Chapter 2
Data and Methodology
2.1 Data used in the study

2.1.1 Rainfall Data

2.1.1.1 All India Summer Monsoon Rainfall Data

The All India Summer Monsoon Rainfall for the period 1871 to 2003 is obtained from the official website of Indian Institute of Tropical Meteorology (IITM), Pune. The rainfall series is an area-weighted average for 306 rain-gauge stations in India. The network of rain-gauge stations is made in such a manner that the network provides one representative station per district having a reliable record for the longest possible period. The network selected under these constraints consists of 306 almost uniformly distributed stations for which rainfall data are available from 1871. The hilly regions consisting of four meteorological subdivisions of India which are parallel to Himalayan mountain range have not been considered in view of the meager rain-gauge network and low areal representation of a rain-gauge in a hilly area. Two island subdivisions far away from mainland have also not been included. Thus, the contiguous area having network of 306 stations over 29 meteorological subdivisions measures about 2,880,000 sq.km, which is about 90 percent of the total area of the country. The detailed description is given in Mooley et al., (1981); Parthasarthy et al., (1987, 1993, 1995); Pant et al., (1997).

2.1.1.2 All India Daily June to September Rainfall Series

IITM has derived time series of the daily averaged 1June to 30September rainfall of India using data of more than 300 rain-gauge stations well distributed over the whole of India. We used this data set kindly provided by IITM to study durations of strong and weak monsoon rainfall. Daily rainfall data for India as a whole for the period 1 June to 30 September 1901-1989 were prepared from grid data (Kripalani et al., 1991) and for the period 1990-2002 as updated by IITM from the All India Weather Summary prepared by the India Meteorological Department.
2.1.1.3 Summer Monsoon Rainfall for South and North Kerala

The rain-gauge network of Kerala whose data for the period 1901-1980 used in this study is shown in fig. 2.1. The dividing latitude between south and north Kerala is about 10°N. South Kerala has 44 rain gauge stations and north Kerala has 31 stations with long records of rainfall. Ananthakrishnan and Soman (1988) using the data of daily rainfall at these stations constructed daily rainfall series for south Kerala and north Kerala by averaging the rainfall at individual stations. This data set for 80 years (1901 to 1980) was donated to Cochin University of Science & Technology by Dr. Soman of Indian Institute of Tropical Meteorology (IITM). For the period 1981-1996, average of 39 stations in south Kerala and 24 stations in north Kerala were used. The daily data for the period (1981-1996) was obtained from the India Meteorological Department. Figure 2.2 gives the stations used for the period 1981-1996. The southwest monsoon rainfall of Kerala for the period 1901 to 1996 is calculated as an area-weighted average of south and north Kerala time series. This time series is used for studies of interannual variability, decadal variability and long-term trend of Kerala rainfall in chapter-4.

![Fig. 2.1: The rain-gauge network of Kerala for the period 1901-1980.](image-url)
2.1.1.4 Kerala Summer Monsoon Rainfall Data

The monthly (January - December) area weighted rainfall series for each of the 29 meteorological subdivisions is available in the IITM website for the period 1901-2003. The monsoon rainfall for Kerala (June-September) is obtained from this site. The sub-divisional rainfall has been prepared by assigning the district area as the weight for each rain-gauge station in that subdivision. The data for the recent period 1991-2003 are preliminary estimates based on the sub-divisional means supplied by India Meteorological Department (IMD), which are in turn based on a variable network. However, IMD data have been rescaled to conform to the long-term means of the respective subdivisions in the IITM-IMD data set. This data set of Kerala rainfall of the monsoon season is used in chapter-5 on Long Range Forecasting.

The climatological mean (1901-1950) value for July rainfall is obtained from IMD publication Annual and Seasonal Rainfall Normals and number of rainy days – PartV

2.1.1.5 Pentad Rainfall

Normal pentad rainfall of 160 stations over India and its adjoining sea areas for the entire year comprising 73 pentads is given in Ananthakrishnan et al.
The pentad rainfall values are an average of 50 years (1901-1950). The details of the stations selected in this study are given in Table 2.1.

