Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression

by

Silvestre Aguilar-Martinez

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science in The Faculty of Graduate Studies (Physics)

The University Of British Columbia (Vancouver)

December, 2008

© Silvestre Aguilar-Martinez 2008
Abstract

Nonlinear and linear regression models were developed to forecast the sea surface temperature anomalies (SSTA) across the tropical Pacific ocean. The methods used were, Bayesian neural networks (BNN), support vector machines for regression (SVR) and linear regression (LR). The predictors of the models were a particular combination of the principal components of sea level pressure (SLP) and sea surface temperature (SST) data at lead times ranging from 3 to 15 months. Two data sets corresponding to the time periods, 1950-2005 and 1980-2005 were independently studied. The later data set contained the standardized volume of warm water (WWV) integrated across the tropical Pacific basin as an extra predictor. The seasonal cycle was removed from both data sets prior to their introduction into the regression models. Various graphical and statistical tools were used to compare the performance of these models, in particular the correlation scores.

The results reveal that although there appeared to be nonlinear elements of the SSTA response to the previous SLP and SST conditions, this response is of second order. The large amplitude variance represented by the first principal component (PC) of the SST field can be adequately expressed as a linear response. Consistently in both data sets, the advantages in the nonlinear modelling manifested mainly through the second PC of the SST variability. The spacial distribution of correlation skills between the predicted and observed SST fields corresponding to the BNN and SVR models show some nonlinear structures in various parts of the domain. Despite some differences in the temporal and spacial distribution of these correlation skills, there was no decisive advantage between using BNN or SVR. Adding the WWV to the predictor set reduced the marginal advantage of the nonlinear methods over the eastern region of the tropical Pacific, in particular at lead times of 9-15 months. These results generally agree with linear recharge oscillator models, and the hypothesis of a linear relationship between thermocline depth and SST in the eastern region.
# Table of Contents

Abstract ................................................................. ii

Table of Contents .................................................. iii

List of Tables ......................................................... v

List of Figures ......................................................... vi

1 Introduction ......................................................... 1
   1.1 Basic Concepts ............................................... 1
   1.2 Brief History ................................................ 4
   1.3 Atmosphere-Ocean Dynamics ............................... 6
   1.4 Atmospheric-Oceanic data ................................ 11
       1.4.1 Sea Surface Temperature ............................. 12
       1.4.2 Sea Level Pressure .................................. 13
       1.4.3 Warm Water Volume ................................. 13

2 Linear Multivariate Statistical Analysis ..................... 15
   2.1 Basic Statistical Concepts in Data Analysis .......... 15
       2.1.1 Mean, variance and standard deviation .......... 16
       2.1.2 Covariance ......................................... 17
   2.2 Principal Component Analysis ............................ 17
   2.3 Predictors and Predictand ............................... 21
   2.4 Interpretation of Principal Components ............... 23

3 Statistical Forecast Models ...................................... 26
   3.1 Linear Regression .......................................... 26
# Table of Contents

3.2 Neural Networks .................................................. 28  
3.2.1 Training and Optimization ................................ 29  
3.2.2 Dimensionality ................................................. 31  
3.3 Support Vector Machines ....................................... 32  
3.3.1 Radial Basis Function ....................................... 36  
3.3.2 Implementation of SVR for SST Forecast .................. 36  

4 Forecast Results .................................................... 38  
4.1 The 1950-2005 Data Set ......................................... 38  
4.2 The 1980-2005 Data Set ......................................... 45  
4.2.1 Discussion and Conclusion ................................. 56  
4.3 Further Research ................................................. 64  

Bibliography ............................................................ 65  

Appendices  
A Time Series for El Niño 4, 3.4, 3, and 1+2 .................... 68
List of Tables

4.1 The selected number of hidden nodes is listed for every validation segment. The leftmost entries in the list indicate recent periods, with earlier periods back to 1950 to the right. ............... 40
4.2 Hidden nodes results 1980-2005 .......................... 49
4.3 Values of the hyperparameter set \{\gamma, c, \epsilon\}. The (*) denotes the values after WWV was added to the predictor set. The large value of the hyperparameters \epsilon(*) reflects the use of nonstandardized predictands in the models. The relative large value of \epsilon is due to the large amplitude of the PCs (Figure 2.2). ............... 51
List of Figures

1.1 Niño regions ................................................. 2
1.2 Recharge Oscillator diagram ................................. 9

2.1 The First 5 Modes of SST ................................. 24
2.2 The First 5 PCs of SST ................................. 25

3.1 feed-forward multilayer perceptron ................. 30
3.2 ε-insensitive error norm ................................. 33

4.1 1950-2005, correlation scores for the first five PCs of SST .... 39
4.2 Correlation and error for El Niño regions (1950-2005) .... 43
4.3 BNN: Performance contour map (1950-2005) ........ 44
4.4 SVR: Performance contour map (1950-2005) ........ 46
4.5 1980-2005: correlation scores for the first five PCs of SST .... 47
4.6 Correlation scores for WWV as a single predictor model .... 48
4.7 1980-2005: Correlation Scores for El Niño 4, 3.4, 3, and 1 + 2 .... 52
4.8 1980-2005: RMSE Scores for El Niño 4, 3.4, 3, and 1 + 2 ........ 53
4.9 Times series: Niño 3.4, BNN and LR .................. 54
4.10 Times series: Niño 3.4, SVR and LR .................. 55
4.11 BNN: Performance contour map (1950-2005) ........ 57
4.12 SVR: Performance contour map (1950-2005) ........ 58
4.14 SVR: Performance contour map (1950-2005) (+WWV) .... 60
4.15 LR: Performance contour map .......................... 61

A.1 Times series: Niño 4, BNN and LR .................. 69
A.2 Times series: Niño 4, BNN and LR (+WWV) ............ 70
List of Figures

A.3 Timesseries: Niño 4, SVR and LR ........................................ 71
A.4 Timesseries: Niño 4, SVR and LR (+WWV) .......................... 72
A.5 Timesseries: Niño 3.4, BNN and LR ............................... 73
A.6 Timesseries: Niño 3.4, SVR and LR ............................... 74
A.7 Timesseries: Niño 3, BNN and LR ................................. 75
A.8 Timesseries: Niño 3, BNN and LR (+WWV) ........................ 76
A.9 Timesseries: Niño 3, SVR and LR ................................. 77
A.10 Timesseries: Niño 3, SVR and LR (+WWV) ........................ 78
A.11 Timesseries: Niño 1 + 2, BNN and LR .......................... 79
A.12 Timesseries: Niño 1 + 2, BNN and LR (+WWV) ............... 80
A.13 Timesseries: Niño 1 + 2, SVR and LR .......................... 81
A.14 Timesseries: Niño 1 + 2, SVR and LR (+WWV) ............... 82
Chapter 1

Introduction

1.1 Basic Concepts

The El Niño-Southern Oscillation (ENSO) refers to the large scale interaction between the tropical Pacific Ocean and the atmosphere at interannual time scales. El Niño/La Niña is the persistent onset of warm/cold surface waters along the tropical central-eastern Pacific Ocean. The Southern Oscillation is the atmospheric component of the ENSO signal, which is characterized by an out-of-phase oscillation of sea level pressure between the western and central-eastern Pacific ocean. According to the Australian Bureau of Meteorology, the Southern Oscillation Index is defined as the normalized air pressure difference at sea level between the island of Tahiti in the Pacific, and Darwin Australia.

\[
\text{SOI} = \frac{\Delta P - \overline{\Delta P}}{\sigma_{\Delta P}},
\]

(1.1)

where \( \overline{\Delta P} \) and \( \sigma_{\Delta P} \) are respectively the mean and standard deviation of the pressure difference \( \Delta P \).

The "normal" conditions of the ocean-atmosphere system are defined by the climatology [16]. Given the observation record, normal conditions in the tropical Pacific are defined by:

- Low pressure over northern Australia
- Strong easterly trade winds
- Warm waters on the western Pacific and Indian ocean; cold waters on the eastern Pacific off the Peruvian coast
- Strong upwelling of cold water along the Peruvian coast
Chapter 1. Introduction

- Strong storm activity over the warm waters in the west

Deviations from this state often result in the development of El Niño, which is defined as a sustained sea surface temperature (SST) anomaly greater than 0.5°C for a period of more than five months across the central tropical Pacific ocean, a region commonly known as Niño 3.4, see Figure 1.1 [1]. The state of ocean-atmosphere system during El Niño is characterized by the following, not exhaustive set of conditions:

- Increase in sea level pressure over northern Australia
- Weakening of the trade winds
- Warm air rises over the Peruvian Coast causing torrential rains
- Cessation of upwelling along the Peruvian coast and deepening of the thermocline

![Figure 1.1: Map of Niño regions](image)

La Niña is the complementary phase, when the sea surface temperatures across the central tropical Pacific ocean are unusually low. There is an inherent asymmetry of the ENSO cycle, the cold phase anomalies (La Niña) are typically not as strong as its warm counterpart. This is discussed in more detail in [2].

Depending on the intensity of a El Niño event, its effects can be global and often with catastrophic consequences. For example, the interruption of upwelling
along the Peruvian coast has a devastating effect on the fishing industry in South America, given that the upwelling of cold, nutrient rich waters is responsible for sustaining the fish populations. Another consequence is an increase in sea level pressure in the western tropical Pacific which can cause the failure of the monsoons in India and the potential for subsequent outbreaks of famine. In fact the famine of 1899 following the failure of the monsoons in India propelled Sir Gilbert Walker, General Director of Observations in India, to investigate the mechanism responsible for the interannual variation of the monsoons [16]. Around the globe El Niño can be associated with a multitude of catastrophes, ranging from wild fires produced by drier than normal conditions, mudslides and floods on normally arid regions [16].

These, and other far reaching consequences of El Niño makes predictions of high practical interest. However, its irregularity in temporal and spacial extensions makes forecasts difficult. A variety of methods have been employed to predict the sea surface temperature anomalies in the tropical Pacific which is the most prominent feature of the ENSO signal. Physical, statistical and physical-statistical hybrid models have been developed for ENSO forecast [12][23]. It has been shown that statistical models can attain comparable forecast skill to physical models [20][23].

So far relatively few statistical models have been designed to explore the nonlinear features of ENSO [23]. An and Jin (2004), suggested that the asymmetry between El Niño/La Niña is a characteristic of the nonlinearities inherent in the ENSO cycle. In this work we have developed a simple linear regression model and two nonlinear statistical models to forecast the SST anomalies over the tropical Pacific Ocean. One such nonlinear model is known as Bayesian neural network (BNN) which uses a set of adaptive basis functions to extract the nonlinear relationships between input data, also known as predictors, and output data (predictand). Due to the general ambiguity associated with neural network training [13], a kernel method known as support vector machine was also used. The kernel methods store the training inputs so that later they can be compared against the set of test inputs. Following this comparison which is characterized by a so called kernel function an output (forecast) can be produced [3]. Chapters 4 and 5 discuss the models and their implementation in more detail. The predictors (inputs) in this case constitute a history of the sea level pressure (SLP) and SST conditions previous to the pre-
Observations and theoretical models suggest that the warm water volume (WWV) anomaly over the tropical Pacific plays a central role in the ENSO cycle [14] [9] [6]. Therefore, the WWV was included in the models as an extra predictor.

