are correlated according to a first-order moving average process. That is there is only correlation between adjacent wet days.

Attention is also paid to the first-order autoregressive process, defined by Eq. (7). This process has widely been used in the past as a model for persistence in daily rainfall sequences (LEVERT, 1960).

The spectral density of the first-order autoregressive process can be obtained from (9). For the general rainfall model Eq. (6) was applied to the autocorrelation coefficients. The method of calculation of the autocorrelation coefficients was given by BUISHAND (1977a, IV.4.2 and IV.4.3).

Fig. 5 compares smoothed estimates of the spectral density with the theoretical values of the general rainfall model for the winter and summer season. There are some differences in the spectra of these seasons, since in summer there is less autocorrelation. In summer the spectrum of the rainfall model fits well: for the winter season there is some indication for lack of long-term persistence. The estimates at the low frequencies depend strongly on the choice of M. For small M there is serious bias.

Fig. 6 compares some theoretical spectra of fitted processes for the winter season. The processes considered are: a first-order autoregressive process, the general rainfall process mentioned previously, and a simplified version of this process, namely with iid rainfall amounts within a wet spell. The spectra of the first two processes only differ at the low frequencies: the first-order autoregressive process shows less long-term persistence. When rainfall amounts are assumed to be iid the spectrum becomes more flat, indicating that the lower order autocorrelation coefficients are too small.

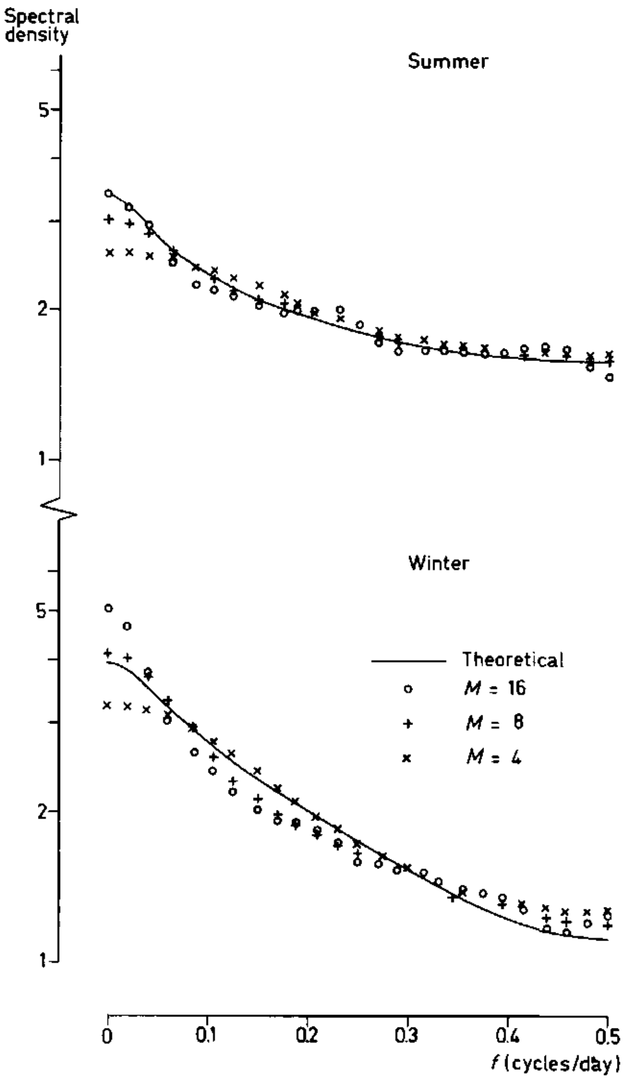

FIG. 5. Smoothed spectral density estimates and the spectral density of a general rainfall model for the rainfall sequence of Winterswijk.

2.3.4. *Comparison with variance-time analysis*

Since the initial point of the spectrum equals the asymptotic slope of the variance-time diagram, a rough estimate of $V_x(n)$ for large n can be based on $0.5\, n\, \tilde{C}^{+}_{xx}(0)$. Because the variance-time diagram has a small negative intercept (BUISHAND, 1977a) $V_x(n)$ is somewhat overestimated by this method.

Table 1 gives estimates of $V_x(30)$ based on $\tilde{C}^{+}_{xx}(0)$ and the estimate directly obtained from the historic series (BUISHAND, 1977a, Table III.6.3). The estimate

