Drought Predictions: Applications in Australia

Geraldine H. Wong

Thesis submitted for the degree of

Doctor of Philosophy

in

Statistics

at

The University of Adelaide

Discipline of Statistics, School of Mathematical Sciences,

Faculty of Engineering, Computer and Mathematical Sciences

January 9, 2010

Chapter 1

Introduction

1.1 Droughts

Drought, in regions of the world where the majority of the population relies on subsistence agriculture, can result in widespread famine and death, for example in most parts of Africa. In contrast to an agricultural definition of drought, meteorological drought is typically defined as a prolonged period where there is below average rainfall which leads to water supply deficiency. A notable example of this was the 1 in 900 year drought in West Yorkshire, an area that usually receives abundant rainfall, in 1995 when water from Keilder Reservoir had to be transported to West Yorkshire to meet the daily needs of the public. While droughts in relatively wet developed countries do not lead to famine, they are nonetheless aggravating. The water industry was criticized over the handling of this drought event in Yorkshire [49, 88].

Australia, having experienced three major droughts over the past century, namely the Federation Drought which affected most of the country particularly in New South Wales (NSW), Queensland and Victoria from 1895 to 1902 [12]; the World War II drought affecting eastern Australia and parts of Western Australia from 1937 to 1945 [14]; and the current drought, is susceptible to both agricultural and meteorological drought. Figure 1.1 displays a time series plot of the total annual rainfall in Australia from 1900 to 2008, with a 5-year moving average curve indicating low rainfall averages during the Federation and World War II droughts, as a comparison to the current drought.

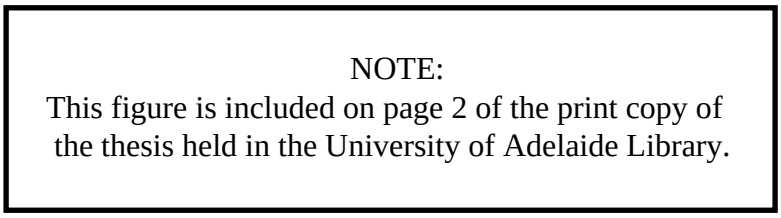


Figure 1.1: Annual rainfall of Australia with 5-year moving averages shown in black curve, 1900-2008 [11].

Drought is likely to become an even more vital issue because of: increasing population; potential increased demand per capita in countries with a relatively high standard of living; and projected temperature increase and decrease in rainfall in some areas. In early 2009, the United Nations (UN) reported that the world population will reach 6.8 billion in the current year, 313 million more than in 2005 and with projections that the population will reach 9 billion in 2050. The increase of international migrants in the world from 75 million in 1960 to almost 191 million in 2005 adds further pressure on water supplies on certain countries, in particular Australia. These figures are shown in Figures 1.2 and 1.3. In addition, the Intergovernmental Panel on Climate Change (IPCC) published temperature statistics and global precipitation trends over the last decade, which indicated an increase in temperature and decrease in precipitation in parts of world (Figures 1.4 and 1.5). Unlike Australia's rainfall, temperature is now higher and thus has implications for water supply because of evaporation.

Indisputably, the effects of droughts are especially devastating to the agricultural and social economy and have been considered the world's costliest natural disaster [61]. In 2006, extreme droughts cover about 3% of world's land area and this is predicted to spread to about 10% by 2050 [17]. It is vital that appropriate water resource infrastructure and management is established to mitigate these effects.

Being surrounded by three oceans, Australia's dry and highly variable climate is one of the major reasons why droughts have such a serious effect on the economy. It is widely reported that rainfall statistics over the current twelve-year drought are similar to the two previous major drought events [131]. Surrounded by the Indian,

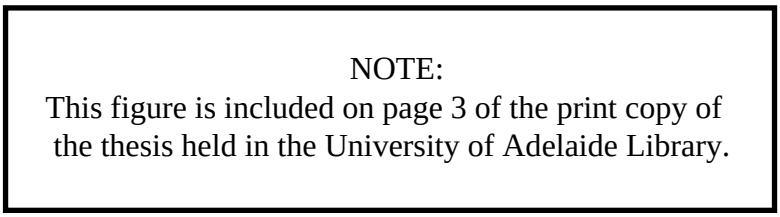


Figure 1.2: Population of the world according to different projections, 1950-2050 [130].

Figure I. Trends in the number of international migrants for the world and major development groups, 1960-2005

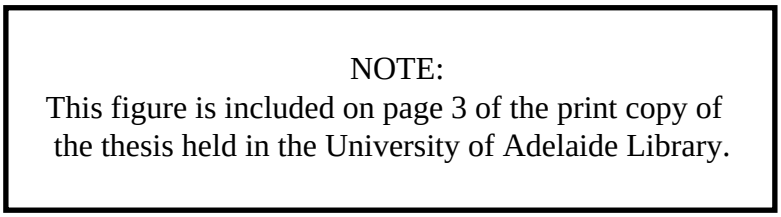


Figure 1.3: Trends in number of international migrants (in millions) for the world, 1960-2005 [129].

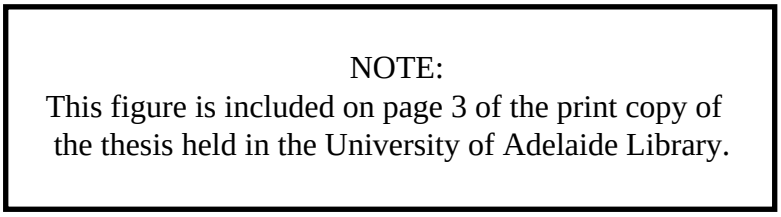


Figure 1.4: Variations of the Earth's surface temperature for the past 140 years, 1860-2000 [53].

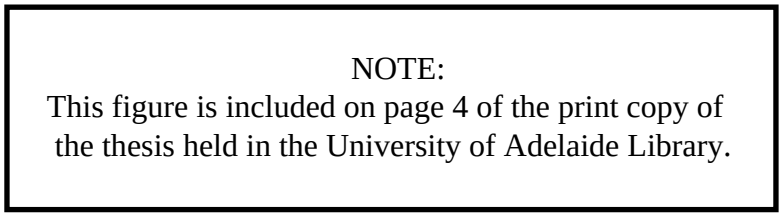


Figure 1.5: Global precipitation trends, 1900-2000 [54].

Southern and Pacific Ocean, Australia's variable climate is significantly influenced by ocean currents and pressure systems, in particular the well documented El-Niño Southern Oscillation (ENSO). Taking into account the influence of these global climatic phenomenon, extreme low rainfall can be more accurately predicted.

An intense and long drought can cause considerable damage to the local economy. Australia's agricultural sector is an important part of the national economy and other parts of the economy are highly interdependent with the agricultural economy. This sector constitute a large part of the Australian export market. As an example of drought impact in Australia, the direct loss to agricultural production from the 2002-03 drought year compared to the 2001-02 year was estimated at $7.36 billion AUD, about 1% of the GDP in 2004 [15]. A similar value of $6.2 billion AUD was estimated for the 2006-07 drought year in 2006 [8]. Improved forecasts of characteristics of a potential drought will allow the relevant authorities to better prepare and plan, for example, irrigation management. In addition to this, better information regarding the spatial extent of the drought will provide environmental agencies and environmentalists a better perspective of the impact on wildlife.

1.2 Objectives of Research

The aim of the research is to predict drought occurrence in Australia and there are two specific objectives. The first is to model droughts for risk assessment purposes, for long term planning of reservoirs and other water storages, and for government policies on water entitlements. The second objective is to investigate the feasibility of short term forecasts of droughts, particularly for farmers.

In order to study droughts, it is necessary to define them in quantitative terms, in particular their spatial extent, intensity and duration. Therefore a multivariate approach is taken to allow for the correlation structure of these features to be modelled. For the first objective, copulas are used. Copulas are multivariate uniform distributions, which allow for marginal and joint behaviour of variables to be modelled separately and their applications in hydrology is recent.

The second objective is to make short term forecasts of droughts at various spatial resolutions across Australia. The timing of such forecasts is important because farmers plant at specific times of the year. Regression methods with the predictor variables including autoregressive terms and climatic indicators, copula-based methods, and a commercial method for predicting a rainfall distribution from Southern Oscillation Index (SOI) categories are investigated and compared.

1.3 Thesis outline

An outline of the thesis structure is given in Figure 1.6 to explain the role of each chapter. Conventional drought measurements are introduced in Chapter 2 and the performance of each technique is compared when identifying meteorological drought. A brief background of important climatic oscillations and their influence on drought and low rainfall is provided. The climatic indices which are commonly used to identify and quantify these global climatic phenomena are also discussed.

For a better understanding of recent drought research, previous studies on modelling and predicting droughts are reviewed in Chapter 3. Attempts of early drought criteria aimed at quantifying droughts based on a standard universal measure, are examined. To account for various drought categorizations, several drought indicators were later developed to overcome limitations arising from existing criteria. This led to the con-

struction of popular drought quantification methods such as the Palmer Drought Index (PDI) and Standardized Precipitation Index (SPI). If drought is described only in terms of the duration, an index of being below a threshold, other important drought features such as the corresponding severity and peak intensity are not considered. Traditional multivariate approaches which describe the correlation structure between these drought characteristics are reviewed and their limitations are discussed. Recent copula applications in hydrology are examined as an improvement over traditional methods. The bivariate and trivariate Archimedean copulas are analyzed, and use of a more general Archimedean copula form is introduced to account for several dependence parameters between pairs of drought variables. The elliptical copula family is also presented. Finally, this chapter reviews approaches to forecasting drought using stochastic models. Previous studies which take advantage of relationships between drought and global climatic phenomena such as ENSO, are also investigated and evaluated.

Chapters 4 and 5 focus on the modelling and simulation of Australian drought using copulas. The theory and structure of the Elliptical and Archimedean copula families are presented. Tail dependence is defined for each copula family, since this is crucial when examining the dependence structure of extreme values, in this case droughts. To demonstrate copula modelling, monthly regional precipitation data from a rainfall district in NSW is used. The drought characteristics of duration, severity and peak intensity, are first evaluated and appropriate marginal distributions are fitted to each characteristic to transform them on a uniform scale. Trivariate Gaussian and asymmetric Gumbel-Hougaard copulas are then fitted to compare their modelling performance. Goodness-of-fit tests and upper tail dependence measures from simulated replicates are also carried out as an examination of their performance. The results demonstrate that the Gumbel-Hougaard copula is more appropriate for modelling the drought characteristics, particularly when upper tail dependence exists.

Having established the utility of copula models, Chapter 5 investigates the effect of ENSO on the dependence structure of these same drought characteristics. Daily precipitation data from two regional districts in NSW, on either side of the Great Dividing Range, are segregated into three states, El-Niño, Neutral and La-Niña, according to the prevailing SOI. Gumbel-Hougaard copulas and t trivariate copulas from the Elliptical copula family are fitted to the droughts in the three states. The effect of the ENSO state on the estimated copula parameters is presented and discussed. The goodness-

of-fit of the Gumbel-Hougaard and t copulas are compared, and the limitations of the two copula models are discussed. The times between drought events are also analyzed according to the ENSO state they occur in. The fitted copulas are used to estimate annual recurrence intervals of (i) at least one of the three variables, and (ii) all three variables exceeding critical values taking account of the mixture of states.

Having modelled the behaviour of drought and discussed its nature, the remaining chapters are devoted to forecasting drought in Australia. Adaptive stochastic models and their performance are evaluated in Chapter 6. Using precipitation data and its corresponding SPI values from three rainfall gauges in NSW, stochastic models predicting the three-month SPI at lead times of one and two months are compared on the basis of root mean squared error (RMSE) and a RMSE restricted to drought periods. The benefits of including SOI and MEI in the models are investigated. Similar models for predicting the twelve-month SPI at a lead-time of six months are likewise compared.

Chapter 7 examines popular probabilistic rainfall forecasting techniques which take into account global climatic indicators, and provides alternatives for comparison. The first part of this chapter investigates the Rainman software, developed by Stone [119], that provides probabilistic forecast of seasonal rainfall based on the changes in SOI and has been used extensively by farmers in Australia. This forecasting strategy which takes into account the change in SOI between April and May to forecast June to October rainfall, is compared with the newly developed strategy of using only the SOI of May. Significance testing is performed for all pixels across Australia based on multiple comparison techniques, that allow for a large number of two-site comparisons. Comparisons are made by computing the percentage of pixels across Australia for which a significant difference between these extreme categories exist. A third strategy compares the June to October rainfall in El-Niño years with that in La-Niña years. This form of classification is recommended to be used with physically-based climate models which are capable of predicting the ENSO state up to two years ahead [21]. The second part of Chapter 7 examines the correlations between the global sea-surface temperature (SST) anomalies with monthly rainfall in Australia. The highest correlated SST pixels are chosen for a particular rainfall location in Australia and multiple regression models are fitted for predictions.

The copula concept is again applied in Chapter 8, together with the SOI categorizations introduced in the previous chapter, to improve June to October rainfall predictions in neighbouring rainfall districts on the eastern coast of Australia. Appropriate marginal

distributions are fitted to district rainfall and the multivariate dependence structure between these variables is modelled through a copula. Separate copulas are fitted to historical data, segregated according to their SOI states, and their parameters are estimated using the Maximum Likelihood Estimation (MLE) method. Statistical tests demonstrate that there is statistical significant difference between the fitted copulas in the different SOI state. Lastly, relevant conclusions and findings of this research are reported in Chapter 9. Further research extensions and recommendations are also suggested.

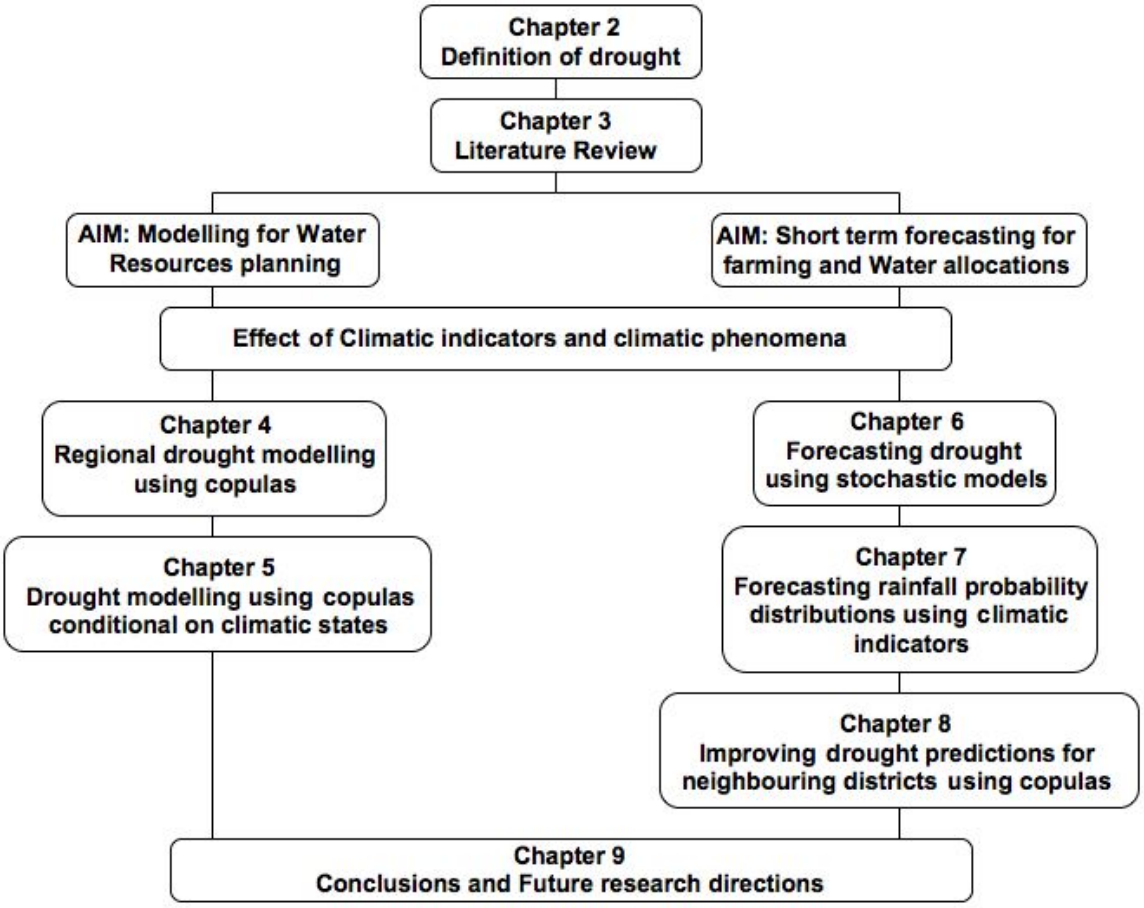


Figure 1.6: Outline of thesis structure

Chapter 2

Defining drought and global climatic phenomena

Droughts are considered to be the world's costliest natural disaster [61] and, in particular, have been a prevalent feature of Australia's climate. This climatic hazard has affected all regions of the world to different degrees, from threatening lives in Africa to loss of crops in Australia and the imposition of water restrictions for the urban population in the society. In order to reduce its effects, it is vital that a system of forecasting for the occurrence of drought is established. This chapter aims to introduce prominent climatic phenomenon and measurements of drought and examine the influence of global climatic phenomenon and anomalies affecting Australia's highly variable rainfall.

2.1 Measuring drought

The extent of damage which result from droughts affects various sectors of the society to differing degrees. This is due to varying water demands by the population. Hence, defining drought is problematic, since it holds different meanings to different sectors of the society. This gives rise to the issue of establishing a universal definition. In general, drought is categorized using either meteorological, hydrological, agricultural or socio-economic definitions [136]. Each of these categories has its own basis when defining drought in terms of physical aspects. Meteorological drought is defined based on the amount of rainfall, whereas hydrological drought is determined by the shortage of water in reservoirs and rivers, and agricultural drought is measured based on the

short term dryness of the surface layers which result in reduction of crop yield. Unlike the three physical classifications of drought, the socio-economic drought focuses on the consequences of drought and will depend on the economic activities of the society. Table 2.1 summarizes the common drought indices associated with the four main drought categories [51, 61]. Figure 2.1 displays the overlap of the effects of each drought category. The onset of meteorological drought is the first and is the driving force for the other drought categories, however, it must be remembered that it is usually a relative term. The onset of an agricultural drought may lag the meteorological drought, depending on previous moisture present in surface soil layers. The effects of hydrological drought will persist long after a meteorological drought ended [51].

Table 2.1: Drought categories and their common drought indices

Drought category	Variables analyzed	Examples of drought indices
Meteorological	rainfall	Discrete and cumulative precipitation anomalies, rainfall deciles, Palmer Drought Severity Index (PDSI), drought area index, SPI
Hydrological	shortage in bulk water supply, snowpack, streamflow with an allowance for temperature	Total water deficit, Cumulative streamflow anomaly, Palmer Hydrological Drought Index (PDHI), Surface Water Supply Index
Agricultural	soil moisture deficit	Crop Moisture Index (CMI), Palmer Moisture Anomaly Index, Computed soil moisture, Moisture Adequacy Index
Socio-economic	supply and demand of some economic good affected by meteorological, agricultural and hydrological drought	Many various indices

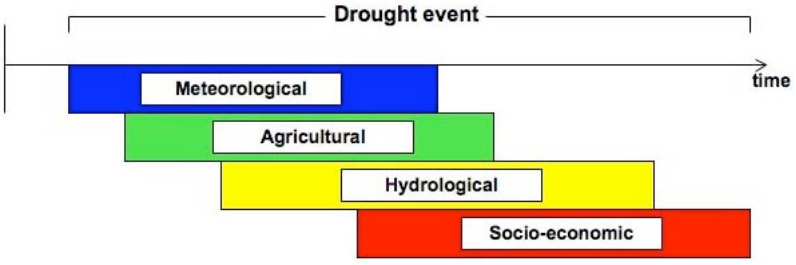


Figure 2.1: Possible overlap of drought phases (categories)

2.1.1 Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (SPI) is a dimensionless index, developed by McKee [70], as an indicator for the identification and analysis of drought. It defines rainfall deficiency relative to the expected distribution for a location and season, and is therefore usually considered to be seasonally adjusted and insensitive to topography. Precipitation shortage can be computed on multiple time scales, for the purpose of describing both short and longer-term drought, and to allow for drought conditions to be described over a range of meteorological, hydrological and agricultural settings. More importantly, SPI provides a platform for standardization and aims to provide consistency between drought occurrences at different locations.

The calculation of the SPI from monthly (t) rainfall totals (r_t) for a chosen period (m) is based on the backward moving average of length m calculated on month t:

$$y_t = \frac{r_t + r_{t-1} + \cdots + r_{t-m+1}}{m} \tag{2.1}$$

Suitable probability distributions are then fitted to the $\{y_t\}$ for each calendar month (j). The two-parameter Gamma distribution is commonly chosen and its probability distribution function (*pdf*) is:

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \leq x < \infty,\ \alpha > 0,\ \beta > 0. \tag{2.2}$$

where α is the shape parameter, β is the scale parameter and $\Gamma(\alpha)$ is the gamma function. McKee [70] defined the SPI (S_t) as a composition of the inverse cumulative distribution function (inverse *cdf*) of the standard Normal distribution and the *cdf* of the fitted Gamma distribution, that is $S_t = \Phi^{-1}(F(y_t))$.

Since separate Gamma distributions are fitted to each month, the SPI is a seasonally adjusted index with a standard normal marginal distribution. It follow3s that approximately 95% of SPI is expected to lie in the range of -1.96 to $+1.96$, with negative values representing relatively dry periods and positive values indicating relatively wet periods. On this scale, droughts are defined as periods during which the SPI remains below a threshold of -1 and ends when the SPI rises above this threshold. [70] .

The SPI is typically calculated for 1 up to 24 month time scales, which requires the length of the moving average in Equation (2.1) to be modified accordingly. The SPI(3) is commonly used as a measure of relatively short-term drought while a longer-term

drought is adequately described by SPI(12) or SPI(24). However, the same principles apply for any chosen duration. Figures 2.2(a), (b), (c) and (d) shows the time series of the monthly rainfall, SPI(3), SPI(6) and SPI(12) respectively from District 63 rainfall data in NSW.

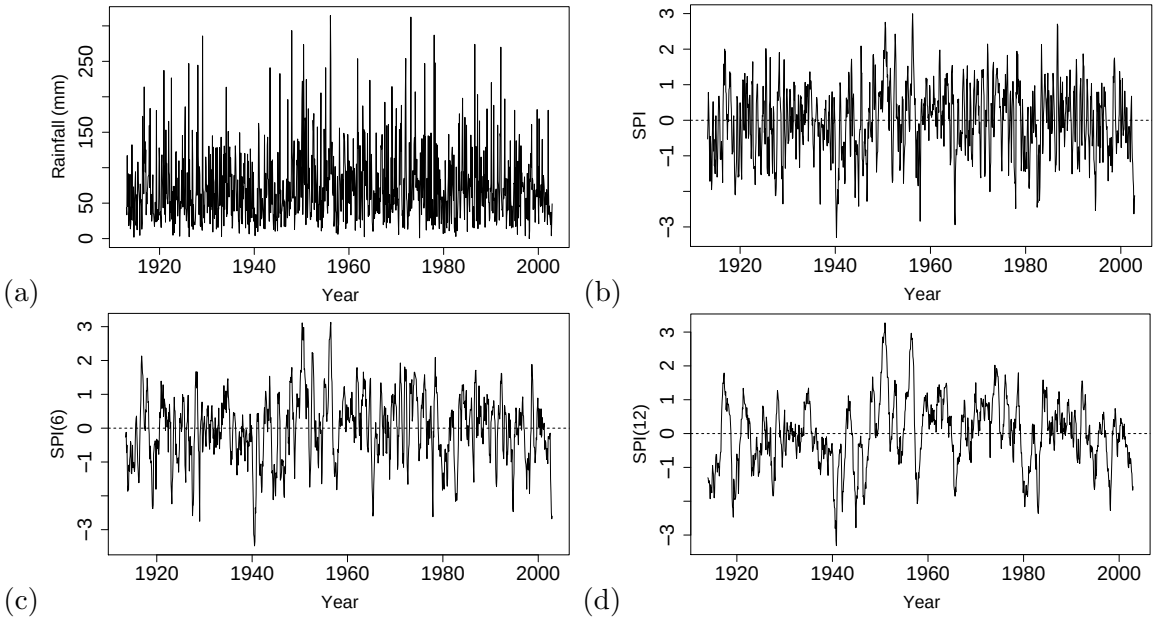


Figure 2.2: Time series of (a) monthly rainfall; (b) SPI (3); (c) SPI (6) and (d) SPI (12)

The effect of transforming rainfall to SPI on different time scales is evident in Figure 2.2. The peaks and troughs in Figure 2.2(a) do not necessarily correspond directly to a peak or trough in Figure 2.2(b), since each month's rainfall is transformed separately, and the SPI measurement is calculated relative to a given month and hence is a measure of deviation from the mean zero. Observe that as SPI increases from SPI(3) to SPI(6) and SPI(12), there is longer escalation above and below zero and the smoothing effect is more apparent, from taking longer moving averages.

2.1.2 Palmer Drought Severity Index (PDSI)

Named after Palmer [87] and widely used by the American Meteorological Society, it was created to "measure the cumulative departure of moisture supply". Similar to the SPI, the PDSI is also dimensionless with a range of -4 to 4, where negative values signify water deficiency and drought risk.

The PDSI is designed to compute meteorological drought and the computation procedure is as follows [51, 47]: First, a monthly hydrologic accounting is carried out us-

ing historic records of precipitation and temperature. Palmer [87] used a two-layered model for soil moisture computations and made assumptions pertaining to the field capacity and transfer of moisture to and from the layers. The results from the hydrologic accounting system are first summarized to obtain coefficients of evapotranspiration, recharge, runoff and loss, which are dependent on the climate of the analyzed area. The data series are then reassessed using the derived coefficients to determine the amount of moisture required for normal weather, which Palmer [87] called Climatological Appropriate for Existing Conditions (CAFEC) precipitation. Monthly departures from normal weather conditions are then transformed to moisture anomaly indices by a standardizing factor designed to account for variations in climate at different locations. These moisture anomaly indices are known as the Palmer Z Index.

Guttman [47] demonstrates the above procedure using the following equation to compute the CAFEC precipitation P_c :

$$P_c = aPE + bPR + cPRO - dPL$$

where PE is the potential evapotranspiration computed using Thornthwaite's method [125] when PE is greater than the monthly precipitation. Potential recharge (PR) refers to the amount of moisture required to replenish the soil to field capacity. PL is the potential loss, which is the amount of moisture which could be lost from the soil to evapotranspiration if the precipitation was zero. The potential runoff (PRO) is defined as the difference between potential precipitation and potential recharge. The coefficients, a, b, c, and d are obtained by dividing the mean actual quantities by the mean potential quantities.

The moisture anomaly index Z, which essentially determines the relative departure from the average moisture conditions for that month is then expressed as:

$$Z = (\text{observed precipitation} - P_c) \times f$$

where f is the standardizing factor. The PDSI for month t is then calculated by the equation:

$$\text{PDSI}_t = 0.897\text{PDSI}_{t-1} + 0.333Z_t$$

where Z_t is the precipitation deficit. From the above equation, it is observed that the

PDSI for a given month depends heavily on the previous month's PDSI, hence containing the memory of previous moisture conditions.

