A Spatial and Temporal Analysis of Australian Climate Fields

Jennifer L. Kesteven

Centre for Resource and Environmental Studies

November, 1998

A Thesis submitted for the degree of Doctor of Philosophy of the Australian National University

Declaration

This thesis represents independent and original research except as indicated below and in the acknowledgements.

The ANUSPLIN suite of programs was written by Dr. Hutchinson and the brief descriptions in Appendix B are from the programs' documentation. The Australian DEM and coastline were also provided by Dr. Hutchinson.

The circular GIF viewer program was written by Jim Throssell and developed in conjunction with him.

A very brief presentation of the data homogeneity and a short description of the space time models were presented in a paper given at the International Geographic Systems and environmental modelling conference held in Santa Fe, New Mexico, January 1996.

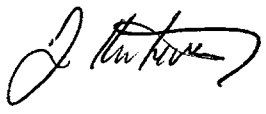
Parts of the section on the Australian climate and some of the maps were used in:

Laughlin G. 1997. *The user's guide to the Australia coast*, New Holland, Sydney.

The surfaces were briefly described and proposed in:

Larson J.W., Taylor J.A., Kesteven J.L, Hutchinson M.F, Bates G.T. and Oglesby R.J. 1994. Regional-scale climate studies of Australia using RegCM2, a paper presented to the International Congress on Modelling and Simulation, Newcastle, November 27-30, 1995.

A variation of the music files was exhibited at the Canberra School of Art, 18-29 August 1997 in "Monga, An Exhibition of Field Work in Monga State Forest.



(Jennifer L. Kesteven)

Acknowledgements

I would like to thank all my supervisors for their encouragement and the interest they have shown in this work.

I have been lucky to have had Michael Hutchinson as a supervisor, he has given freely of his time and invaluable advice. This thesis would not have been possible without him and his interpolation programs. He has been a great friend and the best of supervisors.

Janette Lindesay who joined the supervisory board at a difficult time and fitted in very easily. I would like to thank her for her climatological advice and her friendship.

John Taylor, especially for his interest and encouragement to investigate. He also provided the bulk of the hardware used in the thesis and gave me an opportunity to run some global climate models.

I would like to thank Ken Johnston and Chris Trevitt all those cups of coffee and support in the early stages of my thesis.

There are many people at CRES I would like to thank for providing an interesting and stimulating environment within which this research was undertaken. Henry Nix, firstly, as director of CRES for making available the Centre's resources, and for creating a great environment which is supportive and stimulating. Mark Greenaway for maintaining the computer system, which was essential for the successful output of the thesis, and with help when things go wrong. Rob McArthur provided help in the initial computer learning stage and set me onto the PERL programming language. I would like to thank all of the support staff. I would particularly like to thank Helen, John Mulquiney and John Gallant for the fun and the intellectual stimulus and George for the sustenance.

Jim Throssell for providing the circular viewer presented on the CD-ROM, and for his thought-provoking visits.

Finally, I would like to thank my family and friends for their constant encouragement and support throughout this course of study

This work would not have been possible without the assistance of an Australian Postgraduate Research Award.

Abstract

This study is an exploration of the methods of spatial and temporal analysis and extension, focused on an investigation of Australian climate fields. The analysis is presented as a progression from the one-dimensional temporal analysis of site data, to the investigation of the spline spatial interpolation method, to higher dimensional spatio-temporal modelling. This study also forms the foundation for further spatio-temporal modelling of other variables, particularly rainfall.

In this thesis, mathematical models of climatic variables are devised for the purpose of producing longer-term forecasts. The models constructed predict the spatial distribution of climatic variables for the Australian continent. An appreciation of the space-time relationships of climatic variables is developed through a series of models that use different methods of approaching these relationships.

Table of Contents

List of Tables

List of Figures

List of abbreviations and acronyms used

ACT	Australian Capital Territory
AVG	monthly average
BoM	Bureau of Meteorology
°C	degrees Celsius
CD	computer disk
DALR	dry adiabatic lapse rate
DEM	Digital Elevation Model
DST	daylight saving time
ELR	environmental lapse rate
ENSO	El Nino Southern Oscillation
ERIN	Environment Resource Information Network
GCM	Global Climate Model
GCV	generalised cross validation
GIS	Geographic Information System
hPa	hectopascals
ITCZ	Intertropical Convergence Zone
K	Kelvin
km	kilometre
LST	local standard time
LTA	Long-term monthly averages
MIDI	Musical Instrument Digital Interface
mm	millimetre
MSE	mean square error
MSL	mean sea-level
MSR	mean square residual
RHO	smoothing parameter
rms	root mean square (residuals)
RTGCV	square root of the generalised cross validation
RTMSE	square root of the mean square error
RTMSR	square root of the mean square residual
RTVAR	square root of the error variance estimate
SALR	saturated adiabatic lapse rate
SFLTA	surface long term average
SOI	Southern Oscillation Index
SST	sea surface temperature
T	dry-bulb temperature
T_d	Dew-point temperature
T_{min}	minimum temperature
TMSE	true mean square error
T_w	wet-bulb temperature
UTC	Coordinated Universal Time or Universel Temps Coordonné
VAR	error variance estimate

Chapter Title Page Images

Chapter 1: Rainbow, near Braidwood, NSW

Chapter 2: Dawn, Meringo, NSW

Chapter 3: Fresh snow fall, Little Thredbo Homestead, NSW

Chapter 4: Thredbo, NSW

Chapter 5: Storm, South Coast, NSW

Chapter 6: Storm, Canberra, ACT

Chapter 7: Sunset, Lismore, NSW

Chapter 8: Sunset, Balranald, NSW

All Photographs taken by Jennifer Kesteven.

Chapter 1. An Analytical Framework for the Spatial and Temporal Extension of Climate Site Data

Introduction

Climate plays a crucial role in determining earth surface processes. It exerts fundamental controls on the distribution of natural vegetation, on soils and on the activities of people. As the world's population increases, the demand for food, water and energy steadily increases, while the ability to meet those needs will remain subject to the vagaries of climate. The world is becoming increasingly dependent on the stability of the present seemingly 'normal' climate as it approaches full utilization of the water, land and air which supply the food and receive the wastes. Managing this climate dependency requires coordinated management of the world's resources and a thorough knowledge of the behaviour of climate.

Climate reflects the long-term patterns of weather and thus meteorological observations are fundamental to the development of an understanding of climate and climate change. It is essential that the relevant data be collected on a long-term basis in order to acquire the necessary statistics of climate. Due to the need for improved weather forecasts, the Australian network of meteorological observation stations has grown extensively since the production of the first weather map in 1877 (Colls and Whitaker 1990). Thousands of pieces of meteorological data are collected around Australia on a daily basis; however, they are not available for all times for every location across the continent. These data, then, must be extended in both space and time to adequately document the climatic events that have occurred in the past, and to monitor the climate processes presently occurring.

This study is an exploration of the methods of spatial and temporal analysis and extension, focused on an investigation of Australian climate fields. The analysis is presented as a progression from the one-dimensional temporal analysis of site data, to the investigation of the thin plate smoothing spline spatial interpolation method, to higher dimensional spatio-temporal modelling. This study also forms the foundation for further spatio-temporal modelling of other variables, particularly rainfall.

The broad objectives of the study are:

- the investigation of the characteristics of point data;
- the development of an Australian station climate database developed for specific UTC times;
- the extension of the station climate so that it is spatially and temporally complete;
- to contribute to the development of techniques for using the thin plate smoothing spline for the spatial extension, interpolation and prediction of climatic variables;
- the investigation of mapping techniques used to describe spatio-temporal databases and an assessment of climatic variability and change;
- the development of models which attempt spatial-temporal prediction, by applying time series models to point data, and spatially extending the model parameters; and

- the spatial-temporal prediction, by integration of the spatial and temporal dimensions, of the climatic variables. This involves the development of a space-time model which deals with the space and time dimensions simultaneously.