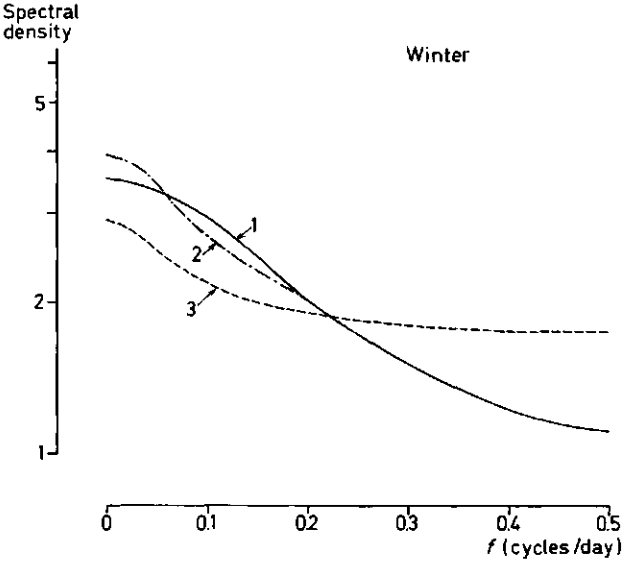

FIG. 6. Spectral densities of some processes fitted to the rainfall sequence of Winterswijk. Process 1 is a first-order autoregressive process; Process 2 is a general rainfall model, and Process 3 is a simplification of it, namely with independent and identically distributed rainfall amounts within a wet spell.

based on the initial point of the spectrum heavily depends on the choice of M: when M is small (that is when too few autocorrelation coefficients are considered) the variance is seriously underestimated. For large M (M at least 16) there is a nice correspondence between the two estimates.

TABLE 1. Estimates of $V_x(30)$ for the wet-dry sequence and the rainfall sequence of Winterswijk.

	Wet-dry sequence (variance in days^2)		Rainfall sequence (variance in mm^2)	
	winter	summer	winter	summer
Based on $\tilde{C}^{+}_{xx}(0)$ with				
$M = 4$	13.8	13.4	751	978
$M = 8$	18.0	17.0	949	1127
$M = 16$	21.6	20.0	1162	1278
$M = 32$	23.0	20.5	1326	1294
Directly estimated from the historic sequence	22.0	18.9	1167	1286

2.4. CONCLUDING REMARKS

It was noticed earlier that the autocovariance function, the variance-time function and the power spectrum are equivalent functions of second-order moments. One may ask, what are the advantages and disadvantages of these different techniques? A drawback of autocorrelation analysis is that it hardly gives information about long-term persistence. On the contrary the variance-time diagram is very sensitive to the length of the tail of the correlogram. When use is made of both correlogram and variance-time analysis one gets a clear insight into the persistence of a rainfall sequence.

The spectrum may give both information about short-term and long-term behaviour. Moreover the sample properties of spectral estimates are simpler than those of autocorrelation coefficients. But when using spectral analysis one encounters the problem of smoothing. Besides spectra are much harder to interpret than correlograms or variance-time diagrams. These drawbacks make spectral analysis intractable for hydrologists.

3. BIVARIATE SPECTRAL ANALYSIS

3.1. The cross spectrum

3.1.1. *Definition of the cross spectrum of a bivariate stochastic process*

The definition of the cross spectrum will be based on a stationary (or at least second-order stationary) bivariate process. As an example of a bivariate stochastic process one may take the rainfall sequences at two different stations. When these rainfall sequences are considered for a particular season (e.g. the winter) they are approximately stationary and the theory given below is applicable.

Now let $\{X_t, Y_t\}$ be a stationary bivariate stochastic process with mean $\{\mu_x, \mu_y\}$ and variance $\{\sigma_x^2, \sigma_y^2\}$, then its cross covariance function is defined by

$$\gamma_{xy}(k) = E(X_t - \mu_x)(Y_{t+k} - \mu_y). \tag{28}$$

Division of $\gamma_{xy}(k)$ by $\sigma_x\sigma_y$ gives the cross correlation function $\rho_{xy}(k)$.

From the definition it follows

$$\gamma_{xy}(-k) = \gamma_{yx}(k). \tag{29}$$

So unlike the autocovariance function the cross covariance function needs not to be an even function. But one may split this function into an even and an odd part.

$$\gamma_{xy}(k) = \lambda_{xy}(k) + \psi_{xy}(k) \tag{30}$$

with

$$\lambda_{xy}(k) = {}^1/_2\{\gamma_{xy}(k) + \gamma_{xy}(-k)\} \text{ being an even function}$$
$$\psi_{xy}(k) = {}^1/_2\{\gamma_{xy}(k) - \gamma_{xy}(-k)\} \text{ being an odd function.}$$

The Fourier transform of the cross covariance function is called the cross spectrum

$$\Gamma_{xy}(f) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\, e^{-2\pi i f k} \quad , -{}^1/_2 < f \leqslant {}^1/_2. \tag{31}$$

For the even part of the cross covariance function one may define

$$\Lambda_{xy}(f) = \lambda_{xy}(0) + 2\sum_{k=1}^{\infty} \lambda_{xy}(k) \cos 2\pi f k \quad , -{}^1/_2 < f \leqslant {}^1/_2 \tag{32}$$

(the co-spectrum) and for the odd part

$$\Psi_{xy}(f) = 2\sum_{k=1}^{\infty} \psi_{xy}(k) \sin 2\pi f k \quad , -{}^1/_2 < f \leqslant {}^1/_2 \tag{33}$$

(the quadrature spectrum). Substitution of (30) in (31) results in

$$\Gamma_{xy}(f) = \Lambda_{xy}(f) - i\Psi_{xy}(f). \tag{34}$$

From the Eqs. (32), (33) and (34) one sees immediately

$$\Lambda_{xy}(f) = \Lambda_{xy}(-f) \tag{35}$$

$$\Psi_{xy}(f) = -\Psi_{xy}(-f) \tag{36}$$

$$\Gamma_{xy}(f) = \bar{\Gamma}_{xy}(-f) \tag{37}$$

where the bar denotes the complex conjugate.