1.2 Brief History

In 1889, via a small bulletin article, Dr. Luis Carranza, president of the Lima geographical society, first called attention to a warm counter current flowing north-south along the Peruvian coast. Due to its prominence around Christmas, fishermen from the ports of Paita and Pacasmayo called this counter current “El niño Jesús”, Spanish for the “child Jesus”. Accompanied with particularly intense occurrences of this current was unusually high levels of rainfall over the Peruvian coastal and thousands of kilometers inland. These rains brought to the area a rare display of flora and fauna to the otherwise barren desert landscape. The local population began to call these, años de abundancia (years of abundance)[16]. A few years later, in 1899 the failure of the monsoon rains in India caused a severe famine which prompted Sir Gilbert Thomas Walker to study the interannual variations of the monsoons. His analysis of land and ocean data led to the discovery of an interannual seesaw oscillation of atmospheric pressure between the Indian Ocean and the eastern tropical Pacific Ocean [16]. Walker called this fluctuation of air pressure the “The Southern Oscillation”. Furthermore, He also discovered correlations of the Southern Oscillation with rainfall patterns over India and the tropical Pacific region [11]. Walker believed that predictability of the Indian monsoon lay on understanding the large scale phenomenon partially revealed by the Southern Oscillation. Inadequate ocean surface data hindered Walker from discovering the effect of the Southern Oscillation on oceanographic variations i.e. sea temperature variations. Walker also documented a correlation between the weakening of trade winds and heavy rainfall in the central equatorial Pacific. It was not until the early 1940s during a particularly strong año de abundancia that the weak trade winds were associated with the warm surface waters that extended from the Peruvian coast far into the central Pacific Ocean. During the 1950s Bjerknes et al. from the University of California, Los Angeles proposed that the exceptionally warm sur-
face waters off the coast of Peru during años de abundancia were a result of large scale changes in the circulation of the entire tropical Pacific Ocean [16].

Bjerknes in 1966 put together the observations obtained after the 1940s El Niño event to conceive a positive feedback mechanism involving a large scale interaction between ocean and atmosphere. Simply put, his mechanism uses Walker’s Southern Oscillation to generate a closed circuit of air flow between east and west. Cooled air in the upper atmosphere sinks over the cold surface waters of the eastern tropical Pacific, then zonal pressure gradients forces air to flow westward as trade winds where it collects moisture and rises over the warm western waters as rain clouds. The circuit closes as the flow returns to the east through the upper atmosphere [4]. The mechanism that drives the Walker Circulation depends on heat exchange between the ocean and atmosphere. The temperature gradients over the tropical Pacific determines the air flow, thus interannual variations of the sea surface temperature causes the Southern Oscillation, conversely the Southern Oscillation causes sea surface temperature variations i.e. El Niño/La Niña [4]. This way, Bjerknes was able to summarize many decades of observations into a large-scale ocean-atmosphere interaction, known now as El Niño-Southern Oscillation (ENSO).

In subsequent decades, starting in the 1970s, scientist set out to construct various dynamical models based on the same fundamental thermodynamical ideas behind Bjerknes’ conceptual model of air-sea interactions. In 1975 Dr. K. Wyrtki from the University of Hawaii published his model on the dynamics of ENSO. He proposed that the built up of warm water on the western tropical Pacific Ocean was caused by “excessively strong southeast trades” [24], two years before the onset of El Niño. This would cause a steep slope of sea level, leading to eastward flow upon relaxation of the wind stress. Since then, many attempts have been made to model the oscillatory nature of ENSO. Many versions of a recharge oscillator model have been published, in which varying mechanisms of the coupled air-sea system produce a built up of warm water (charging phase-El Niño). This heat is then dispersed into higher attitudes (discharge phase-La Niña) producing changes in the precipitation patterns [9] [6]. Other semi-periodic models have tried to incorporate random atmospheric disturbances to explain the spacial and temporal variability of ENSO [16]. The most significant source of high frequency atmospheric forcing is the
Madden-Julian Oscillation (MJO). MJO generates easterward propagating Kelvin waves, then El Niño conditions in turn enhance MJO [16]. Others, attribute the quasi-periodic and irregular nature of ENSO to various nonlinear processes such as, nonlinear zonal advection, tropical instability waves and even biological physical feedback [2].

1.3 Atmosphere-Ocean Dynamics

General circulation models stipulate that the Southern Oscillation (SO) is caused by interannual sea surface temperature variations [16]. Moist air rises over low pressure regions, called convective zones where heavy rainfall is observed [16]. The trade winds blow from high to low pressure regions driving warm surface waters to the convective zones which are located where the sea surface temperatures are highest. The tropical Pacific convective zone is located over the western region and is responsible for the monsoon rains over India, Indonesia, Malaysia and Northern Australia. This zonal atmospheric circulation, characterized by the air rising over the convective zone and sinking over the regions of high air pressure and low sea surface temperatures is known as the Walker circulation [16]. Changes in location of the warm water has a significant influence on the location of the convective zones and hence on the large scale rainfall patterns. During El Niño warm surface waters expand eastward, pushing the convective zones to the east.

Bjerknes (1969) proposed a mechanism upon which later physical models of El Niño have been built. He stated that a slight relaxation of the trade winds could in turn allow the eastward advection of warm waters. This would further reduce the intensity of the trade winds as the convective zone expands towards the east. The effect is amplified until the trade winds stop, or blow eastward and the warm pool reaches the coast of Peru. Similarly, the effect could start with a shift in the surface waters due to anomalous eastward advection.

This positive feedback mechanism rendered the development of various coupled ocean-atmosphere models [24][9][6]. This interaction, according to [16] can be quantitatively described by relating the zonal and meridional atmospheric sur-
Chapter 1. Introduction

face wind stress \((\tau^x, \tau^y)\) to the zonal and meridional \((U, V)\) ocean currents

\[
(\tau^x, \tau^y) = \alpha(U, V),
\]  

(1.2)

where \(\alpha\) is the constant of proportionality. An equation of state is also needed to relate the atmospheric heat source \(Q\), to the temperature of the ocean surface \(T\).

\[
Q = \beta T,
\]  

(1.3)

with \(\beta\) being a constant. As stated in Philander (1990), to relate the atmospheric heat source to ocean dynamics, \(T\) must depend on the oceanic variables. The simplest assumed form of this dependence is given by

\[
T = s\eta,
\]  

(1.4a)

\[
T_t + u\bar{T}_x = -rT,
\]  

(1.4b)

where the constant \(s\) quantifies the linear dependence between sea surface temperature, \(T\) and \(\eta\), the depth of the thermocline. The term \(u\bar{T}_x\) in (1.4b) represents mean zonal advection, \(u\) is the zonal current speed and \(\bar{T}_x\) is the temperature gradient averaged over the meridional direction, and \(-rT\) is the damping term with \(r = \text{constant}\). Equations (1.4) summarize how changes in the thermocline depth cause a zonal redistribution of warm water and vice-versa [16].

Jin (1997) took Bjerknes positive feedback process and the works of Cane, Zeabiak and Wyrtki, and convincingly demonstrated from their hypotheses that ENSO is a natural mode of oscillation of the tropical Pacific ocean-atmosphere system [9]. The central idea to most recent models is understanding the role of the thermocline depth in ENSO dynamics. Assuming mass adjustment of warm water through Sverdrup transport to higher latitudes, the change in thermocline depth in the western equatorial Pacific \(\eta_W\) can be written as:

\[
\frac{d\eta_W}{dt} = -r\eta_W - F_r.
\]  

(1.5)

The first term on the right hand side is the damping characterized by the constant \(r\), while the second term, \(F_r\) represents the wind forcing [9]. Furthermore, the
difference in thermocline depth between east and west is approximately equal to \( \tau \), some value proportional to the zonally integrated wind stress over the equatorial band defined by one Rossby radius of deformation from the equator. Explicitly this can be written as follows,

\[
\eta_E - \eta_W = \tau. \tag{1.6}
\]

SST is another indicator of the thermocline depth. In the east, water from the subsurface is constantly being upwelled through the climatological trade winds. The weakening of the trade winds and the onset of the SST anomalies produce changes in upwelling and horizontal advection processes. These in turn cause an increase in thermocline depth [9]. These processes can be summarized by the following linear equation.

\[
\frac{dT_E}{dt} = -c T_E + \gamma \eta_E + \delta \tau_E, \tag{1.7}
\]

where changes over the averaged central-eastern tropical Pacific SST \( T_E \) are dependent upon the damping term \(-c T_E\) and is also proportional to the local thermocline depth and the averaged zonal wind stress \( \tau_E \) over where the SST anomaly takes place [9]. The atmospheric response \((\tau, \tau_E)\) to the SST anomaly \( T_E \) can be approximated by

\[
\begin{align*}
\hat{\tau} &= b T_E, \tag{1.8a} \\
\tau_E &= d T_E, \tag{1.8b}
\end{align*}
\]

where \( b \) and \( d \) are constants [9]. Substituting (1.8) and (1.6) into (1.5) and (1.7), it can be shown that,

\[
\begin{align*}
\frac{d\eta_W}{dt} &= -\tau \eta_W - \alpha b T_E, \tag{1.9a} \\
\frac{dT_E}{dt} &= R T_E + \gamma \eta_W, \tag{1.9b}
\end{align*}
\]

where \( \alpha \) is a constant of proportionality relating \( \tau \) and \( \tau_E \), and \( R = -c + \gamma b + \delta d \). Equations (1.9a) and (1.9b) constitute Jin's linear coupled system between thermocline depth and SST.
Chapter 1. Introduction

Figure 1.2: Recharge oscillator model where (a) and (c) correspond respectively to the warm and cold phases (El Niño/La Niña). The rectangular region represents the tropical Pacific region, with the elliptical region denoting the SST anomaly in the east. The filled arrows describe the wind stress anomaly $\tau$, and the unfilled arrows indicate the heat discharge/recharge from the equatorial band. $h$ is the thermocline depth. Panels (b) and (d) represent the transition phases [9].

The physics of this coupled oscillator successfully reproduces some essential features of the ENSO cycle as well as the Bjerknes positive feedback process. Figure 1.2 illustrates the essential features of his model. Initially the SST anomaly induces westerly wind forcing which is in phase with the deepening of the thermocline in the eastern tropical Pacific (Fig 1.2a). At this point the ocean adjustment
processes governed by zonally integrated Sverdrup transport unloads the equatorial band from its heat content resulting in a negative thermocline depth anomaly over the entire tropical Pacific (Fig 1.2b). The now shallow thermocline permits seasonal upwelling of cold water in the eastern Pacific. The positive feedback process intensifies the easterly trade winds, leading to anomalous upwelling of cold water in the eastern Pacific, this is known as the cold phase (La Niña) (Fig 1.2c) [9]. This coupled oscillator model is known as the discharge/recharge oscillator model of ENSO. The warm (cold) phase, Figure 1.2a (1.2c) represents El Niño (La Niña) phenomenon.