2.1.3 Other drought classification

Rainfall deciles are used by Australia's Bureau of Meteorology (BoM) to measure meteorological drought. This form of classification is used when the mean is considered a poor reference for typical conditions and the median is used instead [61]. Rainfall deciles are calculated by first ranking the observed precipitation totals for the preceding three months against past records. The district is defined to be 'drought affected' when the sum falls within the lowest decile (ninth percentile). Figure 2.3 shows the rainfall deciles map produced by BoM for 1-month, 6-month, 12-month and 24-month periods up till August 2008.

Figure 2.3: Rainfall deciles map of (a) 1-month; (b) 6-month; (c) 12-month and (d) 24-month (BoM, 2008)

If only the current month is considered, rainfall shortage in Australia does not seem to be a major issue in August 2008, except for parts of Western Australia and north-east of Queensland. In contrast, there is widespread rainfall shortage around the country when rainfall deciles are calculated for the preceding 6 months. Similar observations are drawn for both 12 and 24 month periods.

2.2 Global climatic phenomena

Global climatic phenomena and extreme weather patterns have had a long history. Since the early 20th century, British researchers have been interested in the links between atmospheric phenomena and extreme weather patterns [3]. Such research paved the way of future studies, for instance Sir Gilbert Walker [132, 133], who described the El-Niño Southern Oscillation (ENSO) phenomenon. These global climatic phenomena have an extensive influence on weather patterns in Australia, hence it is important to understand their role, which may improve drought predictions. The following sections present a brief background for each climatic phenomenon and a description of their mechanism.

2.2.1 El-Niño Southern Oscillation (ENSO)

The term "El-Niño" is the name given by Peruvian sailors to a seasonal, warm southward-moving current along the Peruvian coast every few years [45]. Scientifically, the El-Niño Southern Oscillation is the global interaction between the ocean and the atmosphere. The two extreme phases that result from this interaction are commonly known as El-Niño and La-Niña. The El-Niño phenomenon tends to occur every 2 to 7 years and its effects can last up to 2 years [31]. Walker first described the "see-saw" fluctuations in atmospheric and rainfall in the Indo-Pacific region as the Southern Oscillation (SO) [132, 133]. This caused a pressure increase in locations around the Indian region and coincided with a decrease in pressure and rainfall over the Pacific region. This movement of air pressure from a high to low pressure region is known as the Walker circulation. Figure 2.4(a) shows the system of Walker circulation and Figure 2.4(b) shows the movements in air pressure and ocean during normal conditions.

During an El-Niño phase, trade winds are weak and air pressure falls over central and eastern Pacific. This is the result of warming sea-surface temperatures that develop in this region, which leads to a rise in air pressure over the Indian Ocean and Australia, bringing rain along with it. These prolonged movements of the atmospheric and ocean patterns cause extensive droughts across Australia, southern Africa, northern India and Southeast Asia [102] and torrential rains in northern Peru and southern Ecuador [45]. This El-Niño behaviour is shown in Figure 2.5(a).

On the other hand, when a La-Niña phase occurs, the resulting effects of this phe-

nomenon are opposite in nature to those described above. In general, it is characterized by wetter than usual conditions in the Australasian region, southern Africa and northern India and a reduction in rainfall observed in central and northern Pacific. Figure 2.5(b) shows the development of the La-Niña event.

(a) (b)

Figure 2.4: (a) Walker circulation; (b) Atmospheric and ocean movements during normal conditions [71].

(a) (b)

Figure 2.5: Atmospheric and ocean movements during (a) El-Niño conditions; (b) La-Niña conditions [71].

2.2.2 Inter-decadal Pacific Oscillation (IPO)

Another source of natural climate variability that influences Australia's rainfall, is the IPO. This phenomenon behaves in a similar pattern to the Pacific Decadal Oscillation (PDO) that measures the changes in SST and sea-level pressure (SLP) in the Pacific Ocean, but with a 15 to 30 year cycle. Folland *et. al.* [39] describes the IPO as a quasi-symmetric Pacific-wide appearance of the PDO that describes both North and

South Pacific. A number of indices for the IPO are derived from other sources through Empirical Orthogonal Function analyses of SSTs [91]. Figure 2.6 shows the time series plot of the IPO index over the last century, where there is a noticeable shift between the positive (warm) phases and negative (cold) phases.

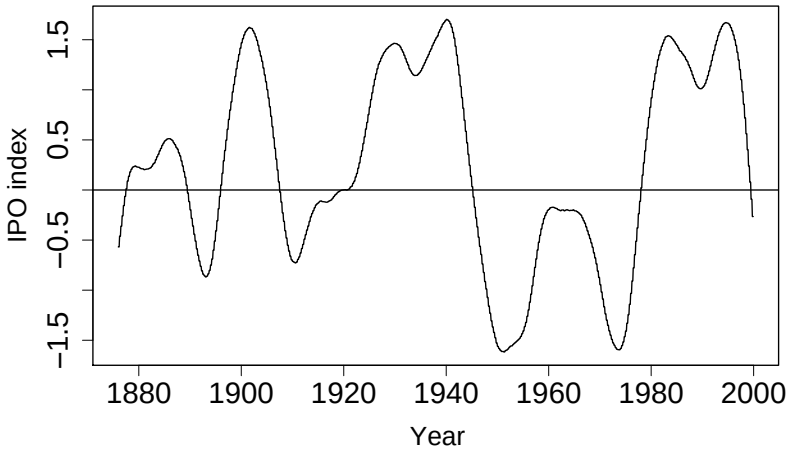


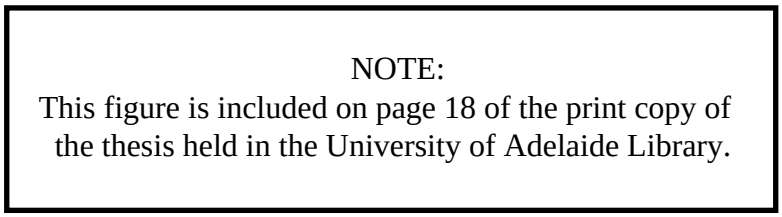
Figure 2.6: Time series of IPO index, 1876-1999

When the IPO index is positive, SSTs in the tropical Pacific increases while the SSTs to the south and north are cold [91]. The opposite behaviour is noticed when the IPO index is in the negative phase. Folland *et. al.* [39] noted that a shift in the IPO changes the location and strength of ENSO. When this shift occurs, the South Pacific Convergence Zone, which is characterized by a band of low-level convergence, cloudiness and precipitation lying from the west Pacific warm pool towards French Polynesia, moves northeast during El-Niño episodes and southwest during La-Niña events. The same movement takes place during positive IPO and negative IPO phases respectively. Hence the impact of ENSO on Australia varies in relation to the IPO.

2.2.3 Other global climatic phenomena

Another recently identified ocean-atmospheric interaction affecting climate is the Indian Ocean Dipole (IOD). This pattern of inter-annual variability over the Indian Ocean region is characterized by anomalously low SST off Sumatra in Indonesia and warmer than usual SST in the western Indian Ocean, accompanied by winds and precipitation anomalies [104]. This interaction which oscillates between positive and negative phases, appears to be independent of ENSO and accounts for about 12% of SST variability.

The IOD phases are determined by the Dipole Mode Index (DMI), which is defined as the standardised SST anomaly difference between tropical western Indian Ocean (50° - 70°E, 10°S - 10°N) and the tropical south-eastern Indian Ocean (90° - 110°E, 10°S - equator). The positive phase is characterized by the weakening of winds and an upwelling of waters over Indonesia and north of Australia, resulting in a decrease in SST and reducing the amount of moisture absorbed and transported across Australia. The consequence is below average rainfall and droughts in Australia and Indonesia. Correspondingly, the opposite is observed in eastern Indian Ocean, where there is cooling of waters. During the negative phase, there is a reversal of these conditions, with warmer water and higher than usual precipitation in Australia. These descriptions are represented in Figure 2.7 [128].



(a) (b)

Figure 2.7: Schematic figures of the (a) negative phase and (b) positive phase of the IOD

Recently, Saji *et. al.* [105] demonstrated that the IOD has an impact on winter climate in the southern hemisphere during June to October. Using atmospheric general circulation models, Ashok *et. al.* [7] also showed that cold SST anomalies prevailing west of the Indonesian archipelago during positive phases, introduces an anticyclonic circulation over most of Australia. Partial correlations between DMI and Australian rainfall showed significant influence on winter rainfall over western and southern Australia, suggesting that a possible forecasting relationship could be developed.

2.3 Measuring global climatic phenomena

Identifying the natural phenomena that affect Australia's rainfall pattern is a critical step. The next step then is to establish the climate indices which are characteristics of these natural phenomena. Explanations on the measurement and calculation of each

global climate indices and their effects on rainfall are provided in this section.

2.3.1 Southern Oscillation Index (SOI)

A common measure of the ENSO phenomenon is the SOI. The SOI is the difference in sea level pressure across the equatorial Pacific Ocean. There are various methods of calculating the SOI. The most popular is the Troup SOI which is the standardized anomaly of the Mean Sea Level Pressure (MSLP) difference between Tahiti and Darwin. The BoM calculates this as follows [10]:

$$SOI = 10 \times \frac{Pdiff - Pdiffav}{SD(Pdiff)}$$

where
$Pdiff$ = (average Tahiti MSLP for the month)−(average Darwin MSLP for the month)
$Pdiffav$ = long term average of *Pdiff* for the month
$SD(Pdiff)$ = standard deviation of *Pdiff* for the month calculated over the long term.

The mean sea level pressure is calculated from long-term monthly observations. The Troup SOI has a range from −35 to +35, and is usually calculated monthly, since daily observations of the SOI do not provide useful information about the current climate state.

Commonly, sustained periods of negative SOI values are used to define El-Niño events, while persistent periods of positive SOI values are taken to define La-Niña events. There are various ENSO classification methods, which make use of SOI. Chiew *et. al.* [24] classified El-Niño as years during which the 12 month (April to March) average SOI falls below −5 and La-Niña as years when that average lies above 5. However, Ropelewski and Halpert [103] defined El-Niño events as years where the SOI five-month running mean remains below minus half standard deviations for 5 months or longer between April and March and La-Niña events occur when this five-month running mean is above half the standard deviations for 5 months or longer. Figure 2.8 shows a time series plot of the SOI data obtained from Australian BoM, from 1913 to 2002.

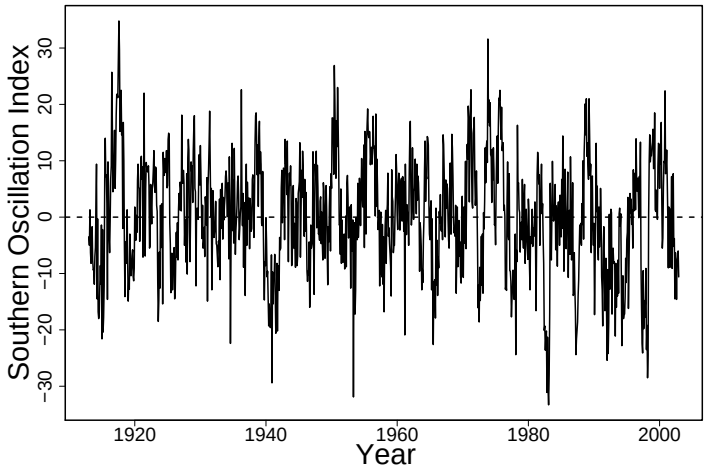


Figure 2.8: Time series plot of SOI, 1913-2002

2.3.2 Multivariate ENSO Index (MEI)

Another recent development of climatic indicators associated with the ENSO phenomenon is the Multivariate ENSO Index (MEI). The MEI is constructed by combining several observed climate parameters: sea-level pressure, zonal and meridional components of surface wind, sea surface temperature, surface air temperature and total cloudiness fraction of the sky. The inclusion of these climate parameters makes the MEI less susceptible to non-ENSO related variability since the index is able to explain the ocean-atmospheric interactions better than indices that rely only on a single variable [63].

The calculation of MEI starts by computing moving averages of length 2 months for each of the twelve months. Then, spatial filtering of individual fields into clusters, through principal component analysis of the six observed climate parameters is carried out. The variance in each climate parameter is normalized and the first principal component of the covariance matrix of the combined climate parameters is obtained [137]. Seasonal values are then standardized according to each season and to the 1950-93 reference period.

Continuously negative MEI values represent La-Niña events, while positive values indicate El-Niño events. Figure 2.9 displays the monthly MEI values from 1950 to September 2008. Long escalations above 0 is observed during the early 1980s and most of 1990s. These positive values of the MEI coincides with the occurrence of droughts

occurring in Australia in 1982 to 1983 and also the severe drought that started in the second half of 1991 which escalated in 1994 and 1995. Hence, the MEI could have some influence on the occurrence of droughts in Australia.

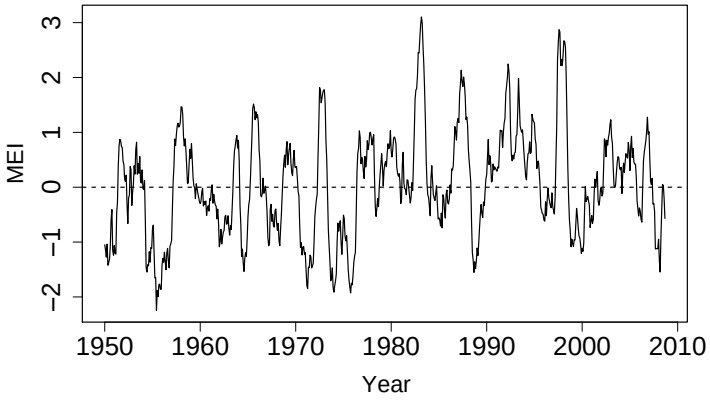


Figure 2.9: Time series plot of MEI, 1950-2008

2.3.3 Sea-surface temperatures (SST)

The SST anomalies are another climatic indicator that measure the ENSO phenomenon and are the differences between the observed SST and climatological SST. The low frequency variation in SST due to the thermal inertia of the ocean, makes the SST a suitable prediction tool for the prediction of the ENSO phenomenon.

There are four regions where SST anomalies are measured in the world and they are classified as follows with their respective longitudinal and latitudinal directions given: Niño 1 + 2(0 - 10 °S)(80 ° - 90 °W), Niño 3(5 °N - 5 °S)(90 ° - 150 °W),Niño 4(5 °N - 5 °S)(160 °E - 150 °W), Niño 3.4(5 °N - 5 °S) (120 ° - 170 °W). Trenberth [127] compared the SST anomalies from Niño 3 and Niño 3.4 regions and discovered that the SST anomalies obtained from the Niño 3.4 region gave the best match of classifying ENSO periods accurately when compared to historical records. Niño 3.4 region is most commonly used by environmental agencies, since changes in SST anomalies in this area is critical in influencing rainfall in a large region in the western Pacific.

An El-Niño event is defined when the 5-month moving average of the SST anomalies exceed +0.4°C for at least 6 consecutive months. Conversely, a La-Niña event is described as when this moving average falls below −0.4°C. The 5-month moving average is used to reduce the amount of variation not associated with ENSO. The time series plot of the SST anomalies from Niño 3.4 region is displayed in Figure 2.10. The

thresholds of $+0.4°C$ and $-0.4°C$ are denoted by dotted lines on the graph.

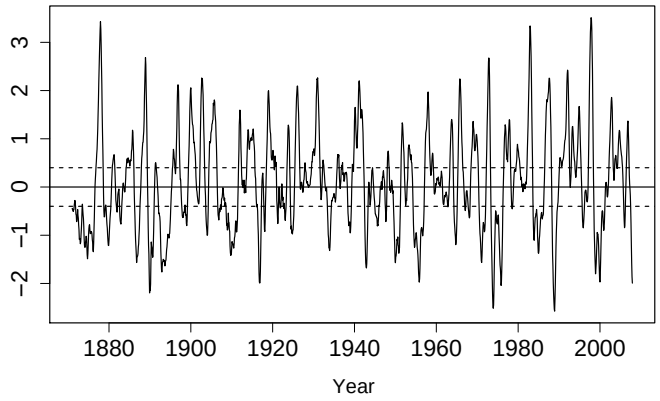


Figure 2.10: Time series plot of monthly SST anomalies in Niño 3.4 region, 1871-2007

2.4 Summary of chapter

Different drought definitions and their measurement techniques were discussed in this chapter. Large-scale global climatic phenomena known to affect Australia's rainfall were also introduced. The large-scale global interaction between the atmosphere and ocean was described in detail. In the last section, three climatic indices that measures the ENSO phenomenon were introduced. Different measurement techniques of the ENSO phenomenon based on each of these climatic indices were discussed. Having introduced the global climatic indices and established their relationship to droughts in Australia, the next chapter reviews and examines previous studies that have developed prediction methods based on these relationships.

Chapter 3

Literature Review

Whichever drought definition is taken, it is highly dependent on rainfall. Rainfall usually varies seasonally but the ability to make further prediction over a period of months or more is hampered by the observation that future rainfall is only slightly correlated with rainfall to date. The most successful approaches are based on finding influential climatic conditions that tend to persist. Some areas are relatively highly affected, for example, the east coast of Australia.

Quantifying the occurrence and impact of drought requires establishing a set of drought criteria to define it. With several definitions being developed over the last century, it is necessary to review them and compare their applicability in this study. Following which, the literature review discusses previous research which concentrate on modelling techniques, as well as forecasting and simulation methods that take advantage of the relationship between various global climatic phenomena and rainfall.

3.1 Univariate modelling of drought

Over the past century, an extensive range of definitions and indicators have been formulated and established to model and quantify drought, and consequently to model its corresponding impact. The fact that drought affects a broad spectrum of the society have led to the categorization of drought by the American Meteorological Society in 1997 into meteorological, hydrological, agricultural or socio-economic groups as discussed in Chapter 2 [6]. To determine such droughts, indicators which rely on different physical properties, such as soil moisture and rainfall, were constructed. Given the diverse nature of these four drought categorized groups, it is difficult for a

generic definition to be developed.

One of the earliest attempt at a standard drought measure was Friedman [41], who cited four criteria that any drought index should meet, in his 1957 drought study of Texas. First, the drought index should be on an appropriate time-scale for the case study; it should be a quantitative measure of large-scale drought conditions; the index should also be able to be applied on the case study; and the availability of a long accurate historical record of this index.

Other examples of early drought criteria include: (a) 15 consecutive days with no rain; (b) 21 days or more with rainfall less than one-third of normal and (c) annual rainfall less than 75% of the normal [51]. Similar criteria were being used in other countries. However, these criteria used were either site-specific or were measured from a different hydrological viewpoint. These sentiments were also expressed in 1894 by Abbe [1] who noted that: "*From an agricultural point of view, a drought is not merely a deficiency of rainfall, but a deficiency of water available for the use of growing crops. Thus a drought affecting agriculture is a complex result of many considerations. Therefore, from both an agricultural and engineering point of view, it is impracticable to define the intensity of a drought in general and exact terms.*" Re-emphasizing this later in 1906, Henry [52] also concluded that, "*In general, climatological statistics alone fail to give a sufficient accurate conception either of the duration or intensity of drought.*" Hence, more information and considerations surrounding drought measurement were required to address these issues.

Several indicators were developed in the early twentieth century to overcome the above deficiencies. Munger [80] defined his drought index as the number of consecutive days where 24-hour rainfall were below 1.27mm, based on his assessment of Oregon's yearly forest fire risk in 1916. He further constructed a graphical technique to calculate the corresponding intensity. Another example is the seasonal Marcovitch Index [66] developed in a study over eastern United States in 1930, and is based on an equation dependent on both temperature and rainfall. These drought indices are suitable for short-term measurement of drought.

So far, these indices are only based on rainfall measurements which have failed to measure agricultural drought, which results from soil moisture deficit affecting vegetation. Difficulties in direct measurements of evapotranspiration and evaporation which are crucial in determining agricultural drought, led Thornthwaite [124] to develop the pre-

cipitation effectiveness index in 1931 which takes into account mean daily temperature and latitude as an approximation to water loss through evaporation. Following this, Thornthwaite and Mather [126] identified the need to determine drought severity in this context, based on comparison of water demand with water supply and the idea of "moisture adequacy" evolved.

This set the framework for Palmer's hydrologic two-layer model in 1965 [87], which incorporates precipitation, moisture supply and demand from 31 counties comprising the western Kansas and central Iowa. He developed two variants of his index to measure both meteorological (Palmer Drought Severity Index (PDSI)) and hydrological (Palmer Drought Hydrological Index (PDHI)) drought, by setting a different time lag criterion between the drought inducing meteorological conditions ending and the environment recovering from a drought. Julian and Fritts [59] considered this "the most satisfactory solution to the problem of combining precipitation and temperature as predictor variables."

In a thorough examination of the PDSI method in 1984, Alley [4] provided substantial amount of literature which applies PDSI for the measurement of drought spatially and temporally in the United States and acknowledges its merits in addressing the most elusive properties of drought: their intensity and starting and ending times. However, he cites some limitations surrounding the arbitrary assumptions of the water balance model which is vital to the computation of PDSI, and the complexity of separating factors such as the beginning and end of droughts, appropriate weighting of antecedent conditions and drought severity from specific impacts and their economic consequences. Furthermore, the PDSI method is more suitable for the characterization of agricultural drought since the model relies on soil moisture.

Though the PDSI has been widely used in agricultural assessments and drought policy decisions in the United States, previous researchers have examined and noted many of its limitations surrounding its application [4, 116, 69, 47, 48, 50]. In an attempt to address these problems, McKee [70] developed the SPI as a simpler alternative in 1993 using data from Colorado. He cited five important practical issues for the analysis of drought: (1) timescale, (2) probability, (3) precipitation deficit, (4) application of the definition to precipitation and (5) relationship of the definition to drought impacts. Given the variety of time scales over which the SPI can be measured upon, the frequency and duration of a drought can be evaluated.

A number of papers have compared the SPI with the PDSI using a range of statistical methods and evaluation criteria. Guttman [47] used spectral analysis in 1998, to compare historical time series of the PDSI with time series of the corresponding SPI from 1035 sites in the United States, and results demonstrate that the PDSI has a long-term moisture memory, is highly variable and is complicated to calculate, making it hard to interpret the representation of the index. The SPI, on the other hand, is easy to interpret using a probabilistic concept and is spatially consistent. Results from similar studies in Iran, China and East Africa, which evaluated the SPI with other local drought indices showed that the SPI performed better in drought identification and identification of onset, and is spatially and temporally consistent [139, 78, 84]. In a 2002 paper, Keyantash and Dracup [61] selected two test regions in the United States: Willamette Valley and North central climate, and used a weighted set of six evaluation criteria to examine the most prominent indices for the assessment of severity of different forms of drought. Evaluation scores indicate that the SPI is a highly valuable estimator of drought severity for meteorological drought.

Nonetheless, previous research on drought indices has focused on meteorological, hydrological and agricultural applications, which depend solely on measurements of physical characteristics, for example precipitation. Dracup *et.al.* [32] also recognized that most research was limited to specific basins or particular historical droughts, and consequently have ignored the study of droughts in terms of severity, magnitude and duration. However, they noted that research carried out by Yevjevich in 1967 [141], which described hydrologic droughts by their duration, areal extent, severity, probability of recurrence, and initiation and termination, was an exception. They applied this concept to annual streamflow and further developed his theory by characterizing each hydrologic drought event with its attributes: duration, severity and magnitude. Frick *et. al.* [40] in their 1990 study of the impact of prolonged droughts on water supplies for the City of Fort Collins, noted the effect of changing water demands on the severity of drought. They observed that many drought definitions are only relevant to specific water uses and area and the lack of areal drought coverage. Hence, they identified the need to describe drought in terms of duration, cumulative deficit and average annual shortage. This form of drought description facilitated the derivation of drought frequency and recurrence analysis.

3.2 Multivariate modelling of hydrological events

The previous section demonstrated that the analysis of drought on a univariate scale only provides limited information for risk, frequency analysis of extreme events and its spatial impact on the surroundings. Previous studies for example, [98] and [28], have focused on univariate cases. Complex hydrological phenomena such as droughts are generally characterized by several correlated variables and by their joint behaviour, and their inherent damage can be expressed as a function of several random variables. Hence, these early studies are unrealistic, since univariate frequency analysis will not provide a valid assessment of the probability of occurrence of an extreme event and consequently, can lead to either an over or underestimation of risk associated with the hydrological event.

Attempts towards using a multivariate approach to describe this dependence relationship are found in significant research in multivariate rainfall and flood frequency analysis. In 2000, Yue [144] modelled the joint distribution of peak rainfall intensity and depth data from two meteorological stations in Japan, using a bivariate normal distribution. Many hydrological examples take advantage of the simple construction of the joint Gaussian *cdf*, however, a distinct limitation is that all marginal distributions must be normal. Yue [144] overcame this restriction through the Box-Cox transformation of the data.

The multivariate Gaussian distribution is convenient for modelling multivariate events as it has been extensively studied and is tractable for many variables. It is tempting to put data into this framework by using Box-Cox transformations. A more general transform for data that can be thought of as being measured on a continuous scale is to fit any continuous *cdf* F then transform to uniform $U = F(X)$ and then to normality $Z = \Phi^{-1}(U)$. Essentially, this is fitting a Gaussian copula, which is discussed later.

Further bivariate research into modelling the joint distribution using non-normal distributions were also carried out. In separate case studies in 2000, Yue demonstrated the use of bivariate lognormal, Gumbel mixed and Gumbel logistic distributions for multivariate rainfall frequency analysis using data from both Niigata and Tokushima meteorological stations in Japan [142, 143]. Similarly in 2001, Yue applied bivariate gamma, bivariate extreme value distribution and bivariate lognormal distribution for multivariate flood frequency analysis using data from two river basins in Quebec, Canada [146, 145, 147].

These traditional multivariate approaches, although straightforward, have several drawbacks. First, the hydrological variables have the same type of marginal probability distribution. Second, the parameters of these traditional multivariate distributions also describe the dependence between the variables considered. Finally, the mathematical formulation becomes more complex when the number of variables is increased. In fact, many hydrological variables are correlated, do not follow the normal distribution (unless transformed) and do not have the same type of marginal distributions with other hydrological variables. Zhang and Singh [150] further stated that it is difficult to distinguish between the marginal and joint behaviour of the variables when using these methods.

The copula function is a simple, flexible and general tool which addresses these issues and is used to perform multivariate frequency analysis of hydrological data. Copulas were first introduced in 1959 by Sklar [115] and have been used extensively in the area of finance and econometrics [22, 36], but hydrological and environmental applications are more recent. In 2003, De Michele and Salvadori [27] used the Frank two-dimensional copula function to joint Pareto margins to model rainfall mean intensity and duration. They also applied the Archimedean copulas to hydrological data to investigate the relationship between univariate and bivariate return periods and thus introduced the concept of "primary" and "secondary" return periods, for the purpose of frequency analysis [106]. In another early copula contribution in hydrology, Favre *et. al.* [37] developed a method for modelling extreme values using elliptical and Archimedean copulas and described some of its benefits in 2004. The possibility of studying the marginal behaviours and global dependence structure separately makes the copula method an efficient and a favourable tool for modelling. Furthermore, most dependence structures are measured by canonical Pearson's coefficient of linear correlation. Salvadori and DeMichele [107] however, note that in some cases, this form of dependence may not exist and hence this parameter is not able to capture any non-linear dependence between variables. The copula function offers such an alternative to measuring non-linear correlation.

