CHAPTER 2

OBSERVATIONS, MODEL AND ANALYSIS

2.1 Introduction

The approach in this thesis is to employ and synthesize analysis based on observed climate data, simulations from comprehensive model (Atmospheric General Circulation Model AGCM) and robust analysis.

This chapter describes the different sources and techniques adopted for the observational analysis, model configuration and the analysis procedure.

2.2 Observed data

We have considered multi-source data at different scales to ensure robustness of the results and present multiple lines of evidence. The major data sets used are outlined below.

2.2.1 NCEP/NCAR reanalysis data

The National Centres for Environmental Prediction (NCEP) and National Centre for Atmospheric Research (NCAR) have initiated a project (denoted “reanalysis”) to produce a retroactive record of more than 50 years of global analyses of atmospheric fields in support of the needs of the research and climate monitoring communities. This effort involved the recovery of land surface, ship, rawinsonde, pibal, aircraft, satellite, and other data. These data were then quality controlled and assimilated with a data assimilation system kept unchanged over the reanalysis period.
Chapter 2 Observations, Model and Analysis

During the earliest decade (1948–57), there were fewer upper-air data observations and they were made 3 h later than the current main synoptic times (e.g., 0300 UTC), and primarily in the Northern Hemisphere, so that the reanalysis is less reliable than for the later 40 years. The reanalysis data assimilation system continues to be used with current data in real time (Climate Data Assimilation System or CDAS), so that its products are available from 1948 to the present. The products include, in addition to the gridded reanalysis fields, 8-day forecasts every 5 days, and the binary universal format representation (BUFR) archive of the atmospheric observations. The products can be obtained from NCAR, NCEP, and from the National Oceanic and Atmospheric Administration/Climate Diagnostics Center (NOAA/CDC). For examining the characteristics and circulation anomalies associated with the active/breakevents, Rajeevan et al. (2010) have used the daily re-analysis data of NCEP/NCAR (Kalnay et al. 1996). The daily data of rainfall, mean sea level pressure, zonal and meridional wind at 1000 hPa, 850 hPa and 200 hPa levels and relative humidity data were used for the analysis. First time, this study has elucidated the deference in the vertical meridional circulation between the active spells with moist convection and intense break events with a heat trough type circulation with the help of NCEP-NCAR reanalysis data. Malik et al. (2011) used zonal and meridional wind components and surface rainfall data from the NCEP/NCAR reanalysis dataset from 1951 to 2004 with 2.5 degree resolution for analysis of spatial and temporal extreme monsoonal rainfall over south Asia using complex networks. Dimri (2012) used 2.5° resolution NCEP Reanalysis II wind speed data for model evaluation and performance check.

2.2.2 APHRODITE data

A daily gridded precipitation dataset covering a period of more than 57 year was created by collecting and analyzing rain gauge observation data across Asia through the activities of the Asian Precipitation—Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) project (Yatagai 2009). APHRODITE’s daily gridded precipitation is presently the only long-term, continental-scale, high-resolution daily product. These data products are based on data collected from 5,000–12,000 stations through the Global Telecommunication system network and is used for most daily gridded precipitation products. The basic algorithm for development of this data set is based on Xie et al. (2007). Under APHRODITE project a high resolution (0.25 degree X 0.25 degree and 0.5 X 0.5
Chapter-2 Observations, Model and Analysis