The management and manipulation of megabytes of data have been essential to this study. Computer-based methods facilitate data analysis and spatial extension, and provide images which can enhance our capacity to view and interpret the climate information. Given the importance of high quality, spatially homogeneous data to the study of climatic variation and climate change, the work reported in this thesis aims to provide a spatially homogeneous Australian climatic database and to examine the temporal variability in that database. The climatic database developed here contains more than 20,000 graphics, animations of all the spatial time series, and the continental anomaly time series transformed and presented as an audible signal which can be heard in conjunction with the animated time series. Due to the practicalities of producing a thesis it is not possible to include these in the hard copy. There is, however, a computer disk (CD) attached to the thesis which includes the climatic database described here (see Appendix A). The CD also includes a version of the thesis which is linked interactively to the graphics. The use of the CD version is encouraged as it provides additional dimensions (visual and aural) to the material documented in the thesis.

Overview

All data have locations in time and space. Sometimes, when doing analysis on these data, it is necessary to take these locations explicitly into account. In few subjects can this be more true than in meteorology and climatology, as most variables in the atmospheric sciences can be viewed as spatio-temporal processes (e.g. sets of pressure and temperature measurements from a number of sites over a period of time). Space-time statistical models are an attempt to extend the historical data in time, and to interpolate across space. Estimation of climatic variables at unsampled times or unsampled spatial locations requires the extension of existing interpolation techniques into the space-time domain. The problem is simplified if the time component is integrated out; the study then is reduced to the problem of interpolation and prediction in space only. Similarly the spatial dependence can be removed, allowing the application of time series methods.

As Ripley (1984) points out, advances in data logging are revolutionising not only areas of the natural sciences such as meteorology, but also the statistical methods which serve them. Meteorologists have available, particularly in recent years, volumes of data which are truly enormous in comparison with many other areas in which statistical methods are applied. Data are routinely collected on temperature, pressure, humidity and precipitation, wind, and other weather conditions. These data are collected by a variety of fixed or transient systems, including weather stations, ships, aircraft, buoys, balloons, radar and, most recently, satellites.

In developing methods for the spatial and temporal determination of climatic variables, this study recognises the importance of the availability of data stations with sufficiently long records so that spatial interpolation errors are not large. The range of data available directly affects the choice of spatial and temporal scales. The ground-based network is the only source of daily surface weather data with extensive temporal coverage and near-complete continental land coverage. This data set, however, contains both spatial and temporal discontinuities which are discussed in Chapters 2 and 3.

It has been suggested that monthly space-time models, which may be reasonably well determined from standard meteorological networks, are sufficient to resolve much ecological and hydrological behaviour, at relatively fine spatial resolutions of a few kilometres (Hutchinson 1995a and 1995b). To achieve the objectives of the present study an integrated analysis of monthly pressure and temperature fields across Australia from January 1952 to December 1990 has been undertaken.

Weather is the state of the atmosphere at a particular point in time over a given region, while climate is a synthesis or generalisation of the weather observed over a longer time period. Meteorology and climatology represent the study of weather and climate respectively(Sturman and Tapper 1996). The different time scales have given rise to different methods used for spatial and temporal extension of the data. One method of spatial extension of meteorological data is to use purely subjective techniques. This method is still used by many practicing meteorologists, especially for pressure data. The data are reduced to mean sea-level (MSL) values at the station by local table calculation and contours are drawn by hand to represent the MSL pressure field. Daily pressure maps are still drawn in this way and produced alongside computer generated maps. More recently, weather radar has been used to present three-dimensional cross-sections of storms, and remotely sensed data from satellites are used to map cloud patterns and cloud-top temperatures. Satellite data will become increasingly valuable as new methods of storing and analyzing large volumes of data are developed.

A variety of techniques have been developed to address the spatial extension of climatic variables. With the availability of improved computing power, a new generation of powerful methodologies is available. Common to all methods is the notion that climatic variables vary in some systematic and hence predictable manner across space. Interpolation in space (normally without reference to time) is a major application of 'splines' and 'kriging' statistical tools, developed and popularised in hydrology, geology and agrometeorology. These tools involve explicit modelling of the spatial autocorrelation of the data. The different techniques of doing this are investigated, and the use of spline functions is discussed, in Chapter 3. Quantifying the uncertainty associated with the data sets and providing methods to assess the effect of this uncertainty in modelling applications is an important area of investigation.

Studies of space-time variation in atmospheric science overlap with research in hydrology and ecology. They cover a wide range, both methodologically and in the nature of the problems investigated. Many studies in spatio-temporal modelling have concentrated on the spatial domain, by treating the time replications effectively as independent. Others have concentrated on the time domain, by ignoring the spatial interactions.

Neither the exclusively temporal nor the spatial approach are entirely satisfactory. Geostatistics offers a variety of methods to model data processes as realisations of random functions. These procedures have been applied primarily to spatial data. Applying such space-orientated approaches to spatio-temporal processes, however, may lead to the loss of valuable information in the time dimension.

Although apparently straightforward, this solution presents both theoretical and practical problems as there are major differences between temporal and spatial phenomena. Rouhani and Myers (1990) and Rouhani and Wackernagel (1990) suggest that these practical and theoretical problems are as follows:

1) One-dimensional temporal data are ordered, a characteristic not usually found in two- or three-dimensional spatial data, which do not exhibit such order as past, present and future. Also, space and time are measured on two different scales which cannot be compared in a physical sense.

2) The spatial locations available are often sparsely distributed, whereas the time series associated with each location are relatively long. The accuracies of estimated temporal and spatial structures are likely to be quite different.

3) Space-time data can often exhibit temporal periodicity and spatial non-stationarity. A variety of temporal periodicities have been observed, such as periodic seasonal cycles, pseudo-periodic climatic cycles, and non-periodic long term trends. The periodic component can be approached as a part of the trend, or the series can be treated as stationary. The spatial non-stationarity may be ignored in some cases; in others, however, it raises serious doubts about the spatial homogeneity assumption.

In response to these problems, a new approach to the modelling of spatio-temporal data is proposed here. The differences in scale between the spatial and temporal data are discussed, and a method to overcome these differences, involving a combination of physical and statistical procedures, is presented. The method involves the transformation of the time scale to a spatially distributed representation of the variable in question, in an attempt to deal with the apparent spatial non-stationarities and temporal periodicities in the meteorological data sets. This study proposes that the space-time covariance structure of climate can be used to develop a space-time model which incorporates time into the spline spatial model. This approach extends the existing spatial techniques by integrating the space and time domains. That is, the proposed solution is to consider the spatio-temporal phenomenon as a realisation of a spline function which incorporates one dimension for time with the remaining dimensions in space (usually two or three).

This chapter continues with, first, an overview of the Australian climate in terms of the variables of interest (pressure, temperature and rainfall), to give an understanding of the variation and the expected patterns of the climate data used for the analysis. This is followed by a discussion of the basis for selected spatial and temporal scales and some of the trade-offs that are required, and an overview of the proposed methodology.