Further for the co-spectrum holds

$$\int_{-1/2}^{1/2} \Lambda_{xy}(f)\, df = \lambda_{xy}(0) = \gamma_{xy}(0). \tag{38}$$

Thus the co-spectrum shows how the lag zero cross covariance is distributed over the various frequencies.

The cross spectrum is often written in complex polar notation

$$\Gamma_{xy}(f) = \alpha_{xy}(f) \exp(2\pi i \phi_{xy}(f)) \tag{39}$$

with $\alpha_{xy}(f)$: the cross amplitude spectrum
$\phi_{xy}(f)$: the phase spectrum.
From (34) and (39) it follows

$$\alpha^2_{xy}(f) = \Lambda^2_{xy}(f) + \Psi^2_{xy}(f) \tag{40}$$

and

$$\tan \phi_{xy}(f) = -\Psi_{xy}(f)/\Lambda_{xy}(f). \tag{41}$$

Instead of the cross amplitude spectrum one may use the coherency spectrum

$$\kappa_{xy}(f) = \alpha_{xy}(f)/\sqrt{\Gamma_{xx}(f)\,\Gamma_{yy}(f)}. \tag{42}$$

Just like the correlation coefficient the coherency spectrum is a normalized measure of linear dependence. It shows the correlation between the two sequences $\{X_t\}$ and $\{Y_t\}$ as a function of frequency.

3.1.2. *Relation to covariances and correlation coefficients of n-day totals*

It was seen earlier that for large n the variance of n-day totals could be derived from the initial point of the power spectrum. Analogously, the initial point of the cross spectrum (or the co-spectrum) gives information about the asymptotic behaviour of the covariance of n-day totals. Let $V_{xy}(n)$ be the covariance of n-day totals of the sequences $\{X_t\}$ and $\{Y_t\}$, then

$$V_{xy}(n) = \sum_{k=-(n-1)}^{n-1} (n - |k|)\gamma_{xy}(k). \tag{43}$$

When $V_{xy}(n)$ is plotted versus n the slope of the curve is

$$S_{xy}(n) = V_{xy}(n+1) - V_{xy}(n) = \sum_{k=-n}^{n} \gamma_{xy}(k). \tag{44}$$

When n tends to infinity one obtains

$$\lim_{n\to\infty} S_{xy}(n) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) = \Gamma_{xy}(0) = \Lambda_{xy}(0) \tag{45}$$

since $\Psi_{xy}(0) = 0$. So the asymptotic slope of the curve is equal to the initial point of the co-spectrum and of the cross spectrum.

Now for large n holds

$$V_x(n) = n\Gamma_{xx}(0) + o(n) \tag{46}$$

$$V_y(n) = n\Gamma_{yy}(0) + o(n) \tag{47}$$

$$V_{xy}(n) = n\Gamma_{xy}(0) + o(n) \tag{48}$$

where $o(n)$ stands for functions $f(n)$ of n with the property that $\lim_{n\to\infty} f(n)/n = 0$.

So the correlation coefficient becomes

$$R_{xy}(n) = \frac{V_{xy}(n)}{\sqrt{V_x(n)\, V_y(n)}}$$

$$= \frac{n\Gamma_{xy}(0) + o(n)}{\sqrt{\{n\Gamma_{xx}(0) + o(n)\}\,\{n\Gamma_{yy}(0) + o(n)\}}}$$

$$= \kappa_{xy}(0) + o(1) \tag{49}$$

where $o(1)$ tends to zero when n tends to infinity. So the correlation coefficient of n-day totals tends to the initial point of the coherency spectrum.

3.2. Estimation of cross spectra

The sample cross covariance function for a realization $(x_1, y_1), \ldots, (x_N, y_N)$ of a stationary bivariate stochastic process is defined by

$$c_{xy}(k) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{j=1}^{N-k} (x_j - \bar{x})(y_{j+k} - \bar{y}) & , k = 0, 1, \ldots, N-1 \\ \dfrac{1}{N}\displaystyle\sum_{j=-k+1}^{N} (x_j - \bar{x})(y_{j+k} - \bar{y}) & , k = -(N-1), \ldots, -1, 0 \end{cases} \tag{50}$$

where $\bar{x}$ and $\bar{y}$ are the means of the N x_js and N y_js, respectively. The even part of $c_{xy}(k)$ will be denoted by $\ell_{xy}(k)$ and the odd part by $q_{xy}(k)$.