Recent hydrological research using the Archimedean family of copulas demonstrated an improvement on the traditional multivariate methods. Zhang and Singh [150, 151] derived both bivariate flood and rainfall frequency distributions using the copula method for flood data from both Amite River at Denham Springs and Ashuapmushuan River in Canada, and rainfall data from Amite river basin in Louisiana respectively. They

then compared the fitting of four Archimedean copulas to the bivariate normal distributions. In both case studies, the copula-based distribution had a better fit to the observed data as compared to the bivariate normal probability model. Conditional return periods based on the copula models were computed and were found to show significant differences for different conditional values. Similarly, to demonstrate the advantages of bivariate modelling of drought, Shiau [113] employed monthly precipitation data from a rainfall gauge station in Southern Taiwan to derive drought severity and duration and applied the copula approach to these variables. Bivariate drought analysis such as joint probabilities and bivariate return periods were subsequently performed in this study.

With the bivariate copula approach showing more promising modelling results in hydrology, research was extended to the trivariate case. Using the Gumbel-Hougaard copulas, Zhang and Singh derived both the trivariate flood and rainfall frequency distributions using their respective characteristics data from similar river basins in Louisiana, and employed these distributions to determine the conditional return periods for both sets of data [153, 152]. Likewise, the copula based trivariate distribution fits the empirical joint distribution better than the trivariate normal distribution for both flood and rainfall data.

Although these copulas did show a vast improvement in modelling, the assumption of having similar dependence for all pairs of variables in the copula can be restrictive. In the bivariate case, only one dependence parameter for the Archimedean copula model is required to model the dependence structure between two variables. When there are more than two variables, the use of one dependence parameter between all pairs of variables is restrictive, since the correlations between any pair of variables are identical. For many hydrological variables, this is unrealistic since they are often correlated and thus show different mutual structures of dependence and degrees of association. To circumvent this, Joe [57] and Embrechets [36] described and applied a more general version of the Archimedean copulas called fully nested or asymmetric. In general, for a trivariate case, the 3-variable asymmetric copula is composed from two bivariate one-parameter copulas. Hence, this copula possesses two dependence parameters. The practicality and flexibility of this class of copula were demonstrated in the multivariate flood frequency analysis by Serinaldi and Grimaldi in 2006 [46], and in their later research using sea-wave data the following year [109]. They further justified the fit of this copula using goodness-of-fit tests and stressed that this form of copula

should only be used when two correlations are equal and lower than the third. Other authors also took advantage of this copula class by extending it to a 4-dimensional characterization of sea-state statistics [29]. The advantage of this approach is its ability to describe how each variable affects the behaviour of others and that the construction of the multivariate distribution only depend on suitable conditional probabilities [29]. Overall, they concluded that this model provide a more general and versatile statistical tool. Some limitations highlighted by Serinaldi and Grimaldi [109] include the fact that this form of copula is not able to describe all mutually different dependence structures, a result of the construction of the Archimedean copulas.

Alternate copula families such as the elliptical family, allow different dependence structures between variables to be modelled. The two commonly used copulas in this family are the Gaussian and Student's t-copulas. The applications of the elliptical copulas in hydrology have been relatively limited [99, 117, 30]. Renard and Lang [99] justified the use of the Gaussian copula to locate an observed event on a probabilistic scale and to express the severity of a hydrological extreme event as a non-exceedance probability or in terms of return period. The main advantage of the Gaussian copula is its simplicity. However, they proposed that two aspects have to be examined before fitting the Gaussian copula. First, the marginal distributions and the dependence structure have to be 'in agreement', that is consistency between the observed data and the Gaussian copula is required. The other aspect which is of importance is the "asymptotic dependence properties of the data". This aspect is crucial especially when calculating the risk of extreme hydrological events, in this case, drought. The main drawback of the Gaussian copula, is its inability to model tail dependence. For this reason, Renard and Lang [99] caution the use of the Gaussian copula when computing low probabilities and computation outside the observed range.

Over the course of fitting the dependence structure using copulas, one should be reminded of the importance of tail dependence, since the modelling of extremal events is measured through the tail dependence. Poulin *et. al.* [90] brings to attention that the chosen copula should reflect this dependence in the extremes and highlights the importance of taking into account this quantity in multivariate frequency analysis using copulas, in order to estimate the risk adequately. To demonstrate the importance of copula choice, the joint return period and conditional density of extreme events were computed for several copula families. In their conclusion, they found that five copula families overestimated the return period of joint extreme events. The failure to take

into account tail dependence in risk estimations can lead to an under-estimation of this risk and cause major consequences in water management and planning.

To improve the dependence model, alternative copula families such as the t-copula may be more suitable when tail dependence is observed in the data. A number of recent papers in 2003 such as Mashal *et. al.* [67] and Breymann *et. al.* [18] showed that the empirical fit of the t-copula is good and was found to be superior to the Gaussian copula [30]. They cite one reason for the success of the t-copula is its ability to model dependence of extreme events. The other advantage of the t-copula is that it can be used to model a large number of correlated variables.

The t-copula has a symmetric upper and lower tail dependence, making it a favourable copula choice. This copula is not widely used in hydrology and may prove to be an improvement to current copula approaches.

To date, the application of copulas has focused on modelling the dependence structure of rainfall, flood and sea-storm characteristics. The only application to drought studies has been bivariate [113], which may not provide additional information when frequency analysis of droughts is carried out. Hence, incorporating another drought variable may give more information on risk assessment. As mentioned, in the trivariate Archimedean case, separate correlations between pairs of variables are mutually different and is more realistic to use the asymmetric class of copulas to model this correlation structure. Other suitable copulas, which have not been applied to drought studies, are the Gaussian and t-copula, which allow for the correlations of each pair of variables to be modelled separately. In the next section, the influence of global climatic phenomena on drought is examined including, both oceanic and atmospheric elements. The incorporation of these phenomena when choosing and fitting of copulas, may improve the characterization of droughts and thus provide more reliable predictions and exact risk estimates of extreme drought events.

3.3 Long-term simulation

Given the highly stochastic nature of drought, flexible models have to be constructed to effectively predict its onset and termination, particularly during cropping season. For water resource planning and management to be carried out, long term historical record of hydrological information such as rainfall, are required. In reality, these historical

records are often unavailable. Hence, the aim of stochastic modelling in hydrology is to generate simulations which are statistically similar in nature to the observed hydrological data, particularly its dependence properties. Copulas can be applied for this purpose, where return periods of extreme events are derived and the conditional probability of a drought occurring can also be computed using simulations from the fitted copula.

3.4 Short-term forecasting using global climatic phenomena relationships

This section explores and reviews various short-term prediction techniques which may be helpful in making crop planting decisions.

3.4.1 Time series and stochastic models

Stochastic models offer a useful statistical method for drought forecasting. The most popular models for this task are regression models and autoregressive integrated moving average (ARIMA) models. Many hydrologic time series are time dependent, and thus may be modelled using such models. The popularity of the ARIMA model is evident from the successful applications to time series forecasting to many areas [134]. Yurekli *et. al.* [148] lists the following advantages of the ARIMA model over other stochastic models, for example its forecasting capability, flexibility and also considers serial correlations among observations. Also, the model allows for a systematic checking at each stage, until an appropriate model is chosen [20]. However, they cite the need for large amount of data required to construct an ARIMA model and that a researcher's experience is important in choosing the appropriate ARIMA model. Some successful applications of the ARIMA models are both in streamflow forecasting [25, 148] in Turkey and rainfall forecasting [38].

Mishra and Desai [73, 74, 75] present three approaches and applied these methods to forecasting drought in the Kansabati basin in India. Their aim in the first study, was to develop stochastic models for forecasting and simulating SPI for different moving lengths of moving averages. In order to take into account seasonality in drought, they fitted the ARIMA model. They cite the small number of model parameters needed to describe the time series, which exhibit non-stationarity both within and across

seasons, as an advantage. Results showed that these linear models were satisfactory in predicting SPI for two-months ahead, and they recommend using longer SPI series to improve predictions and to ascertain future drought severity. This form of stochastic model is also shown to be useful when applied to monthly streamflow forecasting in Iran [77].

Mishra and Desai [75] have, however, considered the above method inadequate in capturing non-stationarities and nonlinearities in hydrologic data, which is a common feature. Ochoa-Rivera [85] further support this method since linear models possessed short-term memory and will not be able to predict the characteristics of drought scenarios. The alternative non-linear solution proposed are the artificial neural networks (ANNs). Following the same concept of regression but with extra predictor variables, these data-driven models are more flexible as they have a large number of parameters and reduce the impact of outlying predictor variables by applying a function such as *tanh*. Also, ANNs work well with data containing measurement errors and with extreme values. The drought predictions from the ANNs reveal a two-month predictive accuracy and increases to a three-month accuracy when dealing with SPI 12 and SPI 24 [75]. Similarly, a good fit was observed with the ANN model in drought predictions performed by Ochoa-Rivera [85].

So far, literature has shown that drought can be forecast using both stochastic and neural network models. Droughts are often seasonal. Hence in 2007, Mishra *et. al.* [76] suggested combing time series models with ANN models so as to characterize autocorrelation structures of drought more accurately. They label this a hybrid model and compared it with the individual models to forecast accuracy for different lead times. An advantage of the hybrid model is that the strength of each model is used to capture different patterns in the data. Results showed that the recursive approach of this hybrid model performs better for short lead time, but when this lead time increases, better forecast is attained using the direct approach.

The above methods is only temporal in nature and do not account for spatial variability. This makes it very difficult to predict accurately the exact onset and the extent of drought damage. The last method provides a spatial representation of drought predictions, by using the spatial interpolation of rainfall [74]. In this study, Inverse Distance Weighting (IDW) approach was performed to each grid of approximately equal area. This technique applies certain weights to each input by a normalized inverse of the distance from the control point to the interpolated point [112, 96]. In

order to analyse the intensity of the drought for different time scales of SPI, drought severity-area-frequency (SAF) curves based on a set of procedures were developed. From this visual representation, historical droughts of varying degrees were examined and temporal variation of drought was studied. The SAF curves are useful as they represent drought severity and area with respect to the drought return period. Hence, these curves can be used to find out which years have been affected by severe drought [74].

3.4.2 Drought forecasting using ENSO relationships

The world is one large environmental system, and extreme events like droughts and floods are associated with global climatic phenomena. In the following sections, a review of the relationship between droughts and these natural modulations will be analyzed, so as to further improve drought predictions. In Chapter 2, the ENSO phenomenon was introduced and the consequences resulting from both its extremes were described. ENSO has been linked to climate extreme events around the world [89], and there have been considerable number of studies which investigate the relationship between one of ENSO's widely used indicator, SOI and rainfall [68, 102, 103, 24, 154].

Early studies performed by McBride and Nicholls [68] in 1983 indicated good correlations between rainfall and the Southern Oscillation (SO) over eastern Australia, especially during winter and spring. Hence, the periodicity as characterised by the SO pressure variations will be reflected in the rainfall. Other early research in the same year by Rasmusson and Wallace [97] and Shukla and Paolino [114] have also established a relationship between ENSO and mid-latitude precipitation anomalies.

Notable research which examines this association includes Ropelewski and Halpert [100, 101, 102, 103], who associated large-scale patterns of above and below average rainfall with both high and low phases of the SOI for several regions of the globe, and describes these patterns associated with the high index phase of the SOI and compares it to those corresponding to the low index phase in terms of circulation features. Results from their early analysis demonstrated that 15 out of 19 global regions showed evidence of low SOI-precipitation relationships and conversely for the high SOI precipitation relationships, which indicates the linear relationship of rainfall with SOI phase during its extremes over most regions [100, 101, 102].

In a more recent paper in 1996, Ropelewski and Halpert [103] suggested examining

the shifts in rainfall distributions associated with the SO extremes, since these shifts may give more useful information such as conditional probability rainfall forecasts, as compared to mean rainfall values. By using the area-averaged rainfall for the 19 regions to identify the shifts in rainfall distribution, they found noticeable spatial variation in the percentiles of the SOI-precipitation relationships for some regions.

Nonetheless, considerable research investigating the relationships between SOI and Australian rainfall have also been carried using various ENSO indicators [154, 24, 26, 138, 122]. Zhang and Casey [154] used spectral analysis to analyze the cyclic nature of SO pressure and rainfall variations in Australia in 1992. Periodicity in rainfall was found to be associated with SO variations at all times of the year, but the spatial distribution varies seasonally, with the exception of eastern Australia which shows a more consistent influence. This strong persistence in the influence of the SO on Australian rainfall especially in eastern Australia implies there is some predictive value. They pointed out that having identified the phase of the pressure cycle during an El-Niño or La-Niña SO event, it will be possible to provide an assessment of rainfall conditions up to a season ahead for large areas.

Chiew *et. al.* [24] also explored the relationship between ENSO and rainfall and drought in Australia in 1998. Their lag correlation analyses showed that indicators of ENSO have the potential for forecasting spring rainfall throughout eastern Australia and summer rainfall in north-east Australia several months in advance. The effects of El-Niño have the greatest impact over inland eastern Australia and in regions such as south-west Western Australia and coastal NSW. Suppiah [122] also examined the trends and changes in the relationship between SO phenomenon and Australia's rainfall over the historical record and found weak relationships after the mid-1970s. The last century also saw decadal fluctuations and a change in relationship after the mid-1970s. All these researchers looked at temporal relationships between ENSO and rainfall, spatial analysis of this relationship was examined by Wooldridge *et. al.* [138] who carried out spatial and temporal analyses to investigate ENSO influence on rainfall and its intensities, and discovered that the strongest ENSO induced rainfall variability occurred during summer months. The strength of this relationship was found to be affected by topography, where the ENSO induced disparity on rainfall amounts and its intensities were shown to be stronger at lower elevations. This form of information is instrumental when forecasting regional drought with differing elevations.

Although researchers like Opoku-Ankomah and Cordery and Nicholls [86, 26, 82] con-

sidered this relationship between SOI and low rainfall in NSW to be encouraging, they recognized that this is insufficiently consistent to be used for forecasting the exact rainfall 3-months ahead. Hence, Cordery [26] proposed the use of a more local scale phenomena which also numerically defines ENSO, that may exert a stronger influence on rainfall since rainfall is not entirely determined by ENSO alone. One such measurement is the geopotential height (GpH). However, they also cite the short record of the GpH data in the last 50 years as a potential obstacle for the investigation of long term relationships. To overcome this, Cordery [26] partitioned the data based on a low or high SOI to calculate the correlations between rainfall and GpH. They found that rainfall is strongly related to combinations of SOI and GpH from one season before and these relationships account for more than 50% of variance in rainfall. These findings are promising for precipitation forecast and consequently drought predictions, since it will be possible to forecast low rainfall for any location and season in NSW given that the location of the GpH with the greatest effect on rainfall has been found.

Correlation studies examined have so far provided a benchmark level of understanding relationships of rainfall with SOI. Yet, further research into the formulation of certain SOI-rainfall relationship has also evolved. Stone *et. al.* in 1996 [119] developed a system of probabilistic rainfall forecasting based on identifying the lag-relationship between rainfall and SOI patterns. They identified an SOI pattern relating to rainfall which is determined through classification by using principal components analysis and cluster analysis and derived 5 distinct SOI phases. Cumulative probability distributions of subsequent rainfall associated with the SOI phases from a number of locations show that the probability of obtaining a given amount of rainfall is higher following a 'rapid rise' SOI phase as compared to a 'rapid fall' phase. This probabilistic approach provides a general overview of the expected rainfall given a particular SOI phase. However, it should also be cautioned that this method may prove to be effective only in locations that show a distinct difference between rainfall distributions corresponding to different phases.

The influence of the ENSO phenomenon on Australia's rainfall was also shown to fluctuate according to the inter-decadal oscillation in surface temperature, which contributes to the success of an ENSO-based rainfall prediction model [154]. Power *et. al.* [91] investigated the association between this low frequency variability in Pacific sea surface temperatures, known as the IPO, and inter-decadal variability in ENSO teleconnections with Australia. They examined the interdecadal variability from the performance of

an ENSO-based rainfall prediction model and found that the relationship between annual Australian climate variations and ENSO was not strong when IPO increased the SSTs, and resulted in poor predictions. In contrast, this predictive scheme performs well and association between ENSO and Australian climate is strong, when the IPO reduces SSTs. Such an association with fluctuations in rainfall and sea surface temperatures may provide some potential with the inclusion of the IPO on rainfall and ENSO prediction process.

Kiem and Franks [62] also investigated the relationship between the IPO index and multi-decadal variability of drought risk in NSW, by examining the risk of falling below the critical level for a water storage reservoir in 2004. The 90% confidence limits show that the risk of falling below the critical level is 20 times higher in the IPO positive phase compared to in the IPO negative phase, indicating a higher probability of drought when IPO is positive. The frequency at which La-Niña events occurred was also found to be signifiantly higher during the IPO negative phase and these La-Niña events are usually wetter than normal La-Niña events. On the other hand, the number of neutral events that occurred was significantly higher when IPO is positive, indicating a higher rate of ENSO extremes when IPO is negative. These findings are important in showing the association of climate variability in eastern Australia and these multi-decadal climate processes, in particular the IPO phase, which modulates ENSO. Knowledge of the role that IPO plays in modulating ENSO will offer a better perspective on drought risk assessment and can be used as a potential drought indicator that can be used together with ENSO predictions.

3.4.3 Drought forecasting using sea-surface temperature relationships

Variations in monthly and seasonal rainfall in Australia are often associated to ENSO [3, 83, 103]. The previous description of ENSO in Chapter 2 related this climatic phenomena with the see-saw fluctuation in sea-level pressures which is observed in the SO and the corresponding fluctuation in SSTs in the equatorial Pacific. Hence, research into the relationship between SSTs and rainfall, and consequently drought in Australia is logical.

Early studies carried out by Priestley [92] in 1964 and Priestley and Troup [93] in 1966, suggested a relationship between SST and rainfall anomalies. Similarly, Streten [120, 121] observed that years where extensive drought occurred coincided with low

SST over the eastern Indian Ocean and south west Pacific. At the same time, warmer than normal SST occurred in eastern Pacific. This suggest that these extreme weather events may be related to SST anomalies.

In order to investigate if more than one mode of variability affected winter rainfall in Australia, Nicholls [83] used rotated principal component analysis in 1989. Results revealed two large-scale pattern of variations which were correlated to the Indian and Pacific Ocean SSTs and they account for more than half of the variance. Around the same time, Palmer and Branković studied the 1987 and 1988 forecasts of atmospheric flow in the United States, using a complex numerical weather prediction model. The predictive skill of this model varied with regions in the US and they concluded that the SSTs in 1987 and 1988 were instrumental in accounting for the rainfall reduction in the US. These research indicate the ability of SST in particular regions, as a potential predictor of Australian rainfall.

The winter rainfall and SST relationship findings by Nicholls [83] was explored further using the S and T-mode rotated principal component analysis by Drosdowsky in 1993 [33]. They found that predictions of early winter rainfall in parts of southern and eastern Australia is feasible based on the SST from the west Australian coast from summer to early autumn periods. These lagged correlations were strongly seasonal and spatially dependent and were not observed when conventional seasons were used. Hence, these research has shown that spatial orientation affects the SST-rainfall relationships for different seasonal groups.

In a series of papers published in 2000, Sharma [110, 111] presents a new approach which first identifies optimal predictors for a probabilistic forecast model and subsequently using a range of ocean-atmospheric predictor variables in an application to predict the quarterly rainfall in Warragamba Dam in NSW. The first paper introduces the nonparametric mutual information criterion as an effective way of quantifying dependence between two random variables. Based on the above criterion, a partial mutual information (PMI) criterion was then established to quantify the dependence between two variables conditional on the presence of existing predictors, and it proved to be effective and accurate in identifying predictors of all models. The subsequent paper then applies the PMI criterion to identify the predictors of quarterly rainfall for a specific location. They compared two prediction models, one using three commonly used ENSO indices as predictors and the other using only SST anomalies. To evaluate their corresponding predictive skill, seasonal rainfall were forecasted using the general

additive model. Results indicate that the use of SST give more superior predictions amongst the models compared. This can be attributed to the limitation of the ENSO indices, since these indices only account for one of the many factors contributing to low rainfall.

3.5 Summary of Chapter

Drought modelling and prediction techniques have been reviewed in this chapter. Univariate modelling studies showed that these methods were only capable of identifying the occurrence of particular types of drought and focus on duration for a chosen area based on data averaged over the area, but were limited in describing the spatial pattern, drought severity and intensity. A proposed solution to providing additional drought information is the multivariate approach, where copulas can model this distribution and the corresponding dependence structure more adequately than multivariate Gaussian techniques. In addition, these studies have demonstrated that SST and ENSO are promising predictors for droughts in Australia, particularly the eastern coast.

Chapter 4

Regional drought modelling using copulas

Complex hydrological phenomena such as droughts are generally characterized by several correlated variables. Realistic modelling of these characteristics is essential for planning and management of water resources. Copulas provide a general form for multivariate distributions, allowing arbitrary marginal distributions and a wide variety of correlation structures. This chapter introduces the general concept of copulas to model the dependence structure of drought characteristics on a regional scale in Australia, which is fundamental for short term predictions and simulation studies.

4.1 Drought Characteristics

Three significant characteristics of droughts identified are: Duration; Peak Intensity; and Severity. Following Yevjevich [141], duration is defined as the number of days or months between SPI (S_t) which was defined in Chapter 2, at time t falling below -1 and subsequently increasing above -1. The intensity of the drought is defined as the absolute value of the SPI, and the Peak Intensity of a drought is the greatest value of intensity over its duration. Severity is measured from the sum of its intensities over the entire duration of the drought:

$$\int_{t_i}^{t_e} \mid S_t \mid dt$$

where t_i indicates the time when the drought starts and t_e is the time when the drought terminates. Furthermore, a useful derived characteristic is Average Intensity which can be calculated by dividing the severity of a drought over its duration $(t_e - t_i)$. With these definitions, Peak Intensity and Average Intensity must exceed 1 and are dimensionless quantities since they are measured in terms of SPI. Figure 4.1 illustrates three droughts, of which, drought 1 has the largest Peak Intensity (I), drought 2 has the longest duration (D) and drought 3 is the most severe.

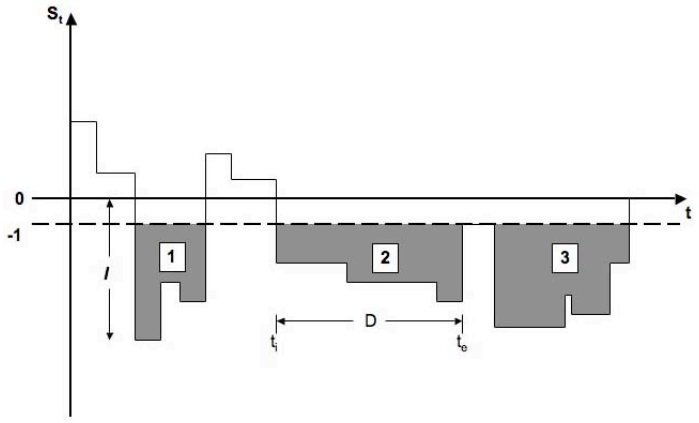


Figure 4.1: Definition of drought characteristics.

4.2 Copulas

4.2.1 Definition

A n-copula is the *cdf* of a multivariate distribution function with all the univariate marginal distributions being uniform on the interval $[0, 1]$, written $U(0, 1)$. It follows that a copula C is a mapping $C : [0, 1]^n \to [0, 1]$. Copulas encompass all multivariate distributions since Sklar's Theorem [115, 81], states that, for an n-dimensional *cdf* H with univariate margins $F_1, \ldots, F_n$, there exists an n-copula C such that

$$\begin{aligned} H(x_1, x_2, \ldots, x_n) &= \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \\ &= C(u_1, \ldots, u_n) \end{aligned} \tag{4.1}$$

where $F_k(x_k) = u_k$ for $k = 1, \ldots, n$ with $u_k \sim U(0, 1)$. Conversely, any choice of copula and F_i is an n-variable *cdf*. The definition of a 2-dimensional copula has a

straightforward visual interpretation. First, a copula is a function C from $[0,1]^2 \rightarrow [0,1]$ and satisfies these conditions: $C(u,0) = 0 = C(0,v)$ and $C(u,1) = u, C(v,1) = v$. In addition, for every u_1, u_2, v_1, v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, the following equation will be non-negative:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Since $C(u,v)$ is in the volume set of $([0,u] \times [0,v])$, we can take $C(u,v)$ to be an assignment of a number in $[0,1]$ to the rectangle $[0,u] \times [0,v]$. Hence, the above equation gives a formula for the number assigned by C to each rectangle $[u_1, u_2] \times [v_1, v_2]$ in $[0,1]^2$ and the number assigned must be non-negative. For a bivariate case, the lower bound of the copulas is $W(u,v) = \max(\mathbf{0}, u + v - 1)$, while the upper bound is $M(u,v) = \min(u,v)$. Furthermore, the continuous random variables X_1, X_2 with copula C are independent if the product copula $C = \Pi(u,v) = uv$.

For a more general case, $C : [0,1]^n \rightarrow [0,1]$ is an n-dimensional copula if it satisfies:

1. $C(u_1, \ldots, u_n) = 0$ when $\mathbf{u} \in [0,1]^n$ has at least one component equating to 0.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, n\}$, $u_i \in [0,1]$.
3. For all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in [0,1]^n$ with $a_i \leq b_i$, then

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \cdots + i_n} C(u_{1i_1}, \ldots, u_{ni_n}) \geq 0$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, n\}$

Correspondingly, the lower bound for the n-dimensional copula is $W(u_1, \ldots, u_n) = \max(\mathbf{0}, u_1 + \cdots + u_n - n + 1)$ and the upper bound is $M(u_1, \ldots, u_n) = \min(u_1, \cdots, u_n)$.

The copula approach allows for marginal distributions and the joint behaviour of variables to be modelled separately. This is done by transforming the univariate variables (x_n) into uniform variables (u_n) by applying a suitable marginal distribution (F_n). The dependence structure between the variables is described by the copula function C. The copula function C express a wide variety of dependence structures and there are many possible families of copula. Another favourable feature of the copula, is that they are invariant for strictly monotone transformations of the random variables. There is a very large number of possible copula families: Two of the more commonly

used, the Elliptical and Archimedean copulas are introduced in the next section.

4.2.2 Elliptical copulas

Embrechets [36] defines the elliptical distribution from which this form of copula is derived. Suppose $\mathbf{X}$ is a n-dimensional random vector with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and some $n \times n$ nonnegative definite, symmetric variance-covariance matrix Σ, there exist a characteristic function $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t})$ of $\mathbf{X} - \boldsymbol{\mu}$ which is a function of the quadratic form $\mathbf{t}^T\Sigma\mathbf{t}$:

$$\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t}) \quad = \quad \phi(\mathbf{t}^T\Sigma\mathbf{t})$$

where ϕ is the unique copula generator. Then, $\mathbf{X}$ has an elliptical distribution with parameters μ, Σ and ϕ, and is denoted as $\mathbf{X}\sim E_n(\boldsymbol{\mu}, \Sigma, \phi)$.

For example, when the generator of $\mathbf{X}$ is $\phi(u) = \exp(-u/2)$, then $\mathbf{X} \sim E_n(\mathbf{0}, \mathbf{I}_n, \phi)$, which is equivalent to the multivariate Normal distribution $N_n(\mathbf{0}, \mathbf{I}_n)$ with all the components $X_i \sim N(0, 1), i = 1, \ldots, n$ being independent. More detailed proofs and theorems relating to the elliptical distribution can be found in Cambanis, Huang and Simons [19]. Since the elliptical distribution is fully described by $\boldsymbol{\mu}, \Sigma$ and ϕ, it also follows that the copula of a elliptically distributed random vector is uniquely determined by the linear correlation matrix of $\mathbf{X}$, denoted by R, where $R_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ and ϕ.