Table 2.1: Details of pentad rainfall as taken from Ananthakrishnan et al. (1971)

<table>
<thead>
<tr>
<th>Station</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>Annual Rainfall (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trivandrum</td>
<td>08°29'N</td>
<td>76°57'E</td>
<td>64</td>
<td>1812</td>
</tr>
<tr>
<td>Allepey</td>
<td>09°33'N</td>
<td>76°20'E</td>
<td>2</td>
<td>3252</td>
</tr>
<tr>
<td>Cochin</td>
<td>09°56'N</td>
<td>76°14'E</td>
<td>3</td>
<td>3407</td>
</tr>
<tr>
<td>Palghat</td>
<td>10°46'N</td>
<td>76°39'E</td>
<td>97</td>
<td>2040</td>
</tr>
<tr>
<td>Kozhikode</td>
<td>11°15'N</td>
<td>75°47'E</td>
<td>5</td>
<td>3176</td>
</tr>
<tr>
<td>Manglore</td>
<td>12°52'N</td>
<td>74°51'E</td>
<td>22</td>
<td>3398</td>
</tr>
<tr>
<td>Mercara</td>
<td>12°25'N</td>
<td>75°44'E</td>
<td>1152</td>
<td>3265</td>
</tr>
<tr>
<td>Aminidivi</td>
<td>11°70'N</td>
<td>72°44'E</td>
<td>4</td>
<td>1504</td>
</tr>
<tr>
<td>Minicoy</td>
<td>08°18'N</td>
<td>73°00'E</td>
<td>2</td>
<td>1640</td>
</tr>
</tbody>
</table>

2.1.1.6 Hourly Rainfall Data

The hourly rainfall data is obtained from the International Hydrology project under the Kerala State Irrigation Department. Full Climatic Stations (FCS) under Water resources Department, established in 2000 under this project take measurements of parameters like rainfall (Siphon Rain Gauge), Pan evaporation, Maximum and Minimum Temperatures, Dry bulb Temperature, Wet bulb Temperatures, instant wind speed & daily average 2 times a day (8.30 am & 5.30 pm).

Continuous recording on graph are taken for some parameters like Sunshine duration, Temperature, Humidity and Rainfall and the graphical readings are converted into hourly values. Section offices under the department
maintain these stations, and the data is observed and validated by trained local personnel using software named SWDES. The data are validated at various levels and stored for future use in the department. We obtained the hourly rainfall for 10 stations in Kozhikode division, 14 stations in Thalassery division, 9 stations in Thrissur division and 9 stations in Chegannur division. Out of these 42 stations, we have utilized 33 stations with missing data less than 10% of the total data length. The station details are given in Chapter 6.

2.1.2 Break/Active Periods

Ramamurthy (1969) has studied the details of break period up to 1967. The main criteria used for the study is Monsoon trough running close to the foot hills of the Himalayas and absence of easterly winds to the north of monsoon trough in the lower tropospheric levels. De et al. (1998) identified break periods from 1968 to 1997 with the help of daily weather charts of India Meteorological Department. He has used the same criteria as by Ramamurthy (1969) with some additional features. During the period studied (1968-1997) there were 193 break days in 33 break spells. Most of these occurred in July and August. Joseph and Sijikumar (2004) have taken the break spells of July and August lasting 3 days or more for the twelve years 1979-1990. These break spells are listed in Table 2.2 and we have used this for our study. An active monsoon spell is defined arbitrarily as one in which for each day of the spell the area averaged zonal wind at 850 hPa in the latitude-longitude box 10°N-20°N and 70°E-80°E in a five day period centered on that day is 15 ms⁻¹ or more (Joseph and Sijikumar, 2004). We have chosen Active days according to this criterion as given by them. The details are given in Table 2.2.
Table 2.2 Active/Break days as given in Joseph and Sijikumar (2004)

<table>
<thead>
<tr>
<th>Year</th>
<th>Active Monsoon</th>
<th>Break Monsoon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>11Aug-16 Aug</td>
<td>-</td>
</tr>
<tr>
<td>1984</td>
<td>14Jun-18 Jun</td>
<td>20Jul - 24Jul</td>
</tr>
<tr>
<td>1985</td>
<td>-</td>
<td>22Aug-25Aug</td>
</tr>
<tr>
<td>1987</td>
<td>-</td>
<td>28Jul-01Aug</td>
</tr>
<tr>
<td>1989</td>
<td>21Jul-26Jul</td>
<td>10Jul-12Jul 29Jul-31Jul</td>
</tr>
<tr>
<td>1990</td>
<td>22Jun-08Jul</td>
<td>08Jul-10 Jul 27Jul-31Jul</td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td>84</td>
</tr>
</tbody>
</table>