As is the case in univariate spectral analysis the Fourier transform of the sample cross covariance function gives an inconsistent estimate of the cross spectrum. Consistent estimates can be obtained by smoothing. Smoothed estimates of the co-spectrum and the quadrature spectrum follow from

$$\tilde{L}_{xy}(f) = \ell_{xy}(0) + 2 \sum_{k=1}^{N-1} w(k)\ell_{xy}(k) \cos 2\pi fk \quad . -{}^1/_2 < f \leqslant {}^1/_2 \qquad (51)$$

$$\tilde{Q}_{xy}(f) = 2 \sum_{k=1}^{N-1} w(k) q_{xy}(k) \sin 2\pi fk \quad . -{}^1/_2 < f \leqslant {}^1/_2 \qquad (52)$$

where $w(k)$ stands for the lag window. In this study use was made of the Tukey window, defined by Eq. (22). Substitution of smoothed estimates of the spectral densities, the co-spectrum and the quadrature spectrum in Eqs. (40), (41) and (42) gives smoothed estimates $\tilde{A}_{xy}(f)$, $\tilde{F}_{xy}(f)$ and $\tilde{K}_{xy}(f)$ of the cross-amplitude spectrum $\alpha_{xy}(f)$, the phase spectrum $\phi_{xy}(f)$ and the coherency spectrum $\kappa_{xy}(f)$, respectively. One usually plots cross spectral estimates for the non-negative frequencies only. Then the estimates of the co-spectrum, the quadrature spectrum and the cross amplitude spectrum should be multiplied by 2: the estimates of the phase spectrum and the coherency spectrum remain unchanged. As in univariate spectral analysis the variance of cross spectral estimates decreases with M. However, for small values of M the estimates may be considerably biased.

3.3. Cross spectral analysis of Winterswijk and Twente data

3.3.1. *Description of the data*

The aviation-base of Twente (52°16′N, 6°54′E) is situated 36 kilometers from Winterswijk. The data used were daily values for the period Dec. 1952–Nov. 1971. These data were mainly obtained from magnetic tapes of the Royal Netherlands Meteorological Institute. From the positions of these rainfall stations and the time-increment used one may not expect time-shifts between the two rainfall sequences. This means that

a. the phase spectrum and the quadrature spectrum are approximately zero at all frequencies.
b. the cross covariance function (and also the cross correlation function) is approximately an even function of lag.

Therefore the emphasis of this study is stressed on the shape of the coherency spectrum.

Because of seasonal variation coherency spectra were estimated for each season separately. Therefore, instead of Eq. (50), use was made of

$$c_{xy}(k) = \begin{cases} \dfrac{1}{nm}\left\{\sum_{i=1}^{n}\sum_{j=1}^{m-k} x_{ij}y_{i,j+k} - \dfrac{1}{n(m-k)} \times \sum_{i=1}^{n}\sum_{j=1}^{m-k} x_{ij}\sum_{i=1}^{n}\sum_{j=1}^{m-k} y_{i,j+k}\right\}, & k \geqslant 0 \\ \dfrac{1}{nm}\left\{\sum_{i=1}^{n}\sum_{j=1}^{m-k} x_{i,j+k}y_{ij} - \dfrac{1}{n(m-k)} \times \sum_{i=1}^{n}\sum_{j=1}^{m-k} x_{i,j+k}\sum_{i=1}^{n}\sum_{j=1}^{m-k} y_{ij}\right\}, & k < 0 \end{cases} \tag{53}$$

with

n: the number of years (here $n = 19$),
m: the length of the season in days (m takes the values 90, 91 or 92),
x_{ij}: the amount of rainfall on the jth day of the ith year for station X,
y_{ij}: the amount of rainfall on the jth day of the ith year for station Y.

So the lag k cross covariance of a particular season is estimated by shifting one rainfall series by k days and then taking the sample covariance of all pairs of observations belonging to the same year. The same procedure was followed for estimating the autocovariances.

3.3.2. *The estimated coherency spectrum*

Fig. 7 shows smoothed estimates of the coherency spectrum for the winter and summer season. For the winter the coherency estimates are larger, indicating that in winter there is a better correlation between the two rainfall sequences. In both seasons the largest coherence is found at the low frequencies.

A crude approximation of the variance of the smoothed coherency estimates (neglecting for instance the influence of non-normality) is

$$\text{var } \tilde{K}_{xy}(f) \approx 0.375 \frac{M}{N}\{1 - \kappa^2_{xy}(f)\} \tag{54}$$

where N stands for the number of data (about nm). (It is not permitted to apply this approximation in the neighbourhood of $f = 0$ and $f = {}^1/_2$. For these frequencies it is better to take twice the value given by (54)).