The Australian climate

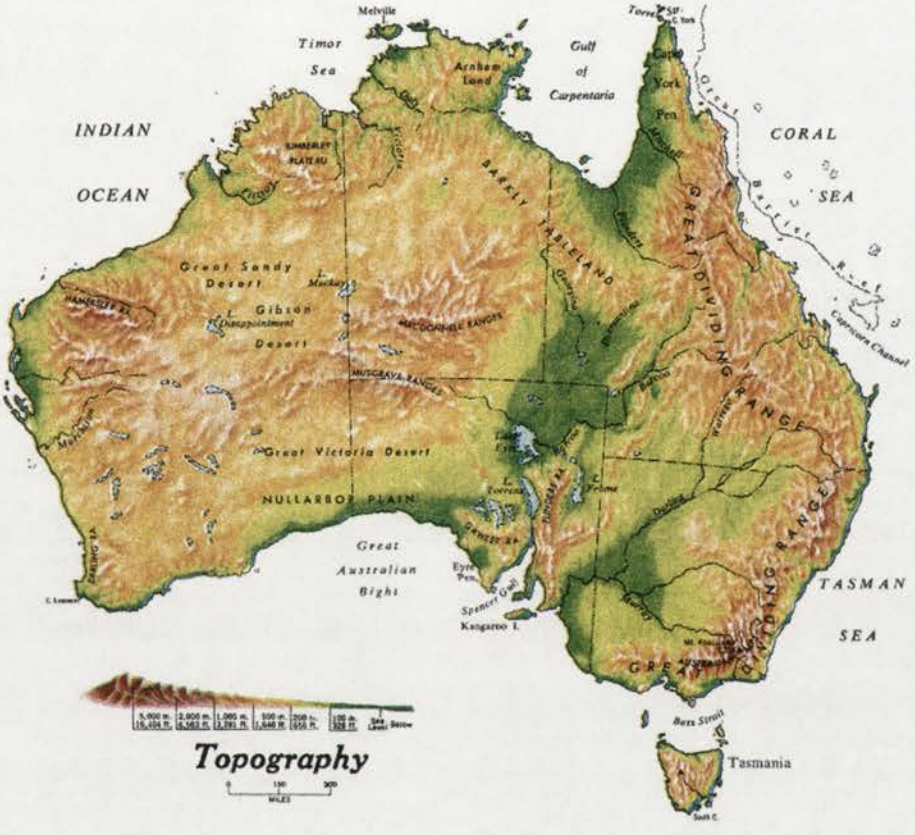


Figure 1.1: Map showing the topography and feature names (source: Hammond World Atlas 1978).

Being a relatively small continent with low relief, surrounded by extensive water-masses (Figure 1.1), Australia does not exert a large influence on the global pattern of pressure and winds, but rather is subject to the winds which are part of the general circulation. It is symptomatic of this predominance of the oceanic influences that the monthly mean pressure at Darwin is taken as an indicator of the Southern Oscillation Index (SOI) (Troup 1967, Allan 1988, Allan *et al.* 1996). However, in the drier latitudes of the southern hemisphere, Australia occupies a larger extent than either of the two other continents. This longitudinal span causes modifications in the air masses that travel over the surface of the land, so that the degree of continentality is greater than in Southern Africa and South America.

The general circulation in the Australian region

Australia spans latitudes from the tropics (10°S) to the lower middle latitudes (45°S) of the southern hemisphere. Much of the continent falls under the influence of the descending air and divergence which is characteristic of the belt of subtropical high pressure. In general this zone comprises vast areas of light winds and gently subsiding air. Rain-bearing clouds are relatively few and the weather is usually fine. It is not surprising that the great deserts of the world are found in this zone.

In the winter months the subtropical high pressure belt is narrower than it is in summer and is broken into cells. The belt lies across the Australian continent around 29°S to 32°S, and a westerly wind regime covers the southern parts of the land mass where fronts embedded in this flow regularly bring periods of unsettled weather. To the south of the high pressure ridge a belt of low pressure characterised by strong westerly winds, can affect southern Australia and especially Tasmania. To the north fine weather prevails under the influence of the tropical easterlies.

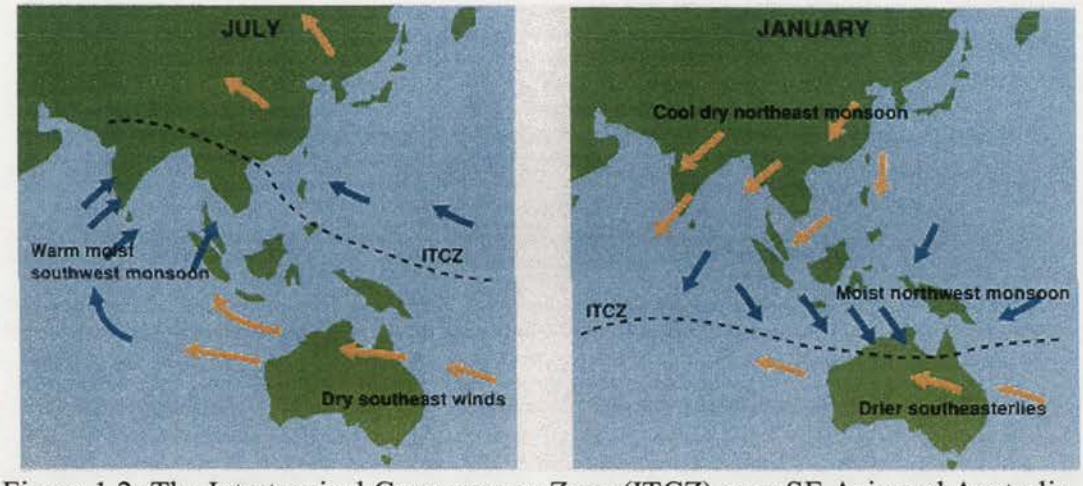


Figure 1.2: The Intertropical Convergence Zone (ITCZ) over SE Asia and Australia in July and January (source: Crowder 1995).

In the summer months the subtropical high pressure belt moves southward to around 34°S to 40°S, allowing warm moist tropical air to move over the northern parts of the continent. Northern Australia has an alternation of monsoonal northwesterlies and tropical easterlies, central and southern Australia are influenced by the dry variable winds of the high pressure belt, and Tasmania is dominated by the mid-latitude westerlies.

In the tropics of the southern hemisphere the trade winds blow from the southeast. The wet season starts when the ITCZ moves over northern Australia and warm moist air moves in from the northwest, replacing the drier air of the southeasterly winds that prevail in the dry season (see Figure 1.2). The wet season usually extends from December to March.

Between latitudes 40°S and 50°S is a zone of strong westerly winds, which develop in these latitudes as the oceans extend around the globe and winds are practically unimpeded by land. Southern Australia is usually on the northern margin of the westerlies and their influence is particularly noted in winter and spring when storms in the belt of extra-tropical cyclones are further north than usual. Western Tasmania is particularly affected by the westerlies which result in windy wet winters.

In the upper troposphere (around 200 hPa) the subtropical jet stream meanders over the continent throughout the year. There are intimate links between high-level flow and surface synoptic disturbances. The jet stream plays a role in creating and steering synoptic disturbances, and it contributes to the maintenance of the general circulation (Reiter 1961). In the Australian region the subtropical jet stream is more prominent than elsewhere, and is a feature over the continent for the entire year. Brook (1982) found the subtropical jet stream a consistent and persistent feature of the atmospheric circulation in the Australian region. Its existence is a direct result of the mechanism responsible for meridional transfer of heat.

There are marked seasonal variations in the subtropical jet stream. It is weaker in summer than in winter. There are also spatial changes in preferred jet-stream location both across the continent and between seasons (Gentilli 1971, Brook 1982). Both the seasonal and latitudinal features and the year to year variations of the jet stream are reflected in the behaviour of the depressions and anticyclones of the Australian region. The pressure trough between successive anticyclones is associated with an equatorward meander of the subtropical jet stream and its subsequent acceleration. Its main effect in winter is the formation of upper-air depressions. The magnitude and latitude of an anticyclone are also closely related to the speed and latitude of the subtropical jet stream (Reiter 1967, Gentilli 1971).