degree) daily rainfall data set was developed for the Asian region. The APHRODITE project has substantially improved the depiction of the areal distribution and variability of precipitation around the Himalayas, southeast Asia, and mountainous regions of the Middle East. The APHRODITE project now contributes to studies such as the determination of Asian monsoon precipitation change, evaluation of water resources, verification of high-resolution model simulations and satellite precipitation estimates and improvement of precipitation forecasts. The APHRODITE project carries out outreach activities with Asian countries, and communicates with national institutions and world data centres. Rajeevan and B hate (2008) compared the IMD high resolution gridded data with APHRODITE rainfall data. Correlation between IMD gridded and APHRODITE data is very high over central and northwest India. However correlation is very low along the west coast of India and some parts of northeast India. The APHRODITE analysis underestimates the rainfall maximum along the west coast and NE India. Otherwise, the differences are mostly within 3 mm/day within the country. Malik et al. (2011) used daily, gridded APHRODITE rainfall data from 1951 to 2007 for analysis of spatial and temporal extreme monsoonal rainfall over south Asia using complex networks. Dimri (2012) used high resolution APHRODITE rainfall and temperature to see diurnal variation in surface temperature at Gulmarg and area averaged precipitation in two climate extremes. Ali et al. (2012) compared APHRODITE rainfall data set with daily real observation data for humid and sub humid regions of Pakistan and reported that for all of the stations, the APHRODITE data underestimates the observed precipitation, particularly in those months in which the precipitation amount is high, with a maximum bias of 271mm for Muzaffarabad for the month of July in the decade 1991–2000. A slight overestimation of APHRODITE data was also observed for some stations (Islamabad, Garhi, Dupatta and Peshawar) with a maximum bias of 40mm for Islamabad in March in the decade 1971–1980. The correlation coefficient values between the two daily datasets varies from very poor (0.05 to 0.15 for Islamabad) to very good (e.g. 0.79 to 0.99 for Astore) for different stations in different decades. The correlation coefficient values become much better in case of mean monthly datasets analysis. Even the very poor results (for example Islamabad) rises to very good (0.98 to 0.99) in this case. Thus it is concluded that observed and APHRODITE datasets are strongly correlated on mean monthly basis as compared to daily basis.
2.2.3 IMD gridded data

Information about spatial and temporal variability of rainfall is very important in understanding the hydrological balance on a global/regional scale. The distribution of precipitation is also important for water management for agriculture, power generation, and drought monitoring. In India peninsula, rainfall received during the monsoon season (June to September) is very crucial for its economy. Real time observation of rainfall distribution on daily basis is required to evaluate the progress and status of monsoon rainfall and to initiate necessary action to control the extreme rainfall events. To fulfil this demand of the research community, India Meteorological Department has developed a high resolution (1° X 1° Lat/Long) gridded daily rainfall data set for the Indian region. For this analysis, daily rainfall data of 6329 stations were considered. Out of these 6329 stations, 537 stations are the IMD observatory stations, 522 stations are under the Hydrometeorology programme and 70 are Agromet stations. Remaining stations are rainfall-reporting stations maintained by state governments. However, there were only 1803 stations with a minimum 90% data availability during the analysis period (1951-2003). IMD has used only those 1803 stations for interpolation in order to minimize the risk of generating temporal inhomogeneity in the data set due to varying station densities. The geographical area, 6.5°N to 37.5°N, 66.5°E to 101.5°E was considered for interpolating the station rainfall data. For the analysis, IMD has followed Shepard (1968) interpolation method. This method is based on the weights calculated from the distance between the station and the grid point and also the directional effects. Standard quality controls were performed before carrying out the interpolation analysis.

Rajeevan and Bhate (2008) have developed a very high resolution daily gridded (0.5° X 0.5° resolution) rainfall data for the Indian region for the period 1971-2005. The high resolution data set has been developed using daily rainfall data from more than 6000 stations from the country. The quality of the data set was tested by comparing the present analysis with other gridded data set developed in Japan. Goswami et al. (2006) used IMD gridded rainfall data for the period 1951-2003 for find out the trend of frequency and magnitude of extreme events. Ghosh et al. (1999) have also used IMD gridded data for study to identify the places that have a significant trend in terms of both rainfall amount and occurrence. Sabin et al. (2013),
Goswami and Gouda (2009, 2010) also used IMD gridded data for validation the variable resolution global climate model simulation of the south Asian monsoon.

2.2.4 IMD Station Data

The IMD has climatological records even for the period prior to 1875, when it formally came into existence. This data is digitized, quality controlled and archived in electronic media at the National Data Centre, Pune. The IMD prepared climatological tables and summaries/atlases of surface and upper-air meteorological parameters and marine meteorological summaries. These climatological summaries and publications have many applications in agriculture, shipping, transport, water resources and industry. IMD station data for Uttarakhand region is used in this study. This data is in ASCII format. We have used rainfall, temperature, wind speed, wind direction and humidity. We are using data from 1965 onwards over the stations in Uttarakhand, over some stations very long term data staring 1901 also considered in this study.

