A Dissertation on

A WAVELET BASED SEARCH FOR PERIODICITIES IN INDIAN MONSOON RAINFALL TIME SERIES

Submitted in partial fulfillment of the requirements
for the award of the degree of

DOCTOR OF PHILOSOPHY
in
MATHEMATICS

by

SARITA AZAD

DEPARTMENT OF APPLIED MATHEMATICS,
DELHI COLLEGE OF ENGINEERING,
FACULTY OF TECHNOLOGY,
UNIVERSITY OF DELHI
INDIA

JANUARY 2008

Acknowledgments

I express my deepest sense of gratitude to my supervisors Late Dr.S.K. Sett and Dr. R. Srivastava for their invaluable guidance and support. I sincerely thank Prof. D. Pental, Vice Chancellor, Delhi University and extend my heartfelt thanks to Prof. D. Goldar, Principal, Delhi College of Engineering. I am thankful to Prof. Promod Kumar, Dean, Faculty of Technology, Delhi University, for helping me get through all the critical times. I also thank Prof. H.C. Taneja, Head of Department of Applied Mathematics and the entire faculty for extending their cooperation and support.

I owe sincere gratitude to my spiritual mentor Dr. Daisaku Ikeda who made me understand the following lines "Always ask yourself: for what purpose should I cultivate wisdom".

Thanks to the kind permission and encouragement I received from Dr.Sett, and his initiative in making the appropriate arrangements, that I have been able to carry out much of the work on my thesis with the help of Prof. Roddam Narasimha, FRS, Engineering Mechanics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore. During my stay at JNCASR I understood the meaning of education and got innumerable opportunities to learn and grow in my subject. I feel indebted to Prof. RN for the meticulous training I received from him not limited to the field of science; and it helped me to develop my humanity. I will be forever thankful to him for his guidance and blessings. I am ever so thankful to Mrs. Narasimha for her warm hospitality.

I am thankful to the Centre for Atmospheric and Oceanic Sciences and Super Computer Education and Research Centre of the Indian Institute of Science(IISc), Bangalore for allowing me to use their facilities. And I am extremely grateful to the JNCASR for providing me immense help and an assistantship with Prof. Narasimha

which supported part of my work in the last four years. I am also thankful to all members of the academic and administrative section of JNCASR for their support and help. I am grateful to the administrative officer, Mr. A. Jaychandra for providing me accommodation in Jawahar visiting guest house. Prof. Anindya Chatterjee of IISc, Prof. Rama Govindarajan of JNCASR and Prof. A.S. Vasudeva Murthy of TIFR Bangalore, have always been a great source of moral support and inspiration. I owe sincere gratitude to Prof. Govindan Rangarajan, Department of Mathematics, IISc, for enabling me to attend various courses in the department. I thank many friends and students at JNC and IISc for their help at various times, and in particular Vignesh of ISI Kolkata, who during his summer research visits to RN was very generous with his knowledge of statistics. I am also thankful to Dr. Subarna Bhattacharyya for all the discussions I had with her. I am thankful to Prof. R.N.Iyengar, IISc, and Prof P.G. Vaidya, National Institute of Advanced Studies, Bangalore, for very useful discussions.

I am indebted to my parents and my brother who inspired me to have aims in life as did my dearest sisters. Friends who helped me immensely throughout this period are Ankur, Ekta, Archana, Preet Sehgal, and Durga. I express my sincere gratitude to Ms. K. Nagarathna for her constant support. In the end I am thankful to my husband Rahul who has been a great support and a constant source of encouragement.

Sarita Azad

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Science, engineering, medicine and many other areas deal with signals acquired in the form of time series [1, 2]. The purpose of time series analysis is to quantify certain features such as periodicities, spectra, attractor dimensions, extreme events, singular points etc. Knowledge of these properties can be useful in efforts towards prediction. The time series data may be modeled as stochastic processes [3], signals mixed with noise [4], deterministic chaos [5] etc. In fact, time series analysis is a very wide subject, and its historical development can be traced back to two main sources, namely, engineering and mathematical statistics. As a result, the subject is influenced with both engineering and statistical concepts and terminology, the former being associated with "spectral" or "frequency domain" approach [6], [7] and the later with "correlation" or "time domain"approach [8]. For the purpose of the present work the frequency domain approach is chiefly adopted. In addition, only direct estimates based on the discrete Fourier transform are pursued. The most important task of Fourier analysis is to estimate the power spectrum of a stationary time series from its autocorrelation function. This has been a classical problem, ever since Wiener [9] showed the Fourier transform relationship between the power spectrum and the autocorrelation function.

Though the Fourier transform has revolutionized the field of time series analysis after the invention of the Fast Fourier Transform (FFT), there are several inherent limitations of the FFT approach. The most prominent limitation is that of frequency resolution, i.e., the inability to distinguish between two spectral peaks

spaced closer than $1/N$, N being the total number of samples in the time series [10]. For such closely spaced frequencies, the so-called multiple-window method has been introduced by [11]. The excellent performance of this method for stationary signals has led several groups to apply the method to time-varying spectrum estimation. Closely spaced frequencies have been successfully separated by [12] and [13]. The asymptotic theory of [14] shows that, in the limit of long series (of duration T, say), frequencies can be separated to $O(T^{-1})$ as $T \to \infty$. References [15], [16] discuss possible frequency separation in terms of signal-to-noise ratio (SNR) and sample length N, using the Cramer-Rao bound.

In the case of the Indian monsoon rainfall data considered here, there is no simple analogue of SNR, and the number of years over which accurate rainfall data are available is $N \sim O(10^2)$, not very large. In the absence of relevant theory, therefore, there is a need to devise new methods whose validity is supported by test cases that mimic the real problem.

For random processes containing multiple scales wavelet based time-scale methods have become useful in recent years [17, 18, 19]. The foundation of wavelets rests on Thomson's key insight that to smooth the spectrum estimate, orthogonal windows, concentrated in the smallest possible region in the domain of interest should be used. These wavelets, expressed in terms of multi-resolution analysis (MRA), then provide a useful tool for decomposing a signal into dyadic scales (i.e. on powers of two); this makes wavelets orthogonal, hence much more efficient. In particular multi-resolution analysis can play a significant role in understanding the spatial and temporal structure of rainfall. In the last decade wavelet based techniques have attracted considerable attention in this area [20, 21, 22, 23, 24]

Most real world problems (e.g. meteorological variables, ECG signals, financial forecasting) exhibit non-stationary behavior. For analyzing the associated time series, Fourier analysis is insufficient since its frequency content does not change with time. The spectral analysis of non-stationary signals can not describe the local transient features due to averaging over the duration of the signal. The inability of the method to describe the evolutionary characteristics of non-stationary processes has lead to a search for tools which allow time and frequency localization beyond customary Fourier analysis. An FFT based method called the short term Fourier

transform (STFT) provides time and frequency localization to establish a local spectrum for any time instant [25]. The key feature of the STFT is the application of the Fourier transform to a time varying signal when the signal is viewed through a narrow window centered at time t. The local frequency content is then obtained at time t. The window is moved to a new time and the process is repeated. High resolution can not be obtained in both time and frequency domains simultaneously by such methods. The window must be chosen for locating both sharp peaks and low frequency features, but the inverse relation between window length and the corresponding frequency band width makes a good choice difficult.

This drawback can be alleviated if one has the flexibility to allow the resolution in time and frequency to vary in a time-frequency plane to reach a multi-resolution representation of the process. Accordingly, the time-frequency window would narrow automatically to observe the high frequency content of a signal and widen to capture low frequency phenomena. This is possible if the analysis is viewed as a set of filter banks consisting of band-pass filters with a constant relative band width. The wavelet transform has turned out to be an excellent mathematical tool for providing such multi-resolution descriptions of non-stationary time series. It has been developed [17, 26] to study a large class of phenomena such as image processing, data compression, chaos, fractals, etc. The wavelet transform has the key property of localization in time (or space) and in frequency (or wave number), contrary to what happens in classical Fourier methods. In fact, the wavelet transform works as a mathematical microscope on a specific part of a signal to extract local structures and singularities. This makes wavelets ideal for handling non-stationary and transient signals as well as fractal-type structures.

1.1 Description of Theory and Tools

In this section the methods used in the thesis are briefly discussed, along with the associated mathematical expressions.

1.1.1 Auto-correlation function

An important guide to the properties of a time series is provided by a series of quantities called sample autocorrelation coefficients [8], which measure the correlation between observations separated by a specific interval in time. These coefficients often provide insight into the probability model that might have generated the data. For a discrete time series $x(t)$ of length N, $t = 0, 1, .., N-1$, we can find the correlation between observations a time interval k apart, which is given by

$$r_k = \frac{\sum_{t=0}^{N-k}(x(t) - \overline{x})(x(t+k) - \overline{x})}{\sum_{t=0}^{N-1}(x(t) - \overline{x})^2}, \tag{1.1}$$

where $\overline{x} = \sum_{t=0}^{N-1} x(t)/N$ = overall mean. This is called the autocorrelation coefficient at lag k. The autocorrelation is a function useful for finding repeating patterns in a time series, such as the presence of a periodic signal which has been buried under noise, or identifying the fundamental frequency of a signal which does not actually contain that frequency component, but implies it with many harmonic frequencies. Autocorrelations are frequently plotted as an autocorrelation function(or acf) against lag k. For a stationary series, the autocorrelation between two points in time, t and $t+k$, becomes smaller as k increases. In other words, the autocorrelation function falls off rather quickly as k increases. However for a non-stationary series this is usually not the case; the autocorrelation function does not become small as k increases.

1.1.2 Power spectral density

A time series can have important characteristics which can be expressed in terms of mathematical functions derived as transforms of the time series. Fourier analysis of a time series is one of the oldest and most widely used techniques in the physical

sciences. The basic idea behind it is to decompose a time series into a number of components, each one of which can be associated with a particular frequency. To define the Fourier transform, let $x(t), t = 0, 1, .., N-1$, denote a discrete-time data sequence of length N. Assume that $x(t)$ has finite energy, which means that

$$\sum_{t=-\infty}^{\infty} \mid x(t) \mid^2 < \infty. \tag{1.2}$$

Then the sequence $x(t)$ possesses a discrete-time Fourier transform (DTFT) defined as

$$x(\omega_k) = \sum_{t=0}^{N-1} x(t) e^{-i\omega_k t} \tag{1.3}$$

where (angular) frequency $\omega_k = 2\pi k/N$, $k = 0, 1, .., N-1$. The corresponding inverse DTFT is then

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} x(\omega_k) e^{i\omega_k t}. \tag{1.4}$$

Both deterministic and stochastic processes can, in principle, be characterized by what is called the power spectrum (or periodogram in the engineering literature) and the power spectral density (PSD) function in the mathematical literature.

The peaks in the power spectrum can give important information that may shed light on the physics of the phenomenon underlying the series $x(t)$. As an example consider daily temperature measurements. The spectrum would evidently contain a peak at a period of one year corresponding to sun's annual migration from the Tropic of Cancer to Tropic of the Capricorn and back. Annual cycles are present in many natural systems and also in the economic system. Along with the annual component there may be several harmonics, half year, quarter year, etc. These arise on account of the fact that the periodic phenomenon is not a pure sinusoidal. In fact one of the first uses of the PSD function was in determining possible "hidden periodicities"in time series by [27]. The PSD function of such a time series can be estimated as

$$\hat{x}(\omega_k) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x(t) e^{-i\omega_k t} \right|^2. \tag{1.5}$$

where $\omega_k = 2\pi k/N$, and $k = 0, 1, ..., N-1$ is the frequency index. These PSD ordinates are symmetric around $N/2$, assuming N even [10], so it is enough to limit k to $N/2$. The value of $\hat{x}(\omega_k)$ is a measure of the contribution to the 'energy' of x by the frequency ω_k. Therefore, the normalization of PSD is given by

$$\sum_{k=0}^{N/2} 2\hat{x}(\omega_k)/\sigma^2 = 1 \tag{1.6}$$

where σ^2 (variance of the time series) is defined as

$$\sigma^2 = \frac{1}{N-1}\sum_{t=0}^{N-1}(x(t) - \overline{x(t)})^2, \tag{1.7}$$

where $\overline{x(t)}$ is the mean of the time series.

1.1.3 Welch periodogram method

The periodogram defined by Eq. 1.5 is an inconsistent estimator of true PSD values [10]. Also, it suffers from large variance. Several methods, traceable to the work of Bartlett, Deniell, Welch, Blackman-Tukey etc., have been devised to obtain good estimates of the PSD function (for details see [7]). We explain here the powerful technique introduced in [28].

To describe the Welch method in a mathematical form (following the explanation in [7]), let

$$x_j(t) = x((j-1)K + t), \tag{1.8}$$

where $t = 0, 1, .., N-1; j = 0, 1, .., S-1$ and $(j-1)K$ is the starting point for the jth observation of the time series $x(t)$. The value recommended for K in the method is $K = N/2$, in which case $S \simeq 2N/M$ data segments (with 50% overlap between successive segments) are obtained.

The windowed periodogram corresponding to $x_j(t)$ is computed as

$$\hat{x_j}(\omega_k) = \frac{1}{NP}\left|\sum_{t=0}^{N-1} \upsilon(t)x_j(t)e^{-i\omega_k t}\right|^2, \tag{1.9}$$

where P denotes the "power "of the temporal window $\upsilon(t)$:

$$P = \frac{1}{N}\sum_{t=0}^{N-1} |\upsilon(t)|^2. \tag{1.10}$$

The Welch estimate of PSD is determined by averaging the windowed periodograms in Eq. 1.9:

$$\hat{x}_W(\omega_k) = \frac{1}{S}\sum_{j=0}^{S-1} \hat{x_j}(\omega_k). \tag{1.11}$$

By allowing overlap between the data segments and hence by getting more periodograms to be averaged in (Eq.1.11), the variance of the estimated PSD is decreased. Also the purpose of introducing the window in the periodogram computation is for gaining control over the bias/resolution properties of the estimated PSD. Therefore, if the data contains sinusoidal components with significantly different amplitudes, then it is good to use a windowed PSD which makes it possible to detect low amplitude components.

1.1.4 Significance tests

Many real-life time series include a combination of periodic or nearly periodic signals embedded in a noise-like background with a continuous spectrum. The first step in any investigation of the temporal structure of such time series is to test the significance of the peaks obtained in the spectrum. For this purpose, a discrete finite reference time series is necessary. The classical statistical model for such a reference is the first-order autoregressive (AR1) process

$$X(t) = \alpha X(t-1) + \varepsilon(t), \tag{1.12}$$

where $t = 1, ...N$ denotes discrete time in units of the sampling interval, $X(0)$ may be taken as 0, α is the lag-1 autocorrelation coefficient ($0 \leq \alpha < 1$), and $\varepsilon(t)$ is a Gaussian white noise process with mean zero and an appropriate variance σ_ε^2. Following [29] the normalized power spectral density function of $X(t)$ is

$$P_k = \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi k/N)}. \tag{1.13}$$

Here $k = 0, 1, .., N/2$ is the frequency index. The shape of the spectrum P_k depends on the value of α. When $\alpha > 0$ (/ < 0), power is concentrated at lower (/ higher) frequencies, giving what is known as a red (/ blue) spectrum, as shown in figure 1.1. When $\alpha = 0$ we obtain the white-noise spectrum with an expectation value of 1 at all frequencies.

If a peak in the power spectrum is significantly above this reference spectrum, then the hypothesis that the peak is a result of statistical fluctuation can be subjected to the χ^2 test with a specified confidence level. Here the number of degrees of freedom is taken as 2, because if $x(t)$ is a normally distributed random variable, then both the real and imaginary parts of its Fourier spectrum are normally distributed [8]. Since the square of a normally distributed variable is chi-square distributed with one degree of freedom, the complex power spectrum is chi-square distributed with two degrees of freedom, denoted by ${\chi_2}^2$. To determine (for example) the 95% confidence level (significant at 5%), one multiplies the reference spectrum of Eq. 1.13 by the 95th percentile value for the χ^2 distribution with two degrees of freedom [29], and checks whether the peak in question exceeds the 95th percentile.

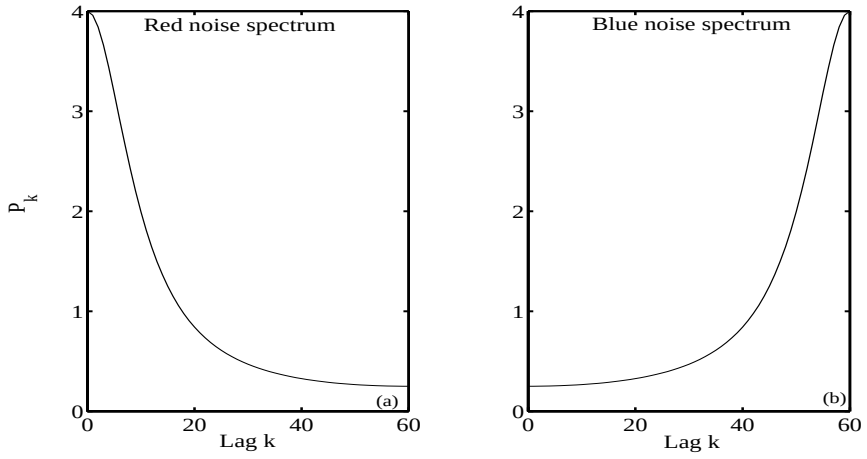

Figure 1.1: (a) Reference spectrum P_k for (a) $\alpha = 0.6$ and (b) $\alpha = -0.6$.

1.1.5 Wavelet analysis

In recent years, the wavelet transform has become an important tool for analysis in areas as diverse as differential equations, image processing and statistics. In this thesis its application in the field of time series [30] is explored. Wavelet analysis, as a relatively new technique, is an important addition to standard signal analysis methods as mentioned recently in the works of [31, 32, 33].

The wavelet transform uses a two-parameter family of basis functions and preserves temporal information along with an indication of dominant frequencies. Wavelets can be either continuous or discrete. The first theoretical results in wavelets had been concerned with continuous wavelet decompositions of functions and go back to the early 1980's [34] [35]. In the continuous case, calculating wavelet coefficients at every possible scale is a fair amount of work, and moreover generates considerable redundancy. However it is often easier to recognize patterns or structures in data using a continuous transform, as shown in [21]. But if the goal is to compress information it is preferable to choose wavelet time-scales based on powers of two (i.e adopt dyadic scaling). This makes wavelets orthogonal, hence much more efficient. They are then known as discrete wavelet transforms (DWT) [36]. These wavelets, expressed in terms of multi-resolution analysis (MRA) [37], then provide a useful tool for approximating such signals as speech, turbulence etc. which contain transient or localized phenomena. Just as the fast Fourier transform made the Fourier transform a practical tool for spectral analysis, multi-resolution analysis has made the discrete wavelet transform a viable tool for computational time-frequency analysis.

The continuous wavelet transform (CWT) of a signal $x(t)$ is given by

$$W(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \overline{\Psi \frac{(t-b)}{a}} dt, \tag{1.14}$$

where b is a shift parameter ($b \in R$) and a is scale parameter ($a \in R^{+}$); the function $\Psi \in L^2(R)$ satisfies the "admissibility condition "

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)^2|}{\omega} d\omega < \infty. \tag{1.15}$$

To find the periodicities in a time series CWT has been extensively used by many workers [38, 39, 40]. Reference [41] have presented a wavelet analysis of sunspot numbers where the principle of detecting periodicities in the signal is discussed. From the peak positions of the wavelet transforms of the sunspot numbers the singular points in the sunspot activity are located. From the time interval between the successive singular points the period of sunspot variation on the given scale is then determined. In fact continuous wavelet analysis has become a common tool for analysing variations of power localized in time within a time series. By decomposing a time series at various scales one is able to determine both the dominant periodicities and their variation in time.

In the discrete case, a and b are restricted to dyadic dilation and translation only, that is, $a = 2^j$ and $b = k$ where $j, k \in Z$:

$$\Psi_{j,k}(t) = 2^{1/2}\Psi(2^j t - k). \tag{1.16}$$

The index j specifies the scale (the level of dilation), and k the location (shift).

The wavelet basis is not unique. A specific function, called a mother wavelet, generates a set of functions called daughter wavelets that can form a basis. For any specially chosen mother wavelet $\Psi(t)$, the corresponding daughter wavelets $\Psi_{j,k}(t)$ form an orthogonal basis in a certain function space so that a function or time series $x(t)$ in that space can be uniquely decomposed as

$$x(t) = \sum_j \sum_k d_{j,k}\Psi_{j,k}(t). \tag{1.17}$$

The coefficient $d_{j,k}$ is found from

$$d_{j,k} = \int_{-\infty}^{\infty} x(t)\Psi_{j,k}\ dt. \tag{1.18}$$

The discrete wavelet transform provides a mechanism to represent data or time series in terms of coefficients that are associated with particular scales and has been found to be a fairly effective instrument for signal analysis.

1.1.6 Multi-resolution wavelet analysis

The MRA concept was initiated in [37, 26]. In the case of discrete wavelets a time series with sample size N can be decomposed at J levels or scales $2^j, j = 1, 2, ..., J{=}\log_2 N$, if the sampling interval is taken as one.

The starting point for such a decomposition (and for the construction of the wavelet bases) is the so-called scaling function. This function should preferably have compact support. Its integer translates $\phi(t-k), k \in Z$ (integer), span a space V_0. A finer space V_j is spanned by the integer translates of the scaled function $\phi(2^j t - k)$. Scaling by 2^j provides the basis functions for the space V_j, and such nesting of the spaces yields a scaling function

$$\phi(t) = \sum_{k \in Z} s_k \phi(2^j t - k) \tag{1.19}$$

with appropriate coefficients $s_k, k \in Z$. The mother wavelet ψ is obtained by building a linear combination of the scaling functions. In order to satisfy orthogonality the coefficients in the linear combination should be properly selected, i.e.

$$\langle \phi(. - k), \psi(. - l)) \rangle = 0, \qquad l, k \in Z \tag{1.20}$$

where angular brackets denote the inner product, defined as $\sum_{k,l} \phi(t-k)\overline{\psi(t-l)}$. The scaling equation (1.19) and the orthogonality condition (1.20) lead to conditions on coefficients q_k which characterize a mother wavelet as a linear combination of the scaled and dilated scaling function ϕ:

$$\psi(t) = \sum_{k \in Z} q_k \phi(2^j t - k). \tag{1.21}$$

Given a fixed mother wavelet, the whole wavelet basis is obtained by scaling and translating the function ψ. The scaled, translated and normalized version of ψ is denoted by

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \tag{1.22}$$

With an orthogonal wavelet ψ the set $\{\psi_{j,k} | j, k \in Z\}$ is an orthogonal wavelet basis. Similarly,

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k). \tag{1.23}$$

As shown in [36], if the wavelet coefficients are given by:

$$a_{J,k} = \langle x(t), \phi_{J,k}(t) \rangle, \tag{1.24}$$

$$d_{j,k} = \langle x(t), \psi_{j,k}(t) \rangle \tag{1.25}$$

then a multi-resolution representation of the signal $x(t)$ can be given by:

$$x(t) = \sum_k a_{J,k}\phi_{J,k}(t) + \sum_j \sum_k d_{j,k}\psi_{j,k}(t) \tag{1.26}$$

where $j = 1, 2, ..., J$ such that J is the maximum scale obtained from the data in hand. Note that when the number of observations is dyadic, the number of coefficients of each type is given as follows:

- at the finest scale 2^1 there are $N/2$ coefficients labeled $d_{1,k}$;
- at the next scale 2^2 there are $N/2^2$ coefficients labeled $d_{2,k}$;
- at the coarsest scale 2^J there are $N/2^J$ coefficients $d_{J,k}$ and $a_{J,k}$.

Therefore, in practice the first step in MRA corresponds to mapping a given time series $x(t), t = 1, ..., N$, to its wavelet coefficients. From these coefficients two components are recovered, first a smoothed version (the so-called approximation), and a second that corresponds to the deviations or the so called details of the signal. The procedure is repeated on the approximations at each scale and hence Eq. 1.26 is obtained at the Jth level of decomposition.

Such a decomposition may be obtained for any specific choice of wavelet. Different families of wavelets have proven to be especially useful in different types of application. They differ with respect to orthogonality, smoothness and other related properties such as vanishing moments or size of support. In particular, to study the frequency resolution of a signal, the Meyer wavelet has been reported to be more suitable [42]. Also, the discrete Meyer wavelet has many advantages (when compared with the Haar, Daubechies2, coiflet2 and Symlet8 wavelets). The most prominent of these advantages is that the Meyer wavelet is windowed in frequency [36], has a smooth Fourier transform, and yields a faster decay of wavelet coefficients

in the time domain. Therefore the discrete Meyer wavelet is preferred in the present work. It is defined in the frequency domain [36] by:

$$\begin{aligned}\hat{\psi}(\omega) &= (2\pi)^{-1/2} e^{i\omega/2} \sin(\frac{\pi}{2}\nu(\frac{3}{2\pi}|\omega| - 1)) \quad \text{if} \quad 2\pi/3 \leq |\omega| \leq 4\pi/3 \\ &= (2\pi)^{-1/2} e^{i\omega/2} \sin(\frac{\pi}{2}\nu(\frac{3}{4\pi}|\omega| - 1)) \quad \text{if} \quad 4\pi/3 \leq |\omega| \leq 8\pi/3 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{1.27}$$

where $\nu(a) = a^4(35 - 84a + 70a^2 - 20a^3)$, $\qquad a \in [0, 1]$.

Definition of MRA using vector spaces.

MRA is a general framework for constructing orthonormal bases. Developed in [37], it has been widely accepted in wavelet research. The following explanation is taken from [43]. Mathematically MRA essentially amounts to constructing a hierarchy of approximations to a function in various subspaces of a linear vector space. To begin, a vector x_N, given in an N- dimensional vector space V_N, can be approximated by a vector x_{N-1} in an $(N-1)$- dimensional vector space V_{N-1}. The best possible approximation for x_N in the least squares sense is made by choosing that vector as x_{N-1} for which the length of the error vector $e_{N-1} \equiv x - x_{N-1}$ is minimum. By the projection theorem stated in [44], the vector e_{N-1} representing minimum error will be orthogonal to every vector in V_{N-1}. Also the vector e_{N-1} can be seen as the amount of "detail" that is lost in going to the "approximation" x_{N-1} for x. As this process continues throughout the entire sequence of subspaces by projecting x_{N-1} on V_{N-2} to yield x_{N-2} etc., a sequence of orthogonal projections of X in the subspaces V_{N-1}, V_{N-2} ....V_1 as x_{N-1}, x_{N-2} ....x_1 respectively is generated. These vector spaces form a nested sequence of subspaces,

$$V_1 \subset V_2 \subset .....V_{N-1} \subset V_N. \tag{1.28}$$

Looking at the projections as approximations to the original vector x, the subspace V_{N-1} contains the finest approximation to vectors in V_N whereas V_1 contains the coarsest approximation. The vector space associated with e_{N-1} is taken as W_{N-1}.

The dimension of W_{N-1} is one. It is shown in [43] that V_{N-1} and W_{N-1} are orthogonal to each other. Any vector in V_N can be expressed as the sum of a vector in V_{N-1} and a vector in W_{N-1}. This is written as

$$V_N = W_{N-1} \oplus V_{N-1}. \tag{1.29}$$

As this process continues, it can be written as

$$V_N = W_{N-1} \oplus W_{N-2} \oplus .....W_1 \oplus V_1. \tag{1.30}$$

Hence any vector x in V_N can be fully reconstructed with details obtained at all levels and the last approximation in V_1, which happens to be the average of the signal over time. The direct sum in Eq. 1.30 represents perfect separation in discrete orthogonal subspaces.

1.1.7 Empirical mode decomposition

A new approach to signal analysis, called empirical mode decomposition (EMD), has been developed in [45] for analysis of nonlinear and nonstationary time series. Here the basis functions, called intrinsic mode functions (IMF), are extracted from the data itself, and their instantaneous frequencies can be calculated using Hilbert spectra. As the procedure generates mode functions with the same number of zeros and extrema, a unique frequency can be associated with each IMF, even though the functions may not be periodic.

The procedure to derive the intrinsic mode functions is the following [45]. The data or time series $\{x(t), t = 0, ..., N-1\}$ is decomposed into independent intrinsic mode functions by a so-called sifting process. The process first identifies all the local maxima and minima of the signal separately. The local maxima are then connected by a cubic spline to define the upper envelope. The procedure is repeated for the local minima to produce the lower envelope. Their mean $m_1(t)$ is determined, and the difference is

$$x(t) - m_1(t) = h_1(t). \tag{1.31}$$

In the second sifting process $h_1(t)$ is treated as the data, then

$$h_1(t) - m_{11}(t) = h_{11}(t). \tag{1.32}$$

We can repeat this sifting procedure k times, until $h_{1k}(t)$ is an IMF, that is

$$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t). \tag{1.33}$$

An IMF must satisfy two conditions:

i. the number of its extrema and zero crossings differ at most by one, and

ii. the mean value of the envelopes defined by the local maxima and minima is zero.

With this definition, the dominant period of oscillation is taken as the average interval between either zeros or extrema for each IMF.

1.2 Historical Background on Monsoon Rainfall

The problem of predicting seasonal monsoon rainfall, and indeed of assessing the degree of predictability in the monsoon, continues to be of great fundamental and practical importance [46]. Detection of significant periodicities in the available rainfall data can be of great value in prediction and has attracted much attention for nearly a century. Reference [47] as well as [48] used mainly classical correlation and power spectral analysis techniques to identify significant periodicities in Indian rainfall distribution. Many recent studies in understanding or prediction of the seasonal rainfall behavior over India are mainly focused on the country as one unit, i.e. they consider all-India statistics. For example [49], using discrete wavelets, has analysed All India Summer Monsoon Rainfall (AISMR) for the period 1871-1998 and its link with the Southern Oscillation (SO). Given the meteorological heterogeneity of India such all-India statistics often give misleading pessimistic about the presence of coherent signals in the Indian monsoon system (see e.g. [24]). Analysis of more homogenous rainfall time series is therefore preferred in this thesis. It is well known that India is meteorologically heterogeneous, so the strongest results are likely to be obtained by analysis of homogeneous regions within India, such as those identified in [50] after an analysis of Indian rainfall over the duration 1871-1990. A detailed analysis in [51] showed a 95% significant periodicity of 2.8 y using the algorithm of

[52] in the homogeneous Indian monsoon (HIM) region identified in [50]. Various other modern methods of spectral analysis (e.g. [7]) to estimate the spectrum of a given time series are available, but have not (to the best of our knowledge) been used for analysis of Indian rainfall data.

An alternate approach would appear to be to use the EMD technique in [45] described above. A comparison of results from a preliminary wavelet analysis and an EMD analysis has shown gratifyingly similar results for major periodicities in the rainfall time series (see Chapter 3 of thesis). However it appears that wavelet techniques may have certain advantages in terms of their mathematical structure, as also pointed out in [53], so a more extensive effort to explore how much further wavelet analysis may be taken is presented in the present work.

The HIM time series has been analysed by various workers using different methods to detect periodicities. Using continuous wavelet maps [21] reported six modes with periods of 3.0, 5.8, 11.6, 20.8, 37.0 and 80.0 y, which revealed for the first time a rather complex temporal structure in the monsoon rainfall. Reference [55] decomposed the same data into five empirical time series of intrinsic mode functions, with associated periods of 2.7, 5.7, 12.0, 24.0 and 60.0 y respectively; these periods are close to those found in [21]. However these periodicities were not tested for the high levels of significance suggested in [54] who proposes that, for a sample of size N, "a good rule of thumb is not to get excited by significance levels less than $1 - 1/N$ ". For HIM rainfall data with $N = 120$, $1 - 1/N = 99.17\%$. The PSD of HIM rainfall is estimated in [51] using the algorithm of [52], and revealed a total of thirteen periodicities, respectively with periods of 2.1, 2.3, 2.5, 2.8, 3.1, 3.5, 4.0, 4.7, 5.7, 7.3, 10.0, 16.0 and 40 y. Of these only the 2.8 y period was statistically significant at 95% confidence against white noise.

In some recent studies [22], [23] and [24] wavelet analysis is used as a new tool to analyse possible connections between monsoon rainfall and solar processes.

These considerations serve to provide motivation for the present work.

1.3 Motivation for the Present Work

This thesis has been inspired by the need to devise more effective search methods for finding periodicities in tropical rainfall, especially the Indian monsoons. The question assumes importance for two reasons. First, it has been shown in a series of studies, beginning with [56], that the inherent predictability of rainfall is appreciably higher in the tropics than at higher latitudes, chiefly because the tropics are 'forced' to a greater extent. One question that arises as a consequence is whether there are hidden periodicities in, say, Indian monsoon rainfall time series. Traditionally the search for periodicities in the monsoon literature has been made by classical power spectral approaches, an extensive analysis is given in [51].

Indian monsoon rainfall is the result of highly nonlinear atmospheric processes that operate and interact over a wide range of scales, and is also a (weakly) non-stationary process in general. Despite this complexity suitably averaged rainfall data reveal patterns and a certain degree of organization both spatially and temporally. Although rainfall is highly variable at very fine scales, e.g. from storm to storm and station to station, larger scales exhibit a hierarchical structure in time and distinct clusters in space. One of the main challenges to hydrologists, meteorologists and climatologists is to measure, model and predict the nature of this variability exhibited by rainfall at different scales. Moreover rainfall data show very low auto-correlation at short lags (e.g. -0.007 at lag 1 in HIM rainfall). For all these reasons it becomes difficult to distinguish a periodic signal from the fluctuations that would be inherent in the relatively small sample sizes available (of order 10^2 y). Now there is an extensive literature that shows how sharp lines in the spectrum of a time series can be identified and tested for significance (e.g. [57], [54]). However, processes like rainfall do not possess sharp lines, but only spectral peaks of varying sharpness, presumably because of nonlinearity, scale interaction and other factors. Hence a major problem lies in testing for statistical significance, in particular in selecting an appropriate reference spectrum against which the peaks found in an estimated spectrum can be assessed. Tests for the significance of the periodic variations against a background of white noise (flat spectrum) are well established [27], [58]. However, it has been noted that the spectra of many geophysical time series frequently show a

continuous decrease of spectral amplitude with increasing frequency, suggesting that they are closer to "red noise" than to white noise [59], [60]. Reference [29] used a first order autoregressive (AR1) process, using the lag-1 autocorrelation coefficient, to model such red noise signatures. Accordingly the AR(1) model is often used to provide a null hypothesis to assess whether or not the power spectral density of a time series is consistent with a likely statistical fluctuation from the reference [29]. Though a number of heuristic or Monte Carlo methods have been devised for reliable identification of the true signal [54], [61], there remains much scope to develop new methodologies tailored to specific applications, and this is what is attempted in the present work.

In the last decade wavelet based techniques have attracted considerable attention in this area because they can be used for non-stationary processes like rainfall. In particular multi-resolution analysis (MRA) can play a significant role in understanding the spatial and temporal structure of rainfall. For this purpose we make use of the multi-resolution properties of discrete wavelets, including their ability to remove interference, for revealing closely spaced spectral peaks. A new procedure, which we first verify on two test signals, is proposed. This is then applied to the time series of homogeneous Indian monsoon (HIM) rainfall annual data. We show that, compared to empirical mode decomposition, discrete wavelet analysis is more effective in identifying closely spaced frequencies if used in combination with classical power spectral analysis of wavelet-based partially reconstructed time series. An effective criterion based on better localization of specific frequency components and accurate estimation of their amplitudes is used to select an appropriate wavelet. It is shown here that the discrete Meyer wavelet has the best frequency properties among the wavelet families considered (Haar, Daubechies, Coiflet and Symlet). The statistical significance of the periodicities thus obtained from the present MRA + PSD technique are assessed using both classical methods and new ones proposed here. Also, a quantitative definition of spectral homogeneity is proposed and subsequently used to identify spectrally homogeneous regions in India. The significance levels of the periodicities are then assessed using also multivariate data analysis.

The discrete wavelet transform (DWT) is extensively used here as a tool of analysis; and our goal is to bridge the gap between theory and practice by

- emphasizing what the DWT actually means in practical terms;
- showing how the DWT can be used to create multi-resolution decomposition of a time series;
- discussing how stochastic models can be used to assess the statistical properties of quantities computed from the DWT;
- presenting substantive examples of wavelet analysis of time series representative of those encountered in geophysical science, in particular of Indian monsoon rainfall.

As periodicities in a time series have been proposed by various workers using continuous wavelets, here we focus on developing wavelet methods in discrete time via standard filtering and matrix transformation ideas.

1.4 Synopsis of the Thesis

The present thesis is arranged into eight Chapters, the contents of which are briefly described below.

Science, engineering, medicine and many other areas deal with signals acquired in the form of time series for the purpose of analysis, for such features as periodicities, spectra, attractor dimensions, extreme events such as singular points etc. The time series data may be modeled as stochastic processes, signals mixed with noise, deterministic chaos etc. The first Chapter presents an introduction to time series analysis. Primarily, classical methods of Fourier spectral analysis are introduced and a brief review of wavelet analysis is given. Wavelet analysis [17] as a relatively new technique is an important addition to standard signal analysis methods as mentioned recently in the work of [33]. The Fourier transform decomposes a given signal into constituent components. Looking at the Fourier spectrum dominant frequencies can be identified; however at any arbitrary frequency temporal localization of any kind cannot be made. Thus, it is not possible to detect the occurrence of any special event, or identify "epochs" during which the signal may have exhibited distinctly

different characteristics, as it might if the signal is non-stationary. This Chapter describes some of the vital differences between the two methods of Fourier and wavelet analysis and all the essential theory and tools used in the thesis. The long history of efforts to seek evidence for periodicities in Indian monsoon rainfall is reviewed and the present status is given.

A detailed description of the rainfall data analysed is presented in Chapter 2. The data selected for study are Indian monsoon rainfall over the six different homogeneous rainfall regions identified in [50], including in particular the homogeneous Indian monsoon (HIM) region. However, over the 120 y period (1871-1990) for which reliable Indian rainfall statistics are available, the analysis is performed first chiefly on the HIM annual time series and then repeated for the other homogeneous regions. We also look at the annual rainfall time series from each of the 14 meteorological sub-divisions constituting the HIM region in detail. It must incidentally be stressed (as noted in [23, 24]) that the often-used all-India summer monsoon rainfall (AISMR) is not the best rainfall index for analyses like the present, as the geographical area it covers is not meteorologically homogeneous, and so physically coherent signals will tend to get diluted due to heterogeneity.

In the recent past HIM rainfall time series has been analysed for periodicities using the power spectral density (PSD) function, empirical mode decomposition (EMD) and continuous wavelet transforms (CWT). Chapter 3 compares and demonstrates the potential of these methods using two specially devised test signals. These signals are intended to simulate the rainfall data, with time taken in years (y) and amplitude in mm, keeping in view the application we shall make in Chapter 4. However the analysis is carried out with normalized time series with zero mean and unit standard deviation for both the test signals and rainfall data, as our objective is to study the temporal structure of the time series and not the magnitude of the rainfall.

The most prominent limitation of Fourier analysis on discrete-time data is that of frequency resolution, i.e., the inability to distinguish between two spectral peaks spaced closer than 2/N, N being the total number of samples in the time series [10]. A wavelet based hybrid technique, where the advantages of wavelet methods in handling nonlinear non-stationary time series are combined with those of spectral

analysis, is proposed here to resolve such closely spaced frequencies. The discrete Meyer, Daubechies and other wavelets are used, and it is found that the discrete Meyer wavelet provides the best results in separating closely-spaced frequencies without yielding spurious peaks. A discussion of these wavelets and their performance in the present application is given. The Chapter concludes with the result that the direct PSD on the test signals merges closely spaced frequencies whereas the PSD on partially reconstructed time series obtained from wavelet coefficients are able to distinguish between these frequencies. This is because during MRA decomposition detail coefficients are obtained by setting low frequency components to zero. This helps in the elimination of sinusoidal interference in the highest frequency band. The detail time series is then subjected to classical Fourier spectral analysis, which is successful in separating closely spaced frequencies provided the associated spectral peaks are sufficiently narrow.

The proposed method is then applied on the HIM rainfall time series annual data in Chapter 4. The use of MRA as a tool for analyzing non-stationarity in the time series is briefly examined. This is required because our method demands combined use of wavelet and Fourier spectral analysis, and the latter is valid only for stationary time series. The method described here helps to isolate the stationary component of the time series, i.e. to de-trend it using long-scale wavelets. Using this approach it is found that the rainfall data is non-stationary only over long periods (of order 60 y). It is further shown that, with the technique proposed, there are in fact some closely spaced periodicities in rainfall that were missed by earlier techniques, which yield only some weighted average values of nearby periods.

Reliable identification of the oscillatory components contained in a short, noisy time series requires effective criteria for assessing statistical significance. Furthermore searching for evidence of predictability in an observational time series provides a starting point for many geophysical investigations since predictability may be very limited in any nonlinear system, whether it is stochastic or deterministic; any trend may vary over time, as may the phase and amplitude of a physical oscillation. Nonetheless, even limited predictability may be highly informative and useful, making analysis techniques, which allow the detection of intermittent trends or oscillations, valuable. Towards this goal, discriminating between signal and noise

is a crucial step.

In Chapter 5, the spectrum of HIM rainfall time series is estimated using the Welch technique [7] and the significance of the peaks detected is tested against classical reference spectrum introduced in [29] and a more recent technique proposed in [62]. If a particular feature of the data spectrum lies well above this reference spectrum, it is often considered to be statistically "significant". By these methods two significant periods, respectively at 2.3 y (at 99%) and 2.8 y (95%) are found to be significant. However other significant periods in HIM rainfall are expected, including around 11 y, as Indian rainfall has a strong solar connection [23]. We present here a new procedure for checking the significance levels of partially reconstructed time series obtained in Chapter 4 at different frequency bands. The auto-correlation coefficient of HIM rainfall series at lag 1 is -0.007, i.e. close to zero. This implies that in the classical technique of devising a reference spectrum based on autocorrelation, the periodicities in HIM rainfall are tested against a nearly flat spectrum (i.e. white noise). However we present strong evidence for the presence of a spectral dip in HIM rainfall around a frequency of 0.25 y^{-1}, and construct a testing procedure that takes this into account. Thus, using MRA we have shown that at the first level of decomposition for high frequencies the r reference spectrum is blue, whereas for low frequencies it is red. Hence the statistical significance of these periodicities must be tested against appropriately colored noise, matching the observed spectrum of rainfall. We conclude the Chapter with the result that the estimated PSD of HIM rainfall time series against the classical reference spectrum shows only one periodicity of 2.3 y above 99% confidence level, whereas using the present method of MRA+PSD against appropriately colored reference spectra, ten periodicities are significant above the 99.9% confidence level, with respective periods of 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3 and 60.6 y.