More recently, another oscillator model was proposed by Clarke et al. (2007) where the key variables are the warm water volume (WWV), i.e. the volume of warm water above the 20°C isotherm within the region bounded by [156°E, 80°W] and [156°S, 5°N], and SST [6]. The physics of Clarke’s model differ slightly from Jin’s recharge oscillator in two aspects. First, according to Clarke’s model the area of maximum ocean-atmosphere coupling takes place in the western-central tropical Pacific Ocean, contrary to Jin’s where only the eastern SST anomaly drives changes in the thermocline (see equations 1.9). Second, zonally integrated wind stress curl anomalies along the northern and southern boundaries of the box in Clarke’s model produce changes in the WWV by meridional and divergent flow [6]. This mechanism is absent in Jin’s oscillator model. The dynamical variable in Clarke’s model are, the SST anomaly, $T$ and the isotherm depth, $\eta$, both averaged over the western-central region, [156°E, 140°W], [5°S, 5°N]. The coupling of these two variables is given by,

$$\frac{\partial \eta}{\partial t} = -\alpha T, \quad (1.10a)$$

$$\frac{\partial T}{\partial t} = \nu \eta, \quad (1.10b)$$

which is Clarke’s couple oscillator model, $\alpha$ and $\nu$ are coupling constants. Measurements of these constants yields a frequency or $\omega = 2\pi/(4.25 \text{ yr})$ which is of the right order for ENSO [6].

The dynamical models offered by [9] and [6] represent ENSO as a coupled ocean-atmosphere system that features periodic and regular oscillations. The ob-
servations, on the other hand reveal an irregular quasi-periodic cycle with tendency to amplify the positive phase (El Niño) more than its negative counter part (La Niña). An and Jin (2004) called attention to the irregularities of ENSO and proposed a low order nonlinear ENSO model which reproduced the observed asymmetries [2]. The heat content in the upper tropical Pacific is given by

\[
\frac{dT_E}{dt} = -a(T_E - T_0) - \frac{2u}{L} (T_W - T_E),
\]

(1.11)

\[
\frac{dT_W}{dt} = -a(T_W - T_0) - \frac{w}{H} (T_W - T'),
\]

(1.12)

where SST is denoted by \(T_E\) and \(T_W\) for the eastern and western region of the tropical Pacific. The other variables include the meridional and vertical velocities \((u, w)\), \(T_0\) is the equilibrium temperature, while \(T'\) denotes the temperature of the mixed layer. \(L\) and \(H\) represent the dimension of the region, \(L\) being the width of the basin and \(H\) the layer depth [2]. The nonlinearities come in by relations among the wind stress \(\tau\), temperature \((T_E, T_W)\) and velocity \((u, w)\) fields.

\[
\tau = -\mu(T_W - T_E)
\]

(1.13)

\[
u = \epsilon \tau
\]

(1.14)

\[
w = \xi \tau
\]

(1.15)

We noticed that this models adds nonlinear coupled dependences between the key variables \(T_E, T_W,\) and \(H\). Solutions for this system requires techniques which lie beyond the aims of this project.

1.4 Atmospheric-Oceanic data

Following more than half of a century of ocean and atmosphere observations [11], [4] have established that large scale ocean-atmosphere interactions are responsible for the alternating anomalous periods of warm/cold surface waters [16]. The key variables of the ocean-atmosphere system governing the low frequency (interannual) variations are sea surface temperature (SST) and sea level pressure (SLP). It has been suggested that high frequency atmospheric forcing characterized by
Chapter 1. Introduction

an index of Madden-Julian Oscillation (MJO) can account for nearly half of the Niño3.4 variance [14]. While the role of high frequency variables such as MJO may be highly significant in accounting for the structural irregularities of ENSO [14], its role was not considered in this work. Instead, two sets of SST and SLP data were used, one covers the period 1950 – 2005, while the other 1980 – 2005. The second period contains warm water volume (WWV) data which is absent from the first set due to its unavailability prior to 1980.

1.4.1 Sea Surface Temperature

Sea surface temperature (SST) is perhaps the most relevant feature of El Niño. The SST data used here comes from the NCEP/National Center for Atmospheric Research (NCAR) reanalysis [10] website, http://www.cdc.noaa.gov/cdc/data.noaa.ersst.html. Monthly average of SST data is available over a global $2^\circ \times 2^\circ$ grid. Version 2 of the extended reconstructed SST (ERSST.v2) used here was available from 1854 to the present [17]. Despite its availability back to 1854, only the periods 1950 – 2005, and 1980 – 2005 were considered. The region of interest is centered in the tropical Pacific (124°E – 70°W, 20°N – 20°S). The data were available in a NetCDF file format; a Matlab program written by graduate student Song Cai of the University of British Columbia, was used to open such data files. Once the data had been opened, smoothing by a three-month running mean was performed, i.e.

$$\tilde{x}_{ijk} = \frac{1}{3} \sum_{m=k-1}^{k+1} x_{ijm},$$

(1.16)

where the variable $\tilde{x}_{ijk}$ denotes the smoothed SST data at latitude $i$, longitude $j$ and time $k$, and $x_{ijk}$ is the original SST value.

For what followed it was necessary to arrange the gridded variables into columns keeping track of their location on the grid. Each column corresponds to a particular month, or more accurately a given three-month average. Two more operations are needed before the data is ready for processing, the first is to eliminate the land data. A given dummy value is assigned to these grid points where SST is not defined, these points are then removed from the data matrix keeping track of their location, i.e. storing their index. This operation was performed using a matlab file...
specifically written for this purpose. The second operation is the removal of the climatological seasonal cycle associated with each data set. A 57 year seasonal cycle extending through January 1948 to December 2004 was removed from the 1950 – 2005 data set, and a 26 year (January 1980-December 2004) climatology was removed from the 1980 – 2005 data set. The end result is a matrix of SST anomalies where the variables are variations above or below the seasonal mean.

\[
X = \begin{pmatrix}
x_{11} & x_{12} & \ldots & x_{1m} \\
x_{21} & x_{22} & \ldots & x_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \ldots & x_{nm}
\end{pmatrix}
\]  

(1.17)

In this format each \(x_{ik}\) represents the SST anomaly at location \(k\) on the grid and time \(l\), the ranges for \(l\) and \(k\) are 1, \ldots, \(n\) and 1, \ldots, \(m\) respectively. Note that the latitude and longitude index \((i,j)\) has been mapped to a single index \(l\).

### 1.4.2 Sea Level Pressure

Monthly Sea Level Pressure data (SLP) available on a 2.5° × 2.5° global grid was downloaded from NCEP Reanalysis. SLP corresponds to the near surface measurements at (0.995 sigma level) [10]. The data used here extends from January 1948 to December 2005 (first set), and January 1980 to December 2005 (second set) over the tropical Pacific ocean (120°E – 70°W, 20°S – 20°N). The data was packaged in the same NetCDF file format as SST, thus was made available for analysis in the Matlab environment. The transformations of running mean, climatology and temporal mean removal described in the previous section were also applied to both sets of SLP data.

### 1.4.3 Warm Water Volume

Warm water volume (WWV) is defined as the volume integral of water above 20°C isotherm between (5°N – 5°S, 120°E – 80°W) [14]. Monthly WWV anomalies were obtained from the Australian Bureau of Meteorology Research Center (BMRC) Ocean Analysis (http://www.bom.gov.au/bmrc/ocean/results/
climocan.htm) for the period of January 1980 to December 2005. A three month running mean was applied to the data followed by removal of the corresponding climatological seasonal mean. The index of WWV according to McPhaden et al. (2006) were inferred from the analysis of mooring and expendable bathythermograph data.
Chapter 2

Linear Multivariate Statistical Analysis

To undertake a statistical analysis of ENSO, one has to deal with large datasets. A serious study must involve a multitude of variables over a significant time interval. These requirements may in some cases render extraction of meaningful information contained in such datasets a difficult task. A necessary processing of the data must be carried out before it becomes manageable for the purposes of studying ENSO. There is a myriad of available options designed to manage data sets involving multiple variables. Among these are:

- Regression analysis, linear and nonlinear
- Canonical correlation analysis (CCA)
- Principal component analysis (PCA)
- Nonlinear CCA and PCA

Our goal from the outset was to reduce the size of the data structures while retaining important information such as variance. A method for reducing the number of variables while keeping the variance information is principal component analysis (PCA). As described in the next section, this linear method allows for a systematic study of our multivariate SST and SLP data.

2.1 Basic Statistical Concepts in Data Analysis

Before one can proceed with the development of PCA, it is necessary to lay down some basic concepts and formulae pertaining to statistical analysis. The following
Chapter 2. Linear Multivariate Statistical Analysis

will not be a rigorous or exhaustive presentation but rather a list of useful and recurrent formulas, for a rigorous treatment, see [21].

2.1.1 Mean, variance and standard deviation

The mean of a discrete variable $X = \{x_1, x_2, ..., x_n\}$ as defined by [21] is the expectation value of $X$ given by,

$$E(X) = \bar{x} = \sum_{i=1}^{n} x_i f_X(x_i).$$ (2.1)

Here the preferred symbol for the mean will be $\bar{x}$. The probability function $f_X(x_i)$ is simply $1/n$, assuming all possible values of $x_i$ are equally probable [21]. Under this simplification, the mean reduces to the simple formula

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$ (2.2)

Another salient feature of datasets is its dispersion or variance, it measures how observations of a variable are distributed about its mean. Mathematically, the variance of a data sample is denoted by the symbol $\sigma^2$ and it is given by the following formula

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$ (2.3)

Sometimes one speaks not of the variable variance but of its square root, known as the standard deviation [21].

Multivariate analysis often requires manipulation of distinct variables into a single regression model. To ensure that differences in units do not produce biased output results, a standardization procedure is necessary. This is accomplished by

$$z_i = \frac{x_i - \bar{x}}{\sigma}.$$ (2.4)

Since $\sigma$ has the same units as $x$, the above operation also eliminates the units of the input variables [8].

16
2.1.2 Covariance

Clearly the equations presented in the previous section pertain only one dimensional, or single variable data sets. A multivariate generalization of these concepts is needed for ENSO analysis. In dealing with multiple variables not only is it necessary to compute the properties of each variable, such as its variance and standard deviation but, we also would like to know how a particular variable changes with respect to changes in another variable. The concept of covariance naturally arises from this need. The covariance between two variables is defined by

$$C(x, y) = \sum_{k=1}^{n} \frac{(x_k - \bar{x})(y_k - \bar{y})}{n - 1}.$$  

(2.5)

Notice that if \( x = y \), the above reduces to the variance formula equation (2.3). These are the essential statistical concepts needed to proceed with the development of principal component analysis.

2.2 Principal Component Analysis

As explained previously sea surface temperature (SST) is the most important variable of this study, thus it will be used for the following illustration of principal components analysis. SST over the region of interest can be represented by the set of measurements \( X(t) = \{ x_1(t), x_2(t), ..., x_n(t) \} \) where \( t = \{ t_1, t_2, ..., t_m \} \), each value of \( x_i \) represents a measurement of sea surface temperature at a location denoted by \( i \) which correspond to a grid point of latitude and longitude. At each location measurements are made a total of \( m \) times. This information can be presented in the more illustrative form of equation (1.17). At certain times during the development of this project, it was necessary to use an alternative notation to that of equation (1.17), such alternative is,

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}. \quad \text{(2.6)}$$
Based on this representation, the $j$ index corresponds to a given observation of the set of variables $\{x_j\}$, where $n$ is the number of grid points.