2.2.5 CPC Merged Analysis of Precipitation (CMAP)

The CPC Merged Analysis of Precipitation (CMAP) is a technique which produces pentad and monthly analyses of global precipitation in which observations from rain gauges are merged with precipitation estimates from several satellite-based algorithms (infrared and microwave). The analyses are on a 2.5°×2.5° latitude/longitude grid with global coverage (90° N-90°S, 0°-360°E) and extend back to 1979. These data sets are comparable similarly combined analyses by the Global Precipitation Climatology Project described by Huffman et al. (1997). This data set consists of monthly averaged precipitation rate values (mm/day). We have used CMAP Precipitation monthly data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd.

2.2.6 MODIS Water Vapour

Moderate Resolution Imaging Spectro-radiometer (MODIS) is an advanced multi-purpose NASA sensor. MODIS is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra passes from north to south across the equator in the morning (10:30 local time), whereas Aqua passes south to north over the equator in the afternoon (13:30 local time). Terra has 36 spectral bands between 0.405µm and 14.385µm. Among the 36 bands, daily and seasonal variations of column atmospheric
water vapour have been observed from MODIS bands located within and around the 0.94μm water level. In our study, we have used Level-3 MODIS Terra (MOD08_M3) monthly and (MOD08_D3) daily global product (near infrared water vapour with clear column) available at spatial resolution of 1 degree \(\text{http://modis-atmos.gsfc.nasa.gov}\) for water vapour analysis in the present study. Gao and Kaufman (1998) have described that MODIS water vapour estimation may have an error up to 10% especially in clouds.

2.3 Model Configuration: Variable Resolution GCM

2.3.1 Model Configuration

We adopted a General Circulation Model (GCM) with variable resolution, which generally allows relatively higher resolution over a chosen domain in a continuous and dynamically consistent manner. The basic formulation and validation of the model adopted here (LMDZ.3 from LMD, France), including formulation of variable resolution, have been described in detail by Sharma et al. (1987), Hourdin et al. (2006), Sadourny and Laval (1984), Laval et al. (1996) and Sabre et al. (2000). They have used a uniform grid to simulate monsoon depressions. The variable resolution, prescribed as a continuous variation (stretching) in the resolution with respect to a point, is implemented through a function like a sine or \(\tanh\). The version used in this study has 144X192 points in the horizontal direction (latitude X longitude) which provides a global resolution of about 2° X 2° in the uniform grid. The vertical coordinate is a hybrid sigma system with 19 levels in the present version. The model physics in the present study uses a diurnal cycle, a land surface module, and the convective parameterization scheme of Tiedtke (1989). The variable resolution version with climatological SST was recently employed to study aspects of monsoon like choice of ensemble for high resolution and long-range forecasting (Goswami and Gouda, 2009) and prediction of date of onset of monsoon (Goswami and Gouda, 2010).

It is well known that GCM simulations are sensitive to the choice of resolution (Lal et al. 1997; Sperber et al. 1994), parameterization scheme (Eitzen and Randall 1999) and other parameters. However, optimization of the model (grid) configurations can significantly improve skill. In this study, we have used optimized model configuration adopted in earlier study for the simulation of Indian monsoon
Chapter-2 Observations, Model and Analysis

(Goswami and Gouda, 2009), zoomed over 15°N, 75°E. The zoom provides a horizontal resolution of about 60 km X 50 km (in longitude and latitude) around the centre, which merged smoothly to about 2° X 2° away from the zoom; a part of the variable grid over the monsoon region is shown in the figure 2.1. With this configuration the study area i.e. Central Himalayan Uttarakhand also has the high resolution.

2.3.2 Physical Processes in GCM

2.3.2.1 Moist Convection

Convection is a crucial driver of the atmospheric circulation and is a key process for the vertical distribution of energy. It is responsible for the bulk of global precipitation. Any atmospheric prediction model with a resolution coarser than a few kilometers needs a convection parameterization to treat non resolved smaller clouds. Sometimes a differentiation between schemes and concepts is not straight forward, e.g. the “moisture adjustment” scheme of Manabe et al. (1965) is both a scheme, but also a concept on which other schemes are based. Organization of convective clouds is one of the most outstanding issues for convection parameterizations. All schemes perform an adjustment of moisture and energy by redistribution and precipitation processes.

