Chapter 2

Data and Methodology

2.1 Datasets

The main periods of analysis are 1950-2010 and 1979-2010, unless specified otherwise, the reason being the increased observed sources and consequently relatively better quality of data since 1950s (e.g. Giese and Ray, 2011). Having said that, the 1950-2010 period is relatively short to study any decadal changes in interannual statistics of an interannual phenomenon. Therefore brief analysis of the TP SSTA for the 1870-2010 period is carried out as well, as keeping in mind the sparse observations available from the pre-1950 period. While Hadley Centre's sea ice and sea surface temperature (HadISST; Rayner et al. 2003) is used mainly, verified the results, in general, using the Extended Reconstruction SST v3 (ERSST; Smith et al. 2008) to account for the sampling issues for these periods. Incidentally, Ashok et al. (2007) also confirm that the results from the analysis of the TP SST data from HadISST are more or less in agreement with those from a corresponding analysis of Optimally Interpolated Sea Surface Temperature (OISST, Reynolds et al. 2002).

In addition, Simple Ocean Data Assimilation (SODA, Carton et al. 2008) and the 850 hPa wind, latent heat flux, shortwave radiation and sensible heat flux data from the NCEP reanalysis (National Centers for Environmental Prediction, Kalnay et al. 1996), for the period 1958-2010 have been used. Ocean Reanalysis System (ORAS4, Balmaseda et al. 2013) and 40-year European Centre for Medium Range Weather Forecasts Reanalysis (ERA-40, Simmons and Gibson 2000) and ERA-Interim reanalysis products (Dee et al. 2011) are used to verify the results. The Climate Research Unit (CRU, Mitchell and Jones 2005) and Global Precipitation Climatology Project V2 (GPCP, Adler et al. 2003) rainfall datasets have been used for the period 1950-2010 and 1979-2010 respectively.
Various projection data from the 32 CMIP5 models have been utilized in this study. Specifically, the outputs from the following two types of CMIP5 multi-model simulations have been analyzed:

a) **Historical (H) simulations (1958-2005):** including both natural and anthropogenic forcings. The natural forcing includes effects such as volcanoes and other natural phenomena in the climate system. The anthropogenic forcing includes the effects of increasing concentration of greenhouse gases (GHG), aerosol emission, etc.

b) **Future projections based on Representative Concentration Pathway (RCP8.5) scenarios (2006-2050):** The RCP 8.5 experiments start and continue from the point where the corresponding H runs end (e.g. in 2005), but the radiative forcing increases throughout 21st century before reaching a level of about 8.5 W/m².

The list of 32 CMIP5 models under consideration and their relevant information is provided in Table 2.1. For each model, the first ensemble member (i.e., r1i1p1) run has been used.

Table 2.1 List of CMIP5 models along with their modeling groups and resolution.