One persistent question that arises in the quest for identifying periodicities is the heterogeneity of the Indian monsoon. In spite of many studies of the problem, and the considerable evidence of this heterogeneity, many analyses still continue to rely on all-India indices. Even the homogeneous regions identified by different workers may contain heterogeneities not considered in the criteria laid down for determining the degree of homogeneity.

The present study starts with the observation that an annual time series for rainfall, such as those presented in [50] for the various homogenous regions within India identified by them, is not just a single sequence of numbers giving one value for each year. The reason is that such regional time series are obtained by taking area-weighted averages over several meteorological sub-divisions, each of which is itself an average over a large number of individual rain-gauges. Now the sub-divisional time series possess substantial cross-correlations with each other, and hence are not independent. Thus, treating them as independent realizations from one homogeneous population is not justifiable, but considering HIM rainfall as just one isolated realization in itself would be too conservative. The main purpose of this chapter is to develop a method based on multivariate analysis that provides a rational procedure for taking these factors into account.

This line of thinking leads us to propose a quantitative definition of spectral homogeneity. This is based on evaluating the mean square deviations of each normalized sub-divisional rainfall from normalized HIM rainfall against the null hypothesis that the rainfall deviations do not differ from those of white noise. When the hypothesis is rejected at a confidence level of 99.5% by the F-test, we consider the deviations to be small enough to permit identification of spectral homogeneity. In particular, as an illustrative demonstration of the proposed method, certain subdivisions in the HIM region showing significantly low spectral variability are grouped into a "Spectrally Homogeneous Indian Monsoon"(SHIM) sub-region. This is proposed as a good candidate for further analysis, as likely to provide the strongest evidence for any periodicity that may be present in any part of India. The main conclusion of this chapter is that the HIM rainfall time series over the time period 1871-1990 shows one 99% significant period of 2.3 y and 95% significant period of 2.8 y, against the white noise reference spectrum. However, by introducing the spectrally homogeneous sub-region SHIM, we have shown that there is a 99.5% significant period of 2.3 y, along with four other significant peaks above 95% confidence at 2.1, 7.5, 13.3 and 60.6 y and a 2.8 y period at $> 94.8\%$..

A new test based on multivariate analysis to find the significance levels of common periodicities among sub-divisions constituting SHIM rainfall is then proposed. The motivation here is one of the key messages of multivariate analysis, namely

that correlated variables must be analysed jointly. Confidence ellipses, also called prediction interval ellipses [82], are used as an exploratory method to investigate the correlation between the four sub-divisions constituting the newly defined SHIM sub-region. The proposed method indicates the presence of seven significant periodicities in SHIM rainfall, five at confidence levels exceeding 95% (respectively 2.1 y at 97.8%, 2.3 y at 95.7%, 2.8 at 98.4%, 7.5 at 99.5%, 24.0 at 99.4%), and 10.9 at 90.5% and 60.6 at 93.2% confidence level. The main conclusion of this chapter is that a common period may be present in all the sub-divisions constituting a region but the significance of their commonalty cannot be established using the classical methods. However it is found that the proposed method of confidence ellipse is a more informative and powerful tool for testing confidence levels of common periodicities in correlated time series. Also we found that the classical test of checking confidence level of a spectral peak is inadequate when the amplitude of longer periods is low. The proposed confidence ellipse technique however shows high confidence levels even for relatively low amplitudes when the cross-correlation is high. Hence the present method is less amplitude-sensitive than other available methods.

In Chapter 7, the analysis performed so far on the HIM rainfall time series is repeated for other homogeneous regions. Earlier classifications of India into homogeneous regions based on rainfall do not take into account spectral characteristics. Using spectral homogeneity as a criterion, we here classify the country into ten spectrally homogenous regions (SHRs). These new regions cut across those defined earlier by other workers. For example, SHR7 has six sub-divisions from the HIM region and one from the PENSI region. SHR7 may be of particular interest as all the constituent sub-divisions show a spectral dip around the frequency of 0.25^{-1}, and most of the spectral peaks in the different sub-divisions nearly coincide with each other. The significance levels of periodicities detected in these regions are assessed. In SHR7 there is reported to be one periodicity above 95% by the direct PSD whereas the present MRA+PSD technique reveals seven periodicities above 99% confidence level, with respective periods of 3.3, 5.7, 8.6, 13.3, 10.9, 24.0 and 30.3 y.

Chapter eight outlines a possible explanation for the results that we obtain and discusses new vistas for further and future investigations.

Chapter 2

The Data Analysed

2.1 Details of Rainfall Data

The present analysis involves the time series of Indian monsoon rainfall annual data for the period 1871-1990 spanning 120 years. The rainfall indices have been obtained from [64] and are also available at the Indian Institute of Tropical Meteorology web site: www.tropmet.res.in. The details about how these data were procured and pre-processed are as follows. A network of 306 rain-gauge stations across India, almost uniformly distributed, each providing one representative station per district having a reliable record since 1871, formed the basis for the creation of the data set [66], [65]. Figure 2.1 shows the rain gauge network considered. In view of the limitations of areal representativeness of rain gauges in hilly areas, four meteorological sub-divisions parallel to the Himalayan mountain range, numbered 2, 12, 15 and 16 in figure 2.1, were excluded from the study. Also excluded from Parthasarathy's network of rain gauges are two sub-divisions representing island territories far from the mainland, Andaman & Nicobar (sub-division No.1) and Lakshwadeep (sub-division No.35). The area covered by the data set measures 2.88 x 10^6 km^2, amounting to about 88% of the total area (3.292 x 10^6 km^2) of the country.

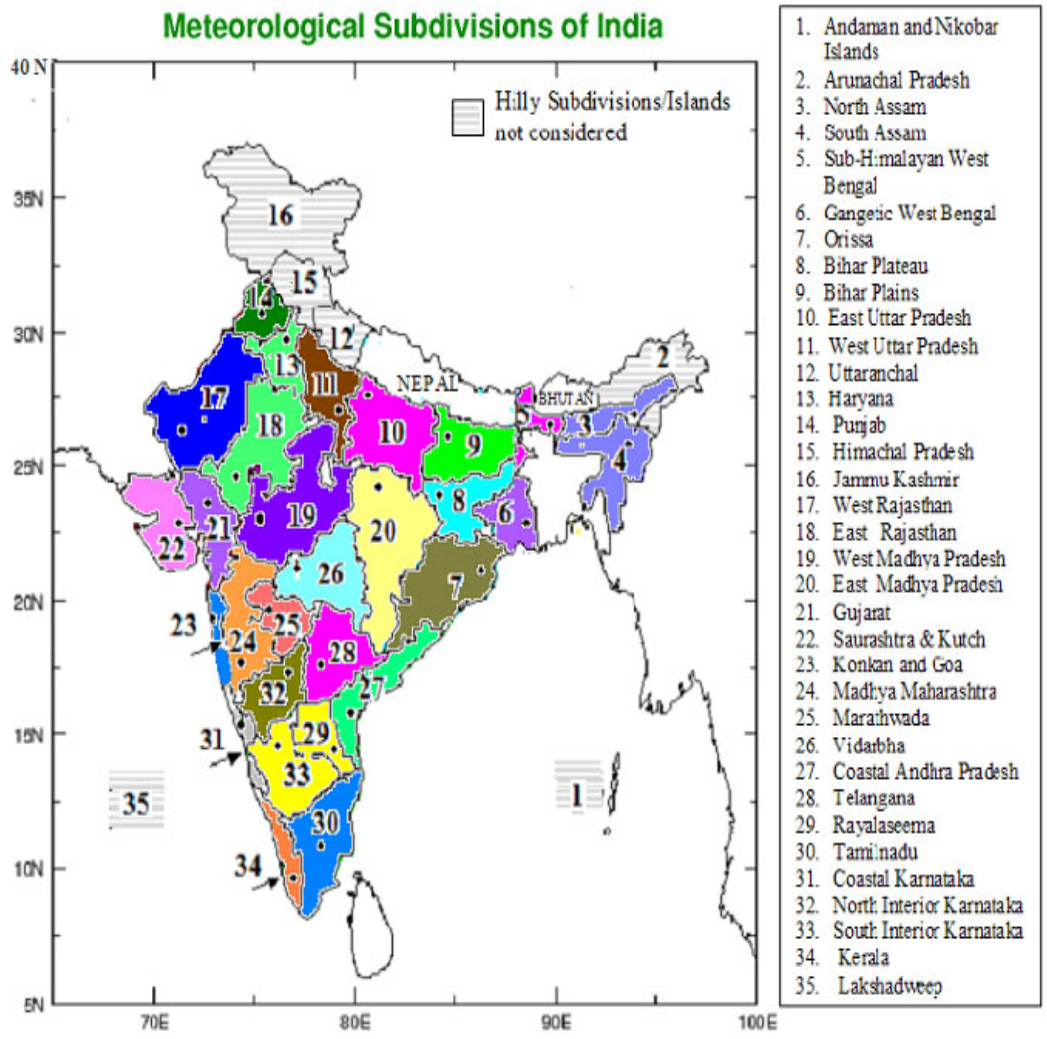

Figure 2.1: Map of India showing the meteorological sub-divisions defined by the India Meteorological Department (IMD). Sub-divisions 1, 2, 12, 15, 16 and 35 are not considered in the rainfall data set analysed here.

2.1.1 Homogeneous regions in India

Apart from the 29 meteorological sub-divisional monsoon rainfall series [50] have also prepared rainfall series for 5 special regions over the country considered homogeneous on the basis of similarity in rainfall characteristics and association of subdivisional monsoon rainfall with regional/global circulation parameters during 1871-1990. These five homogeneous regions (shown in figure 2.2) are: (i) North West India (NWI); (ii) West Central India (WCI); (iii) Central Northeast India (CNEI); (iv) North East India (NEI) and (v) Peninsular India (PENSI). Apart from these five homogeneous regions, one more broad region called the homogeneous Indian monsoon (HIM) region has also been identified and the monsoon rainfall time series for this region was also prepared. The map of HIM region is given in figure 2.3. It is to be noted that NWI together with WCI form the HIM region. The particulars of the six homogeneous regions in terms of the sub-divisions considered in the preparation of rainfall time series, as well as some other pertinent statistical details based on 1871-1990 data, are given in table 2.1.

Table 2.1: Details of homogeneous monsoon regions of India.

Region	Sub-divisions considered	Area, India %	Mean rainfall (mm)	% of Annual
Northwest India	13,14,17,18,21,22 (6)	22.0	545.4	89.9
West Central India	19,20,23-26,28,32 (8)	33.4	1081.1	86.3
Central Northeast India	7-11 (5)	19.9	1203.1	83.3
Northeast India	3-6 (4)	9.3	2066.2	68.7
Penisular India	27, 29-34 (6)	15.4	1156.2	57.0
Homogeneous India Monsoon	13,14,17-26,28,32 (14)	55.5	868.3	87.0

It may be noted that although rainfall data are available for the period 1990 to 2005 also, we have limited our analysis to the period 1871-1990.

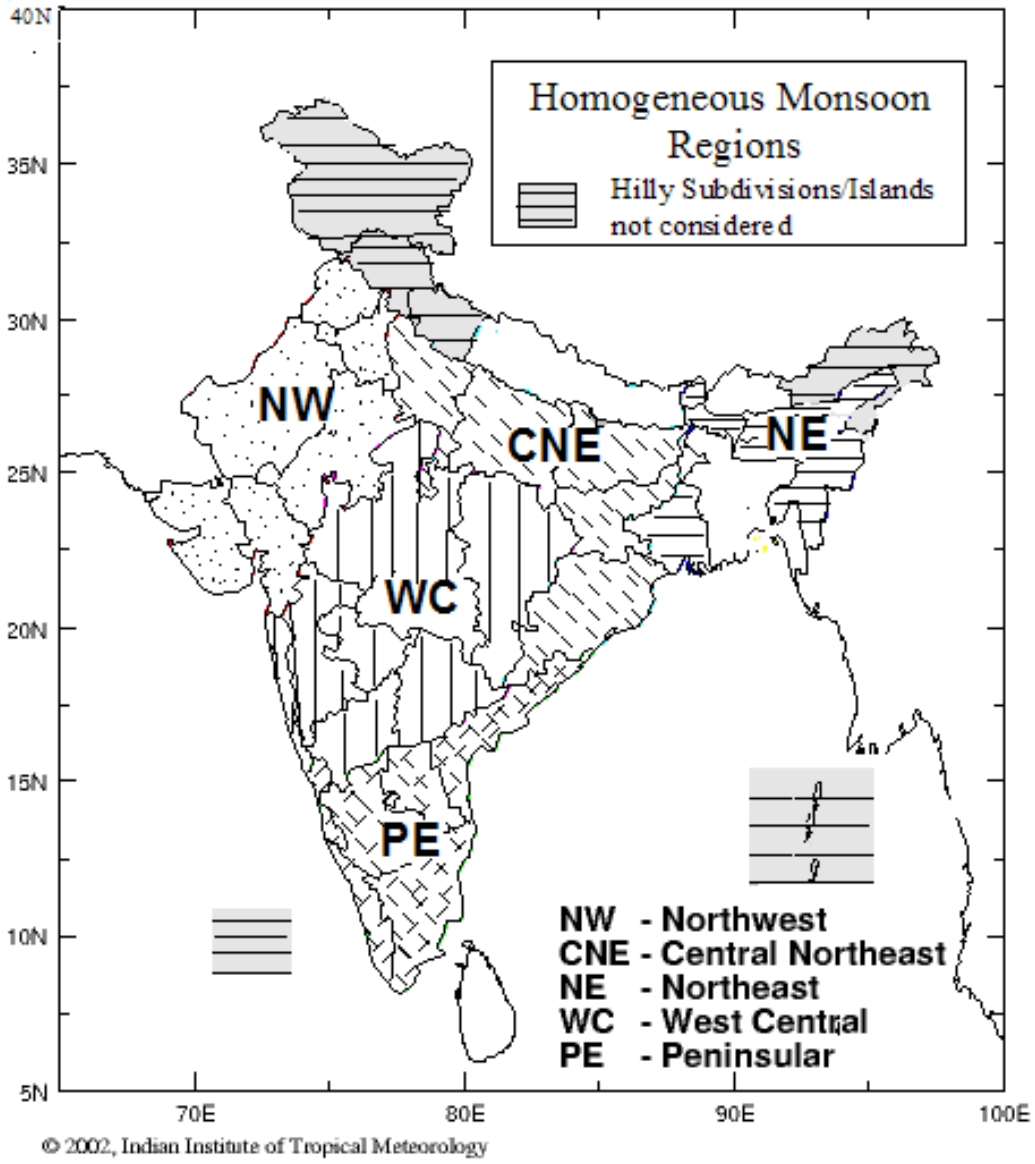

Figure 2.2: Homogeneous monsoon regions of India.

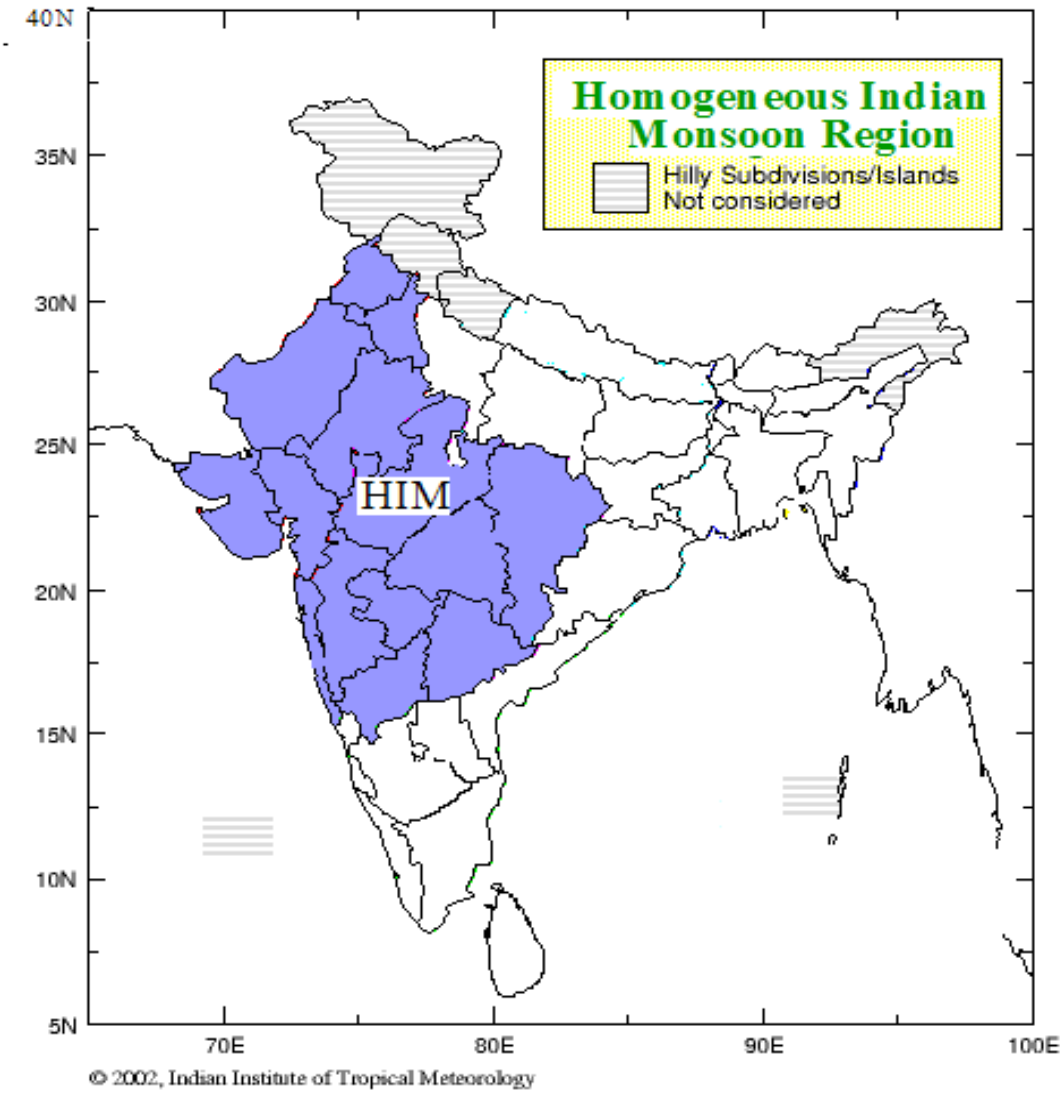

Figure 2.3: Homogeneous monsoon regions of India.

2.1.2 Homogeneous Indian monsoon rainfall

The HIM rainfall annual data from 1871-1990 is shown in figure 2.4.

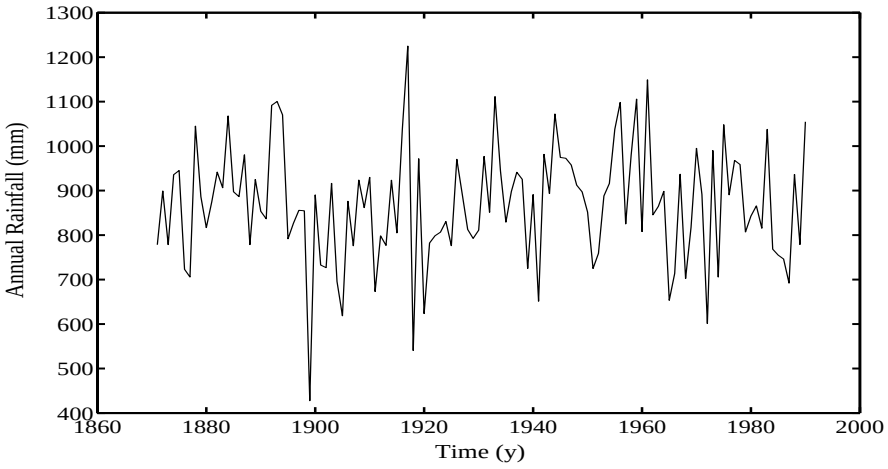

Figure 2.4: HIM rainfall time series, 1871-1990.

Out of the six homogeneous regions mentioned earlier, the most extensive is the homogeneous Indian monsoon (HIM) region. This region covers a geographical area (northwest and central part of India, amounting to about 55% of the total land area of the country) which, based on various criteria, may be considered to be meteorologically homogeneous, and hence may be expected to yield a physically coherent rainfall signal. In particular the HIM region may be considered to be by far the most representative index of the component of Indian rainfall dominated by the dynamics of the south west monsoon. HIM rainfall is an area-weighted average over an ensemble of 14 rainfall time series, each of which represents one of the 14 sub-divisions constituting the HIM region. These 14 sub-divisions are Haryana, Punjab, West Rajasthan, East Rajasthan, West M.P., East M.P., Gujarat, Saurashtra, Konkan, Madhya Maharashtra, Marathwada, Vidarbha, Telangana, and North Interior Karnataka. In this thesis the main data analysed is the time series of HIM rainfall and the 14 constituent sub-divisions over the period 1871-1990 provided in [64]. The rainfall time series of 14 sub-divisions constituting HIM are shown in figures 2.5 and 2.6. Table 2.2 shows the details of these sub-divisions, their areas and

the number of rain gauges considered in calculations of the sub-divisional average.

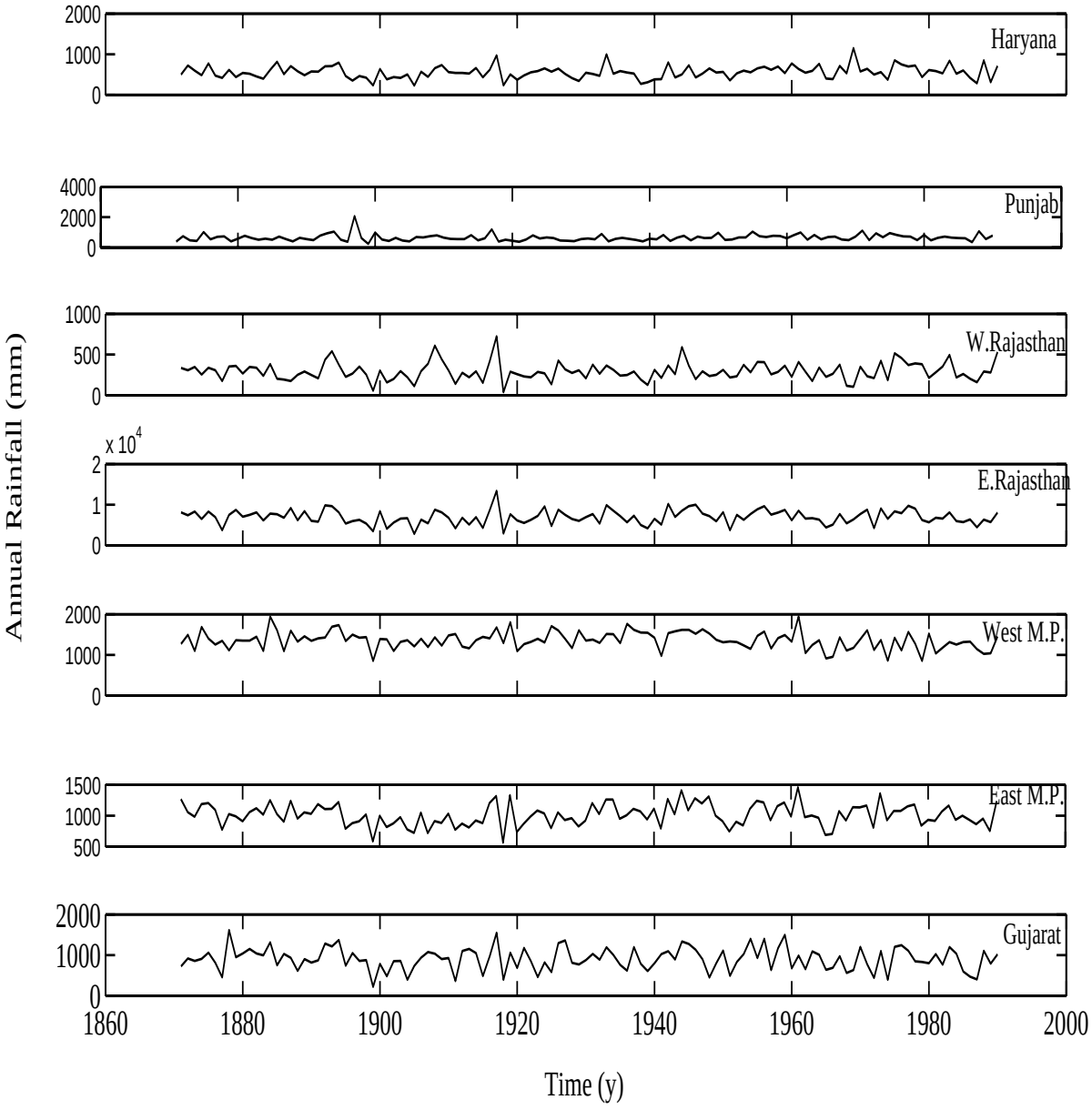

Figure 2.5: The rainfall time series of 7 sub-divisions constituting HIM region.

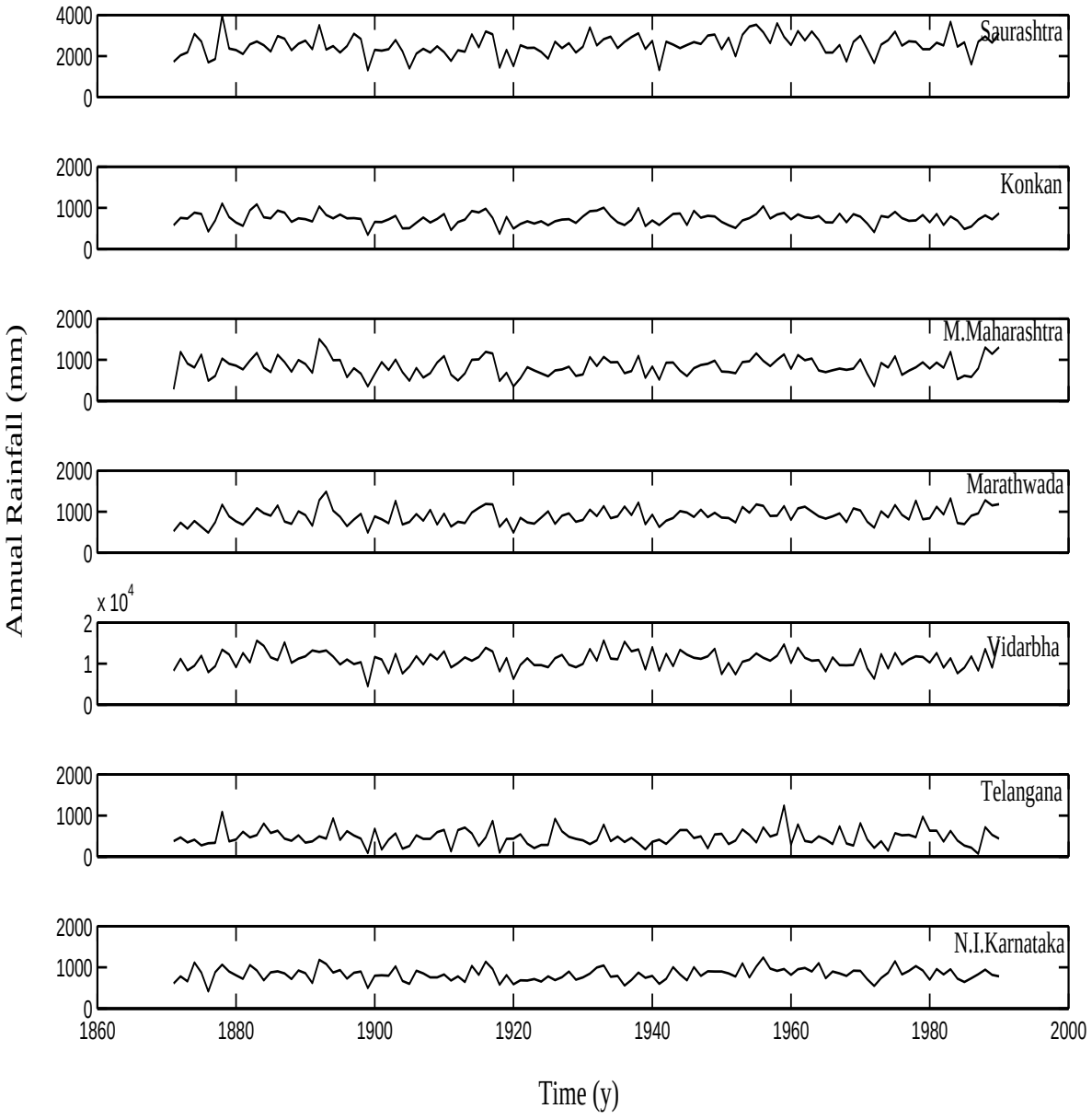

Figure 2.6: The rainfall time series of 7 sub-divisions constituting HIM region.

Table 2.2: Details of 14 sub-divisions constituting HIM region in terms of the area they covered.

S.no	Sub-division	Area (km^2)	No. of rain gauges	Average area per rain gauge (km^2)
1.	Telangana	114,726	9	113 x 113
2.	West M.P.	232,315	29	90 x 90
3.	East Rajasthan	147,128	17	93 x 93
4.	Vidarbha	97,537	8	110 x 110
5.	Konkan	34,095	5	83 x 83
6.	East M.P.	224,296	14	127 x 127
7.	West Rajasthan	195,086	9	14 x 147
8.	Marathwada	64,525	5	114 x 114
9.	Haryana	45,698	12	62 x 62
10.	Saurashtra	109,950	7	125 x 125
11.	Gujarat	86,034	11	88 x 88
12.	Madhya Maharashtra	115,309	9	113 x 113
13.	North Interior Karnataka	79,895	6	115 x 115
14.	Punjab	50,376	10	71 x 71

2.1.3 Correlations among sub-divisional rainfall data

The present study is based on the observation that an annual time series for rainfall, such as those presented in [50] for various homogenous regions within India, is not merely a sequence of independent numbers comprising one value for each year. The reason is that such time series are obtained by taking averages over meteorological sub-divisions (say n in number) of order 10, and rainfall time series of order 10^2 rain-gauges. Now the sub-divisional time series are not independent, but instead possess substantial cross-correlations with each other. Thus, treating them as independent realizations from one homogeneous population is not justifiable, but considering HIM rainfall as just one realization would be too conservative. The correlation coefficients between HIM rainfall and the 14 sub-divisional rainfall series for the period 1871-1990 are presented in table 2.3. The correlation coefficients are positive in all sub-divisions and significant at 1% level with the confidence intervals shown also in table 2.3. Further, the inter-correlations among these 14 sub-division are calculated and found to be positive for all. The inter-correlations of 29 sub-divisional monsoon rainfall series are discussed in [65, 67]. These studies show that neighboring sub-divisions are in general highly positively correlated. However the two sub-divisions of Assam state are negatively correlated with the rest of the sub-divisions in NEI.

Table 2.3: Correlation coefficients between rainfall in HIM region and that in its constituent 14 sub-divisions.

S.no	Subdivision	Corr.coefficient	Confidence interval
1.	Punjab	0.47	[0.32, 0.60]
2.	North Interior Karnataka	0.59	[0.46, 70.0]
3.	East M.P.	0.59	[0.46, 0.70]
4.	Haryana	0.60	[0.48, 0.71]
5.	Saurashtra	0.62	[0.50, 0.72]
6.	Telangana	0.67	[0.55, 0.76]
7.	Marathwada	0.68	[0.57, 0.76]
8.	Madhya Mahrashtra	0.70	[0.60, 0.78]
9.	Konkan	0.70	[0.60, 0.78]
10.	West Rajasthan	0.73	[0.63, 0.80]
11.	Gujarat	0.75	[0.66, 0.82]
12.	Vidarbha	0.76	[0.67, 0.83]
13.	East Rajasthan	0.79	[0.63, 0.80
14.	West M.P.	0.81	[0.74, 0.87]

Chapter 3

Analysis of Test Signals

[1]

3.1 Introduction

As mentioned in Section 1.1, closely spaced frequencies have been separated using different techniques in the recent past and pose an important problem in time series analysis. The main aim of this chapter is to devise a new wavelet based method for identifying closely spaced frequencies, and validate the method by test cases that mimic the time series of our interest, namely Indian rainfall. There are of course other methods [14] using maximum likelihood and regression techniques that separate closely spaced frequencies, but these assume that the time series is stationary, whereas the method proposed here, analysing each level, does not need to assume stationarity of the whole signal but only of the decomposition on shorter scales that may be subjected to spectral analysis. (This procedure is satisfactory for monsoon rainfall whose non-stationarity is slight; we shall return to this question in Chapter 4.)

A mathematical description of three powerful methods, namely power spectral density (PSD) function, multi-resolution analysis (MRA) and empirical mode decomposition (EMD), has already been given in Chapter 1. In the sections to follow,

[1]Some of the material in this chapter has appeared in Azad *et al.* (2007a).

the performance of each of these methods is compared on specially devised test signals. A hybrid technique using MRA, which takes advantage of both spectral and wavelet analysis for separating closely spaced frequencies, is then proposed. This method is shown to perform better than classical spectral analysis or EMD in resolving closely spaced frequencies, especially if a suitable wavelet is chosen.

3.2 The Test Signals

To make a critical assessment of the methods mentioned above, two test signals, each a sum of sinusoidal functions with prescribed frequencies f_i and amplitudes a_i, are constructed using the equation

$$x(t) = \sum_i a_i \sin 2\pi f_i t, \qquad t = 1, ..., 256. \tag{3.1}$$

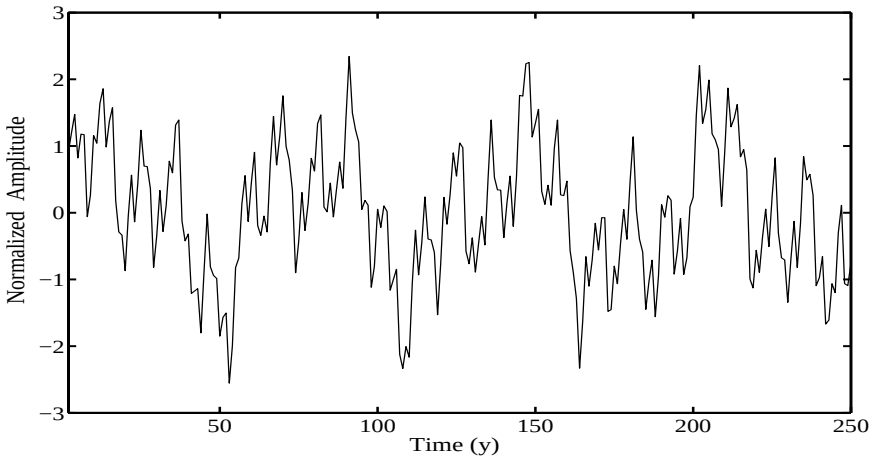

Figure 3.1: Time series of test signal 1.

These signals are intended to simulate the rainfall data, with time taken in years (y) and amplitude in mm, keeping in view the application we shall make in Chapter 4. However the analysis is carried out with normalized time series with zero mean and unit standard deviation for both the test signals and rainfall data, as our objective is to study the temporal *structure* of the time series and not the

Level	Frequency (y^{-1})	Normalized Amplitude
1.	0.33	0.10
2.	0.20	0.07
3.	0.09	0.18
4.	0.035	0.15
5.	0.0156	0.20

Table 3.1: Test signal 1 with given frequencies and normalized amplitudes.

magnitude of the rainfall. The sinusoids constituting test signals 1 and 2 are shown in figures 3.1, 3.2, and their details are listed in tables 3.1, 3.2.

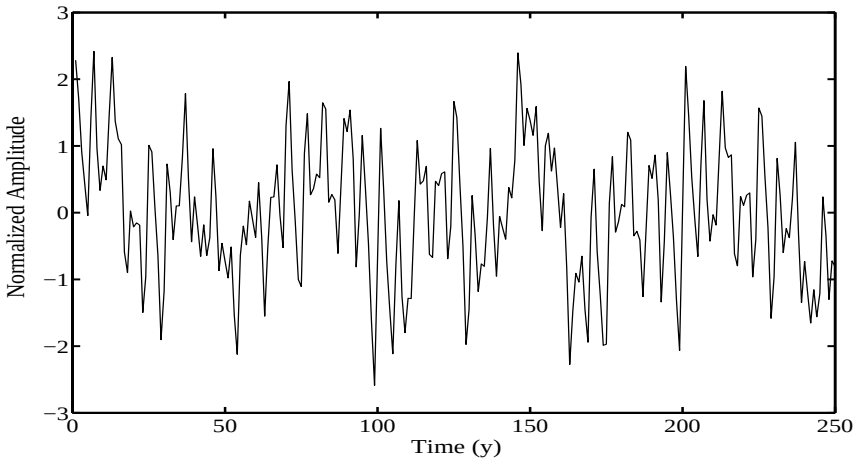

Figure 3.2: Time series of test signal 2.

3.2.1 EMD on the test signals

The EMD analysis was carried out using the method of [68], and the code available at http://perso.ens-lyon.fr/patrick.flandrin/emd.html. First EMD is performed on test signal 1. Figure 3.3 shows 8 intrinsic mode functions (IMF) obtained as a result of the sifting process described in Section 1.1. The corresponding periods are

Level	Frequency (y^{-1})	Normalized Amplitude
1.	0.33	0.18
2.	0.20	0.19
	0.17	0.17
	0.16	0.19
3.	0.09	0.20
4.	0.035	0.18
	0.028	0.18
5.	0.0156	0.19

Table 3.2: Test signal 2 with given frequencies and normalized amplitudes.

calculated by the method of zero crossings. Table 3.3 shows that all the periods are captured quite accurately by this method.

However, [45] explain that in separating two very close frequencies, spline fitting takes thousands of iterations and the extracted IMF's are severely contaminated. As a result, the instantaneous frequencies of the components overlap to form a single peak located near the mean value. It is further reported in [45] that the cubic spline fitting has both overshoot and undershoot problems. Because of this some asymmetric wave forms can still exist no matter how many times the data are sifted.

In order to analyse such difficulties test signal 2 (table 3.2) is constructed and investigated. This signal has three closely spaced high frequencies at 0.2, 0.17 and 0.16 y^{-1} and two low frequencies at 0.035 and 0.028 y^{-1}. Results are shown in table 3.4. IMF 2 captures a frequency 0.175 y^{-1}, which is some weighted average of the three frequencies 0.2, 0.17 and 0.16 y^{-1} present in test signal 2. At IMF 4 the frequency captured is 0.033 y^{-1}, which again is some weighted average of the frequencies 0.035 and 0.028 y^{-1}. It seems plausible that the way frequencies of each decomposition are calculated by counting zero crossings gives a weighted average period which is unable to separate any closely spaced frequencies that the original time series may contain. This example shows that at any time in the sifting

process a function may exist (for example a waveform comprising two closely spaced frequencies) that may satisfy the IMF criteria but still cannot be taken as a single frequency component. Hence some frequency information can be missed out as the method of counting zero crossings gives an average over closely spaced frequencies; this is the problem of mode mixing [45].

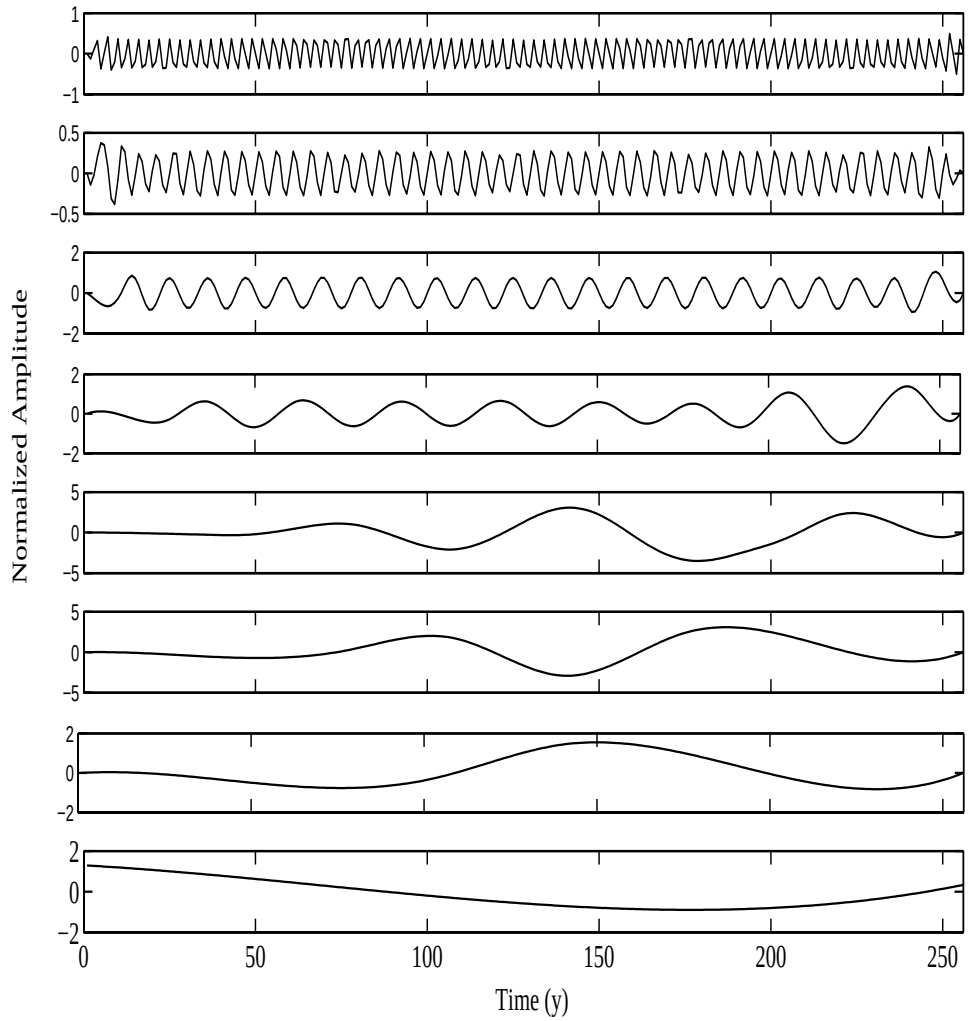

Figure 3.3: EMD on test signal 1, IMF1 (top) to IMF8 (bottom).