Convection scheme of Tiedtke (1989) is adopted in the model and these schemes are mass flux schemes but they are different in characteristic closure and triggering of convection. The scheme of Tiedtke takes moisture convergence for the closure and the triggering of convection is due to moisture convergence and convective instability. The large-scale condensation is called immediately after moist convection. The physical parameterization package also includes a prognostic equation for cloud water. It solves two prognostic equations for liquid/ice water content and cloud fraction.
Figure 2.1: Part of the structure of the zoom grid with the centre of zoom at 15°N and 75°E. The total number of grid points is 144X192 latitude and longitude; the highest resolution near the centre of the zoom is ~50 km (variable), which merges to a uniform 2° X 2° away from zoom (stretched coordinate).

2.3.2.2 Radiation Scheme

The energy balance of the earth is fundamental for the understanding of the genesis and evolution of the global circulation. The earth's thermal conditions are determined by the absorption of solar radiation and the emission of terrestrial radiation. The pole-ward transport of heat in the atmosphere and oceans and the global hydrological cycle are driven by the geographical distribution of the energy balance components.

Recent investigations have brought to light substantial uncertainties in the distribution of solar energy within the climate system and its representation in GCMs. Particularly, the partitioning of energy absorption between the atmosphere and surface, and between the cloud-free atmosphere and clouds is insufficiently established and currently disputed. We use a combination of the best available models and collocated surface and top of atmosphere measurements for the model simulation.
2.3.2.3 Land Surface Processes

The model uses a diurnal land-surface module. Atmosphere-land interactions are also quite important for simulation and prediction of atmospheric processes. Unlike SST, forcing functions associated with land surface processes are very complex because of the interaction among ground temperature, soil moisture, and all components of surface energy budget. In GCM experiments, it is difficult to simply treat land surface processes in a universal way and results are likely to differ depending on what land surface parameters are specified. Furthermore, in the real system, SST and land surface processes themselves are likely to be mutually interactive (Meehl, 1994, 1997; Yang, 1996).

2.3.3 Initial and Boundary Conditions

In the present study, three initial conditions (i.e. APR-01, APR-15 and MAY-01) from NCEP reanalysis are used to generate the simulation and ensemble of these three member initial conditions were considered using the ensemble methodology adopted by earlier study (Goswami and Gouda, 2009) to identify an optimum choice of the ensembles for the long range rainfall simulatio. For validating the model simulated parameters over the study region, different observational datasets have been used. These include the daily gridded rainfall data from India Meteorological Department (Rajeevan et al., 2006) which is available in 1°X1° latitude-longitude grid over India for the period (1951-2003), APHRODITE data and station level data collected from IMD National Climate Data Centre.

2.3.4 Simulation Methodology

In our study we have consider ensemble of simulation carried out by three initial conditions (APR-01, APR-15 and MAY-01) adopted from the NCEP reanalysis.

The ensemble is done by using the following algorithm

\[ R_{\text{ENS}}(i,j,t) = \frac{1}{m} \sum_{i=1}^{m} R_{m}(i,j,t) \]  (2.1)

Where \( R_{m}(i,j,t) \) the simulation is obtained using the \( m \) initial condition

The analysis of the simulation is carried out by the following methods.
Chapter 2 Observations, Model and Analysis

To avoid effects due to bias between observation and simulation, we consider anomalies in rainfall as % of respective mean

\[ R_s(n) = \frac{R(n) - \bar{R}}{\bar{R}} \times 100 \]  

(2.2)

In addition to the correlation coefficients (CC) between observations and simulation, we shall also consider phase synchronization between the observed and the simulated anomalies:

\[ P(n) = \frac{R_{ao}(n) R_{as}(n)}{|R_{ao}(n)| |R_{as}(n)|} = \begin{cases} 1, & P(n) \geq 0 \\ 0, & other\wise \end{cases} \]  

(2.3)

Thus phase synchronization for year \( n \) is 1 if both anomalies have the same sign. It was noticed that the two observations (IMD and APHRODITE) showed significant differences; the differences in the basic statistical quantities like the mean and the standard deviation in different observations are quite high. In what follows, therefore, we shall consider composite observations to minimize the uncertainty in evaluation due to spread in observations. We have considered composite formed through an equal-weight averaging for both observations and simulations.

2.4 Analysis and Validation

The following parameters were considered for evaluation and statistical analysis.

2.4.1 Statistical Measures

Mean

The sample mean is a measure of the central tendency of a given data set. If \( x_1, x_2, x_3, \ldots, x_N \) represent the sequence of the observations where \( N \) is the number of observations. So mean is calculated as

\[ \bar{X} = \frac{1}{N} \sum_{j=1}^{N} x_j \]  

(2.4)

Where \( \bar{X} \) is the sample mean.