<table>
<thead>
<tr>
<th>Model</th>
<th>Institution</th>
<th>Resolution (latitude×longitude)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS1.0</td>
<td>Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia</td>
<td>300 × 360</td>
</tr>
<tr>
<td>ACCESS1.3</td>
<td>Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia</td>
<td>300 × 360</td>
</tr>
<tr>
<td>BCC-CSM1-1</td>
<td>Beijing Climate Center, China Meteorological Administration, China</td>
<td>232 × 360</td>
</tr>
<tr>
<td>CCSM4</td>
<td>National Center for Atmospheric Research, USA</td>
<td>384 × 320</td>
</tr>
<tr>
<td>CanESM2</td>
<td>Canadian Centre for Climate Modeling and Analysis, Canada</td>
<td>192 × 256</td>
</tr>
</tbody>
</table>
### Chapter 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Institution</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESM1-BGC</td>
<td>NSF-DOE-NCAR</td>
<td>384 x 320</td>
</tr>
<tr>
<td>CESM1-CAM5</td>
<td>NSF-DOE-NCAR</td>
<td>384 x 320</td>
</tr>
<tr>
<td>CESM1-WACCM</td>
<td>NSF-DOE-NCAR</td>
<td>384 x 320</td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>Centre National de Recherches Meteorologiques and Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique</td>
<td>292 x 362</td>
</tr>
<tr>
<td>CNRM-CM5-2</td>
<td>Centre National de Recherches Meteorologiques and Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique</td>
<td>292 x 362</td>
</tr>
<tr>
<td>CSIRO-Mk3.6.0</td>
<td>Commonwealth Scientific and Industrial Research Organization in collaboration with the Queensland Climate Change Centre of Excellence, Australia</td>
<td>189 x 192</td>
</tr>
<tr>
<td>EC-EARTH</td>
<td></td>
<td>292 x 362</td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>Geophysical Fluid Dynamics Laboratory, USA</td>
<td>200 x 360</td>
</tr>
<tr>
<td>GISS-E2-H</td>
<td>NASA Goddard Institute for Space Studies, NY</td>
<td>90 x 144</td>
</tr>
<tr>
<td>GISS-E2-H-CC</td>
<td>NASA Goddard Institute for Space Studies, NY</td>
<td>90 x 144</td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>NASA Goddard Institute for Space Studies, NY</td>
<td>90 x 144</td>
</tr>
<tr>
<td>GISS-E2-R-CC</td>
<td>NASA Goddard Institute for Space Studies, NY</td>
<td>90 x 144</td>
</tr>
<tr>
<td>HadCM3</td>
<td>Met Office Hadley Centre, UK</td>
<td>144 x 288</td>
</tr>
<tr>
<td>HadGEM2-AO</td>
<td>National Institute of Meteorological Research/Korea</td>
<td>216 x 360</td>
</tr>
<tr>
<td></td>
<td>Meteorological Administration, South Korea</td>
<td></td>
</tr>
<tr>
<td>HadGEM2-CC</td>
<td>Met Office Hadley Centre, UK</td>
<td>216 x 360</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>Met Office Hadley Centre, UK</td>
<td>216 x 360</td>
</tr>
<tr>
<td>INM-CM4</td>
<td>Institute for Numerical Mathematics, Russia</td>
<td>340 x 360</td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>Institute Pierre-Simon Laplace, France</td>
<td>149 x 182</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>Institute Pierre-Simon Laplace, France</td>
<td>149 x 182</td>
</tr>
<tr>
<td>MIROC5</td>
<td>Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan</td>
<td>224 x 256</td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>Max Planck Institute for Meteorology (MPI-</td>
<td>220 x 256</td>
</tr>
</tbody>
</table>

*Linear and non-linear evolution and interannual variability of sea surface temperature in the tropical Pacific*
2.2 Methods

As analysis is focused on the interannual time scale, monthly anomalies are calculated by removing the monthly climatology and detrended the observed times series and model outputs by using a linear least square method. The climatology from the observations and that for the H simulations is based on the period 1958-2005, while for RCP 8.5 simulations the climatology is based on the period 2006-2050. The observation, H and RCP 8.5 simulations are detrended separately. In order to check how sensitive the results are to the detrending method, a quadratic detrending technique has also been tested. However, since the main results of the work are found to be robust with regard to these different detrending methods, all the results presented here are based on linear detrended data.

For convenience, we refer to the boreal summer (June-August; henceforth JJA), fall (September-November; henceforth SON), winter (December through following February; henceforth DJF) and spring (following March-May; henceforth MAM) seasons simply as summer, fall, winter and spring.

The methods used in this study comprise of linear methods such as correlation analysis, EOF analysis, REOF analysis, Wards clustering, and non-linear methods such as Self Organizing Map (SOM), and composite analysis. The details follow:

As mentioned earlier, a composite analysis of the detrended TP SSTA during El Niños, and a similar analysis addressing the El Niño Modoki events is carried, respectively. Composite method (e.g. von Storch and Zwiers 1999) is an excellent

Linear and non-linear evolution and interannual variability of sea surface temperature in the tropical Pacific
empirical technique to identify the distinct and significant features of any climate phenomenon, and to establish seasonal evolutions. While the results from EOF analysis can sometimes be subject to artificiality associated with orthogonality, composite analysis is a good tool to reconfirm or reject the results from the EOF analysis. The technique is very useful from this context to examine the existence and reoccurrence of any distinct climate phenomenon that is seasonally phase-locked (e.g. Yamagata et al. 2004; Behera et al. 2003b; Saji et al. 2003). As discussed by von storch and Zwiers in their text book, composite analysis approach does not make any assumptions about the link between the index being composited (in our case SSTA), and its expected mean value. Therefore, we retain the composite analysis as a general method.