4.2.2.1 Gaussian copulas

A frequently used member of the elliptical copula family is the Gaussian copula where the multivariate Gaussian copula with n variables is expressed as:

$$C_P(u_1, \ldots, u_n) \quad = \quad \mathbf{\Phi^n}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)) \tag{4.2}$$

where $\mathbf{\Phi^n}$ indicates the standard multivariate Normal *cdf* with correlation matrix P, and Φ^{-1} denotes the inverse *cdf* of the univariate standard Normal distribution. For a bivariate case, the Gaussian copula is expressed as

$$C_\rho(u_1, u_2) \quad = \quad \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\} dsdt \tag{4.3}$$

where ρ is the correlation matrix.

Simulating from the Gaussian copula is straightforward using the following algorithm [36]. Since only trivariate cases are considered in this thesis, simulations of trivariate random variables are given:

1. Find A such that $A = A^T$ and $AA^T = P$. This can be performed using the Cholesky decomposition A of P.
2. Simulate n independent random variates z_1, z_2, z_3 from $N(0, 1)$.
3. Set $\mathbf{x} = A\mathbf{z}$.
4. Set $u_i = \Phi(x_i), i = 1, 2, 3$.
5. $(u_1, u_2, u_3)^T \sim C_P$.

4.2.2.2 Student's t-copulas

Another important member in this family is the t-copula. A multivariate t distribution with ν degrees of freedom, mean vector $\boldsymbol{\mu}$ and a positive-definite covariance matrix Σ is defined by the *pdf*

$$f(\mathbf{x}) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d \mid \Sigma \mid}} \left(1 + \frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}} \tag{4.4}$$

where $\mathbf{x}$ is a d-dimensional random vector. The above t-distribution of $\mathbf{x}$ can then be denoted by $\mathbf{x} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$. Notice that Σ is not the variance-covariance matrix of $\mathbf{x}$. For example, if $d = 1$, $\boldsymbol{\mu} = 0$ and $\Sigma = 1$, the standard univariate t-distribution is observed which has mean 0 and variance $\nu/(\nu - 2)$. In general, if $\mathbf{Y} \sim N(\mathbf{0}, \Sigma)$ and $W \sim \chi^2_\nu$ and is independent of $\mathbf{Y}$ and

$$\frac{\mathbf{Y}}{\sqrt{W/\nu}} = \mathbf{x} - \boldsymbol{\mu}$$

then $\mathbf{x}$ has the multivariate t-distribution $t_d(\nu, \boldsymbol{\mu}, \Sigma)$. Since the copula remains invariant under any series of strictly increasing transformations of the components of $\mathbf{x}$ [30], it follows that the copula of $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ is identical to $t_d(\nu, \mathbf{0}, P)$, where P is the correlation matrix corresponding to the variance-covariance matrix Σ. It follows from the definition of a copula (Equation 4.1) that the t-copula of $\mathbf{x}$ with ν degrees of freedom and correlation matrix P can be written as

$$C^t_{\nu,P}(\mathbf{u}) = C^t_{\nu,P}(u_1, \ldots, u_d) = t^n_{\nu,P}(t^{-1}_\nu(u_1), \ldots, t^{-1}_\nu(u_d))$$

Then the t-copula expression is given by

$$C^t_{\nu,P}(\mathbf{u}) = \int_{-\infty}^{t_v^{-1}(u_1)} \cdots \int_{-\infty}^{t_v^{-1}(u_d)} \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d \mid P \mid}} \left(1 + \frac{\mathbf{x}' P^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} \mathrm{d}\mathbf{x} \quad (4.5)$$

where t_ν^{-1} is the inverse *cdf* (quantile function) of a univariate t_ν distribution. Embrechets [36] gives an algorithm for the generation of trivariate random variables from a t-copula, $C^t_{\nu,P}$ is as follows:

1. Find the Cholesky decomposition of A of P.
2. Simulate three independent random variates z_1, z_2, z_3 from $N(0,1)$.
3. Simulate a random variate w from χ^2_ν independent of z_1, z_2, z_3.
4. Set $\mathbf{y} = A\mathbf{z}$.
5. Set $\mathbf{x} = \frac{\sqrt{\nu}}{\sqrt{w}}\mathbf{y}$.
6. Set $u_i = t_v(x_i)$, where $i = 1, 2, 3$.
7. $(u_1, u_2, u_3)^T \sim C^t_{\nu,P}$.

4.2.3 Archimedean copulas

Unlike the elliptical copula family which do not possess closed form expressions and are unable to model complex dependence structures, the Archimedean copula family has closed form expressions and is able to model a variety of dependence structures displayed by hydrological variables [36]. However, one distinct disadvantage of using multivariate extensions of this copula family is the absence of a free parameter, as a result of the construction of this copula, which forces at least two correlation pairs to be equal. Definitions and fundamental properties of this family are introduced in this section.

An essential component in the construction of the Archimedean copula is the copula generator φ, which is unique for different copulas. Here, a general definition of this function and its corresponding inverse is provided.

Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$, such that $\varphi(1) = 0$. Then the pseudo-inverse of φ is the function $\varphi^{[-1]}$ of t that exists from $[0,\infty]$ to $[0,1]$, where the $\varphi^{[-1]} = \varphi^{-1}$ for $0 \leq t \leq \varphi(0)$ and 0 otherwise.

Furthermore, $\varphi^{[-1]}$ is continuous and non-increasing on $[0,\infty]$ and strictly decreasing on $[0,\varphi(0)]$. In addition, $\varphi^{[-1]}(\varphi(u)) = u$ on $[0,1]$ and likewise $\varphi(\varphi^{[-1]}(t)) = t$ for $0 \leq t \leq \varphi(0)$, and $\varphi(0)$ otherwise. Hence, if $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$.

Given the above definition of φ as being continuous with a strictly decreasing function from $[0,1]$ to $[0,\infty]$, and with $\varphi^{[-1]}$ as the pseudo-inverse of φ as defined above. There exists a function C from the unit square $[0,1]\times[0,1]$ to $[0,1]$ which can be expressed as:

$$C(u_1,u_2) \;=\; \varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)) \tag{4.6}$$

Hence, C satisfies the boundary conditions: $C(u_1,0)=0=C(0,u_2)$ and $C(u_1,1)=u_1, C(1,u_2)=u_2$, for a copula. It can be shown that it also satisfies the volume condition and is a copula [81].

The simplest Archimedean copula is the Archimedean symmetric (one-parameter) copula, which is of the form:

$$C(u_1,\ldots,u_n) \;=\; \varphi^{-1}(\varphi(u_1)+\cdots+\varphi(u_n)) \tag{4.7}$$

where φ is the unique generator of the copula and $u_1,\ldots,u_n$ are uniform random variables. Nelsen [81] and Joe [57] provide an extensive list of Archimedean copulas. Table 4.1 shows some common examples of different generators which lead to several important copulas.

Table 4.1: Bivariate Archimedean copulas and their corresponding generators

Copula	Generator $\varphi(t)$	Parameter θ	Bivariate Copula $C(u_1,u_2)$
Clayton	$t^{-\theta}-1$	$\theta>0$	$(u_1^{-\theta}+u_2^{-\theta}-1)^{-1/\theta}$
Gumbel-Hougaard	$(-\ln t)^{\theta}$	$\theta\geq 1$	$\exp\{-[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{1/\theta}\}$
Frank	$\ln\left(\frac{e^{\theta t}-1}{e^{\theta}-1}\right)$	$\theta\neq 0$	$\frac{1}{\theta}\ln\left(1+\frac{(e^{\theta u_1}-1)(e^{\theta u_2}-1)}{e^{\theta}-1}\right)$

For a trivariate example, the Gumbel-Hougaard copula is one relatively well-known example:

$$C(u_1,u_2,u_3) \;=\; \exp\{-[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}+(-\ln u_3)^{\theta}]^{1/\theta}\} \tag{4.8}$$

The Archimedean copula with its generator φ has a few favourable properties, such as C is symmetric. That is $C(u,v)=C(v,u)$ for all u,v in $[0,1]$. The other property is

that C is associative: $C(C(u,v),w) = C(u,C(v,w))$ for all u, v, w in $[0,1]$. Embrechets [36] gives the proof of this property.

Embrechets [36] gives a general algorithm to simulate from an Archimedean copula. Let $U_1, \ldots, U_n$ have copula C, then the conditional *cdf* of U_k given the values of $U_1, \ldots, U_{k-1}$ can be shown to be:

$$\begin{aligned} C_k(u_k|u_1,\ldots,u_{k-1}) &= \Pr\{U_k \leq u_k | U_1 = u_1, \ldots, U_{k-1} = u_{k-1}\} \\ &= \frac{\partial^{k-1} C_k(u_1,\ldots,u_k)}{\partial u_1 \cdots \partial u_{k-1}} \Big/ \frac{\partial^{k-1} C_{k-1}(u_1,\ldots,u_{k-1})}{\partial u_1 \cdots \partial u_{k-1}} \end{aligned} \tag{4.9}$$

For a trivariate case, the algorithm is:

1. Simulate a random variate u_1 from $U(0,1)$.
2. Simulate a random variate u_2 from $C_2(\cdot \mid u_1)$.
3. Simulate a random variate u_3 from $C_3(\cdot \mid u_1, u_2)$.

The conditional copulas $C_2(u_2 \mid u_1)$ and $C_3(u_3 \mid u_1, u_2)$ require the partial differentiation of Equation (4.9) with respect to u_1, and both u_1 and u_2 and it is advisable to use a computer algebra package such as **R** [95] since computation of these expressions can be cumbersome. Appendix A gives the derivations of these expressions.

4.2.4 Asymmetric Archimedean copulas

The above symmetric Archimedean copula suffices to describe the dependence structure between two variables, but when there are more than two variables, this form is restrictive, since the correlations between any pair of variables are identical. For most hydrological variables, such an assumption is unrealistic. To circumvent this, an asymmetric Archimedean copula can be constructed by nesting symmetric copulas and is expressed as [57, 36]:

$$\begin{aligned} C(u_1,\ldots,u_n) &= C_1(u_n, C_2(u_{n-1},\ldots,C_{n-1}(u_2,u_1)\ldots)) \\ &= \varphi_1^{-1}(\varphi_1(u_n) + \varphi_1(\varphi_2^{-1}(\varphi_2(u_{n-1} + \ldots \\ &+ \varphi_{n-1}^{-1}(\varphi_{n-1}(u_2) + \varphi_{n-1}(u_1))\ldots))) \end{aligned} \tag{4.10}$$

where $C_{n-1}(u_2, u_1) = \varphi_{n-1}^{-1}(\varphi_{n-1}(u_2) + \varphi_{n-1}(u_1))$. As an illustration, the 3-variable case is given by:

$$\begin{aligned} C(u_1, u_2, u_3) &= C_1(C_2(u_1, u_2), u_3) \\ &= \varphi_1^{-1}(\varphi_1(\varphi_2^{-1}(\varphi_2(u_1) + \varphi_2(u_2)) + \varphi_1(u_3))) \end{aligned} \tag{4.11}$$

Based on the symmetric and associative properties of the Archimedean copulas in Theorem 4.3, the multivariate copulas in Equations (4.10) and (4.11) are also copulas. The 3-variable asymmetric copula in Equation (4.11) is composed from two bivariate one-parameter copulas C_1 and C_2, where C_2 is the copula describing the dependence between variables u_1 and u_2 and the outer copula C_1 is a function of the inner copula and u_3. With this construction, the bivariate marginal copulas $C_1(u_1, u_3)$ and $C_2(u_2, u_3)$ are identical and it follows that the correlations between the inner variables and the outer variable are identical. The construction can be extended for n variables, where $n(n-1)/2$ bivariate marginals are required and there are $n-1$ distinct generators φ_i. Equation (4.7) is a special case of Equation (4.10) when all the generators are equal [36, 46].

For the trivariate case, the asymmetric Gumbel-Hougaard copula is uniquely defined by its generator $\varphi_i(t) = (-\ln t)^{\theta_i}$ with $\theta_i \geq 1$:

$$C_1(C_2(u_1, u_2), u_3) = \exp[-([(-\ln u_1)^{\theta_2} + (-\ln u_2)^{\theta_2}]^{\frac{\theta_1}{\theta_2}} + (-\ln u_3)^{\theta_1})^{\frac{1}{\theta_1}}] \tag{4.12}$$

where θ_1 is the measure of dependence for the pairs (u_1, u_3) and (u_2, u_3) and θ_2 represents dependence of (u_1, u_2). For a proper 3-dimensional copula to exist, $\theta_1 \leq \theta_2$, that is variables that are more nested within a copula have a higher dependence between them.

4.2.5 Tail dependence

Linear correlation is common measure of linear dependence between variables from an elliptical distribution, for example the multivariate gaussian distribution. Although it is straightforward to calculate, this form of dependence is non-invariant under strictly increasing non-linear transformations and hence will not be able to capture non-linear dependence. The dependence structure displayed by heavy tailed distributions like the Archimedean copula cannot be measured adequately by linear dependence.

Tail dependence is introduced in [36, 57, 81] to explain the amount of dependence represented by the extreme values in the upper right quadrant tail or lower left quadrant tail of a bivariate distribution, and is of practical importance in hydrology.

Suppose X and Y are continuous random variables with *cdf* F and G respectively, then the upper tail dependence, λ_U is the limit of the conditional probability that Y is more than the $100u$-th percentile of G given that X is more than the $100u$-th percentile of F as u approaches 1.

$$\lambda_U = \lim_{u \to 1} \Pr\{Y > G^{-1}(u) \mid X > F^{-1}(u)\} \tag{4.13}$$

where $\lambda_U \in [0, 1]$. A bivariate copula C has upper tail dependence when $\lambda_U \in (0, 1]$ and no tail dependence when $\lambda_U = 0$. It follows that $\Pr\{Y > G^{-1}(u) \mid X > F^{-1}(u)\}$ is equivalent to

$$\begin{aligned} \lambda_U &= \frac{1 - \Pr\{X \leq F^{-1}(u)\} - \Pr\{Y \leq G^{-1}(u)\} + \Pr\{X \leq F^{-1}(u), Y \leq G^{-1}(u)\}}{1 - \Pr\{X \leq F^{-1}(u)\}} \\ &= \frac{1 - 2u + C(u, u)}{1 - u} \\ &= \frac{\bar{C}(u, u)}{1 - u} \end{aligned}$$

where $\bar{C}$ is the survival copula of the copula C [57]. This result will be similar if $\Pr\{X > F^{-1}(u) \mid Y > G^{-1}(u)\}$ is calculated in Equation (4.13). Similarly, the lower tail dependence λ_L is defined as the limit of the conditional probability that Y is less than or equal to the $100u$-th percentile of G, given that X is less than or equal to the $100u$-th percentile of F as u tends to 0 and is expressed as

$$\begin{aligned} \lambda_L &= \lim_{u \to 0} \Pr\{Y \leq G^{-1}(u) \mid X \leq F^{-1}(u)\} \\ &= \lim_{u \to 0} \frac{C(u, u)}{1 - u} \end{aligned} \tag{4.14}$$

Elliptical copulas such as the Gaussian copulas are symmetrical distributions and it can be shown that there is no upper and lower tail dependence [36]. The t-copula, however, has both upper and lower tail dependence which are equal. An example of the Archimedean copulas with upper tail dependence that can be derived, is the Gumbel copula with upper tail dependence $2 - 2^{1/\theta}$.

The calculation of tail dependence can be extended for a multivariate case, where there

are more than 2 variables. Jajuga [55] defines the upper and lower tail dependence for a trivariate copula example as follows:

$$\begin{aligned} \lambda_U &= \lim_{u\to 1} \Pr\{Y > G^{-1}(u), Z > H^{-1}(u) \mid X > F^{-1}(u)\} \\ &= \lim_{u\to 1} \frac{\bar{C}(u,u,u)}{1-u} \end{aligned} \tag{4.15}$$

$$\begin{aligned} \lambda_L &= \lim_{u\to 0} \Pr\{Y \leq G^{-1}(u), Z \leq H^{-1}(u) \mid X \leq F^{-1}(u)\} \\ &= \lim_{u\to 0} \frac{C(u,u,u)}{1-u} \end{aligned} \tag{4.16}$$

4.3 Application to rainfall in NSW

Trivariate Gaussian and Gumbel-Hougaard copulas are fitted separately to drought characteristics data from a district in NSW, and are described in this section. Tail dependence for both copula families are investigated and the goodness-of-fit of the data to the different copulas is assessed.

4.3.1 Preliminary analysis of data

The Australian BoM divides the continent into 99 numbered rainfall districts, where sites of relatively similar rainfall climates are grouped. For this application, monthly rainfall from District 63 (Central Tablelands) in NSW from 1913 to 2002 is analyzed to derive drought characteristics and occurrences. There were 78 droughts over the 90 year period. Figure 4.2(a) shows the location of District 63 in NSW and Figure 4.2(b) shows the distribution of the time between the end of one drought and the start of another in terms of months.

The monthly SPI(3) is calculated based on the description in Chapter 2.1.1 and the time series is displayed in Figure 4.3. Drought characteristics are then derived according to Section 4.1. Tables 4.2 and 4.3 display the relevant statistics and correlations of the drought characteristics. On average, droughts in District 63 last for about 2.5 months over the investigated period, with an average peak intensity of 1.6. Correlations between Severity and Duration are found to be the strongest amongst the variable pairs. Based on Table 4.4, correlations between these three drought characteristics and inter-drought time, before or after the drought, are small, and none are

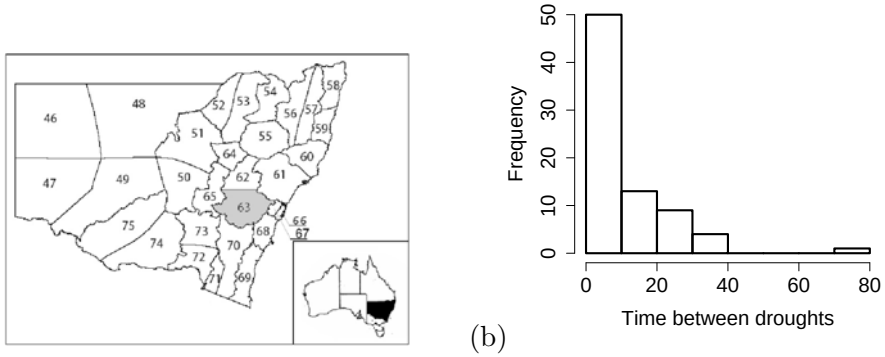


Figure 4.2: (a) New South Wales (NSW) rainfall districts (b) Histogram of months between the end of one drought and start of the next.

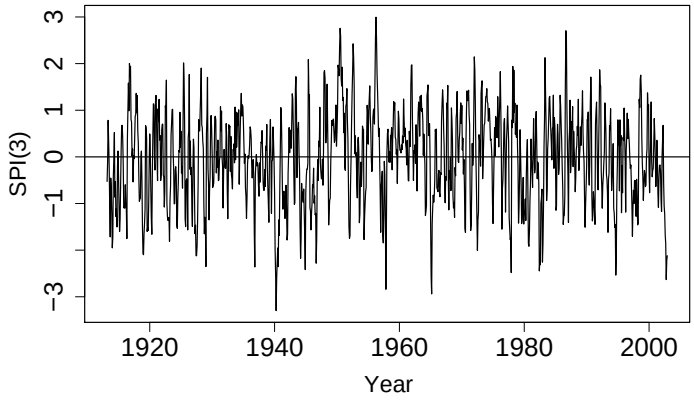


Figure 4.3: Time series plot of SPI(3) of District 63

statistically significant at the 5% level according to the p-values. Figures 4.4(a), (b) and (c) shows the correlation plots of inter-drought time with Duration, Intensity and Severity respectively. Consequently, time between droughts is realistically modelled independently of the other three variables.

Table 4.2: Means and Standard deviations of drought characteristics in District 63.

Characteristic	Severity	Duration (months)	Peak Intensity
Mean	3.72	2.42	1.63
Standard deviation	3.20	1.62	0.49

Table 4.3: Correlations of drought characteristics.

Variable pairs	Correlations
Severity, Duration (months)	0.97
Severity, Peak Intensity	0.83
Duration (months), Peak Intensity	0.76

Table 4.4: Correlations of inter-drought time and drought characteristics.

Variable pairs	Correlations	p-value
Inter-drought time, Duration (months)	-0.11	0.33
Inter-drought time, Peak Intensity	-0.19	0.10
Inter-drought time, Severity	-0.13	0.26

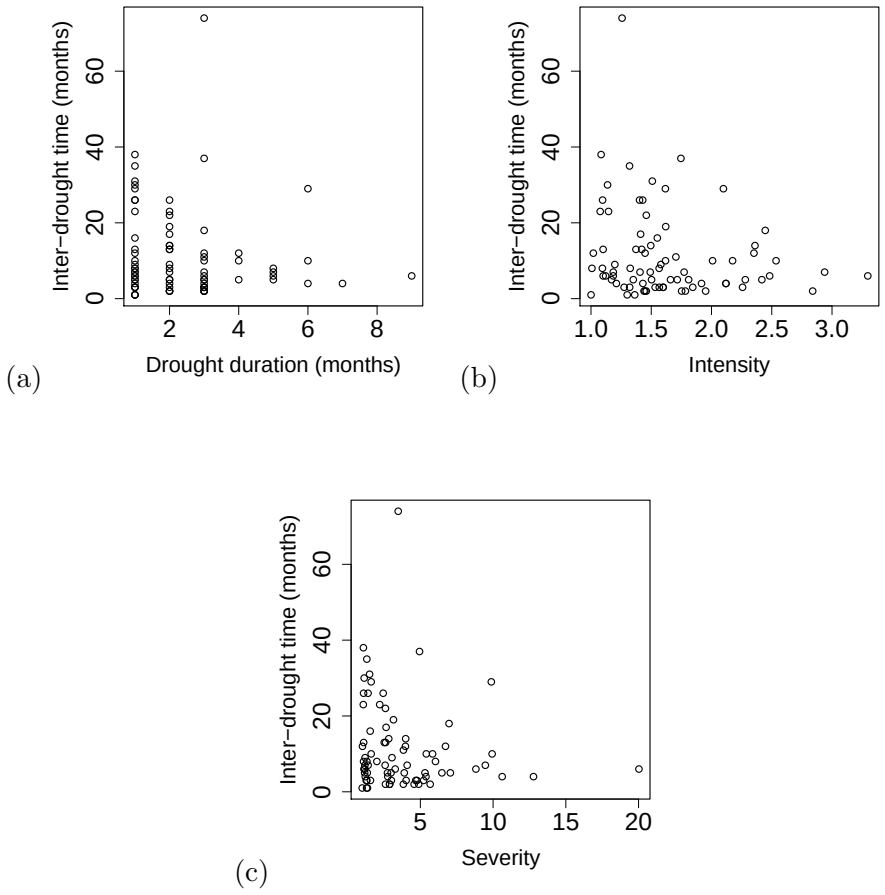


Figure 4.4: Correlation plots of inter-drought time with (a) Duration (b) Intensity and (c) Severity.

4.3.2 Fitting marginal distributions to drought characteristics

With the copula approach, it is possible to fit separate individual marginal distributions to the three drought characteristics. The first step is to identify appropriate marginals for each of the three variables; severity, duration and peak intensity. The Akaike Information Criterion (AIC) is used to assess the goodness of fit of a particular distribution to the data set. The AIC is defined as $\text{AIC} = 2k - 2\ln(L)$, where k represents the number of parameters in the statistical model and L is the maximized value of the likelihood function for the estimated model, and thus modifies the maximum likelihood criterion by penalising for the number of parameters. This forms the first of the two-step Inference Function of Margins (IFM) approach, where the copula parameter θ is estimated. A number of marginal distributions such as Lognormal, Weibull, Normal and Gamma distributions were fitted. The three-parameter Weibull distribution with the following *cdf* is found to be the best fit to all the drought characteristics:

$$F(x) \quad = \quad 1 - \exp[-(x-L)^k/\theta] \qquad\qquad L \leq x,$$

where L, θ and k are the threshold, scale and shape parameters respectively. Table 4.5 shows the maximum likelihood estimates of the parameters and the corresponding AIC values of the fitted Weibull distribution. Severity, duration and peak intensity are defined as u_1, u_2 and u_3 following their transformations to uniform variables. Table 4.6 displays the correlation between the drought variables. Figure 4.5(a)-(c) shows the probability plots of severity, peak intensity and duration fitted by the three parameter Weibull marginals. In Figure 4.5(c), points are superimposed as the data are restricted to integers. Overall, the Weibull distribution provides a reasonably good fit to the drought variables.

Table 4.5: Estimated parameters of Weibull distributions.

Characteristic	Threshold	Scale	Shape	AIC
Severity	0.9904	2.463	0.8244	311.5
Duration (months)	0.99	0.9375	0.5291	161.8
Peak Intensity	0.99	0.688	1.311	81.13

Table 4.6: Correlations of drought variables after transformation to uniform marginals.