Mean Monthly value

\[ \bar{X}(m) = \frac{1}{N_y} \sum X(m, n_y) \]  

(2.5)
Chapter 2 Observations, Model and Analysis

Where \( X(m, n_y) \) represents monthly value of the variable such as rainfall or temperature for month \( m \) and year \( n_y \).

**Mean Seasonal value**

\[
\bar{X}(s) = \frac{1}{N_y} \sum X(s, n_y)
\]  

(2.6)

Where \( X(s, n_y) \) represents seasonal value of the variable such as rainfall or temperature for season \( s \) and year \( n_y \).

**Standard Deviation**

The standard deviation measures the dispersion of sample values around the mean. If \( x_1, x_2, x_3, \ldots, x_N \) represent the sequence of the observations, where \( N \) is the total number of observations, the unbiased estimate of standard deviation of this sequence may be determined as:

\[
\sigma = \left[ \frac{1}{N-1} \sum_{j=1}^{N} (x_j - \bar{x})^2 \right]^{1/2}
\]  

(2.7)

Variance is the square of standard deviation. The coefficient of variance is a dimensionless dispersion parameter equal to the ratio of the standard deviation and the mean.

So coefficient of variance is represents as:

\[
CV = \frac{\sigma}{\bar{x}}
\]  

(2.8)

**Correlation co-efficient**

The relation between the random variables \( x \) and \( y \) is called the simple or bivariate correlation. The correlation coefficient is the most commonly used statistical parameter for measuring the degree of association of two linearly dependent variables. Correlation coefficient is defined as

\[
r = \frac{\sum_{j=1}^{N} x_j y_j - N \bar{x} \bar{y}}{S_x S_y (N-1)}
\]  

(2.9)
Chapter 2 Observations, Model and Analysis

Where \( S_x \) and \( S_y \) are the standard deviation of \( x_i \) and \( y_i \), respectively. Correlation coefficient \( r \) varies from -1 to +1. The correlation coefficient is unity only if all points fall on a straight line. A positive value of \( r \) means that \( y \) increases with the increase of \( x \). A negative value of \( r \) means that \( y \) decreases with an increase of \( x \). If the points \((x_i, y_i)\) falls along a circle, \( r \) is zero. However in this case there is a high correlation for the function which has circle as the regression line.

In the present study the correlation analysis have been carried out for following cases

1. The correlation between observed sources to see the relation between the multi-source data.
2. Correlation between observed and simulation analysis.
3. In the present case, we consider seasonal cycle and anomaly for the whole Uttarakhand and at different station scale. The seasonal and monthly cycle are considered. The significance of correlation coefficient also presented in the result.

2.4.2 Analysis Methodology

2.4.2.1 Climatology

Climatology is commonly known as the study of our climate. Climatology is also defined as the long-term average of a given variable, often over longer time periods of 50-100 years. Climatologies are frequently employed in the atmospheric sciences, and may be computed for a variety of time ranges. A monthly climatology produce a mean value for each month and a daily climatology produce a mean value for each day, over a specified time range. So climatology is the long-term average of a variable. It is a periodic mean function of the data. It is computed over a variety of time ranges (daily, monthly and decade).

The climatology of a parameter \( p(x, y, t) \) is represented as

\[
\overline{p}(x, y) = \frac{1}{N} \sum_{t=1}^{N} p(x, y, t) \tag{2.10}
\]

where \( t \) is the time (in day, month or year) and \( N \) is the no. of time steps.

2.4.2.2 Trend Analysis

Trend analysis is an active area of interest for both hydrology and climatology in order to investigate climate change scenarios and enhance climate impact research.
Trend detection in precipitation time series is crucial for planning and designing regional water resource management.