In this work, the canonical ENSO events are represented by the well-known Niño 3 index (area-averaged SSTA over 5°N-5°S, 150°W-90°W). We have adapted the EMI, after Ashok et al. (2007), as:

\[ \text{EMI} = [\text{SSTA}]_A - 0.5*[\text{SSTA}]_B - 0.5*[\text{SSTA}]_C \]  

The square bracket in Equation (1) represents the area-averaged SSTA over each of the regions A (165°E-140°W, 10°S-10°N), B (110°W-70°W, 15°S-5°N), and C (125°E-145°E, 10°S-20°N), respectively.

Apart from the composite analysis, broadly following Takahashi et al. (2011), the seasonal evolution of various El Niño, and El Niño Modoki events in boreal summer, fall, winter and spring is explored. The similarity/distinction between the El Niño Modokis and the canonical El Niños is studied by examining the simultaneous temporal evolution of the gravest two EOF components of the TP SSTA associated with the particular events from summer till spring seasons. The simultaneous inter-seasonal tracking of the magnitudes of both of the gravest principal components (PCs) accommodates non-linear evolution. This approach is different from the linear approach of following only one of the two PCs in a single season (in the current work, the notations PC1 and PC2 designate the respective PC of the EOF1 and EOF2 modes of the TP SSTA of a particular season, respectively). Comparison of the inter-seasonal evolution provides an independent dimension of change, in addition to the changes in PC1 and PC2. Further, using the amplitudes of
the PC1 & PC2 for each El Niño and El Niño Modoki event, cluster analyses of the
TP events is carried out during boreal summer till spring to verify our findings. The
cluster analysis method (Ward 1963; Wilks 2011; Endo et al. 2012) is a hierarchical
method, which carries out clustering based on minimum squared Euclidean distance.
The method has been routinely used in climate studies (e.g. Endo et al. 2012). In
addition, to examine whether the Modoki events arise as artifact due to the
orthogonality constraint associated with the EOF analysis (e.g. Wilks 2011; von
Storch and Zweiers 1999), an EOF analysis of the TP SSTA is carried out for the
1979-2010 period, and compare the results with a varimax-based rotational EOF
method (Richman 1986; Mestas-Nunez 2000).

Further, following Giese and Ray (2011), the histogram of center of heating
in the TP SST for the 1950-2010 period is examined. The location of the maximum
heating is represented by the Center of Heat Index (CHI, see equation 1, Giese and
Ray 2011 for details); the CHI is “a temperature-weighted center (in terms of
longitude) of the area over which the warm anomaly exists only if this warm area is
greater than or equal to the area of the Niño 3.4 region”. A Lilliefors test is
performed to test the normality of the data. The Lilliefors test is similar to the
Komogorov-Smirnoff test but does not require a predetermined cumulative
distribution function to test the null hypothesis.

The mechanistic evolutions of both flavors of El Niños is examined. Briefly,
we look into the distinct dynamical and thermodynamic feedbacks which are
calculated following Xiang et al. (2012) and Liu et al. (2011) during canonical and
Modoki El Niños.

The distinct dynamical feedbacks during canonical and Modoki El Niño events are
represented based on the following three parameters, which are listed below:

1. Air-sea coupling strength \( R(u, T) \); where \( u (\text{m s}^{-1}) \) is the zonal wind
anomaly at 850 hPa and \( T (\text{K}) \) is the Niño3.4 SSTA.
2. Wind-thermocline coupling strength \( R(u, D_{20}) \); where \( u (\text{m s}^{-1}) \) is the
zonal wind anomaly at 850 hPa and \( D_{20} (\text{m}) \) is the thermocline depth anomaly.
3. Thermocline-subsurface temperature coupling strength \( R(T_e, D_{20}) \)
where \( D_{20} (\text{m}) \) is the thermocline depth anomaly and \( T_e (\text{K}) \) is estimated at the
depth below the mixed layer and above thermocline. 

(R (u, T)), R (u, D20), and R (Te, D20) are the linear slopes between the respective parameters.

The thermodynamic air–sea feedback processes involve the feedback between the heat fluxes and the SST. The SST-wind-evaporation and SST-cloud-shortwave radiation are dominant contributors for El Niño evolution (Xie and Philander 1994, Wang et al. 1999). The thermodynamic coupling strength is estimated based on the slope between Niño3.4 SSTA and the heat fluxes (latent, shortwave and sensible heat flux) in the region of Niño3.4.