To obtain estimates whose variance is approximately independent of $\kappa_{xy}(f)$ Fisher's z-transform was applied.

$$\tilde{Z}_{xy}(f) = {}^1/_2 \ln\{(1 + \tilde{K}_{xy}(f))/(1 - \tilde{K}_{xy}(f))\}. \tag{55}$$

The variance of $\tilde{Z}_{xy}(f)$ is approximately

$$\text{var } \tilde{Z}_{xy}(f) \approx 0.375 \frac{M}{N}. \tag{56}$$

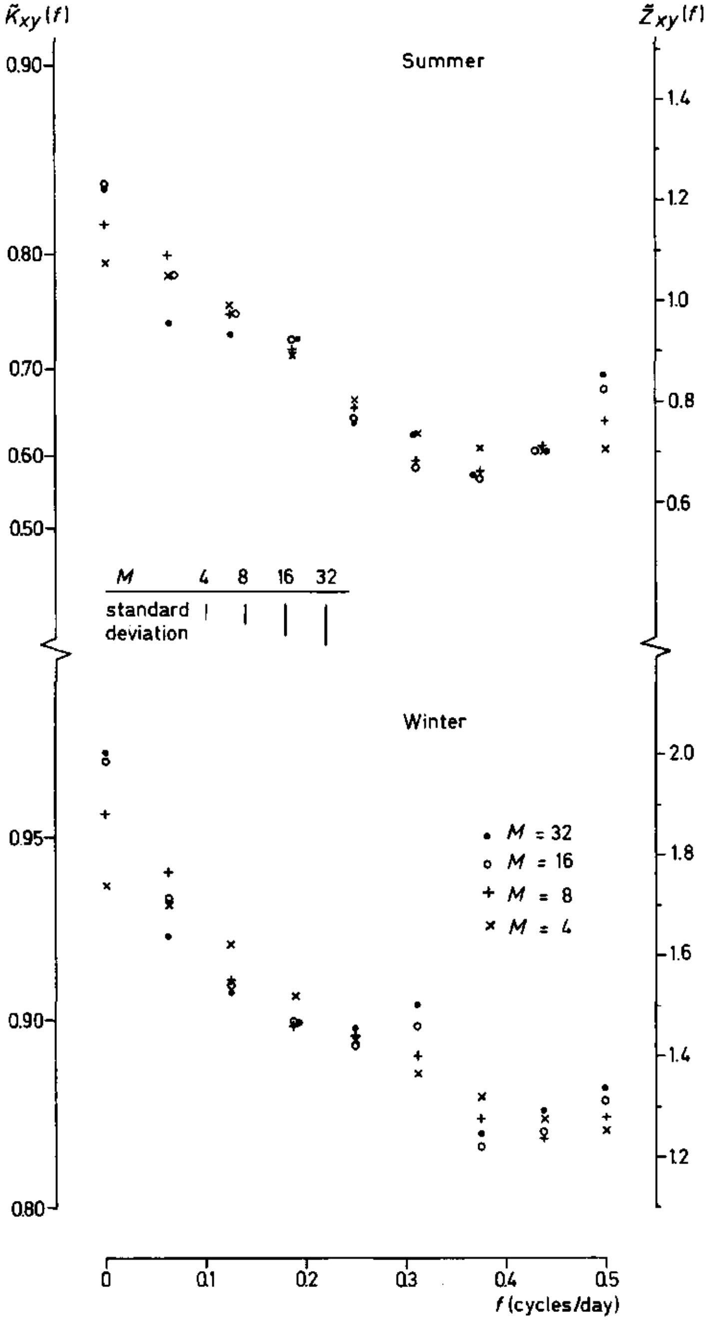

FIG. 7. Smoothed coherency estimates for the rainfall sequences of Winterswijk and Twente.

The standard derivations given in Fig. 7 are based on this approximation.

The estimates at the zero frequency can be compared with correlation coefficients of n-day totals for large n. For 30-day totals the estimated correlation coefficient is 0.962 for the winter and 0.843 for the summer (BUISHAND, 1977b). These values correspond nicely with the initial point of the estimated coherency spectrum for M not too small (e.g. $M = 16$).

3.3.3. *Comment on a simple model for cross correlations*

In hydrology it is often assumed that the cross correlation function is separable in a space and a time component (RODRÍGUEZ-ITURBE and MEJÍA, 1974). That is the cross correlation function takes the form

$$\rho_{xy}(k) = \rho^*(d)\,\rho(k). \tag{57}$$

Here $\rho^*(d)$ denotes the lag zero cross correlation coefficient as a function of the distance d, and $\rho(k)$ denotes the lag k autocorrelation coefficient. Since $\rho^*(d)$ is usually less than 1 the lag k cross correlation coefficient is smaller than the lag k autocorrelation coefficient.