(a) Linear Trend

A common form of a linear equation in the two variables \( x \) and \( y \) is

\[
y = mx + c
\]

(2.11)

Where \( m \) and \( c \) are constants. The origin of the name linear comes from the fact that the set of solutions of such an equation forms a straight line in the plane. In this particular equation, the constant \( m \) determines the slope or gradient of that line and \( c \) the constant term determines the point at which the line crosses the \( y \)-axis. The different methods adopted for the trend analysis are given below:

(i) Sen’s Method

Sen’s method for the estimation of slope requires a time series of equally spaced data. Sen’s method proceeds by calculating the slope as a change in measurement per unit change in time, the slope between measurements collected at the same time is not calculated. Sen’s estimator of slope is simply given by the median slope, shown below as:

\[
\text{Sen’s Estimator of slope} = \text{median slope} = Q'
\]

\[
= Q'_{(N'+1)/2} \text{ if } N' \text{ is odd,}
\]

\[
= (Q'_{(N'/2)} + Q'_{((N'+2)/2)})/2 \text{ if } N' \text{ is even}
\]

Where \( N' \) = number of calculated slopes

Sen’s Method also allows determination of whether the median slope is statistically different from zero. A confidence interval is developed by estimating the rank for the upper and lower confidence interval and using the slopes corresponding to these ranks to define the actual confidence interval for \( Q' \). For a two-sided confidence interval about the median slope, first we find the \( S_{\text{statistics}} \) for a two-tailed normal distribution test. For a two-sided confidence interval of 95%, \( Z_{(1-0.05)} = Z_{0.975} = 1.96 \). Then, estimate the variance of the Mann-Kendall statistics (\( \text{VAR} (S) \)) as developed by Kendall.

\[
\text{VAR} (S) = \left[ n(n-1)(2n+5) - \sum_{p=1}^{q} t_p (t_p - 1)(2t_p + 5) \right]/18
\]

(2.12)

Where: \( n = \text{number of data points} \)
Chapter-2 Observations, Model and Analysis

t_p = the number of tied for the p^th value
q = the number of tied values

To estimate the range of ranks for the 95% confidence interval, we find C using:

\[ C_a = Z_{(1-a/2)} \cdot \sqrt{VAR(S)} \]  
(2.13)

Using the value of equation, the ranks of the lower (M_1) and (M_2+1) confidence limits is obtained using:

\[ M_1 = (N-C_a)/2 \]  
(2.14)

\[ M_2 = (N+C_a)/2 \]  
(2.15)

Finally, the slopes corresponding to M_1 and M_2+1 are chosen as the lower and upper confidence limits, respectively. The median slope is then defined as statistically different from zero (for the selected confidence interval) if the zero does not lie between the upper and lower confidence limits.

(b) Normalized Trend

In the present analysis the normalized trend is defined as the trend expressed as the percentage of mean or standard deviation.

\[ X_{TN} = \frac{T_x}{\sigma_x} \times 100 \]  
(2.16)

Where \( T_x \) and \( \sigma_x \) represent respectively the coefficient of linear trend and the standard deviation in X for the corresponding period.

2.4.2.3 Robust locally weighted scatter plot smoothing technique (RLWSPS)

Robust Fit is a programme that carries out robust smoothing of data using locally weighted scatter plot smoothing (LOWESS), or robust fitting of data to a polynomial. The robust version of this technique LOWESS is an extension of local polynomial (LP) smoothing developed by Cleveland (1979). Like LP it uses a time window centered on each time point, but instead of using the average location (value) over the window, it fits a low-order polynomial (usually a straight line or a parabola) to the data at the window. Once such a polynomial is fitted (using simple or weighted least squares) the smoothed location at time t is the value of the polynomial at that point. In Simple polynomial fitting, all data points within the window contribute equally to the fit whereas in weighted polynomial fitting used here (LOWESS) the data points close to the point which is actually being fitted, (i.e. the point at the centre of the window),
carry more weight in the fitting procedure than data points nearer to the edge of the window. The usual weighting function is a tricubic function. Mainly choosing the window's width and the degree of the polynomial controls the strength of LOWESS smoothing. The advantage of LOWESS is that it does not distort peaks in the data as much as the later technique.

2.4.2.4 Anomaly

Anomaly shows percentage from which the values are positively or negatively deviated for the respective time (months or year) from the corresponding long term mean values. The deviation of a measurable unit, (e.g., temperature or precipitation) in a given region over a specified period from the long-term average, often the thirty year mean, for the same region. Anomalies, or the deviation from the mean, are created by subtracting the value for a particular time from the corresponding climatological values.

For Rainfall parameter the anomaly is given as:

\[ R_A(i, j, n) = R(i, j, n) - \bar{R}(i, j) \]  \hspace{1cm} (2.17)

Where \( \bar{R}(i, j) = \frac{1}{N} \sum_{n=1}^{N} R_A(i, j, n) \) is the mean at location \((i, j)\) at a given time scale (daily, monthly and seasonal)

As in the present study, both observation and simulations are used so we define the anomalies of both the data.