Correlations	Pearson	Kendall's tau
u_1, u_2	0.958	0.860
u_1, u_3	0.859	0.752
u_2, u_3	0.710	0.615

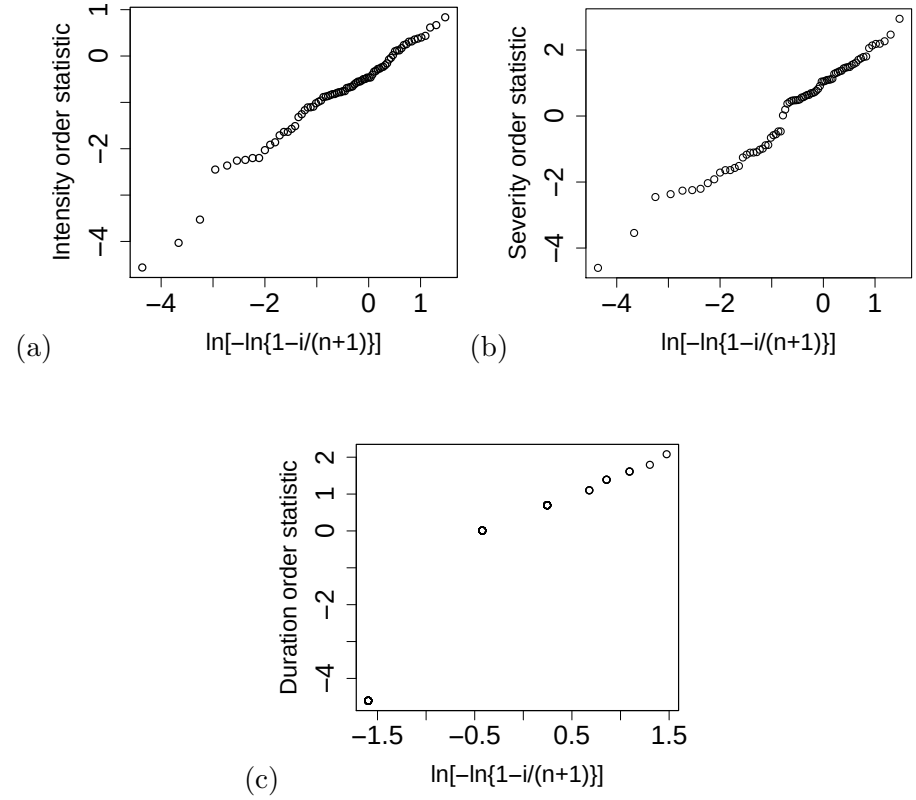
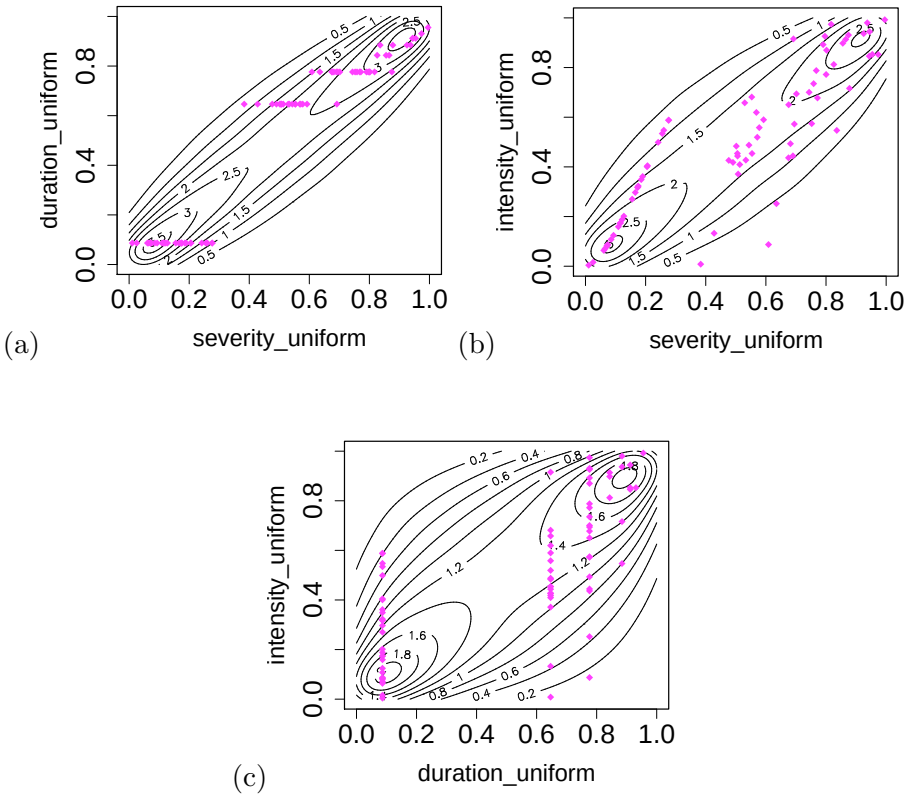


Figure 4.5: Weibull probability plots of (a) Intensity (b) Severity and (c) Duration.

4.3.3 Fitting trivariate Gaussian copulas

The Gaussian copula is fitted by taking the standard Gaussian inverse *cdf* of each element of the (u_1, u_2, u_3) triples and then calculating the three sample Pearson correlations: $r_{uv} = 0.934$, $r_{uw} = 0.867$ and $r_{vw} = 0.695$. Simulations of 100,000 deviates are made using the multinomial Gaussian distribution function in **R** [95, 140]. Figure 4.6 is a contour plot, fitted to the simulated data, representing the bivariate copula probability density function, with the historic data shown as points. Figure 4.7 is a plot of the simulated data from the copula back-transformed to severity against duration. The 90% percentiles emphasize the association of the variables in the upper tails.

To measure upper tail dependence, the Pearson correlation of points falling above the 90th percentiles is calculated (Table 4.7).



(a) (b) (c)

Figure 4.6: Empirical contour plot of (a) u_1 against u_2, (b) u_1 against u_3 and (c) u_2 against u_3 from the Gaussian copula.

The correlations of the simulated data from the Gaussian copula (Table 4.7) are close to the Pearson correlations of the historic data shown in Table 4.6. Note that the Gaussian copula is capable of modelling separate correlations for each variable pair. However, the upper tail correlations of the simulated data are fairly weak, especially between the pairs of severity-intensity and duration-intensity.

4.3.4 Fitting trivariate Gumbel-Hougaard copulas

Archimedean copulas are able to model upper tail dependence structure more effectively, as noted in Section 4.2.5. The trivariate asymmetric Gumbel-Hougaard copula introduced in Equation (4.12) is chosen to fit the drought characteristics, where $\varphi_i(t) = (-\ln t)^{\theta_i}$ and $i = 1, 2$. Recall that for a proper three dimensional copula to exist, the nested variable pairs within the copula have to have higher correlations. In

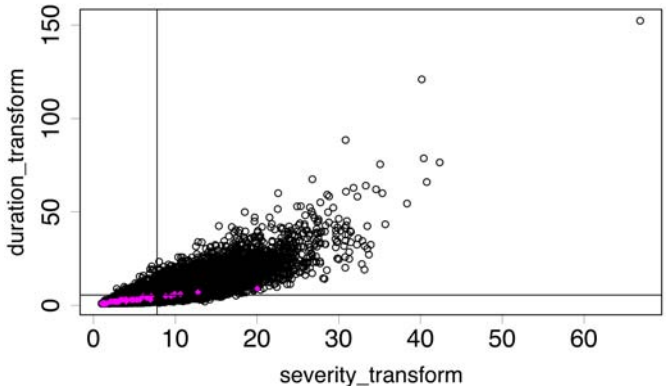


Figure 4.7: Plot of back-transformed simulations of severity against duration from Gaussian copula and 90% quantiles of margins, and historic data.

this example, severity and duration have the strongest correlation and hence are the nested variables. A constraint of this copula form as mentioned earlier, is the lack of a free parameter. As a result, correlations between u_1 and u_3, and u_2 and u_3 are identical.

The second step of the IFM method is to estimate the parameters of the copula $\hat{\theta}_1$ and $\hat{\theta}_2$ using Maximum Likelihood Estimation (MLE). Simulating variables from this trivariate asymmetric Gumbel-Hougaard copula requires the algorithm in Section 4.2.3 and Equation (4.9). Following this algorithm, u_1 is first simulated from $U(0,1)$ before u_2 is simulated from $C_2(u_2 \mid u_1)$ and u_3 is obtained from the conditional distribution of $C_3(u_3 \mid u_1, u_2)$. The conditional copulas $C_2(u_2 \mid u_1)$ and $C_3(u_3 \mid u_1, u_2)$ requires the partial differentiation of Equation (4.12) with respect to u_1 and, both u_1 and u_2 respectively. The parameters using this method are found to be $\hat{\theta}_1 = 2.18$ and $\hat{\theta}_2 = 3.97$. Figure 4.8 shows the corresponding contour plots of the simulations, of length 100,000, and the historic data.

Figure 4.9 shows the plot of the transformed simulations of severity against duration with respect to the historic data. Correlations calculated from the simulations of the trivariate Gumbel-Hougaard copula are shown in Table 4.7. By comparing these correlations to the Pearson correlations in Table 4.6, it is observed that these correlations are slightly lower. However, the upper tail dependence is captured by the Gumbel-Hougaard copula. This is important in the case of extreme events where positive correlations between variables with high values have serious consequences. The proportion of points in that upper tail is also higher than for the Gaussian copula.

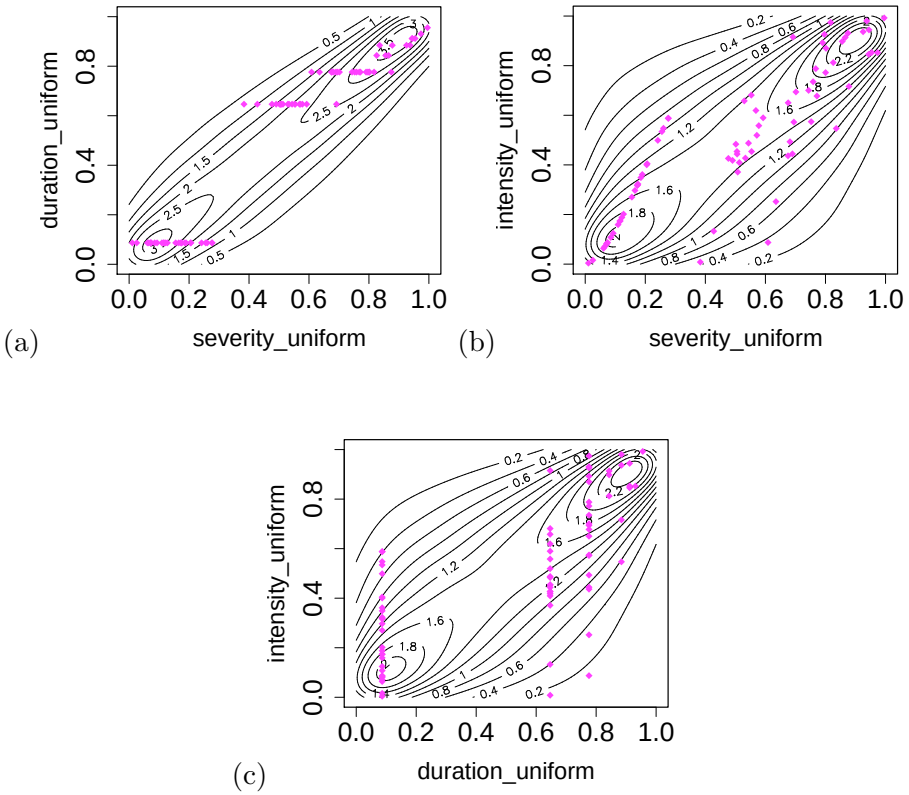


Figure 4.8: Empirical contour plot of (a) u_1 against u_2, (b) u_1 against u_3 and (c) u_2 against u_3 from a Gumbel-Hougaard copula.

4.3.5 Goodness-of-fit tests

In order to evaluate the fitted copula, goodness-of-fit tests are performed. Several goodness-of-fit tests are described in Genest *et. al.* [42]. The simplest is to compare the observed data, with a generated data set from the copula density. This is provided in the form of contour plots in Figures 4.6 and 4.8, with the superimposed observed data. Another is a plot of the empirical copula, C_n evaluated at the observed data, (u_{1i}, u_{2i}, u_{3i}), against the fitted copula, $\tilde{C}$, evaluated at the observed data $\tilde{C}(u_{1i}, u_{2i}, u_{3i})$. The empirical copula is:

$$C_n(u_{1i}, u_{2i,} u_{3i}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\left(\frac{R_i}{n+1} \leq u_{1i}, \frac{S_i}{n+1} \leq u_{2i}, \frac{T_i}{n+1} \leq u_{3i}\right) \quad (4.17)$$

where n is the sample size, $\mathbf{I}(A)$ denotes the indicator variable of the logical expression A and taking the value 0 if A is false and 1 is A is true, and the ranks of the ith observed Severity, Duration and Peak Intensity data are represented as R_i, S_i and T_i

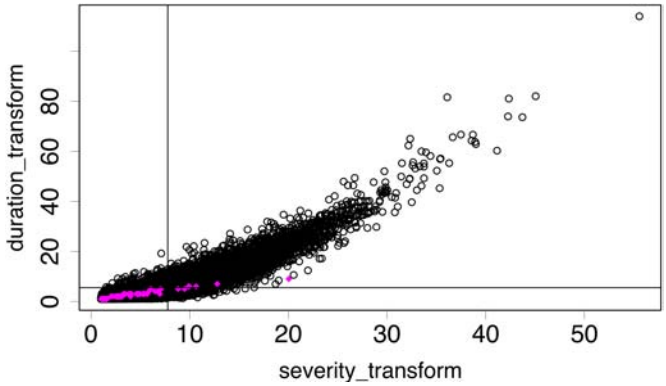


Figure 4.9: Plot of back-transformed simulations of severity against duration from Gumbel-Hougaard copula and 90% quantiles of margins, and historic data.

Table 4.7: Descriptive statistics of simulations from Gumbel-Hougaard and Gaussian copulas, u_1, u_2 and u_3 are severity, duration and intensity after transformation to uniform distributions.

Variable pairs	Correlation		Upper tail correlation		Proportion	
	Gumbel	Gaussian	Gumbel	Gaussian	Gumbel	Gaussian
u_1, u_2	0.913	0.928	0.788	0.586	0.0813	0.0746
u_1, u_3	0.729	0.856	0.561	0.416	0.0660	0.0633
u_2, u_3	0.728	0.678	0.557	0.232	0.0664	0.0455

respectively. The measure of fit is based on how close the points are to the diagonal line.

Figure 4.10 compares the goodness-of-fit between the Gumbel-Hougaard (circles) and Gaussian (diamonds) copulas for this set of data. The Gumbel-Hougaard copula provides a better fit than the Gaussian copula, because the points are closer to the diagonal. The lower left steep linear clusters are a consequence of many drought of one-month duration.

4.3.6 Discussion

This analysis have shown that employing the symmetric Archimedean copula with one dependence parameter is unrealistic for the case where there are more than two hydrological variables, since not all variable pairs possess similar dependence structures. The alternative is to apply the asymmetric Archimedean copula. The drawback of this form of copula, is the lack of one free parameter for one hydrological pair, since the two outer pairs of variables share the same dependence parameter. To overcome this

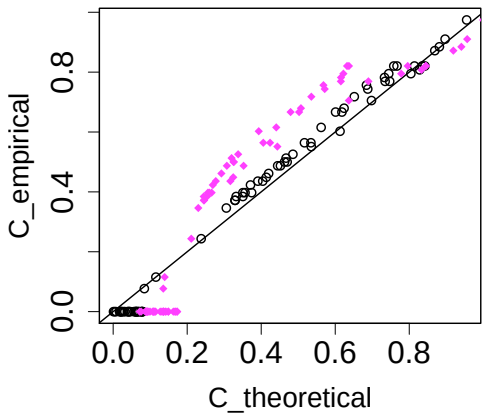


Figure 4.10: Plot of values from theoretical Gumbel-Hougaard copula (circles) and theoretical Gaussian copula (diamonds) against empirical copula

constraint, the Gaussian copula allows for separate dependence parameters between a large number of variable pairs. However, this copula is inadequate in describing tail dependence, which is important when examining the behaviour of extreme events. The t-copula, which is another copula from the elliptical copula family is capable of describing symmetric tail dependence and will be examined in the following chapter.

4.4 Summary of chapter

Copulas and their properties were introduced in this chapter for modelling the dependence structure between drought characteristics and to provide a description of the drought characteristics in terms of a small number of parameters. This chapter extended the application of copulas to the trivariate asymmetric case using a case study of rainfall district in NSW, where previous drought studies have only investigated the bivariate symmetric copula. The trivariate asymmetric copula allows for more realistic modelling of drought characteristics, since the dependence parameter for one pair of variables is different from the dependence parameter between either of the variables in that pair and a third variable.

Initial analyses reveal that times between the end of one drought and the beginning of the next, appear to be independent of these drought characteristics. Consequently, drought events can be modelled with a trivariate copula and an independent distribution of times. The difference between the Gaussian and Gumbel-Hougaard copulas is striking in Figure 4.10, but less so in the bivariate contour plots (Figures 4.6 and 4.8). The incapability of modelling tail dependence using the Gaussian copula is reflected

in Table 4.7, and comparing Figures 4.6 and 4.8.

Chapter 5

Regional drought modelling using copulas conditional on climatic states

The relationship between drought characteristics was established using copulas in Chapter 4. Global climatic variation such as ENSO is known to have a considerable effect on regional precipitation and drought in Australia [24]. In Eastern Australia, for example, climate variability is particularly influenced by the ENSO phenomenon and drought severity is associated with this phenomenon. Data from two rainfall districts in NSW, either side of the Great Dividing Range, are categorized into three states; El-Niño, Neutral and La-Niña, according to the prevailing SOI. Whereas previous studies and the previous chapter use monthly rainfall data, this research uses daily rainfall data to avoid the need to express duration as an integer number of months. Furthermore, the incorporation of climatic states in copula models has yet to be investigated in earlier studies. Gumbel-Hougaard copulas and t-copulas are first fitted to the droughts in the three states. The effect of the state on the estimated copula parameters is presented and discussed. The goodness-of-fit of the Gumbel-Hougaard and t-copulas are compared, and limitations of the two copula models are discussed. Finally, the fitted copulas are used to estimate recurrence intervals of at least one of the three variables exceeding critical values, and the recurrence intervals of all three variables exceeding critical values, taking into account the mixture of states.

5.1 Study Region

For this study, two rainfall districts are chosen on either side of The Great Dividing Range. District 58 is located east and District 63 is located west of the Great Divide. Figure 5.1 shows the location of these two rainfall districts in NSW [9].

The Great Dividing Range is Australia's largest mountain range, which stretches over more than 3500 km from the north-east of Queensland through the eastern coastline of NSW, moving into Victoria. The Great Dividing Range is an important factor in the amount of rainfall received in districts west and east of it due to the different strength of ENSO signals, and in general, the west receives significantly lower rainfall than the east [24]. District 58, north of NSW, covers an area of approximately 1767 km^2, with elevation of between 140 m to 160 m and has a mild sub-tropical climate. District 63 is also known as the Southern Tablelands, with an average elevation of 794 m which has been extensively cleared and used for grazing purposes. Initial regression of rainfall from districts east and west of the Great Divide shows strong association with ENSO indicators ($p < 0.005$).

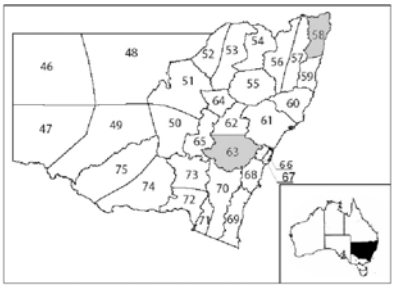


Figure 5.1: New South Wales (NSW) rainfall districts

5.2 Categorization of drought characteristics

Studies in Chapter 2 and 3 have established that global climatic oscillations particularly ENSO, have a major effect on Australia's rainfall due to its location. In Australia, it is usual to use persistently negative values of the SOI as an indicator of an El-Niño event, and such events are associated with droughts, while La-Niña events are characterised by positive values of SOI. Each year is classified as being in the El-Niño, La-Niña or Neutral state, based on the method proposed by Ropelewski and Halpert [103]. The five-month running means of the SOI are calculated and El-Niño events

are defined as any year where the SOI five-month running means remain below −0.5 standard deviations for five months or more during the water year (April to March the following year). In the same manner, years where the five-moving running means of the SOI is above 0.5 standard deviations for five months or longer are classified as a La-Niña event.

The sustained rainfall deficits in the respective districts are classified as occurring in an El-Niño (EN), La-Niña (LN) or Neutral (N) state, depending on the years in which they appear. If a given low rainfall period starts in a particular year and continues to the following year, the length of the event in each year is then counted and the event is categorized as occurring in the year that has the maximum number of days. The study period from 1900 to 2002 has 32 years categorized as El-Niño, 31 years as La-Niña and 40 years as Neutral state. For the purpose of this research, comparisons between the peripheral states, El-Niño and La-Niña, are focused on. Tables 5.1 and 5.2 provide the relevant statistics and correlations of the rainfall deficits based on the ENSO state.

Table 5.1: Means and Standard deviations of drought by ENSO state

Drought Variable	Average Intensity			Peak Intensity			Duration (days)		
	EN	N	LN	EN	N	LN	EN	N	LN
District 58									
No. of events	101	116	65	101	116	65	101	116	65
Events per year	3.16	2.90	2.10	3.16	2.90	2.10	3.16	2.90	2.10
Mean	1.27	1.21	1.19	1.47	1.35	1.35	24.7	18.3	21.1
Standard deviation	0.27	0.21	0.19	0.52	0.39	0.40	33.6	26.9	28.2
District 63									
No. of events	89	82	41	89	82	41	89	82	41
Events per year	2.78	2.05	1.32	2.78	2.05	1.32	2.78	2.05	1.32
Mean	1.30	1.23	1.25	1.56	1.39	1.45	40.5	20.4	22.4
Standard deviation	0.30	0.20	0.20	0.59	0.37	0.40	55.0	24.3	26.0

5.3 Statistical Analysis

5.3.1 Rainfall and Droughts

Daily precipitation data for Districts 58 and 63, from 1900 to 2002 are used for this analysis, Figures 5.2(a) and 5.2(b).

Table 5.2: Correlations between drought characteristics given ENSO state

Correlations	Pearson			Kendall's tau		
	EN	N	LN	EN	N	LN
District 58						
Peak Intensity, Average Intensity	0.92	0.96	0.95	0.86	0.87	0.85
Peak Intensity, Duration (days)	0.75	0.86	0.87	0.62	0.65	0.61
Average Intensity, Duration (days)	0.70	0.80	0.79	0.57	0.59	0.50
District 63						
Peak Intensity, Average Intensity	0.94	0.95	0.93	0.88	0.83	0.88
Peak Intensity, Duration (days)	0.86	0.84	0.89	0.72	0.69	0.69
Average Intensity, Duration (days)	0.82	0.79	0.84	0.69	0.62	0.64

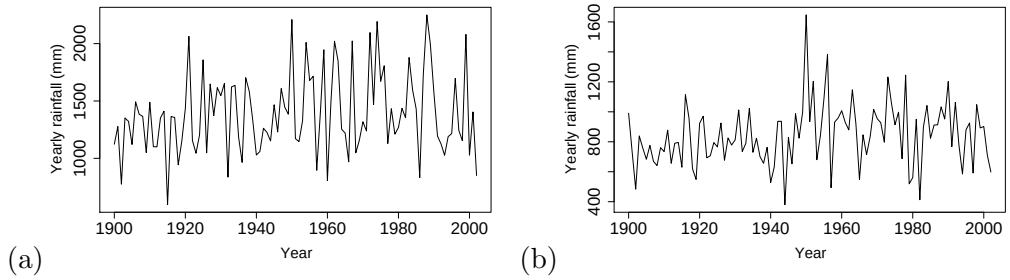


Figure 5.2: Yearly rainfall for (a) District 58 (b) District 63

To determine the presence of a trend or an association with ENSO indicators, regression analysis were performed using a measure of annual drought impact. Annual drought impact, over a water year, is defined as the sum of all intensity values less than the drought threshold of -1. From this analysis, there is no evidence of a trend in the annual drought impact over the 103 year period in either rainfall district, although there is evidence of a significant association with ENSO indicators. Results of this regression are given in Appendix B. Similar analyses carried out on non-drought periods did not indicate a linear trend but maintained an association with ENSO indicators. Turning to individual droughts, Tables 5.3 and 5.4 provide an overview of the short-term drought characteristics from both rainfall districts. Droughts in District 63 are slightly more intense and are slightly longer. The correlations between Average Intensity and Peak Intensity are similar in both districts, while the correlations between Average Intensity and Duration, and Peak Intensity and Duration are slightly higher in District 63.

Table 5.3: Means and Standard deviations of drought characteristics without segregating into ENSO states

Drought	Average Intensity		Peak Intensity		Duration (days)	
Variable	District 58	District 63	District 58	District 63	District 58	District 63
No. droughts	282	212	282	212	282	212
Mean	1.22	1.26	1.39	1.47	21.2	29.2
Std.	0.23	0.25	0.45	0.48	29.8	41.3

Table 5.4: Correlations of drought variables without segregating into ENSO states

Correlations	Pearson		Kendall's tau	
	District 58	District 63	District 58	District 63
Peak Intensity, Average Intensity	0.94	0.94	0.86	0.86
Peak Intensity, Duration (days)	0.81	0.85	0.64	0.70
Average Intensity, Duration (days)	0.75	0.81	0.57	0.65

5.3.2 Inter-drought duration

The times between the end of one drought and the beginning of the next is crucial when simulating for possible drought occurrences, since it provides information on drought rate. The inter-drought durations for each district have been classified as occurring in an El-Niño, La-Niña or Neutral state using the categorization method in Section 5.2. Table 5.5 summarizes the statistics of the inter-drought durations in Districts 58 and 63.

Table 5.5: Means and Standard deviations of inter-drought durations (days) given ENSO state

Drought Variable	EN	N	LN
District 58			
No. of inter-drought durations	105	112	65
Mean (days)	91.1	101	164
Standard deviation (days)	136	161	225
District 63			
No. of inter-drought durations	79	88	45
Mean (days)	91.4	146	250
Standard deviation (days)	120	211	266

In both districts, there are more droughts per year during El-Niño years and fewer

droughts per year during La-Niña years. It follows that the mean times between droughts are shortest during El-Niño years and longest during La-Niña years. These differences in mean are statistically significant in District 63 ($p = 0.002$), based on analysis of variance of the logarithms of inter-drought durations. The differences do not reach statistical significance in District 58 although they are in the expected direction ($p = 0.27$), from results of analysis of variance of the logarithms of inter-drought durations. The distributions are reasonably approximated as Weibull, based on the AIC values of the fitted distributions given in Table 5.6.

Table 5.6: AIC values of fitted distributions to inter-drought durations.

Distribution	EN	N	LN
District 58			
Gamma	1125	1221	765
Exponential	1160	1260	795
Weibull	1118	1215	763
District 63			
Gamma	843	1016	587
Exponential	874	1055	589
Weibull	841	1014	584

5.4 Marginal distributions of drought variables

As in Chapter 4, suitable marginal distributions for the three variables; Average Intensity, Peak Intensity and Duration, for the two rainfall districts are selected using the AIC. A number of marginal distributions such as Lognormal, Weibull, Normal, Exponential and Gamma distributions were fitted. The best fitting distributions for each of the variables are given in Tables 5.7 and 5.8, along with estimated parameters. Figures 5.3 and 5.4 also show the histograms of the drought characteristics and the fitted distributions with estimated parameters given in Tables 5.7 and 5.8 respectively, indicating a reasonable fit. After fitting the marginal distributions, the correlations between the transformed uniform variables $u_i = F(x_i)$, where $i = 1, 2, 3$, are determined. Table 5.9 shows the Pearson correlation coefficient for the three variables in both districts given their ENSO state. The transformation has little effect on the correlations and the strongest association is between Average Intensity and Peak Intensity for both districts.