2.1.3 Eurasian Snow Cover Extent

The monthly values of the Eurasian Snow cover extent (SCE) are obtained from the Rutgers University from their web site [http://climate.rutgers.edu/snowcover/](http://climate.rutgers.edu/snowcover/). The data set also contains time series of monthly snow cover extent (SCE) for North America and the Northern Hemisphere from 1966 to 2005, based on snow-cover reconstruction and NOAA satellite data. The reconstruction method used in situ snow depth and daily climate data from the U.S., Canada, China, and the former Soviet Union (FSU) to generate a monthly snow-cover index, which was closely related to satellite-derived estimates of SCE in certain months. Details of the reconstruction are given in Brown (2000). Snow cover between 1966 and 1971 was reanalyzed at the Rutgers University Climate Lab using daily gridded composites of visible
imagery for the eastern and western hemispheres of the Northern Hemisphere. Surface resolution of the imagery is approximately 25 km. The imagery was supplemented with daily reports of snow depth at several thousand stations in the U.S., Canada, China and the former Soviet Union, gridded to 1° x 1° grid cells using all reports from within a given cell. Daily surface weather charts also provided information on cloud cover, precipitation and temperature. Infrared imagery and the above ancillary information were employed in many areas to confirm interpretations made from visible data. The weekly maps were digitized to the National Meteorological Center Limited-Area Fine Mesh grid. This is an 89 x 89 cell Cartesian grid laid over a polar stereographic projection of the Northern Hemisphere. Cell resolution ranges from 16,000 to 42,000 square kilometers (this product has also been regridded to the equal area EASE-grid in a CD rom for 1978-2001 and is distributed by the National Snow and Ice Data Center). Each grid cell in the digitized product has a binary value. Cells with at least 50% of their surface covered with snow were considered snow covered. All other cells were considered snow free. The raw NOAA gridded 89x89 data and the GSL reanalysis gridded data are both utilized in creating a unique Northern Hemisphere snow cover product. Using Perl software created at the GSL, weekly and monthly 89x89 grid cell charts are generated. In this procedure, weekly areas are calculated from digitized snow files, and monthly values are calculated by weighting the weekly areas according to the number of days of a map week falling in the given month. During this process, the raw NOAA 89x89 grid data, as well as the reanalysis gridded data, are subjected to filtering through the corrected land mask created here at the GSL. The result is an accurate grid cell product which details Northern Hemisphere snow cover data over the last 38 years.

2.1.4 Quasi-Biennial Oscillation (QBO) Zonal Wind Index

The monthly values of the QBO zonal wind index for 30hPa is obtained from the web site: http://tao.atmos.washington.edu/data_sets/ for the period January 1953 - September 2001. The index is the concatenation of values at
Canton Island (3°S, 172°W) for Jan 1953 - Aug 1967; Gan/Maledives (1°S, 73°E) for Sep 1967 - Dec 1975; and Singapore (1°N, 104°E) for Jan 1976 - Sep 2001. Naujokat (1986) documents the data, uncertainties in the early years due to the lack of daily data, the change of reference stations, and the general features of the QBO. Marquardt and Naujokat (1997) describe the update of this index.

2.1.5 Sunspot number

Daily observations of sunspot numbers started at the Zurich Observatory in 1749 and with the addition of other observatories continuous observations were obtained starting in 1849. The sunspot number is calculated by first counting the number of sunspot groups and then the number of individual sunspots. The "sunspot number" is then given by the sum of the number of individual sunspots and ten times the number of groups. Since most sunspot groups have, on average, about ten spots, this formula for counting sunspots gives reliable numbers even when the observing conditions are less than ideal and small spots are hard to see. The monthly sunspot numbers are obtained from http://science.msfc.nasa.gov/ssl/pad/solar/sunspots.htm.