Significant deviations from the mean flow patterns occur as a function of longitude in both summer and winter. Apart from the major seasonal changes, some periods are found to be marked by strictly zonal flow with a single jet stream, while others have pronounced wave patterns or split jet configurations, with a single strong jet in the west and a double jet further east. The transition from the single to the double jet tends to occur at a preferred site for the formation of slow-moving intense 'blocking' anticyclones and of mid-tropospheric low-latitude depressions which play a large role in the rainfall regime of the Tasman Sea area (Brook 1982).

The southern hemisphere blocking systems are not well defined in monthly mean circulation fields, since they are generally less persistent or of smaller amplitude than in the northern hemisphere, due to the lack of major land masses in the southern mid-latitude regions (Coughlan 1983). They are, nevertheless, sufficient to deflect propagating wave disturbances towards or away from the southern parts of Australia. Lejenas (1984) and Kidson (1988, 1991) found the maximum frequency of blocking in the Australia-New Zealand region occurred in April and June-August. Blocking is also associated with El Nino Southern Oscillation (ENSO) events.

Winds in the middle troposphere (at the 500 hPa level) and above are the counterparts of the surface winds, balancing equatorward surface flows by poleward winds, and vice versa. In the southern hemisphere the mean tropospheric flow departs little from a purely westerly direction, although there is a particularly strong westerly flow at 30°S to 45°S. The westerlies are stronger and more extensive in the southern than the northern hemisphere, due to fewer land obstructions. They are also stronger in winter than in summer as the temperature gradient between the equator and the poles increases. On any given day, however, the circulation pattern may display many small-scale weather disturbances travelling in an easterly direction.

Surface pressure patterns and wind in the Australasian region

As previously discussed, upper-level features play a major role in the creation and steering of the surface synoptic systems that produce the regional weather that is aggregated to provide the climate. Before looking at the role of these surface systems there is a need to understand that the origin of the air masses, which the surface pressure differences transport, which will have a predominant effect on the weather. Australia is affected by most of the major types of air masses: cold subpolar maritime air originating over the Southern Ocean, hot dry subtropical continental air originating over the Australian continent, and warm moist Pacific and Indian subtropical maritime air originating over the Pacific and Indian oceans. Figure 1.3 shows that the air masses have different limits of penetration into the continent depending on the time of year (Johnson 1992).

The representation of the surface circulation for long periods (months, seasons, etc.) by charts of anticyclonicity and cyclonicity was developed during the early 1950s, to help solve the problem of extended long-range forecasting in Australia. Charts of seven-year averages (1946-1952) of monthly seasonal anticyclonicity and cyclonicity and their description were published by Karelsky (1954). Updated charts for the 15 years 1946-1960 were later published by Karelsky (1961) and also for the period 1952-1963 (Karelsky 1965). Charts of 23 year averages (1965-1987) of monthly anticyclonicity and cyclonicity have been published by Leighton and Deslandes (1991a, 1991b).

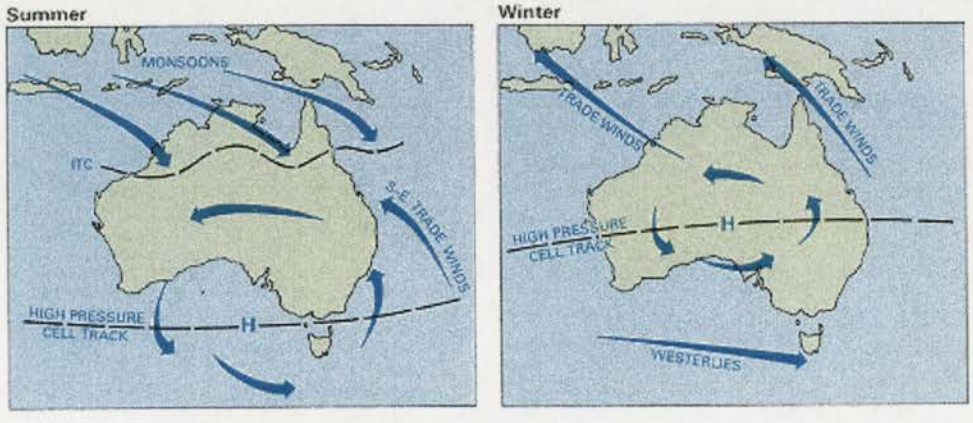


Figure 1.3: Air masses of Australia (source: Johnson 1992).

At the surface the main atmospheric features of the Australian region which influence the weather are the high pressure systems or areas of divergence which tend to bring settled weather, especially to areas under the influence of the center of the system. Low pressure systems or areas of convergence to the south of the continent, with their associated fronts and relatively colder air masses, tend to bring most of south-east Australia's winter weather.

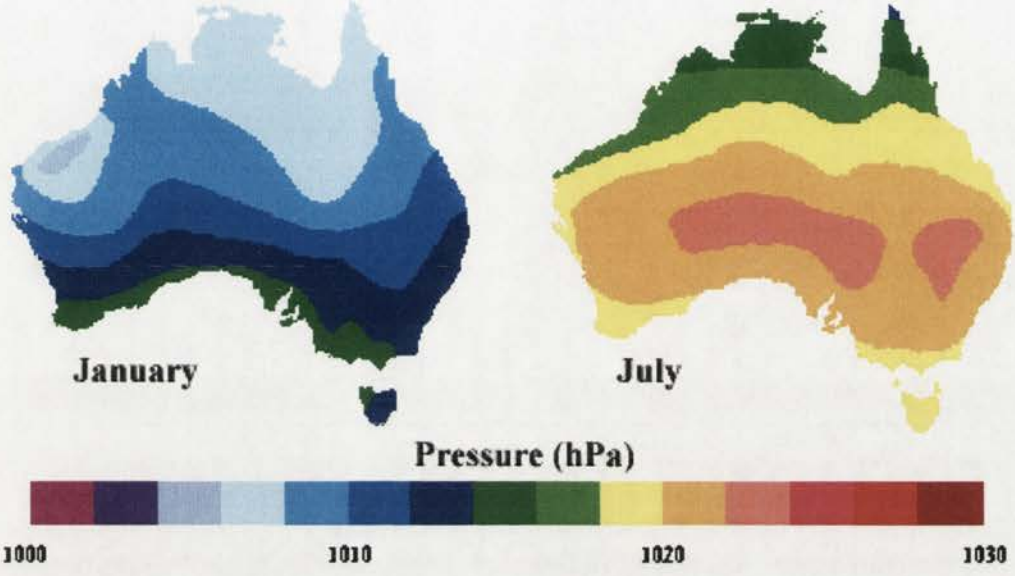


Figure 1.4: Mean pressure in January and July, 0900 local standard time (LST) (adapted from Gentilli 1971).

In winter, the cooling of the land mass causes an increase in the density of the overlying air, and a consequent alteration in the barometric pattern. Figure 1.4 is adapted from Gentilli (1971) and shows the mean January and July surface pressure fields. The high-pressure belt moves northwards with the seasonal change, with a majority of the anticyclone tracks being between 29°S and 32°S (Karelsky 1954, 1961 and 1965, Gentilli 1971 and Streten 1982). There is, however, latitudinal variation; usually the track is at its furthest north over most of Australia in June-July, but, due to the greater lag in ocean temperatures off the east coast, the northward deflection can be retarded there until August-September.

During the winter the proximity of the mid-latitude depressions to the Australian mainland increases very rapidly to its July maximum (Karelsky 1965) with cyclonic centres passing near Tasmania. Any large depression well to the south of Australia produces a strong southerly flow, which may be as cold as 2°-3°C when it reaches the southern shore in winter.