The covariance matrix for the SST field is

$$C_{ij} = \sum_{k=1}^{m} \frac{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{m - 1}. \quad (2.7)$$

The advantage of the principal component method is the systematic reduction of the number of variables needed to represent the variance of a data set. This approach follows [8], where, in the language of linear algebra a change of basis is applied to the original data set namely, $X$ in (1.17). The new set of variables denoted by $Z$ is designed to better exhibit the properties of variance. One can think of this geometrically as an axis rotation from $\{x_1, x_2, ..., x_n\}$ to $\{z_1, z_2, ..., z_n\}$, [8]. The matrix that performs the transformation is

$$E = \begin{pmatrix}
    e_{11} & e_{12} & \cdots & e_{1n} \\
    e_{21} & e_{22} & \cdots & e_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    e_{m1} & e_{m2} & \cdots & e_{mn}
\end{pmatrix}. \quad (2.8)$$

following the notation of [8]. The transformation in matrix and index notation, respectively

$$Z = EX \quad \text{or} \quad z_{ik} = \sum_{j=1}^{n} e_{ij}x_{jk}. \quad (2.9)$$

The purpose of this transformation is to find the $e_{ij}$s such that the variance of $z_1$ is maximized. Recall that the variance of $z_1$ is

$$\text{var}(z_1) = \sum_{i=1}^{m} \frac{(z_{1i} - \bar{z}_1)^2}{m - 1} \quad \text{where} \quad \bar{z}_1 = \sum_{i=1}^{m} \frac{z_{1i}}{m}. \quad (2.10)$$

Substituting (2.9) into $\bar{z}_1$ yields

$$\bar{z}_1 = \sum_{i=1}^{m} \sum_{k=1}^{n} \frac{e_{1k}x_{ki}}{m} = \sum_{k=1}^{n} e_{1k}\bar{x}_k. \quad (2.11)$$
The variance of $z_1$ is

$$\text{var}(z_1) = \sum_{k=1}^{n} \sum_{l=1}^{n} e_{1l} e_{1k} \sum_{i=1}^{m} \frac{(x_{ki} - \bar{x}_k)(x_{li} - \bar{x}_l)}{m - 1}$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{n} e_{1l} e_{1k} C_{kl}. \quad (2.12)$$

Where we identify $C_{kl}$ as the covariance matrix defined by (2.7). Using once again the notation of [8] we can write (2.12) in matrix notation

$$\text{var}(z_1) = e_1^T C e_1 \quad \text{provided} \quad e_1 = [e_{11}, \ldots, e_{1n}]^T. \quad (2.13)$$

Recalling that our plan was to chose $e_1$ such that $\text{var}(z_1)$ is a maximum, however as shown in (2.12) the variance of $z_1$ is dependent upon the length of $e_1$, to solve this issue the following constraint is imposed [8].

$$e_1^T e_1 = 1. \quad (2.14)$$

The technique of choice to solve an optimization problem subject to constraints is the method of Lagrange multipliers. Optimization is performed on the Lagrange function $L$ [8],

$$L = e_1^T C e_1 - \lambda (e_1^T e_1 - 1). \quad (2.15)$$

One can show that the $e_1$ must be an eigenvector of the covariance matrix and the eigenvalue corresponds to the variance of $z_1$ [8].

$$\lambda = e_1^T C e_1 = \text{var}(z_1). \quad (2.16)$$

The process is repeated for $z_2$ subject to, $e_2^T e_2 - 1 = 0$ with the additional constraint that $\text{cov}(z_1, z_2) = 0$. It turns out that this condition amounts to the orthogonality relation

$$e_2^T e_1 = 0. \quad (2.17)$$

The new Lagrangian function which incorporates the additional constraints takes the
Chapter 2. Linear Multivariate Statistical Analysis

following form

\[ L = e_2^T C e_2 - \lambda (e_2^T e_2 - 1) - \gamma e_2^T e_1. \]  (2.18)

Finding the stationary points of the Lagrangian function i.e setting the derivatives of \( L \) to zero, followed by matrix product with \( e_1^T \) on the left leads to an eigenvalue equation for \( e_2 \) [8]. Therefore,

\[ \lambda = e_2^T C e_2 = \text{var}(z_2). \]

Previously it was found that the eigenvector \( e_1 \) which maximizes \( \lambda = \lambda_1 \), similarly the eigenvector that corresponds to the second largest eigenvalue is \( \lambda = \lambda_2 \). The same procedure is used to find the variances of \( z_3, z_4, \ldots, z_n \), which are arranged according to \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0 \) and their sum equals the variance of the original variable matrix

\[ \text{var}(X) = \sum_{i=1}^{n} \lambda_i. \]

These expressions make evident the primary goal of principal component analysis which is to cast the data matrix into a new set of variables where their characteristic variances are ordered. In some cases only a subset of \( \{z_i\} \) can account for most of the variance inherent in the set \( \{x_i\} \). Our statistical study of ENSO hinges on this special feature of PCA. What remains is to find the coefficients corresponding to the projections in the vector space spanned by the set \( e_i \)

\[ X = E^T A \quad \text{or, in index notation} \quad x = \sum_{i=1}^{m} a_i(t) e_i. \]  (2.19)

The projection coefficients \( a_i(t) \) can be easily computed using the orthogonality of condition \( e_i^T e_j = \delta_{ij} \) to yield

\[ a_i(t) = e_i^T x. \]  (2.20)
Chapter 2. Linear Multivariate Statistical Analysis

To clarify the notation,

\[ E^T = \begin{pmatrix} e_{11} & e_{21} & \cdots & e_{m1} \\ e_{12} & e_{22} & \cdots & e_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{mn} \end{pmatrix} = (e_1 \ e_2 \ \cdots \ e_m) \quad \text{and,} \]

\[ A = \begin{pmatrix} a_1(t_1) & a_1(t_2) & \cdots & a_1(t_m) \\ a_2(t_1) & a_2(t_2) & \cdots & a_2(t_m) \\ \vdots & \vdots & \ddots & \vdots \\ a_m(t_1) & a_m(t_2) & \cdots & a_m(t_m) \end{pmatrix}. \quad \text{(2.21)} \]

The columns of the matrix \( E^T \) are the set of eigenvectors of the covariance matrix \( C \). In the literature they are known as principal vectors, spacial patterns, or EOFs (Empirical Orthogonal Functions). The set \( a_i(t) \) are known as the principal components (PCs), or time series. Another important property of PCA is that the PCs \( \{a_i\} \) are uncorrelated i.e

\[ \text{cov}(a_i, a_j) = 0, \quad \text{(2.22)} \]

[8].

2.3 Predictors and Predictand

As discussed in the previous section, one of the advantages of using PCA is the reduction of the number of variables needed to represent the variance of a multivariate data set. Explicit use of this property has been made while handling ENSO data. Thus this process defined the variables used in the models. Nevertheless, a presentation of the problem needs to be provided before discussion of the pertinent variables.

The solution to the so called regression problem is of primary interest. This problem can be presented via the following equation

\[ y = f(x), \quad \text{(2.23)} \]
where the objective is to find the form of $f$ that better represents the relationship between the variable sets $x$ and $y$. The first set are called predictors or inputs, the later set, namely $y$ are called predictands (outputs). The relationship $f$ can be \textit{linear} or \textit{nonlinear}. The particular choices for $f$ used here are discussed in the next chapter.

The predictand, SST in this case, corresponds to the variable whose variability over time one is attempting to forecast. Recall that SST is given by a relatively large data matrix given by equation (1.17), but PCA allows representation of this data by a smaller set of synthetic variables. Therefore, PCA was performed on the SST data matrix (equation 1.17), and the first five PCs served as the predictands in our regression models. Notice that the PCs ($a_i(t)$) provide the temporal variations of the variable field, whereas the EOFs ($e_i$) account for the spatial variability. Furthermore, since the PCs are uncorrelated, forecasts can be made independently for each one (equation 2.22) [8]. The first five PCA modes of SST account for 59.85%, 10.82%, 7.98%, 3.51%, and 2.85% respectively, of the total variance for the 1950 — 2005 SST data set, and 60.33%, 15.04%, 7.95%, 3.10%, and 2.40% for the 1980 — 2005 data set. The PC time series ($a_i$) for 1950-2005 are given in Figure 2.2, while the spacial patterns are given in Figure 2.1.

To complete the formulation of the regression problem we need to provide the input variables \textit{i.e.} $x$ in equation (2.23). The predictors are generated from the PCs of SST and SLP data over a relative short duration prior to the predicted period. In other words, the predictors are the first 9 and 7 PCs of SST and SLP, respectively, with the SLP PCs also supplied at time lags of 3, 6, and 9 months. This procedure incorporates a temporal profile of SST and SLP conditions into the inputs of the regression models [23][14]. A more specific form of the regression expression appropriate to the forecast models is

$$y(t + \Delta_{\text{lead}}) = f(x_1(t), x_2(t), x_2(t - \Delta t_1), x_2(t - \Delta t_2), x_2(t - \Delta t_3)), \quad (2.24)$$

where $x_1$ denotes the 9 leading PCs of SST, while $x_2$ denotes the 7 leading PCs of SLP, with lag times $\Delta t_1 = 3$ months, $\Delta t_2 = 6$ months, and $\Delta t_3 = 9$ months. The lead time $\Delta_{\text{lead}}$ as defined in [18] is the time interval between the center of the latest predictor to the center of the predicted period [23].
2.4 Interpretation of Principal Components

According to [8], the first three modes of the linear expansion of the SST field by PCA are associated with El Niño/La Niña phenomenon. Given that the first three modes account for approximately 75% of the total variance, it is clear that this phenomenon is the dominant feature of the SST variability. Mode 1, (see Figure 2.1a) indicates strong variability in the central-eastern equatorial Pacific, increasing from 160°E to the eastern boundary. The peak of the variability occurs at the coast of Peru and extends slightly southward, which is where it was first identified as a warm current flowing south [16]. The time series PC1, Figure 2.2a is the projection coefficient which multiplies every point on the field in EOF1. From Figure 2.2a we can notice that temperatures in the central-eastern tropical Pacific reached their highest levels in 1982/1983 and 1997/1998 El Niño events. Mode 2/EOF2 (Figure 2.1b) corresponds to a E-W “standing mode” mentioned in [15] and [22]. This mode reveals the internal structure of SST variability with positive and negative anomalies separated roughly by the 130°W meridian, see Fig. 2.1b. Its time series indicates that PC2 was positive during both El Niño and La Niña events, and the maximum anomalies also occurred during the two strongest El Niño events. For mode 3, the peak occurs in the central equatorial Pacific at the 150°W meridian (Figure 2.1c). Shown in Figure 2.2c, the time series of mode 3 illustrates a steady increase in intensity starting at around 1970 [8].
Chapter 2. Linear Multivariate Statistical Analysis

Figure 2.1: The first five spatial modes (EOFs) for the SST field (1950 to 2005) with positive contours indicated by solid curves, negative contours by dashed curves and zero contours thickened. Contours are multiplied by a factor of 100.
Figure 2.2: The first five principal components of SST (multiplied $10^{-2}$) for 1950 to 2005.
Chapter 3

Statistical Forecast Models

The high level of complexity of the ocean-atmosphere system limits a complete analytical development of its evolution. In certain practical cases, we may be more interested in forecasting the SST anomalies than in modeling ocean-atmosphere interactions. According to [20] statistical models yield forecast skills that are comparable to the dynamical models despite their relative simplicity [23]. Many linear statistical models have been used for ENSO forecasting, such as: multivariate linear regression, and canonical correlation analysis [23][20]. However, recent evidence suggests that ENSO may involve nonlinear interactions [2], therefore a nonlinear statistical treatment is more adequate for such a case. The following sections describe the various statistical models used; the nonlinear methods are Bayesian neural networks (BNN) and support vector machine for regression (SVR). The results were compared with multivariate linear regression.