2.1.6 NCEP/NCAR Reanalysis

The Global NCEP/NCAR reanalysis data set (Kalnay et al., 1996) is used in this study. Reanalysis is different from the ‘traditional’ data sets in two fundamental ways: (1). an atmospheric general circulation model (AGCM) is an integral component of the analysis system and (2). a wide range of observations are used. Thus, the reanalysis not only gives potentially very useful dynamical quantities that cannot be determined by subjective analysis, but may be more accurate than such traditional analyses, particularly in the data sparse regions. However, the differences in the AGCMs and the analysis methods will give rise to differences in reanalysis. Several intercomparison studies have been made to realize the magnitude and nature of this ambiguity in NCEP/NCAR reanalysis.
The NCEP/NCAR reanalysis is a joint venture between NCEP and NCAR to produce a multi-decadal record of global atmospheric analysis with unchanged data assimilation system. The assimilation system used observations from the COADS surface marine data sets, the rawinsonde network, satellite soundings (the Tiros Operational Vertical Sounder, TOVS data), aircraft data and satellite (GMS, GOES and METEOSAT) cloud drift winds. These data were subject to stringent quality control; (Kalnay et al; 1996). The NCEP/NCAR Reanalysis has three major modules (1). Data decoder and quality control (QC) preprocessor (2). Data assimilation module with an automatic monitoring system and (3). Archive module (fig. 2.3).

The preprocessor minimizes the need for reanalysis re-runs due to the many data problems that frequently appear, such as data with wrong dates, satellite data with wrong longitudes etc. The preprocessor also includes the preparation of the surface boundary conditions (SST, Sea Ice etc). For the analysis module, the Spectral Statistical Interpolation Scheme (SSI) is used, which is a three dimensional variational technique (Derber et al 1991, Parrish and Derbur 1992). An important advantage of the SSI is that the balance imposed on the analysis is valid throughout the globe, thus making unnecessary the use of nonlinear normal mode initialization. Recent enhancements such as improved error statistics and the use of full tendency of the divergence equation in the cost function (replacing the original linear balance of the increments constraint), have also been included (Derber et al 1991, Parrish and Derbur 1992). A T62/28 level global spectral model corresponding to an approximate grid point spacing of 208 km, with 28 vertical levels was used in the assimilation system. The model has 5 levels in the boundary layer and about 7 levels above 100hPa. The lowest model level is about 5hPa from the surface and the top level is at about 3hPa. The reanalysis gridded fields have been classified into four classes, depending upon the relative influences of the observational data and the model on the gridded variable (Table 2.3).
Reanalysis outputs are available in 17 standard pressure levels (hPa), 11 isentropic surfaces (K) and 28 sigma levels. The horizontal resolution is 2.5° longitude 2.5° latitude. The standard pressure levels (hPa) are 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10.

The parameters used from NCEP/NCAR data sets are zonal (U) and Meridional (V) wind at 850hPa, 200hPa and 50hPa pressure levels, Vertically Integrated Moisture (VIM), 2-m air temperature and Sea Surface Temperature (SST). The wind data has a rating A, which means that they are strongly influenced by the observed data and the influence of the model used to derive the grid point values is minimal, whereas VIM and 2-m temperature has a rating B. SST analysis is on a nearly 1.9° x 1.9° latitude-longitude grid. The analysis is produced both daily and weekly, using 7 days of in situ data (ship and buoy) and bias-corrected satellite SST data. SST data from the UK Met office are used up to 1988 after which SST from Reynolds (CPC) are used. Since there is strong trend in SST we have calculated SST anomalies for each month removing a 11-year running mean to get a monthly data series for 1955 to 1998.

2.1.7 NOAA OLR Data

Originally the data are from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA Polar Orbiting Spacecraft. The data are taken from the Interpolated OLR Data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA from their website (http://www.cdc.noaa.gov). The OLR data for a period 1974-2003 is used, with the exception of 1978. The data contains a major gap of several months during 1978 due to the failure of satellite. The data resolution are at 2.5° × 2.5° latitude-longitude (Gruber and Krueger, 1984). The data is in Wm⁻².
Fig. 2.3: Schematic illustration of the main components of the NCEP/NCAR Reanalysis system (NMC has changed to NCEP) (Kalnay et al, 1996)
Table 2.3: Classification of NCEP/NCAR reanalyzed fields.