Deep and complex lows south of the continent maintain a westerly airflow. However, if these depressions move north the rapid passage of cold fronts gives surges of cold southerly air, and, if upper-level moisture is high, general precipitation results. As the trough moves eastward and the next anticyclone becomes dominant, winds abate, skies clear, nocturnal radiation increases and widespread frosts result. Low pressure

systems moving north parallel to the east coast can result in gales and strong south-easterly winds along the east coast.

Temperatures on the Australian continent

There is a great diversity of surface air temperature ranges across Australia due to the size and location of the continent. This diversity (Figure 1.5) stems from the effects of latitude, altitude and the degree of continentality, and ranges from humid tropical heat in the north to the dry heat of the centre and to alpine winters in the highlands of the south-east. The Australian Alps can experience subzero blizzard conditions in the winter, which contrasts with the occasional +50°C temperatures over the inland during summer. The lowest recorded minimum temperature was -23°C at Charlotte Pass, New South Wales in 1992, and the highest recorded maximum shade temperature of 53.1°C occurred at Cloncurry, Queensland in 1989.

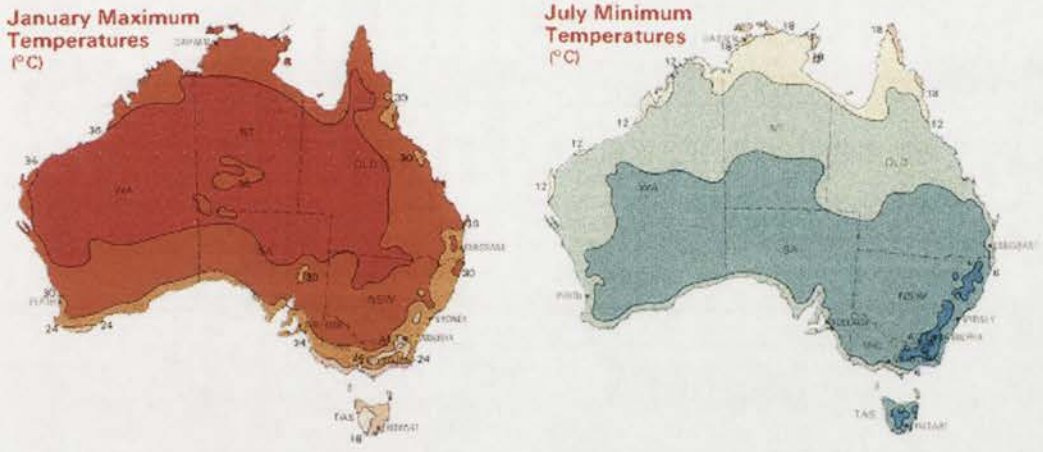


Figure 1.5: Australian January maximum and July minimum air temperatures over Australia (source: Johnson 1992).

When considering the continent as a whole, Australia may be considered the hottest continental mass despite higher temperatures occurring in both parts of Asia and Africa. The hottest time of the year is progressively later from north (November) to south (February), corresponding in a general way with the time of most effective insolation (Gentilli 1971).

Rainfall

Australia is a dry continent with more than half of its land area considered desert. Less than 30% of the continent receives sufficient rainfall to support general agriculture (Colls and Whitaker 1990). The relatively narrow coastal strip receives most of the rainfall and the dry interior is mainly due to the subsiding air of the subtropical high pressure belt.

When the subtropical high pressure belt moves northward in winter, dry southeasterly winds prevail over northern Australia, and southern Australia comes under the influence of the winter westerlies. Cold fronts associated with depressions to the south of the continent move eastwards across southern Australia and are often accompanied by strong winds, heavy rain showers and snow in the alpine areas. Areas exposed to the westerlies, such as the southwest corner of Western Australia, southwestern Victoria, and particularly western Tasmania, have heavy winter rainfalls, but most of the southeast part of the continent is more sheltered and rainfall more evenly distributed around the year.

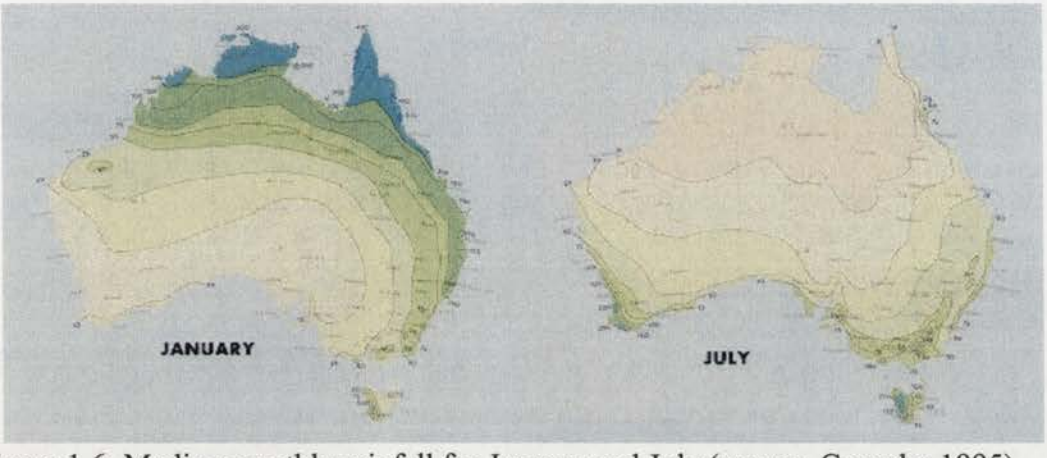


Figure 1.6: Median monthly rainfall for January and July (source: Crowder 1995).

When the subtropical high pressure belt moves over southern Australia in summer, the ITCZ and summer monsoon move in over northern Australia (see Figure 1.2). Moist northerly winds bring showers, thunderstorms and the threat of tropical cyclones to northern Australia.

Precipitation patterns are always complex in mountainous areas and their nature varying enormously from one mountainous region to another depending on the characteristics of the region (including height, orientation and location) and its interaction with the prevailing winds and storms that occur in the region. Australia is relatively flat and its mountains are small compared with those of all the other continents. Nevertheless, the Great Dividing Range and the escarpments which extend along the eastern seaboard, and the Alps of the southeast, are major factors in determining the climate of the continent and the distribution of water resources. The higher rainfall along the east coast of Australia is primarily due to the inflow of moist air from the ocean and the orographic influence of the Great Dividing Range. The heaviest rainfalls occur where moisture-laden winds are forced to flow up slopes of mountain ranges, and very arid areas can occur in the lee of mountains. The heaviest falls do not always occur at the tops of the highest mountains, as much of the moisture carried by winds will have fallen as rain on the lower slopes, and the air at higher altitudes can hold less moisture. The ranges also enhance the very heavy rain from tropical cyclones, particularly when these storms move slowly over the ocean close to the coast. The heavy rainfalls in the coastal region of far north Queensland

are the result of the southeast trade winds gathering moisture as they move over the Pacific Ocean and releasing large volumes of water as they strike the coast and move up the slopes of the ranges. Very heavy rainfalls also occur in winter in the mountains of western Tasmania as this region is exposed to the westerlies. Hutchinson and Bischof (1983) and Hutchinson (1995a) have shown the importance of incorporating vertically exaggerated topography in rainfall interpolation and analyses.

Space and time scales

The word 'scale' is used in many contexts and often connotes different aspects of space and time. Scale is the spatial or temporal dimension of an object or process. It should not be confused with cartographic scale which is the degree of spatial reduction indicating the length used to represent a larger unit of measure. The monthly time scale and a spatial resolution of ten kilometres for the Australian continent have been selected as the temporal and spatial scales for this study.

