Chapter 2: 

Data and Methodology

2.1 Introduction

In the present study hindcast reanalysis wave data generated at the Environmental Modeling Center, NOAA by adopting the third generation wave model Wavewatch III version 2.22 has been used. The data sets covering a span of 13 years from 1998 to 2010 was used to study the variability and trends of the wave height in Indian Ocean. With the aid of Empirical Orthogonal Function (EOF) analysis and Fast Fourier Transformation (FFT) analysis, an attempt has been made to identify the variability modes of the ocean and obtain their characteristics.

The second important parameter used for the present study was wind speed at different levels of the atmosphere. The other parameters like sea surface temperature (SST), sea level pressure (SLP) were also used to describe the climatic patterns. The detailed description of the parameters (data sets) used along with wave data is given below.

2.2 Significant Wave Height

In the present study the SWH was obtained from *in-situ* measurements, satellite altimetry and hindcasted model. The details are discussed in the following sections.
2.2.1 *In-situ* measurements

The SWH data was acquired from Indian National Center for Ocean Information Services (INCOIS, Hyderabad, India) data base. This data was measured using wave rider buoys moored at three locations (Table 2.1). The details about the *in-situ* measurements are given in the Table 2.1. The wave measurements using the buoys DS01 and MB10 were used for validation of the model data. The wave spectrum measurements at Visakhapatnam station were used to study the characteristics of sea and swell.

**Table 2.1: Details of *in-situ* measurement locations and duration.**

<table>
<thead>
<tr>
<th>Buoy ID</th>
<th>Location</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS01</td>
<td>69.3° E, 15.5° N</td>
<td>Jan-2000 to Dec 2001</td>
</tr>
<tr>
<td>MB10</td>
<td>85.0° E, 12.5° N</td>
<td>Jan-2005 to Mar-2007</td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td>83.27° E, 17.63°N</td>
<td>01.06.2009 to 31.05.2010</td>
</tr>
</tbody>
</table>

**Wave rider buoy:**

The wave spectrum measurements at Visakhapatnam station were obtained from the data base of INCOIS. The Directional Wave Rider Buoy (DWRB, Make: Datawell, Netherlands) is a surface floating metal sphere (Figure 2.1) design to measure height and direction. It consists of wave motion sensor based on a stabilized platform, accelerometers, and magnetic compass. Waves are measured by integrating the vertical acceleration. To avoid measurements of unwanted accelerations, the accelerometer is mounted on a stabilized platform having a natural period of 40 seconds. By this means the sensitive axis was kept within a few degrees of the vertical. Though the suspension of the stabilized platform is fragile, the construction as a whole has proven to be very reliable. This reliability...
is accomplished by making the platform and the accelerometer neutrally buoyant in the sensor fluid.

![DWRB deployed at 20 m depth region off Visakhapatnam (East coast of India).](image)

DWRB measures wave height for wave periods of 1.6 to 30 seconds with an accuracy of 0.5 % of measured value. It has an internal data logger to store the data within the instrument. It is also equipped with GPS for monitoring and tracking through HF link. The data collected by DWRB was transmitted through HF and satellite link. The additional technical specifications of the DWRB are given in the table 2.2. When the moored buoy follows the waves, the force of the mooring line may change resulting in a maximum error of 1.5% in the measurement of surface elevation. Also, if the wavelength is less than 5 m, the buoy will not follow the wave amplitude and hence will not measure the wave.
Table 2.2: Technical specifications of Datawell Directional Wave Rider Buoy  
(Source: www.datawell.nl).

| Resolution and Accuracy | Heave | Range: | -20 m to +20 m, resolution: 0.01 m  
| Accuracy: | < 0.5 % of measured value after calibration  
| Period: | < 1.0 % of measured value after 3 year  
| | 1.6 s to 30 s  
| Direction | Range: | 0° to 360°, resolution 1.4° (1 binary degree)  
| Heading error: | 1.4° to 2° (depending on latitude) typical 0.5°  
| Period: | 1.6 s to 30 s (free floating)  
| Water temperature | Range: | -5 °C to +46 °C, resolution: 0.05 °C  
| Accuracy: | < 0.1 °C (sensor accuracy)  
| Sensor and Processing | Type | Datawell stabilized platform sensor, performing heave and direct pitch and roll measurements combined with a 3D fluxgate compass and X/Y accelerometers.  
| Sampling | 8 – channel, 14 bit @ 3.84 Hz  
| Processing | 32 bits microprocessor system  
| Standard features | Integrated datalogger | Compact flash module 512 Mb  
| LED Flashlight | Antenna with integrated LED flasher, colour yellow (590 nm), pattern 5 flashes every 20 s, standard length 35 cm  
| GPS Position | 12 channel, fix every 30 min, precision < 10 m  
| Optional features | Datawell HF link | Frequency range 25.5 to 35.5 MHz (35.5 to 45.0 MHz on request)  
| Iridium / Argos | Satellite communication  
| GSM | Mobile communication  
| Solar power system | Solar panel combined with Boostcap capacitors  
| Hull painting | Brantho Kurrux “3 in 1” paint system (no anti-fouling)  
| Radar reflectors | Two reflectors mounted on hatchcover (retrofittable)  
| Hull diameter | 0.7 m and 0.9 m (excluding fender)  
| General | Material | Stainless steel AISI316 or Cunifer10  
| Weight | Approx. 105 kg (0.7 m), approx. 225 kg (0.9 m)  
| Batteries | 0.7 m diam. Operational life 1 year, 1 section of 15 batteries  
| 0.9 m diam. Operational life 3 years, 3 section of 15 batteries  
| Type: Datacell RC20B (200 Wh black)  
| Temperature range | Operating: -5 °C to + 35 °C  
| Storage: -5 °C to + 40 °C (+ 55 °C short term, weeks only) |
2.2.2 Satellite data