The normalized daily rainfall anomalies for day \( l \) and at location \((i, j)\) from observation and simulation are respectively defined as

\[ R_{ADO}(i, j, l, n) = \frac{R_{DO}(i, j, l, n) - \bar{R}_{DO}(i, j, l)}{\bar{R}_{DO}(i, j, l)} \times 100 \]  \hspace{1cm} (2.18)

\[ R_{ADS}(i, j, l, n) = \frac{R_{DS}(i, j, l, n) - \bar{R}_{DS}(i, j, l)}{\bar{R}_{DS}(i, j, l)} \times 100 \]  \hspace{1cm} (2.19)

Where \( \bar{R}_{DO}(i, j, l) = \frac{1}{N} \sum_{n=1}^{N} R_{DO}(i, j, l, n) \) and \( \bar{R}_{DS}(i, j, l) = \frac{1}{N} \sum_{n=1}^{N} R_{DS}(i, j, l, n) \) are the \( N \) year mean daily rainfall from observation and simulation respectively.

39
Where $R_{DO}(i,j,l,n)$ and $R_{DS}(i,j,l,n)$ represent the daily rainfall values from observation and simulation.

### 2.4.2.5 Rainfall extremes (Categories)

For many applications, a rainfall category in terms of excess or deficit from the normal rainfall can be a valuable input, for seasonal rainfall at all-India level, we have consider three categories for rainfall:

- **Normal**: $-1\sigma \leq R_A \leq 1\sigma$
- **Excess**: $R_A > 1\sigma$
- **Deficit**: $R_A < -1\sigma$

Where $R_A$ is the normalized rainfall anomaly expresses as % of standard deviation

### 2.4.2.6 Composite Analysis

Composite is defined as average of a variable taken over specially-selected time periods with a common characteristic. It is computed over a variety of time ranges (e.g., monthly, seasonal). Composite is a "selective climatology". Generally we take composite of excess, deficit and normal years of rainfall. Also in some cases the composite of El-Nino and La-Nina are considered.

### 2.4.2.7 Downscaling

The General Circulation Model (GCM) as well as reanalysis product are available at very coarse resolution. Downscaling of the parameters carried out using the regridding techniques which basically uses a Spline interpolation. Here we have downscaled the parameters to smaller region like Uttarakhand scale also the station level data being extracted and used in the study.

### 2.4.2.8 Wind Analysis

Wind analysis are carried out using the horizontal wind both meridional and zonal wind at various vertical level. The resultant wind at different pressure level is computed as

$$w = \sqrt{(u^2 + v^2)}$$  \hspace{1cm} (2.20)

where $u$ and $v$ are the $x$ and $y$ component of the wind respectively at different height

The resultant wind direction also used in the studies using the relation $\tan \Theta = v/u$
The Wind Divergence is an important weather parameter

Wind divergence is a vector field used to understand the atmospheric circulation, which is defined as

$$D = \text{div} \vec{U} = \left( \frac{du}{dx} + \frac{dv}{dy} \right)_z$$ (2.21)

$u$ and $v$ are the $x$ and $y$ components of the wind and $z$ indicates the height. Positive divergence means area is increasing and negative divergence means area is decreasing. In mathematical term negative divergence defined as convergence.

2.5 Study Region

The locations considered cover most part of CHU but with varying degree of density of observations (Figure 2.2). The 32 locations represent altitudes from a few hundred meters (like Kashipur) to high altitudes of more than 2000 meters. In terms of longitudes, the region covered is between 78°E-82°E; the latitudes vary from about 28°N to 32°N (Table 2.1).