The (R (u,T)), (R (u,D20)), and (R(Te, D20)) for the dynamical and the corresponding slope for thermodynamical coupling strength are estimated based on seasonal (June-August, September-November, December-February and following March-May, respectively) mean values of these parameters for canonical and Modoki El Niño years. The zonal wind anomaly at 850 hPa is averaged over central equatorial Pacific (160°E–130°W, 5°S–5°N, which we refer to as ‘R1’ in the ensuing discussion). The D20 and Te are averaged in the eastern equatorial region (180°–80°W, 2°S–2°N; we refer to this region as ‘R2’). The slope values above 90% confidence are shown as bars filled with colors.

It is emphasized that, unless specifically mentioned, linearly detrended SSTA, and detrended data of all the other datasets is used over all the periods. This is keeping in mind the caution espoused by L’Heureux et al. (2012), that the linear trends in the central TP SSTA do not always seem to be associated with a corresponding opposite trend in the tropical eastern Pacific SSTA. Incidentally, a similar disconnection in the sea level pressure trends across the TP has been documented by Nicholls (2008), who finds a trend in the Southern Oscillation Index for the March-May months, but due to only a trend in Darwin pressures, with no trend in Tahiti pressures. Some of the analysis is repeated by filtering out decadal component (periodicities above 7 years) using a Seasonal-Trend Decomposition Procedure.

The SOM algorithm developed by Kohonen (1990) is applied in our studies. A SOM is a type of artificial neural network (ANN) clustering technique in which the clustering or mapping is achieved using unsupervised learning methods.
Chapter 2

It produce a low-dimensional (typically two-dimensional) and discretized representation of the input data of the training samples, called a map. The map usually consists of one or two dimensional neurons. This method has also been used in climate and weather studies expensively (e.g. Sahai and Chattopadhyay 2006; Chattopadhyay et al. 2008; Borah et al. 2012; Sahai et al. 2014). The neurons map the large and (statistically) less informative input vectors to a more concise informative output vector space. Thus essentially it is a clustering algorithm. Such clustering is done without any pre-information or supervision and the neurons adjust themselves according to the information provided by the data. Also the clustered maps have neighborhood relationship among each other, i.e. maps are arranged in such a way that topologically similar patterns are placed adjacent to each other, that is unique in SOM only. The importance of using SOM is that it assumes the data is continuous, yet the nonlinearity is well taken into account and captures the similar states. The other important advantage is that the lesser SOM nodes are allocated when the data is sparse in a region (Hewitson and Crane 2002; Chattopadhyay et al. 2008).

The four indices used in the SOM classification (Table 2.1) are constructed from the HadISST and SODA dataset for a period of 51 yr (1958–2008). For the actual computation of the input reference vectors of the SOM the standardized and smoothed anomaly values of all the indices are used. Henceforth all the discussion will be based on the result of using the smoothed standardized values. The nonlinear combination of all the indices should sufficiently represent the complex interannual variation of the SSTA and the indices themselves have the capability of capturing the seasonality. We have considered the data for the first three months starting from 1958 (e.g. Jan, Feb, March). Thus we have three months data for each of four variables (indices) i.e. 12 inputs. The next data will start from Feb, March, April. Thus we get continuous data which will help to track the evolution of TP SST. Distinct phases of ENSO will be identified by a lattice of 3x3 SOM nodes.
Table 2.2. List of indices and the corresponding variable used for that index.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Nino3 index: SSTA</td>
</tr>
<tr>
<td>2.</td>
<td>NinoB index: D20</td>
</tr>
<tr>
<td>3.</td>
<td>EMI index: SSTA</td>
</tr>
<tr>
<td>4.</td>
<td>EMI index: D20</td>
</tr>
</tbody>
</table>

To assess the performance of the different CMIP5 coupled models in simulating the observed tropical SST interannual variability, a method based on EOF analysis inspired from the work of Dommenget and coworkers is used (Dommenget, 2007; Bayr and Dommenget 2013; Wang et al. 2015). The principle of the method is as follow. To compare the spatial structure of SST variability of two different datasets, an objective method is to define a set of (spatial) vectors, as a common basis, project the two data sets onto this common basis and compare the amount of variance described by each vector of the basis for the two datasets. Note that these computations require that the two datasets share the same spatial grid. Consequently, we have first interpolated all the CMIP5 outputs onto the HadISST grid, i.e 1° x 1° for further analysis.