In the present study the data from the multi mission merged products of Archiving, Validating and Interpreting Satellite Oceanographic data (AVISO) were used. The merged significant wave height data was generated using Interim Geophysical Data Records (IGDR) for each satellite. Data products were cross calibrated using OSTM/Jason-2 as reference mission. The merged significant wave heights were with a spatial resolution of 1.75° x 1° (lat, long) for global grids covering a period from December 2009 to November 2010.

2.2.3 Model data

In the present study the SWH was obtained from NOAA Wavewatch III version 2.22. This model version was tested for global grids (Tolman, 2002). The model data were extracted for the period from February 1997 to December 2010. The data sets were available in 3hrs time interval with 1°x1.25° (lat, long) spatial resolution for global grids.
2.3 Description of Wavewatch III

The full-spectral third-generation wind-wave model Wavewatch III was developed at the Marine Modeling and Analysis Branch (MMAB) of the Environmental Modeling Center (EMC) of the National Centers for Environmental Prediction (NCEP) (Tolman, 2009). It was based on Wavewatch I and Wavewatch II as developed at Delft University of Technology, and NASA Goddard Space Flight Center, respectively. This section briefly describes the governing equations used in the model.

Propagation

The balance equation used in the model in conservation form with velocities inside the derivatives to conserve the total wave energy.

The balance equation for the spectrum $N(k, \theta; x, t)$ as used in model is given as

$$\frac{\partial N}{\partial t} + \nabla_x \cdot \mathbf{v} N + \frac{\partial}{\partial k} \frac{\partial}{\partial \theta} \mathbf{v} N = \frac{S}{c_g}$$  ---- (2.1)

where,

$$\mathbf{v} = c_g + \mathbf{U},$$

$$\dot{k} = - \frac{\partial \sigma}{\partial d} \frac{\partial d \sigma}{\partial s} \frac{\partial d}{\partial s} - k \cdot \frac{\partial \mathbf{U}}{\partial s},$$

$$\dot{\theta} = - \frac{1}{k} \left[ \frac{\partial \sigma}{\partial d} \frac{\partial d \sigma}{\partial m} - k \cdot \frac{\partial \mathbf{U}}{\partial m} \right],$$

Where $N$ is the spectrum, $c_g$ is group velocity, $\theta$ is direction of the wave, $k$ is wave number, $s$ is a coordinate in the direction of $\theta$, and $m$ is perpendicular to $s$. Equation (2.1) is valid for a Cartesian grid. For large-scale applications, this equation is usually transferred to a spherical grid, defined by longitude and latitude, but maintaining the definition of the local variance (WAMDIG, 1988).
Source terms

The general processes of generation, dissipation and non-linear wave-wave interactions that are implemented in model are described in this section. In deep water, the net source term $S$ is generally considered to consist of three parts, a wind-wave interaction term $S_{in}$, a nonlinear wave-wave interactions term $S_{nl}$ and a dissipation (‘whitecapping’) term $S_{ds}$ (Eqn 2.2). The input term $S_{in}$ is dominated by the exponential growth term, and this source term generally describes this dominant process only. For model initialization, and to provide more realistic initial wave growth and linear input term $S_{in}$ can also be considered in model. In shallow water additional processes have to be considered, most notably wave-bottom interactions $S_{bot}$ (Shemdin et al., 1978). In extremely shallow water, depth-induced breaking ($S_{db}$) and triad wave-wave interactions ($S_{tr}$) become important. Also available in model are source terms for scattering of waves by bottom features ($S_{sc}$) and a general purpose slot for additional, user defined source terms ($S_{xx}$).