Though this seems a reasonable correlation model it might be impractical for daily rainfall sequences in the Netherlands since its coherency spectrum is constant at all frequencies. This fact follows from the following relation between the cross spectrum $\Gamma_{xy}(f)$ and the autospectra $\Gamma_{xx}(f)$ and $\Gamma_{yy}(f)$

$$\begin{aligned}\Gamma_{xy}(f) &= \sum_{k=-\infty}^{\infty} \rho_{xy}(k)\, e^{-2\pi i f k} \\ &= \rho^*(d) \sum_{k=-\infty}^{\infty} \rho(k) e^{-2\pi i f k} \\ &= \rho^*(d)\,\Gamma_{xx}(f) = \rho^*(d)\,\Gamma_{yy}(f)\end{aligned} \tag{58}$$

(for brevity of notation it is assumed that the variance at both stations equals 1). Inserting this relation in (42) gives that $\kappa_{xy}(f) = \rho^*(d)$ for all frequencies.

3.4. CONCLUDING REMARK

It was noticed that the height of the smoothed coherency estimates decreases with the frequency f. This is an indication for an increase of the correlation coefficient of n-day totals with n. But this information could also be obtained by estimating these correlation coefficients for different values of n (BUISHAND, 1977b). This last technique may be preferred to cross spectral analysis since the results are more readily interpretable.

APPENDIX

A1. The sample spectrum at the zero frequency

From (17) it follows for $C_{xx}(0)$

$$C_{xx}(0) = c_{xx}(0) + 2 \sum_{k=1}^{N-1} c_{xx}(k). \tag{A1}$$

Substitution of (16) results in

$$C_{xx}(0) = \frac{1}{N}\left\{ \sum_{j=1}^{N} (x_j - \bar{x})^2 + 2 \sum_{k=1}^{N-1} \sum_{j=1}^{N-k} (x_j - \bar{x})(x_{j+k} - \bar{x}) \right\}. \tag{A2}$$

Putting $m = j + k$ in the second term in braces and changing the order of summation one obtains

$$\begin{aligned} C_{xx}(0) &= \frac{1}{N}\left\{ \sum_{j=1}^{N} (x_j - \bar{x})^2 + 2 \sum_{m=2}^{N} \sum_{k=1}^{m-1} (x_{m-k} - \bar{x})(x_m - \bar{x}) \right\} \\ &= \frac{1}{N}\left\{ \sum_{j=1}^{N} (x_j - \bar{x})^2 + 2 \sum_{m=2}^{N} \sum_{k=1}^{m-1} (x_k - \bar{x})(x_m - \bar{x}) \right\} \\ &= \frac{1}{N}\left\{ \sum_{j=1}^{N} (x_j - \bar{x}) \right\}^2 \doteq 0. \end{aligned} \tag{A3}$$

A2. The spectral density of an alternating renewal process

First, an expression for the spectral density is derived under the assumption that wet spells have a geometric distribution. Thereafter the result is extended to an alternating renewal process.

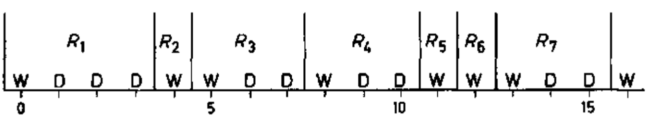

FIG. 8. Realization of a wet-dry process for $t = 0(1)16$. Wet days are denoted by W and dry days by D. The recurrence times R_1, R_2, ... are the waiting times between successive wet days.

Fig. 8 shows a realization of a wet-dry process for $t = 0(1)16$. The waiting times between successive wet days are denoted by R_1, R_2, ... and are called recurrence times. If the recurrence times are idd random variables the process is called a renewal process. Then the process is independent of its history whenever a wet day occurs and the lengths of wet spells have a geometric distribution. Let $\{f_n\}$ denote the distribution of the recurrence times. Further, let u_k be the probability of a wet day at $t = k \geqslant 0$ given that a wet day occurred at $t = 0$. (Note that $u_0 = 1$.)

It is convenient to introduce the generating functions

$$F(s) = \sum_{k=1}^{\infty} f_k\, s^k \tag{A4}$$

$$U(s) = \sum_{k=0}^{\infty} u_k\, s^k. \tag{A5}$$

It can be shown (FELLER, 1968, XIII.3) that the following relation exists between these generating functions

$$U(s) = 1/(1 - F(s)) \quad ,|s| < 1. \tag{A6}$$

Since the autocovariance function $\gamma_{xx}(k)$ is symmetric in k Eq. (2) can be written as

$$\Gamma_{xx}(f) = \gamma_{xx}(0) + \sum_{k=1}^{\infty} \gamma_{xx}(k)\, e^{2\pi ifk} + \sum_{k=1}^{\infty} \gamma_{xx}(k)\, e^{-2\pi ifk}. \tag{A7}$$