Climates, ecosystems, and societies interact over a wide range of temporal and spatial scales. For temporal scales of a year or shorter, the space-time structure of atmospheric phenomena has been reasonably well studied. There is a clear link between space and time scales and the nature of processes that dominate across the space and time spectrum (Orlanski 1975). These are summarised in Figure 1.7, which shows the range of scales and typical phenomena observed in the atmosphere. A consideration of the spatial scales of some important climatic phenomena suggests that a grid size on the order of 10 square kilometres and the use of the monthly time step would allow an adequate representation of key meso-scale processes that are related to synoptic scale meteorological events. Many other processes operate on finer spatial scales, but data limitations can preclude continental modelling at such scales.

The tremendous burden of sampling spatial variables adequately often means that existing data sources must be used. Sometimes available data determine research designs and space-time scales. This is especially true for broad-scale problems. Mathematical and statistical considerations may affect the selection of scales. Data handling thresholds are intertwined with time and space scales. This data-handling threshold has been moved to higher time-space resolution by technology. However, time and financial considerations constrain spatial scales, the number of variables considered and the time scales used.

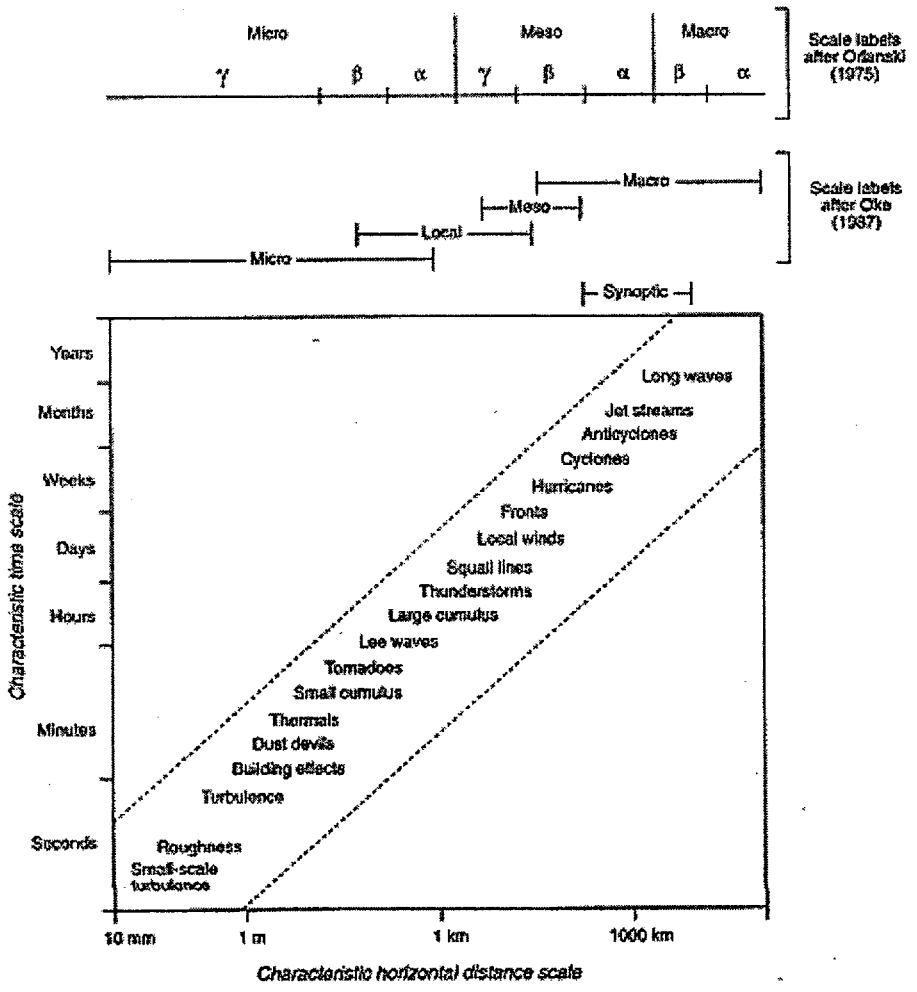


Figure 1.7: Characteristic time and space scales, with associated atmospheric phenomena (source: Sturman and Tapper 1996).

The geographic literature is rich in philosophical discussions of spatial scales and methodological solutions for dealing with scale (e.g. Harvey 1969). There is, however, widespread agreement that, with changes in scale, different variables assume different degrees of importance. Moreover the value of a phenomenon at a particular place is usually dependent on causal processes operating at differing scales. The literature of the physical climatology of the earth's surface also illustrates the pragmatic problem of matching time and space scales (Flohn 1981, Lewis 1995), as well as determining the nature of the variables which are important. The methodological solutions already in existence need not be reinvented to more fully incorporate the spatial dimension in the modelling process and for the extension to broader scales of spatial analysis.

Steyne *et al.* (1981) make the interesting point that disciplines concerned primarily with processes, such as meteorology, are able to switch scales with relative ease (e.g. Gedzelman 1985). On the other hand, disciplines dealing with phenomena are often

restricted to the size of the phenomenon. It should be clear, however, that the addition of the spatial dimension in the study of nearly any process or phenomenon may involve a variety of trade-offs. Models for broader-scale patterns result in reduced predictive accuracy at specific points or places, and the models are only as good as the finest resolution spatial data available.

Meentemeyer (1989) claims that researchers with similar propensities select similar scales and seem therefore to group together, and that perhaps this is caused by dominant paradigms, data sources, and other realities. Perhaps a better way to decide on proper scales is their efficacy in applications, that is not just studying weather and climate in isolation but applying it to modelling biodiversity or agricultural distributions. Nevertheless it is worth noting Clark's (1985) warning, that "no simple rules can automatically select the 'proper' scale for attention".

A major constraint on the choice of time step for this study was the computational requirements of an analysis of continental climate data. The spatial distribution of monthly means can be reasonably determined from standard meteorological networks. Monthly surfaces and monthly mean surfaces have often been used in climatology (e.g. Allan and Haylock 1993, Kong 1995) and are used by the Bureau of Meteorology as a method of providing summary statistics (Bureau of Meteorology 1968-1989). Atmospheric variables exhibit temporal variations consisting of both regular periodic components and chaotic or random fluctuations. The former provide a good basis for prediction, while the latter cause real problems for both short- and long-term forecasting.

The monthly time scale incorporates much of the spatial variability of climate and is sufficient to resolve much of the seasonal variation in ecological and hydrological activity and hence the spatial variability of dependent biological activity (Nix 1986, Booth *et al.* 1987, Mackey *et al.* 1989 and Hutchinson *et al.* 1992). The BIOCLIM program has successfully used the spatial distribution of monthly mean climatic variables to predict the spatial distribution of natural plant and animal species (Nix 1986). The monthly time scale has also been used as a basis for agro-climatic classifications (Hutchinson *et al.* 1992).

Effective determination of the spatial distribution of mean climatic variables is also a necessary first step towards the development of stochastic models of the weather for more refined assessments of the impacts of climate (Hutchinson 1987, 1995b). This determination provides a basis for spatially interpolating and simulating actual weather values at monthly, weekly and even daily time steps (Hungerford *et al.* 1989, Hutchinson 1995b). The broader spatial scale of monthly data also goes some way to meeting the broad spatial scales of Global Climate Models (GCMs), and climatological means play a crucial role in validating general circulation models (Gates *et al.* 1992).

While the monthly time step is appropriate for the assessment of the Australian climate, it is important to note that there are limitations to the physical interpretability of this

approach. As measurement time scales are lengthened, the estimate of variability of the phenomena is low because the monthly values vary less than the daily and/or hourly values upon which they are based. As the underlying processes tend to have shorter time scales, erroneous inferences about lower levels from the higher levels of aggregation can be made. For example, many factors are intricately related to the spatial distribution of surface air temperature on a local scale. Hourly air temperature distribution is directly influenced by factors such as solar radiation, precipitation, atmospheric movement on a larger scale, and various sizes of eddies. However, mean air temperature distributions during longer periods, such as the monthly mean, are mainly affected by the factors such as topography, vegetation and soil wetness. This is because spatial variations in the former factors are averaged over longer periods and thus their effects on spatial prediction of the mean air temperature disappear.