![Figure 2.2: The locations of the observations considered in the study](image)
Chapter-2 Observations, Model and Analysis

Table 2.1: -Stations considered in the study

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Lat</th>
<th>Long</th>
<th>Altitude (m asl)</th>
<th>Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29.22</td>
<td>78.93</td>
<td>183</td>
<td>Kashipur</td>
</tr>
<tr>
<td>2</td>
<td>28.92</td>
<td>79.5</td>
<td>205</td>
<td>Rudrapur</td>
</tr>
<tr>
<td>3</td>
<td>29.08</td>
<td>80.12</td>
<td>255</td>
<td>Tanakpur</td>
</tr>
<tr>
<td>4</td>
<td>29.85</td>
<td>77.88</td>
<td>274</td>
<td>Roorkee</td>
</tr>
<tr>
<td>5</td>
<td>29.95</td>
<td>78.17</td>
<td>299</td>
<td>Hardwar</td>
</tr>
<tr>
<td>6</td>
<td>29.22</td>
<td>79.52</td>
<td>348</td>
<td>Haldwani</td>
</tr>
<tr>
<td>7</td>
<td>30.27</td>
<td>79.25</td>
<td>396</td>
<td>Karanprayag</td>
</tr>
<tr>
<td>8</td>
<td>29.4</td>
<td>79.12</td>
<td>482</td>
<td>Ramnagar</td>
</tr>
<tr>
<td>9</td>
<td>30.15</td>
<td>78.6</td>
<td>556</td>
<td>Deoprayarag</td>
</tr>
<tr>
<td>10</td>
<td>30.2</td>
<td>78.77</td>
<td>564</td>
<td>Srinagar</td>
</tr>
<tr>
<td>11</td>
<td>30.3</td>
<td>78.08</td>
<td>654</td>
<td>Raipur</td>
</tr>
<tr>
<td>12</td>
<td>30.32</td>
<td>78.03</td>
<td>682</td>
<td>Dehradun</td>
</tr>
<tr>
<td>13</td>
<td>30.33</td>
<td>78.25</td>
<td>735</td>
<td>Bhogpur</td>
</tr>
<tr>
<td>14</td>
<td>30.55</td>
<td>79.58</td>
<td>792</td>
<td>Joshimath</td>
</tr>
<tr>
<td>15</td>
<td>30.4</td>
<td>78.08</td>
<td>975</td>
<td>Rajpur</td>
</tr>
<tr>
<td>16</td>
<td>30.73</td>
<td>78.45</td>
<td>1170</td>
<td>Uttarkashi</td>
</tr>
<tr>
<td>17</td>
<td>29.75</td>
<td>78.53</td>
<td>1387</td>
<td>Kotdwar</td>
</tr>
<tr>
<td>18</td>
<td>29.6</td>
<td>79.67</td>
<td>1585</td>
<td>Almora</td>
</tr>
<tr>
<td>19</td>
<td>29.33</td>
<td>80.08</td>
<td>1638</td>
<td>Champawat</td>
</tr>
<tr>
<td>20</td>
<td>30.15</td>
<td>78.78</td>
<td>1646</td>
<td>Pauni</td>
</tr>
<tr>
<td>21</td>
<td>29.58</td>
<td>80.22</td>
<td>1646</td>
<td>Pithoragarh</td>
</tr>
<tr>
<td>22</td>
<td>29.77</td>
<td>80.05</td>
<td>1676</td>
<td>Berinag</td>
</tr>
<tr>
<td>23</td>
<td>29.83</td>
<td>78.68</td>
<td>1692</td>
<td>Lansdown</td>
</tr>
<tr>
<td>24</td>
<td>30.25</td>
<td>79.25</td>
<td>1707</td>
<td>Bironkhol</td>
</tr>
<tr>
<td>25</td>
<td>30.5</td>
<td>79.25</td>
<td>1861</td>
<td>Ukhimath</td>
</tr>
<tr>
<td>26</td>
<td>30.55</td>
<td>79.57</td>
<td>1875</td>
<td>Joshimath Observetory</td>
</tr>
<tr>
<td>27</td>
<td>30.7</td>
<td>77.87</td>
<td>1905</td>
<td>ChakrataKalsi</td>
</tr>
<tr>
<td>28</td>
<td>29.63</td>
<td>79.43</td>
<td>1951</td>
<td>Ranikhet</td>
</tr>
<tr>
<td>29</td>
<td>29.4</td>
<td>79.47</td>
<td>1953</td>
<td>Nainital</td>
</tr>
<tr>
<td>30</td>
<td>30.45</td>
<td>78.08</td>
<td>2042</td>
<td>Mussoorie</td>
</tr>
<tr>
<td>31</td>
<td>29.75</td>
<td>80.25</td>
<td>2286</td>
<td>Chaukuri</td>
</tr>
<tr>
<td>32</td>
<td>29.47</td>
<td>79.65</td>
<td>2311</td>
<td>Mukteshwar</td>
</tr>
</tbody>
</table>