Level	Frequency (y^{-1})	Corresponding freq. in test signal 1
IMF1	0.32	0.33
IMF2	0.19	0.20
IMF3	0.089	0.09
IMF4	0.035	0.035
IMF5	0.0156	0.0156

Table 3.3: EMD with count of zero crossings on test signal 1.

Level	Frequency (y^{-1})	Corresponding freq. in test signal 2
IMF1	0.30	0.33
IMF2	0.175	0.20
		0.17
		0.16
IMF3	0.085	0.09
IMF4	0.033	0.035
		0.028
IMF5	0.0156	0.0156

Table 3.4: EMD with count of zero crossings on test signal 2.

3.2.2 MRA on the test signals

The procedure adopted here goes with the description of MRA given in Section 1.1. The MRA technique defined there can be implemented using standard routines in the wavelet tool box of *Matlab*. There are many types of mother wavelets that can be used in practice; a total of six wavelets have been considered here: Haar; discrete Meyer (dmey); Daubechies of order 2 (db2); Daubechies of order 10 (db10); Coiflet of order 2 and Symlet of order 2. These wavelets are shown in figure 3.4. A discussion of these wavelets can be found in [36]. In this subsection the performance of the dmey wavelet is evaluated, and its superiority is explained in terms of better localization of each specific frequency component and an accurate measurement of its amplitude at each level of MRA decomposition.

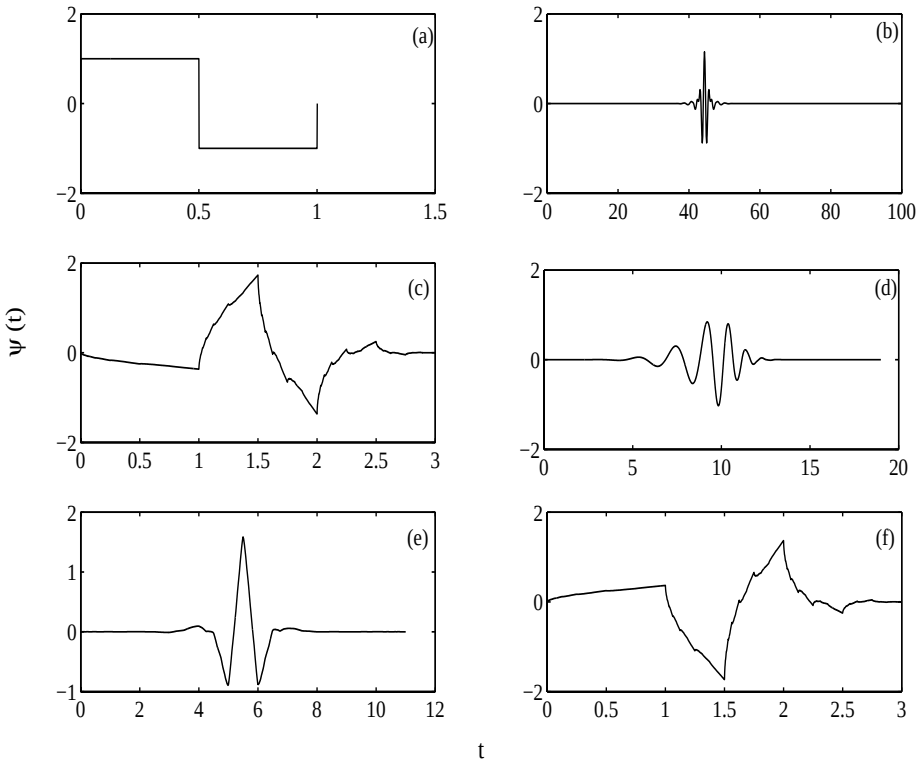

Figure 3.4: Examples of mother wavelets: (a) Haar; (b) Discrete Meyer wavelet; (c) Daubechies of order 2; (d) Daubechies of order 10; (e) Coiflet 2; (f) Symlet 2 .

In the case of discrete wavelets a time series with sample size N can be decomposed at J levels or the scales $2^j, j = 1, ..., J = \log_2 N$, if the sampling interval is taken as one. The procedure goes with the description of MRA given in Chapter 1, where the orthogonal projection on lower subspaces is achieved by recursive sums and the error terms are obtained from differences. The first step produces two sets of coefficients: approximation coefficients X_{N-1} and detail coefficients e_{N-1}. At the end of the first step, the number of samples reduces to $N/2$, so to keep their number the same as the original size the detail and approximate coefficients are reconstructed by padding them with $N/2$ zeros [30]. Hence at the first level, instead of X_{N-1} the reconstructed approximate coefficients X^+_{N-1} and the detail coefficients e^+_{N-1} are considered.

The test signal 1 ($N = 256$ samples) is decomposed at $2^j, j = 1, ..., 8$ scales using the dmey wavelet. The first partially reconstructed approximate series is obtained by losing the highest frequency, the time series of which makes the 1st detail series as shown in figure 3.5, i.e.

$$X_N - e^+_{N-1} = X^+_{N-1}. \tag{3.2}$$

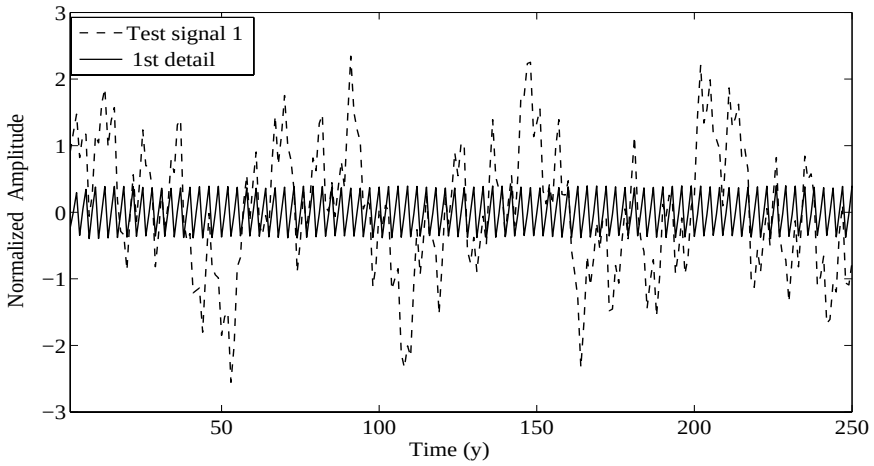

Figure 3.5: Test signal 1 plotted along with its first reconstructed detail, using dmey wavelet.

Note that e^+_{N-1} and X^+_{N-1} would fully reconstruct the original signal by the projection theorem [44]. Also it is observed that the amplitude of the first reconstructed

detail is equal to the amplitude given to the frequency 0.33 y^{-1} in the test signal 1 (table 3.1).

In similar manner X_{N-1} is decomposed into X^+_{N-2} and e^+_{N-2} and so on. Thus at the end of J levels of decomposition X_N can be written as

$$X_N = e^+_{N-1} + e^+_{N-2} + e^+_{N-3} + ...e^+_1 + X^+_1. \tag{3.3}$$

Figure 3.6 shows the reconstructed approximations and details obtained at 8 levels of decomposition using the dmey wavelet.

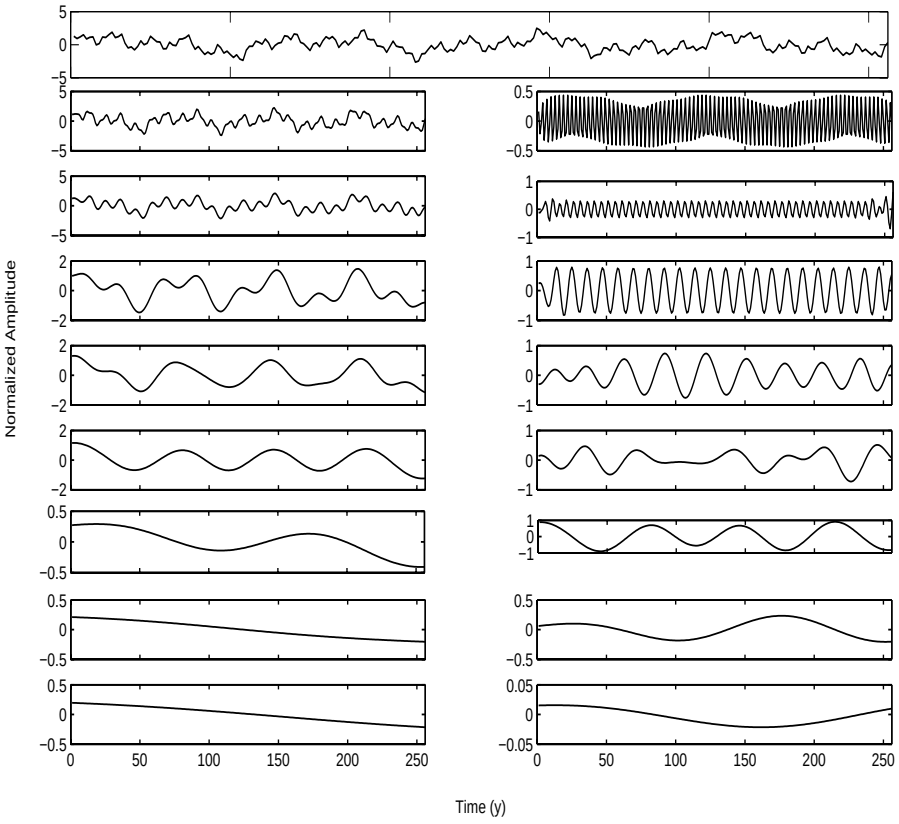

Figure 3.6: Test signal 1 is plotted with its partially reconstructed approximations (left) and details (right) from level 1 (top) to level 8 (bottom).

However, the first four partially reconstructed details and the fourth partially reconstructed approximation are considered for the present analysis (see figure 3.7),

as the test signal 1 can be fully reconstructed by a decomposition upto four levels with an error of only 10^{-4}. Also it is shown in the next subsection that upto the fourth level of decomposition all the frequencies present in the test signal 1 are captured (see table 3.5).

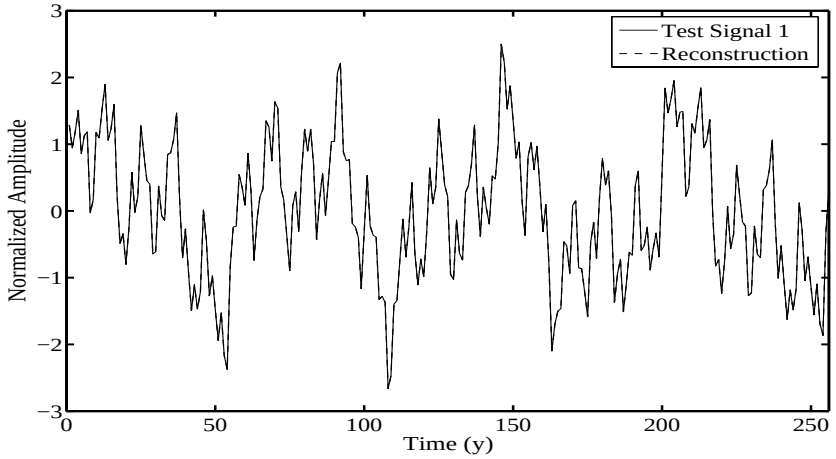

Figure 3.7: Test signal plotted along with its four details and fourth partially reconstructed approximation, using dmey wavelet.

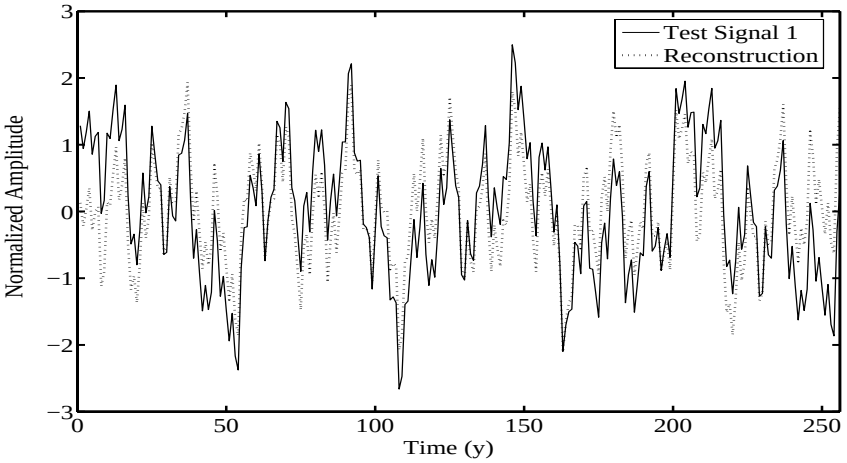

Figure 3.8: Test signal 1 plotted along with its five partially reconstructed details, using dmey wavelet.

If we use five details instead, figure 3.8 shows that the reconstruction has an error of as much as 10%.

The same procedure is repeated with the db2 wavelet, and results are shown in figure 3.9 for the first level of detail. It is clear from the figure that the partially reconstructed detail time series at this level shows more than one frequency, whereas the dmey wavelet could capture only one (figure 3.5).

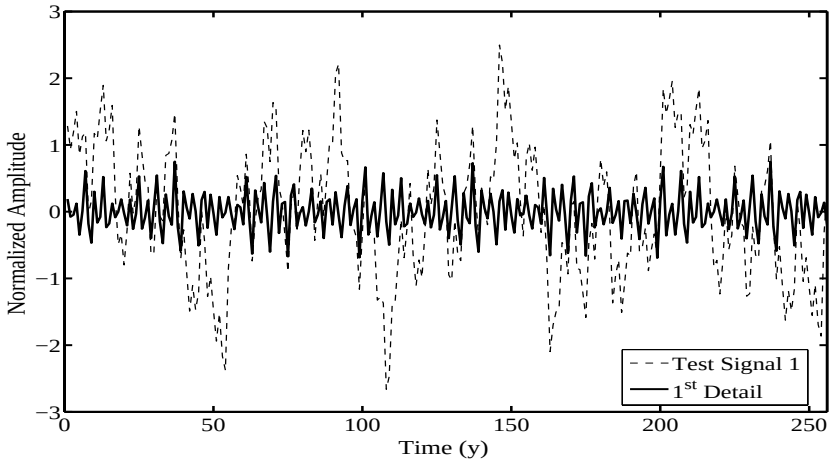

Figure 3.9: Test signal 1 plotted along with first reconstructed detail, using db2 wavelet.

This difference in performance between the two wavelets is analysed in the next subsection.

3.2.3 Spectral analysis of wavelet decompositions

The main aim here is to show that classical spectral analysis applied to MRA decompositions captures closely spaced frequencies better than a 'direct' spectral analysis of the original signal (as described in classical texts such as [7].

Once the decomposition is achieved, the frequencies are defined over a particular range of scales at each level in MRA. This is because (in terms of the time interval between successive points in the discrete time series, say one year), the first level periods in MRA decomposition are in the scale range 2-4 y, the second level in the range 4-8 y, and so on, with dyadic scaling. But in order to get specific values

for the frequencies we revert to the spectral approach. Thus, the power spectral density (PSD) functions of the reconstructed detail and approximate coefficients of test signal 1 are first estimated, treating them as time series in their own right. The Welch modified periodogram method [28] with Hanning window (described in Section 1.1) is used for this purpose. The rationale behind the present approach will be explained towards the end of this section.

Results obtained from MRA on test signal 1 with the dmey wavelet are shown in table 3.5. Here level 1 shows the results of frequencies obtained from partially reconstructed detail e^+_{N-1} and reconstructed approximation X^+_{N-1}. It is clear from the table that at the first level of detail, the highest frequency (0.33 y^{-1}) is separated out from the rest of the frequencies, which are together captured at the first level of approximation. This highest frequency is the detail lost in making the first approximation of the test signal 1. Similarly the next highest frequency (0.2 y^{-1}) is separated out at the second level of decomposition. The results for the 8 levels are shown in table 3.5.

It is clear from the table that it is sufficient to consider the decomposition till the fourth level as all the relevant frequencies are individually captured; the test signal 1 can of course be fully reconstructed with this decomposition by virtue of Eq. 1.30.

On the other hand the db2 wavelet yields false frequencies at the first level itself, i.e. along with the frequency 0.33 y^{-1} it suggests other frequencies at 0.41, 0.3 and 0.16 y^{-1} which are not present in test signal 1 (see table 3.6). Also, with the db2 wavelet there is repetition of frequencies at various levels, whereas the dmey wavelet is able to separate the highest frequency at each level without any mixing of frequencies. This confirms the characteristic of the dmey wavelet that it is localized in the frequency domain, as explained in [36].

The same procedure is carried out with four other wavelets: Haar, db10, coiflet2 and Symlet8. In all these cases results obtained are similar to those with db2. Hence it is concluded that the dmey wavelet works satisfactorily on test signal 1, and this wavelet will therefore be adopted for further analysis.

Next test signal 2 (table 3.2) is considered. Now it is not possible to resolve frequencies spaced closer than $1/N$ in classical spectral analysis, and frequencies

Level	Details	Approximations
	Frequency (y^{-1})	Frequency(y^{-1})
1.	0.33	0.20
		0.09
		0.035
		0.0156
2.	0.20	0.09
		0.035
		0.0156
3.	0.09	0.035
		0.0156
4.	0.035	0.0156
5.	0.026	0.016
	0.016	
6.	0.016	0.0083
7.	0.0083	0.0039
8.	0.0039	0.0039

Table 3.5: PSD on MRA of test signal 1, using dmey wavelet.

Level	Details	Approximations
	Frequency (y^{-1})	Frequency(y^{-1})
1.	0.41	0.20
	0.33	0.09
	0.30	0.035
	0.20	0.0156
	0.16	
2.	0.34	0.16
	0.30	0.09
	0.20	0.035
	0.16	0.0156
	0.0.09	
3.	0.16	0.035
	0.09	0.0156
4.	0.046	0.0156
	0.035	

Table 3.6: PSD on MRA of test signal 1, using db2 wavelet.

closer than $2/N$ either miss out or merge into a single entity to give an average value [10]. It is shown here that such frequencies can be distinguished using the present MRA technique. By applying PSD directly on the test signal it is found that the closely spaced higher frequencies 0.2, 0.17 and 0.16 y^{-1} are captured, but the lower frequencies 0.035 and 0.028 y^{-1} merge to give an average value of 0.031 y^{-1}. This is because the separation between these lower frequencies is 0.007 y^{-1}, which is less than $2/N = 0.0078$ y^{-1}, whereas in the case of the higher frequencies PSD is able to capture both 0.17 and 0.16 y^{-1} as the frequency separation of 0.01 y^{-1} between them is higher than $2/N$. Results obtained by estimating the PSD of MRA decompositions using the dmey wavelet are shown in figure 3.10. At the second level of MRA, PSD captures all the three high frequencies, and at the fourth level of approximation the closely spaced lower frequencies are also captured. It is recalled that in direct PSD the frequencies 0.035 and 0.028 y^{-1} could not be separated.

A plausible explanation or theoretical justification for the ability to separate closely spaced frequencies using the approach proposed here would go as follows. Once it is decided that the time series being handled is nonlinear and possibly non-stationary, the wavelet technique is clearly an attractive option. However separating closely spaced frequencies by direct inspection of a wavelet map poses difficulties. In the first place even a pure sinusoid leads to wavelet coefficients over a band of time scales, so separating two closely-spaced frequencies this way would not be practical. Secondly, nonlinearity, non-stationarity, and the regime transitions that may occur as a consequence can cloud the issue. Finally even when a time series contains a pure sinusoid the interference (or leakage) it produces can make contributions not only within its own band but in neighboring bands as well [69]. This problem leads to serious difficulties in many technological applications; for example when there is interference by the mains frequency in an electronic signal that is the object of analysis. Much work has been done in this area, for example in [69, 70]. Here a variety of different methods for eliminating interference are reviewed and new ones suggested. In our application the objective is to separate frequencies for reasons already stated in the introduction. Here we fall back on the advantages of Fourier analysis, which has greater localization power in frequency space and should therefore be able to pick out sharp peaks when leakage is minimized.

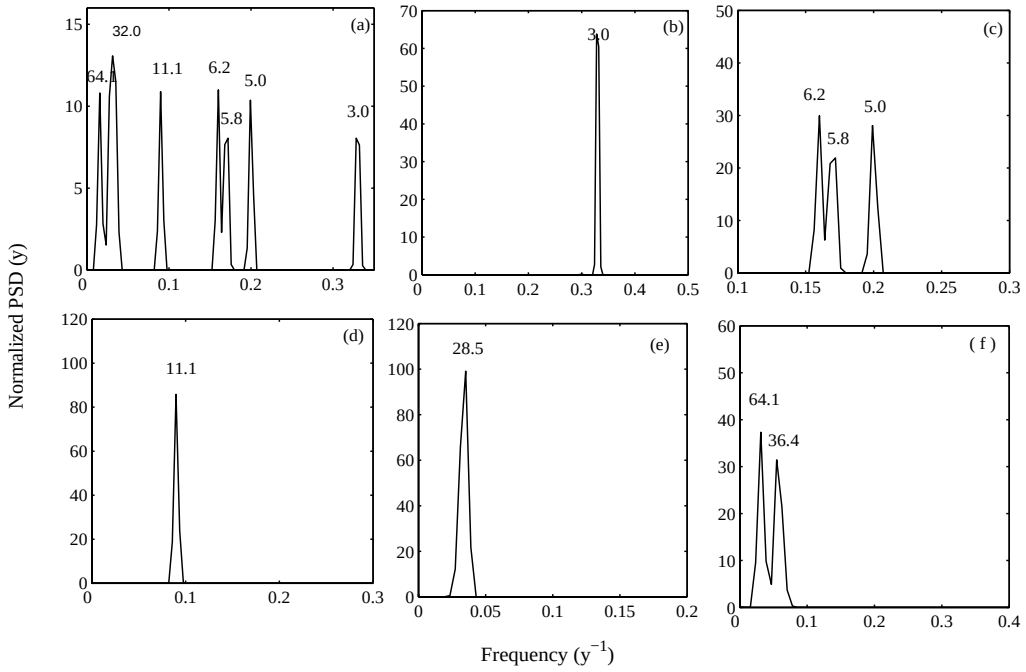

Figure 3.10: PSD on MRA of test signal 2 (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: Periods(y) are marked at each peak.

What is proposed here is thus a hybrid 'MRA+PSD' technique where the advantages of wavelet methods in handling nonlinear non-stationary time series are combined with those of spectral analysis by the following procedure. At first level of MRA decomposition the original time series X_N is divided into two partially reconstructed time series e^+_{N-1} and X^+_{N-1} as defined in the last section. Here e^+_{N-1} is obtained by setting approximate coefficients of the low frequency components to zero. This helps in the elimination of sinusoidal interference in the highest frequency band. The e^+_{N-1} is then subjected to classical Fourier spectral analysis, which is successful in separating closely spaced frequencies provided the associated spectral peaks are sufficiently narrow. This procedure is repeated for all levels and successively interference is removed from higher frequency bands by setting the corresponding approximate coefficients to zero.

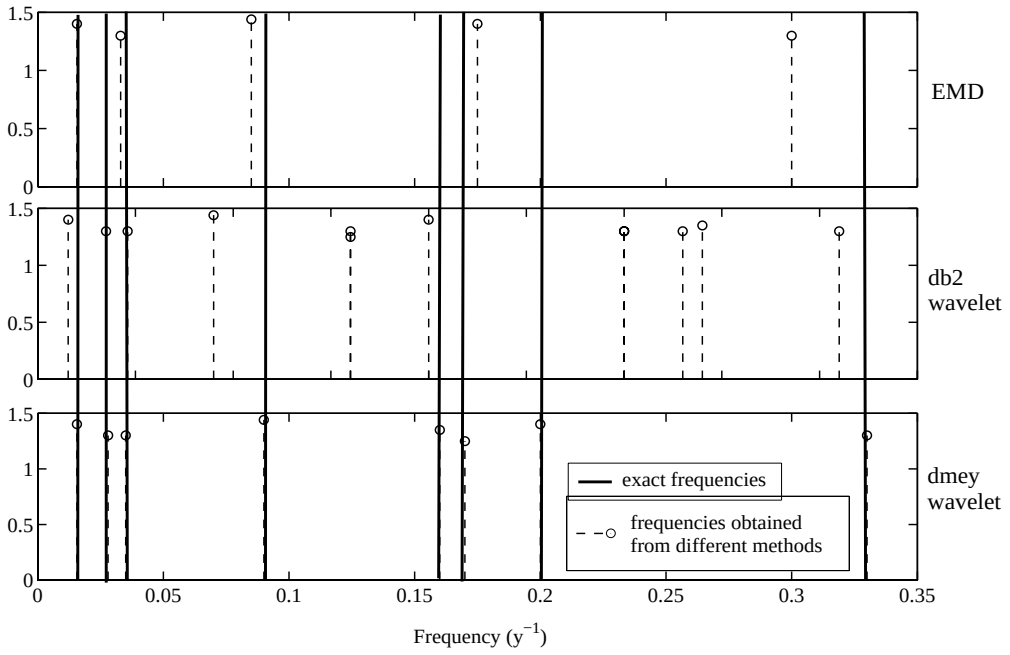

Figure 3.11: The exact frequencies given on test signal 2 are shown by bold vertical lines, whereas dotted lines show frequencies captured by EMD, db2 and the dmey wavelet. Ordinates represent normalized amplitudes.

3.2.4 Conclusions

The results from the analysis of the test signals before going on to treating rainfall data are briefly summarized here.

- When frequencies are widely separated (as in test signal 1) both EMD and MRA work well. However, when frequencies are closely spaced, as in test signal 2, EMD will capture only average values as it counts zero crossings, whereas PSD on MRA (with the dmey wavelet) is able to distinguish between closely spaced frequencies and capture them separately.

- Also, the direct PSD on test signal 2 merges closely spaced frequencies when the separation between them is less than $2/N$. However PSD on wavelet coefficients is able to distinguish between these frequencies.

- Comparing the performance of dmey, db2 and four other wavelets on the specially devised test signals it is found that dmey offers the best performance in separating closely-spaced frequencies without yielding spurious peaks. Figure 3.11 shows the difference in performance of EMD, db2 and dmey wavelet. The full vertical lines are the exact frequencies present in test signal 2; frequencies obtained from EMD and the present MRA using db2 and dmey wavelets are shown by dotted lines. It is clear from the figure that frequencies obtained from the dmey wavelet closely match with the exact frequencies.

- Other methods, including those based on maximum likelihood and regression techniques and on the Cramer-Rao bound can in principle resolve frequencies even when they are spaced closer than $1/N$, but they assume stationarity, large samples and high signal-to-noise ratio.

Chapter 4

Identifying Closely Spaced Periodicities

[1]

4.1 Introduction

The MRA+PSD method is now applied to the normalized time series of HIM annual rainfall data over the 120 y period 1871-1990 (described in Chapter 2). The use of MRA as a tool for analysing non-stationarity in the time series of the HIM rainfall is examined. This is required because our method demands combined use of wavelet and Fourier spectral analysis, and the latter is valid only for stationary time series. The method we describe helps us to isolate the stationary component of the time series, i.e. to de-trend it using long-scale wavelets. Using this approach it is found that the rainfall data is non-stationary only over long periods of order 60 y. Thus, application of spectral analysis on wavelet-based partially reconstructed time series, as used in Chapter 3, is justified. Using the discrete Meyer wavelet function, the normalized HIM rainfall annual data is decomposed into seven dyadic scales corresponding to 2, 4, 8, 16, 32, 64 and 128 y. However, the first four details and the fourth approximation are sufficient for the present analysis. As was explained for the

[1]Much of the material in this chapter has appeared in Azad *et al.* (2007a).

test signals in Chapter 3, the HIM time series can also be almost fully reconstructed by a decomposition upto four levels with an error of only 10^{-4}. Also it is shown that upto the fourth level of decomposition all the frequencies present in the rainfall are captured. Figure 4.1 shows the normalized HIM rainfall time series with its MRA decomposition at four levels.

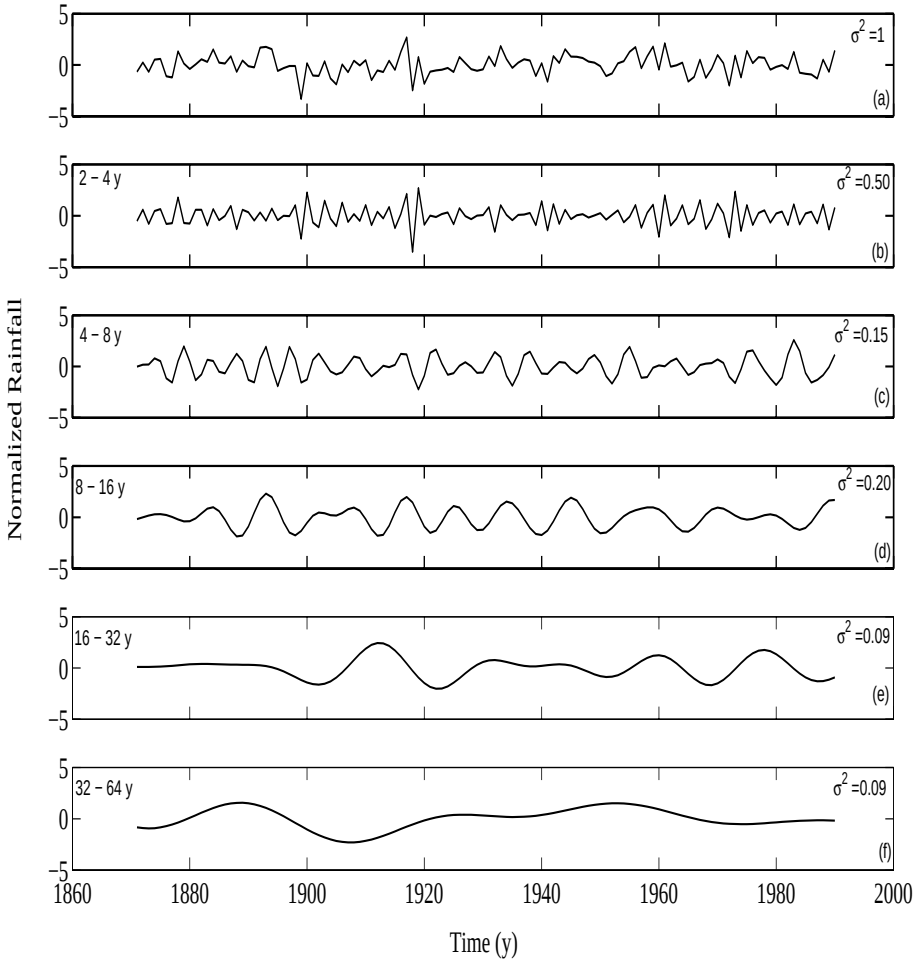

Figure 4.1: MRA of HIM rainfall data. (a) Normalized HIM time series from 1871-1990; (b) 1st partially reconstructed detail; (c) 2nd partially reconstructed detail; (d) 3rd partially reconstructed detail; (e) 4th partially reconstructed detail; (f) 4th partially reconstructed approximation.

Figure 4.2 shows how MRA divides the variance of the HIM time series into additive contributions from different levels. Note that the periods in the range 2-4 y at scale 1 and periods in the range 8-16 y at scale 3 together account for 72% of the total variance of HIM rainfall.

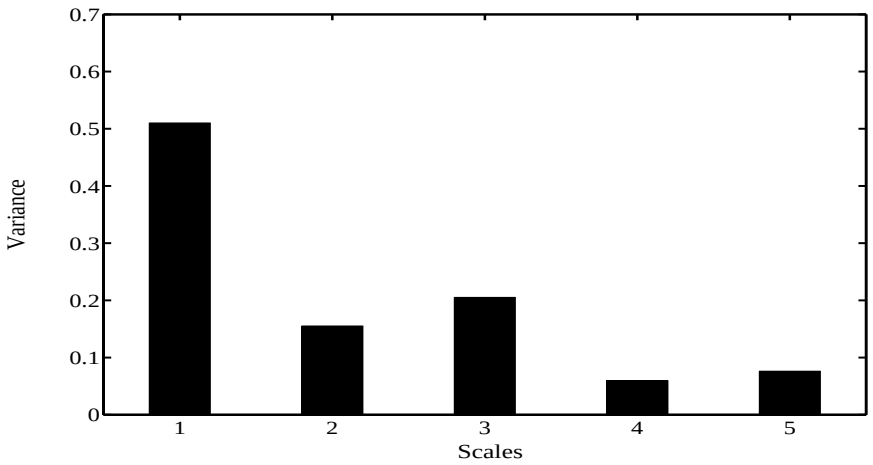

Figure 4.2: Contribution of variance at different scales for HIM region: scale 1 corresponds to 1st partially reconstructed detail; scale 2 corresponds to partially reconstructed 2nd detail, upto 4 details and scale 5 corresponds to 4th partially reconstructed approximation.

4.2 Non-Stationarity Tests

A random process is stationary if its statistical characteristics are invariant under time shifts, that is, if they remain the same when t is replaced by $t + \triangle$, where $\triangle$ is arbitrary. A weaker type of stationarity (called second order stationarity [8]) is that in which only the first– and second–order moments depend solely on time differences. Here we test for second order non-stationarity associated with significant differences in the mean value and/or variance of the time series under consideration. For non-stationary random processes wavelet based methods [71, 72] have been shown to be useful in recent years. However the non-stationarity present in the Indian rainfall is not very strong as assumed in these methods. The present analysis first uses MRA as a tool to identify the range of wavelet scales over which stationarity can be

assumed.

Recently reference [23] using time domain analysis have shown that HIM rainfall time series exhibits non-stationarity. By calculating the mean difference over two distinct test periods it was shown by them that the rainfall during 1933–1964 is higher than that during 1878–1913 with 97% confidence by z-test. As an interesting extreme case, the Konkan sub-division (on the west coast) is illustrated in figure 4.3 (taken from reference [23]) by plotting rainfall data over 1871–1990. This time series appears irregular and random, but one can already notice the marked difference in average rainfall between the two test periods. The difference is significant at the 99.9% confidence level. Indeed the non-stationarity of the data is clearly visible in the plot itself. (Incidentally similar marked differences are seen all along the west coast).

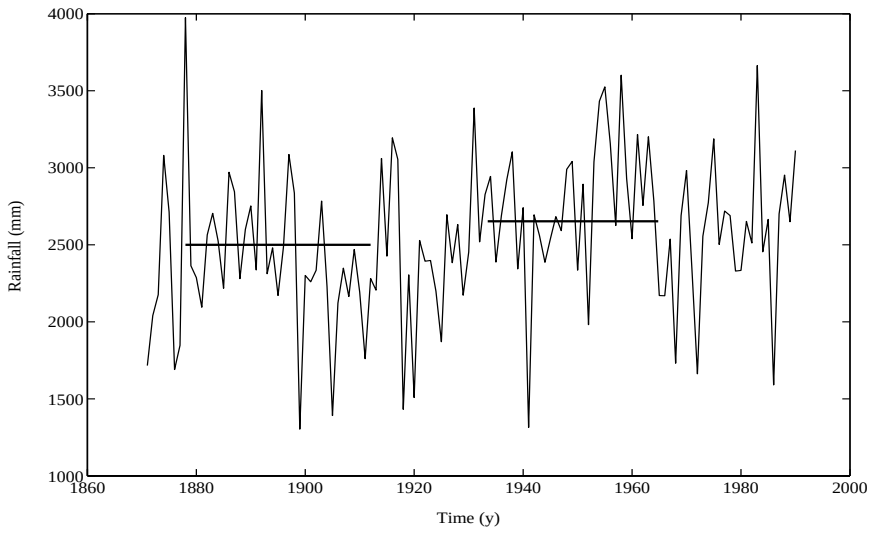

Figure 4.3: Time series of Konkan sub-visional rainfall, indicating test periods which show difference in mean.

Though the rainfall time series is non-stationary, we can however show, by applying statistical tests on MRA decomposition at different levels, that it can be expressed as a sum of stationary and non-stationary time series. That is, for short periods (at high frequencies) rainfall data actually behaves as a stationary time

Lag(y)	Autocorrelation coefficient
1.	-0.007
2.	0.180
3.	0.083
4.	-0.117
5.	-0.036
6.	-0.0001

Table 4.1: Autocorrelation coefficients of HIM rainfall at different lags.

series, whereas for longer periods (at low frequencies) it exhibits non-stationarity. Therefore the present analysis uses MRA as an exploratory tool to identify the range of wavelet scales over which stationarity can be assumed.

For this purpose, the autocorrelation coefficients of HIM rainfall at various lags are calculated using Eq. 1.1 and values are listed in table 4.1.

It is clear from the table 4.1 that the autocorrelation coefficient for HIM rainfall is quite low at various lags, for example at lag-1 it is -0.007, i.e. very close to zero. The implication of such a small autocorrelation is discussed in the next chapter. Here we look at autocorrelation coefficients of partially reconstructed time series at four levels for 119 lags, as shown in figure 4.4. It is clear from the figure that for levels 1-3, the time series moves occasionally away from its mean, but eventually returns to it, hence shows characteristics of stationarity. For the rest of the levels, however, it oscillates slowly showing periodic behavior.

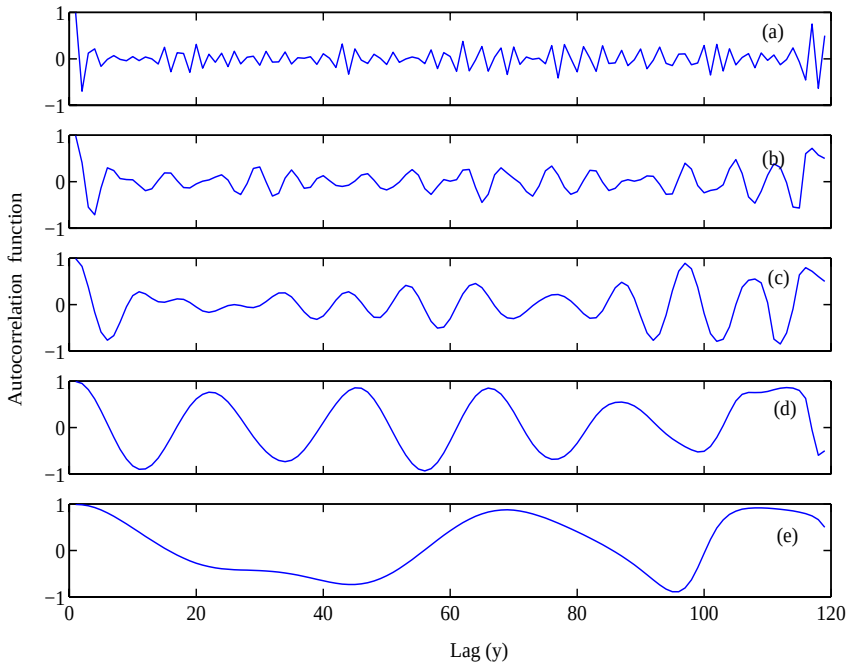

Figure 4.4: Auto-correlation coefficients for l19 lag values at four levels of (a) first reconstructed detail; (b) second reconstructed detail; (c) third reconstructed detail; (d) fourth reconstructed detail; (e) fourth reconstructed approximation.

4.2.1 Two statistical tests

Suppose we wish to test for possible non-stationarity in the HIM rainfall time series on a specified time scale. To do this we adopt the following procedure:

- Select two appropriate test periods with n_1 and n_2 samples respectively. (e.g. from 1871-1930 and 1931-1990, for a 60 y time scale, $n_1 = n_2 = 60$).

- First, test against the null hypothesis that the mean over the two test periods is the same. If μ_1 and μ_2 are the observed mean values over the two test periods, form the z-test statistic

$$Z = \frac{\mu_1 - \mu_2}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}. \tag{4.1}$$

where s_1^2, s_2^2 are estimators of the variances over the two test periods. The null hypothesis that $\mu_1 - \mu_2 = 0$ is rejected at confidence level $100(1-\alpha)\%$ if $|Z| > z_{\alpha/2}$ for an equal-tails test, where $|Z|$ is the absolute value of Z. (Values of $z_{\alpha/2}$ are tabulated in [73], for example.)

- Next, test for variance, adopting the null hypothesis that the variance over the two test periods is the same. This can be done by the F-test, using the statistic

$$F = \frac{s_1^2}{s_2^2}. \tag{4.2}$$

This ratio follows the F-distribution with $\gamma_1 = n_1 - 1$ and $\gamma_2 = n_2 - 1$ degrees of freedom. The null hypothesis is rejected at confidence level $100(1-\alpha)\%$ if the computed F value is greater than the tabulated value of $F_\alpha(\gamma_1, \gamma_2)$ (e.g. [73]).

4.2.2 Results

The two tests were performed on the two consecutive populations of 60 samples each from the reconstructed details and approximation of MRA decomposition at four levels; i.e. two test periods over 1871–1930 and 1931–1990 are analysed for stationarity. Results are shown in tables 1 and 2 respectively. It is clear from the tables that for both the tests the null hypothesis is rejected at 99.9% confidence level for the 60 year period at the fourth approximation of MRA decomposition. At shorter scales the hypothesis is not rejected. Hence we conclude that HIM rainfall exhibits non-stationarity only over long periods of order 60 y. Figure 4.5 shows the stationary component of HIM rainfall obtained from adding the partially reconstructed time series at four levels of detail. This reconstructed stationary time series accounts for 93% of the variance of the total rainfall. We conclude that the application of spectral analysis on the MRA decomposition, which we shall describe in the next section, is justified upto the 16-32 y band as till this band rainfall exhibits stationarity.

	μ_1^a	μ_2^b	Z	% confidence
$\hat{d}_1$	-0.0043	0.0024	-0.05	$< 10\%$
$\hat{d}_2$	-0.0036	0.023	-0.08	$< 10\%$
$\hat{d}_3$	-0.0085	0.156	-0.42	30%
$\hat{d}_4$	-0.006	0.008	-0.26	20%
$\hat{\alpha_4}$	-0.107	0.09	-3.83	99.9%

Table 4.2: z-test on HIM rainfall data and its MRA decomposition at different scales with % confidence at which null hypothesis is rejected. μ_1^a: Mean rainfall over test period 1871:1930; μ_1^b:Mean rainfall over test period 1931–1990.