Methodology

The methodology applied in this study approaches the task of developing space-time climate models from standard meteorological networks in five steps.

The first step is to produce a standardised database appropriate for use in interpolation of climatic variables on a continental scale. This includes obtaining data to provide an adequate sample of the spatial variation in the climatic variables of interest, and the determination of the criteria for station selection. The derivation of estimates for standard times is necessary for the removal of temporal discontinuities caused by the collection of data in different time zones, and to overcome discontinuities in the climatic record as a result of the introduction of daylight saving time at some stations.

The second step is the investigation of methods of spatial interpolation and an assessment of the performance of the thin plate splines in extrapolating site climatic data (for times and locations for which data are available). This is necessary because an appropriate statistical technique is essential for the production of the climatically realistic MSL surfaces. This is also important in obtaining an understanding of the interpolation process and to assess the options available in this process. It is closely linked to the third step and both of these are required for the fourth and final steps.

The third step is to apply the thin plate smoothing spline technique to spatially interpolate the point data to produce MSL surfaces. The surfaces produced in this step form the 'primitives' which are the basis of the Australian climate atlas on the CD. The 'primitive' surfaces are then used to derive new climatic variables by applying 'logical' models. The techniques available for the investigation of large spatio-temporal data sets are considered and some different techniques suggested.

The fourth step uses the 'primitive' surfaces in the development of spatial regression and lag regression models which operate at a specific monthly time step. This was

considered necessary to test the integrity of the spline function in producing the surfaces and to provide some useful insight into the interpolation of parameter surfaces. The procedure is based on regression techniques applied to spatial data and on the use of the thin plate splines to spatially extend parameter surfaces. These models, which attempt spatial-temporal prediction by applying regression and lag regression models to surface point data, and then spatially extending the model parameters. While these models were developed to test the integrity of the spline function in parameter surface construction, they also provide some useful models which can be used as a base against which the results from the next step can be compared.

The fifth step is the development of the space-time spline model, which models simultaneously the spatial and temporal anomaly correlations of climatic variables. The step is important, as the contemporaneous incorporation of time into the spatial model is not often done. The space-time spline model, while still under development, shows promise for modelling climatic variables at the monthly time scale. It also provides useful insights into the use of spline functions and has led to the formulation of a new approach to the integration of space and time scales.

The approach used in this investigation conveniently isolates different spatial and temporal aspects of climate model development. In particular it investigates the spatial and temporal dimensions separately before an attempt is made to integrate the dimensions. It identifies the climatic patterns existing over the Australian continent for the period from 1952 to 1990, and provides an insight into the spatio-temporal patterns and the modelling of these climatic variables.

Conclusions

This thesis applies the analytical framework as outlined above for the simultaneous spatial and temporal extension of climatic site data to develop models of particular variables and to support the future development of further models, including space-time models of rainfall. The eventual aim in the study is the incorporation of the space and time scales in a space-time spline cube. This work has involved a number of major tasks:

- obtaining meteorological data and the production of a climatic database;
- an assessment of the appropriateness of the spline interpolation method for the spatial extension of climatic variables;
- the development of a spatial database for the Australian mainland, containing gridded estimates of primary climate elements for the period 1952-1990. This includes an assessment of the validity of the surfaces, statistical checks on the homogeneity of the database, and an assessment of climate variability and change;

- the production of a 2.5 x 2.5 degree grid of estimates of climatic variables, which were used to establish the correlation parameters for the models of climatic variables, and the spatial interpolation of those parameters;
- the production of the regression and lag regression models which effectively extend the temporal dimension by interpolating the model parameters in space; and
- the production of the space-time spline model for pressure and dew-point temperature which requires the suitable transformation of the time dimension so that the climatic variable can be modelled simultaneously in two-dimensional space and one-dimensional time.

The remaining chapters report on these major tasks and associated activities. Chapter 2 discusses the temporal and spatial climate database, beginning with a brief discussion of the primary climatic elements and their measurement. This is followed by a discussion of pressure and temperature lapse rates. Management and homogeneity problems associated with large climatic data sets are then considered. The methods used to produce the new pressure and temperature station database and problems associated with the spatial and temporal coverage of the new station data set are then presented. This is followed by a discussion of the rainfall station data and the use of percentiles to represent the rainfall distribution.

Chapter 3 reviews some of the methods of statistical interpolation presently in use for the spatial extension of site climatic data. A full investigation of the spline technique is followed by an introduction to the ANUSPLIN computer programs. The chapter continues with a validation of the package's ability to represent meteorologically realistic daily pressure patterns and an evaluation of its ability to reproduce Bureau of Meteorology rainfall decile maps. An assessment of the options available in the ANUSPLIN package, including the incorporation of elevation in the climate surfaces, then follows.

Chapter 4 begins with an example of fitting a series of Australian pressure surfaces, with particular attention being given to the ability of the spline function to detect data inhomogeneities, and to appraise the adequacy of the meteorological station network. This is followed by a discussion of the errors associated with the climate surfaces and of the output statistics generated in the production of the surfaces. This is succeeded by an assessment of the lapse rates determined from the spline functions. Finally, long-term monthly averages and the problems associated with incomplete time series, and ways of overcoming these problems, are discussed. The chapter concludes with a discussion of fitting climate surfaces to rainfall data and the rainfall percentile surfaces.

Chapter 5 begins with a description of the methods used to produce the surfaces that were interpolated from the station data and which form the base of the Australian spatio-temporal climate database. The surfaces derived from these 'primitive' surfaces and the 'logical' models used to create them are then presented. Discussion

of the mapping techniques used to describe the database starts with a consideration of grid-point data and an analysis of the trends which are apparent in the spatial time series. This is followed by an overview of visual and aural mapping techniques used to describe spatio-temporal data, with some of the different techniques presented.

Chapter 6 describes the application of the surfaces to the development of spatial models at the monthly time step. This chapter begins with a description of the methodology used to calculate the correlation coefficients for a spatial regression model. It describes the process of spatially extending these coefficients and how they have been coupled to the spatial database, to produce the spatial regression model. A similar methodology is used to produce a spatial lag regression model which effectively extends the temporal dimension. Other models are also suggested.

Chapter 7 presents the spatio-temporal spline model. The space-time spline model depicts processes of two-dimensional space (longitude and latitude) that are played out along a third (temporal) dimension. The processes required for transformations in the time dimension are described. Two different representations of the 'time' dimension are used in the spatio-temporal spline model. The scaling of the 'time' dimension was found to be critical and several methods for selecting the appropriate scale are examined. The chapter concludes with a comparison of the spatio-temporal spline model with the 'primitive' surfaces and the lag regression model output.

Chapter 8 concludes the thesis with a summary of findings, comments on future research directions, and a discussion of possible applications of the methodology to climatological and environmental issues.

Chapter 2. The Spatio-temporal Station Climate Database

Introduction

In Chapter 1 an analytical framework was presented for the spatial and temporal extension of meteorological site data. Key elements of this framework are the availability of data and the determination of the criteria for data selection. The Australian Bureau of Meteorology's station three-hourly surface meteorological records and the monthly rainfall records were the foundation of the spatio-temporal climate database. The Australian Bureau of Meteorology (1990b) station dictionary was used for station location information.