Table 5.7: Estimated parameters of marginal distributions, District 58

Drought Variable	Distribution	Estimated Parameters
El-Niño		
Peak Intensity	Gamma	shape = 0.88, rate = 1.85
Average Intensity	Weibull	shape = 0.99, scale = 0.27
Duration (days)	Lognormal	mean = 2.75, std deviation = 1.53
La-Niña		
Peak Intensity	Weibull	shape = 0.84, scale =0.32
Average Intensity	Gamma	shape = 0.96, rate = 4.98
Duration (days)	Lognormal	mean = 2.65, std deviation = 1.63

Table 5.8: Estimated parameters of marginal distributions, District 63

Drought Variable	Distribution	Estimated Parameters
El-Niño		
Peak Intensity	Weibull	shape = 0.90, scale = 0.53
Average Intensity	Weibull	shape =0.99, scale = 0.30
Duration (days)	Lognormal	mean=2.83, std deviation =1.40
La-Niña		
Peak Intensity	Gamma	shape = 1.21, rate =2.70
Average Intensity	Gamma	shape=1.43, rate = 5.84
Duration (days)	Weibull	shape=0.90, scale=21.2

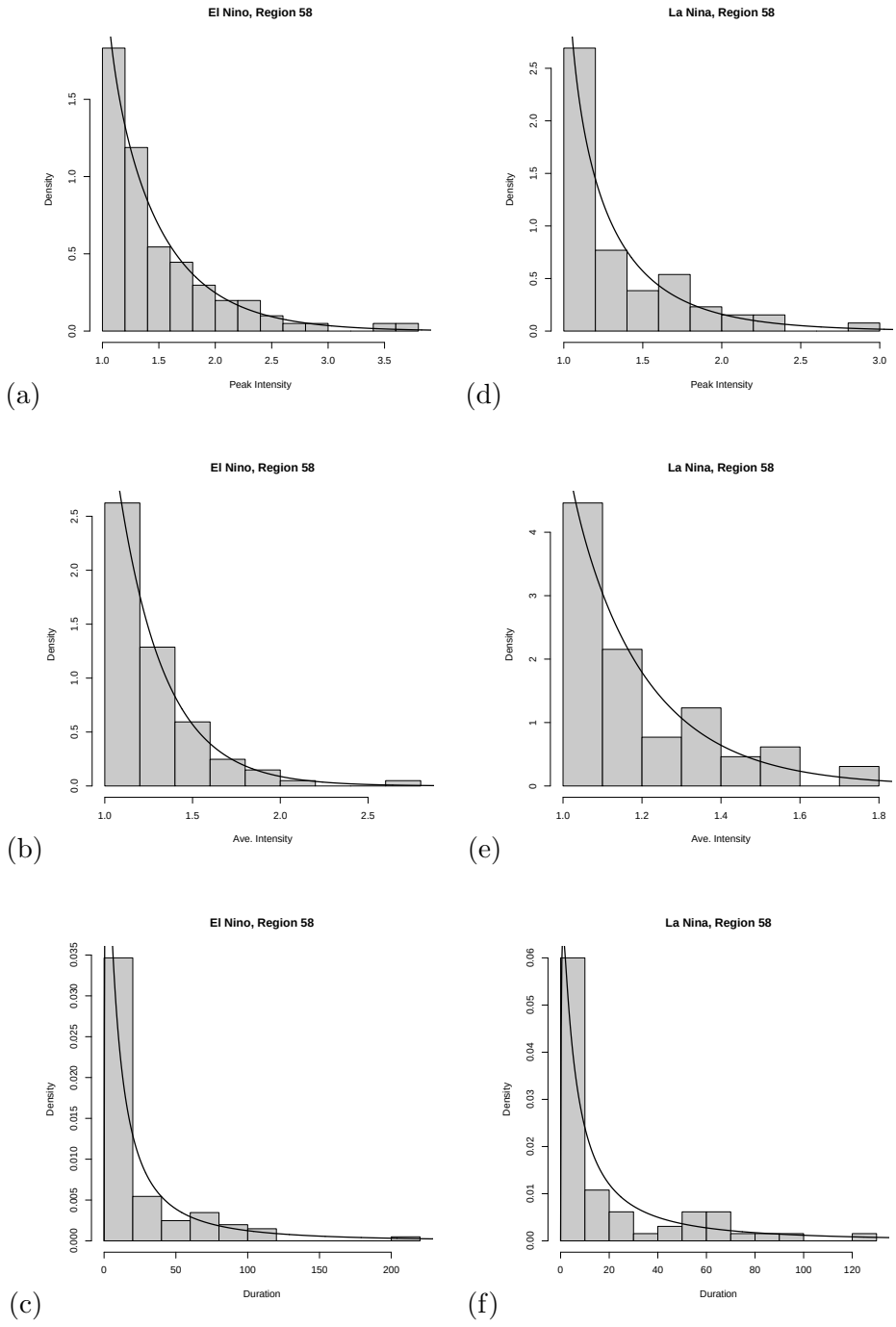


Figure 5.3: District 58: Histograms and fitted distributions of drought characteristics during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f).

The Gumbel-Hougaard copula has a restriction that the correlations between u_1 and u_3, and between u_2 and u_3 are equal. The statistical significance of the difference in the corresponding sample correlations was ascertained by Monte-Carlo simulation (1000 triplets generated). For District 63, there was no evidence of a difference at the

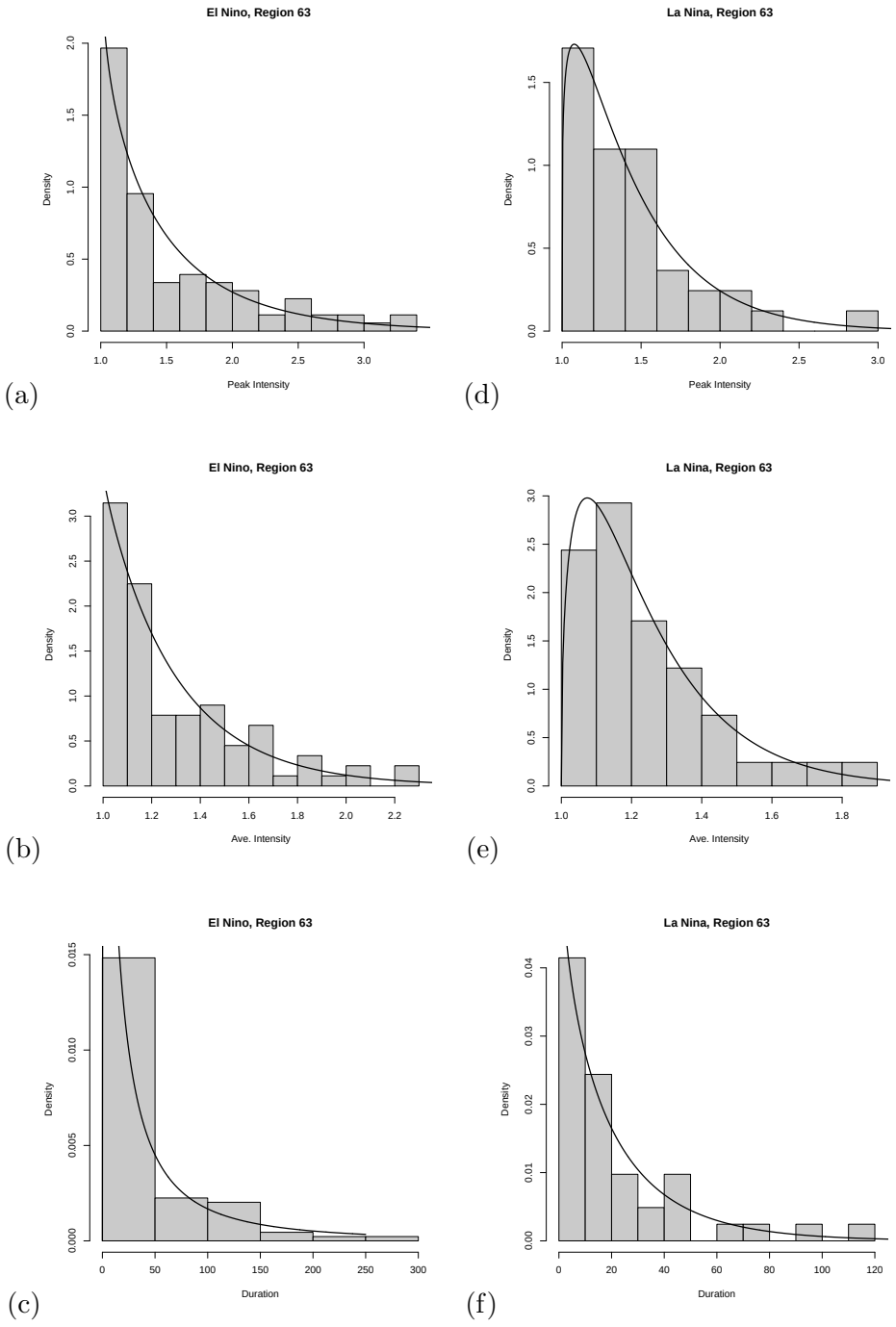


Figure 5.4: District 63: Histograms and fitted distributions of drought characteristics during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f).

20% level, but the difference was significant beyond the 1% level in District 58.

Table 5.9: Correlations of transformed uniform drought variables given ENSO state

Correlations	Pearson	
	El-Niño	La-Niña
District 58		
Peak Intensity, Average Intensity	0.97	0.96
Peak Intensity, Duration	0.83	0.83
Average Intensity, Duration	0.77	0.72
District 63		
Peak Intensity, Average Intensity	0.97	0.97
Peak Intensity, Duration	0.90	0.86
Average Intensity, Duration	0.87	0.83

5.5 Fitting Trivariate Gumbel-Hougaard Copulas

The parameters θ_1 and θ_2 need to be estimated for the Gumbel-Hougaard copula from Equation (4.12) given below, which was also used in Chapter 4. Let the transformed drought variables of Peak Intensity, Average Intensity and Duration be denoted as u_1, u_2 and u_3. Dependence parameters of this Gumbel-Hougaard asymmetric copula, $\hat{\theta}_1$ and $\hat{\theta}_2$ are determined using MLE, which requires the *cdf* of the copula. This is best found by differentiation using a symbolic algebraic package. Appendix C shows the **R** program used to obtain the best parameter estimates. The estimated parameters for both districts given their ENSO state and the standard errors for each of the parameters are given in brackets in Table 5.10.

$$C_1(C_2(u_1,u_2),u_3) \;=\; \exp[-([(-\ln u_1)^{\theta_2}+(-\ln u_2)^{\theta_2}]^{\frac{\theta_1}{\theta_2}}+(-\ln u_3)^{\theta_1}]^{\frac{1}{\theta_1}}]$$

The statistical significance of the difference in estimates of a parameter θ obtained from the La-Niña state and El-Niño state, $\hat{\theta}^{LN}$ and $\hat{\theta}^{EN}$ respectively, is approximately gained by referring the test statistic

$$\frac{\hat{\theta}^{LN}-\hat{\theta}^{EN}}{\sqrt{[\hat{sd}(\hat{\theta}^{LN})]^2+[\hat{sd}(\hat{\theta}^{EN})]^2}} \tag{5.1}$$

to a standard Normal distribution. For District 58, there is no evidence to suggest that either θ_1 or θ_2 are different between the El-Niño and La-Niña state, but in District 63,

Table 5.10: Estimated parameters of trivariate Gumbel-Hougaard copula given ENSO state in District 58 and 63, with standard errors given in brackets.

District	El-Niño	La-Niña
District 58	$\hat{\theta}_1^{EN} = 4.1$	$\hat{\theta}_1^{LN} = 4.3$
	(0.2)	(0.3)
	$\hat{\theta}_2^{EN} = 6.3$	$\hat{\theta}_2^{LN} = 6.5$
	(0.5)	(0.7)
District 63	$\hat{\theta}_1^{EN} = 5.1$	$\hat{\theta}_1^{LN} = 4.9$
	(0.3)	(0.3)
	$\hat{\theta}_2^{EN} = 7.4$	$\hat{\theta}_2^{LN} = 6.0$
	(0.7)	(0.6)

there is evidence that θ_2 is higher in the El-Niño state (Table 5.11).

Table 5.11: Test statistic and corresponding p-value (in brackets) for testing the hypotheses of the equality of parameters in Gumbel-Hougaard copulas in El-Niño and La-Niña states.

District\Parameter	θ_1	θ_2
District 58	0.55	0.23
	(0.29)	(0.41)
District 63	-0.47	-1.52
	(0.32)	(0.06)

The contour plots of the marginal bivariate copula density functions between pairs of u_1(Peak Intensity) and u_2(Average Intensity), u_1 and u_3(Duration) and u_2 and u_3, for both El-Niño and La-Niña states, in District 58 and 63 are shown in Figures 5.5 and 5.6 respectively. The observed data for both these data are superimposed on the contour plots as circles. The effect of the estimated parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ on the shape of the contours generated from the copula is observed particularly between the ENSO states in District 63. In general for both Districts 58 and 63, the contour lines representing the probability of 0.001 in El-Niño state fail to include at least one or more extreme observations. There is no major difference in the contours between both ENSO states in District 58, due to its coastal location relative to the Great Divide. The difference in rainfall received in the El-Niño and La-Niña states is not large in District 58.

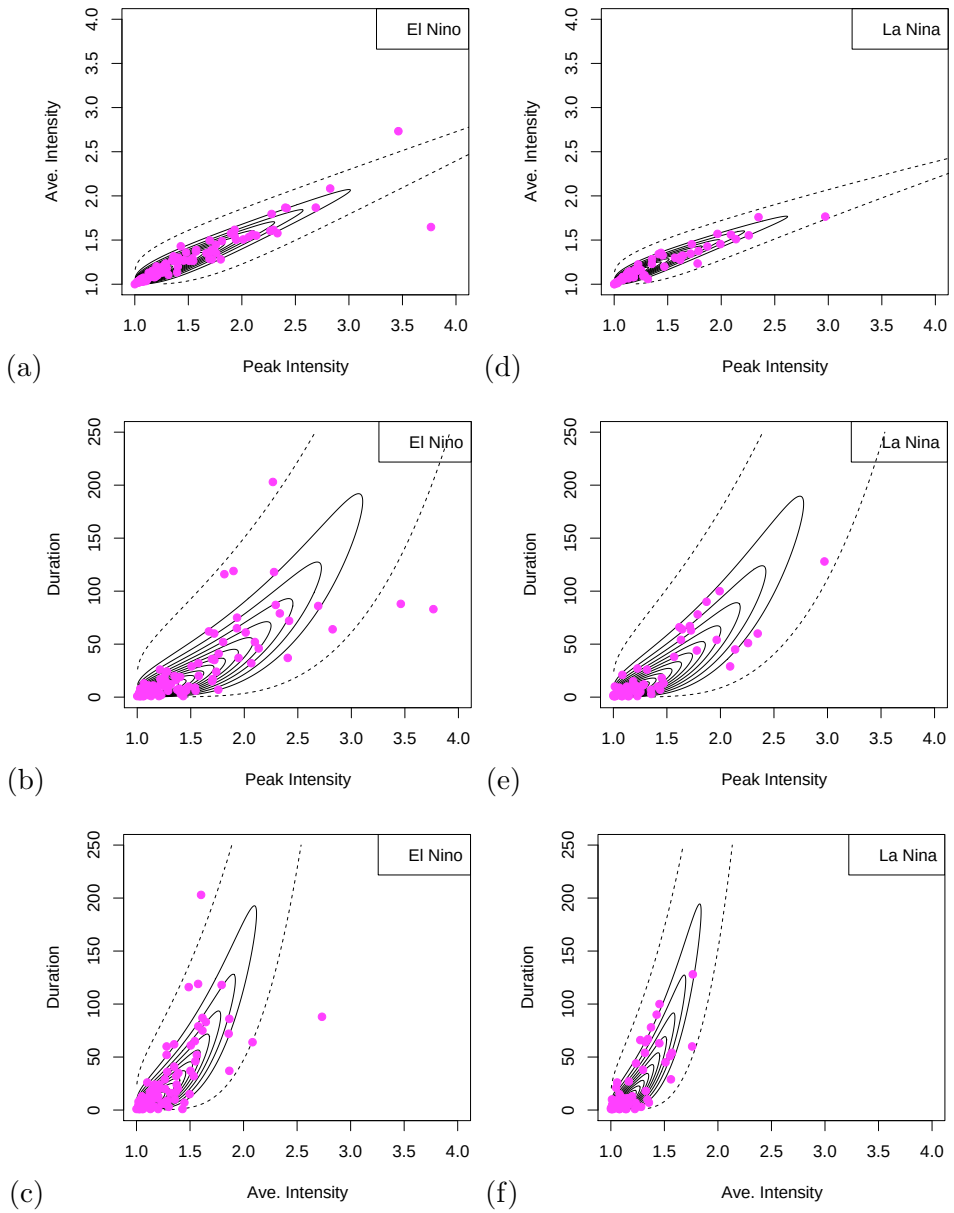


Figure 5.5: District 58: Marginal bivariate *pdf* contours from the Gumbel-Hougaard copula and observed pairs(circles) during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f). Contour lines represent 0.05 probability increments, where the outermost line is 0.05 and dotted lines represent 0.001.

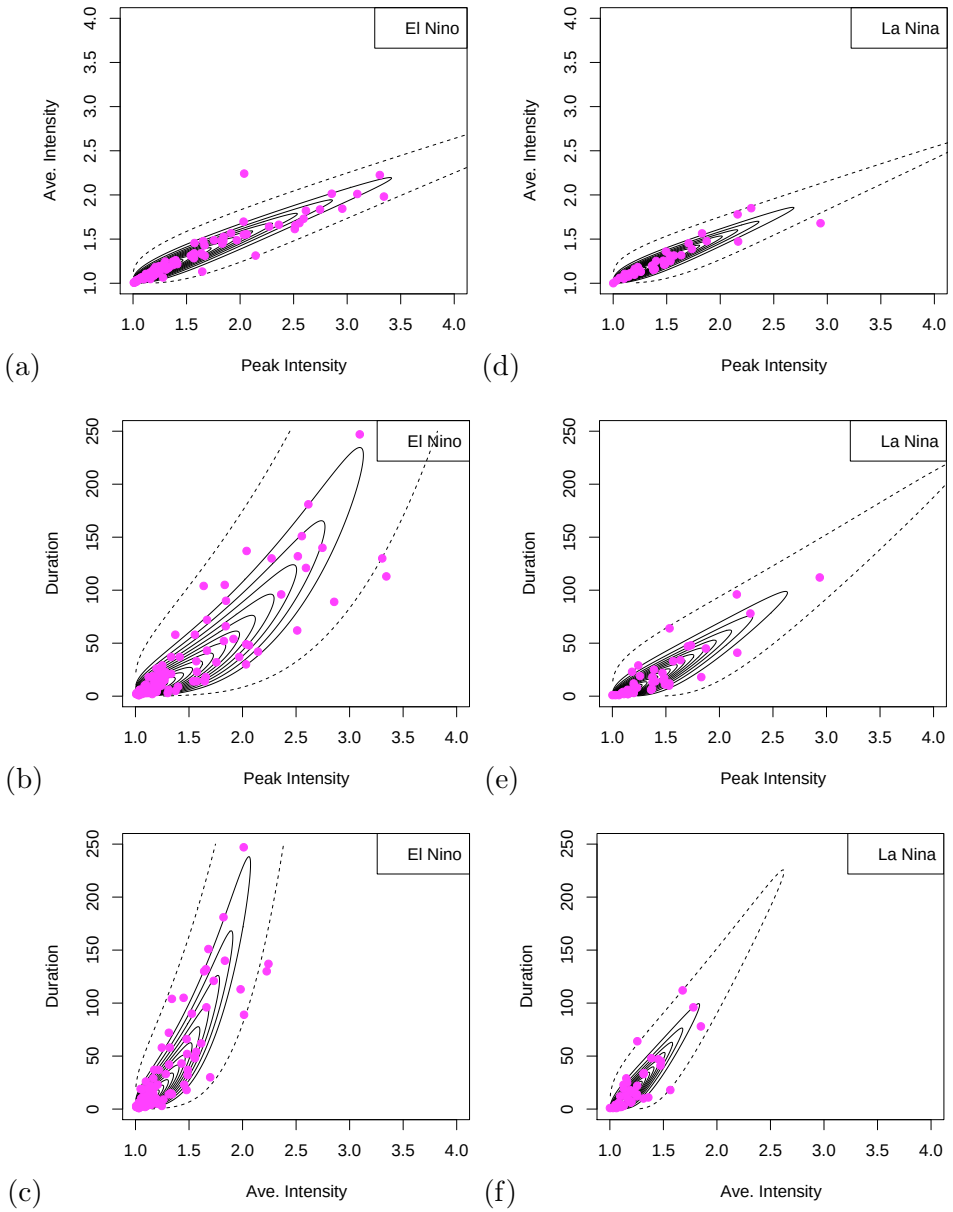


Figure 5.6: District 63: Marginal bivariate *pdf* contours from the Gumbel-Hougaard copula and observed pairs(circles) during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f). Contour lines represent 0.05 probability increments, where the outermost line is 0.05 and dotted lines represent 0.001.

During La-Niña conditions in District 63, the observed data is less disperse towards the upper tails as compared with those occurring during El-Niño conditions. Hence, the copula density produced during the La-Niña state is more highly correlated and more compact than the observed data in the El-Niño state.

5.6 Fitting Trivariate t-copulas

Given that the appropriate marginal distributions have been fitted to the drought characteristics data and transformed to the uniform scale, the R matrix in Equation 4.5 was estimated by the sample correlations and the fitted matrix was checked to be positive definite. The justification for setting the degrees of freedom ν to 10 for this application is explained in Section 5.8 by considering the sensitivity with respect to this parameter. Figures 5.7 and 5.8 show the contour plots of t-copula density for both ENSO states in District 58 and 63 respectively.

There is no observable difference in the fitted t-copulas between the El-Niño and La-Niña state in District 58, which again can be attributed to its coastal location relative to the Great Divide. On the other hand, the difference in fitted t-copulas between both ENSO states in District 63 is noticeable and is consistent with the finding in Section 5.5. In general, the fitted t-copulas performed relatively well in modelling the observations, since the contours are able to capture the extreme events.

Overall, there is a distinct difference in the shape of the t-copula density as compared to the Gumbel-Hougaard copula density. For both districts, the contours in the lower tail are more spread out than what is observed from the data. The dependence in the upper tail for the t-copula is as strong dependence as the Gumbel-Hougaard copula, since the contours are more spread out.

5.7 Measures of goodness-of-fit

Recall in Section 4.3.5 of Chapter 4 that the most straightforward method is to compare the observed data, with the copula density. In Figures 5.5 to 5.8, the copula density is represented by contours, with observed data superimposed. Another method is to plot the empirical copula, C_n evaluated at the observed data, (u_{1i}, u_{2i}, u_{3i}) also defined in Section 4.3.5, against the fitted copula, $\tilde{C}$, evaluated at the observed data $\tilde{C}(u_{1i}, u_{2i}, u_{3i})$. The empirical copula for this application is:

$$C_n(u_{1i}, u_{2i,}u_{3i}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}\left(\frac{R_i}{n+1} \leq u_{1i}, \frac{S_i}{n+1} \leq u_{2i}, \frac{T_i}{n+1} \leq u_{3i}\right) \qquad (5.2)$$

where n is the sample size, $\mathbf{I}(A)$ denotes the indicator variable of the logical expression A and taking the value 0 if A is false and 1 if A is true, and the ranks of the

ith observed Peak Intensity, Average Intensity and Duration data are represented as R_i, S_i, T_i respectively. The measure of fit is based on how close the points are to the diagonal line.

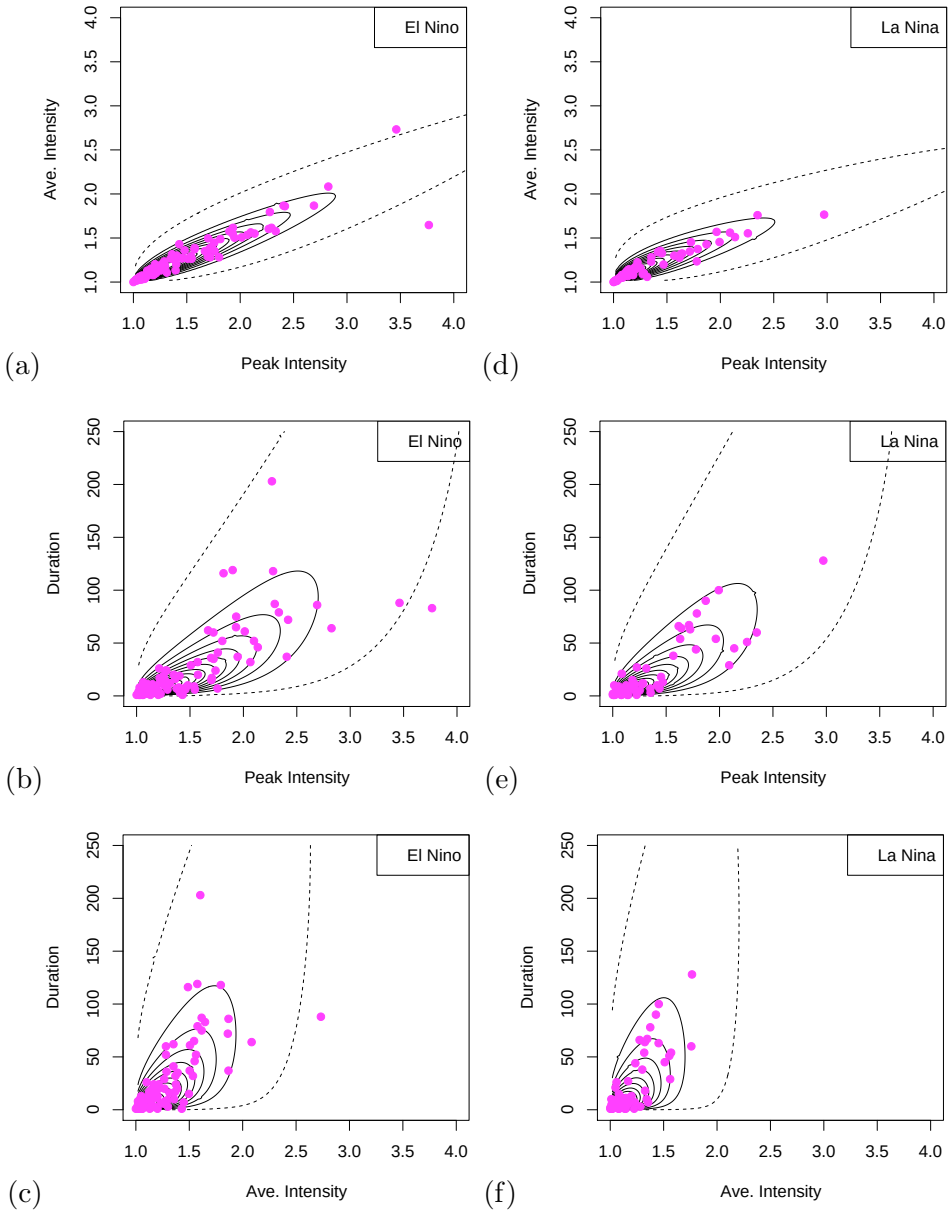


Figure 5.7: District 58: Marginal bivariate *pdf* contours from the t-copula and observed pairs(circles) during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f). Contour lines represent 0.05 probability increments, where the outermost line is 0.05 and dotted lines represent 0.001.