As the Dommenget (2007) method is relatively new, we shall briefly explain a few more details of the same. This technique facilitates the definition a common basis for two datasets. It employs the leading EOF modes estimated from one dataset, as these leading EOF modes give a synthetic depiction of the main modes of variability in this dataset. In the next step, a slightly modified procedure proposed by Wang et al. (2015) is adopted, by projecting the second dataset onto these (spatial) EOF modes (computed from the first dataset) and estimate the amount of variance that the reference EOF-modes explain in this projected data set. Here, the HadISST EOF-modes are taken as the reference modes and project the time series of the thirty two H simulations onto these reference modes. Therefore, comparing the overall spatial structure of variability essentially means to estimate how much variance each...
of the reference EOF modes explains in both data sets. As can be perceived, for each EOF-mode, there exists a difference between the explained variance (in the first dataset) and the projected explained variance (in the second dataset), subject to the fidelity of the simulations. This information will be utilized to generate any statistics necessary to rank the CMIP5 models, based on their respective performance in replicating the observed interannual variability of the climate system. The statistics we use, following Wang et al. 2015, are the normalized Root Mean Square Error (RMSE) value. The RMSE for the SST from the two sets is defined as follows:

Define the RMSE statistic equation

\[
\text{RMSE}_{\text{EOF}} = \sqrt{\sum_{i=1}^{n} (\lambda_p - \lambda_o)^2 / (\sum_{i=1}^{n} (\lambda_o)^2)}
\]  

(2)

Where

\[ \lambda_p \] is projected explained variance from a model

\[ \lambda_o \] is explained variance of observation

\[ n = \text{the first eight modes} \]

The normalization allows a better comparison of the RMSE values among different models or dataset with different sampling uncertainties (Wang et al. 2015). A small RMSE value for any model suggests that the model simulates the observed variability of the climate system well. On the other hand, a RMSE value of 100% corresponds to errors that are as big as the eigen values of the selected EOF modes, used to define the RMSE statistic. As per Wang et al. (2015), in most of the cases, this occurs due to a critical mismatch between the leading EOFs in the two datasets or, alternatively, when modes are mixed or inverted between the two datasets giving an improper distance between the two datasets.

A question that comes to mind is why we use only 8 degrees of freedom. Following Bretherton et al. (1999), Wang et al. (2015), we adopt the concept of the effective spatial number of degrees of freedom in order to determine the number of EOFs to be used to compute the RMSE. However, the number of spatial degrees of freedom is a highly non-robust (and not well-defined) quantity, which is difficult to estimate objectively in practice (Sterl et al. 2007). Here, we have thus computed the RMSE
statistics from the first eight modes, as well as the first five modes, from the monthly SST HadISST dataset for the period 1958-2005 and different domains. We have chosen this after ascertaining that the RMSE values don't change much from five to eight modes. Therefore, for the purpose of grouping the CMIP5 models in different categories according to their performance in simulating the observed SST variability, RMSE values computed from first eight modes are used. Finally, standard regression and correlation analyses are used to document the teleconnection patterns and their possible changes in a

2.3 Indices

- **Nino3**: Nino3 is the average SST in the region bounded by 5°N to 5°S, from 150°W to 90°W of Pacific ocean.

- **Nino3.4**: Nino3.4 is the average SST in the region bounded by 5°N to 5°S, from 170°W to 120°W of Pacific ocean.

- **EMI** = \([\text{SSTA}]_A - 0.5 \times [\text{SSTA}]_B - 0.5 \times [\text{SSTA}]_C\)
  where
  
  \[A = (165°E-140°W,10°S-10°N)\]
  \[B = (110°W-70°W, 15°S-5°N)\]
  \[C = (125°E-145°E, 10°S-20°N)\]

- **IOBM**: The IOBM is the SST averaged over the region 40°E to 110°E and 20°S to 20°N of Indian ocean.

- **IOD**: The IOD is the difference between averaged SST over western box (50°E-70°E, 10°S-10°N) and eastern box (90°E-110°E, 10°S-0) of Indian ocean. warming climate.