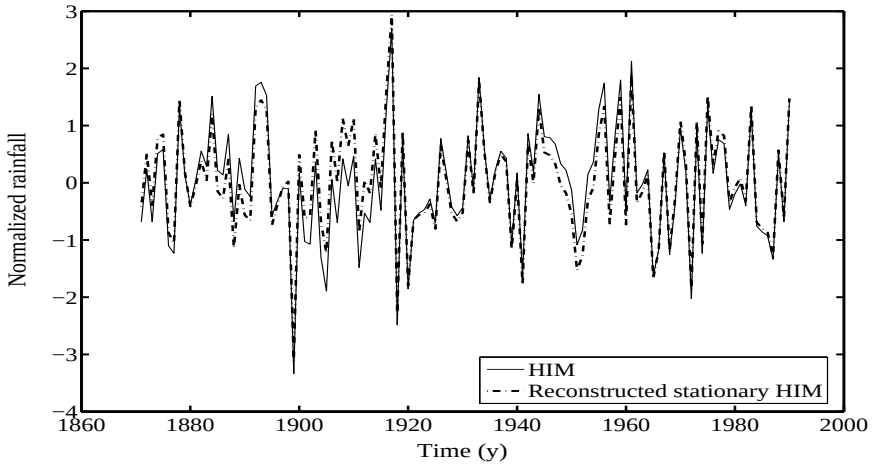

Figure 4.5: Normalized HIM rainfall time series compared with its stationary part obtained from MRA.

	s_1^{2a}	s_2^{2b}	F	% confidence
HIM	1.07	0.91	1.18	90%
$\hat{d}_1$	0.55	0.48	1.14	70%
$\hat{d}_2$	0.15	0.16	1.06	70%
$\hat{d}_3$	0.19	0.17	1.12	70%
$\hat{d}_4$	0.10	0.07	1.43	90%
$\hat{a_4}$	0.12	0.04	2.99	99.9%

Table 4.3: F-test on HIM rainfall data and its MRA decomposition at different scales with % confidence at which null hypothesis is rejected. s_1^{2a}:Variance of rainfall over test period 1871–1930; s_2^{2a}: Variance of rainfall over test period 1931–1990.

4.3 Identifying Closely Spaced Periodicities

As was done for the two test signals in Chapter 3, the PSD of each of the partially reconstructed time series of HIM rainfall obtained from the MRA decomposition is now estimated. Results are shown in figure 4.6 and table 4.4. It is immediately clear from the figure that PSD on MRA yields two more periods, respectively at 8.6 and 30.3 y. Their absence in the directly computed PSD (see figure 4.6) is consistent with the criterion cited earlier, as the frequency separation between the two period pairs (8.6 - 10.0) y and (30.3 - 60.6) y is each less than $2/N = 0.0167$ y^{-1}.

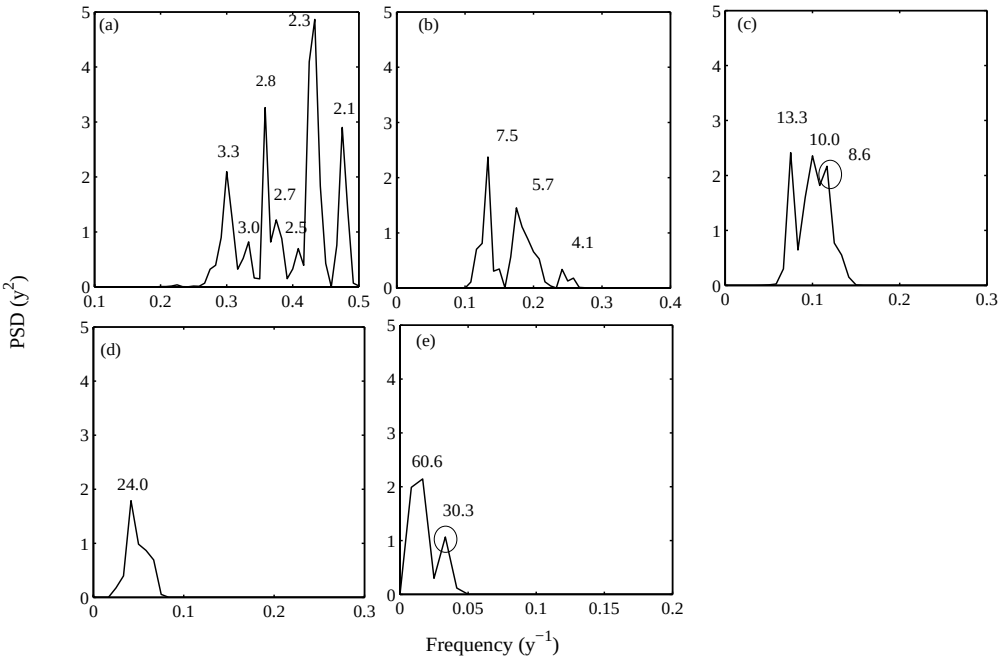

Figure 4.6: PSD on MRA of HIM (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: circled periodicities are the new periodicities which were missed out by direct PSD. Periods(y) are marked at each peak.

In figure 4.7, the direct PSD of normalized HIM rainfall is plotted against that of the MRA reconstructed time series (using the spectra of four levels of details and the

level	Details			Approximations		
	Frequency(y^{-1})	Period(y)	Amplitude	Frequency(y^{-1})	Period(y)	Amplitude
1.	0.475	2.1	2.9	0.24	4.1	0.3
	0.433	2.3	4.8	0.175	5.7	1.4
	0.384	2.6	2.1	0.15	6.6	0.4
	0.358	2.7	3.2	0.133	7.5	2.2
	0.3	3.3	2.1	0.1	10.0	2.6
				0.075	13.3	2.8
				0.0416	24.0	1.5
				0.0165	60.6	2.0
2.	0.24	4.1	0.3	0.116	8.6	2.1
	0.175	5.7	1.4	0.1	10.0	2.4
	0.15	6.6	0.4	0.075	13.3	2.8
	0.133	7.5	0.6	0.0416	24.0	1.5
				0.0165	60.0	2.0
3.	0.116	8.6	2.1	0.066	15.1	0.2
	0.1	10.0	2.4	0.0416	24.0	1.5
	0.075	13.3	2.9	0.0165	60.6	2.0
4.	0.066	15.1	0.2	0.033	30.3	0.3
	0.0416	24.0	1.5	0.0165	60.0	2.0
5.	0.033	30.3	0.3	0.0083	119.0	1.2
	0.0165	60.0	0.3			
6.	0.0084	119.0	0.6	0.0084	119.0	0.07
7.	0.0084	119.0	0.8	.0084	119.0	0.02

Table 4.4: Periods(y) obtained in HIM rainfall from PSD on MRA at various levels of details and approximations.

fourth approximation). (Note of course that the area under both the PSD functions is the same. Also, the peaks from the present method are sometimes higher e.g. at 13.3 and 7.5 y, and sometimes lower, e.g. at 24.0 and 10.0 y.) The striking difference is the 8.6 y period, which was missed by direct PSD.

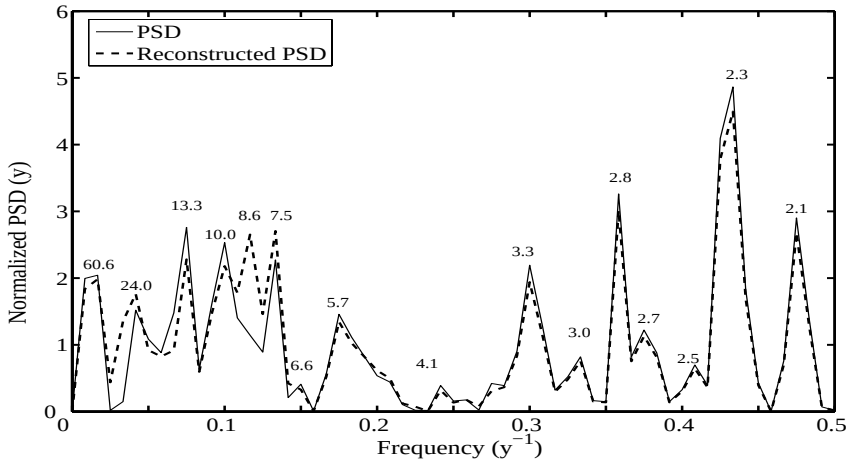

Figure 4.7: The direct PSD of normalized HIM rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

4.3.1 Handling border distortion

Classically, the DWT is defined for sequences with length of some power of 2, and different ways of extending the analysis to samples of other sizes to avoid border distortion have been proposed. These include zero-padding, smooth padding, periodic extension, and symmetric boundary replication (symmetrization), as discussed in [36]. Bearing in mind that the sample size of rainfall data is 120 y and we decompose it at eight levels, a total of 128 samples was needed, so different padding schemes have been considered in this work. When decomposing this time series using MRA we experience border distortion at various levels in terms of change in amplitudes in the PSD's of reconstructed time series with respect to the PSD of HIM rainfall time series. Hence we have applied periodic extension on the first two details, and symmetrization on the the next three decompositions.

The important question that arises at this stage concerns the statistical signifi-

Level	CWT	zero crossings on EMD	PSD on MRA
1.	3.0	2.7	2.1, 2.3, 2.5, 2.7, 2.8, 3.0, 3.3
2.	5.8	5.7	4.1, 5.7, 6.6, 7.5
3.	11.6	12.0	8.6, 10.0, 13.3
4.	20.8	24.0	24.0
5.	37.0		30.3
6.	80.0	60.0	60.0

Table 4.5: Periods (y) obtained by continuous wavelet transform (Narasimha and Kailas 2001), Empirical mode decomposition (Iyengar and Kanth 2004) and PSD on Multi-resolution analysis (present technique) of HIM.

cance of the peaks so discovered. We shall return to this issue in detail in the next chapter.

4.3.2 Comparison with different methods

As already described in Section 1.2, the HIM time series has been analysed by various workers using different methods to detect periodicities. Using continuous wavelet maps, reference [21] reported six modes with periods of 3.0, 5.8, 11.6, 20.8, 37.0 and 80.0 y; using EMD, reference [55] found five empirical time series of intrinsic mode functions with associated periods of 2.7, 5.7, 12.0, 24.0 and 60.0 y respectively; in [51] the PSD is estimated using the Blackman-Tukey [52] algorithm, and detected thirteen periodicities, respectively with periods of 2.1, 2.3, 2.5, 2.8, 3.1, 3.5, 4.0, 4.7, 5.7, 7.3, 10.0, 16.0 and 40 y; of these only the 2.8 y period was statistically significant at 95% confidence against white noise. Table 4.5 compares the periodicities obtained by the different methods. From this comparison the following are observed:

- At the 1st level of MRA decomposition the present method captures the seven very close periods of 2.1, 2.3, 2.5, 2.7, 2.8, 3.0, 3.3 y, whereas CWT shows a period of 3.0 y and EMD of 2.7 y, which are some weighted averages of the periods obtained from MRA.

- At the 2nd level the present MRA analysis captures the four periods of 4.1, 5.7, 6.6 and 7.5 y, whereas the only period captured by CWT and EMD is close to 5.7 y, which is again some weighted average of the periods obtained from MRA.
- At the 3rd level also MRA captures the nearby periods of 8.6, 10.0 and 13.3 y, whereas CWT shows a period of 11.6 y and EMD of 12.0 y.
- At the 4th level MRA and EMD both capture a period of 24.0 y.
- At the 5th level MRA captures a period of 30.3 y whereas CWT has 37.0 y.
- At the last level a period of 60.6 y is obtained by MRA and EMD whereas CWT shows a period of 80.0 y. But these must be treated with great caution, as the total length of the data accommodates fewer than 2 cycles in this period range, and both end effects and non-stationarity can be important.

4.4 Summary and Conclusions

Based on the work reported here the following conclusions can be drawn:

- Classical spectral analysis methods cannot separate frequencies spaced closer than $1/N$, where N is number of sample points in a time series. Moreover frequencies which are closer than $2/N$ often get merged to yield only a weighted average value.
- The MRA technique can be applied to non-stationary time series and can be used as a tool to identify and separate frequencies which are closer than $2/N$.
- Application of spectral analysis on reconstructed wavelet coefficient time series is justified by the fact that rainfall is stationary except over long periods of order 60 y.
- By analysing performance on two test signals, we explicitly demonstrate that the EMD and CWT methods counting zero crossing intervals capture only average periods of closely spaced frequencies, whereas the present MRA method

is able to capture many more frequencies. The same is true in the case of HIM rainfall; PSD on MRA decomposition is able to yield periods close to each other missed in earlier work by CWT and EMD methods, which yield only some weighted average values of nearby periods.

- There are therefore grounds to believe that certain periodicities in rainfall have been missed by earlier techniques.

- In case of HIM rainfall, the present technique reveals 17 periodicities at 2.1, 2.3, 2.5, 2.7, 2.8, 3.0, 3.3, 4.1, 5.7, 6.6, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3, 60.0 y. Of these the periodicities at 8.6 and 30.3 y are absent in all earlier work.

- The important question of statistical significance of these periodicities is discussed in Chapter 5.

Chapter 5

Statistical Significance Tests

[1]

5.1 Introduction

Many real-life time series include a combination of periodic or nearly periodic signals embedded in a noise-like background with a continuous spectrum. The first step in any investigation of the temporal structure of a time series is therefore to identify any significant periodicities that may be present in the given time series. Though various modern methods of spectral and wavelet analysis [36] have been developed in recent years to estimate the spectrum of a time series, reference [59] explain that the progress has been hindered by a lack of effective statistical tests to discriminate between potential oscillations and anything but the simplest form of noise, that is, "white " (independent, identically distributed) noise, in which power is independent of frequency. Here a major problem lies in testing for statistical significance, in particular in selecting an appropriate background spectrum against which the peaks found in an estimated spectrum can be compared. For this purpose different signal detection methods have been developed [74, 75, 59, 76, 61, 60].

In this chapter, the spectrum of HIM rainfall annual time series, as estimated in the previous chapter, is tested for the significance of the peaks observed in the

[1]Some of the material in this chapter is accepted and in press, Azad *et al.* (2007b).

estimated spectrum. The test is carried out against the classical reference spectrum proposed in [29] (explained in Section 1.1) as well as a more recent one due to [**?**]. The power of the MRA technique using a combination of wavelet and Fourier tools has already been demonstrated in the previous chapter: the technique identified 17 peaks in the spectrum of HIM rainfall time series. It is our objective here to assess the significance levels of the periodicities thus obtained against appropriate colored background spectra at each scale.

5.2 White Noise Reference Spectrum

Reference [29] have derived a formula for calculating the experimental red noise spectra to be used in constructing a null hypothesis for significance testing, as explained in Chapter 1 (Eq. 1.13). For HIM time series it is found that the lag-1 autocorrelation coefficient used in the Eq. 1.13 is very small ($\alpha = -0.007$), so the reference spectrum described is very close to white.

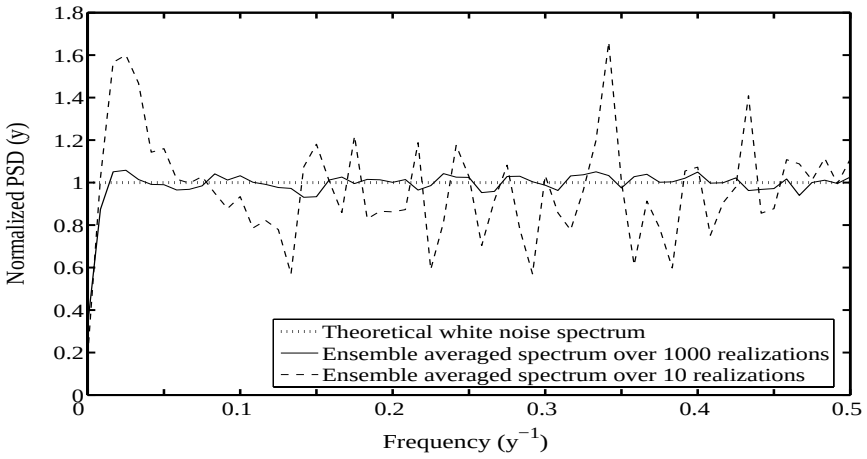

Figure 5.1: Theoretical white noise spectrum and the one obtained from ensemble averaged PSD of 10 and 1000 realizations of white noise.

In figure 5.1, the white noise spectrum obtained from Eq. 1.13 is plotted for lag-1 auto-correlation. Also PSD function of one realization of Gaussian white noise (generated using the algorithm of [77]) of 120 samples is estimated using the Welch

technique described in Section 1.1. As references [61] and [22] point out, a single, short realization of a noise process can have a spectrum that appears to differ greatly from the theoretical one used as a classical reference spectrum. It is only the average of such sample spectra over many realizations that will tend to the true spectrum of the ideal noise process. To demonstrate this, figure 5.1 shows spectra obtained from 120 samples of the white noise averaged over 10 and 1000 such realizations.

5.3 Significant Periodicities against Classical Reference Spectrum

As mentioned in the previous chapter, the HIM time series has been analysed by various workers using different methods to detect periodicities. However neither of these studies attempted to assess the statistical significance of their results. The only significant periodicity reported in [51], using the Blackman-Tukey algorithm [52] for Fourier spectral analysis, is 2.8 y at 95% confidence level against a reference white noise spectrum.

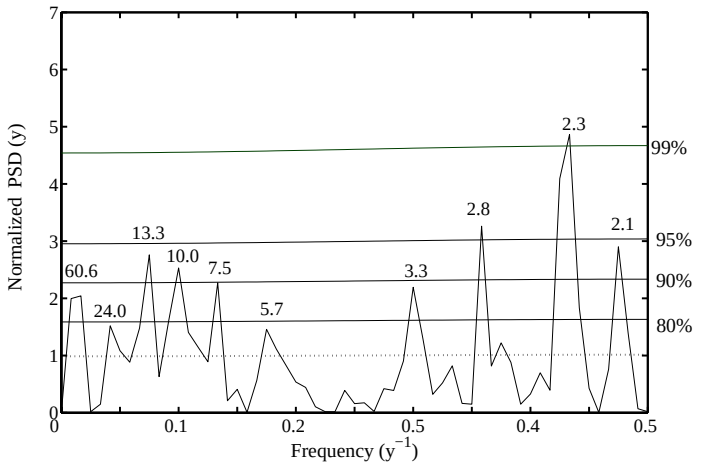

Figure 5.2: HIM spectrum obtained from Welch technique and reference spectrum (dashed line) from Eq. 1.13 at different confidence levels (horizontal lines). Periods (y) are marked at each peak.

However it was suggested in [11] that periodicities should be tested for high levels of significance i.e., for a sample of size N, he proposes that "a good rule of thumb is not to get excited by significance levels less than $1 - 1/N$ ". For HIM rainfall data with $N = 120$, $1 - 1/N = 99.17\%$. We estimate the PSD function of HIM rainfall time series using the Welch technique, described in Section 1.1. The significance test is illustrated in figure 5.2 using the nearly white reference spectrum for $\alpha = -0.007$ obtained from Eq. 1.13 (the area under the PSD function and the reference spectrum is of course the same). It is seen that the test yields two significant periods, 2.3 y at 99% and 2.8 y at 95% confidence level.

S.no	Sub-division	Auto Correlation		
		Lag-1	Lag-2	Lag-3
	HIM	-0.007	0.18	0.08
1.	Haryana	-0.02	0.14	0.06
2.	Punjab	-0.13	-0.10	0.28
3.	West Rajasthan	0.06	-0.04	0.04
4.	East Rajasthan	0.002	0.11	0.05
5.	West M.P.	0.04	0.27	0.10
6.	East M.P.	-0.0001	0.13	0.17
7.	Gujarat	-0.02	-0.003	0.01
8.	Saurashtra	-0.07	0.004	0.09
9.	Konkan	0.11	0.03	0.007
10.	Madhya Maharashtra	0.18	-0.05	-0.04
11.	Marathwada	0.2	-0.03	0.03
12.	Vidarbha	-0.06	0.23	0.07
13.	Telangana	0.15	0.09	0.02
14.	North Interior Karnataka	0.10	-0.04	0.13

Table 5.1: Auto-correlation coefficient at various lags for HIM and 14 sub-divisional rainfall during 1871-1990.

The PSD function for each of the 14 sub-divisional rainfall time series constitut-

ing HIM may be estimated by the same method. The auto-correlation coefficient for HIM and for the 14 sub-divisional rainfall time series are calculated at three values of lag and are given in table 5.1. For example, East M.P. shows -0.0001 value at lag-1, whereas at lag-3 the value is 0.17. Interestingly, the autocorrelation at higher lags is often greater than at lag-1 (West Rajasthan, Konkan, Madhya Maharashtra, Marathwada, Telangana however behave differently: higher-lag autocorrelations are lower. We shall return to these points later.) The significance test is performed for each sub-division exactly as for HIM, using the values of α from table 5.1. The results are shown in figures 5.3–5.5. The periodicities obtained in the 14 sub-divisional rainfall time series are listed in table 5.2.

S.no	Sub-division	90% Significance	95% Significance
1.	Haryana	2.8, 8.6, 10.9	2.8, 8.6
2.	Punjab	2.8, 60.6	2.8, 60.6
3.	West Rajasthan	2.3, 2.8, 3.3, 8.6, 12.0	2.8
4.	East Rajasthan	2.3, 2.8, 3.3, 8.6, 13.3	2.3, 3.3, 13.3
5.	West M.P.	2.1, 2.3, 2.8, 7.5, 13.3, 60.6	2.1, 2.3, 13.3, 60.6
6.	East M.P.	2.1, 10.9, 60.6	2.1, 60.6
7.	Gujarat	2.3, 3.3, 4.1, 4.7, 12.0	2.3, 4.7
8.	Saurashtra	2.3, 2.9, 3.0	2.3, 2.9, 3.0
9.	Konkan	2.3, 2.8, 3.0, 20.0	2.3, 20.0
10.	Madhya Maharashtra	2.3, 2.8, 3.0, 6.0, 15.1	6.0
11.	Marathwada	7.5, 24.0	
12.	Vidarbha	2.3, 24.0	2.3, 24.0
13.	Telangana	2.1, 2.3, 2.8, 7.5, 10.0	7.5, 10.0
14.	North Interior Karnataka	2.6	2.6

Table 5.2: Periodicities present in the monsoon rainfall of 14 sub-divisions at two significance levels.

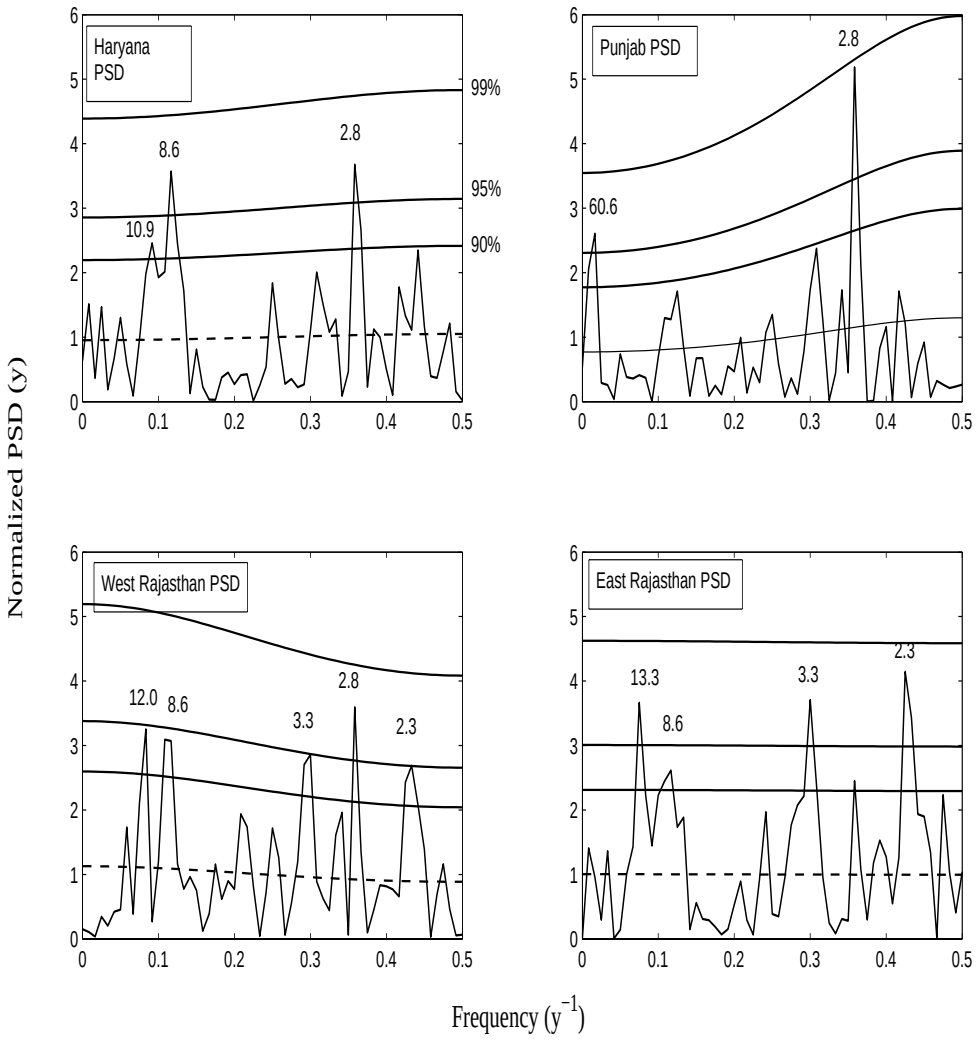

Figure 5.3: Power spectrum of normalized time series of sub-divisional rainfall.

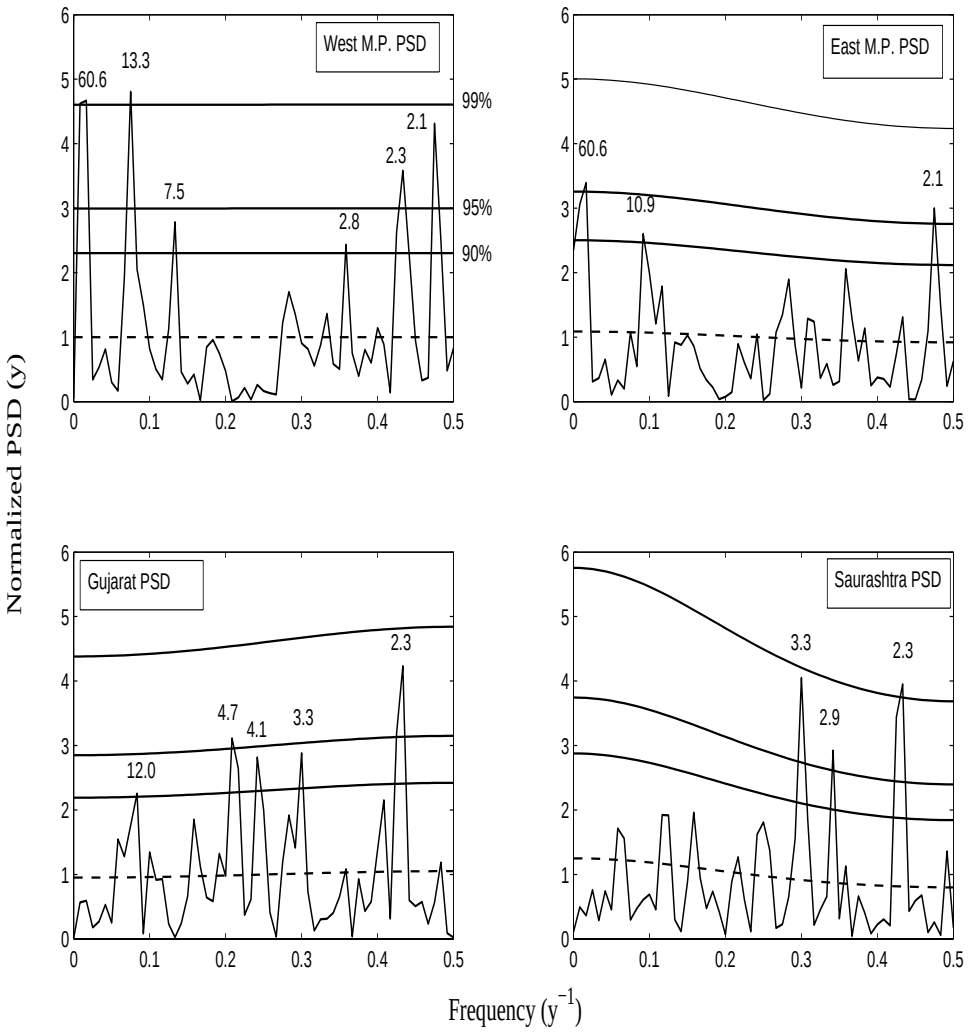

Figure 5.4: Power spectrum of normalized time series of sub-divisional rainfall.

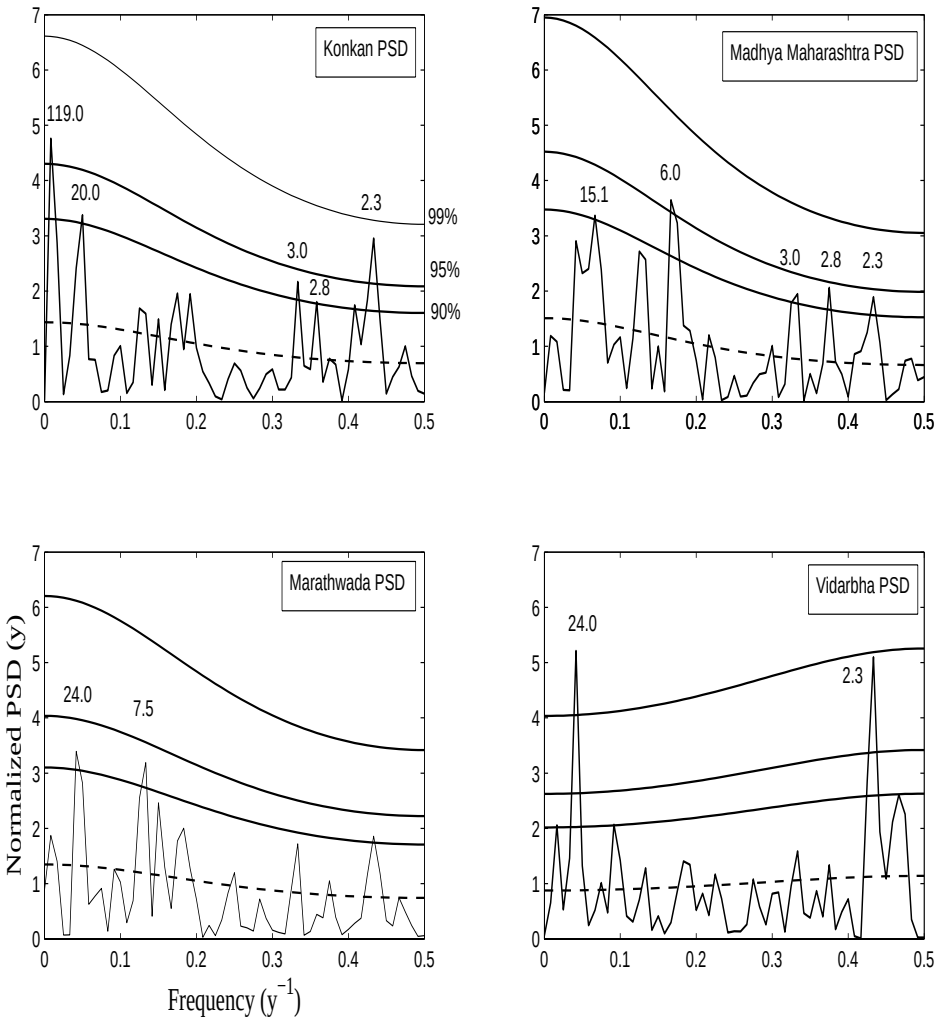

Figure 5.5: Power spectrum of normalized time series of sub-divisional rainfall.

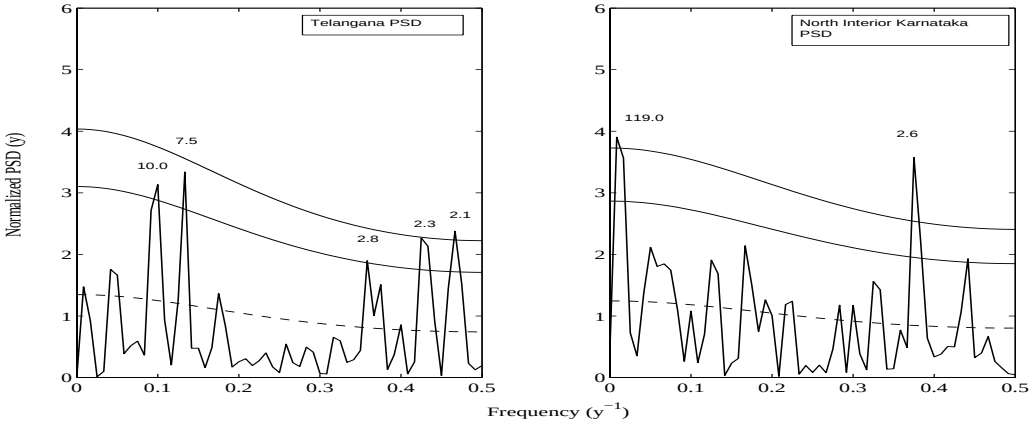

Figure 5.6: Power spectrum of normalized time series of sub-divisional rainfall.

5.4 The Spectral Dip in Some Sub-divisional Rainfall Data

The spectrum in figure 5.2 shows relatively high power at both lower and higher frequencies, and suggests that there is a spectral dip in a band around the frequency 0.25 y^{-1}. That this spectral dip is not a mere statistical fluctuation is shown by the following argument.

First we calculate the average PSD values in the frequency band 0.19 to 0.29 y^{-1} (table 5.3) for each sub-division. We now use a t-test. For testing against the null hypothesis that the mean $\mu = a$, where a is a specified number, on the basis of a random sample of size n, compute

$$t = \frac{\overline{x} - a}{s/\sqrt{n}} \tag{5.1}$$

where s is the sample standard deviation. For a one-sided test of the null hypothesis $\mu = a$ against the alternative $\mu < a$, use the rejection region as $t > t_{\alpha,n-1}$. From table 5.1, the power over the frequency band 0.19 to 0.29 y^{-1} has an average value ($\overline{x}$) of 0.67 and standard deviation (s) of 0.31. Therefore the value of t (with $a = 1$), comes to be 3.92 from Eq. 5.1. Also, the tabulated t-value for $n - 1$ degrees of freedom is 3.56 [73]. Therefore the null hypothesis that the spectral dip in the

frequency range 0.19–0.29 in the HIM spectrum is a result of statistical fluctuation in the white noise reference spectrum of unity is rejected at a confidence level of 99.5%.

Furthermore, HIM rainfall is the weighted average of 14 time series, one from each of the 14 meteorological sub-divisions constituting the HIM region. It is found that seven out of the 14 sub-divisional time series also exhibit a spectral dip around the frequency 0.25 y^{-1}. This incidentally raises the question to what extent the HIM region is spectrally homogeneous; this question is considered in the next chapter. Meanwhile we conclude that the reference spectrum must take account of the spectral

S.no	Sub-division	Average of PSD values
1.	Haryana	0.48
2.	Punjab	0.56
3.	West Rajasthan	1.10
4.	East Rajasthan	0.96
5.	West M.P.	0.50
6.	East M.P.	0.65
7.	Gujarat	1.45
8.	Saurashtra	0.83
9.	Konkan	0.51
10.	Madhya Maharashtra	0.48
11.	Marathwada	0.49
12.	Vidarbha	0.58
13.	Telangana	0.28
14.	North Interior Karnataka	0.54

Table 5.3: Average PSD value in the frequency band 0.19–0.29 y^1, for each sub-division.

dip and hence there is a need for devising a more appropriate background spectrum.

5.5 Monte Carlo Reference Spectra

One method proposed in [62] for devising an appropriate reference spectrum breaks the given time series into a number of blocks of the same length, and then randomly scrambles their order. This preserves short-term correlations but causes the timing of major events to be randomly re-arranged with respect to one another.

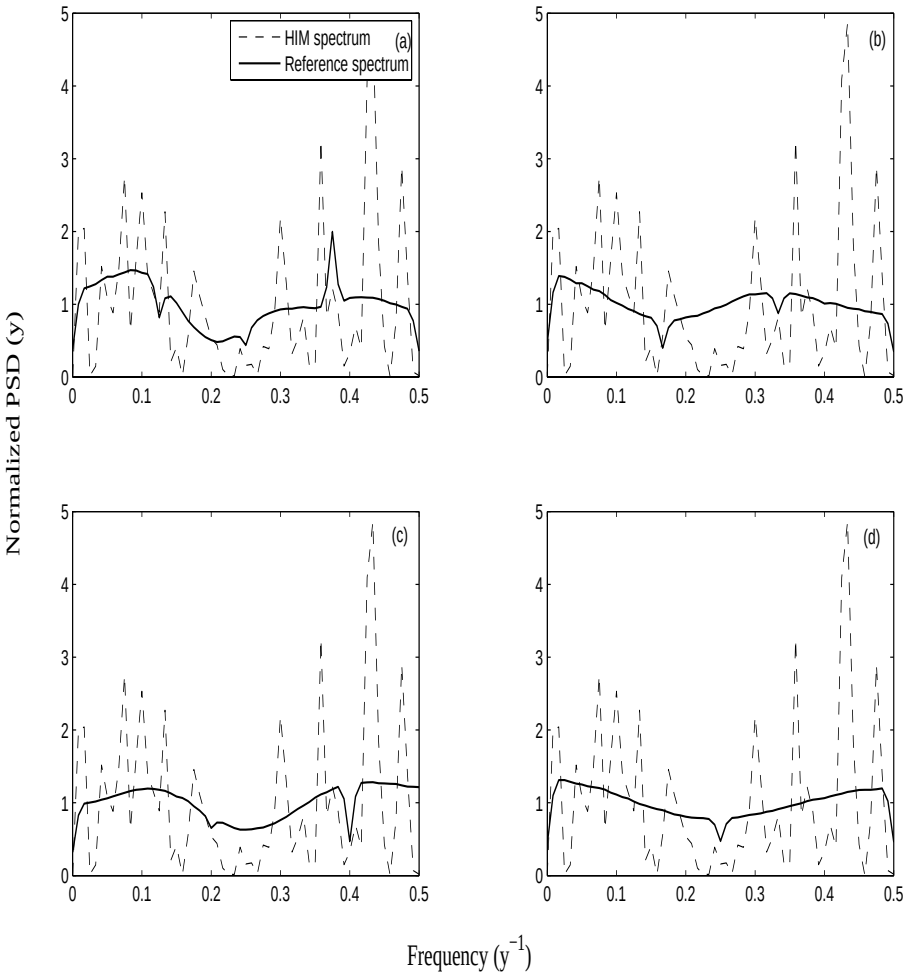

Figure 5.7: Reference spectrum obtained from the scrambling of HIM rainfall time series. (a) bin size 8 y; (b) bin size 6 y; (c) bin size 5 y; (d) bin size 4 y.

We scramble the time series of HIM rainfall by choosing 15, 20, 24 and 30 blocks

of size 8 y, 6 y, 5 y and 4 y bin respectively. For each block size the reference spectrum is obtained by averaging the PSD's over 10,000 Monte Carlo simulations of the scrambled rainfall data. Results are shown in figure 5.7, where the HIM spectrum is plotted against different reference spectra. It may be noted that in each of the reference spectra a sharp 'spike' at a frequency 1/bin size cycles/y is present. For example (see figure 5.7), for a bin size of 6 y, symmetric sharp dips can be observed at 0.16 and 0.34 = (0.50 – 0.16) y^{-1} in the reference spectrum. These can be attributed to a small enhancement at that frequency due to the choice of the 6 y dividing width.

However, all four spectra in figure 5.7 show a broad dip around 0.25 y^{-1} (figure 5.7 (b) least of all). It is observed from the figure that the reference spectrum obtained from the bin size 4 y seems appropriate as it takes the shape of the HIM spectrum.

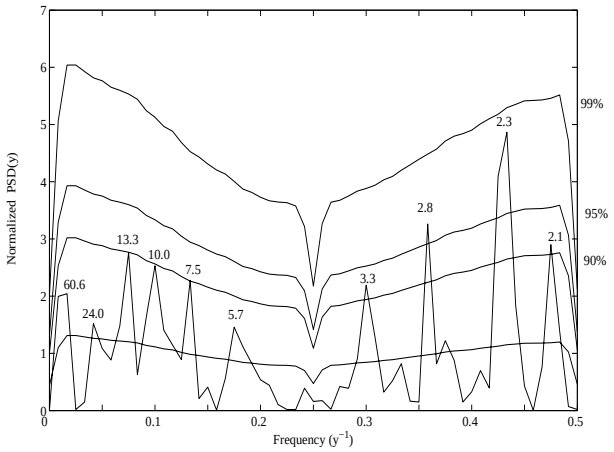

Figure 5.8: HIM spectrum obtained from Welch technique and reference spectrum (dashed line) obtained from scrambling of HIM data at different confidence levels. Periods (y) are marked at each peak.

Using this reference spectrum for further analysis, the PSD of HIM and the significance levels of the peaks are shown in figure 5.8 by curved lines. Based on this background, we find that the 2.3 and 2.8 y period are statistically significant above a confidence level of 95%.

5.5.1 Spectral analysis on MRA

We present below a new procedure for checking the significance levels of the partially reconstructed time series obtained in Section 3 based on wavelet coefficients at different frequency bands. The procedure automatically takes account of the general nature of the rainfall spectrum, including in particular the dip at around 0.25 y^{-1}.

The procedure is as follows. The PSD of each of the partially reconstructed time series obtained from MRA at four levels is estimated. From figure 5.9, it is clear that PSD at each scale is a part of the HIM spectrum (Fig. 5.2) in that particular band. Reference spectra using Eq. 1.13 are formulated for the value of the lag-1 autocorrelation coefficient at each scale. It is found that, unlike for the total HIM rainfall, the reference spectra at different scales are colored. For example, at the first level of decomposition the autocorrelation coefficient of the partially reconstructed time series is $\alpha = -0.69$, hence blue; for the second level $\alpha = 0.40$, hence red. The results are shown in figure 5.9. This method shows 10 periodicities above the 99% confidence level (Figure 5.9): 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3, 60.6 y, using in each case the appropriate colored reference spectrum. It is to be noted that the periodicity at 2.3 y is no longer highly significant, as it appeared to be in figure 5.2, but 2.8 and even 3.3 y are significant at 99.9% confidence level.