3.1 Linear Regression

Linear regression constitutes the simplest nontrivial way to relate two variables, or two sets of variables. Linear regression for a single output variable (SST) can be represented mathematically by

\[ y(x, \beta) = \beta_0 + x_1 \beta_1 + \ldots + x_n \beta_n, \]

\[ y(x, \beta) = \sum_{j=1}^{n} X_{ij} \beta_j + \beta_0. \]  

This expression can be equivalently represented in matrix notation by \( Y = X^T \beta \) [3]. Recall that by convention, the lower case characters denote a single observation of a single variable, while lower case bold characters represent a single
observation of the full variable set. Bold upper case characters denote a matrix of observations (time series) of the variable set. Generalization to multiple outputs is straightforward and can be done by trading the vector of parameters $\beta$ by a matrix, but since our problem (forecasting SST) involves only a single output we will use the linear model as expressed in equation (3.1).

The output $y$ in this regression model is given by a linear combination of the inputs (predictors) $\{x_i\}$. The goal in this case is to choose the set of parameters $\beta$ such that their product with input variables $\{x_i\}$ most closely predicts the target value $y$. To accomplish this, the data set is first divided into training and test sets. The training set constitutes 90% of the time series for every variable; the test set constitutes the remaining 10% of the data. In other words the time series data is split into 10 segments, training is performed using 9 segments and forecasts are made on the remaining segment. The process is repeated until all 10 segments have been forecasted. This procedure is known as cross validation. During every training session the $\beta$s are chosen such that sum of squared residuals is minimized [3]. In order for the following to coincide with the notation used in other texts such as [3], the matrix of inputs $X$ contains $n$ variables along its rows, and each row represents one out of a total of $m$ observations. Then,

$$ E = \sum_{i}(y_i - \hat{y}_i)^2, $$

(3.2)

where,

$$ \hat{y}_i = \sum_j X_{ij}\beta_j + \beta_0, $$

(3.3)

is the model output and $y_i$ is the target or observed value. Equation 3.2 is minimized with respect to $\beta_i$, by $\partial E/\partial \beta_i = 0$ [3]. Before preceding it is convenient to transform our variables according to $X_{ij} \rightarrow (1, X_{ij})$ and $\beta_j \rightarrow (\beta_0, \beta_j)$ [3]. These transformation allow for equations (3.2) and (3.3) to be cast in a more compact form

$$ E = \sum_{i}(y_i - \sum_{j=0}^{n} X_{ij}\beta_j)^2 $$

(3.4)
Chapter 3. Statistical Forecast Models

The extremum is given by the following

\[
\frac{\partial E}{\partial \beta_i} = \sum_i 2(y_i - \sum_{j=0}^n X_{ij} \beta_j) (\sum_{j=0}^n -X_{ij} \delta_{ij}) = 0,
\]

\[
\sum_i 2(y_i - \sum_{j=0}^n X_{ij} \beta_j) X_{ij} = 0,
\]

\[
\sum_i 2(X_i^T y_i - \sum_j X_i^T X_{ij} \beta_j) = 0.
\]

The result can be more easily expressed in matrix notation

\[
2(X^T y - X^T X \beta) = 0,
\]

\[
\beta = (X^T X)^{-1} X^T y. \tag{3.5}
\]

This gives the sets of parameters \( \beta \) that minimizes the mean squared error, equation (3.4) [3] for every validation set.

### 3.2 Neural Networks

Attention is now devoted to one of the nonlinear regression models used in the forecast of SST. This model consists of a feed forward two-layer perceptron neural network (NN) model with Bayesian regularization [3]. As described in [3], the multilayer perceptron is a particular representation of the general regression equation

\[
y = \sum_{j=1}^N w_j \phi_j(x), \tag{3.6}
\]

where \( \phi_j(x) \) corresponds to a set of nonlinear basis functions. In certain large scale applications, the number of basis function can become very large and prohibitively costly in terms of computational resources [3]. An elegant solution is to fix the number of basis function to a relatively small number \( h \), at the cost of allowing the basis functions to depend on some input parameter. This leads to the basic neural
network structure [3]

\[ H_i = H(\sum_{j=1}^{n} w_{ij}x_j + w_{i0}) \text{ where, } i = 1, 2, \ldots, h. \] (3.7)

Notice that \( H \) depends on the parameter set \( w_{ij} \) and the bias parameter \( w_{i0} \). The number or adaptive parameters at the first layer is given by \( n + 1 \). Our choice for basis function, following [3] is the hyperbolic tangent function

\[ H_i = \tanh(\sum_{j=1}^{n} w_{ij}x_j + w_{i0}). \] (3.8)

These expressions together with the assignments \( X_{ij} \mapsto (1, X_{ij}), \) and \( w_{ij} \mapsto (w_{i0}, w_{ij}) \) lead to the neural network model used here

\[ y_k = \sum_{i=1}^{h} \tilde{w}_{ki} \tanh(\sum_{j=0}^{n} w_{ij}x_j) + \tilde{w}_{k0}. \] (3.9)

In this case there is only one output variable (SST), therefore only one index is needed to describe the parameters, namely \((\tilde{w}_{0}, \tilde{w}_{i})\). With this model the total number of adaptive parameters to be determined by training are \((n + 1)h + 1\). The diagram in Figure 3.1 provides a representation of the particular neural network model used in this model.

### 3.2.1 Training and Optimization

Training of the neural network begins with a random choice of the model parameters \((w_{i0}, w_{ij})\) and \((\tilde{w}_{0}, \tilde{w}_{i})\) as well as a set number of hidden neurons \( h \). From this choice of parameters an error function such as (3.2) can be computed. However, given the flexibility of NN to conform to any continuous function \( y(x) \) [3], choosing \( w \), and \( \tilde{w} \) based on optimization of the error function (3.2) would lead to a problem known as overfitting, which occurs when the network model fits into the noise of the data. To reduce the problem of overfitting, optimization is performed
hidden nodes

Figure 3.1: General diagram of a feed-forward two layer perceptron.

on a regularized error function such as,

\[ E = \frac{\beta}{2} \sum_{n=1}^{N} (\hat{y}(x_n, w) - y_n)^2 + \frac{\alpha}{2} w^T w, \]  

(3.10)

where the second term is a weight penalty or regularization term, and the hyperparameters \( \alpha \) and \( \beta \) are chosen based on the Bayesian framework, hence the name Bayesian neural network [3]. Equation (3.10) is known as Bayesian regularization, and it was used in the training of the neural network via the MATLAB program "trainbr" from the Neural Network tool box.

Finally, the following procedure was used to select the optimum number of hidden neurons \( h \). As in the case of linear regression the data was partitioned into 10 segment of which 9 were used for training, and one was reserved for testing. For what follows, the labels D9 and D1 are given respectively to the 9 and 1 data segments. The D9 set was further partitioned into two sets, one reserved for training and the other for validation. Roughly, 2/3 of the data was used for training, which
by the way was chosen at random in blocks of 12 observations minimum size. This procedure was introduce to further avoid overfitting due to the autocorrelation in the data. The remaining 1/3 of the data was used for validation. For clarity, the labels DT and DV are respectively given to a particular random choice of training and validation data.

With the value of $h = 1$, and a given random pick of DT and DV, the error function (3.10) was optimized a total of 50 times. Since the function (3.10) contains many local minimum, there were potentially 50 different models. The model with lowest mean squared error (MSE) over the training set DT was kept, this error was given the label $t_e$. This model was then applied to the corresponding validation set and the MSE over this set was also computed, the name given to this value was $v_e$. The set $(t_e^1, v_e^1)$ characterized this particular choice of DT, DV and $h = 1$. Next, $h$ was set to 2 and the whole process was repeated without changing DT and DV to obtain $(t_e^2, v_e^2)$. This process was carried out for $h$ ranging from 1 to 8, such that in the end there were 8 different sets of pairs $(t_e, v_e)$; the $h$ corresponding to the lowest validation error ($v_e$) is stored as $h_1$. At this point a different DT and DV were randomly chosen from the same D9 and the process was repeated until $h_2$ was determined. This was done a total of 20 times which implies the set $s = \{h_1, h_2, ..., h_{20}\}$. Ideally all the $h_i$s should be the same since they come from random picks of the same data set D9. Finally, the mode of set $s$ was chosen as the optimum number of hidden neurons that represnts the data D9 [13]. With this choice of $h$ a total of 50 models were trained using the entire D9 set and tested on D1. The performance was evaluated in terms of the ensemble average. This entire process was repeated until all 10 segments had been forecasted.

### 3.2.2 Dimensionality

Severe difficulties can arise in problems involving high dimensionality [3]. The problem in a sense results from an exponential increase in the "size" of mathematical spaces with an increase in dimensions [3]. In this case, there is a total of 37 time series corresponding to the predictors in the model. The adaptive parameters needed to train such a model must be found in a hypervolume of $(37 + 1) \times h + 1$ dimensions. The neural network approach used here to forecast SST is therefore
Chapter 3. Statistical Forecast Models

...threatened by the curse of dimensionality.

To circumvent this issue, an expedient solution due to [23] was used; the 37 time series were compressed by computing a second PCA, also known as extended empirical orthogonal functions (EEOF) analysis, or singular spectrum analysis (SSA) [23]. The first 12 SSA PCs were taken to be the final predictors in the model. One notices that this technique yields a considerable reduction in the dimensions, which in turn yields a reduction of the parameter space volume. The volume is proportional to the number of parameters raised to the power of the number of inputs. Therefore,

\[
\frac{V_{37}}{V_{12}} \propto \frac{N^{37}}{N^{12}} = N^{25},
\]

for the special case where \(N\) parameters are possible along each dimension, the 37-input parameter space is \(N^{25}\) times larger than the 12-input parameter space.

3.3 Support Vector Machines

Support vector machine for regression (SVR) constitutes the other nonlinear model used in our forecast of SST. SVR is a particular example of kernel methods [3]. This type of model stores the input vectors during the training process, which are used later during the prediction phase.

The method works in the following way. Consider a linear regression model

\[
\hat{y}_i = \phi^T(x_i)w + b.
\]

The index \(i\) runs over the time series subset of the training data. The set of training input vectors \(x_i\) are mapped to a feature space represented by a set of fixed nonlinear basis functions \(\phi_k(x)\). Afterwards, linear regression is performed on the feature space characterized by the weight vector \(w\) to produce output, \(\hat{y}\) [3]. The relative simplicity of expression (3.12) hides the complication involved in evaluating a large, possibly infinite set of basis functions \(\phi_k(x)\). Luckily, there is a way to elude this complication.