For a renewal process it can be shown that (BUISHAND, 1977a, IV.4.1)

$$\gamma_{xx}(k) = \frac{1}{\mu}\left(u_k - \frac{1}{\mu}\right) \quad , k \geqslant 0 \tag{A8}$$

in which μ stands for the mean recurrence time. Substitution of (A8) in (A7) results in

$$\Gamma_{xx}(f) = \frac{1}{\mu}\left\{1 - \frac{1}{\mu} + \sum_{k=1}^{\infty}\left(u_k - \frac{1}{\mu}\right) e^{2\pi ifk} + \sum_{k=1}^{\infty}\left(u_k - \frac{1}{\mu}\right) e^{-2\pi ifk}\right\}. \tag{A9}$$

The problem is that $\Sigma u_k s^k$ and Σs^k do not converge for s on the unit circle. Bur for renewal processes relevant in hydrologic time series $\Sigma \tilde{u}_k s^k$, with $\tilde{u}_k = u_k - 1/\mu$, converges for $|s| = 1$. Let $\tilde{U}(s)$ denote the generating function of the $\tilde{u}_k s$, that is

$$\tilde{U}(s) = \sum_{k=1}^{\infty} u_k\, s^k \tag{A10}$$

then the power spectrum becomes

$$\Gamma_{xx}(f) = \frac{1}{\mu}\left\{1 - \frac{1}{\mu} + \tilde{U}\,(e^{2\pi if}) + \tilde{U}\,(e^{-2\pi if})\right\}. \tag{A11}$$

For $|s| < 1$ holds

$$\tilde{U}(s) = U(s) - 1 - \frac{s}{\mu(1-s)} = \frac{F(s)}{1-F(s)} - \frac{s}{\mu(1-s)} \tag{A12}$$

using (A6).

Since $\Sigma \tilde{u}_k s^k$ converges for $|s| \leqslant 1$, $\tilde{U}(s)$ is a continuous function for $|s| \leqslant 1$ (Abel's theorem) and hence

$$\tilde{U}(e^{2\pi i f}) = \frac{F(e^{2\pi i f})}{1 - F(e^{2\pi i f})} - \frac{e^{2\pi i f}}{\mu(1 - e^{2\pi i f})} \tag{A13a}$$

$$\tilde{U}(e^{-2\pi i f}) = \frac{F(e^{-2\pi i f})}{1 - F(e^{-2\pi i f})} - \frac{e^{-2\pi i f}}{\mu(1 - e^{-2\pi i f})}. \tag{A13b}$$

Substitution of (A13) in (A11) gives

$$\Gamma_{xx}(f) = \frac{1}{\mu}\left\{1 + \frac{F(e^{2\pi i f})}{1 - F(e^{2\pi i f})} + \frac{F(e^{-2\pi i f})}{1 - F(e^{-2\pi i f})}\right\}. \tag{A14}$$

This relation corresponds to the expression given by BARTLETT (1963) for the spectrum of counts in a continuous renewal process.

One can not apply Eq. (A14) directly for those values of f for which $F(e^{2\pi i f}) = 1$. For renewal processes encountered in hydrology (with a so-called non-periodic distribution of the recurrence times) this situation only occurs for $f = 0$. But at the zero frequency the power spectrum equals the asymptotic slope of the variance-time diagram, see Eq. (15). From renewal theory it is well known (see for instance FELLER, 1968, XIII. 6) that this asymptotic value equals σ^2/μ^3 in which σ^2 stands for the variance of the recurrence times.

The result given above can be extended for application to an arbitrary alternating renewal process. Let $\{f_n^{(w)}\}$ and $\{f_n^{(d)}\}$ stand for the distributions of the lengths of wet and dry spells: their generating functions are denoted by $F^{(w)}(s)$ and $F^{(d)}(s)$, respectively. The means of the lengths of wet and dry spells are denoted by $\mu^{(w)}$ and $\mu^{(d)}$, and the variances by $\mu_2^{(w)}$ and $\mu_2^{(d)}$, respectively. Further, for a stationary alternating renewal process one can define the conditional probability $h_k^{(wd)}$ of a dry day at $t = k \geqslant 0$, given that a wet day occurred at $t = 0$. (Note that $h_0^{(wd)} = 0$.) An analogous definition can be given for the probability $h_k^{(ww)}$. It is convenient to introduce the generating functions

$$H^{(wd)}(s) = \sum_{k=0}^{\infty} h_k^{(wd)} s^k \tag{A15a}$$

$$H^{(ww)}(s) = \sum_{k=0}^{\infty} h_k^{(ww)} s^k. \tag{A15b}$$

These functions can be expressed in the generating functions of the lengths of wet and dry spells (BUISHAND, 1977a, IV.2.2), for instance

$$H^{(wd)}(s) = \frac{s\,\{1 - F^{(w)}(s)\}\,\{1 - F^{(d)}(s)\}}{\mu^{(w)}(1-s)^2\,\{1 - F^{(w)}(s)F^{(d)}(s)\}}. \tag{A16}$$

Since $h_k^{(wd)} + h_k^{(ww)} = 1$, one has

$$H^{(wd)}(s) + H^{(ww)}(s) = 1/(1-s). \tag{A17}$$

For the autocovariances it can be shown that (BUISHAND, 1977a, IV.4.1)

$$\gamma_{xx}(k) = q^{(w)}\,(h_k^{(ww)} - q^{(w)}) \quad , k \geqslant 0 \tag{A18}$$

in which $q^{(w)} = \mu^{(w)}/(\mu^{(w)} + \mu^{(d)})$ stands for the probability of a wet day in a stationary alternating renewal process.