<table>
<thead>
<tr>
<th>Class</th>
<th>Relative influence of Observational Data and Model on Reanalysis Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Strongly influenced by observational data (most reliable)</td>
</tr>
<tr>
<td></td>
<td>[e.g. upper air temperature and wind]</td>
</tr>
<tr>
<td>B</td>
<td>Model has very strong influence than observational data</td>
</tr>
<tr>
<td></td>
<td>[e.g. humidity and surface temperature ]</td>
</tr>
<tr>
<td>C</td>
<td>Derived solely from model fields forced by data assimilation to remain close to the atmosphere.</td>
</tr>
<tr>
<td></td>
<td>[e.g. clouds, precipitation, and surface fluxes]</td>
</tr>
<tr>
<td>D</td>
<td>Obtained from climatological values and does not depend on model</td>
</tr>
<tr>
<td></td>
<td>[e.g. plant resistance, land-sea mask]</td>
</tr>
</tbody>
</table>

2.1.8 Mesoscale Modeling

The Fifth-Generation NCAR/Penn State Mesoscale Model MM5 (Version 3) is used to study the orographic control on Kerala’s monsoon rainfall. The MM5 is the latest in a series that developed from a mesoscale model used by Anthes at Penn State in the early ’70’s that was later documented by Anthes and Warner (1978). It is a three dimensional nonhydrostatic model with multiple-nest capability and a four-dimensional data assimilation capability (Dudhia, 1993).

The model is supported by several auxiliary programs, which are referred to collectively as the MM5 modeling system. A schematic diagram representing
complete modeling system is shown in figure. 2.4. Terrestrial and isobaric meteorological data are horizontally interpolated (programs TERRAIN and REGRID) from a latitude-longitude mesh to a variable high-resolution domain on either a Mercator, Lambert conformal, or polar stereographic projection. Since the interpolation does not provide mesoscale detail, the interpolated data may be enhanced (program RAWINS or little r) with observations from the standard network of surface and rawinsonde stations using either a successive-scan Cressman technique or multiquadric scheme. Program INTERPF performs the vertical interpolation from pressure levels to the sigma coordinate system of MM5. Sigma surfaces near the ground closely follow the terrain, and the higher-level sigma surfaces tend to approximate isobaric surfaces. Since the vertical and horizontal resolution and domain size are variable, the modeling package programs employ parameterized dimensions requiring a variable amount of core memory. Some peripheral storage devices are also used (PSU/NCAR MM5 User's guide, 2002).

The modeling system usually gets and analyzes its data on pressure surfaces, but these have to be interpolated to the model's vertical coordinate before being input to the model. The vertical coordinate of the MM5 model is terrain following (fig.2.5) meaning that the lower grid levels follow the terrain while the upper surface is flat. Intermediate levels progressively flatten as the pressure decreases toward the chosen top pressure.

A dimensionless quantity $\sigma$ is used to define the model levels where

$$\sigma = \frac{p - pt}{ps - pt}$$

(2.1)

$p$ is the pressure, $pt$ is a specified constant top pressure, $ps$ is the surface pressure.
It can be seen from the equation and figure 2.5 that \( \sigma \) is zero at the top and one at the surface, and each model level is defined by a value of \( \sigma \). The model vertical resolution is defined by a list of values between zero and one that do not necessarily have to be evenly spaced. Commonly the resolution in the boundary layer is much finer than above, and the number of levels may vary from ten to forty, although there is no limit in principle.

The horizontal grid has an Arakawa-Lamb B-staggering of the velocity variables with respect to the scalars. This is shown in figure 2.6 where it can be seen that the scalars (T, q etc.) are defined at the center of the grid square, while the eastward (u) and northward (v) velocity components are collocated at the corners. The center points of the grid squares will be referred to as cross points,
and the corner points are dot points. Hence horizontal velocity is defined at dot points, for example, and when data is input to the model the preprocessors do the necessary interpolations to assure consistency with the grid.

MM5 contains a capability of multiple nesting with up to nine domains running at the same time and completely interacting. The nesting ratio is always 3:1 for two-way interaction. "Two-way interaction" means that the nest's input from the coarse mesh comes via its boundaries, while the feedback to the coarser mesh occurs over the nest interior.

To run any regional numerical weather prediction model requires lateral boundary conditions. In MM5 all four boundaries have specified horizontal winds, temperature, pressure and moisture fields, and can have specified microphysical field (such as cloud) if these are available. Therefore, prior to running a simulation, boundary values have to be set in addition to initial values for these fields. The boundary values come from analysis at the future times, or a previous coarser-mesh simulation (1-way nest), or from another model's forecast (in real-time forecasts). For real-time forecasts the lateral boundaries will ultimately depend on a global-model forecast. In studies of past cases the analysis providing the boundary conditions may be enhanced by observation analysis (RAWINS or little r) in the same way as initial conditions are. Where upper-air analysis are used the boundary values may only be available 12-hourly, while for model-generated boundary conditions it may be a higher frequency like 6-hourly or even 1-hourly.