The SVR method hinges on minimizing a particular regularized error norm defined in terms of a weighted cost function [3]. The cost function contains two
regimes, one called \( \epsilon \)-insensitive, and linear. For the \( \epsilon \)-insensitive regime, errors which are smaller than \( \epsilon \), where \( \epsilon \) is a free model parameter, are set to zero, and therefore are not considered. The linear regime assigns a linear cost for errors greater than \( \epsilon \), this can be appreciated in Figure 3.2. Mathematically, these condition can be summarized by

\[
F = C \sum_{n=1}^{N} E_{\epsilon}(\hat{y}(x_n) - y_n) + \frac{1}{2} w^2,
\]

(3.13)

where,

\[
E_{\epsilon}(\hat{y}(x) - y) = \begin{cases} 
0, & \text{if } |\hat{y}(x) - y| < \epsilon \\
|\hat{y}(x) - y| - \epsilon, & \text{otherwise}
\end{cases}
\]

(3.14)

The constant \( C \) in front of the sum of errors term is an inverse regularization parameter, \( N \) is the number of observations [3]. The error sum in (3.13) can be expressed only in terms of nonzero error contributions by introducing a new set of variables, called slack variables. These are defined by those points whose prediction overestimate/underestimate the target value by an amount greater than \( \epsilon \). The
slack variables are denoted by $\xi_i$ and $\hat{\xi}_i$ respectively. An illustration of this idea is given in Figure 3.2. Our regularized error function (3.13) cast in terms of slack variables is

$$F = C \sum_{i=1}^{M} (\xi_i + \hat{\xi}_i) + \frac{1}{2} \|w\|^2.$$  

(3.15)

The variable $M$ is the number of observations for which the input vector $x_i$ produced a target value $\hat{y}$ outside the $\epsilon$-insensitive region. Following Bishop (2006), equation (3.15) must be minimized subject to the constraints $\xi_i \geq 0$ and $\hat{\xi}_i \geq 0$. The problem of optimization subject to constraints lends itself naturally to the use of Lagrange multipliers. Nonzero Lagrange multipliers $\alpha_i$, $\hat{\alpha}_i$, $\beta_i$, and $\hat{\beta}_i$ and a Lagrangian function $L$ are introduced.

$$L = F - \sum_{i=1}^{M} (\alpha_i \xi_i + \hat{\alpha}_i \hat{\xi}_i) - \sum_{i=1}^{M} \beta_i (\epsilon + \xi_i + \hat{\xi}_i - y_i)$$

$$- \sum_{i=1}^{M} \hat{\beta}_i (\epsilon + \hat{\xi}_i - \hat{y}_i + y_i).$$

(3.16)

Substitution of (3.12) into the the above equation followed by setting the derivatives with respect to $w$, $b$, $\xi_i$, and $\hat{\xi}_i$ to zero yield the following solution for the model parameters based in terms of the Lagrange multipliers [3].

$$w = \sum_{i=1}^{M} (\alpha_i - \hat{\alpha}_i) \phi(x_i),$$

(3.17)

$$\sum_{i=1}^{M} (\alpha_i - \hat{\alpha}_i) = 0,$$

(3.18)

$$\alpha_i + \beta_i = C,$$

(3.19)

$$\hat{\alpha}_i + \hat{\beta}_i = C.$$  

(3.20)

The above conditions are combined to eliminate the variables $\beta$ and $\hat{\beta}$. The La-
Chapter 3. Statistical Forecast Models

The Lagrangian function then takes the form

\[
L = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} (\alpha_i + \hat{\alpha}_i)(\alpha_j + \hat{\alpha}_j)\phi^T(x_i)\phi(x_j) - \epsilon \sum_{i=1}^{M} (\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^{M} (\alpha_i + \hat{\alpha}_i)y_i. \quad (3.21)
\]

Clearly, from (3.19), (3.20), and the constraints \( \alpha, \hat{\alpha} \geq 0 \) and \( \beta, \hat{\beta} \geq 0 \) we can arrive at the box constraints [3]

\[
\begin{align*}
0 \leq \alpha_i & \leq C, \quad (3.22a) \\
0 \leq \hat{\alpha}_i & \leq C. \quad (3.22b)
\end{align*}
\]

The set of nonzero multipliers \( \alpha \), and \( \hat{\alpha} \) determine the data points \( x_i \) that contribute to prediction. These data points are known as the support vectors and correspond to those points which lie outside the \( \epsilon \)-insensitive tube [3]. The prediction output is obtained by combining equations (3.17) and (3.12) into

\[
\hat{y}(x) = \sum_{i=1}^{M} (\alpha_i - \hat{\alpha}_i)\phi^T(x_i)\phi(x_i) + b. \quad (3.23)
\]

Note that the model output as well as the Lagrangian function (3.21) depend only on the inner product of the basis functions, namely, \( \phi^T(x_i)\phi(x_j) \). This result, replaces the need to evaluate a large set of complicated nonlinear functions with that of defining a proper inner product. In the literature this inner product is known as the kernel function.

\[
K(x_i, x_j) = \phi^T(x_i)\phi(x_j) \quad (3.24)
\]

There is a myriad of kernel functions available, the one used here belongs to a particular class traditionally used by beginners called homogeneous kernel functions. These are characterized by a dependence on magnitude of the difference between two input vectors, i.e. \( K(x, x') = K(\|x - x'\|) \).
Chapter 3. Statistical Forecast Models

3.3.1 Radial Basis Function

As discussed in [3], a necessary and sufficient condition for a kernel function $K(x, x')$, to be valid is that the matrix formed by evaluating $K$ at all possible points in the input space be positive definite. Clearly, this requirement is met if we consider the linear kernel, $K(x, x') = x^T x'$. Furthermore, more complicated kernels can be constructed from known valid kernels [3], in particular the $\exp(K(x, x'))$ is valid, provided that $K(x, x')$ is valid. With this in mind, we consider the following kernel function

$$K(x, x') = \exp(-\|x - x'\|^2/2\sigma^2) \tag{3.25}$$

This is known as the Gaussian kernel or radial basis function, for its dependence on $\|x - x'\|$. Notice that

$$\|x - x'\|^2 = x^T x + 2x^T x' + (x')^T x'$$

Expansion of (3.25) in terms of these series of inner products leads to a product of kernels of the type $\exp(x^T x')$ which are valid kernels [3].

3.3.2 Implementation of SVR for SST Forecast

The process of implementing SVR was similar to that used in Baysian neural network and linear regression training. A ten-fold cross validation scheme was devised to forecast the five leading PCs of SST. The model predictors were the first twelve singular spectrum analysis (SSA) PCs of the combined SLP and SST data set. As before the input data is given by the matrix $X$, $x_i$ corresponds to the $i^{th}$ observation of input vectors, where $i = 1, 2, \cdots, m$. For a given set of observations of both input and target values the Lagrangian function (3.21), subject to the box constraints (3.22), was maximized. There are three free hyperparameters in our model which can be adjusted. The value chosen for $\epsilon$ determines the number of relevant points used in building the model and thus, determines the number of support vectors set by $\alpha_i$ and $\hat{\alpha}_i$. The inverse of the hyperparameter $C$ controls the amount of weight penalty or regularization, as seen in (3.15). Lastly, the kernel function hyperparameter, denoted by $\gamma = 1/2\sigma$ governs the spatial scale of the model, with larger $\gamma$ allowing a finer spatial scale, as $\sigma$ controls the width of the
Chapter 3. Statistical Forecast Models

Gaussian kernel in (3.25).

The implementation was carried out by a C++ routine interphased with MATLAB designed by Chih-Chung Chang and Chih-Jen Lin from the Computer Science department at the National Taiwan University, Taipei. The library of support vector machines (SVM) was available for download at http://www.csie.ntu.edu.tw/~cjlin/libsvm. Ten different matlab scripts were written for every cross-validation data segment with the SVM library [5] used to carry out the optimization procedure. The training process was conducted by a seven-fold cross validation procedure for every one of the ten pairs of training and testing sets. A given range of the free parameters $\epsilon$, $C$, and $\gamma$, were probed after which the mean square error (MSE) of each of the seven validation sets evaluated the performance of the model. Finally, the parameters $\{\epsilon, C, \gamma\}$ that yielded the lowest MSE during validation were chosen for predictions over the test data. This process was conducted until the entire range had been tested. The search for the hyperparameters was exhaustive, first choosing a grid to identify the region of minimum MSE, then a finer grid for that region until a minimum was identified at the higher resolution level. The results of the SVR method, the Bayesian neural network (BNN) and linear regression (LR) are discussed in the following chapter.
Chapter 4

Forecast Results

Individual forecasts of the first five PCs of SST were conducted using linear regression (LR), Bayesian neural networks (BNN) and support vector machines for regression (SVR). Forecasts were made at lead times ($\Delta t_{\text{lead}}$) of 3, 6, 9, 12 and 15 months, and the performance of these methods was compared. This analysis was applied to two distinct data records, 1950 – 2005 and 1980 – 2005. The later data set contains an extra predictor, namely, the warm water volume index (WWV). This allowed us to evaluate the importance of WWV on model performance as well as to study the nature of its relationship with the ENSO signal [14][15][6][9].

4.1 The 1950-2005 Data Set

For this set, the results of the forecasts of SST support the conclusions reached by Tang et al. (1999). The nonlinear forecast models offer only a marginal advantage over linear regression in the prediction of SST dynamics over the tropical Pacific ocean at seasonal time scales. One possible explanation for the strength for the linear method in comparison to BNN and SVR is that the ENSO signal is basically linear at these time scales. The correlation skills for individual SST PC forecasts are presented in Figure 4.1.

The first PC is by far the most significant, accounting for $\sim 60\%$ of the total variance. Thus, the forecast of PC1 is the most important as one would like for the models to successfully predict the time series associated with the largest mode of variability. Recall that PC1 has to do with large scale variability over the central-eastern tropical Pacific ocean as it indicates the appearance of warm (or cool) surface waters over Niño 3.4, 3, and 1 + 2 regions (Figure 2.1a). As illustrated in Figure 4.1, correlation skills for PC1 show no significant difference between the various models at all lead times. The temperature variability in the
eastern-central topical Pacific appears to be dominated by a linear relationship between SST and extra-seasonal SLP and SST patterns at 3 to 21 months prior to the predicted month.

Figure 4.1: Cross-validated correlation skill for the 5 leading PCs of SST. The white, black and gray bars denote skill for LR, SVR and BNN, respectively. These scores correspond to the period 1950-2005.
Chapter 4. Forecast Results

The second mode of SST variability associated with PC2 illustrates the east-west contrast in the equatorial Pacific (Figure 2.1b). This second mode was shown in Figure 4.1b to be the most nonlinear aspect of the ENSO signal at short lead times of 3-9 months. Correlation skills between PC2 and the nonlinear forecasts are superior to those of linear regression at these lead times, with BNN showing the maximum correlation scores. The nonlinearity of PC2 is apparent in the cross-validated selection of the parameter $h$ during model training. Recall that the nonlinearities in the BNN framework are shared between the parameter set $\{w\}$ and the number of hidden modes $h$ [13]. The only mode for which $h = 1$ was not selected during the optimization procedure was PC2 at 3, 6, and 9 month lead times. The choice for $h$ in these cases was 2, see Table 4.1.

Table 4.1: The selected number of hidden nodes is listed for every validation segment. The leftmost entries in the list indicate recent periods, with earlier periods back to 1950 to the right.