Therefore it appears that the advantage of using PSD on MRA decomposition instead of applying it directly to the time series is twofold:

- First, as found in Chapter 4, the PSD on reconstructed time series based on wavelet coefficients is able to distinguish between certain closely spaced periods at different scales which are missed by the direct PSD on the time series. Two such significant periodicities, namely 8.6 and 30.3 y, are present in figure 5.9 at 99.9% confidence level but are absent in the direct PSD.

- Also, using MRA decomposition and partial reconstructions, periodicities may be tested against the appropriate colored reference spectra in each scale band. For example at the first level of decomposition for high frequencies the reference spectrum is blue whereas for low frequencies it is red. This procedure reveals many periodicities that are significant at high confidence levels. (For a

related but alternative formulation of spectrally matched noise time series and its use in wavelet cross-spectral analysis, see Bhattacharyya and Narasimha 2007.)

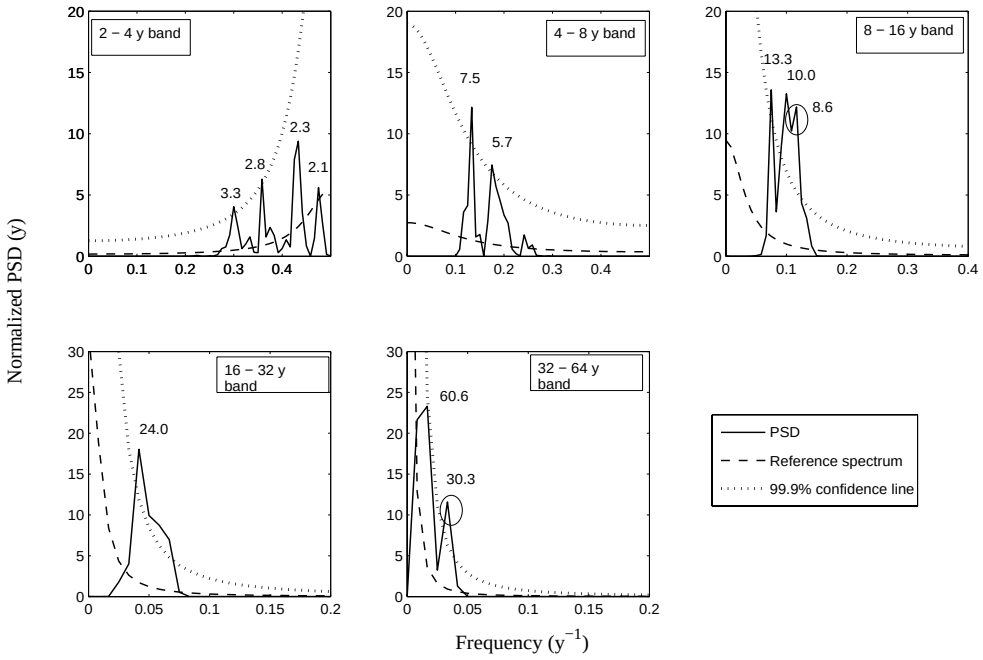

Figure 5.9: Significance test on MRA of HIM rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation.

5.6 Conclusions

- The auto-correlation coefficient of HIM rainfall series at lag 1 is -0.007 which is close to zero. This implies that, in the standard procedure currently used, the periodicities in HIM rainfall are tested against a reference spectrum that is nearly flat (i.e. white noise). However we have presented strong evidence for the presence of a spectral dip in HIM rainfall around a frequency of 0.25 y^{-1}, and construct a testing procedure that takes this into account. Thus,

using MRA we have shown that at the first level of decomposition for high frequencies the reference spectrum is blue, whereas for low frequencies it is red. Hence the statistical significance of these periodicities must be tested against appropriately colored noise, matching the observed spectrum of rainfall.

- The classical significance test on the PSD of HIM rainfall time series shows only one periodicity of 2.3 y above 99% confidence level, whereas the CWT and EMD methods of [21] and [55] suggest respectively 3.0, 5.8, 11.6, 20.8, 37.0, 80.0 y and 2.7, 5.7, 12.0, 24.0, 60.0 y periods. Here ten periodicities are significant above 99.9% confidence level using the present method of analysing PSD of reconstructed time series based on MRA. These periods are 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3, 60.6 y. The significant peaks account for 64% of the variance of the total rainfall. However the 60.6 y periodicity is close to the boundary of stationarity, and should be interpreted with caution.

- Table 5.4 compares the periodicities reported in HIM rainfall using different methods.

Table 5.4: Periods (y) obtained from different methods.

Level	PSD Kumar (1997)	CWT Narasimha & Kailas (2001)	EMD Iyengar & Kanth (2004)	PSD + MRA Fig. 5.9	PSD + Monte Carlo, Fig. 5.8	PSD + AR(1) Fig. 5.2
1.	2.8(95%)	3.0	2.7	2.1(80%) 2.3(95%) 2.8(99.9%) 3.3(99.9%)	2.3(95%) 2.8(95%)	2.3(99%) 2.8(95%)
2.		5.8	5.7	4.1 (80%) 5.7(99.9%) 7.5(99.9%)		
3.		11.6	12.0	8.6(99.9%) 10.0(99.9%) 13.3(99%)		
4.		20.8	24.0	24(99.9%)		
5.		37.0		30.3(99.9%)		
6.		80.0	60.0	60.6(99.9%)		

Chapter 6

Spectral Homogeneity

[1]

6.1 Introduction

The main aim of this thesis is to search for significant periodicities in Indian monsoon rainfall annual data over the 120 y period 1871 to 1990. One persistent question that arises in the quest for identifying periodicities is the heterogeneity of the Indian monsoon. In spite of many studies of the problem, and the considerable evidence of this heterogeneity, many analyses still continue to rely on all-India indices. Even the homogeneous regions identified by different workers may contain heterogeneities not considered in the criteria laid down for determining the degree of homogeneity. For example, we saw in Chapter 5 that only 7 out of the 14 sub-divisions in the HIM region exhibit a characteristic spectral dip at 0.25 y^{-1}. Using heterogeneous data will mask inherent periodicities in the phenomenon. Towards that end we here propose a quantitative definition of spectral homogeneity, and use this concept to identify a spectrally homogeneous (SHIM) sub-region within the HIM region.

The new methodology proposed here is based on the observation that homogeneous zone rainfall time series are derived by an averaging process over numerous

[1]Some of the material in this chapter is to be communicated, Azad and Narasimha (2008).

other time series of the same length, involving $O(10)$ sub-divisional time series and $O(10^2)$ rain gauge records. The sub-divisional rainfall time series cannot be treated as independent of each other because of significant cross-correlations among them. The methodology proposed here is based on multivariate data analysis that takes into account these cross-correlations among the sub-divisions constituting the SHIM sub-region.

6.2 Testing for Homogeneity

Homogeneity is a term often used to describe the common patterns in a given system. The basic hypothesis (which needs to be tested) is that the data is coming from one and the same process. In mathematical terms : the data are realizations from one and the same probability distribution function. Statistical procedures are available to check the homogeneity in a data set. Many attempts have been made to classify India into different groups which can be considered as homogeneous with respect to the variation of rainfall [78, 79]. The techniques most commonly used for grouping stations or grid points to define homogeneous regions are cluster analysis, principal component analysis, and correlation analysis.

The India Meteorological Department (IMD) divides the country into 29 meteorological sub-divisions. Reference [50] group 14 sub-divisions (151 rain gauge stations) into a HIM region. The (seasonal) monsoon rainfall of these 14 sub-divisions is more than 80% of the annual amount. The series are highly interrelated and their relationships with regional/global circulation parameters are also similar.

Though much work has been done to study the temporal and spatial variability of the Indian monsoon [80, 78], spectral methods have played no direct role in identifying homogeneous regions in India. This chapter is an attempt to quantify the spectral homogeneity within the HIM region. As we have seen, HIM rainfall represents an area-weighted average over an ensemble of the 14 sub-divisions Haryana, Punjab, West Rajasthan, East Rajasthan, West M.P., East M.P., Gujarat, Saurashtra, Konkan, Madhya Maharashtra, Marathwada, Vidarbha, Telangana, and North Interior Karnataka. Among the 29 sub-divisions into which India is divided, those comprising the HIM region have been selected using the following criteria in [50]:

- contiguity of area,
- contribution of monsoon seasonal rainfall to the annual amount,
- inter-correlations of sub-divisional and All-India monsoon rainfall and
- relationships between sub-divisional monsoon rainfall and regional /global circulation parameters.

6.3 Notation and Methodology for Analysis of Spatial Homogeneity

The motivation for this section is to explore whether the variation among sub-divisional rainfall time series significantly differs from what may be expected in different realizations of a white noise process. The confidence level with which deviations of a sub-region from white noise can be identified provides a parameter relevant to the exercise. The F test explained in subsection 4.2.1 is used for this purpose.

The normalized time series of each of the 14 sub-divisional rainfall data is defined as $r_i^{(m)}$ in sub-division m and year i ($i = 1, ..., 120, m = 1, ..., 14$), and the normalized HIM rainfall as $r_i^{(HIM)}$. The mean square deviation of sub-divisional from HIM rainfall is

$$s^2_{r^{(m)}} = \frac{1}{119}\sum_i (r_i^{(m)} - r_i^{(HIM)})^2 \tag{6.1}$$

Similarly, 14 realizations of white noise (120 samples each) are generated using an algorithm of [77], and their normalized time series is denoted by $n_i^{(m)}(i = 1, ..., 120, m = 1, ..., 14)$. The ensemble average of these realizations is then defined as

$$\langle n_i \rangle = \sum_m n_i^{(m)}/14. \tag{6.2}$$

The mean square deviation of each white noise realization from its ensemble average is taken as

$$s_n^2(m) = \frac{1}{119}\sum_i (n_i^{(m)} - \langle n_i \rangle)^2, \tag{6.3}$$

and the ensemble average of these deviations is

$$\langle s_n^2(m)\rangle = \frac{1}{14}\sum_m \frac{1}{119}\sum_i (n_i^{(m)} - \langle n_i\rangle)^2. \tag{6.4}$$

Results and conclusions

The mean square deviation of each sub-divisional rainfall from HIM rainfall is calculated using Eq. 6.1. Similarly an ensemble average of white noise deviations is then calculated using Eq. 6.4 and an average value from 1000 such ensembles is calculated as 1.457, which is denoted as s^2_{noise} and can be taken as a population statistic.

To check the null hypothesis that the two variances are the same, the F-test is formulated by considering the ratio of mean square deviation of each sub-divisional rainfall with that of the calculated average value for white noise, $F = s^2_{noise}/s^2_{r^{(m)}}$. Results are given in table 6.1. It is clear from the table that the mean square deviation of each sub-divisional rainfall is significantly lower than that for white noise. We see that in 13 sub-divisions (Punjab is the exception) this difference lies in the confidence level range $[97.5, 99.9\%)$, hence the 13 sub-divisions form a spatially homogeneous group very largely supporting the definition of [50].

6.3.1 Notation and methodology for analysing spectral homogeneity

We now introduce the notion of spectral homogeneity among a set of M sub-divisions constituting a region. The PSD function, defined as $\hat{r}^{(m)}(\omega_k), m = 1, ..., M, k = 1, ..., K$ (frequency index), for the normalized sub-divisional rainfall time series $r_i^m,\ i = 1, ..., N$ for sub-division m, is first estimated using the Welch technique. The mean square deviation of the sub-divisional PSD from the PSD of regional rainfall is defined as

$$s^2_{\hat{r}^{(m)})} = \frac{1}{K-1}\sum_{\omega_k} (\hat{r}^{(m)}(\omega_k) - \hat{r}(\omega_k))^2 \tag{6.5}$$

S.no	Sub-division	$s^2_{r^{(m)}}$ $m = 1, ..., 14$	$F = s^2_{noise}/s^2_{r^{(m)}}$	Rejection probability(%)
1.	West M.P.	0.37	3.93	> 99.9
2.	East Rajasthan	0.41	3.55	> 99.9
3.	Vidarbha	0.47	3.11	> 99.9
4.	Gujarat	0.49	2.97	> 99.9
5.	West Rajasthan	0.54	2.69	> 99.9
6.	Konkan	0.59	2.46	> 99.9
7.	Madhya Mahrashtra	0.59	2.46	> 99.9
8.	Marathwada	0.63	2.31	> 99.9
9.	Telangana	0.66	2.20	$[99, 99.5)$
10.	Saurashtra	0.75	1.94	$[99, 99.5)$
11.	Haryana	0.79	1.84	$[99, 99.5)$
11.	East M.P.	0.81	1.79	$[97.5, 99)$
13.	North Interior Karnataka	0.81	1.79	$[97.5, 99)$
14.	Punjab	1.09	1.37	$[75.0, 90.0)$

Table 6.1: F-test for the spatial deviations of each sub-divisional rainfall with the calculated value from white noise $s^2_{noise} = 1.457$.

where $\hat{r}(\omega_k)$ is the PSD function of the ensemble average rainfall over the region

$$\langle r_i \rangle(t) = \sum_m r_i^{(m)}(t)/M. \tag{6.6}$$

Similarly, the PSD of M realizations of white noise, each with N =120 samples as rainfall, is estimated and the mean square deviation of the PSD of each white noise realization from the PSD of the ensemble-average (i.e. white) is then taken as

$$s^2_{\hat{n}^{(m)}} = \frac{1}{K-1} \sum_{\omega_k} (\hat{n}^{(m)}(\omega_k) - \hat{n}(\omega_k))^2. \tag{6.7}$$

6.4 Spectrally Homogeneous Indian Monsoon Rainfall as Example

Before presenting a classification of India into spectrally homogeneous regions, it might be instructive to analyse HIM data and examine the homogeneity among the spectra of the 14 sub-divisions.

To do this, the spectral mean square deviation of each sub-divisional rainfall from HIM rainfall is calculated using Eq. 6.5. Similarly an ensemble of 14 white noise deviations (of sample size 120) is then calculated using Eq. 6.8. The average over 1000 such ensembles is found to be 1.82, which is s^2_{noise} and is taken as a population statistic.

We now calculate the F ratio from spectral mean square deviation of each sub-divisional rainfall with that of the averaged value for white noise, $F = s^2_{noise}/s^2_{\hat{x}(r^{(m)})}$, $m = 1, ..., 14$. Results are given in table 6.2. From the table we group those sub-divisions where spectral mean square deviations for rainfall are significantly lower than that of s^2_{noise}. It is found that the 4 sub-divisions Telangana, West M.P., East Rajasthan and Vidarbha form a group with a rejection probability $> 99.5\%$. We shall call this group as the spectrally homogeneous Indian monsoon (SHIM) sub-region. Figure 6.1 shows the SHIM sub-region which represents a contiguous area embedded in the HIM region. We can therefore assert, with very high confidence ($> 99.5\%$), that the four sub-divisions mentioned constitute a strongly spectrally homogeneous sub-region.

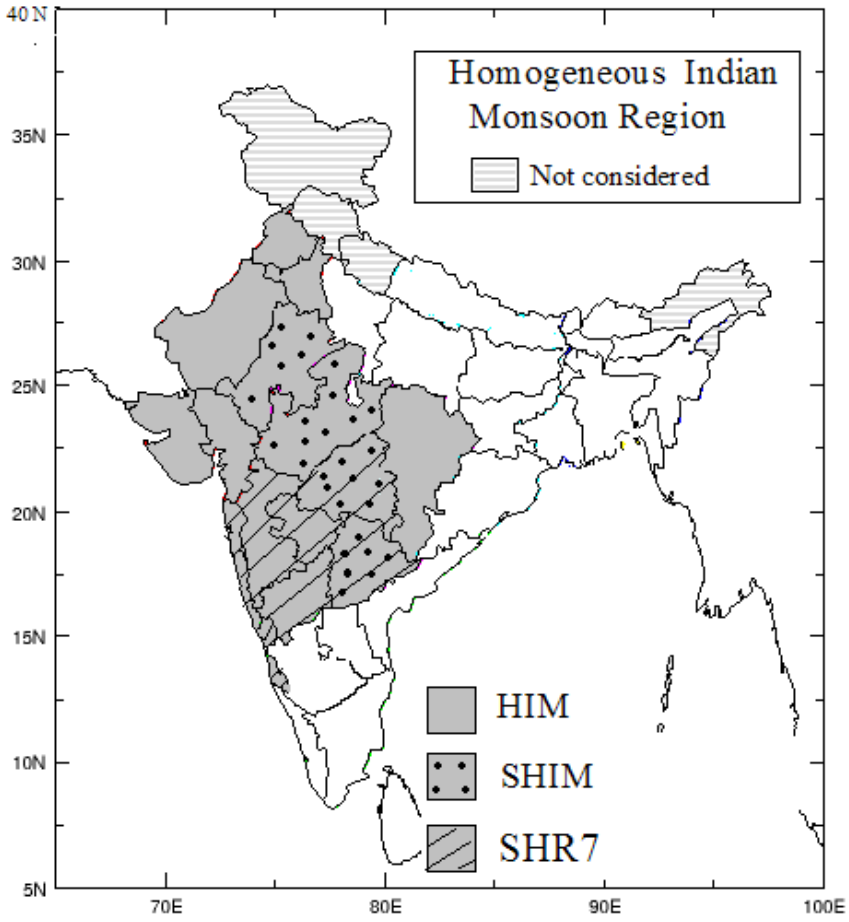

Figure 6.1: Spectrally homogeneous Indian monsoon (SHIM) sub-region within the HIM region.

S.no	Sub-division	$s^2_{\hat{x}(r^{(m)})}$	$F = s^2_{noise}/s^2_{\hat{x}(r^{(m)})}$	Rejection probability(%)
1.	Telangana	0.715	3.09	> 99.9
2.	West M.P.	0.735	3.01	> 99.9
3.	East Rajasthan	0.771	2.87	$[99.5, 99.9)$
4.	Vidarbha	0.860	2.57	$[99.5, 99.9)$
5.	Konkan	0.947	2.33	$[97.5, 99.0)$
6.	East M.P.	1.07	2.05	$[95.0, 97.5)$
7.	West Rajasthan	1.08	2.04	$[95.0, 97.5)$
8.	Marathwada	1.148	1.92	$[95.0, 97.5)$
9.	Haryana	1.160	1.190	$[90.0, 95.0)$
10.	Saurashtra	1.168	1.89	$[90.0, 95.0)$
11.	Gujarat	1.176	1.88	$[90.0, 95.0)$
12.	Madhya Mahrashtra	1.231	1.79	< 90.0
13.	North Interior Karnataka	1.293	1.71	< 90.0
14.	Punjab	1.381	1.60	< 90.0

Table 6.2: F-test formulated from the spectral deviations of sub-divisional rainfall.

6.4.1 Discussion and conclusions

- Standard tests treat homogeneous zone rainfall (HZR) time series as a set of 120 points, whereas they are actually based on an averaging process over numerous other time series (sub-divisions, stations : $O(10)$ to $O(10^2)$).

- It is not possible to assert that sub-division / station time series are drawn from a single population/universe, but they do possess certain structural similarities among themselves.

- From sub-divisional data, with very high confidence ($> 99.5\%$), one can assert: (a) strong spectral homogeneity among four sub-divisions of the set of 14 constituting the HIM region, and (b) distinction from the behavior of white noise.

- It is possible to identify a "Spectrally Homogeneous Indian Monsoon (SHIM)"sub-region within the HIM region defined in [50].

- It appears reasonable to consider SHIM sub-divisional data as quasi-realizations from a spectrally homogeneous ensemble.

6.5 Periodicities Re-visited in SHIM Sub-region

The time series of SHIM rainfall is now prepared by averaging over the time series of the 4 sub-divisions constituting it. The autocorrelation coefficients of SHIM rainfall at various lags is calculated using Eq. 1.1, and the values are listed in table 6.3. As seen in Chapter 4 the autocorrelation coefficient for HIM rainfall was quite low at various lags, for example at lag-1 it was -0.007, therefore the reference spectrum of Eq. 1.13 was close to white. Here for SHIM rainfall it is 0.017 at lag-1, still low; hence the reference spectrum (Eq. 1.13) for SHIM rainfall is also close to white.

As was done for the HIM rainfall time series in Chapter 5, the PSD of SHIM time series is also estimated. Figure 6.2 shows the result of the significance test against the classical reference spectrum (Eq. 1.13).

Lag(y)	Autocorrelation coefficient
1.	0.017
2.	0.212
3.	0.050
4.	-0.090
5.	-0.025
6.	-0.065

Table 6.3: Autocorrelation coefficients of SHIM rainfall at different lags.

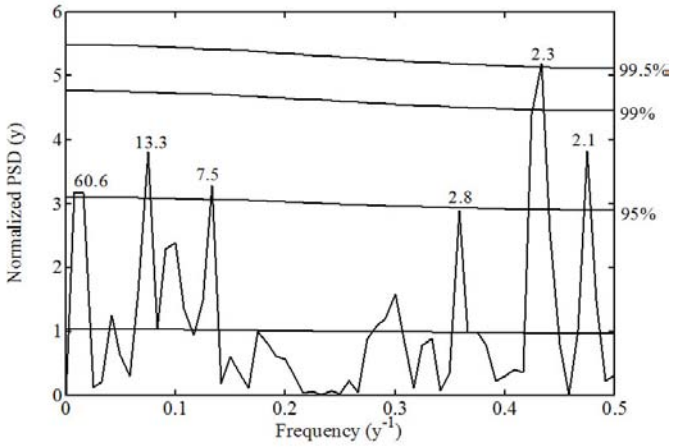

Figure 6.2: Significant periodicities in SHIM rainfall. Periods (y) are marked on each peak.

From figure 6.2 we observe that compared to the HIM spectrum (figure 5.2), the SHIM spectrum shows higher amplitudes at lower frequencies. For example four periods of 2.1, 7.5, 13.3 and 60.6 y are above the 95% confidence line in SHIM, whereas 7.5 y is at 90% and the 13.3 y at 94% in HIM. Also the 2.3 y period in SHIM rainfall (figure 6.2) is at 99.5% confidence whereas it is 99% in the case of HIM rainfall (figure 5.2).

Figure 6.3 shows the estimated PSD of rainfall in each of the 4 sub-divisions

constituting the SHIM sub-region.

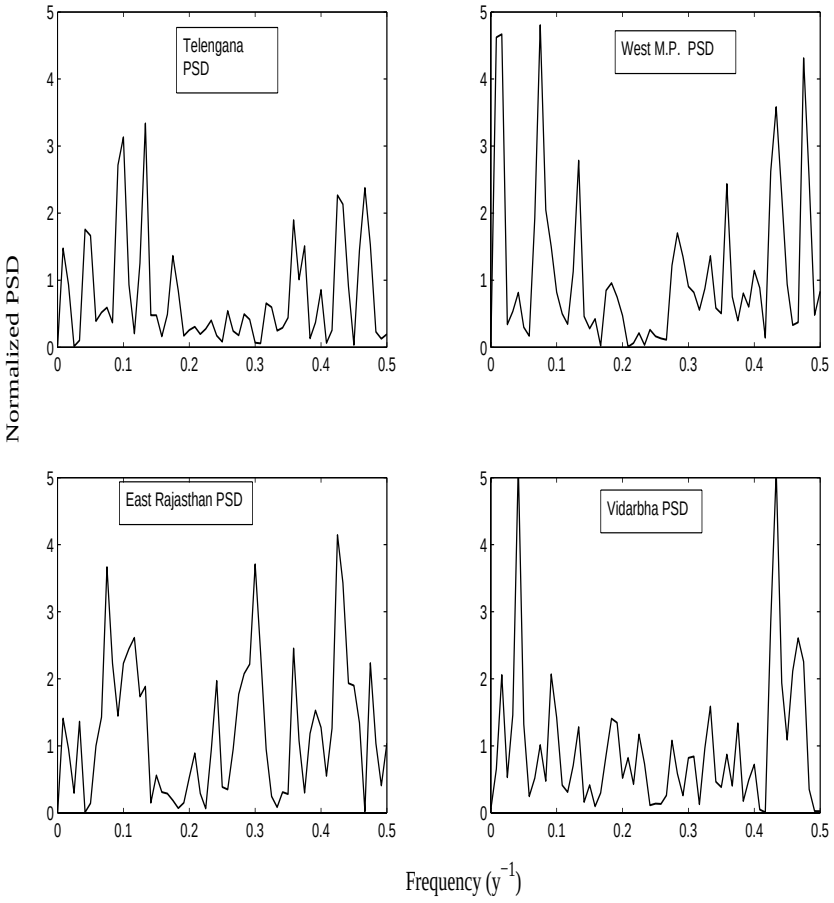

Figure 6.3: Estimated PSD of 4 sub-divisions constituting SHIM.

In Chapter 5 it was shown that 7 out of the 14 sub-divisions constituting the HIM region show a spectral dip around a frequency of 0.25 y^{-1}. However here in the SHIM sub-region we find all 4 sub-divisions showing a spectral dip, three of them around the frequency 0.25 y^{-1} (Fig. 6.3), and the fourth (East Rajasthan) showing it over the slightly lower frequency band 0.13 to 0.26 y^{-1}. Hence we conclude that spectral homogeneity in the SHIM sub-region is supported in several ways. The HIM region covers 55.4% of the total area of country, and the SHIM sub-region 37% of the HIM region, i.e. 20.5% of the whole country.

We now apply the MRA+PSD approach on SHIM rainfall. That is, as was done for HIM rainfall in Chapter 5, the PSD of each of the partially reconstructed time series obtained from MRA of SHIM rainfall at four levels is estimated. The results are shown in figure 6.4. There are 11 significant periodicities in the SHIM rainfall above 99% confidence level, namely 2.3, 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3 and 60.6. y.

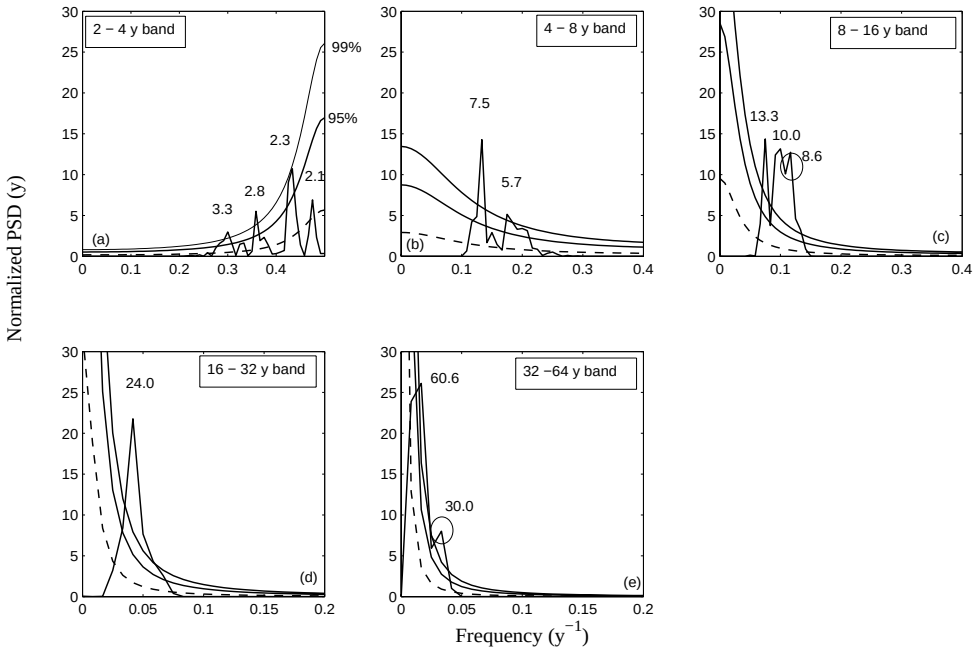

Figure 6.4: Significance test on MRA of SHIM rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: circled periodicities are the new periodicities which were missed out by direct PSD. Significant periods(y) are marked at each peak.

Also as was done for HIM rainfall in Chapter 5, in figure 6.5, the direct PSD of normalized SHIM rainfall is plotted against that of the MRA reconstructed time series (using the spectra of four levels of details and the fourth approximation). (Note of course that the area under both the PSD functions is the same.) The

striking difference is the 8.6 y period, which is missed by the direct PSD.

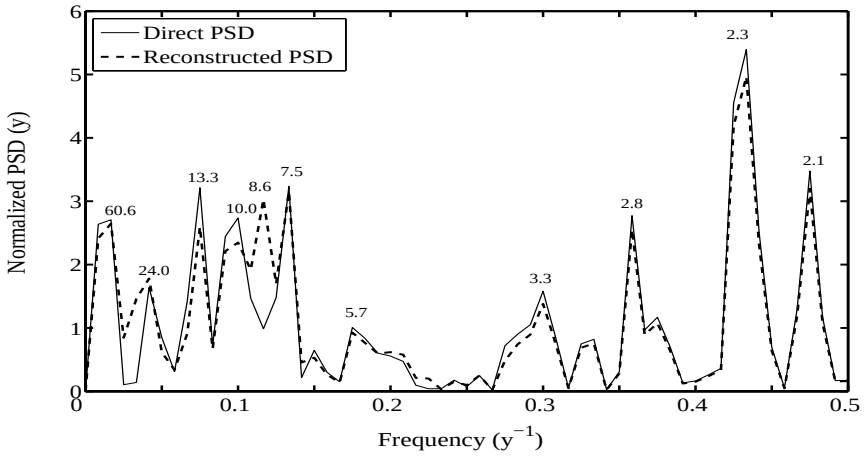

Figure 6.5: The direct PSD of normalized SHIM rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

The same method of MRA+PSD is applied on the 4 sub-divisions that constitute the SHIM sub-region. Results are shown in figures 6.6 –6.9. The significant periodicities above 99% confidence level are listed in table 6.4. In Telangana sub-divisional rainfall MRA+PSD yields 7 significant peaks above 99% confidence level: 2.8, 5.7, 7.5, 8.6, 10.0, 24.0 and 30.3 y. In West M.P. sub-divisional rainfall there are 9 significant peaks above 99% confidence level: 2.8, 3.5, 5.5, 7.5, 8.6, 13.3, 17.2, 30.3 and 60.6 y. In East Rajasthan sub-divisional rainfall there are 7 significant peaks above the 99% confidence level: 3.3, 4.0, 4.7, 7.5, 8.6, 13.3 and 17.2. In Vidarbha sub-divisional rainfall there are 6 significant peaks above the 99% confidence level: 4.5, 5.4, 8.6, 10.9, 24.0 and 60.6 y. Also, the direct PSD of the 4 sub-divisional rainfall are plotted in figures 6.10–6.13 against that of the MRA reconstructed time series (using the spectra of four levels of details and the fourth approximation). It is seen from the figures that the 8.6 y period is captured from the MRA+PSD method but is missed by the direct PSD.

S.no	Sub-division	Periodicities
1.	Telangana	2.8, 5.7, 7.5, 8.6, 10.0, 24.0, 30.3
2.	West M.P.	2.8, 3.5, 5.5, 7.5, 8.6, 13.3, 17.2, 30.3, 60.6
3.	East Rajasthan	3.3, 4.0, 4.7, 7.5, 8.6, 13.3, 17.2
4.	Vidarbha	4.5, 5.4, 8.6, 10.9, 24.0, 60.6
5.	SHIM	2.3, 2.8, 3.3, 5.7, 7.5 8.6, 10.0, 13.3, 24.0, 30.3, 60.6

Table 6.4: Significant periodicities (above 99% confidence line) in the 4 sub-divisional rainfall time series constituting the SHIM sub-region, using MRA+PSD technique

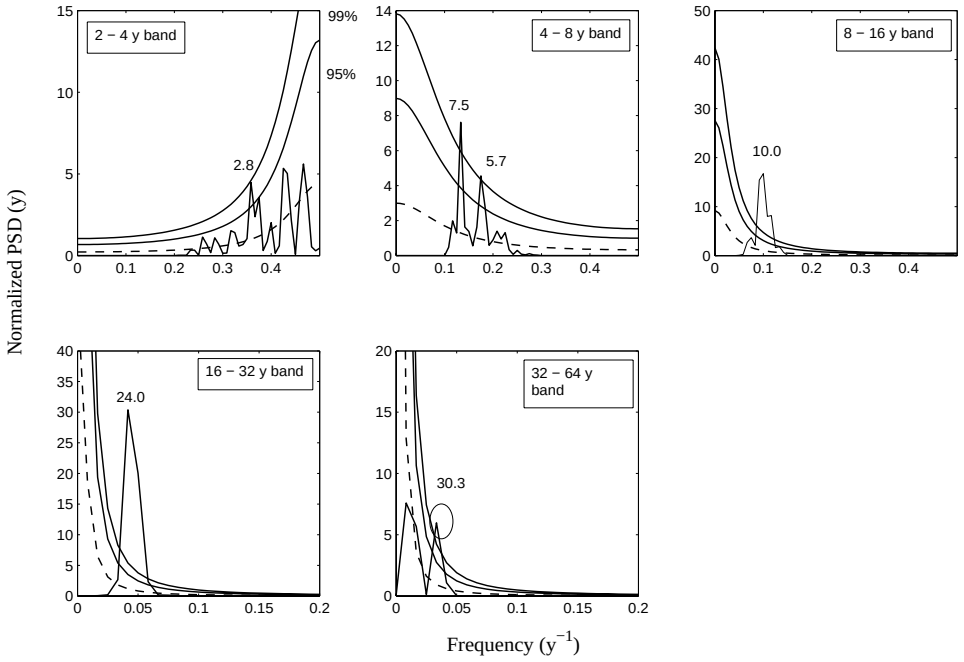

Figure 6.6: Significance test on MRA of Telangana sub-divisional rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: circled periodicities are the new periodicities which were missed out by direct PSD. Significant periods(y) are marked at each peak.

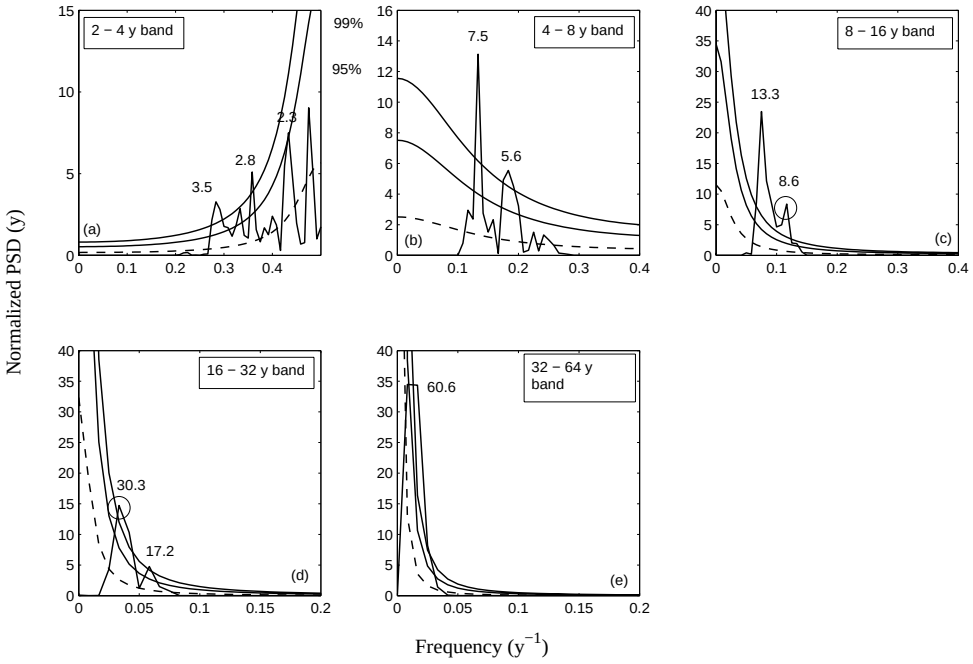

Figure 6.7: Significance test on MRA of West M.P. sub-divisional rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: circled periodicities are the new periodicities which were missed out by direct PSD. Significant periods(y) are marked at each peak.

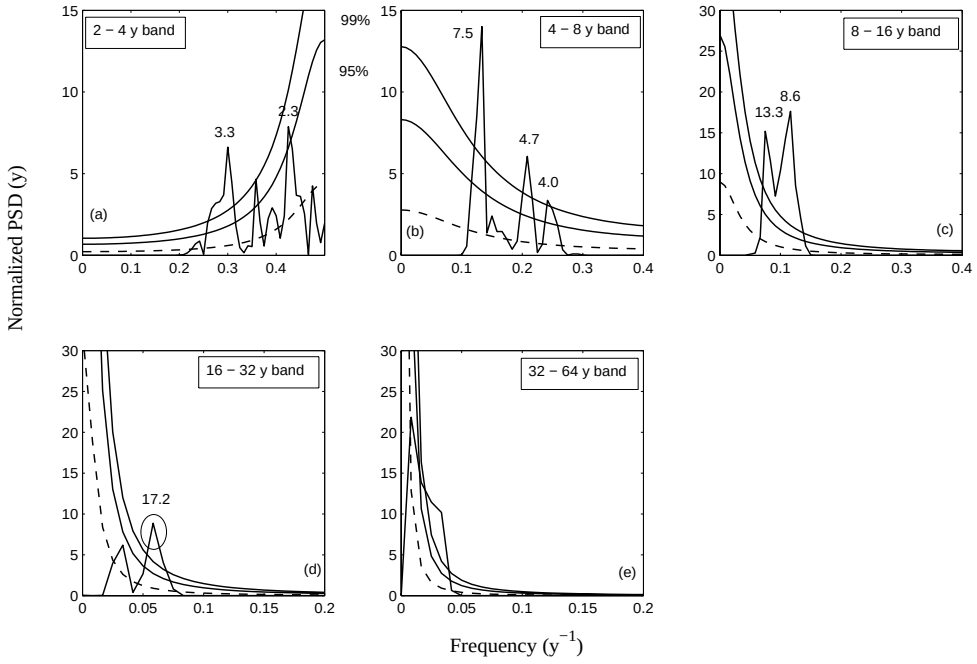

Figure 6.8: Significance test on MRA of East Rajasthan sub-divisional rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: circled periodicities are the new periodicities which were missed out by direct PSD. Significant periods(y) are marked at each peak.

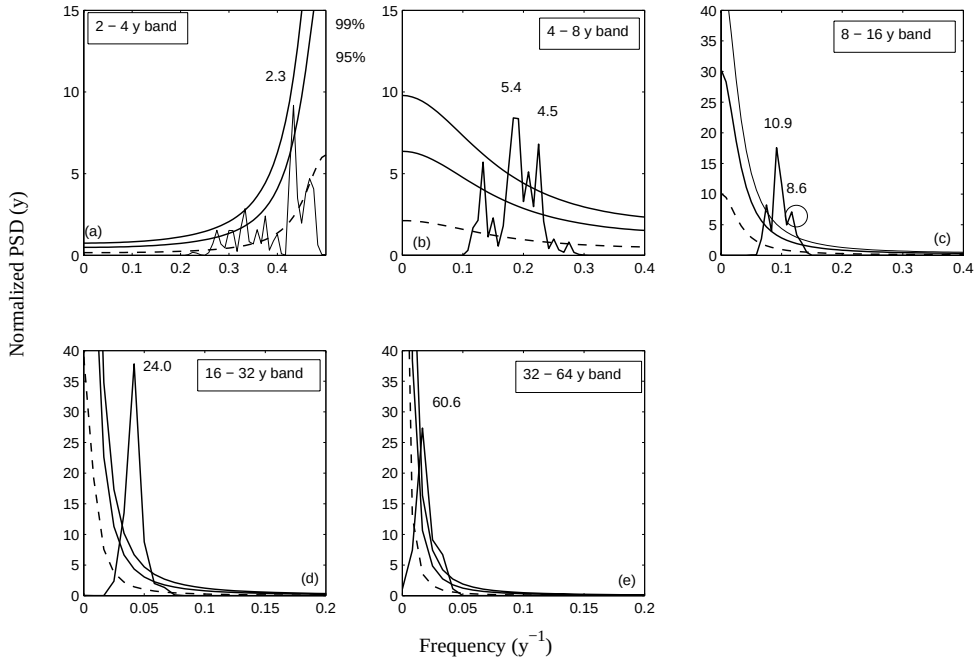

Figure 6.9: Significance test on MRA of Vidarbha sub-divisional rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Significant periods(y) are marked at each peak.