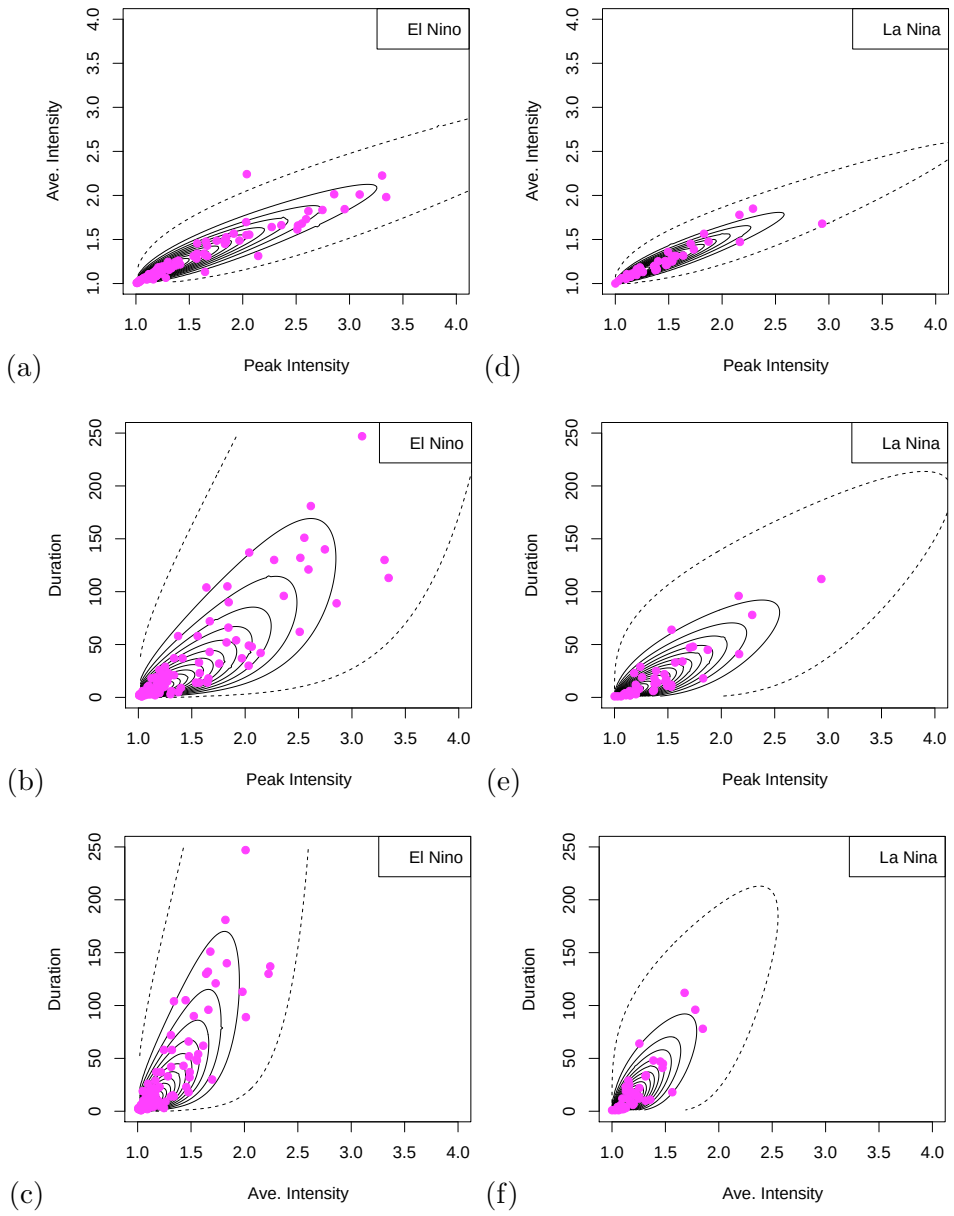


Figure 5.8: District 63: Marginal bivariate *pdf* contours from the t-copula and observed pairs(circles) during El-Niño state (a), (b) and (c); and La-Niña state (d), (e) and (f). Contour lines represent 0.05 probability increments, where the outermost line is 0.05 and dotted lines represent 0.001.

Figures 5.9(a) and (b) show the plots for District 58 and 63 respectively using the Gumbel-Hougaard copula. Figures 5.10(a) and (b) show the plots for the t-copula. The El-Niño state and La-Niña state are distinguished by circles and diamonds respectively. The correlations (r) are calculated for each district and Table 5.12 reports the values of $1-r$, along with the maximum absolute vertical distance of a point from the diagonal shown in italic. Additionally, Table 5.13 gives the RMSE values calculated between

the empirical and theoretical Gumbel-Hougaard and t-copulas for both El-Niño and La-Niña states.

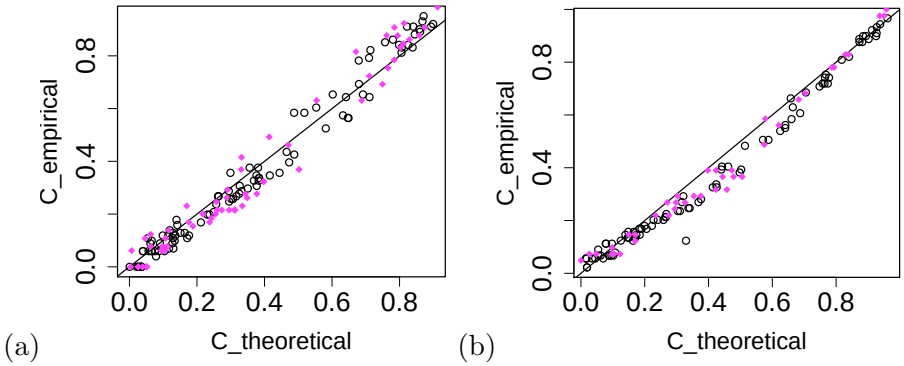


Figure 5.9: Fitted Gumbel-Hougaard copula against empirical copula for El-Niño (circles) and La-Niña (diamonds) in (a) District 58, (b) District 63

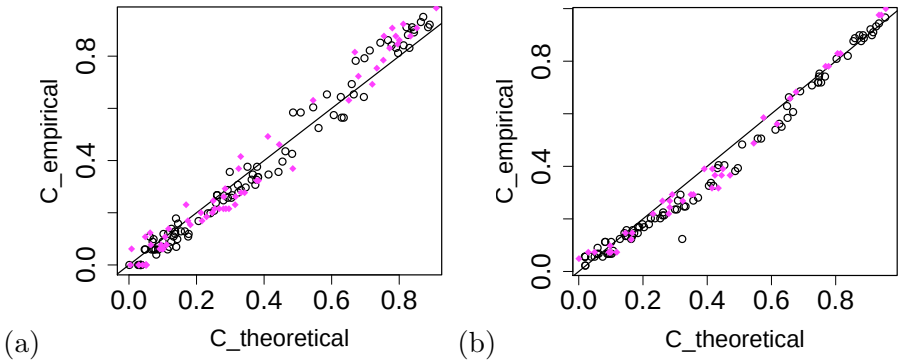


Figure 5.10: Fitted t-copula against empirical copula for El-Niño (circles) and La-Niña (diamonds) in (a) District 58, (b) District 63

For District 58, both plots show points lying close to the diagonal, indicating a satisfactory fit. There is some indication of slight curvature for all the plots in District 63, implying that the cumulative probabilities of the empirical copula are slightly less than those from the fitted copula. The numerical summaries in Table 5.12 are close for both copulas in both districts, and do not indicate that the fits of the theoretical copulas to observations from District 58 are any better than those for District 63.

Based on the RMSE values in Table 5.13, it appears that there is a better fit of the Gumbel-Hougaard copula in the El-Niño state compared to the La-Niña state for both districts. This observation only applies to District 58 when comparing the t-copulas for the two ENSO states, while in District 63, there is a similar RMSE value for both ENSO

Table 5.12: 1-r between empirical and theoretical copula with maximum absolute distance from diagonal line in brackets

District	El-Niño		La-Niña	
	GH copula	t-copula	GH copula	t-copula
District 58	0.0103	0.0093	0.0165	0.0126
	(*0.1082*)	(*0.1110*)	(*0.1448*)	(*0.1467*)
District 63	0.0096	0.0086	0.0136	0.0105
	(*0.2058*)	(*0.1987*)	(*0.1406*)	(*0.117*)

Table 5.13: RMSE values between empirical and theoretical Gumbel-Hougaard and t-copulas for both El-Niño and La-Niña states.

District	El-Niño		La-Niña	
	GH copula	t-copula	GH copula	t-copula
District 58	0.0473	0.0481	0.0620	0.0588
District 63	0.0524	0.0470	0.0559	0.0463

states. Overall, the t-copula fits slightly better than the Gumbel-Hougaard copula in terms of these numerical summaries.

5.8 Trivariate Return Periods and Probability of Exceedance

For effective drought mitigation planning to be carried out, it is essential that environmental and government agencies have sufficient information on the probability of a drought exceeding a certain magnitude for each of the drought characteristics, and consequently the corresponding return period of such droughts. The return period of a drought of some specified magnitude is defined as the mean time between such drought events. Salvadori and De Michele [106] gives a detailed explanation of return periods.

For some given percentile, upper 10% and 1% are considered here, let c_1, c_2 and c_3 represent the percentiles of the annual marginal distributions of Peak Intensity, Average Intensity and Duration respectively. The three states are denoted by: El-Niño (1), Neutral (2) and La-Niña (3). The proportion of time for which these ENSO states occur over the 103 year study period is denoted by ω_i where $i = 1, 2$ or 3 and $\sum \omega_i = 1$. The drought rate per year for each ENSO state is represented by λ_i. The rate of occurrence for each ENSO state and the drought rate per year within each ENSO state, for both

districts, are displayed in Table 5.14. Then the trivariate drought *cdf* $H_i(c_1, c_2, c_3)$ for $i = 1, 2, 3$ representing the ENSO state can be combined to obtain the overall annual *cdf* $F(c_1, c_2, c_3)$ as follows [72]:

$$\begin{aligned}
&F(X_1 < c_1, X_2 < c_2, X_3 < c_3) \\
&\quad = \Pr(\text{year being in state 1})\{\Pr(0 \text{ state 1 drought in a year}) \\
&\quad + \Pr(1 \text{ state 1 drought in a year}) \times H_1(c_1, c_2, c_3) + \cdots \\
&\quad + \Pr(k \text{ state 1 droughts in a year}) \times [H_1(c_1, c_2, c_3)]^k\} \\
&\quad + \Pr(\text{year being in state 2})\{\Pr(0 \text{ state 2 drought in a year}) \\
&\quad + \Pr(1 \text{ state 2 drought in a year}) \times H_2(c_1, c_2, c_3) + \cdots \\
&\quad + \Pr(k \text{ state 2 droughts in a year}) \times [H_2(c_1, c_2, c_3)]^k\} \\
&\quad + \Pr(\text{year being in state 3})\{\Pr(0 \text{ state 3 drought in a year}) \\
&\quad + \Pr(1 \text{ state 3 drought in a year}) \times H_3(c_1, c_2, c_3) + \cdots \\
&\quad + \Pr(k \text{ state 3 droughts in a year}) \times [H_3(c_1, c_2, c_3)]^k\} \\
&\quad = \sum_{i=1}^{3} \omega_i e^{-\lambda_i(1-H_i(c_1,c_2,c_3))}
\end{aligned} \tag{5.3}$$

where k is the maximum possible number of droughts in a year.

Table 5.14: Rate of ENSO state and drought rate per year for each ENSO state, 1900-2002

ENSO state	ω_i	$\lambda_{58,i}$	$\lambda_{63,i}$
El-Niño (1)	0.31	3.16	2.78
Neutral (2)	0.39	2.90	2.05
La-Niña (3)	0.30	2.10	1.32

The derivation of the *cdf* in Equation (5.3) is used to determine the annual probabilities of exceeding some threshold (AEP). Consider the following two calculations, (i) the probability of exceeding any one threshold of the drought characteristic:

$$AEP^{\cup} = \Pr(X_1 > c_1 \cup X_2 > c_2 \cup X_3 > c_3) = 1 - F(c_1, c_2, c_3) \tag{5.4}$$

and (ii) the probability of exceeding all thresholds at the same time:

$$\begin{aligned} AEP^{\cap} &= \Pr(X_1 > c_1 \cap X_2 > c_2 \cap X_3 > c_3) \\ &= 1 - F(c_1, \infty, \infty) - F(\infty, c_2, \infty) - F(\infty, \infty, c_3) \\ &+ F(c_1, c_2, \infty) + F(c_1, \infty, c_3) + F(\infty, c_2, c_3) - F(c_1, c_2, c_3) \end{aligned} \tag{5.5}$$

where for example, $F(c_1, \infty, \infty)$ corresponds to a univariate annual marginal distribution and $F(c_1, c_2, \infty)$ is a bivariate marginal distribution. The return period or average recurrence interval (ARI) is the reciprocal of the exceedance probability. The upper 10th percentile of the annual marginal distribution of Peak Intensity, c_1 is found using Equation (5.3) by replacing $H_i(c_1, c_2, c_3)$ by $H_i(c_1, \infty, \infty)$ and equating the right side of the equation to 0.9. The other c_i's are found in a similar manner.

The sensitivity of $AEP^{\cap}$ and $AEP^{\cup}$ to the parameter ν of the t-copula can be investigated. Table 5.15 shows the upper 10th percentile probabilities for both District 58 and 63 for varying degrees of freedom. The probabilities do not change significantly when the degrees of freedom are increased, which helps justify the choice of 10 as the degrees of freedom in Section 5.6.

Table 5.15: Probabilities of exceedence corresponding to the upper 10th percentile in District 58 and 63

Degrees of Freedom	$AEP^{\cap}$	$AEP^{\cup}$
District 58		
5	0.055	0.148
10	0.053	0.150
15	0.053	0.151
20	0.052	0.152
District 63		
5	0.061	0.142
10	0.059	0.144
15	0.059	0.145
20	0.059	0.145

Table 5.16 shows the corresponding return periods for both exceedance probabilities for Districts 58 and 63. Overall, there is a longer return period in District 63 as compared to District 58. If all variables were independent, this would correspond to a return period for the event $AEP^{\cup}_{indep} = 3.558$, for the quantile being 0.9. On the other hand, the return period for the event $AEP^{\cap}_{indep} = 1/0.001 = 1000$ if the variables were

all independent. Considering the case when the quantile is 0.9, the return periods for the event $AEP^{\cup}$ when the dependence structure is taken into account is more than double that of $AEP^{\cup}_{indep}$ for all districts and copula models. The difference is even greater when $AEP^{\cap}$ for all cases are compared to $AEP^{\cap}_{indep}$.

For comparison, Table 5.17 shows the corresponding return periods calculated from the t-copula. Again, there is an overall longer return period for District 63. When the return periods obtained from both the Gumbel-Hougaard copula and t-copula are compared, return periods calculated for the event $AEP^{\cap}$ are lower using the Gumbel-Hougaard copula but are higher when computing the return periods for the event $AEP^{\cup}$. When the return periods for $AEP^{\cup}$ are compared separately for Gumbel-Hougaard and t-copulas, there is very slight difference in return periods between Districts 58 and 63. However, this is not the case when the return periods for $AEP^{\cap}$ are compared.

Calculating return periods which only involve univariate cases can be misleading and may result in incorrect analysis of drought risk, when correlations exist among variables. The copula approach provides a framework for the joint dependence structure of the drought characteristics to be analyzed. Hence, return periods for droughts exceeding a particular threshold can be determined through the use of copulas and the drought risk can be established.

Table 5.16: Return period using trivariate Gumbel-Hougaard copula given probability of exceedance for Districts 58 and 63

Return period (years)	Quantile	Return period for $AEP^{\cup}$ (years)	Return period for $AEP^{\cap}$ (years)
District 58			
10	0.9	7.6	10.9
100	0.99	68	142
District 63			
10	0.9	7.8	13.4
100	0.99	70	162

5.9 Summary of Chapter

The trivariate Gumbel-Hougaard copula, with Gumbel marginal distributions, includes the trivariate Gumbel distribution as a special case and in some cases, there is a phys-

Table 5.17: Return period using trivariate t-copula given probability of exceedance for Districts 58 and 63

Return period (years)	Quantile	Return period for $AEP^{\cup}$ (years)	Return period for $AEP^{\cap}$ (years)
District 58			
10	0.9	6.8	11.7
100	0.99	60	192
District 63			
10	0.9	6.9	16.8
100	0.99	61	225

ical rationale for choosing Gumbel distributions to model extreme values of environmental variables. However, the form of the copula has more to do with mathematical tractability than physical arguments. The trivariate Gumbel-Hougaard copula has the substantial limitation that the two outer correlations are identical. In contrast, the trivariate t-copula has no restriction on the three correlations beyond the variance-covariance matrix being positive definite. However, the Gumbel-Hougaard copula has fewer parameters to be estimated and precise estimation of the degrees of freedom in the t-copula is elusive. Both copulas have the property of tail dependence, which is generally considered appropriate for modelling hydrological extremes. The contour plots in Figures 5.5 to 5.8 indicate that both copulas provide a reasonable fit to the data with only one or two points lying beyond the 0.001 contour, with the t-copula having fewer such outliers.

There is also little difference between the Gumbel-Hougaard copula and t-copulas in terms of the goodness-of-fit tests, but there is a noticeable difference in the predicted ARI of all three variables exceeding the annual marginal upper 1% points. In practice, the t-copula should be used if there is evidence that the outer correlations differ. If there is no such evidence, and assuming equal outer correlations are plausible, use of either copula may be justified. Comparing the results from fitting both copulas can provide a sensitivity analysis.

The marginal distributions of the drought variables are quite different for the different climate states, droughts being generally more frequently and more severe in the El-Niño state. The effects of the climate states on the correlation structure modelled by the copula however are more subtle. For District 63, the contours of the fitted marginal bivariate distributions are visibly more dispersed in the El-Niño state than they are

in the La-Niña state. Furthermore, for District 63, the estimated copula parameters are significantly different between the two ENSO states. For District 58, there is little visible difference. This difference between District 58 and 63 may be attributable to their location relative to the Great Dividing Range.

Chapter 6

Single site drought forecasting using adaptive stochastic models

Having modelled the relationship between drought characteristics, this thesis moves on to methods of forecasting drought. In Chapter 2, a relationship was observed between rainfall on the eastern coast of Australia and climatic indicators such as the SOI. This chapter aims at investigating the effect of including such climatic factors for both short and longer term drought predictions, in stochastic models.

Forecasts of SPI(3) with a lead time of one month and SPI(12) with a six-month lead are made at three rainfall gauges in NSW. For SPI(3), four forecasting methods are compared: ARMA(3,0); ARMA(3,3); ARMA(3,0) with climatic indicators SOI and MEI; rainfall on past values of rainfall, SOI and MEI, with rainfall being combined with known rainfall at times t and $t-1$ to give a forecast of SPI(3) at time $t+1$. For SPI(12) with a six-month lead time, three models are considered: autoregression model of SPI(12) at time $t+6$ on SPI(12) at times $t, \ldots, t-6$; as the preceding model with SOI and MEI; as preceding model with SOI and MEI with higher weight given to large negative SPI(12) in the fitting.

6.1 Preliminary analysis

6.1.1 Statistics of monthly rainfall data

Monthly rainfall data from three individual rainfall gauges in District 61 in NSW were selected: Clarence Town, Dungog Post Office and Blackville. District 61 was chosen

because this region supplies a high percentage of water to agricultural industries in the surrounding regions. Also, the main agricultural industries in this district are wheat, grain and beef. Both Clarence Town and Dungog Post Office rainfall gauges lie within the Williams River Catchment, on the coastal fringe of NSW 200km north of Sydney. Blackville is located in the Upper Hunter region, about 256km inland from the other two rainfall gauges. Figure 6.1 shows the location of the first two gauges and Table 6.1 gives the statistics for the rainfall data. The monthly rainfall record shows that the three rainfall stations receive roughly similar rainfall. Overall, Clarence Town station receives the highest mean monthly rainfall. Figures 6.2, 6.3 and 6.4 shows the time series of the monthly rainfall of the record period from Clarence Town, Dungog Post Office and Blackville stations respectively. Figure 6.4 appears to have fewer extreme events as compared to the other two gauges. Also, from these figures, there is no apparent linear trend. Results from regression analysis provides no evidence of a linear trend (Table 6.2), so none is included in the regression models. Similar seasonal variation is present for all three stations.

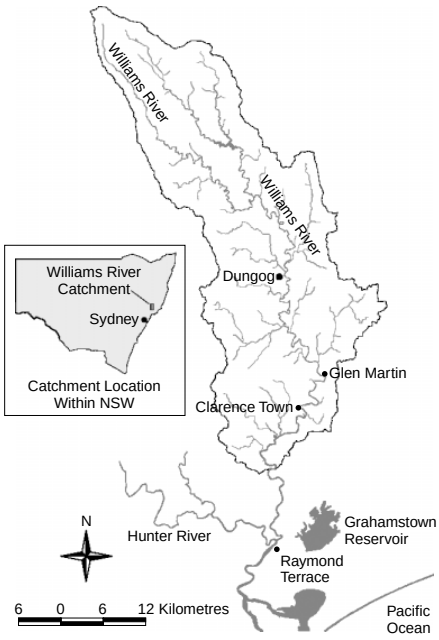


Figure 6.1: The Williams River Catchment [62].

6.1.2 Statistics of climatic indicators

The monthly SOI data from 1913 to 2002 obtained from the Bureau of Meteorology (BoM) website [13] and monthly MEI data from 1950 to 2003 derived from Wolter

Table 6.1: Basic statistics of monthly rainfall from rainfall gauges.

Rainfall Gauge	Record period	Mean rainfall (mm)	Median rainfall (mm)	Standard deviation
Clarence Town	1895-2002	89.01	65.9	80.4
Dungog Post office	1897-2000	82.46	61.2	73.6
Blackville	1885-2001	74.68	58.2	63.43

Table 6.2: Results from regression models with fitted linear trend.

Rainfall Gauge	Trend coefficient	p-value of regression model
Clarence Town	0.00005	0.99
Dungog Post office	0.009	0.15
Blackville	0.002	0.64

and Timlin [137] are used in this chapter. Table 6.3 displays the statistics for the SOI and MEI. Figure 6.5 shows the time series plot of the monthly SOI. There is a strong evidence of a slight decreasing linear trend over this period (-0.003 per month). It is observed from Figure 6.6, that there is a slight increasing trend (0.002, p-value ¡ 0.05) in the monthly MEI time series plot. Further regression results can be found in Appendix D.

Table 6.3: Basic statistics of climatic indicators

Climatic indicator	Record period	Mean	Median	Standard deviation
SOI	1913-2002	0.049	0.250	10.12
MEI	1950-2003	0.047	0.0005	0.99

6.2 Short-term forecasting for one-month ahead

In this section, regression models and autoregressive moving average (ARMA) models are fitted to both rainfall and SPI(3), to predict SPI for short-term drought. In Chapter 2, SPI was introduced as a measure of meteorological drought and SPI(3) is often used to measure short-term droughts. Due to the different record length of rainfall and climatic indicators, the following forecasting models are developed based on data from 1950 to 1989 and one-month ahead predictions are made from 1990 to 1999 using the

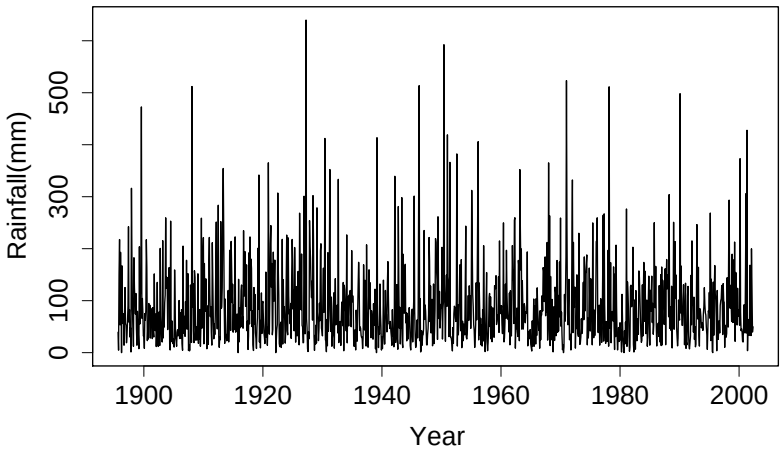


Figure 6.2: Time series of monthly rainfall for Clarence town station, 1895-2002.

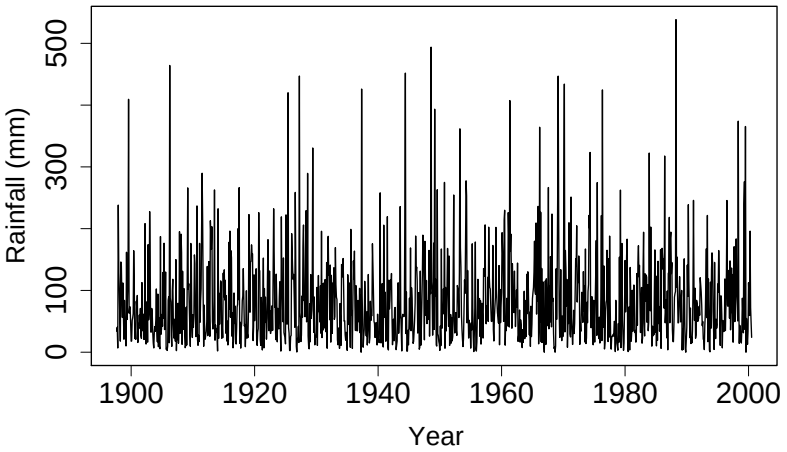


Figure 6.3: Time series of monthly rainfall for Dungog post office station, 1897-2000.

model fitted to 1950-1989 data. These predicted values are then compared to the actual rainfall received during that period, to assess the forecasting performance of these models.

6.2.1 ARMA Model for SPI(3)

Time series models, particularly the ARMA models are a straightforward method for simulating and forecasting the stochastic nature of rainfall and consequently drought. In general, ARMA models are formed by combining both the moving average (MA) and autoregressive (AR) processes. An ARMA process of order (p, q) which combines p AR terms and q MA terms is then defined as:

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + Z_t + \beta_t Z_{t-1} + \cdots + \beta_q Z_{t-q} \tag{6.1}$$

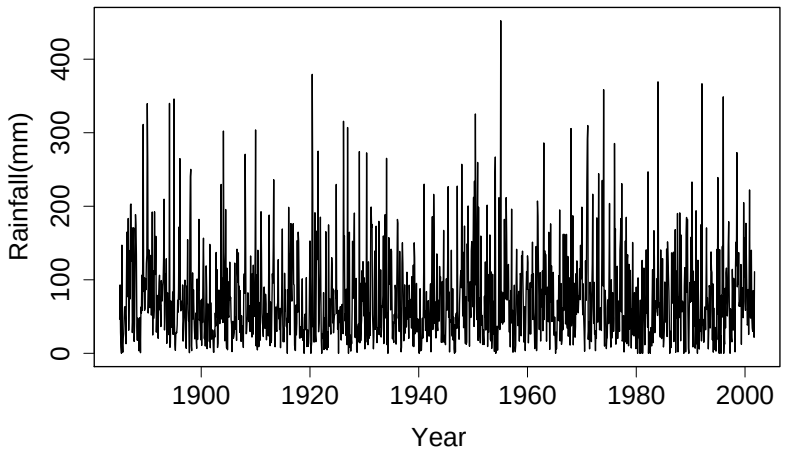


Figure 6.4: Time series of monthly rainfall for Blackville station, 1885-2001.

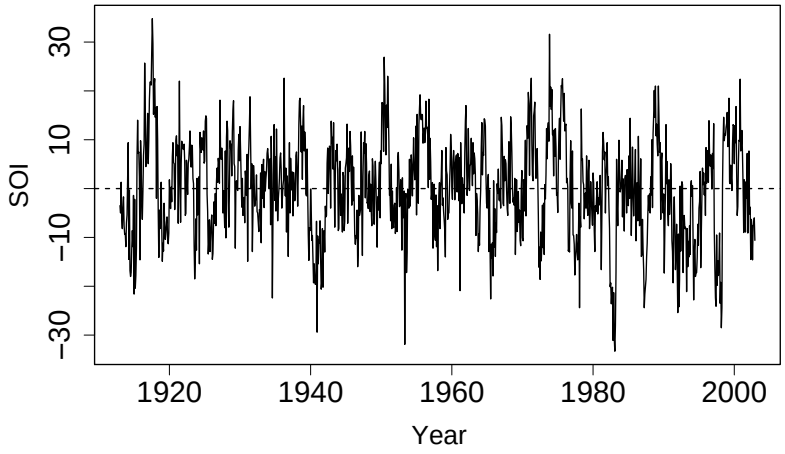


Figure 6.5: Time series of monthly SOI, 1913-2002.

where Z_t is a discrete, random process with mean 0 and variance σ_Z^2, Y_t is the observed time series and α_i, β_i are constants.