Set $\tilde{h}_k^{(ww)} = h_k^{(ww)} - q^{(w)}$ and let $\tilde{H}^{(ww)}(s)$ be the generating function

$$\tilde{H}^{(ww)}(s) = \sum_{k=1}^{\infty} \tilde{h}_k^{(ww)} s^k \tag{A19}$$

then the power spectrum becomes

$$\Gamma_{xx}(f) = q^{(w)}\{1 - q^{(w)} + \tilde{H}^{(ww)}(e^{2\pi i f}) + \tilde{H}^{(ww)}(e^{-2\pi i f})\}. \tag{A20}$$

For $|s| < 1$ holds

$$\begin{aligned}\tilde{H}^{(ww)}(s) &= H^{(ww)}(s) - 1 - q^{(w)}s/(1-s)\\ &= -H^{(wd)}(s) + (1 - q^{(w)})\,s/(1-s).\end{aligned} \tag{A21}$$

Since $\Sigma \tilde{h}_k^{(ww)} s^k$ converges for $|s| = 1$ it is permitted to apply this result on the unit circle. Then the power spectrum becomes

$$\Gamma_{xx}(f) = -q^{(w)}\{H^{(wd)}(e^{2\pi i f}) + H^{(wd)}(e^{-2\pi i f})\} \tag{A22}$$

in which $H^{(wd)}(e^{2\pi i f})$ and $H^{(wd)}(e^{-2\pi i f})$ can be obtained from (A16).

This relation can not be applied directly for $f = 0$. For this frequency holds

$$\Gamma_{xx}(0) = \frac{(\mu^{(d)})^2\,\mu_2^{(w)} + (\mu^{(w)})^2\,\mu_2^{(d)}}{(\mu^{(w)} + \mu^{(d)})^3} \tag{A23}$$

being the asymptotic slope of the variance-time diagram (BUISHAND, 1977a, IV.5.2).

As an example assume that the wet-dry process is a Bernoulli process. That is the probability of a day being wet or dry does not depend on the situation of previous days. Let p denote the probability of a day being wet and $q = 1 - p$. For $k \geqslant 1$ holds $h_k^{(wd)} = q$ and $h_k^{(ww)} = p$, since the process has no memory. Then it follows

$$H^{(wd)}(s) = qs/(1-s) \tag{A24}$$

and thus the power spectrum becomes

$$\Gamma_{xx}(f) = -p\left\{q\,\frac{e^{2\pi i f}}{1 - e^{2\pi i f}} + q\,\frac{e^{-2\pi i f}}{1 - e^{-2\pi i f}}\right\} = pq. \tag{A25}$$

So the power spectrum is constant at all frequencies (Note that *pq* is just the variance of the process.)

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to A. OTTEN who made some valuable comments.

REFERENCES

BARTLETT, M. S. (1963). The Spectral Analysis of Point Processes. Journal of the Royal Statistical Society, Series B, **25**, 264–296.

BUISHAND, T. A. (1977a). Stochastic Modelling of Daily Rainfall Sequences. Mededelingen Landbouwhogeschool Wageningen 77-3.

BUISHAND, T. A. (1977b). De variantie van de gebiedsneerslag als functie van puntneerslagen en hun onderlinge samenhang. (The variance of the amount of rainfall on an area related to point rainfall amounts). Mededelingen Landbouwhogeschool Wageningen 77–10.

FELLER, W. (1968). An Introduction to Probability Theory and its Applications, Vol. I, Third edition, John Wiley & Sons, Inc., New York.

JENKINS, GWILYM M. and WATTS, DONALD G. (1969). Spectral Analysis and its applications. Holden-Day, San Francisco.

LEVERT, C. (1960). Twee statistische modellen van persistentie in klimatologische tijdreeksen (Two statistical models of persistency in climatological time series). WR 60-4 (III–251) KNMI, De Bilt.

NEAVE, HENRY R. (1972). A Comparison of Lag Window Generators. Journal of the American Statistical Association, **67**, 152–158.

RODRÍGUEZ-ITURBE, IGNACIO and MEJÍA, JOSÉ M. (1974). On the Transformation of Point Rainfall to Areal Rainfall. Water Resources Research, **10**, 729–735.