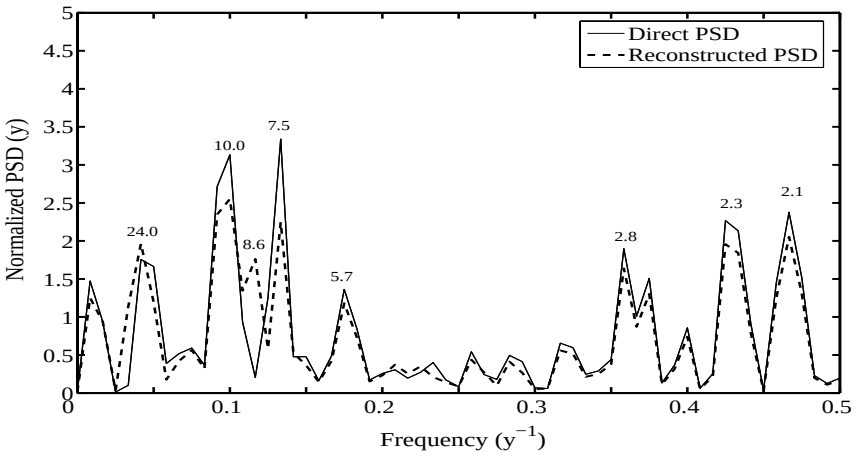

Figure 6.10: The direct PSD of normalized Telangana rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

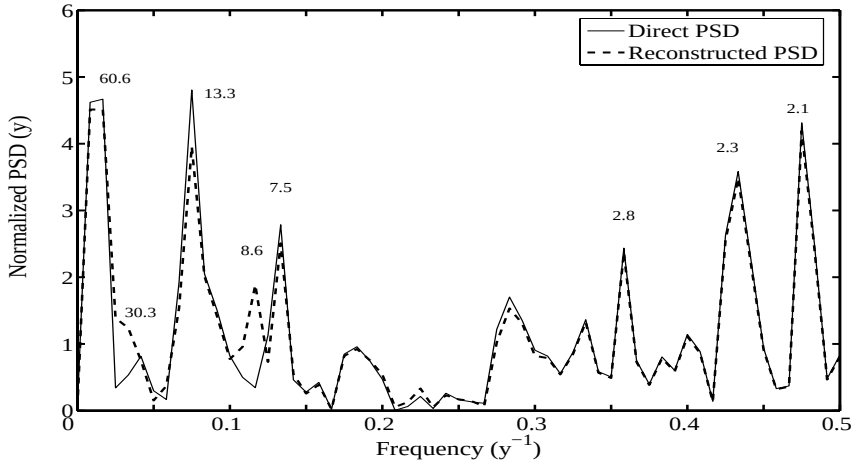

Figure 6.11: The direct PSD of normalized West M.P. rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

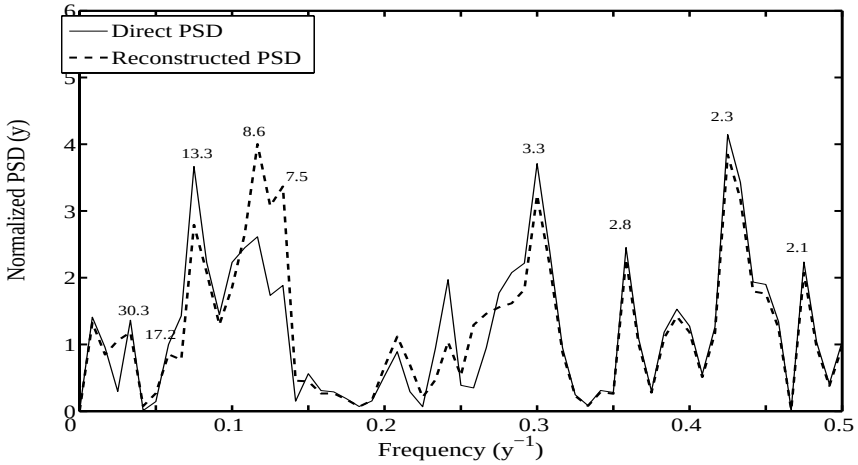

Figure 6.12: The direct PSD of normalized East Rajasthan rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

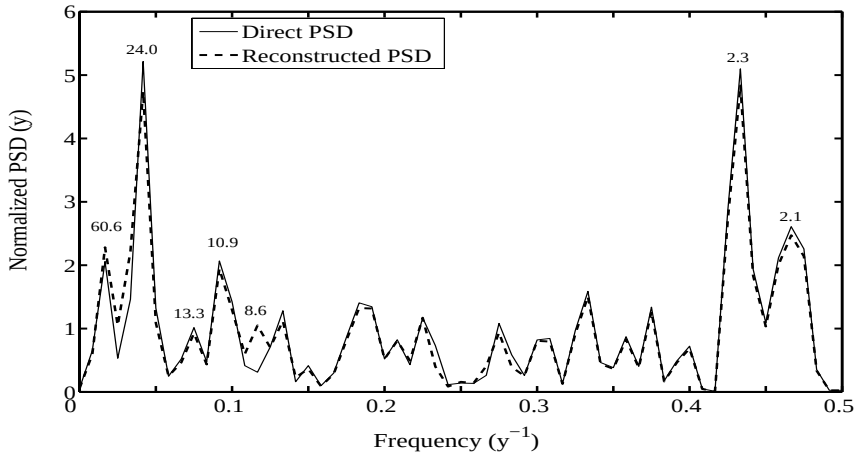

Figure 6.13: The direct PSD of normalized Vidarbha rainfall is plotted against reconstructed PSD obtained from MRA, using Welch technique.

In the next section, we consider the question of common periodicities among the SHIM sub-divisions.

6.6 Multi-variate analysis

It has already been remarked in the introduction that power spectral density analysis for rainfall demands a new approach, as most rainfall time series subjected to such analysis are averages of some kind over a large number of rain-gauge stations. When spectra are evaluated for sub-divisions whose rainfall exhibits significant cross-correlations one question that arises is whether the different sub-divisions have common periodicities, and if so the associated confidence levels. This needs multivariate analysis [81], which is based on the idea that correlated variables must be analysed jointly. When analysing pairs of time series, one also needs to know whether a cross-correlation is statistically significant.

6.6.1 Significance tests and confidence ellipse approach

Confidence ellipses, also called prediction interval ellipses, are an exploratory method of investigating the relationship between two variables by bivariate analysis. They are often used for hypothesis testing and for detecting outliers [82]. A confidence ellipse is analogous to the confidence interval of a univariate distribution, and is defined as the smallest possible (sub-)area in 2-dimensional space in which the common value of the variables analysed should be found with a certain probability (e.g. 95%).

The main aim here is to devise a new method to test the significance levels of common periodicities (if any) among sub-divisional rainfall time series using the confidence ellipse technique. Before doing this we validate the method by test cases that mimic the time series of our interest, namely Indian rainfall. It is to be noted that the Hotelling T^2 statistic [83] used here to find the significance levels of a common periodicity is more commonly used for problems related to the study of means [84], and is a multivariate generalization of the univariate Student's t-test.

To assess the significance of a common period p that may be present in the two time series, say x_i and y_i, $i = 1, ..., N$, we propose the following procedure.

- We test the null hypothesis H_0: there is no common period between the two given time series.

- Normalize x_i, y_i with zero mean and unit standard deviation.

- Generate sine and cosine time series of size $N = 120$ and period p :

$$a_i = \sin(2\pi i/p), \quad b_i = \cos(2\pi i/p), \quad i = 1, ..., N. \tag{6.8}$$

- Normalize a_i, b_i with zero mean and unit standard deviation.

- Calculate

$$X_1 = \sum_i \frac{a_i x_i}{N}; \; Y_1 = \sum_i \frac{a_i y_i}{N} \tag{6.9}$$

$$X_2 = \sum_i \frac{b_i x_i}{N}; \; Y_2 = \sum_i \frac{b_i y_i}{N} \tag{6.10}$$

- Compute the Hotelling T^2 [83]

$$T^2 = N \; [X_1 \; X_2 \; Y_1 \; Y_2] \; \Sigma^{-1} [X_1 \; X_2 \; Y_1 \; Y_2]^T \tag{6.11}$$

where the covariance matrix with four degrees of freedom is

$$\Sigma = \begin{bmatrix} 1.0 & 0.0 & r & 0.0 \\ 0.0 & 1.0 & 0.0 & r \\ r & 0.0 & 1.0 & 0.0 \\ 0.0 & r & 0.0 & 1.0 \end{bmatrix} \tag{6.12}$$

where $r = \sum_i (x_i y_i)/N$ is the correlation between the two time series x_i and y_i.

- At $(1 - \alpha)\%$ confidence level, the Hotelling test consists of accepting the hypothesis H_0 if, and only if,

$$T^2 \leq \chi^2(\gamma, 1 - \alpha) \tag{6.13}$$

where $\chi^2(\gamma, 1 - \alpha)$ is the chi-square distribution with γ degrees of freedom.

- The related $(1-\alpha)\%$ confidence region is the set of points that simultaneously satisfy

$$[X_1\ X_2\ Y_1\ Y_2]\ \Sigma^{-1}[X_1\ X_2\ Y_1\ Y_2]^T \leq \chi^2(\gamma, 1-\alpha) \tag{6.14}$$

  which is the interior of a 4-dimensional ellipsoid with $\gamma = 4$ degrees of freedom.

- To construct ellipses at various confidence levels, a subroutine is written in *Matlab*, which uses the chi-square cumulative distribution function with γ degrees of freedom to obtain all values of (X_1, Y_1) or (X_2, Y_2) that lie on the confidence ellipse for the given confidence level.

- The commonality of periods in the two time series can then be seen in terms of the points within the elliptic region of given confidence level.

Next we construct the test cases. For this purpose, two cross-correlated time series of Gaussian white noise (each of 120 samples, the same as in rainfall time series), with a specified correlation coefficient, are generated using the algorithm of [77]. Two realizations of such white noise, with a cross correlation of 0.32, are normalized and shown in figure 6.14. Their PSD's are estimated using the Welch technique [7] and are shown in figure 6.15.

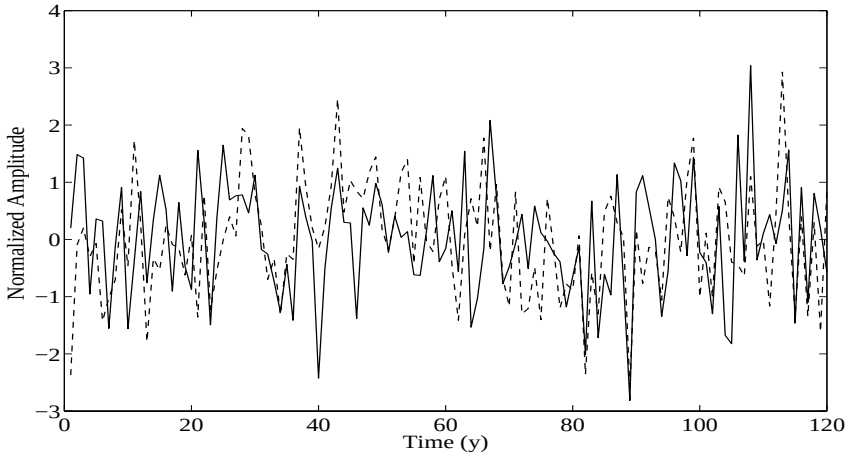

Figure 6.14: Two cross-correlated realizations of white noise (correlation coefficient is 0.32), each of sample 120.

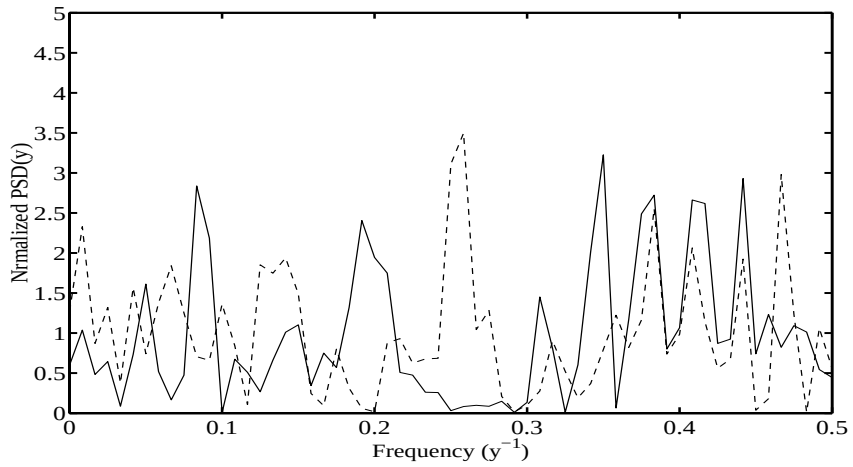

Figure 6.15: Estimated PSD's of two cross-correlated realizations of white noise.

Now a sine wave of period 10 y is imposed on each noise signal, but with an amplitude 0.4 in one case and 0.8 in the other. The two signals so obtained are then renormalized with zero mean and unit standard deviation. The estimated PSD's of these new time series are shown in figure 6.16. It is clear from the figure that the two realizations of white noise now have a common period of 10 y.

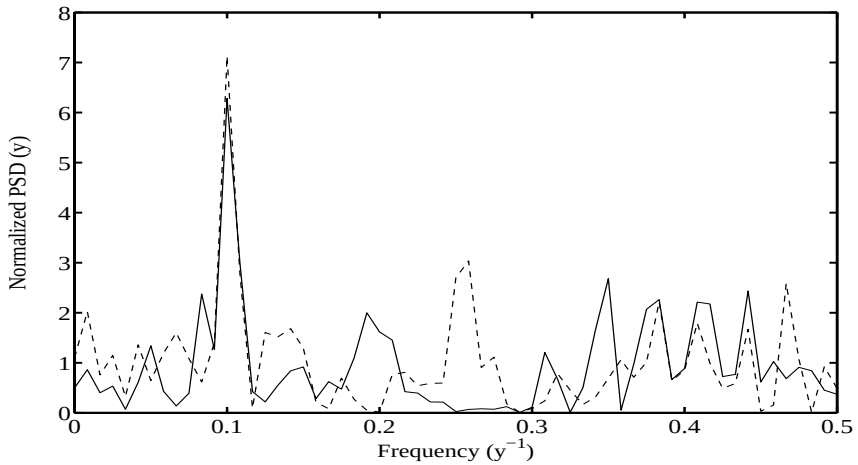

Figure 6.16: Estimated PSD's of two cross-correlated realizations of white noise with a sine wave of period 10 y imposed on them.

To check the significance level of the common period 10 y, we first compute

(X_1, Y_1) and (X_2, Y_2) from Eqs 6.10 and 6.11 respectively. The value of the T^2 statistic is then calculated as 29.62 using Eq. 6.12, which is greater than $\chi^2(0.9999, 4) = 23.5$. Hence we reject the null hypothesis and the common periodicity 10 y is 99.99% significant. Figure 6.17 shows the ellipses in the X_1Y_1 plane for various confidence levels. The point indicated by a star (*) shows the position of the variable (X_1, Y_1), which corresponds to null hypothesis rejection with $> 99.99\%$ confidence level for common period 10 y.

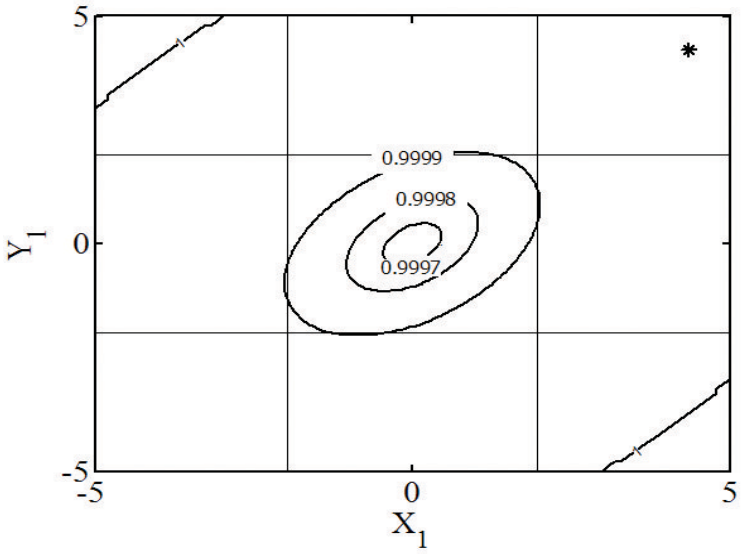

Figure 6.17: Ellipses at various confidence levels representing common period (10 y) between the two cross-correlated white noise realizations. The point indicated by a star (*) shows the position of the variable (X_1, Y_1), which corresponds to null hypothesis rejection with 99.99% confidence level for the period.

If the cross-correlations are not high, or if the amplitudes of the imposed sine waves are low (e.g. 0.04, 0.08), then the confidence levels for common periodicity are not generally high for all realizations of cross-correlated white noise.

6.6.2 Ellipse construction in rainfall

To show how the confidence ellipse methodology can be used to find the significance levels of common periodicities among sub-divisions constituting SHIM sub-region, we consider as example the normalized time series of Telangana and West M.P., two SHIM sub-divisions. Their computed correlation coefficient is 0.35, the error band being [0.18, 0.50] at 99.99% confidence level. The common periodicities among the two sub-divisional rainfall time series are 2.1, 2.3, 2.8, and 7.5 y (see table 5.2).

To check the significance of any common period, say $p = 7.5$ y, we first compute (X_1, Y_1) and (X_2, Y_2) from Eqs. 6.10 and 6.11 respectively. The T^2 statistic value is then calculated as 14.9 using Eq. 6.12, where

$$\Sigma = \begin{bmatrix} 1.0 & 0.0 & 0.35 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.35 \\ 0.35 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.35 & 0.0 & 1.0 \end{bmatrix}. \tag{6.15}$$

The null hypothesis that there is no common period between the two time series is rejected because the calculated value of T^2 is greater than the tabulated value of $\chi^2(0.995, 4)$, which is 14.0. Figure 6.18 shows the ellipses for various confidence levels. The point indicated by a star (*) shows the position of the rainfall variable (X_1, Y_1), which corresponds to null hypothesis rejection with 99.5% confidence level for the period 7.5 y.

Similarly the null hypothesis is rejected with confidence level 94.2% and 80% for periods 2.3 and 2.8 y respectively. Using the same procedure, the null hypothesis rejection probabilities for all periods in sub-divisional rainfall in SHIM sub-region are displayed in table 5. We note seven significant periodicities, five at confidence levels exceeding 95% (respectively 2.1 y at 97.8%, 2.3 y at 95.7%, 2.8 at 98.4%, 7.5 at 99.5%, 24.0 at 99.4%), and 10.9 at 90.5 % and 60.6 at 93.2% confidence level.

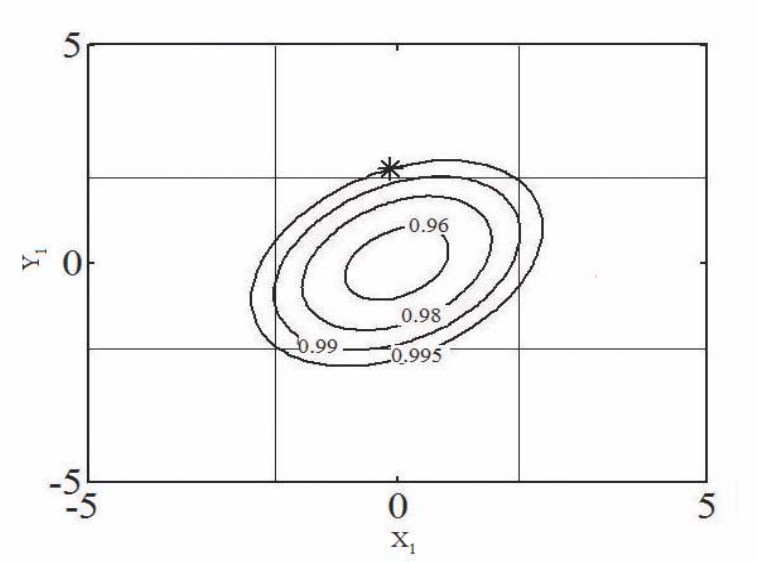

Figure 6.18: Ellipses at various confidence levels representing common period (7.5 y) between Telangana and West M.P. rainfall. The point indicated by a star (*) shows the position of the rainfall variable (X_1, Y_1), which corresponds to null hypothesis rejection with 99.5% confidence level for the period.

Table 6.5: Significance levels (quoted in brackets) for common periodicities in the rainfall time series among four sub-divisions constituting the SHIM sub-region, using confidence ellipses. Periods with maximum confidence levels are in bold.

Sub-division pair	Telangana	West M.P.	East Rajasthan
Vidarbha	2.1 (32.8%)	**2.3 (95.7%)**	2.8 (90.0%)
	2.3 (73.1%)	2.8 (75.6%)	7.5 (92.5 %)
	7.5 (96%)	7.5 (92.9%)	13.3 (66.8%)
	10.9 (90.5%)	**24.0 (99.4%)**	
	24.0 (99.1%)	**60.6 (93.2%)**	
East Rajasthan	2.3 (52.3%)	**2.1 (97.8%)**	
	2.8 (98.4%)	2.8 (82.2%)	
	7.5 (99.5%)	7.5(90.5%)	
		13.3 (75.1%)	
West M.P	2.3 (80%)		
	2.8(96%)		
	7.5 (99.5%)		

6.7 Conclusions

- The homogeneous regions identified by Parthasarthy *et al.* (1993) are based on observations which take into account the spatial characteristics of sub-divisional monsoon rainfall but ignoring possible spectral similarities or dissimilarities among these sub-divisions.

- Tested against the classical reference spectrum of Eq. 1.13, HIM rainfall time series over the time period 1871-1990 shows one 99% significant period of 2.3 y and a 95% significant period at 2.8 y. However, by introducing the spectrally homogeneous sub-region SHIM, we have shown that there is a 99.5% significant period of 2.3 y, along with four other significant peaks above 95% confidence at 2.1, 7.5, 13.3 and 60.6 y and a 2.8 y period at $> 94.8\%$..

- Using confidence ellipse technique a new method to detect significant common periodicities in the cross-correlated sub-divisional rainfall is proposed. We report here seven significant periodicities in SHIM rainfall, five at confidence levels exceeding 95% (respectively 2.1 y at 97.8%, 2.3 y at 95.7%, 2.8 at 98.4%, 7.5 at 99.5%, 24.0 at 99.4%), and 10.9 at 90.5 % and 60.6 at 93.2% confidence level.

- It is to be noted that a common period may be present in all the sub-divisions constituting a region but the significance of their commonalty can not be established using the classical method (as described by Eq.1.13). However it is found that the proposed method of confidence ellipse is a more informative and powerful tool for testing confidence levels of common periodicities in correlated time series.

- We found that the classical test of checking confidence level of a spectral peak is inadequate when the amplitude of longer periods is low. The proposed confidence ellipse technique however shows high confidence levels even for relatively low amplitudes when the cross-correlation is high. Hence the present method is less amplitude-sensitive than other available methods.

Chapter 7

Spectrally Homogeneous Monsoon Regions

[1]

7.1 Introduction

As mentioned in Chapter 2, the five homogeneous regions identified by Parthasarathy *et al.* (1993) are: North East India(NEI), Central North East India (CNEI), Peninsular India (PENSI), West Central India (WCI) and North West India (NWI).

	NEI	CNEI	PENSI	WCI	NWI
Lag-1	0.09	-0.04	-0.01	0.04	-0.04
Lag-2	-0.04	-0.05	-0.19	0.20	0.04
Lag-3	0.11	0.06	0.02	0.07	0.09
Lag-4	0.12	0.02	-0.08	-0.09	-0.08

Table 7.1: Auto-correlation coefficient at three lags for homogeneous regions during the time period 1871-1990.

[1]Some of the material in this chapter is to be communicated, Azad and Narasimha (2008).

The time series for these homogeneous regions (obtained from Parthasarathy *et al.* 1995) are displayed in figure 7.1. Also given are their auto-correlation coefficient values in table 7.1.

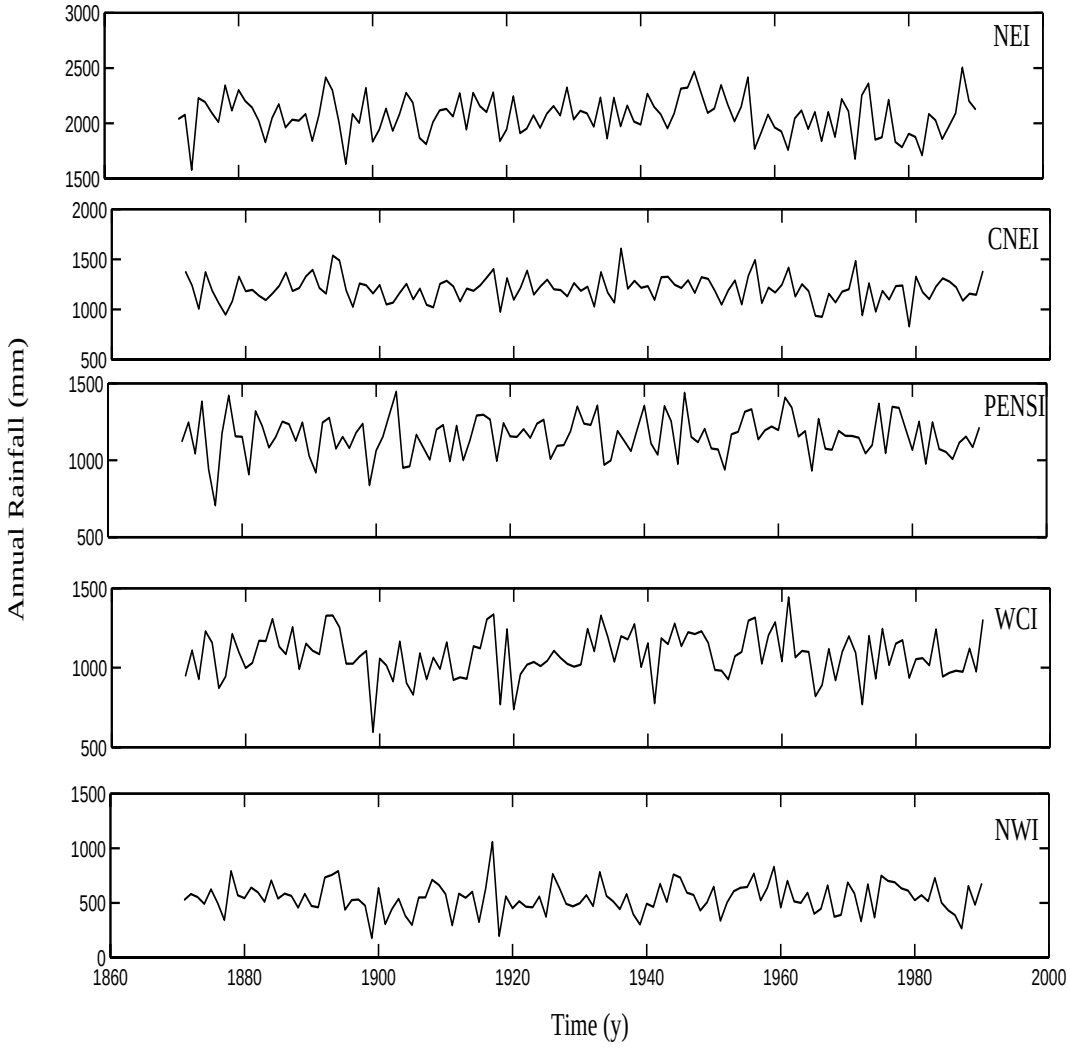

Figure 7.1: Time series of 5 homogeneous regions.

7.2 MRA of Homogeneous Regions

The MRA decomposition of NEI, CNEI, PENSI, WCI and NWI rainfall at four details plus the fourth approximation are obtained by following the MRA procedure adopted for the HIM rainfall in Chapter 4. The partially reconstructed time series from 1871-1990 are shown in figures 7.2 to 7.6.

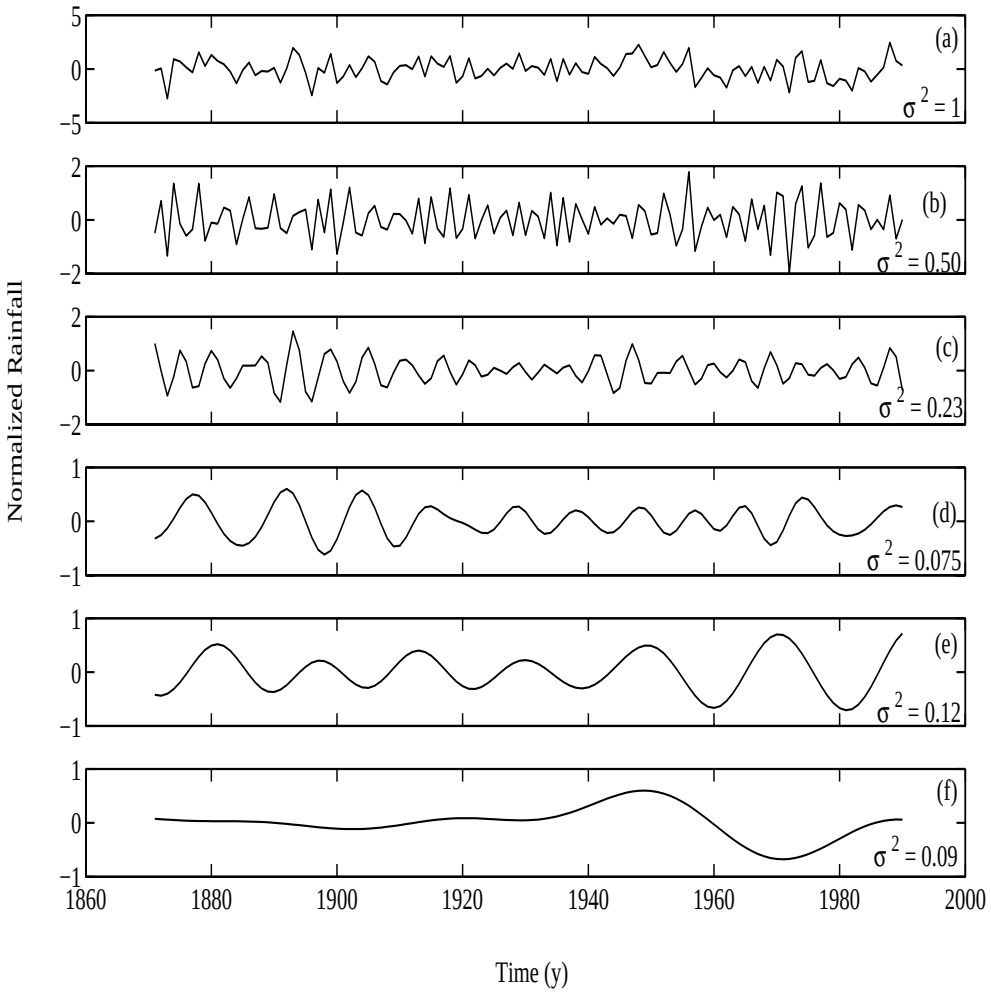

Figure 7.2: MRA of NEI rainfall data. (a) Normalized NEI time series from 1871-1990; (b) 1st reconstructed detail; (c) 2nd reconstructed detail; (d) 3rd reconstructed detail; (e) 4th reconstructed detail; (f) 4th reconstructed approximation.

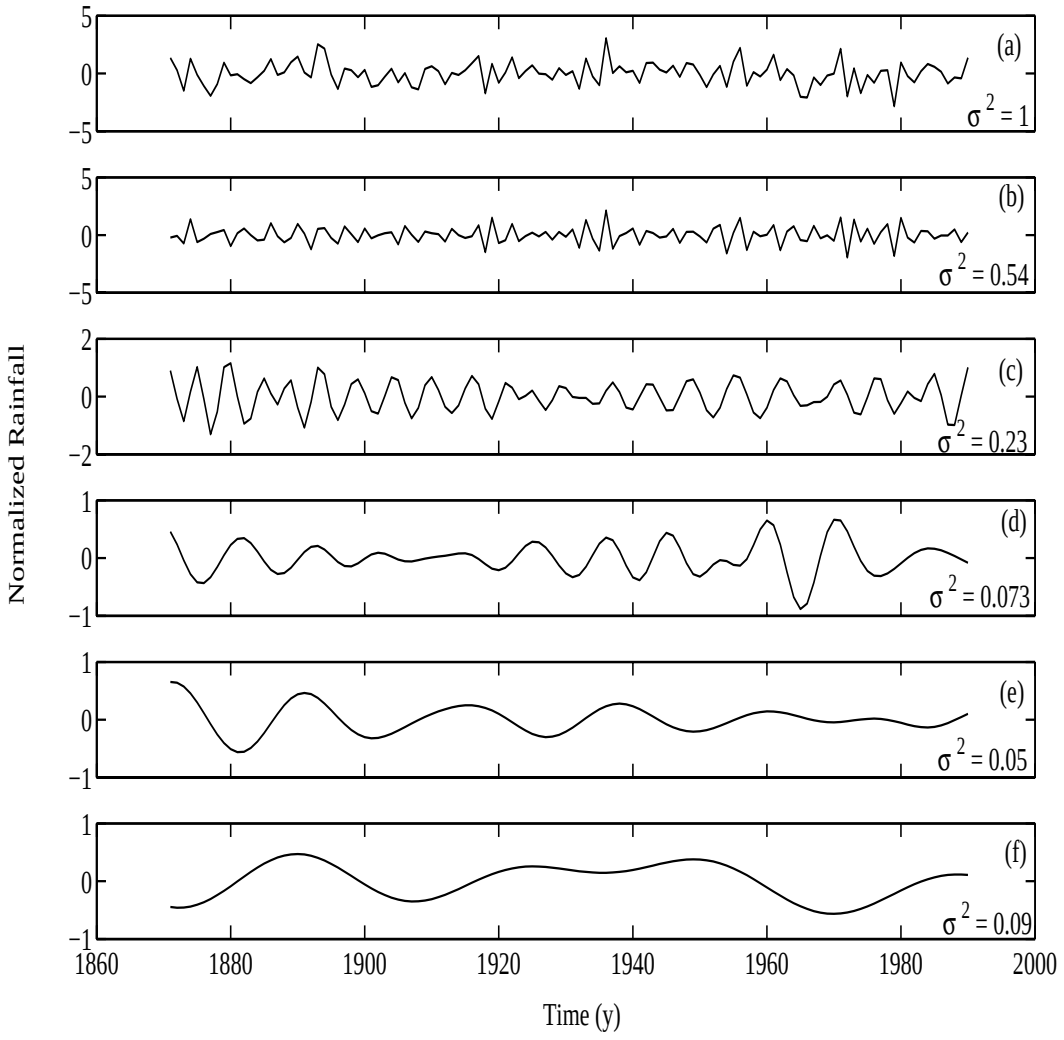

Figure 7.3: MRA of CNEI rainfall data. (a) Normalized CNEI time series from 1871-1990; (b) 1st reconstructed detail; (c) 2nd reconstructed detail; (d) 3rd reconstructed detail; (e) 4th reconstructed detail; (f) 4th reconstructed approximation.

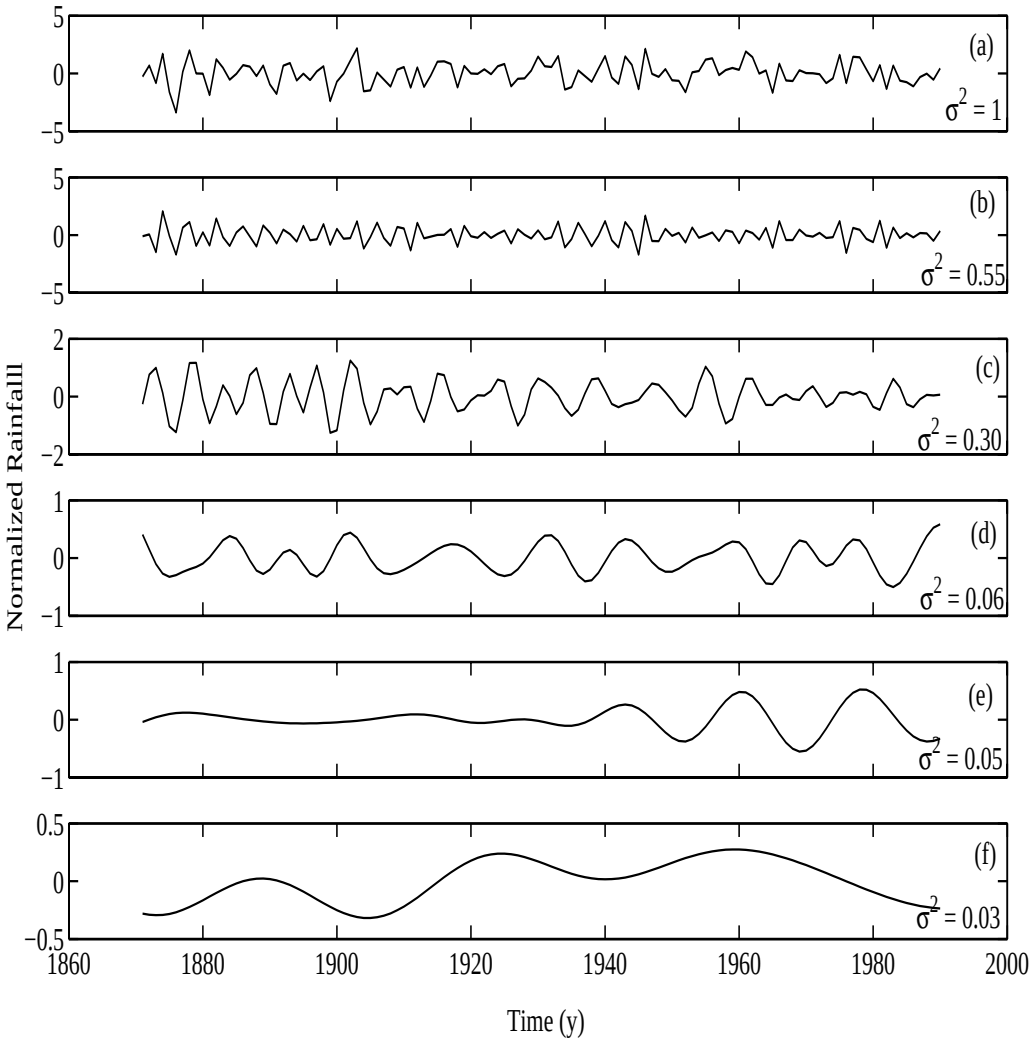

Figure 7.4: MRA of PENSI rainfall data. (a) Normalized PENSI time series from 1871-1990; (b) 1st reconstructed detail; (c) 2nd reconstructed detail; (d) 3rd reconstructed detail; (e) 4th reconstructed detail; (f) 4th reconstructed approximation.

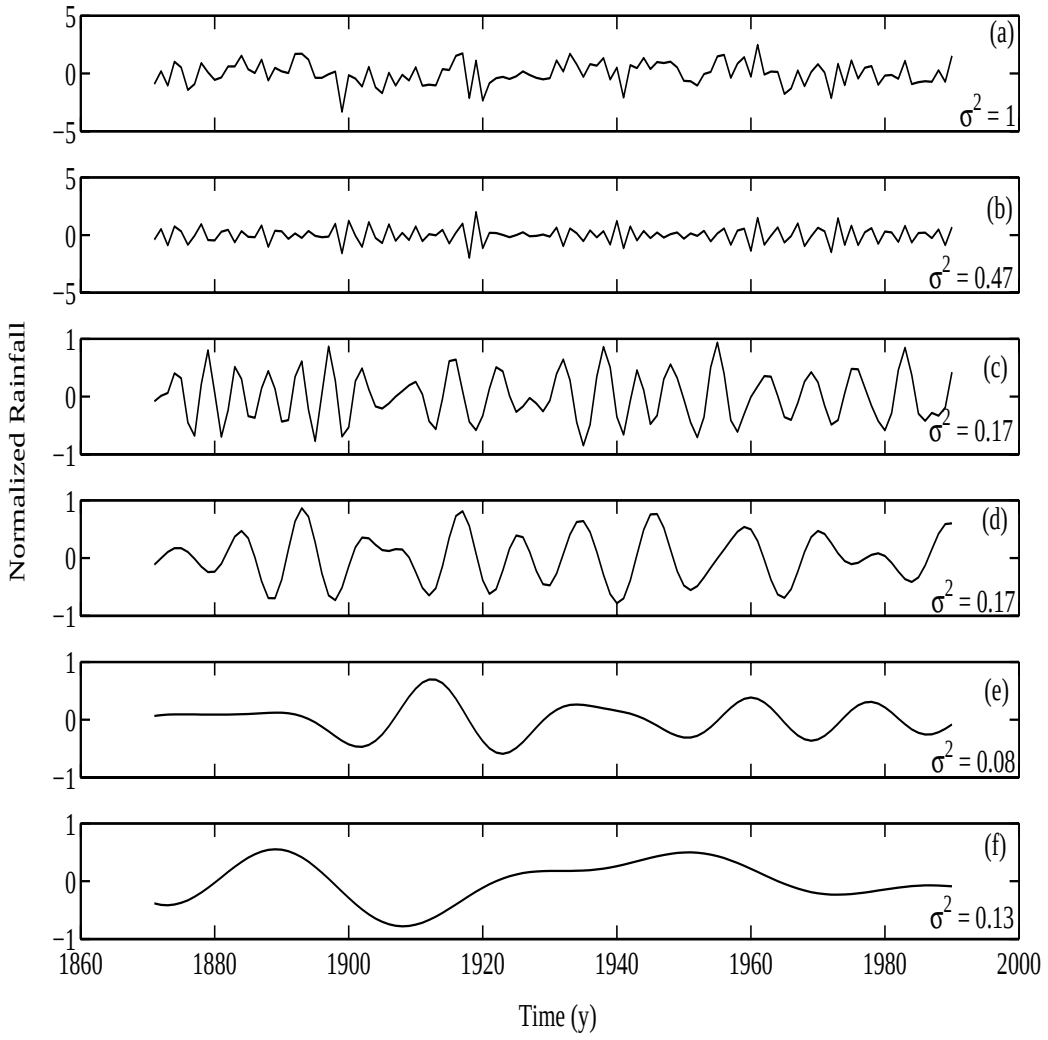

Figure 7.5: MRA of WCI rainfall data. (a) Normalized WCI time series from 1871-1990; (b) 1st reconstructed detail; (c) 2nd reconstructed detail; (d) 3rd reconstructed detail; (e) 4th reconstructed detail; (f) 4th reconstructed approximation.

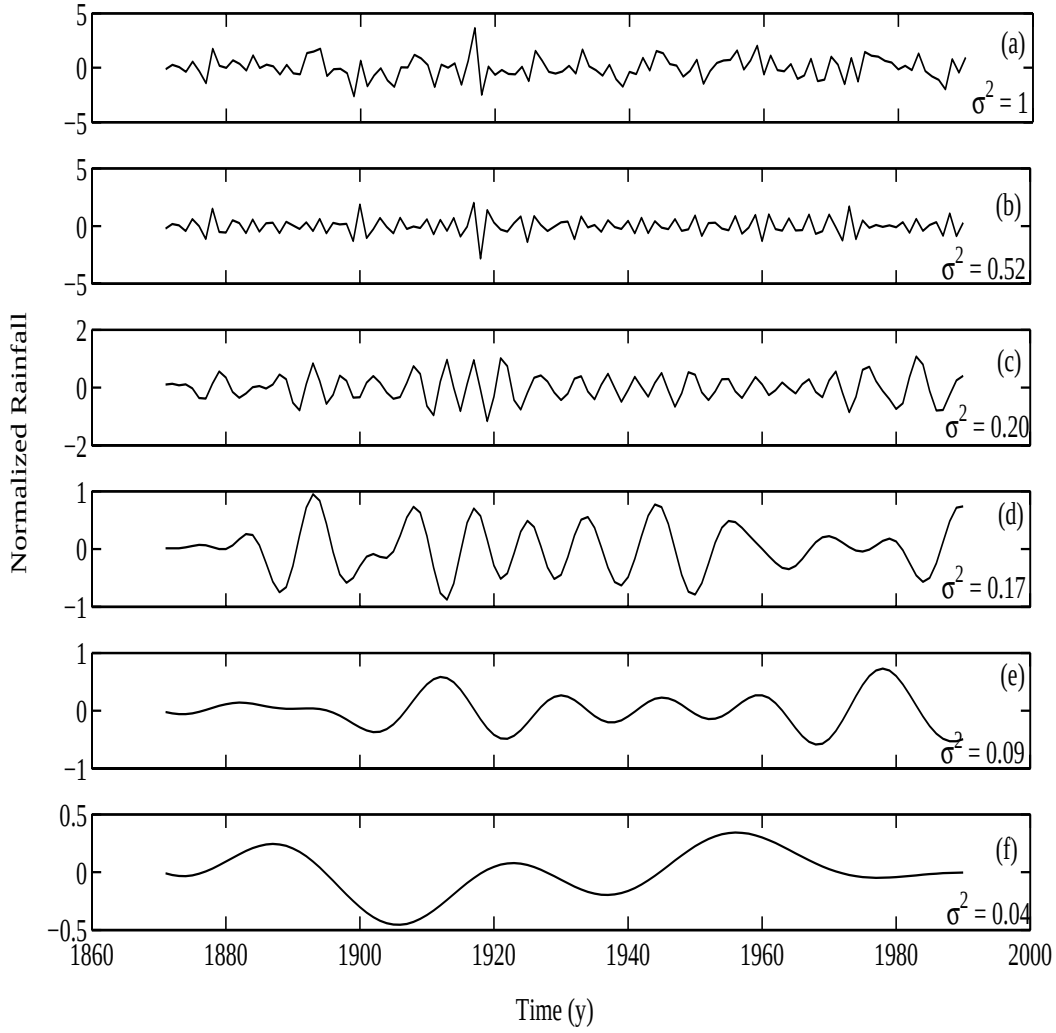

Figure 7.6: MRA of NWI rainfall data. (a) Normalized NWI time series from 1871-1990; (b) 1st reconstructed detail; (c) 2nd reconstructed detail; (d) 3rd reconstructed detail; (e) 4th reconstructed detail; (f) 4th reconstructed approximation.