This defines the general source terms used in model as

$$ S = S_{in} + S_{in} + S_{nl} + S_{ds} + S_{bot} + S_{db} + S_{tr} + S_{sc} + S_{xx} \quad --- \quad (2.2) $$

Other source terms are easily added. These source terms are defined for the energy spectra. In the model, however, most source terms are directly calculated for the action spectrum. The treatment of the nonlinear interactions defines a third-generation wave model. $S_{in}$ and $S_{ds}$ represent separate processes, but should be considered as interrelated, because the balance of these two source terms governs the integral growth characteristics of the wave model. Two combinations of these basic source terms are available, those of WAM cycles 1 through 3 and the
parameterizations of Tolman & Chalikov (1996) and those of WAM cycle 4. Linear input, shallow water source terms or source terms describing special physical processes are considered to be “additional” source terms.

Model setup

The present study is based on the hindcast wave data of NOAA Wavewatch III version 2.22 of Environmental Modeling Center, NOAA. The minimum water depth in the model is set for 25m and bottom friction constant is set for 0.038 (Tolman, 2002). Winds used to force the model were from the operational Global Data Assimilation Scheme (GDAS) and the aviation cycle of the Medium Range Forecast model (Kanamitsu, 1989; Kanamitsu et al., 1991; Derber, Parish and Lord, 1991; Caplan et al., 1997). The winds were converted to 10m height assuming neutral stability. The SST as needed in the stability correction for wave growth was obtained from the GDAS. Bathymetric grid files contain an array of water depth values or land flags (zero depth) at the wave model resolution. Obstruction grids consist of two arrays representing the degree of meridional and zonal blocking of wave energy propagation due to subgrid topographic features (Tolman, 2003).

2.4 COADS Climatology

To describe spatial climatological patterns of surface parameters like SST, sea level pressure, zonal surface wind and meridional surface wind a well accepted climatological data set called Comprehensive Ocean Atmosphere Data Set (COADS) was used. The 12-month climatology of the above parameters were derived from 1946 to 1989 COADS data base. The spatial resolution for this data was 1˚x1˚.
2.5 **Indian Monsoon Index**

A study conducted by Wang et al. (1999; 2001) suggested that the strength of the Indian monsoon can be represented by Indian Monsoon Index (IMI). The IMI is the difference between the zonal wind at 850 hPa over southern Arabian Sea and over the northern Indian continent. In the present study monthly IMI computed from the wind data obtained from National Center for Environmental Prediction (NCEP), NOAA is used. The IMI was computed using equation 2.3. A typical choice of monsoon indices of Indian monsoon and western north pacific monsoon is shown in figure 2.2.

\[ \text{IMI} = U_{850} (40^\circ\text{E}-80^\circ\text{E}, 5^\circ\text{N}-15^\circ\text{N}) - U_{850} (70^\circ\text{E}-90^\circ\text{E}, 20^\circ\text{N}-30^\circ\text{N}) \quad (2.3) \]

---

**Figure 2.2:** A typical figure showing the choice of indices. Boxes in red color represents the regions used to compute IMI, boxes in blue color represents the regions used to compute DMI and box in green color shows the region of southern ocean wind in the present study.
2.6 Dipole Mode Index

The magnitude of the IOD can be represented by an index called Dipole Mode Index (DMI), which is the difference between mean spatial SST over western equatorial Indian Ocean and over south eastern equatorial Indian Ocean. In the present study the monthly DMI data for the period 1998 to 2010 was obtained from Japan Agency for Marine Earth Science and Technology (JAMSTEC). DMI computed using equation 2.4.

\[
DMI = SST (50^\circ E-70^\circ E, 10^\circ S-10^\circ N) - SST (90^\circ E-110^\circ E, 10^\circ S-0^\circ N) \quad -- (2.4)
\]

2.7 Nino Index

The data on intensity of El Nino/La Nina was obtained from the Climate Prediction Center, NOAA. The tropical Pacific was divided into a number of regions named Nino 1, 2, 3, 4, and 3.4 (which encompasses part of both region 3 and 4). Nino 1 was the area defined by 80\(^\circ\)W-90\(^\circ\)W and 5\(^\circ\)S-10\(^\circ\)S, Nino 2 by 80\(^\circ\)W-90\(^\circ\)W and 0\(^\circ\)-5\(^\circ\)S, Nino 3 by 90\(^\circ\)W-150\(^\circ\)W and 5\(^\circ\)N-5\(^\circ\)S, Nino 4 by 150\(^\circ\)W-160\(^\circ\)E and 5\(^\circ\)N-5\(^\circ\)S, Nino 3.4 by 120\(^\circ\)W-170\(^\circ\)W and 5\(^\circ\)N-5\(^\circ\)S. The focus of much activity related to ENSO has been the Nino3 region. This has been, for instance, the primary predicted ENSO-related quantity by models verified by observed data (Trenberth, 1997). Hence among these indices Nino3 was considered in the present study.