<table>
<thead>
<tr>
<th>PC</th>
<th>Leadtime (mon)</th>
<th>No. of Hidden Nodes, $h$</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2, 2, 2, 2, 2, 2, 2, 2, 2, 2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2, 2, 2, 2, 2, 2, 2, 2, 2, 2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2, 1, 2, 2, 1, 2, 2, 2, 2, 2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1, 1, 1, 2, 1, 1, 2, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
</tbody>
</table>
As explained in the development of the BNN model, each of the ten segments of the PC time series was forecasted using the remaining 9 segments. The resulting optimum value of $h$ for each of the ten segments is listed in column 3 of table 4.1. The list given in column 3 is also ordered, showing the most recent period on the left and the more distant past on the right. Only PC2 and PC5 at some lead times selected a model with $h$ higher than the minimum value of $h = 1$.

The PC3 time series corresponds to a mode of variability centered at the Niño 3.4 region [190E-120W, 5N-5S]. This mode accounts for the small out of phase SST oscillation between the Niño 3.4 region and the rest of the domain (Figure 2.1). One feature of this mode is its relatively constant correlation skill from 3 to 12 month lead times. An interesting note about the PC3 time series is its noticeable steady increase over the given time period. The amplitude of this mode was smaller during the 1950s compared to the present with its maximum amplitude occurring during the 1997-98 El Niño episode (Figure 2.2). This observation is in accordance with global warming scenarios and their effect on ENSO, see [7].

These correlation results for PCs 4-5 indicates no advantage of the nonlinear methods over linear regression. The cross validated BNN choice of hidden nodes was 1 for PCs 1, and 3-5 (with the exception of PC5 at 3 month lead time). Recall

<table>
<thead>
<tr>
<th>PC</th>
<th>Leadtime (mon)</th>
<th>No. of Hidden Nodes $h$</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 2, 1, 2, 1, 3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 3</td>
<td>1</td>
</tr>
</tbody>
</table>

| 5  | 3              | 2, 2, 2, 2, 1, 1, 1, 2, 2, 2 | 2    |
|    | 6              | 2, 1, 1, 1, 1, 2, 2, 2, 1 | 1    |
|    | 9              | 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 | 1    |
|    | 12             | 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 | 1    |
|    | 15             | 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 | 1    |
Chapter 4. Forecast Results

that 1 hidden node is the simplest nonlinear model allowed by the neural network method. If \( h = 1 \) equation (3.9) reduces to

\[
y = \tilde{w} \tanh\left( \sum_{j=0}^{n} w_j x_j \right) + \tilde{w}_0,
\]

(4.1)

which means there is only one adaptive activation function.

Forecasts of the 5 PCs were combined with their corresponding EOFs shown in Figure 2.1 to reconstruct the SST anomalies over the domain. These anomalies were then averaged over the different Niño regions (4, 3.4, 4, and 1 + 2) (Figure 1.1). Correlation scores between the observed and predicted indices was computed for all models as well as the root mean squared error (RMSE). Shown in Figure 4.2, the obtained results are nearly identical for all 3 models. The region with lowest correlation skill corresponds to the Niño 1 + 2 (Fig. 4.2d) which is also the region of largest variability. These results are also very similar to those obtained in [18] where LR, CCA (canonical correlation analysis) and neural networks were compared.

Finally, the model performance was evaluated at each grid point over the domain. Observation and predicted time series anomalies were compared via correlation scores. A contour map of correlation skills for BNN and SVR is provided in Figures 4.3, and 4.4, respectively. These figures also contain the difference in skill between the nonlinear and linear model. As expected, all of our methods show maximum correlation in the central-eastern tropical Pacific at lead times of 3 and 6 months. At longer lead times the correlation skills decrease while showing maximum values straddling the equator along the 120°W meridian. The right column of panels in Figures 4.3 and 4.4 shows the difference in correlation skills.

It was observed, in the case of BNN that at short lead times (3 and 6 months), these differences are positive over most of the domain. In particular, Figure 4.3d resembles spacial mode 2 where the maximum increased correlation regions correspond roughly to the regions of maximum variability (Figure 2.1b). However, the correlation skill differences are only of the order of 0.01-0.1. Similarly, SVR records a marginally positive difference in the correlation skill over some regions of the domain, particularly over the region centered at 15°S between the 158°E
Chapter 4. Forecast Results

Figure 4.2: Correlation scores and root mean square error (RMSE) for the different Niño regions.

(a) Nino4 [160E–150W, 5N–5S]
(b) Nino4

(c) Nino3.4 [170W–120W, 5N–5S]
(d) Nino3.4

(e) Nino3 [150W–90W, 5N–5S]
(f) Nino3

(g) Nino1+2 [90W–80W, 0N–10S]
(h) Nino1+2
Figure 4.3: BNN forecast performance and its difference from deviation from the linear models. The contour interval is 0.1 in the right column and 0.02 in the left, with positive (negative) contours indicated by solid (dashed) curves, and the zero contour thickened.
and 150°W meridians shown in Figures 4.4d and 4.4f. It also shows a consistent negative correlation difference over the western tropical Pacific for lead times of 3-9 months. Both nonlinear models lose to linear regression at 15-month lead.

4.2 The 1980-2005 Data Set

A more in depth analysis was carried out for the shorter data record, namely the 1980-2005 set, due to the availability of a warm water volume index (WWV) from 1980 to the present. The WWV index was included as an extra predictor in all the aforementioned regression models. For comparison purposes two sets of models were developed, one without the WWV index. It was noted that the most intense El Niño events ever recorded occurred during the 1980-2005 period, viz, the 1982-83 and 1997-98 events (Figure 2.2). The nonlinear terms in equations (1.11) and (1.12) have been found to play a significant role in enhancing the warm cycle during these very strong El Niño events [2]. Hence for the later period of 1980-2005, with more intense ENSO events, the nonlinear methods could be expected to outperform linear ones.

As before, the correlation skills for the PCs of SST corresponding to the 1980-2005 record are given in Figure 4.5. The scores of the linear regression, support vector machines, and Bayesian neural network models are assigned the respective colors, white, black and gray. Two correlation scores are provided for each method. The bar on the left represents the score prior to the inclusion of the WWV index. Before WWV is included as a predictor, BNN provides better correlation scores for the forecast of PC1 than SVR, at lead times of 9, 12 and 15 months. SVR shows marginally higher scores than BNN and LR at short lead times (3 and 6 months). Inclusion of the WWV index in the forecast of PC1 positively influences the correlation scores for all models, reducing, or eliminating the differences among them. This change is expected given that the WWV along with SST are the main variables in most theoretical recharge oscillator models [6][9].

Since one is interested in knowing if the SST response to WWV could be governed by nonlinear dynamics, the performance of the models was tested using only the WWV index as input. The forecast results for PC1 are given in Figure 4.6a, which indicate a linear dependence between SST and WWV, in accordance with
Chapter 4. Forecast Results

Figure 4.4: SVR performance and correlation difference SVR and LR. The contour interval is 0.1 in the right column and 0.02 in the left.
Figure 4.5: Cross-validated correlation skills for the 5 leading PCs of SST. The white, black and gray bars denote skill for LR, SVR and BNN, respectively. For every method the left(right) bar denotes the score before(after) WWV has been included. These scores correspond to the period 1980-2005.
Chapter 4. Forecast Results

Figure 4.6: Cross-validated correlation skills for PCs 1 and 2 of SST. The white, black and gray bars denote skill for LR, SVR and BNN, respectively. For every method the left bar denotes the score when the model contained the full predictor set (including WWV), whereas the bar on the right denotes the scores when only the WWV was used as a predictor in the models.

The correlation scores for PC1 are displayed adjacent to the values obtained using the full set of predictors, with the score bar on the right obtained from the single predictor (WWV) models. For these models the correlation scores provided by linear regression were shown to be superior to nonlinear regression at all lead times except for 15 months. Furthermore, the PC1 predictions of these models show relatively low scores at very short lead times (3 months).

Once again it was noticed that the nonlinearities of ENSO variability come in through PC2. The correlation skills for PC2 at lead times of 3-9 months are superior for the nonlinear methods, where SVR attained the best performance (Figure 4.5). The number of hidden nodes selected from the training and validation procedure for BNN was again 2 for lead times of 3 and 6 months, while all other PCs selected only 1 hidden node (Table 4.2). The hyperparameters chosen during the training of the SVR model are given in Table 4.3.

It was discovered that in the case of PC2, inclusion of WWV does not improve the correlation skills (Figure 4.5). The forecast of PC2 using WWV as the single predictor yielded low correlation scores for all models, as indicated in Figure 4.6b. This result suggests that the asymmetry of SST between eastern and western trop-
Chapter 4. Forecast Results

The Pacific characterized by EOF2 is largely independent of the 20°C isotherm depth (WWV). The scores for PCs 3-5 are given in Figure 4.5, where inclusion of the WWV index did not significantly affect these PCs.

Table 4.2: Hidden nodes results 1980-2005

<table>
<thead>
<tr>
<th>PC</th>
<th>Leadtime (mon)</th>
<th>No. of Hidden Nodes $h$</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2, 2, 2, 1, 2, 2, 2, 2, 2, 2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2, 2, 2, 1, 1, 1, 1, 1, 1, 1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>2, 2, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1</td>
</tr>
</tbody>
</table>
Chapter 4. Forecast Results

Following the procedure of the previous section, the predicted five PCs were combined to generate the Niño 4, 3.5, 3, and 1 + 2 indices. The correlation skills and root mean squared error (RMSE), before and after WWV was introduced into the input set are given in Figures 4.7 and 4.8, for the various Niño regions. Before the WWV index was added BNN provided a better correlation skill than SVR at lead times of 9-15 months for the Niño 3.4 and 4 regions. Incidentally, correlation skills for the Niño 4 region show the largest observed discrepancies among the different models, with BNN dominating at longer lead times (9-15 months), and SVR at short lead times (3 and 6 months). Incorporating the WWV tended to reduce the differences in correlation among the 3 methods due to the linear coupling between WWV and SST [6]. For region Niño 3.4 the differences in correlation scores before and after the addition of WWV are less pronounced. The superiority of the nonlinear methods is more apparent in the region of largest variability, Niño 1 + 2, over short lead times of 3-9 months. However, the correlation scores for LR show improvements at long lead times for all regions except Niño 4. This is consistent with the forecast results of PC1 (Figure 4.6a), where LR provides better correlation scores at 6-15 months lead time.

An alternative presentation of the results for the various Niño regions is given by plotting simultaneously the time series of observed values and the predicted values of LR and one of the nonlinear models. Figures 4.9 and 4.10 illustrate the time series of the observed El Niño 3.4 compared to the predicted time series of the LR, BNN and SVR models (without WWV as a predictor). Notice that at 3 month lead time all of the models successfully predict the major El Niño episodes of 1982/1983, 1986/1987, 1991/1992, and 1997/1998. Beyond 9 months lead, identification of anomalies becomes difficult, however, the BNN prediction and the observed anomaly closely match for El Niño 1982/1983, 1986/1987, and La Niña 1984/1985, 1988/1989. The time series for the other 3 Niño regions and for Niño 3.4 with WWV added are provided in Appendix A.