The contribution of variance at different scales for all the homogeneous regions is shown in figure 7.7. It is seen that 50% of the variance in all the regions is from the band 2-4 y.

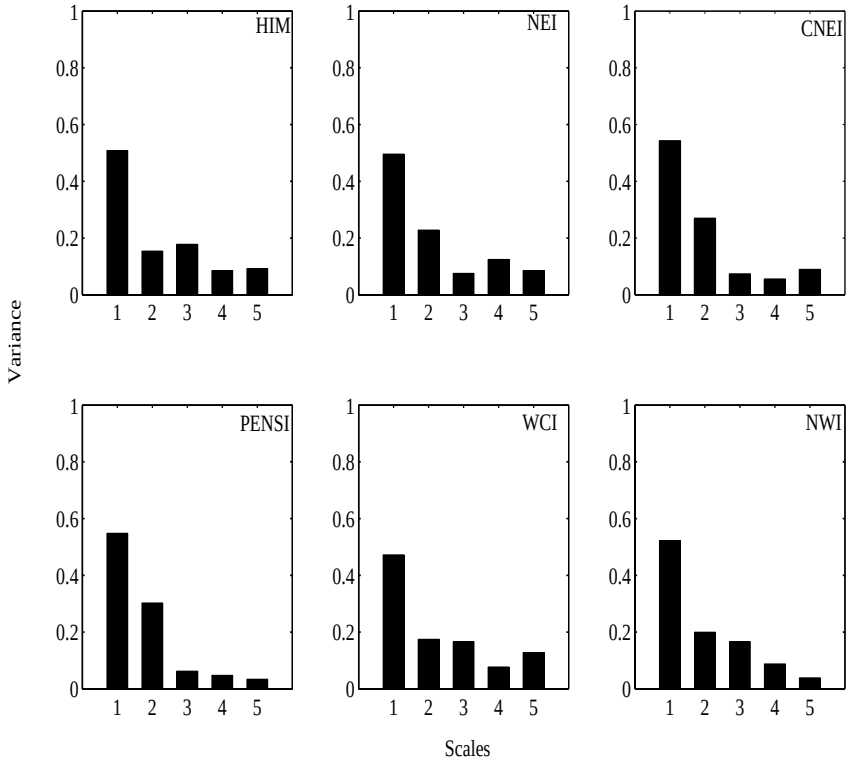

Figure 7.7: Contribution of variance at different scales for homogeneous regions: scale 1 corresponds to 1st partially reconstructed detail; scale 2 corresponds to partially reconstructed 2nd detail, upto 4 details and scale 5 corresponds to 4th partially reconstructed approximation.

7.3 Spectral Analysis of Homogenous Regions

As was done for the HIM rainfall time series in Chapter 5, the PSD functions for all the homogeneous regions are estimated and the results of significance tests are shown in figure 7.8. It is clear from the figure that using the classical significance test (Eq. 1.13), only one period at 7.5 y in the PENSI region is seen to stand above the 99% confidence level. Above the 95% confidence level, CNEI shows a period of 2.8 y, PENSI 2.8 and 7.5 y, WCI 2.1, 2.3, 7.5, 60.6 y and NWI shows 2.3, 2.8 and 3.3 y.

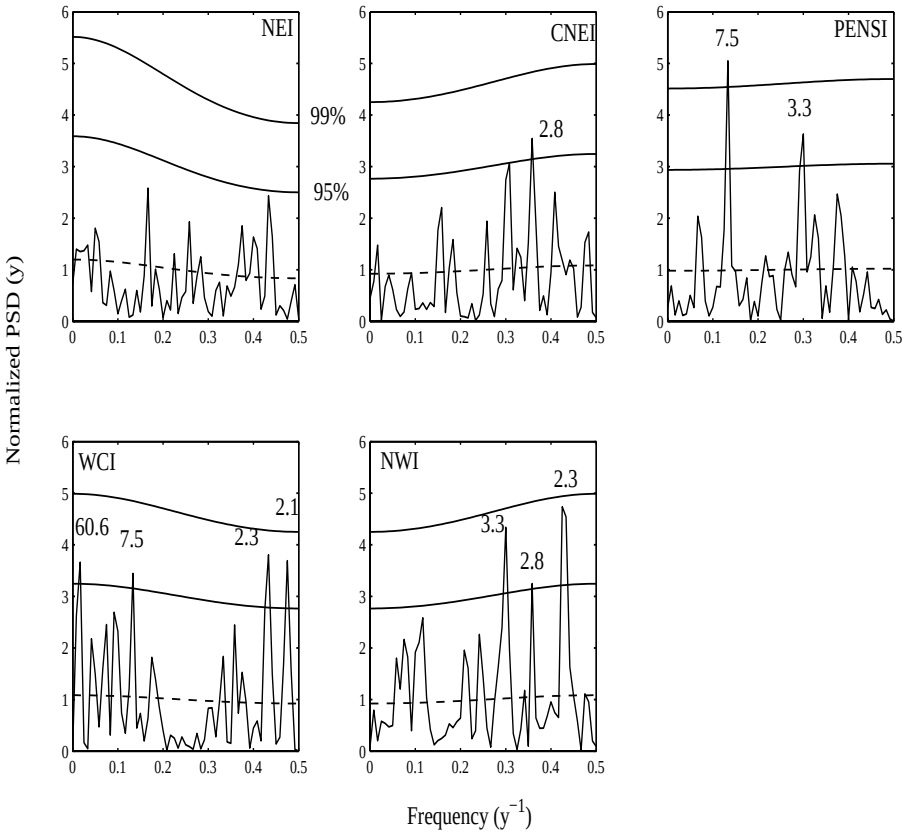

Figure 7.8: The estimated PSD's for 5 homogeneous regions and the reference spectrum (dotted line) with different confidence levels. Significant periods(y) are marked at respective peaks.

7.4 Spectral Analysis on MRA of Homogeneous Regions

In this section spectral analysis on MRA decomposition of all the homogeneous regions is performed and the results are reported. The objective is to find the closely spaced frequencies with a combination of wavelet and spectral analysis as was done for HIM rainfall in Chapter 4.

7.4.1 Results on NEI rainfall

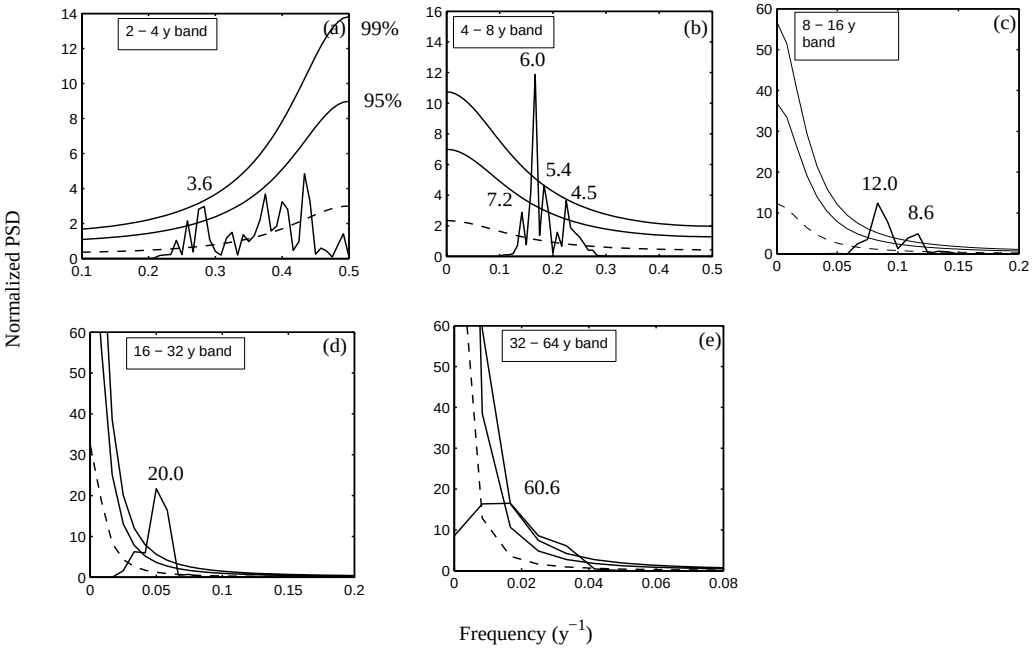

Figure 7.9: Significance test on MRA of NEI rainfall (a) 1st reconstructed detail; (b) 2nd reconstructed detail; (c) 3rd reconstructed detail; (d) 4th reconstructed detail; (e) 4th reconstructed approximation. Note: Significant periods(y) are marked at respective peaks. Reference spectrum is represented by dotted lines with different confidence levels marked.

The PSD of each of the partially reconstructed time series obtained from MRA of NEI rainfall at four levels is estimated. Results are shown in figure 7.9, MRA here does not capture any closely spaced frequencies that are missed by the direct PSD (Fig. 7.8). However PSD on MRA captures 5 periods above 99% confidence level whereas there are no significant peaks in the direct PSD above 95% confidence level. These significant periods are: 6.0, 8.6, 12.0, 20.0 and 60.6 y. Also periods obtained from the EMD technique and the PSD on MRA are compared in table 7.2. As was explained for HIM rainfall in Chapter 4, periods obtained from EMD and zero crossing of NEI rainfall are some averages of the ones obtained from PSD on MRA.

Level	zero crossings on EMD	PSD on MRA
1.	3.08	2.1, 2.3, 2.5, 2.8, 3.0, 3.6
2.	6.0	4.5, 5.4, 6.0, 7.2
3.	10.9	8.6, 12.0
4.	30.0	20.0
5.	40.0	60.6

Table 7.2: Periods (y) obtained by, Empirical mode decomposition and PSD on Multiresolution analysis of NEI rainfall.

Level	zero crossings on EMD	PSD on MRA
1.	3.08	2.1, 2.3, 2.8, 3.3, 4.0
2.	6.32	4.1, 5.5, 6.6
3.	10.9	8.6, 10.9
4.	24.0	24.0
5.	60.0	30.3, 60.6

Table 7.3: Periods (y) obtained by, Empirical mode decomposition and PSD on Multiresolution analysis of CNEI rainfall.

7.4.2 Results on CNEI rainfall

The PSD of each of the partially reconstructed time series obtained from MRA of CNEI rainfall at four levels is estimated. Results are shown in figure 7.10. It is seen that the PSD on MRA separate two closely spaced periods 30.3 and 60.6 y. The absence of the 30.3 y period in the directly computed PSD (Fig. 7.8) is consistent with the criterion cited in Chapter 3, i.e. the frequency separation between the two periods (30.3 - 60.6) y is too low, being less than $2/N = 0.0167\ \mathrm{y}^{-1}$. Also, the direct PSD shows two significant peaks, 2.8 and 3.3 y, above 95% confidence level whereas PSD on MRA captures 8 significant periods above 99% confidence level. These periods are 2.8, 3.3, 5.5, 6.6, 8.6, 10.9, 24.0, 30.3 y. The periodicities obtained from

the EMD technique and the PSD on MRA are compared in table 7.3.

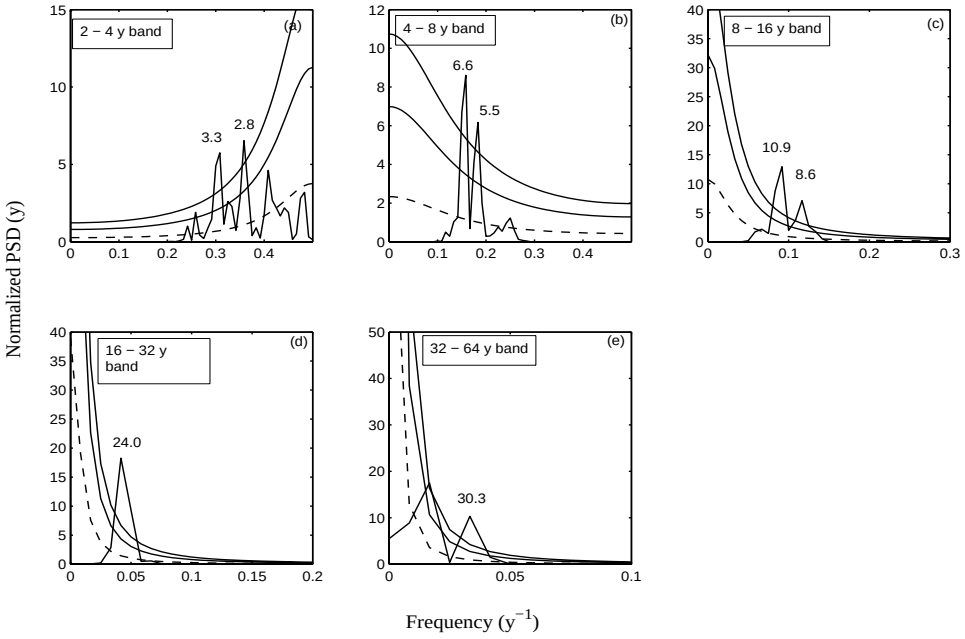

Figure 7.10: Significance test on MRA of CNEI rainfall (a) 1st reconstructed detail; (b) 2nd reconstructed detail; (c) 3rd reconstructed detail; (d) 4th reconstructed detail; (e) 4th reconstructed approximation. Note: Significant periods(y) are marked at respective peaks. Reference spectrum is represented by dotted lines with different confidence levels marked.

7.4.3 Results on PENSI rainfall

The PSD of each of the partially reconstructed time series obtained from MRA of PENSI rainfall at four levels is estimated. Results are shown in figure 7.11. It is seen from the figure that an average of the two periods at 13.3 and 17.2 y is present as 15.1 y in the direct PSD (Fig. 7.8). Also, the direct PSD shows two significant peaks at 3.3 and 7.5 y above 95% and one at 7.5 y above 99% confidence level. The significant periods captured by PSD on MRA above 99% confidence line are: 3.3, 4.5, 7.5, 13.3, 17.2 and 40.0 y. Also periods obtained from the EMD technique and the PSD on MRA are compared in table 7.4.

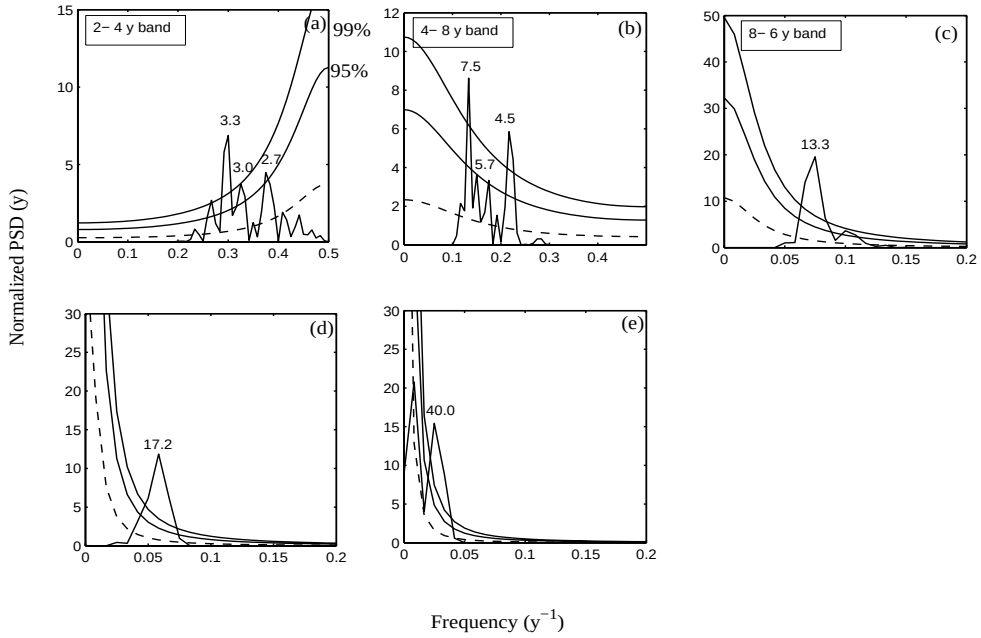

Figure 7.11: Significance test on MRA of PENSI rainfall (a) 1st reconstructed detail; (b) 2nd reconstructed detail; (c) 3rd reconstructed detail; (d) 4th reconstructed detail; (e) 4th reconstructed approximation. Note: Significant periods(y) are marked at respective peaks. Reference spectrum is represented by dotted lines with different confidence levels marked.

7.4.4 Results on WCI rainfall

The PSD of each of the partially reconstructed time series obtained from MRA of PENSI rainfall at four levels is estimated. Results are shown in figure 7.12.

Level	zero crossings on EMD	PSD on MRA
1.	2.72	2.1, 2.3, 2.8, 3.3, 4.0
2.	6.0	4.1, 5.5, 6.6
3.	10.0	8.6, 10.9
4.	20.0	24.0
5.	60.0	30.3, 60.6

Table 7.4: Periods (y) obtained by, Empirical mode decomposition and PSD on Multi-resolution analysis of PENSI rainfall.

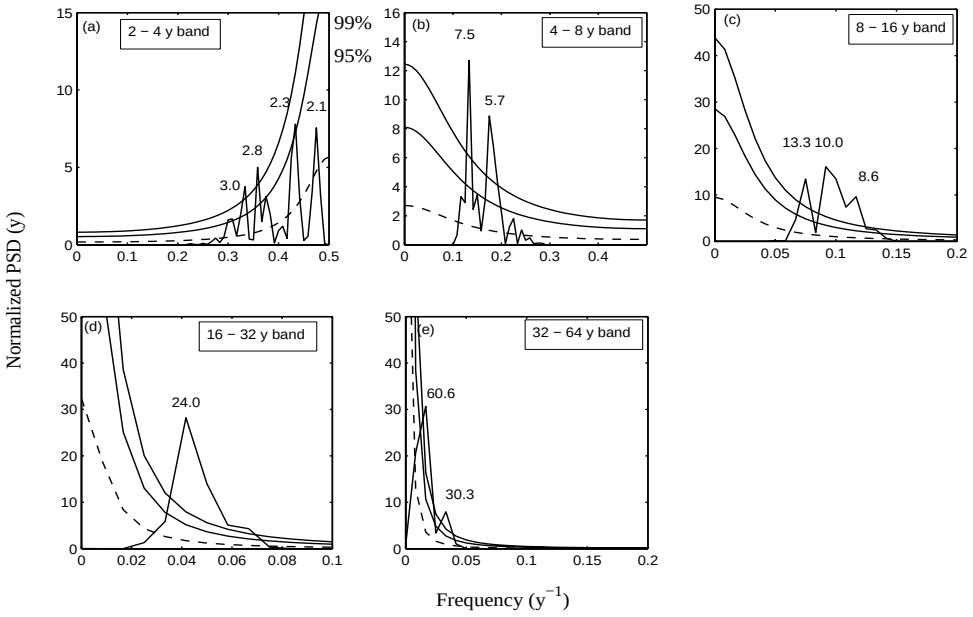

Figure 7.12: Significance test on PSD of MRA of WCI rainfall (a) 1st reconstructed detail; (b) 2nd reconstructed detail; (c) 3rd reconstructed detail; (d) 4th reconstructed detail; (e) 4th reconstructed approximation. Note: Significant periods(y) are marked at respective peaks. Reference spectrum is represented by horizontal dotted lines with different confidence levels marked.

It is seen from the figure that PSD on MRA yields two more periods, respectively

Level	zero crossings on EMD	PSD on MRA
1.	2.61	2.1, 2.3, 2.8, 3.3, 4.0
2.	5.71	4.1, 5.5, 6.6
3.	12.0	8.6, 10.9
4.	24.0	24.0
5.	60.0	30.3, 60.6

Table 7.5: Periods (y) obtained by, Empirical mode decomposition and PSD on Multi-resolution analysis of WCI rainfall.

at 8.6 and 30.3 y. Their absence in the directly computed PSD (Fig. 7.8) is again because the frequency separation between the two period pairs (8.6 - 10.0) y and (30.3 - 60.6) y is each less than $2/N = 0.0167$ y^{-1}. Also, there are 4 significant periods in the direct PSD of WCI (Fig. 6.2) above 95% confidence level, namely 2.1, 2.3, 7.5 and 60.6 y, whereas PSD on MRA yields 10 significant peaks above 99% confidence level: 2.8, 3.0, 5.7, 7.5, 8.6, 10.8, 13.3, 24.0, 30.3, 60.6. Also periods obtained from the EMD technique and the PSD on MRA are compared in table 7.5.

7.4.5 Results on NWI rainfall

The PSD of each of the partially reconstructed time series obtained from MRA of CNEI rainfall at four levels is estimated. Results are shown in figure 7.13. It is seen from the figure that PSD on MRA yields two more periods, respectively at 12.1 and 30.3 y which are absent in the direct PSD (Fig. 7.8). Also, the direct PSD of WCI shows 3 significant periods (Fig. 6.2) above 95% confidence level, at 2.3, 2.8 and 3.3 y, whereas PSD on MRA yields 10 significant peaks above 99% confidence level: 2.8, 3.3, 4.1, 4.7, 8.6, 12.1, 24.0 and 30.3. Also periods obtained from the EMD technique and the PSD on MRA are compared in table 7.6.

Level	zero crossings on EMD	PSD on MRA
1.	2.79	2.1, 2.3, 2.8, 3.3, 4.0
2.	5.45	4.1, 5.5, 6.6
3.	13.3	8.6, 10.9
4.	20.0	24.0
5.	60.0	30.3, 60.6

Table 7.6: Periods (y) obtained by, Empirical mode decomposition and PSD on Multi-resolution analysis of NWI rainfall.

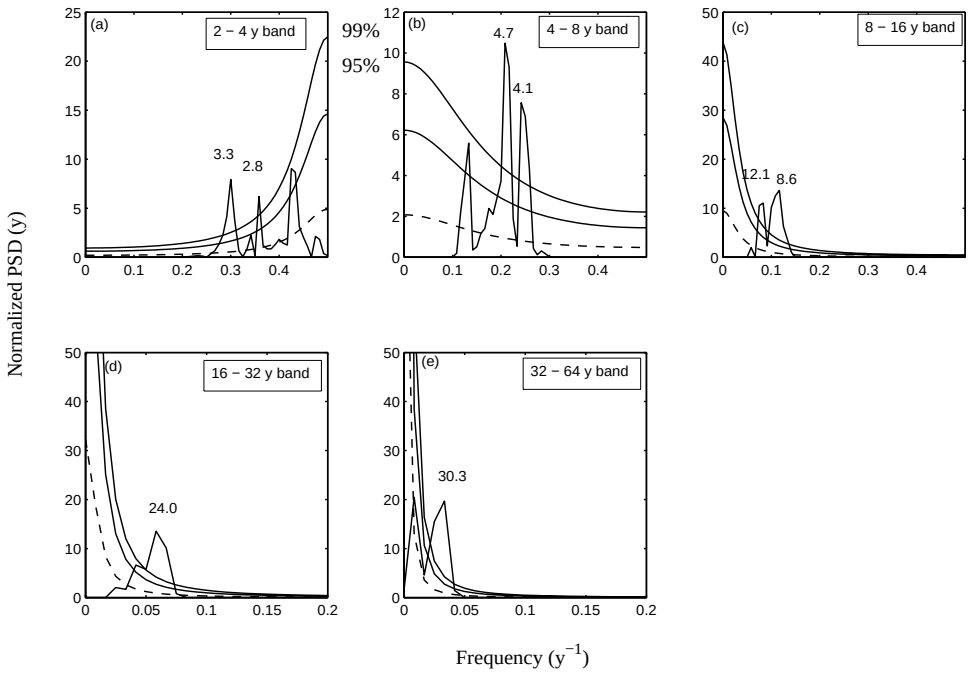

Figure 7.13: Significance test on MRA of NWI rainfall (a) PSD on its 1st reconstructed detail; (b) PSD on its 2nd reconstructed detail; (c) PSD on its 3rd reconstructed detail; (d) PSD on its 4th reconstructed detail; (e) PSD on its 4th reconstructed approximation. Note: Significant periods(y) are marked at respective peaks. Reference spectrum is represented by dotted lines with different confidence levels marked.

7.5 Spectrally Homogeneous Monsoon Regions

As SHIM was identified as an example to our proposed method of spectral homogenity in Chapter 6, we identify in this section ten Spectrally Homogeneous Regions (SHR's) in India, as listed in table 7.7 and also shown in figure 7.14. We have found that there are sub-divisions like Orissa, Kerala and Tamilnadu which are not spectrally homogeneous with any other sub-division.

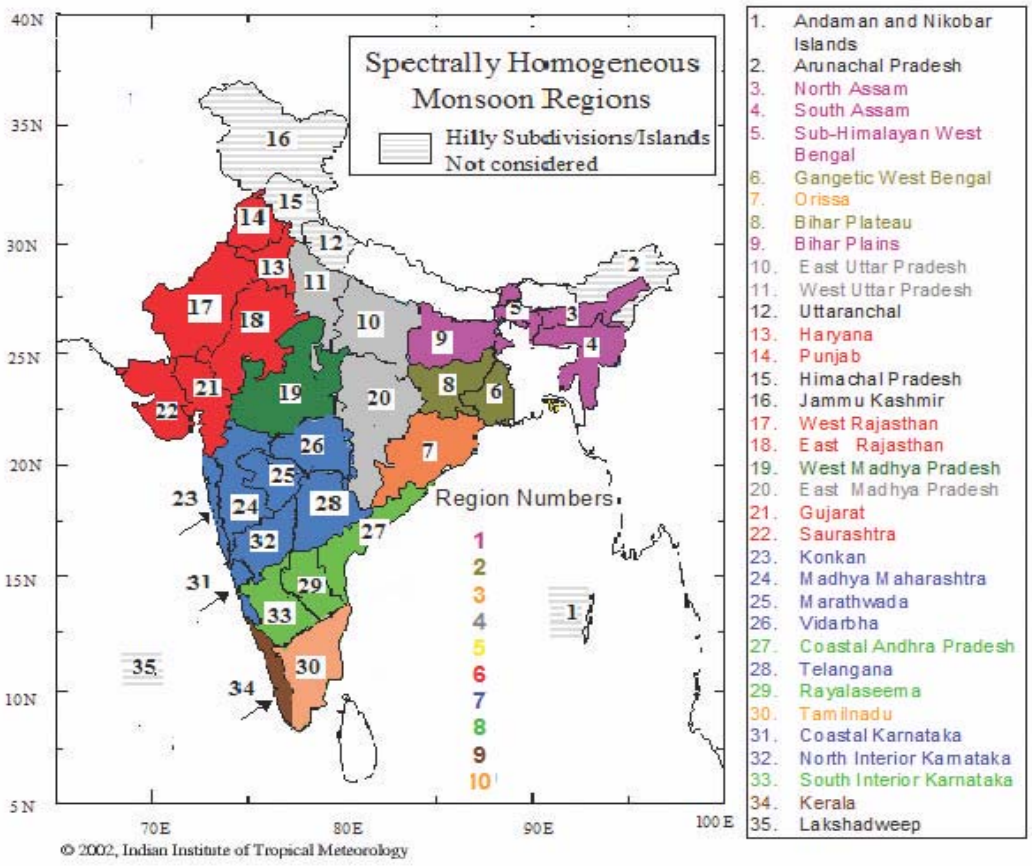

Figure 7.14: Newly defined ten Spectrally Homogeneous Monsoon Regions in India.

Region	Sub-division
SHR1	North Assam (3), South Assam (4), Sub-Himalayan W. Bengal (5), Bihar Plains (9)
SHR2	Gangetic W. Bengal (6), Bihar Plateau (8)
SHR3	Orissa (7)
SHR4	East U.P. (10), West U.P (11) East M.P (20)
SHR5	West M.P (19)
SHR6	Haryana (13), Punjab (14) West Rajasthan (17), East Rajasthan (18) Gujarat (21), Saurashtra (22)
SHR7	Konkan (23), Madhya Maharashtra (24) Marathwada (25), Vidarbha (26) Telangana (28), Coastal Karnataka (31) North Interior Karnataka (32)
SHR8	Coastal Andhra Pradesh (27), Rayalaseema (29) South Interior Karnataka (33)
SHR9	Kerala (34)
SHR10	Tamilnadu (30)

Table 7.7: Ten newly defined Spectrally Homogeneous Monsoon Regions (SHR) identified in India. Sub-divisions numbers according to Fig. 7.14 are given in brackets.

The time series for these spectrally homogeneous regions are displayed in figures 7.15 and 7.16.

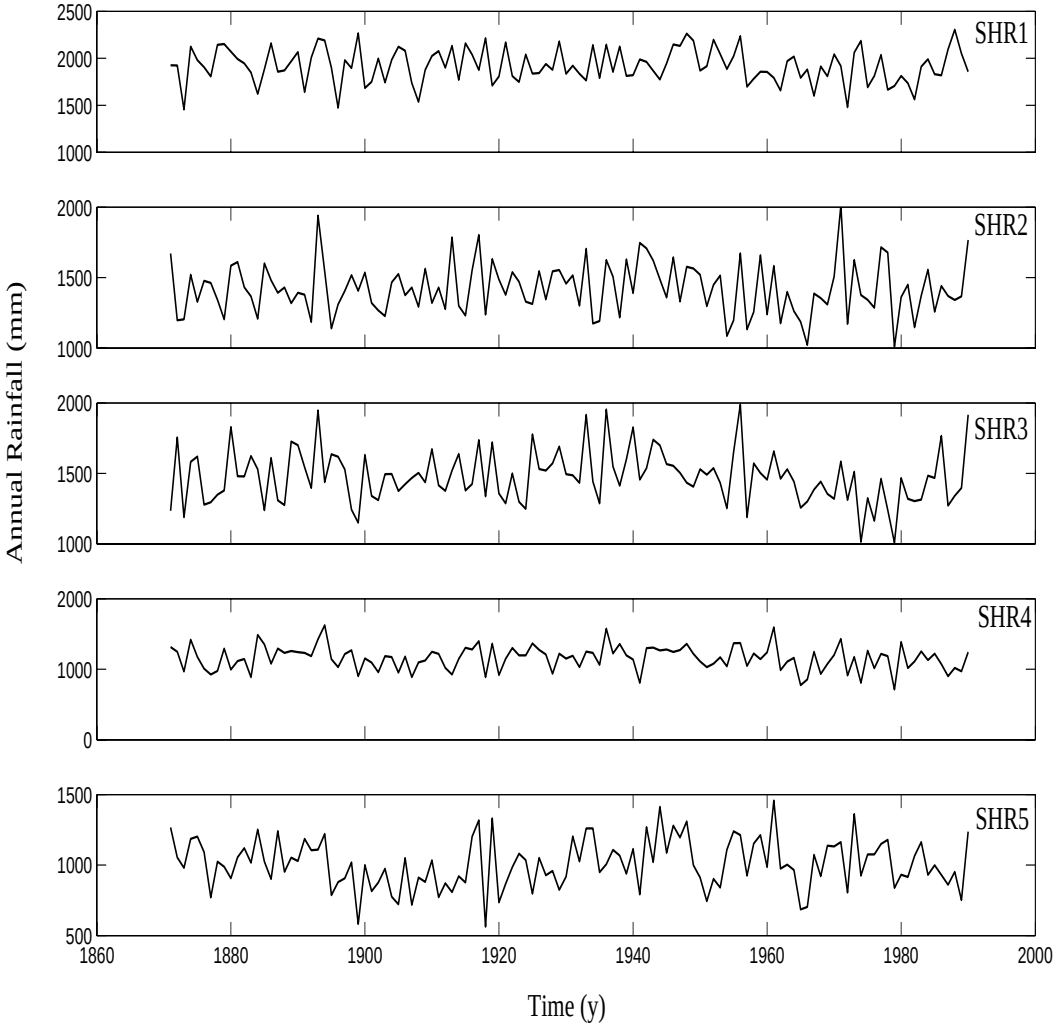

Figure 7.15: Time series of spectrally homogeneous regions.

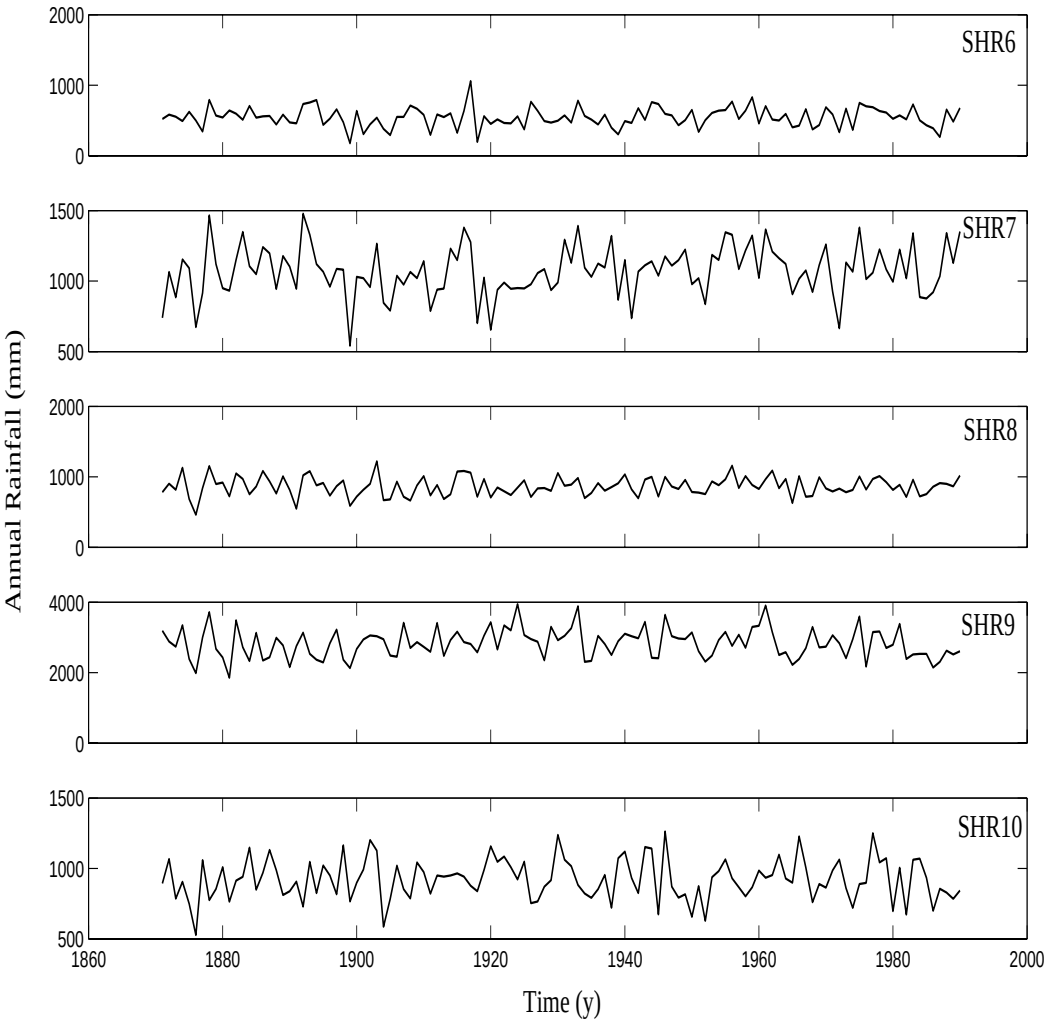

Figure 7.16: Time series of spectrally homogeneous regions.

The values of F-test where sub-divisions departing from white noise $> 99.5\%$ are grouped into spectrally homogeneous regions are given in tables 7.8–7.13.

S.no	Sub-division	Variance	F	Rejection probability(%)
1.	South Assam	0.51	3.88	> 99.9
2.	North Assam	0.45	3.24	> 99.9
3.	Sub-Himalayan W.Bengal	0.53	2.75	> 99.9
4.	Bihar Plains	0.58	2.51	> 99.9

Table 7.8: F-test for the SHR1 (s^2_{noise}=1.47).

S.no	Sub-division	Variance	F	Rejection probability(%)
1.	Bihar Plateau	0.23	4.29	> 99.9
2.	Gangetic W. Bengal	0.29	3.45	> 99.9

Table 7.9: F-test for the SHR2 (s^2_{noise}=0.99).

S.no	Sub-division	Variance	F	Rejection probability(%)
1.	West U.P.	0.33	3.99	> 99.9
2.	East U.P.	0.46	2.83	> 99.9
3.	East M.P.	0.66	2.00	$[99.5, 99.9)$

Table 7.10: F-test for the SHR4 (s^2_{noise}=1.32).

S.no	Sub-division	Variance	F	Rejection probability(%)
1.	W. Rajasthan	0.509	3.18	> 99.9
2.	Saurashtra	0.524	3.09	> 99.9
3.	Punjab	0.577	2.80	> 99.9
4.	Gujarat	0.627	2.58	> 99.9
5.	Haryana	0.700	2.31	> 99.9
6.	E. Rajasthan	0.762	2.12	$[99.5, 99.9)$

Table 7.11: F-test for the spectral deviations of SHR6 (s^2_{noise}=1.62).

S.no	Sub-division	Variance	F	Rejection probability(%)
1.	Marathwada	0.283	5.86	> 99.9
2.	Konkan	0.341	4.86	> 99.9
3.	Telangana	0.411	4.03	> 99.9
4.	Madhya Maharashtra	0.449	3.69	> 99.9
5.	North Interior Karnataka	0.553	3.00	> 99.9
6.	Costal Karnataka	0.627	2.64	> 99.9
7.	Vidarbha	.690	2.40	> 99.9

Table 7.12: F-test for the spectral deviations of SHR7 (s^2_{noise}=1.66).

S.no	Subdivision	Variance	F	Rejection probability(%)
1.	Rayalaseema	0.378	3.48	> 99.9
2.	South Karnataka	0.424	3.10	> 99.9
3.	Coastal Andhra Pradesh	0.446	2.95	> 99.9

Table 7.13: F-test for the SHR8 (s^2_{noise}=1.32).

To demonstrate spectral similarity, the estimated PSD functions of the sub-divisions constituting SHR's (listed in table 7.7) are shown in figures 7.17–7.22 respectively.

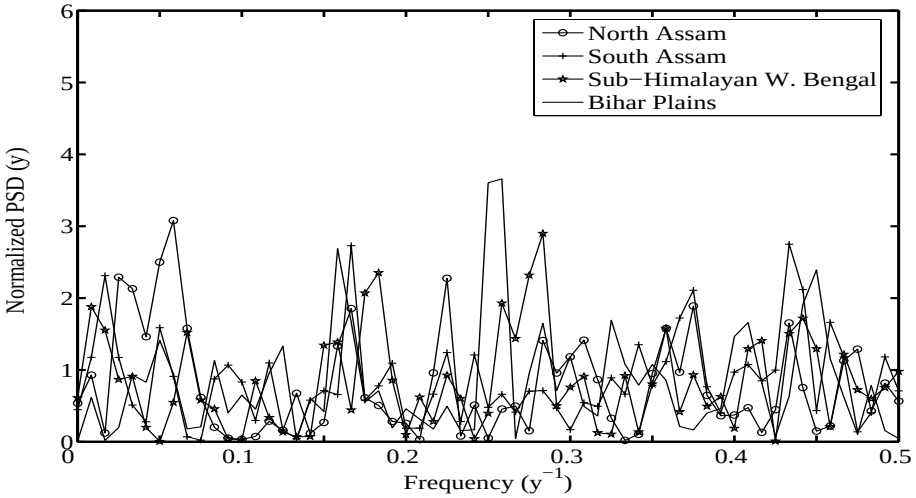

Figure 7.17: Sub-divisional rainfall spectra in SHR1.

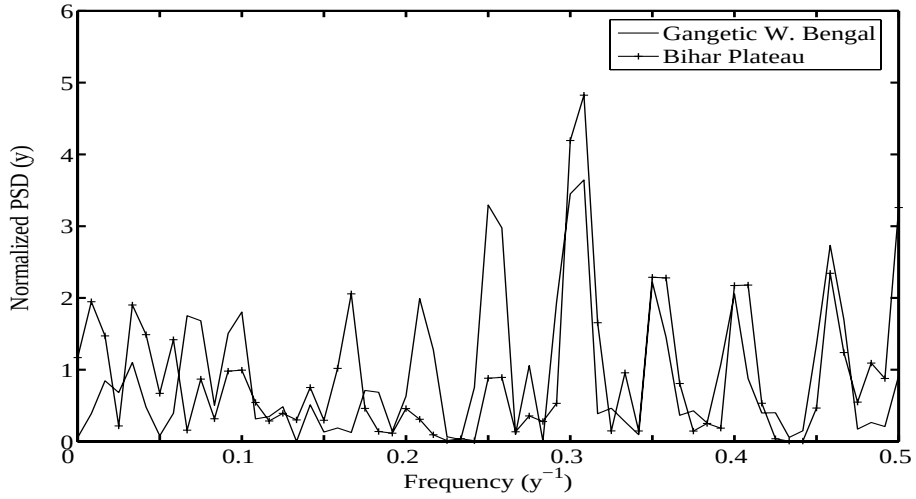

Figure 7.18: Sub-divisional rainfall spectra in SHR2.