The model uses this discrete-time analysis by linearly interpolating them in time to the model time. The analysis completely specifies the behavior of the outer row and column of the model grid. In the next four rows and columns in from the boundary, the model is nudged towards the analysis, and there is also a smoothing term. The strength of this nudging decreases linearly away from the boundaries. To apply this condition, the model uses a boundary file with information for the five points nearest each of the four boundaries at each
boundary time. This is a rim of points from the future analysis described above. The interior values from these analyses are not required unless data assimilation by grid-nudging is being performed, so disk-space is saved by having the boundary file just contain the rim values for each filed. Two-way nest boundaries are similar but are updated every coarse-mesh time step and have no relaxation zone. The specified zone is two grid-points wide instead of one.

MM5 modeling system requires the following data sets to run

- Topography and land use.

- Gridded atmospheric data that have at least the variables; sea-level pressure, wind, temperature, relative humidity and geopotential height at the pressure levels: surface, 1000, 850, 700, 500, 400, 300, 250, 200, 150, 100 mb.

- Observation data that contains soundings and surface reports

Three resolutions of elevation data; 30 minute, 10 minute, and 5 minute are used for this present study. All these data are created from the 30 seconds United States Geological Surveys (USGS) data. Vegetation/land-use data are also in the same resolution of elevation data and from USGS version 2 land-cover data. Data from NCEP/NCAR Reanalysis Projects (NNRP) are used for the atmospheric variables. The NNRP data are at resolution 2.5° x 2.5° latitude-longitude.
Figure 2.5: Schematic representation of the vertical structure of the model.

Dashed lines denote half-sigma levels, solid lines denote full-sigma levels.
Figure 2.6: Schematic representation showing the horizontal Arakawa B-grid staggering of the dot and cross grid points. The smaller inner box is a representative mesh staggering for a 3:1 coarse-grid distance to fine-grid distance ratio.

2.2 Methodology

2.2.1 Wavelet Analysis

Wavelet analysis tool is well suited to study multiscale, non-stationary processes occurring over finite spatial and temporal domain. Since its introduction by Morlet (1983), this technique has found wide application in diverse fields. Wavelet analysis gives the localised variations of power within a time scale by decomposing a time series in time-frequency space and one is able to determine both the dominant modes of variability and how those modes vary in time. Morlet wavelet is defined as the product of a complex exponent wave and a Gaussian envelope:

$$\varphi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^{2/3}}$$

2.2
where $q_0$ is the wavelet value at non-dimensional time $\eta$, and $w_0$ is the wavenumber. The "scaled wavelet" is

$$\varphi\left[\frac{(n'-n)\partial t}{s}\right] = \left(\frac{\partial t}{s}\right)^{1/2} \varphi_o\left[\frac{(n'-n)\partial t}{s}\right]$$

where $s$ is the "dilation" parameter used to change the scale, and $n$ is the translation parameter used to slide in time. The factor of $s^{-1/2}$ is a normalization to keep the total energy of the scaled wavelet constant. Details regarding the analysis procedure are given in Torrence and Compo (1998).

For the Morlet wavelet transform, where the mother wavelet is given by equation 2.2 we first choose the wavenumber $w_0$, which gives the number of oscillations within the wavelet itself. One condition of the wavelet transform is that the average of the wavelet itself must be zero. In practice, if we choose $w_0=6$, then the errors due to non-zero mean are smaller than the typical computer round-off errors (Farge 1992).

In order to adequately sample all the frequencies present in our time series, we have chosen the smallest resolvable scale, $s_0$, which is a multiple of our time resolution, as 2 days. The larger scales (longer periods) are chosen as power-of-two multiples of this smallest scale,

$$s_j = s_0 2^{j\rho}, j = 1, 1, \ldots, j$$

$$j = \frac{\log_2 (N\partial t/s_0)}{\rho}$$

The choice of octave which is logarithmic with the base of 2 as a unit to divide the frequency domain allows us to include a broad range of scales, from very small to very large in an efficient way in a coordinate system with linear interval in octave while logarithmic in frequency scale. In a continuous WT
where more scale decomposition is desired, each octave may be further divided into infinite voices. For this study we have divided each octave into 4 voices.