The SST anomalies were reconstructed by combining the PCs forecasts with their corresponding EOFs. Contour maps of the correlation scores between the predicted and observed anomalies for the nonlinear models are provided in Figures 4.9 and 4.10.
Table 4.3: Values of the hyperparameter set \( \{\gamma, c, \varepsilon\} \). The (*) denotes the values after WWV was added to the predictor set. The large value of the hyperparameters \( c(*) \) reflects the use of nonstandardized predictands in the models. The relative large value of \( \varepsilon \) is due to the large amplitude of the PCs (Figure 2.2).
Figure 4.7: Correlation scores before and after WWV was added to the predictor set for the Niño 4, Niño 3.4, Niño 3, and Niño 1 + 2 regions.
Figure 4.8: RMSE scores before and after WWV was added to the predictor set for the Niño 4, Niño 3.4, Niño 3, and Niño 1 + 2 regions.
Chapter 4. Forecast Results

Nino3.4 SSTA (1980–2005)

(a) 3-month lead

(b) 6-month lead

(c) 9-month lead

(d) 12-month lead

(e) 15-month lead

Figure 4.9: Time series for the Niño 3.4 region, with BNN (black line), LR (thin line) predictions and observed values (thick grey line).
Chapter 4. Forecast Results

Figure 4.10: Time series for the Niño 3.4 region, with SVR (black line), LR (thin line) predictions and observed values (grey line).
4.11 and 4.12. The advantage in the nonlinear modelling of ENSO shows up at 6 and 9 months lead time in the regions with largest variability (Fig. 4.11d and f). At these lead times, SVR reveals advantage of nonlinear modelling off the equator along the dateline (Figure 4.12).

Adding the WWV index as a predictor effectively reduces the differences in prediction skill between the nonlinear and linear models at long lead times. The prediction advantage at lead times of 9-15 months of BNN is reduced when WWV was added as can be seen by comparing corresponding panels of Figures 4.11 and 4.13. Similarly, the difference in skill between the SVR and linear models is also reduced, however at short lead times of 3 and 6 months the positive differences are enhanced in the eastern and western regions of the domain (Figure 4.14). Comparing LR against itself after WWV was added, reveals an increase in correlation skill emerging at long lead times of 9-15 months concentrated on the central-eastern equatorial region of the tropical Pacific (Figure 4.15). This increase in skill corresponds to the region of maximum variability, which seems to indicate that SST in the central-eastern region of the tropical Pacific responds more to changes in the thermocline depth than the western region as formulated in Jin's oscillator model (1.9). As observed through the forecast of PC1 when WWV was the sole predictor (Figure 4.6a) both the linear and nonlinear models indicate a high correlation between thermocline depth and SST variations; LR leads the correlation scores at all lead times except for 15 months.

4.2.1 Discussion and Conclusion

The forecast results of the 5 leading PCs of SST, indicate evidence of nonlinear structures in the ENSO cycle. This nonlinear structure is mainly present in the second mode of the SST field (Figures 4.1 and 4.5). The central and eastern regions displayed the highest nonlinear tendencies before the WWV is included. The effect of WWV on the dynamics of the SST anomaly along the central-eastern tropical Pacific region appears to be governed by linear dynamics as no advantage was observed in the forecasts of the nonlinear models when WWV was included (Figures 4.13, 4.14, and 4.15). Stated in [14], the WWV index embodies the large-scale low frequency dynamics. As such, the response of the SST field to the WWV index
Figure 4.11: BNN forecast correlation performance and its difference from that of the linear model for the period 1980-2005 before WWV was included in the predictor set. The contour interval is 0.1 in the right column and 0.05 in the left.
Figure 4.12: SVR forecast correlation performance and its difference from the linear model for the period 1980-2005 before WWV was included in the predictor set. The contour interval is 0.1 in the right column and 0.05 in the left.
Figure 4.13: BNN forecast performance and its difference from the linear model for the period 1980-2005 after WWV was included in the predictor set. The contour interval is 0.1 in the right column and 0.05 in the left.
Figure 4.14: SVR forecast performance and its difference from the linear model for the period 1980-2005 after WWV was included in the predictor set. The contour interval is 0.1 in the right column and 0.05 in the left.
Chapter 4. Forecast Results

Figure 4.15: LR forecast performance and its difference from LR without WWV, (1980-2005). The contour interval is 0.1 in the right column and 0.05 in the left.
comes with a particular time delay, with a maximum at 6 to 12 months lead time. This observation is in agreement with our results, where the maximum correlation between the SST field and WWV was recorded at 12 months lead time.

The forecasts for the 1950-2005 data indicated a marginally nonlinear response in SST at short lead times over certain areas of the domain. These areas are mainly centered off the key Niño regions along the 180° meridian. This nonlinear structures may point to nonlinear effects of the ENSO cycle at higher latitudes [18], Similar structures appeared in the case of SVR modeling for the later period (1980-2005) but with higher amplitudes of the correlation scores. In general an increase in the performance of the nonlinear models relative to linear regression was observed. As argued by [7], changes in the amplitude and spatial distribution of the SST anomalies could signal a changing nature of El Niño, these changes could be the result of interdecadal variations of the background state, perhaps triggered by global warming. Evidence of these changes could be the recent increased intensity in the warm anomalies, namely the 1982/83 and 1997/98 events. These extreme warm episodes could be responsible for the improved performance of the nonlinear modeling in the shortened time series of 1980-2005 which allowed theses events to stand out.

Nevertheless, the increase in correlation skill achieved by isolating the more extreme SST time series anomalies led to only modest advantages in forecasts skill provided by the nonlinear models. Tang et al. (2000) [18] suggested three possibilities to explain the limited performance of our nonlinear models. One, the ocean-atmosphere interaction over the tropical Pacific Ocean is essentially linear. Two, the time scales, and data resolution used here are inadequate for extracting the nonlinearities of ENSO as stated in the Cane-Zebiak dynamical model. The third possibility addressed in [18] is that neural network training is not yet optimized, this however is unlikely given the use of SVR which reduces the decisions made during training. The two methods yielded similar results.

An important aspect of this study was to investigate the role that warm water volume plays in the development of the SST anomalies. As indicated in [14] the index of WWV embodies the large scale ocean dynamics associated with ENSO. Further confirmation that the volume of warm water (WWV) above the 20°C isotherm integrated across the basin is linearly related to the ENSO cycle on the
times scales investigated here was obtained [14][15]. It was also observed that the WWV index alone accounts for as much of the variance as the entire set of predictors at lead times of 6 months or greater (Figure 4.6a). However, for PC2, WWV alone did not contribute much skill. Likewise, PC3 appears to have no dependence on the WWV index, as correlation skills generally dropped at all lead times for linear and nonlinear regression methods (Figure 4.5). These results suggest that the linear coupled models stated in various versions of the recharge oscillator, e.g. [9], are adequate for modeling the SST and thermocline depth relationship. Furthermore, the maximum correlation scores between SST and WWV occurred at approximately 12 months lead time which is in agreement with [15]. This increase in correlation scores is concentrated mainly in the eastern region of the tropical Pacific which supports the hypothesis that the dynamical modeling should involve the variables, WWV and \( T_E \) (SST in the eastern region) with some time lag (Figure 4.15).

Forecast of the SST anomalies over the tropical Pacific were conducted using linear and nonlinear methods. Representation of the data in terms of principal components significantly reduced the number of variables needed to be handled, thus avoiding issues associated with large dimensionality. The first 5 PCs of SST account for more than 85% of the total variance of the SST field. Independent forecasts of each of these PCs was carried out using LR, BNN, and SVR. Our results show some general nonlinear features of ENSO, but failed to establish a clear advantage in using BNN or SVR over LR in the forecast of SST, particularly in the analysis of the 1950-2005 period. The nonlinear features showing up in the forecasts of SST were of second order in the variance, namely PC2, these are located along the regions of largest variability (central-eastern tropical Pacific). This nonlinear advantage was reduced when WWV was added to the predictor set. We concluded that ENSO, at interseasonal times scales is essentially linear, the nonlinear features account for only a small percentage of the variance. As pointed out by [18], other nonlinear elements of ENSO could be contained in the seasonal cycle. However, the seasonal cycle was eliminated in this analysis.
4.3 Further Research

The forecast model development presented here could be extended to a third nonlinear model known as *Gaussian Processes*. Gaussian Processes is another kernel method, which extends the use of kernels to the realm of probabilistic forecast in the Bayesian setting.

There are many aspects of ENSO left uninvestigated. There is strong evidence that many processes associated with this phenomenon exhibit nonlinear behavior, for example the locking of El Niño events with the seasonal cycle are difficult to explain with linear methods, and the asymmetries between El Niño/La Niña amplitudes. Low order chaotic forcing has been proposed as a mechanism for locking ENSO to the seasonal cycle [19]; asymmetries are attributed to nonlinear dynamical heating [2]. A statistical nonlinear analysis could be applied to investigate the degree, if any, these processes have on the ENSO cycle. Furthermore, Tang *et al.* [18] proposed that nonlinear methods should also be used to study the response of the extratropics to the tropical SST anomalies.

Of particular interest is to study the effect of high frequency atmospheric forcing on ENSO variability as an extension of the present project. McPhaden *et al.* (2006) used a simple linear regression model with an index of the Madden Julian Oscillation (MJO) and the (WWV) index as predictors for the tropical Pacific SST. This simple two predictor model reportedly accounts for $\sim 60\%$ of the total SST variance at 2-3 season lead times [14]. A natural extension of this work would be to apply the nonlinear methods, BNN and SVR to this system to investigate if high frequency forcing characterized by the MJO index has a nonlinear influence on low frequency ENSO dynamics.
Bibliography


Appendix A

Time Series for El Niño 4, 3.4, 3, and 1+2
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.1: Time series for the Niño 4 region, showing the BNN (black line), and LR (thin line) predictions and the observed values (thick grey line).
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.2: Time series for the Niño 4 region, showing the BNN and LR predictions. WWV is included in the predictor set.
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.3: Time series for the Niño 4 region, showing the SVR and LR predictions.
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.4: Time series for the Niño 4 region, showing the SVR and LR predictions. WWV is included in the predictor set.
Figure A.5: Time series for the Niño 3.4 region, showing the BNN and LR predictions. WWV is included in the predictor set.
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.6: Time series for the Niño 3.4 region, showing the SVR and LR predictions. WWV is included in the predictor set.
Figure A.7: Time series for the Niño 3 region, showing the BNN and LR predictions.
Figure A.8: Time series for the Niño 3 region, showing the BNN and LR predictions. WWV is included in the predictor set.
Figure A.9: Time series for the Niño 3 region, showing the SVR and LR predictions.
Figure A.10: Time series for the Niño 3 region, showing the SVR and LR predictions. WWV is included in the predictor set.
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.11: Time series for the Niño 1 + 2 region, showing the BNN and LR predictions.
Appendix A. Time Series for El Niño 4, 3.4, 3, and 1+2

Figure A.12: Time series for the Niño 1 + 2 region, showing the BNN and LR predictions. WWV is included in the predictor set.
Figure A.13: Time series for the Niño 1 + 2 region, showing the SVR and LR predictions.
Figure A.14: Time series for the Niño 1 + 2 region, showing the SVR and LR predictions. WWV is included in the predictor set.