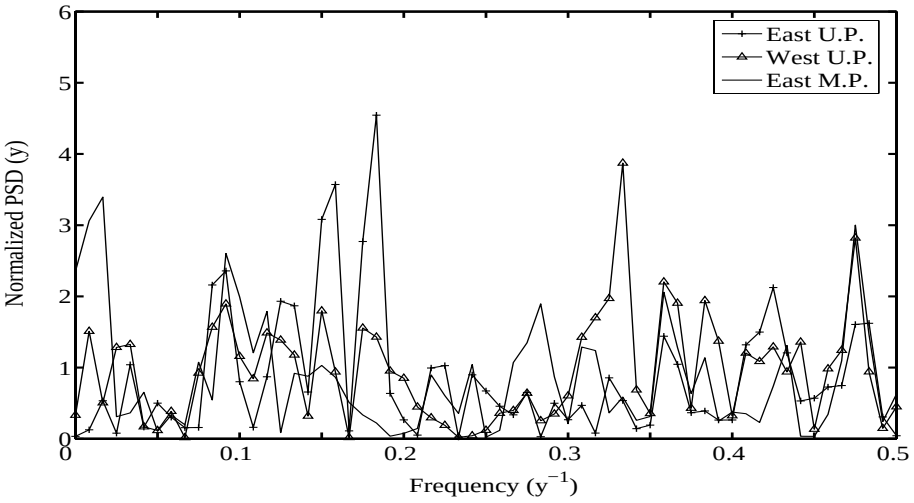

Figure 7.19: Sub-divisional rainfall spectra in SHR4.

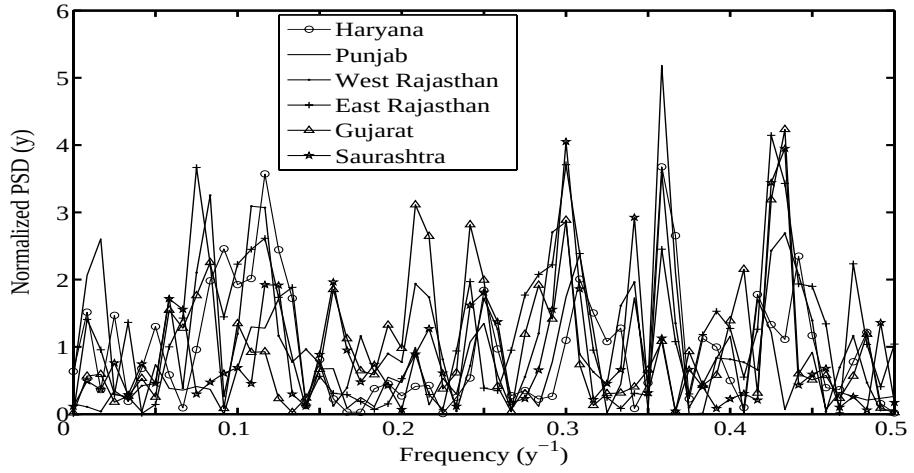

Figure 7.20: Sub-divisional rainfall spectra in SHR6.

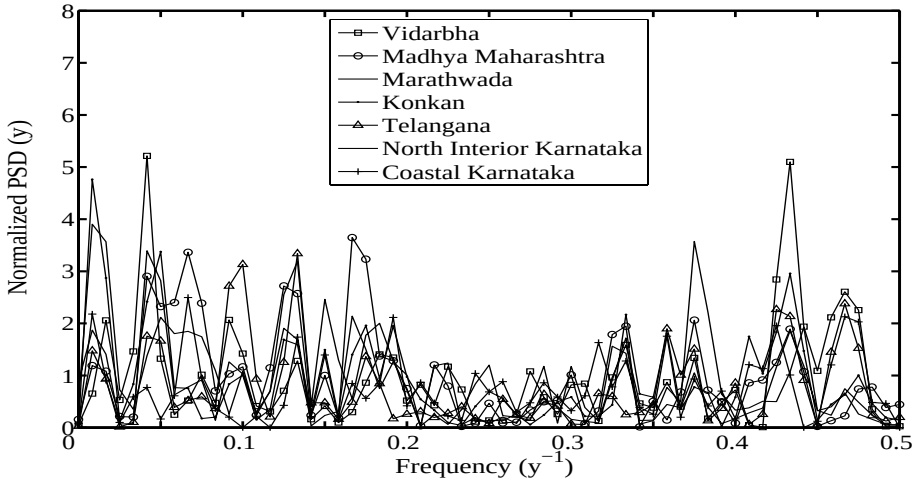

Figure 7.21: Sub-divisional rainfall spectra in SHR7.

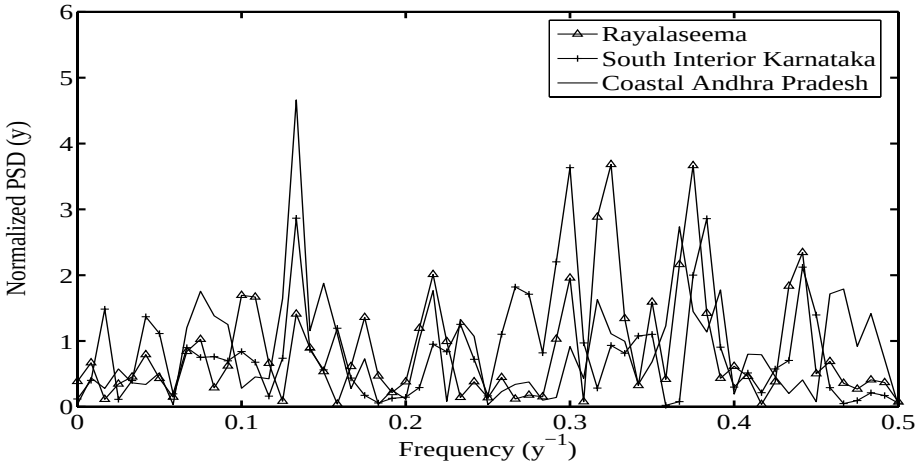

Figure 7.22: Sub-divisional rainfall spectra in SHR8.

In particular SHR7 is considered, the significant periodicities in these sub-divisions above 90% confidence level are listed in table 7.14. It is found that peaks in any frequency band for the sub-divisions constituting SHR7 (Fig. 7.21) are in near-coincidence with those from other sub-divisions which is clearly different from the seven white noise realizations, each of sample size same as rainfall (Fig.7.23).

The similarity in the sub-divisional spectra in SHR7, including the presence of the spectral dip, is visibly evident in figure 7.21, wherase the varaition in mean rainfall in these sub-divisions is dipcted in figure 7.24. The seven sub-divisions constituting SHR7 come from what are often thought to be different regions, as they span the country from Konkan and Coastal Karnataka, to the 'rain-shadow' areas of North Interior Karnataka and Madhya Maharashtra (see Fig. 7.14). Figures 7.21 and 7.23 between them show the power of the concept of spectral homogeneity, and suggest that similar dynamical forcing operate in the region, even though the magnitude of rainfall varies by a factor exceeding four.

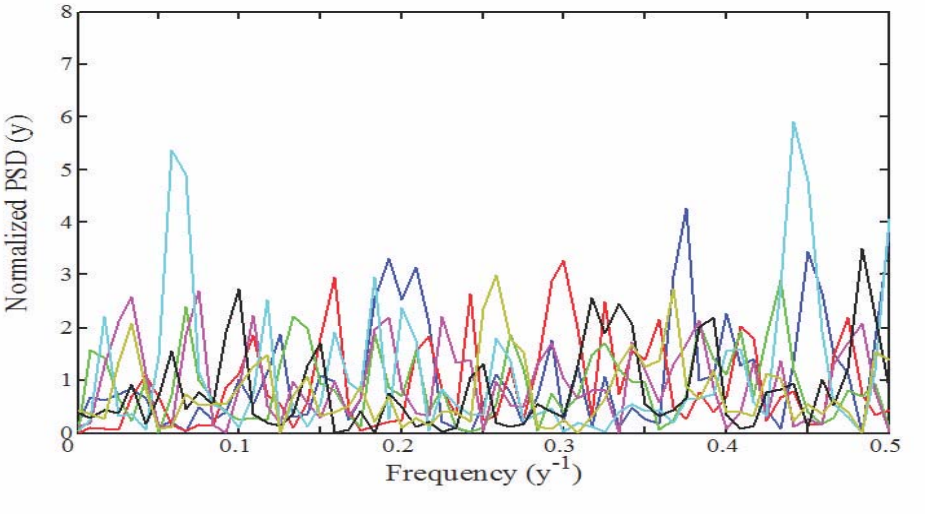

Figure 7.23: White noise spectra of seven realizations, each of sample 120.

To assess the difference between SHR7 and white noise spectra, we define the null hypothesis that the number of singletons (=unrepeated peaks) above 90% significance level in the seven sub-divisional rainfall time series in SHR7 are same as in seven realizations of white noise. On average we found that there are 22.15 unrepeated peaks or singletons in white noise above 90% significance level, with a standard deviation of 3.86. From table 7.14 we see that there are 10 singletons in SHR7 spectra. Using the z test [73] with 6 degrees of freedom, the null hypothesis is rejected at a confidence level of 99.99%.

Sub-division	Periodicities								
Konkan		2.31		3.0				**20.0**	
M. Maharashtra		2.31	2.66	3.0	**5.71** **5.99**		14.99		
Marathwada		2.31			7.51				
Vidarbha		2.31				**10.91**		**23.98**	59.88
Telangana	**2.14**	2.31	**2.79**		7.51	**10.0**			
Coastal Karnataka							14.99		
North I. Karnataka		**2.26**	**2.60** 2.66						59.88

Table 7.14: 90% Significant periodicities(y) in seven sub-divisional rainfall constituting SHR7, obtained from the direct PSD estimated using Welch technique and classical reference spectrum (Eq. 1.13). Singletons are in bold.

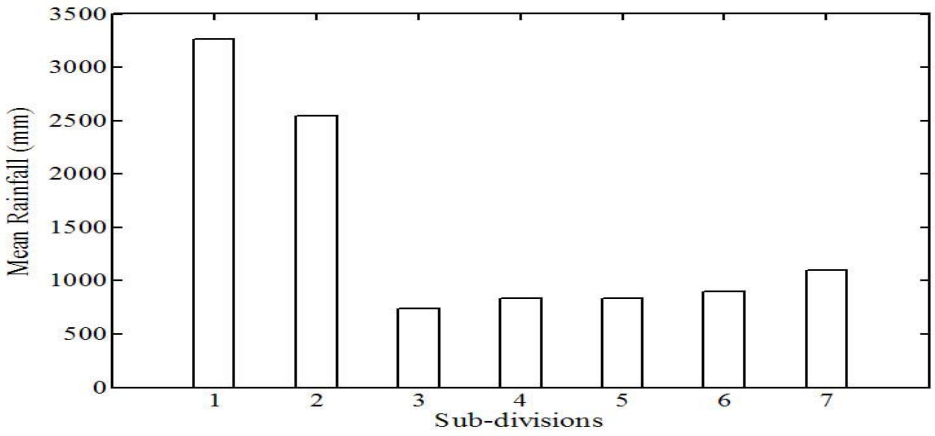

Figure 7.24: Mean rainfall of seven sub-divisional rainfall constituting SHR7: 1=Coastal Karnataka; 2=Konkan; 3= Madhya Maharashtra; 4= North Interior Karnataka; 5=Marathwada; 6= Telangana; 7=Vidarbha.

7.5.1 Periodicities in SHR's

The PSD function for each of the ten spectrally homogeneous rainfall time series is estimated by the Wlech technique. The auto-correlation coefficient for these time series are calculated at three values of lag and are given in table 7.15.

S.no	Sub-division	Auto Correlation		
		Lag-1	Lag-2	Lag-3
1.	SHR1	0.0358	-0.1399	0.0431
2.	SHR2	-0.1371	-0.1206	0.1573
3.	SHR3	-0.0238	0.0148	0.2628
4.	SHR4	-0.0284	0.0621	0.0866
5.	SHR5	0.0417	0.2683	0.1024
6.	SHR6	-0.0534	0.0208	0.1109
7.	SHR7	0.1127	0.0490	-0.0021
8.	SHR8	-0.0511	-0.2089	0.0428
9.	SHR9	0.1096	-0.1337	0.1045
10.	SHR10	0.0253	-0.0312	-0.0112

Table 7.15: Auto-correlation coefficient at various lags for SHR's during 1871-1990.

The significance levels of periodicities obtained in ten SHR's are assessed using the method of direct PSD, MRA+PSD and confidence ellipses. Results are quoted in table 7.16.

Region	PSD (above 95%)	MRA+PSD (above 99.9%)	Multivariate analysis
SHR1	2.3, 6.0	3.6 & 4.5 (99%), 6.0, 8.6 10.9, 20.0, 30.3	2.7(98.3%), 3.3(99.1%) 6.0(97.3%)
SHR2	3.3	3.3, 4.0, 10.0 30.3, 60.6	None
SHR3	2.8, 3.3, 3.6 60.0	2.8 (99%), 3.3, 3.7 4.1 & 6.2 (99%), 8.6, 13.3 (99%) 24.0, 60.6	None
SHR4	2.1, 10.9	2.8 & 6.6 (99%), 7.5 & 60.6 (99%) 5.5, 8.7, 10.9, 24.0, 30.3	10.9(94.5%)
SHR5	2.1(99%), 2.3, 13.3 60.6	2.8 & 5.5 (99%), 3.5, 7.5 8.6, 13.3, 30.3 (99%), 60.6	3.5 (95%)
SHR6	2.3, 2.8, 3.3	2.8(99%), 3.3, 4.0, 5.0 8.6, 10.0, 12.2 17.8, 30.3	2.8(99.94%), 60.0(95%) 17.2(97.2%)
SHR7	2.3	3.3, 13.3, 60.6 (99%) 5.7, 10.9, 8.6 24.0, 30.3	2.8(97.4%), 7.5(98.6%) 10.9 (91%), 6.0(99.3%) 24.0(99.2%)
SHR8	7.5	3.3, 6.7 (99%) 4.6, 7.5, 13.3 24.0, 30.3	7.5(99.85%)
SHR9	3.4	3.4, 4.4, 6.0 (99%), 9.2 17.2	None
SHR10	2.3, 3.7, 10.9	3.4, 4.6 & 5.7(99%) 10.9, 20.0, 40.0	None

Table 7.16: Significant periodicities(y) obtained from different methods in ten newly defined Spectrally Homogeneous Regions (SHR) in India.

7.6 Conclusions

- The analysis carried out on the HIM region is repeated for the other five homogeneous regions identified in [50].

- Using the proposed technique of MRA+PSD periodicities are determined in each of these regions.

- The statistical significance of these periodicities is then assessed against an appropriately colored reference spectra defined in Chapter 5.

- It is found that using the direct PSD, the homogeneous region NEI does not capture any period above 95% confidence level whereas MRA+PSD shows five periods of 6.0, 8.6, 12.0, 24.0 and 60.6 y above 99% confidence level.

- As was found in the HIM region in Chapter 5, it is also seen here that MRA+PSD in the homogeneous regions CNEI, WCI and NWI captures two additional periodicities at 8.6 and 30.3 y at a confidence level exceeding 99%. Each of these period is missed by the direct PSD but is very close to others found by it.

- The period of 10 y is 99% significant in the CNEI, PENSI, WCI and the HIM regions using appropriately colored reference spectra. Direct PSD is unable to capture the 10 y period against the white reference spectra of Eq. 1.13.

- There are some common periodicities (above 99% confidence level) among the homogeneous regions. For example the 2.8 y period is seen in the CNEI, WCI, NWI and HIM regions. Similarly 7.5 y is present in the PENSI, WCI, NWI and HIM regions, 13.3 is common in the PENSI, WCI and HIM regions, 24.0 y is common in the NEI, CNEI, WCI, NWI and HIM regions (i.e all homogeneous regions except PENSI), 60.6 y is common in the NEI, PENSI and HIM regions.

- Using the definition of spectral homogeneity proposed in Chapter 6, ten Spectrally Homogeneous Regions (SHR) are identified in India.

- Table 7.16 compares the periodicities obtained in these regions using the method of PSD, MRA+PSD and confidence ellipses.

- It is seen that SHR7 may be of particular interest as all the constituent sub-divisions here show a spectral dip around the frequency of 0.25 y^{-1}, and most of the spectral peaks in the different sub-divisions nearly coincide with each other (as is clearly seen in figure 7.21).

- In SHR7 there is one periodicity above 95% by the direct PSD whereas the present MRA+PSD technique reveals seven periodicities above 99% confidence level, with respective periods of 3.3, 13.3, 5.7, 10.9, 8.6, 24.0 and 30.3 y.

- Using the method of confidence ellipses we report five significnat common periodicities among the sub-divisions of the SHR7 region: 2.8(97.4%), 6.0(99.3%), 7.5(98.6%, 10.9 (91%) and 24.0(99.2%) condidence level.

Chapter 8

Conclusions, Explanations and Vistas Ahead

8.1 Conclusions

The main aim of this thesis has been to search for significant periodicities in the rainfall received by appropriately defined homogeneous regions of India. The main data analysed is the annual time series of homogeneous Indian monsoon rainfall and its 14 constituent sub-divisions over the period 1871-1990 as provided in [64]. Other regions are also studied, to provide a broad view of spectral characteristics of Indian rainfall. Following are the major conclusions from the thesis:

Using multi-resolution analysis (MRA), HIM rainfall annual data of 120 y is decomposed at 7 scales corresponding to periods of 2, 4, 8, 16, 32, 64, 128 y. A hybrid technique is proposed where the advantages of wavelet methods in handling nonlinear non-stationary time series are combined with the frequency localization property of Fourier spectral analysis by the following procedure. At the first level of MRA decomposition the original time series is divided into two partially reconstructed time series. The detail partially reconstructed is obtained by setting approximate coefficients of the low frequency components to zero. This helps in the elimination of sinusoidal interference in the highest frequency band. This time series is then subjected to classical Fourier spectral analysis, which is successful in separating closely

spaced frequencies provided the associated spectral peaks are sufficiently narrow. This procedure is repeated for all levels and successively interference is removed from higher frequency bands by setting the corresponding approximate coefficients to zero.

A theoretical justification for the ability to separate closely spaced frequencies using the approach proposed here is as follows. Once it is decided that the time series being handled is nonlinear and possibly non-stationary, the wavelet technique is clearly an attractive option. However separating closely spaced frequencies by direct inspection of a wavelet map poses difficulties. In the first place even a pure sinusoid leads to wavelet coefficients over a band of time scales, so separating two closely-spaced frequencies this way would not be practical. Secondly, nonlinearity, non-stationarity, and the regime transitions that may occur as a consequence can cloud the issue. Finally even when a time series contains a pure sinusoid the interference (or leakage) it produces can make contributions not only within its own band but in neighboring bands as well. This problem leads to serious difficulties in many technological applications; for example when there is interference by the mains frequency in an electronic signal that is the object of analysis. Here we fall back on the advantages of Fourier analysis, which may be able to pick out sharp peaks when leakage is minimized.

To summarize, our finding are as follows.

- The Welch technique with Hanning window is a powerful tool for estimating the PSD function of a time series.
- The MRA+PSD technique proposed here is found to be a powerful tool to identify and separate frequencies which are closer than $2/N$ where N is number of sample points in a time series.
- An effective criterion based on better localization of specific frequency components and accurate estimation of their amplitudes is used to select an appropriate wavelet. Comparing the performance of discrete Meyer (dmey), daubechies (db2) and four other wavelets on test signals, it is shown that the discrete Meyer wavelet offers the best performance in separating closely-spaced frequencies without yielding spurious peaks.

- By analysing the performance of the proposed method on specially devised test signals, it is explicitly demonstrated that when frequencies are closely spaced, as in test signal 2, the method of empirical mode decomposition (EMD) will capture only average values as it counts zero crossings, whereas PSD on MRA with the discrete Meyer wavelet is able to distinguish between closely spaced frequencies and capture them separately.

- The same is true in the case of HIM rainfall data; PSD on MRA decomposition is able to yield periods close to each other missed in earlier work by continuous wavelet transform (CWT) and EMD methods, which yield only some weighted average values of nearby periods.

- There are therefore grounds to believe that certain periodicities in rainfall have been missed by earlier techniques.

- The present analysis incidentally uses MRA as a tool also to identify the range of wavelet scales over which stationarity can be assumed.

- The F and z statistical tests are performed on two consecutive populations using the reconstructed details and approximation of MRA decomposition at four levels, of 60 samples each from the two test periods over 1871–1930 and 1931–1990. The data are analysed for stationarity.

- It is shown that the HIM rainfall data exhibits non-stationarity only over long periods of order 60 y. The stationary component of HIM rainfall is obtained from adding the partially reconstructed time series at four levels of detail. This reconstructed stationary time series accounts for 93% of the variance of the total HIM rainfall.

- Application of spectral analysis on partially reconstructed wavelet coefficient time series is justified by the fact that rainfall is stationary except over long periods of order 60 y.

- In case of HIM rainfall, the present MRA+PSD technique reveals 17 periodicities at 2.1, 2.3, 2.5, 2.7, 2.8, 3.0, 3.3, 4.1, 5.7, 6.6, 7.5, 8.6, 10.0, 13.3, 24.0,

30.3, 60.0 y. Of these the periodicities at 8.6 and 30.3 y are not present in any previous work.

- The important question of statistical significance of these periodicities is discussed and spectral peaks are tested against appropriately colored noise.

- The auto-correlation coefficient of HIM rainfall series at lag 1 is $\alpha = -0.007$, i.e. close to zero. This implies that in the classical technique of devising a reference spectrum based on α, the periodicities in HIM rainfall are tested against a nearly flat spectrum (i.e. white noise). However we present strong evidence for the presence of a spectral dip in HIM rainfall around a frequency of 0.25 y^{-1}, and construct a testing procedure that takes this into account. Thus, using MRA we have shown that at the first level of decomposition for high frequencies the reference spectrum is blue, whereas for low frequencies it is red. Hence the statistical significance of these periodicities must be tested against appropriately colored noise, matching the observed spectrum of rainfall.

- The estimated PSD of HIM rainfall time series against classical reference spectra of Eq. 1.13 shows only one periodicity of 2.3 y above 99% confidence level.

- Using the present method of MRA+PSD against appropriately colored reference spectra, ten periodicities are significant above the 99.9% confidence level, with respective periods of 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3 and 60.6 y.

- The significant peaks account for 64% of the variance of the total HIM rainfall. However the 60.6 y periodicity is close to the boundary of stationarity, and should be interpreted with caution.

- One persistent question that arises in the quest for identifying periodicities in Indian rainfall is the meteorological heterogeneity in the data. In spite of many studies of the problem, and the considerable evidence of this heterogeneity, many analyses still continue to rely on all-India indices. Even the homogeneous regions identified by different workers may contain heterogeneities not considered in the criteria laid down for determining the degree of homogeneity.

For example, we show in Chapter 5 that only seven out of the 14 sub-divisions in the HIM region exhibit a characteristic spectral dip at 0.25 y^{-1}. Using heterogeneous data will mask inherent periodicities in the phenomenon. Towards that end we here propose a quantitative definition of spectral homogeneity, and use this concept to identify a spectrally homogeneous (SHIM) sub-region within the HIM region. The new methodology proposed here is founded on the observation that homogeneous zone rainfall time series are derived by an averaging process over numerous other time series of the same length involving $O(10)$ sub-divisional time series and $O(10^2)$ rain gauge records, and so cannot be treated as comprising just a set of isolated individual measurements, one for each year.

- It is found that the four sub-divisions of Telangana, West Madhya Pradesh, East Rajasthan and Vidarbha form a spectrally homogeneous sub-region within the HIM region defined in [50].

- Tested against the classical reference spectrum of Eq. 1.13, HIM rainfall time series over the time period 1871-1990 shows one 99% significant period of 2.3 y and a 95% significant period at 2.8 y. However, by introducing the spectrally homogeneous sub-region SHIM, we have shown that there is a 99.5% significant period of 2.3 y, along with four other significant peaks above 95% confidence at 2.1, 7.5, 13.3 and 60.6 y and a 2.8 y period at $> 94.8\%$..

- Using MRA+PSD, eleven significant periodicities in the SHIM rainfall above 99% confidence level are found, namely 2.3, 2.8, 3.3, 5.7, 7.5, 8.6, 10.0, 13.3, 24.0, 30.3 and 60.6. y.

- Using confidence ellipse technique a new method to detect significant common periodicities in the cross-correlated sub-divisional rainfall is proposed. We report here seven significant periodicities in SHIM rainfall, five at confidence levels exceeding 95% (respectively 2.1 y at 97.8%, 2.3 y at 95.7%, 2.8 at 98.4%, 7.5 at 99.5%, 24.0 at 99.4%), and 10.9 at 90.5 % and 60.6 at 93.2% confidence level.

- We found that the classical test of checking confidence level of a spectral peak is inadequate when the amplitude of longer periods is low. The proposed confidence ellipse technique however shows high confidence levels even for relatively low amplitudes when the cross-correlation is high. Hence the present method is less amplitude-sensitive than other available methods.

- It is concluded that a common period may be present in all the sub-divisions constituting a region but the significance of their commonalty can not be established using the classical method (as described by Eq.1.13). However it is found that the proposed method of confidence ellipse is a more informative and powerful tool for testing confidence levels of common periodicities in correlated time series.

- Earlier methods of classification of India into homogeneous regions based on rainfall do not take into account spectral characteristics. Using spectral homogeneity as a criterion, we here classify the country into ten spectrally homogenous regions. These new regions cut across those defined earlier by other workers. For example, the Spectrally Homogeneous Region7 has (only) six sub-divisions from the HIM region and one from the PENSI region.

- SHR7 may be of particular interest as all the constituent sub-divisions show a spectral dip around the frequency of 0.25 y^{-1}, and most of the spectral peaks in the different sub-divisions nearly coincide with each other.

- The significance levels of periodicities detected in these regions are assessed. In SHR7 there is reported to be one periodicity above 95% by the direct PSD whereas the present MRA+PSD technique reveals seven periodicities above 99.9% confidence level, with respective periods of 3.3, 13.3, 5.7, 10.9, 8.6, 24.0 and 30.3 y.

- In summary, it is found that the wavelet-based methods developed in this thesis suggest the presence of far more significant periodicities than traditional methods of spectral analysis do. The application of the present results to prediction is now receiving attention.

8.2 Explanations

- All-India (India taken as one unit) summer monsoon rainfall data has been considered for various studies to understand the spatial and temporal variations of rainfall within country. However in this thesis we have preferred to analyze various homogeneous zone monsoon rainfall time series. In particular we have devoted considerable attention to the Homogeneous Indian Monsoon (HIM) region, as it is a good rainfall index for an analysis like the present. The reason is that the geographical area it covers is meteorologically homogeneous, and so physically coherent signals tend to show strong periodicities that may not be present in mixed zones.

- The auto-correlation coefficient for HIM rainfall at lag 1 is -0.007, so the reference spectra according to Eq. 1.13 is very close to white. However the HIM spectrum shows relatively high power at both higher and lower frequencies. Therefore, appropriate colored noise spectra obtained from MRA decomposition at different scales appear to be necessary.

- We have concluded that the standard tests used for significance tests can be misleading. This is because

  - The classical reference spectrum (Eq. 1.13) capture only the largest amplitude peak as highly significant; when the amplitude of longer periods is low such tests are unable to assess their significance satisfactorily.

  - Most climatic and geophysical time series tend to have spectra that are not even approximately white, hence the reference spectra used often need to be appropriately colored.

8.3 Vistas Ahead

- New methods need to be devised to assess the significance levels of periodicities in a time series.

- Wavelet based prediction techniques have been extensively used recently in time series analysis. For rainfall data, application of the proposed MRA technique as a tool for prediction is the subject of future work.

- It is a long-debated question whether, and if so by what means, we can distinguish between observed irregular signals that are deterministically chaotic or stochastic. Using wavelets to achieve this goal can be the subject of future work.

- The present work showing several periodicities in regional monsoon rainfall needs to be exploited for rainfall prediction.

Bibliography

[1] Hannan E.J. (1960): *Time Series Analysis*, Methuen, London.

[2] Shumway R. and Stoffer D. (2000): *Time Series Analysis and its Applications*, New York, Springer texts in statistics.

[3] Bartlett M. S. (1966): *An Introduction to Stochastic Processes with Special Reference to Methods and Application*, 2nd Ed. Cambridge University Press, London.

[4] Brillinger D.R. (1975): *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York.

[5] Ott E. (1993): *Chaos in Dynamical Systems*. NewYork: Cambridge Univ. Press.

[6] Bloomfield P. (1976): *Fourier Analysis of Time Series-An Introduction*. New York: Wiely.

[7] Stoica P. and Moses R.L. (1997): *Introduction to Spectral Analysis*, Prentice Hall.

[8] Chatfield C. (1984): *The Analysis of Time Series: An Introduction*, Chapman and Hall, New York 3rd ed., 286 pp.

[9] Wiener N. (1930): Generalised harmonic analysis, *Acta Math.*, **55**, 117–258.

[10] Priestley M.B. (1981): *Spectral Analysis and Time series*, Academic press (London).

[11] Thomson D.J. (1982): Spectrum estimation and harmonic analysis, *Proc. IEEE* **70**, 1055-1096.

[12] Percival D.B. (1993): *An Introduction to Spectral Analysis and Wavelets*, International Workshop on Advanced Mathematical Tools in Meterlogy Torino.

[13] Denison D.G.T, Walden A.T., Balogh A. and Forsyth R.J. (1999): Multitaper testing of spectral lines and detection of the solar roation frequency and its harmonics, *J. R. Statistical Society: Series C* **48** 427-439.

[14] Hannan E.J. and Quinn B.G. (1989): The resolution of closely adjacent spectral lines, *Journal of Time Series Analysis*, **10**, 13-31.

[15] Lee H. (1992): The Cramer-Rao bound on frequency estimates of signals closely spaced in frequency, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **40**, 15081517.

[16] Lee H. (1994): The Cramer-Rao bound on frequency estimates of signals closely spaced in frequency (unconditional case), *IEEE Transactions on Signal Processing*, **42**, 6, 1569-1572.

[17] Daubechies I. (1990): The wavelet transform, time frequency localization and signal analysis, *IEEE transcations on Information Theory*, **36**, 5, 961-1005.

[18] Farge M. (1992): Wavelets transfoms and their applications to turbulence, *Ann. Reviews of Fluid Mechanics* **24**, 395-457.

[19] Ivanov P. Ch., Rosenblum M.G., Peng C.K., Joseph Mietus, Shlomo Havlin, Stanley H.E. and Goldberger A.L. (1996): Scaling behaviour of heartbeat intervals obtained by wavelet-based time series analysis, *Nature*, **383**, 26, 323-327.

[20] Kailas S.V. and Narasimha R. (2000): Quasi cycles in monsoon rainfall using wavelet analysis, *Current Science*, **78**, 592-595.

[21] Narasimha R. and Kailas S.V. (2001): A wavelet map of monsoon variability, *Proc Indian National Science Academy* **67**, A, 3, 327-341.

[22] Bhattacharyya S. (2005): A wavelet analysis of possible connections between solar processes, Indian monsoon rainfall and Enso indices, (Ph.D thesis, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore , India).

[23] Bhattacharyya S. and Narasimha R. (2005): Possible association between Indian monsoon rainfall and solar activity, *Geophysical Research Letters* **32**, L05813.

[24] Bhattacharyya S. and Narasimha R. 2007. Regional differentiation in multidecadal connections between Indian monsoon rainfall and solar activity., *J. Geophy. Res.*, **112**, D24103.

[25] Blackman R.B and Tukey J.W. (1958): *The Measurement of Power Spectra from the point of View of Communications Engineering*, Dover Publications, New York.

[26] Meyer Y. (1992): *Wavelets and Operators* (Cambridge University, Cambridge, translated by D.H. Salinger).

[27] Schuster A. (1898): On the investigation of hidden periodicites with application to a supposed 26 day period of meterological phenomena, *Terrestrial Magnetism*, **3**, 13-41.

[28] Welch P.D. (1967): The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics*, **AU-15 (2)**, 70-76.

[29] Gilman D.L., Fuglister F.J. and Mitchell J.J. (1963): On the power spectrum of "Red noise ", *J.Atmos.Sci* **20**, 182-184.

[30] Percival D.B. and Walden (2000): Wavelet methods for time series analysis, *Cambridge University Press.*

[31] Perrier V., Philipovitch T. and Badevant C. (1995): Wavelet spectra compared to Fourier spectra *Journal of Mathematical Physics*, **36**, 3, 1506-1519.

[32] Grenfell B.T, Bjornstad O.N. and Kappey J. (2001): Travelling waves and spatial hierarchies in measles epidemics, *Nature*, **414**, 13, 716-723.

[33] Anderson D.M., Woodhouse C.A. (2005): Let all the voices be heard, *Nature* **433**, 587-588.

[34] Morlet J., Arens G., Fourgeau E., Giard D. (1982): Wave propogation and sampling theory, *Geophysics*, **47**, 203-236.

[35] Grossmann A. and Morlet J. (1985): Decompositions of functions into wavelets of constant shape and related transforms. In L.Streit, editor, *Mathematics and Physics, lectures on recent results*, World Scientific, River Edge, NJ.

[36] Strang G. and Nguyen T. (1996): *Wavelets and Filter Banks*, Wellesley-Cambridge Press.

[37] Mallat S. (1989): A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pat. Anal. Mach. Intell.* **11**, 674–693.

[38] Friis-Christensen E. and Lassen K. (1991): Lenght of the solar cycle: an indicator of solar activity closely associated with climate, *Science*, **254**, 698-700.

[39] Fligge M., Solanki S.K., Beer J. (1999): Determination of solar cycle lenght variations using the continuius wavelet transform, *Astronomy and Astrophysics*, **346**, 313-321.

[40] Torroslu H. and Kantarcioglu M. (2001): Mining cyclically repeated patterns, *Springer Lectutre notes in Computer Science* **2114**, 83-90.

[41] Feng Bo, Xi-Zeng K.E. and Hua-ling Ding (1998): Wavelet analysis of sunspot numbers, *Chinese Astronomy and Astrophysics*, **22**, 1, 83-91.

[42] Zarei J. and Poshtan J. (2007): Bearing fault detection using wavelet packet transform of induction motor stator current, *Tribology International* **40**, 763-769.

[43] Rao M. and Bhopardikar R. (1998): *An Introduction to Wavelet Transforms* (Addison Wesley).

[44] Frazier M. (1999): *An Introduction to Wavelets Through Linear Algebra*, Springer.

[45] Huang N.E, Shen Z., Long S.R., Wu M.C., Shih H.H., Zheng Q., Yen N.C., Tung C.C., Liu H.H. (1998): The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R Soc. Lond.* **A** 454-903.

[46] Rajeevan M., Pai D.S., and Kumar Anil R. (2005): New statistical models for long range forecasting of southwest monsoon rainfall over India, *NCC Research Report* 1.

[47] Jagannathan P. and Bhalme H.N. (1973): Changes in the patterns of distribution of south-west monsoon rainfall over India associated with sunspots, *Monthly Weather Review* **101 9**, 691-700.

[48] Jagannathan P. and Parthasarthy B. (1973): Trends and periodicities of rainfall over India, *Monthly Weather Review* **101 4**, 371-375.

[49] Kulkarni J.R. (2000): Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon, *International Journal of Climatology* **20**, 89-104.

[50] Parthasarthy B., Munot, A.A. and Kothwale, D.R. (1993): Homogeneous Indian monsoon rainfall: variability and prediction, *Proceedings of Indian Academy of Science (Earth and Planetary Science)* **102**, 121-155.

[51] Kumar K. (1997): Seasonal forecasting of Indian summer monsoon rainfall: diagnostics and synthesis of regional and global signals, (Ph.D thesis, University of Pune).

[52] Blackman R.B and Tukey J.W. (1958): *The Measurement of Power Spectra from the point of View of Communications Engineering* (Dover Publications).

[53] Olhede S. and Walden A.T. (2004): The Hilbert spectrum via wavelet projections, *Proc. R Soc. Lond. A* **460** 955-975.

[54] Thomson D.J. (1990): Time series analysis of holocene climate data, *Philos. Trans. R. Soc. London* A **330** 601-616.

[55] Iyengar R.N. and Raghu Kanth S.T.G. (2004): Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall, *Meteorology and Atmospheric Physics*.

[56] Charney J.G. and Shukla J. (1981): *Predictibility of Monsoons*. In: *Monsoon Dynamics* (J. Lighthill and R.P. Pearce Eds.) (Cambridge University Press) 99–110.

[57] Hasselmann K. (1976): *Stochastic Climate Models* Part I. Theory. Tellus, **28** 6, 473-485.

[58] Fisher R.A. (1929): Tests of significance in harmonic analysis, *Proceedings of the Royal Society of London Series A* **125**, 54-59.

[59] Allen M.R. and Smith L.A. (1996): Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise, *Journal of Climate* **9**, 3373-3404.

[60] Vaughan S. (2005): A simple test for periodic signals in red noise, *Astronomy and Astrophysics* **431**, 391-403.

[61] Ghil *et al.* (2000): Advanced spectral methods for climatic time series, *Reviews of Geophysics* **40**.

[62] Rohde and Muller (2005): Cycles in fossil diversity, *Nature* **434**, 208-210.

[63] Owen J.G. and Chmielewski M.A. (19850: On canonical variates analysis and the construction of confidence ellipses in systematic studies, Syst. Zool. **34(3)**, 366-374.

[64] Parthasarthy B., Munot, A.A. and Kothwale, D.R. (1995): Contributions from Indian Institute of Tropical Meterology, Pune: Monthly and seasonal rainfall series for all-India homogeneous regions and meterological subdivions: 1871-1994, *Research Report no. RR-065, Indian Institute of Tropical Meterology, Pune.*

[65] Parthasarthy B. (1984b): Interannual and long-term variability of Indian summer monsoon rainfall, *Proc. Indian Acad. Sci. (Earth Planet. Sci)* **93**, 371-385.

[66] Mooley D.A. and Parthasarathy B. (1984a): Fluctuations in all-India summer monsoon rainfall during 1871-1978, *Climate Change*, **6**, 287-301.

[67] Subbaramayya I. and Naidu C.V. (1992): Spatial variations and trends in the Indian monsoon rainfall, *Int. J. Climatol.* **12**, 597-609.

[68] Flandrin P., Rilling G. and Goncalves P. (2004): Empirical mode decomposition as a filter bank, *IEEE Signal Process. Lett.* **11** (2004) 112–114.

[69] Xu L. and Yan Y. (2004): Wavelet-based removal of sinusoidal interference from a signal, *Meas. Sci. Technol.* **15**, 1779-1786.

[70] Verbeek J., Rolain Y. and Pintelon R. (1999): Leakage reduction in measurement data contaminated with 50 Hz mains and harmonic frequency components, *IEEE Instrum. Meas. Tech.* **2**, 1054-1058.

[71] Donoho D., Mallat S. and Sachs R.V. (1998): Estimating covariance of locally stationary processes: rates of convergence of best basis methods, *Tech. Rep., Dept. Statistics*, (Standford Univerisity, Standford.

[72] Nason G.P., Sachs R.V. and Kroisadt G. (1999): Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62 (2)** 271-292.

[73] Crow E. L.,F. K. Davis and M. W. Maxfield (1960): Statistics manual: with examples taken from ordnance development, (*Dover, Mineola, N. Y*).

[74] Broomhead D.S. and King G. (1986a): Extracting qualitative dynamics from experimental data, *Physica D*, **20**, 217-236.

[75] Ghil M. and Vautard R. (1991): Interdecadal oscillations and the warminf trend in global temperature time series, *Nature*, **350**, 324-327.

[76] Prokoph A., Fowler A.D. and Patterson R.T. (2000): Evidence for periodicity and nonlinearity in a high-resolution fossil record of long-term evolution, *Geology* **28**, 867-870.

[77] Marsaglia G. and Tsang W.W. 2000: The ziggurat method for generating random variable, *Journal of Statistical Software*, **5 8**.
http://www.jstatsoft.org/vo5/i08/

[78] Gadgil S. and Iyenger R.N. (1980): Cluster analysis of rainfall stations of the Indian peninsula, *Quarterly Journal of the Royal Met. Soc* **106** 873.

[79] Gadgil S. and Joshi N.V. (1983): Climatic clusters of the Indian region, *The Royal Meterological society* **3**, 47-63.

[80] Shukla J. 1987a: Interannual variability of monsoons; In:*Monsoons* (eds) J S Fein and P L Stephens (New York: Wiley and Sons) Chapter 14, 399–464.

[81] Anderson T.W. (2003): *An Introduction to Multivariate Statistical Analysis*, Wiely Third ediction.

[82] Owen J.G. and Chmielewski M.A. (1985): On canonical variates analysis and the construction of confidence ellipses in systematic studies, *Syst. Zool.* **34(3)**: 366-374.

[83] Hotelling H. 1931: The generalization of Students ratio, *Ann. Math.Stat.*, **2**, 360378.

[84] Overall J.E. and Keett C.J. (1972): *Applied Multivariate Analysis*. McGraw-Hill, Inc.

List of Author's Publications

- Sarita Azad and S.K.Sett (2000), A wavelet analysis of singularities of sunspot numbers, International Conference on Design of Experiments, *ISI Delhi Delhi University.*

- Sarita Azad and S.K.Sett (2003), Wavelets detecting chaos in logistic map, Proceedings of National Conference on Nonlinear Dynamics, *Indian Institute of Technology, Kharagpur.*

- Sarita Azad and S.K.Sett (2004), Wavelets analyzing chaos in a time series: a recent approach, poster presentation at International Conference on Nonlinear Phenomena, *Indian Institute of Science, Bangalore.*

- Sarita Azad, R.Narasimha and S.K.Sett (2007a), Multiresolution analysis for separating closely spaced frequencies with an application to Indian monsoon rainfall data, *International Journal of Wavelets, Multiresolution and Information Processing*, **5**, No.5, 735–752.

- Sarita Azad, R.Narasimha and S.K. Sett (2007b), A wavelet based significance test for periodicities in Indian monsoon rainfall, *International Journal of Wavelets, Multiresolution and Information Processing*, in press.

- Sarita Azad and R.Narasimha (2008), Periodicities in Indian monsoon rainfall over spectrally homogeneous regions, *to be Communicated.*