2.8 Data analysis techniques

The model data was obtained in ‘grib’ format for individual months which were the post processed output format of Wavewatch III. These ‘grib’ files were converted to netcdf file format using open source converter and the total data was
merged to make a single data file by using an analysis tool for gridded data called FERRET. The detailed discussion on major methodologies used in the present thesis described in the following sub sections.

2.8.1 Generation of climatology

The daily and monthly climatology was performed as an unweighted average: each non-missing source point contributes 100% of its weight to the destination grid box within which it falls. The probable shift in the data can occur if the source and destination axes were not properly aligned. For example, if a monthly time series has data points at the first of each month and a climatological axis is defined at midmonths, then unweighted modulo averaging will lead to an apparent 1/2-month shift. To avoid situations of this type, linear interpolation was performed to the time axis prior to the generation of climatological data.

2.8.2 Statistical measures for validation

The performance of the model was assessed using the statistical parameters such as Bias, Root Mean Square Error (RMSE), Scatter Index (SI) and correlation coefficient (r) as defined by equations (2.5) to (2.8). The bias of an estimator is the difference between an estimator's expectation and the true value of the parameter being estimated. The RMSE is a frequently used measure of the differences between values predicted by a model and the values actually observed. SI is the ratio of RMSE to the observed mean.

To construct the RMSE, determination of the residuals is required which are the difference between the actual values (x) and the predicted values (y). Squaring the residuals, averaging the squares, and taking the square root gives the RMSE. Then
this RMSE can be used as a measure of the spread of the y values about the predicted y value \( y_i \). The correlation coefficient is a quantity that gives the quality of least squares fitting to the original data.

\[
Bias = \bar{y} - \bar{x} \quad --- \quad (2.5)
\]

\[
RMSE = \sqrt{\frac{\sum(y_i - x_i)^2}{n}} \quad --- \quad (2.6)
\]

\[
SI = \frac{RMSE}{\bar{x}} \quad --- \quad (2.7)
\]

\[
r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad --- \quad (2.8)
\]

### 2.8.3 EOF analysis

The accurate description of the spatial patterns associated to the modes is an important issue in climate research. A useful technique for compressing the variability in this type of time series data is Principal Component Analysis (PCA) (Hotelling, 1933), which is commonly known as Empirical Orthogonal Function (EOF) analysis.

In the present study EOF technique was used to extract the meaningful components of variability in SWH field over Indian Ocean.

Once the anomaly matrix \( X' \) or its weighted version is determined, the covariance matrix is then defined by:

\[
\Sigma = \frac{1}{n-1} X' T X', \quad --- \quad (2.9)
\]
This contains the covariance between any pair of grid points. The aim of EOF is to find the linear combination of all the variables, i.e., grid points, that explains maximum variance. That is to find a direction \( a = (a_1, a_2, \ldots, a_p)^T \) such that \( X'a \) has maximum variability. Now the variance of the time series \( X'a \) is

\[
\text{var}(X'a) = \frac{1}{n-1} \|X'a\|^2 = \frac{1}{n-1} (X'a)^T (X'a) = a^T \Sigma a
\]

To make the problem bounded we normally require the vector \( 'a' \) to be unitary. Hence the problem readily yields:

\[
\max_a (a^T \Sigma a), \text{such that } a^T a = 1
\]

The above solution is a simple eigenvalue problem

\[
\Sigma a = \lambda a
\]

By definition the covariance matrix \( \Sigma \) is symmetrical and therefore diagonalizable. The \( k \)'th EOF is simply the \( k \)'th eigenvector \( a_k \) of \( \Sigma \) after the eigenvalues, and the corresponding eigenvectors, have been sorted in decreasing order. The covariance matrix is also semidefinite, hence all its eigenvalues are positive. The eigenvalue \( \lambda_k \) corresponding to the \( k \)'th EOF gives a measure of the explained variance by \( a_k \), \( k = 1, \ldots, p \). It is usual to write the explained variance in percentage as:

\[
\frac{100\lambda_k}{\sum_{k=1}^{p} \lambda_k} \%
\]

The projection of the anomaly field \( X' \) onto the \( k \)'th EOF \( a_k \), i.e. \( c_k = X' a_k \) is the \( k \)'th principal component (PC) (Hannachi, 2004) is given by

\[
c_k(t) = \sum_{s=1}^{p} x'(t, s) a_k(s)
\]

--- (2.10)