Rainfall in Australia is often characterized by a seasonal component from Figures 6.2, 6.3 and 6.4. For short-term predictions in this case study, SPI(3) is applied to rainfall for all three stations. The calculation of SPI requires the moving average of the rainfall be converted using the Gamma *cdf* with parameters corresponding to the respective month. Hence, this step deseasonalises the SPI(3) and a seasonal component is not required in this ARMA model.

Identifying an appropriate ARMA model is a vital step in model fitting. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was used to test for stationarity in all three stations and the p-values of more than 0.1 for all three stations indicate that stationarity cannot be rejected here, hence differencing is not required. An initial assessment of model

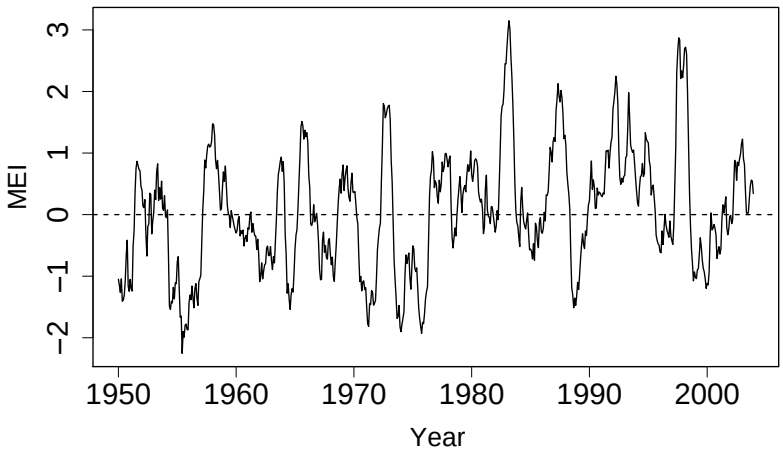


Figure 6.6: Time series of monthly MEI, 1950-2003.

order, that is (p, q), of the ARMA model can be made by visual inspection of the correlogram and partial correlogram. In general, a cut-off in the correlogram at integer q indicates an MA(q) process is a possible model. Similarly, a cut-off in the partial correlogram at integer p suggests an underlying AR(p) process. A more formal process of model selection also relies on the values of p and q that will yield a minimum AIC (Akaike's Information Criterion). Figure 6.7 shows the correlogram and partial correlogram for all three rainfall stations and Figure 6.8 shows the correlogram of the residuals from the fitted ARMA models.

Table 6.4: AIC values of fitted ARMA models for all three stations.

Rainfall Gauge	AIC value of ARMA(3,0) model	AIC value of ARMA(3,3) model
Clarence Town	1168	1087
Dungog Post Office	1200	1119
Blackville	1248	1164

The correlogram plot in Figure 6.7 shows a sine-wave pattern with a cut-off at lag 3, which indicates that fitting a MA(3) model may be suitable. Likewise, the partial correlogram shows a cut-off at lag 3. Correlogram of the residuals from the ARMA(3,0) model shows slight autocorrelation between observations. On the other hand, the residuals from the ARMA(3,3) model are not statistically significant, indicating that the fitted model has removed this autocorrelation. AIC values for the fitted models in Table 6.4, indicate that ARMA(3,0) and ARMA(3,3) were the most suitable models. Hence, an updated ARMA(3,0) and ARMA(3,3) are employed to predict one month ahead. For each one-month ahead prediction, the parameters in the respective ARMA

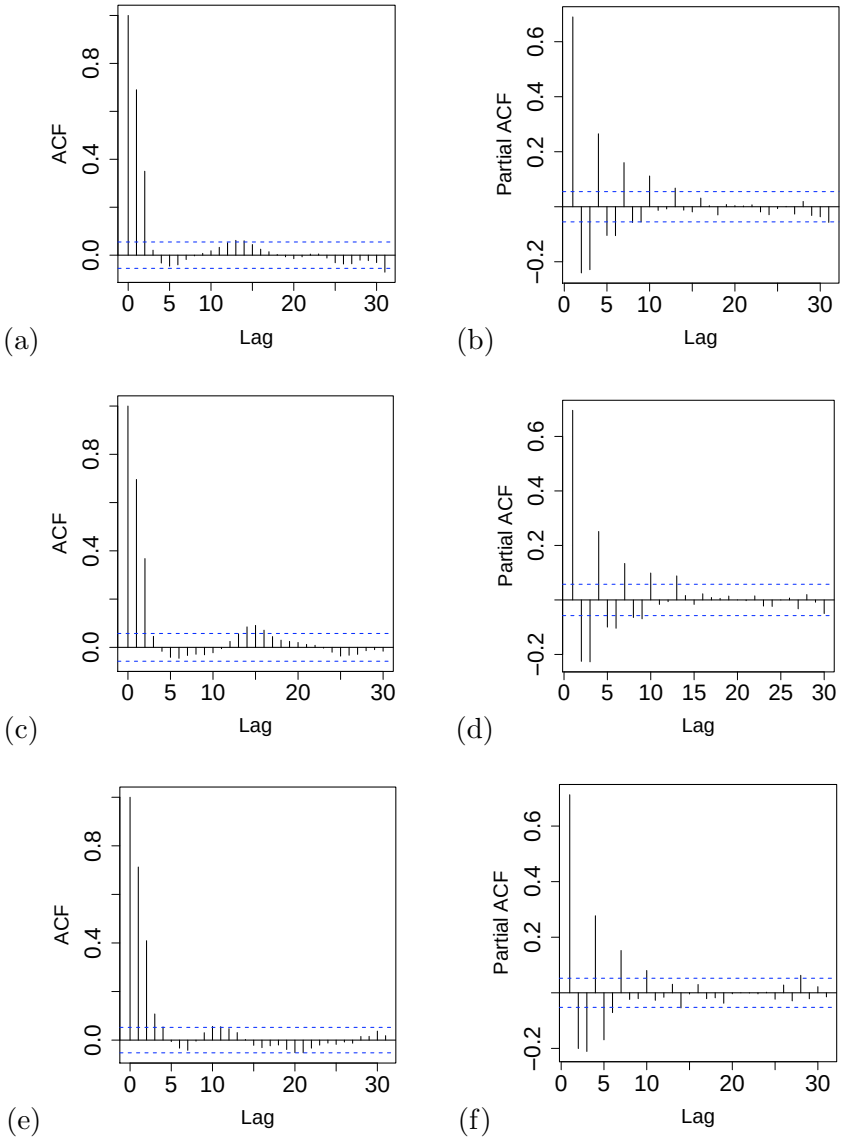


Figure 6.7: Correlogram for (a) Clarence town, (c) Dungog Post office and (e) Blackville; and Partial Autocorrelogram for (b) Clarence Town (d) Dungog Post office and (f) Blackville.

models are revised each time a new month is added.

6.2.2 Regression Model for SPI(3)

To account for the influence of climatic phenomena, an ARMA(3,0) model with climatic indicators is developed as a regression model.

$$\begin{aligned}\hat{S}_{t+1} = \beta_0 + \beta_1 S_t(3) + \beta_2 S_{t-1}(3) + \beta_3 S_{t-2}(3) \\ + \beta_4 SOI5_t + \beta_5 MEI_t + \beta_6 (SOI5_t \times MEI_t)\end{aligned} \tag{6.2}$$

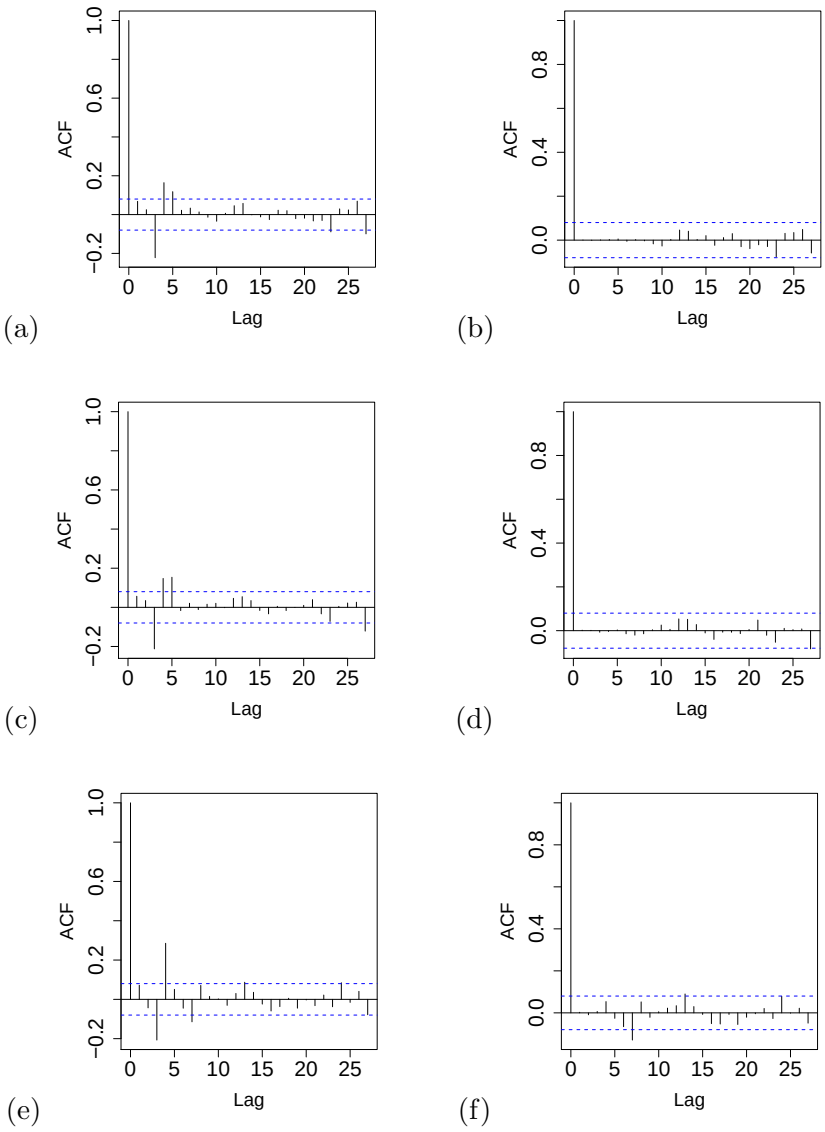


Figure 6.8: Correlogram plots of residuals from ARMA(3,0) (a) Clarence town, (c) Dungog Post office and (e) Blackville; and from ARMA(3,3) models (b) Clarence Town (d) Dungog Post office and (f) Blackville.

Here, SPI(3) at time point t is denoted as S_t and similarly for the other lagged SPI(3) terms. The SOI for five months are averaged and included in the model, as they proved to have a significant effect on SPI(3) ($p = 0.051$) compared to only using SOI from the previous month. This SOI moving average term is denoted as SOI5. Also, the interaction between SOI5 and MEI from lag 1 is significant in this model.

6.2.3 Regression Model for rainfall

The predictions from SPI(3) climatic model are now compared with the climatic model using predictions of one-month ahead rainfall. First, months with zero amount of rainfall is taken to be 0.5mm before the logarithm of the rainfall data is taken, since applying logarithms to zero values would yield infinity. Taking logarithms of rainfall places more emphasis on smaller values of rainfall. To account for seasonality, indicator variables for each month relative to January, are incorporated into the model. That is, if t corresponds to February, then the indicator variable for feb will have a value of 1 and the rest of the indicator variables, $mar, \ldots, dec$ will be zeros.

The predicted logarithm of rainfall is then multiplied by an adjustment factor, c, to give the forecast expected value of rainfall. The adjustment factor is

$$c = \frac{\bar{y}}{mean[\exp(\ln \hat{y}_t)]}$$

where $\ln \hat{y}_t$ is the predicted logarithm of the rainfall and $\bar{y}_t$ is the mean of the actual rainfall. Thus, $\hat{y}_t = c\exp(\ln \hat{y}_t)$. Then $S_{t+1|t}(3)$ is calculated from the predicted $\hat{y}_{t+1}$ and the known y_t and y_{t-1}. The chosen model is as follows:

$$\begin{aligned} \ln(\hat{y}_{t+1}) = \beta_0 + \beta_1 \ln(y_t) + \beta_2 SOI5_t + \beta_3 MEI_t \\ + \beta_4(SOI5_t \times MEI_t) + \beta_5 feb + \cdots + \beta_{15} dec \end{aligned} \tag{6.3}$$

6.2.4 Summary of results

Having analyzed the above stochastic models for one-month ahead predictions of SPI(3), the Root Mean Squared Error (RMSE) is calculated between the observed and predicted SPI(3) for all stations. Furthermore, to quantify the effectiveness of the stochastic models as a drought prediction tool, the RMSE is calculated between the observed and predicted SPI(3), for SPI(3) observations falling below -1, classified as the drought criterion. This RMSE is abbreviated as RMSE(DC). The results for all three rainfall gauges are consistent and Tables 6.5, 6.6 and 6.7 provide the RMSE results from Clarence Town, Dungog Post Office and Blackville respectively.

In general, the ARMA(3,0) model with climatic indicators performs slightly better than the ARMA(3,0) model, emphasizing the influence of climatic indicators. However, the ARMA(3,3) model out-performs the autoregressive model ARMA(3,0) with climatic

indicators, which is in agreement with the partial correlograms and correlograms in Figure 6.7. However, the climatic rainfall model performs better than the ARMA(3,3) model, and has more ability to anticipate a potential drought than the other stochastic models.

For a graphical comparison, all four stochastic methods applied to Clarence Town are shown in Figure 6.9. The improvement in the ARMA(3,0) model when climatic indicators are included is visible from the predictions, although there is still a one-month lagging in predictions. The ARMA(3,3) model further improves the predictions and is capable of identifying sudden drops in SPI(3) and the lagging of one-month appears to be less obvious. Predictions from the rainfall regression model follows fairly close to the observed SPI(3), although it tends to underestimate the severity of the drought for some predictions.

Table 6.5: Clarence Town: RMSE and RMSE associated with drought criterion, for stochastic models

Stochastic model	RMSE	RMSE when observed SPI(3) < -1
ARMA(3,0)	0.56	0.97
ARMA(3,3)	0.53	0.89
ARMA(3,0) and climatic indicators	0.55	0.91
Rainfall regression and climatic indicators	0.48	0.61

Table 6.6: Dungog Post Office: RMSE and RMSE associated with drought criterion, for stochastic models

Stochastic model	RMSE	RMSE when observed SPI(3) < -1
ARMA(3,0)	0.58	0.93
ARMA(3,3)	0.56	0.88
ARMA(3,0) and climatic indicators	0.57	0.88
Rainfall regression	0.50	0.64

Table 6.7: Blackville: RMSE and RMSE associated with drought criterion, for stochastic models

Stochastic model	RMSE	RMSE when observed SPI(3) < -1
ARMA(3,0)	0.71	1.19
ARMA(3,3)	0.66	1.01
ARMA(3,0) and climatic indicators	0.71	1.14
Rainfall regression	0.54	0.71

6.3 Forecasting several months ahead

Longer-term forecasting is essential for water management and planning amongst the agricultural community and forecasts of up to six months ahead can reduce any potential damage to crops brought about by drought. In this section, SPI(12) is utilized and six-month ahead monthly forecasts are made for 1990 to 1999, based on data from Clarence Town.

6.3.1 Autoregressive model

The SPI(12) from months with lags from 6 to 12 are found to be significant in affecting the present SPI(12) value. Hence, the following autoregressive model (AR) is investigated:

$$\hat{S}_{t+6}(12) = \beta_0 + \beta_1 S_t(12) + \beta_2 S_{t-1}(12) + \cdots + \beta_7 S_{t-6}(12) \tag{6.4}$$

Figure 6.10 shows the performance of the predictions against the observed SPI(12). Observe that SPI(12) takes longer escalation above and below the drought threshold of -1, as compared to the fluctuations noticed when using SPI(3). This is attributed to the smoothing effect, which is a result of taking longer moving averages. This model produces a more varied set of forecasts. In some instances, it is able to identify sharp decreases in SPI(12) but fails to recognise the severity of the drought. However, there is still a six-month lag effect.

6.3.2 Autoregressive model with climatic indicators

Figure 6.11 displays the observed and predicted SPI(12) obtained from the regression model. This model is similar to the autoregressive model, but with the addition of SOI and MEI and their interaction term, which has proved to be significant in the model ($p < 0.05$). The regression model is given as follows:

$$\begin{aligned}\hat{S}_{t+6}(12) = \beta_0 + \beta_1 S_t(12) + \beta_2 S_{t-1}(12) + \cdots + \beta_7 S_{t-6}(12) \\ + \beta_8 SOI_t + \beta_9 MEI_t + \beta_1 0(SOI \times MEI)\end{aligned} \tag{6.5}$$

The six-month lagging effect is still present, but the model appears to have more conservative predictions for non-drought periods while placing more emphasis on the

severity of a drought.

6.3.3 Weighted regression model

Weights are applied to the above regression model and each observation is weighted according to their distance from the maximum SPI value, which has the value of 3 in this particular rainfall station. The weight is chosen to be $(S_{t+6}(12) - 3)^2$. This allow for more emphasis or weight to be placed on fitting the large negative values of $S_{t+6}(12)$, which are of crucial importance to farmers.

It is evident from Figure 6.12 that this model produces almost similar forecasts with those in Figure 6.11. The model has a tendency to under-estimate SPI(12) values, which is associated to the weight assigned in the regression model. This model is also able to predict accurately at least two droughts, which makes it a favourable model to predict long-term drought.

6.3.4 Summary of longer term forecasting

Both RMSE and RMSE(DC) are calculated for the above stochastic models. Table 6.8 presents the results for predicting six months ahead. The autoregressive model with climatic indicators performs better than the AR model in terms of predicting the occurrence of a drought, providing further evidence that climatic indicators, SOI and MEI are associated with the occurrence of drought.

Although the weighted regression model with climatic indicators seems to have a higher overall RMSE than the climatic regression model, it is more reliable in predicting an onslaught of a drought given that it has the lowest RMSE when the observed SPI(12) is below -1. There is a trade-off here and since the purpose of this study is to predict the onset of drought, the weighted regression model is more suitable for this purpose.

Table 6.8: Comparison of SPI(12) models in Clarence Town for six months ahead

Stochastic model	RMSE	RMSE when observed SPI(12) < -1
Autoregressive	0.70	0.98
Autoregressive and climatic indicators	0.68	0.76
Weighted regression	0.75	0.55

Rainfall stochastic models were also analyzed to compare their forecasting performance to the regression model in Equation 6.5. The rainfall predictions were then transformed to the corresponding SPI(12) values and the RMSE obtained is 0.80 and the RMSE(DC) is 1.15. To account for the seasonality in rainfall, indicator variables for each month were created in the second rainfall model. There was some improvement to the RMSE(DC) (0.97), however, the RMSE for all the predictions performed worst (1.15).

6.4 Summary of chapter

For the one month ahead SPI(3) forecasts, the strategy of predicting rainfall one month ahead conditioned on past rainfall, SOI and MEI, and combining this prediction with known rainfall in the current and previous months, was the best at all three stations.

The climatic indicators were statistically significant but the practical improvement in forecasts was slight. Longer-term forecasts of drought was investigated using SPI(12) for Clarence Town rainfall gauge. Prediction of SPI(12) at a lead time of six months was substantially improved by inclusion of climatic indices and the forecasting of drought conditions was further improved with the use of weighted regression, which yield a RMSE of almost half that obtained from the autoregressive model when there was an observed drought. The influence of climatic indicators, such as SOI and SST are investigated further in detail in the following chapter, where monthly rainfall from pixels across Australia are used.

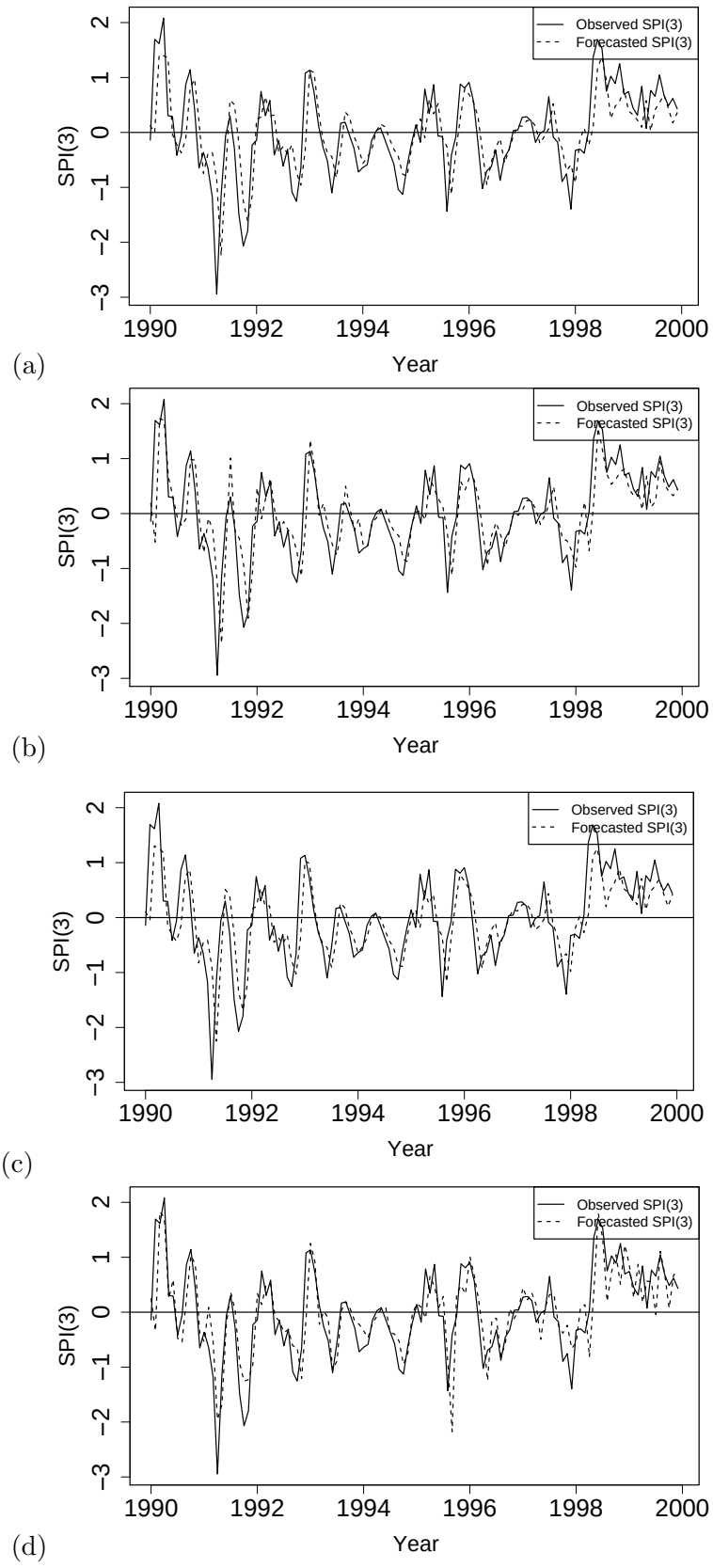


Figure 6.9: Time series plot of observed SPI(3) (solid line) and predicted SPI(3) (dotted line) for Clarence Town using: (a) ARMA(3,0), (b) ARMA(3,3), (c) ARMA(3,0) and climatic indicators and (d) rainfall regression.

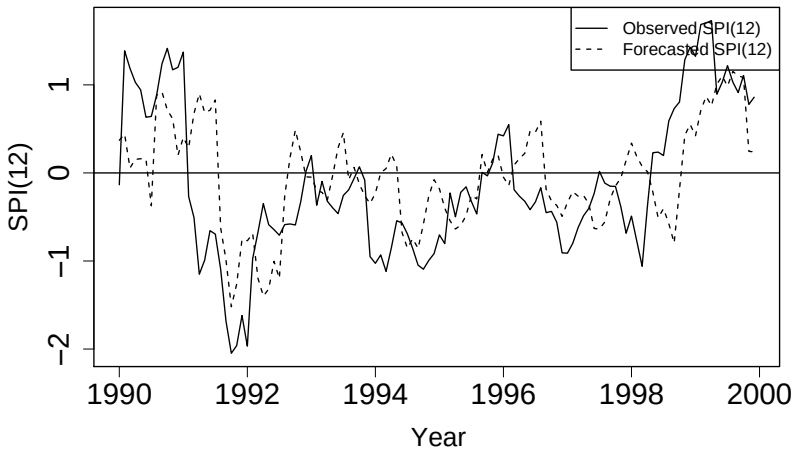


Figure 6.10: AR model: Time series plot of observed SPI(12) (solid line) and predicted SPI(12) (dotted line) for Clarence Town, 1990 - 1999

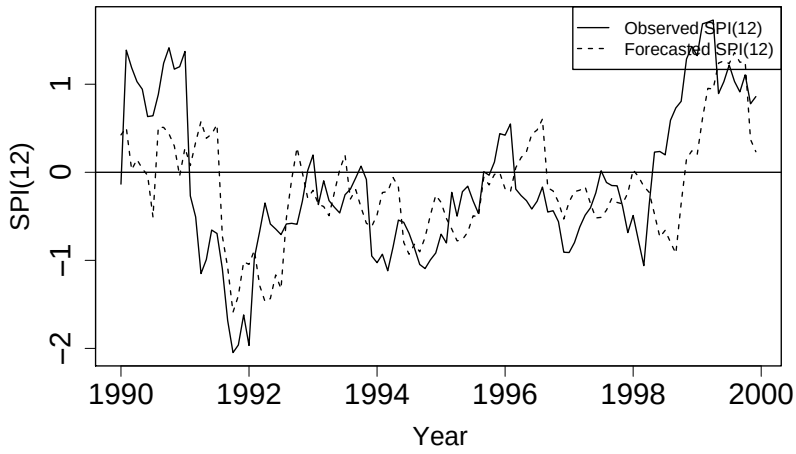


Figure 6.11: Regression model: Time series plot of observed SPI(12) (solid line) and predicted SPI(12) (dotted line) for Clarence Town, 1990 - 1999

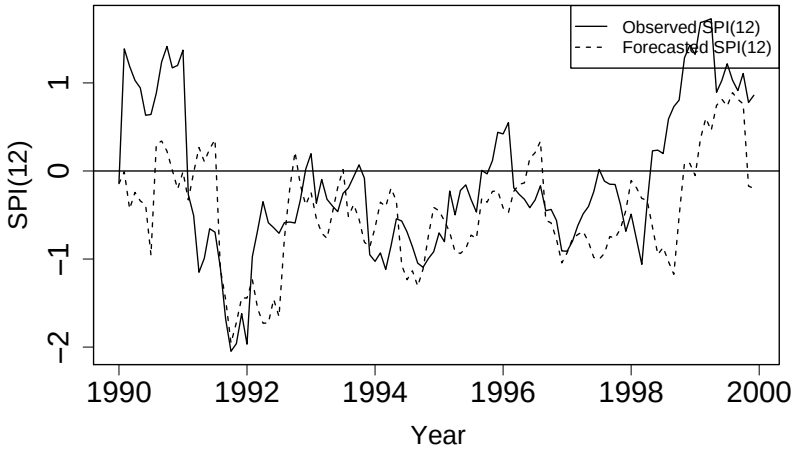


Figure 6.12: Weighted regression model: Time series plot of observed SPI(12) (solid line) and predicted SPI(12) (dotted line) for Clarence Town, 1990 - 1999