To determine significance we choose an appropriate background spectrum as follows:

\[
P_k = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos(2\pi k / N)}
\]

where \( k = 0 \ldots N/2 \) is the frequency index and \( \alpha \) is the lag-1 autocorrelation.

The null hypothesis is defined for the wavelet power spectrum as follows: It is assumed that the time series has a mean power spectrum, given by 2.3; if a peak in the wavelet power spectrum is significantly above this background spectrum, then it can be assumed to be a true feature with a certain percent confidence 90%, 95% or 99%.

2.2.2 Harmonic Analysis

Harmonic analysis consists of representing the fluctuations or variations in a time series as having arisen from the adding together a series of sine and cosine functions. These trigonometric functions are "harmonic" in the sense that they are chosen to have frequencies exhibiting integer multiples of the "fundamental" frequency determined by the sample size of the data series. A common physical analogy is the musical sound produced by a vibrating string, where the pitch is determined by the fundamental frequency, but the aesthetic quality of the sound depends also on the relative contributions of the higher harmonics.

A time series having a sinusoidal character and executing a single cycle over the course of \( n \) observations can be represented as:

\[
y_t = y + C_1 \cos\left(\frac{2\pi}{n}\right)
\]

64
The equation can be written in terms of the phase angle or phase shift $\phi_1$, as

$$y_t = y + C_1 \cos\left(\frac{2\pi t}{n} - \phi_1\right)$$  \hspace{1cm}  \text{(2.5)}$$

The trigonometric identity

$$\cos(\alpha - \phi_1) = \cos(\phi_1) \cos(\alpha) + \sin(\phi_1) \sin(\alpha)$$  \hspace{1cm}  \text{(2.6)}$$

used to compute $C_1$ and $\phi_1$.

Substituting $\alpha = \frac{2\pi t}{n}$ and multiplying both sides by the amplitude $C_1$, we obtain

$$C_1 \cos\left(\frac{2\pi t}{n} - \phi_1\right) = C_1 \cos(\phi_1) \cos\left(\frac{2\pi t}{n}\right) + C_1 \sin(\phi_1) \sin\left(\frac{2\pi t}{n}\right)$$  \hspace{1cm}  \text{(2.7)}$$

$$= A_1 \cos\left(\frac{2\pi t}{n}\right) + B_1 \sin\left(\frac{2\pi t}{n}\right)$$  \hspace{1cm}  \text{(2.8)}$$

where

$$A_1 = C_1 \cos(\phi_1)$$  \hspace{1cm}  \text{(2.9)}$$

and

$$B_1 = C_1 \sin(\phi_1)$$  \hspace{1cm}  \text{(2.10)}$$

The sine and cosine trigonometric functions are periodic, effectively the same phase angle is produced by adding or subtracting a half-circle of angular measure if $A_1 < 0$. The alternative that produces $0 < \phi_1 < 2\pi$ is usually selected.

Equation (2.7) says that it is possible to mathematically represent a harmonic wave either as a cosine function with amplitude $C_1$ and phase $\phi_1$ or as the sum of an unshifted cosine and unshifted sine wave with amplitudes $A_1$ and $B_1$. The amplitude $C_1$ is then

$$C_1 = \left[A^2 + B^2\right]^{1/2}$$  \hspace{1cm}  \text{(2.11)}$$

and phase $\phi_1$. 

65
\[
\phi_i = \begin{cases} 
\tan^{-1} \frac{B_i}{A_i}, & A_i > 0 \\
\tan^{-1} \frac{B_i}{A_i} \pm \pi, \text{ or } \pm 180^0, & A_i < 0 \\
\frac{\pi}{2}, \text{ or } 90^0, & A_i = 0
\end{cases}
\]  
\text{2.12}

Usually \(A_1\) and \(B_1\) are computed from the following equations

\[
A_i = \frac{2}{n} \sum_{i=1}^{n} y_i \cos \left( \frac{2\pi}{n} \right) \]  
\text{2.13}

\[
B_i = \frac{2}{n} \sum_{i=1}^{n} y_i \sin \left( \frac{2\pi}{n} \right) \]  
\text{2.14}