The use of these records to form a gridded data set of monthly and long-term climatic means for the Australian continent for modelling purposes required the development of a spatially and temporally homogeneous data set. Inconsistencies in the records, particularly differences in recording times, were corrected. Data storage and maintenance methods were based on the eventual use of the data. Generally the intention was to output the data sequence in both tabular and graphic form. The production of the gridded data set required an intricate level of organization necessary to store and access very large amounts of data.

This chapter discusses the temporal and spatial climate database beginning with a brief introduction to the primary climatic elements and their measurement. This is followed by a discussion of climatic data and homogeneity problems and the methods used to overcome the inconsistencies in the pressure and temperature station database. A description of the methods used to solve the problems associated with the spatial and temporal coverage for the derived station data set is then presented. Finally, a discussion of the station rainfall data and the use of percentiles to represent the rainfall distribution is given.

The climatological elements and their measurement

The combinations of all the climatic elements at a given moment make up the weather, while the climate depends upon factors, such as geographic location, altitude, aspect, and distance from the coast, which practically do not vary over long periods. The climatic elements have been classified as primary or derived. Primary elements are those which are directly observed or estimated and include such things as temperature, pressure, precipitation, wind velocity and wind direction. Derived elements, some of which can also be determined instrumentally, include variability, duration of characteristic average states of the atmosphere, dew-point temperature, lapse rates of temperatures, frequencies, probabilities, and anomalies. This thesis is concerned with three of the primary elements (pressure, temperature and precipitation) and elements derived from them.

Pressure

The air pressure at a particular height is the weight of the column of air above that height. Thus pressure decreases as altitude increases, particularly in the lower atmosphere. A change of pressure at any particular place indicates an alteration in the amount of air above, and a falling barometer reading implies a net outflow of air from the space directly over the place of measurement. A major reason for measuring atmospheric pressure is to help forecast the weather. To determine pressure differences, pressures must be expressed in a comparable fashion, allowing for variations of height, latitude, temperature and the time of measurement. Maps are drawn to show the air pressure at a reference height or horizontal level, the most common being sea level.

Pressure is normally measured with a Kew type mercury barometer at Bureau of Meteorology stations. There is usually a thermometer attached to the barometer so adjustments can be made to compensate for the changes in density of mercury at different temperatures. The earth's gravity is not constant over the earth's surface so corrections to the observations are required. These gravity and latitude corrections are derived from tables. The barometers can measure pressure changes as small as 0.2 hectopascals (hPa), corresponding to an elevation change of about 2 metres near sea level (Linacre and Hobbs 1977). The air pressure at sea level normally varies between about 970 hPa and 1040 hPa, but values as high as 1084 hPa and as low as 870 hPa have been recorded(Linacre 1992).

Temperature

Reliable thermometers were first developed in the mid 18th century. However, the development of a large scale temperature recording network did not begin until early in the 19th century (Bradley *et al.* 1985). There are several scales for measuring temperature. Before metrication in Australia, in September 1972, the Fahrenheit scale was used. The Celsius scale is in present use, expressed to the nearest 0.1°C. The Bureau of Meteorology (1990a) suggests that the measurement of temperature is uncomplicated and generally contains few errors. The three-hourly data set includes values for dry-bulb, wet-bulb and dew-point temperature.

The temperature is the equivalent of how hot or cold an object is, not how much heat it contains. Thus the air temperature varies greatly depending on where it is measured. Dry-bulb temperatures are usually measured with a mercury-in-glass thermometer, placed in a Stevenson screen for protection from direct sunlight. While there are more sophisticated electrical instruments for measuring temperature, the mercury-in-glass thermometers are still used for most meteorological observations.

The wet-bulb temperature is the lowest temperature to which air can be cooled by evaporating water into it and provides a measure of humidity. The mercury bulb of the wet-bulb thermometer is covered with a muslin bag that is kept permanently

moist. It gives a lower reading than the dry-bulb because of the cooling effect of water evaporating from the muslin bag.

Dew-point temperature (T_d) is the temperature to which the air must be cooled for the moisture present to reach saturation. The dew-point temperature is generally derived theoretically from the dry-bulb and wet-bulb temperature, with a correction for the site's elevation (Bureau of Meteorology 1988). At certain stations over some of the period of interest here (i.e. 1952 to 1990), the dew-point temperature has been given directly by instruments. If the dry-bulb temperature is the same as the dew-point temperature, the air is said to be saturated and the relative humidity is 100 per cent. Theoretically, dew-point temperature does not vary as the air temperature changes and hence it is fairly constant during the course of a day, unless there is a significant change in the origin of the air mass at the station.

Precipitation

Precipitation is collected in a rain gauge, which is a small bucket with a sharp bevelled edge to define the collecting area accurately. A standard rain gauge has a diameter of 200 mm and stands with its rim 300 mm above the ground to prevent the collection of any water that splashes or runs from the surroundings. The rainfall measured by a rain gauge is usually an underestimate of the local precipitation. The degree of underestimation varies with the gauge type and siting, the wind field around the gauge, wetting losses and evaporation from the gauge. The problem of rainfall measurement can be complicated not only by these physical effects, but also by sampling error. A single gauge may not give measurements characteristic of the surrounding area, especially in a region of irregular topography or areas subject to local isolated showers. There are differences in rainfall amounts over one or two kilometres or less which would not be detected by standard network of gauges, which may possibly be tens of kilometres apart due to financial practicability.

Different problems arise in other cases. For example, determining the amount of snowfall is particularly difficult and is confused by drifting snow. It is difficult to assess the amount of water collected by vegetation in a fog or by the ground during storms in remote areas. Various problems relating to rainfall measurement are well known and widely documented (see for example World Meteorological Organisation 1983, Folland 1988).

Sophisticated pluviometers and pluviographs with tipping bucket and other devices are now used to measure rainfall amounts and intensities at key stations. However the standard rain gauge still provides the vast majority of rainfall reports. Amounts of rainfall are measured in terms of the depth of the layer created by spreading the rainfall on a horizontal surface. Rainfall was measured in hundredths of an inch (0.254mm) from the beginning of the period under consideration until 31 December 1973. After that date, measurements were made to the nearest 0.2 mm (Bureau of Meteorology 1990a).The earlier values have been converted to the nearest tenth of a

millimetre (Bureau of Meteorology 1977).

Lapse rates

Dependence on elevation is well known in the case of pressure and temperature. Less systematic "lapse rates" also exist for rainfall. There are large variations from the average lapse rate from place to place, day to day and season to season. Significant differences in temperature lapse rates can arise from large temperature variations in the air layer immediately in contact with the ground, particularly on hot days and on cold nights with little or no wind. Differences can also arise from temperature inversions. Linacre (1992) and Hutchinson (1991a) suggest that lapse rates tend to be smaller in winter than in summer and lower inland than on the coast.

Heat from the surface on the sides of mountains may alter the local lapse rate, especially in light winds. The rate of decrease of pressure with height (see for example Sturman and Tapper 1996 or Barry and Chorley 1987) is given by

$$\frac{dp}{dz} = -g\rho \tag{2.1}$$

where p is pressure, z is the height above the ground, ρ is the density of air, and g is the acceleration due to gravity. Equation 2.1 is the hydrostatic equation, but in this form it is not meteorologically useful since neither density nor height is a directly measured quantity. Assuming that the perfect gas law applies to dry air, pressure p can be expressed as

$$p = \rho RT \tag{2.2}$$

where R is the gas constant for dry air (= 287 $m^2/s^2 deg$) and T is the temperature in K. The quantity for density in equation (2.2) can be substituted into equation (2.1) to give

$$\frac{dp}{dz} = \frac{-gp}{R\overline{T}} \tag{2.3}$$

where $\overline{T}$ is the temperature of the surrounding air mass. If T is the temperature of the parcel of air which is ascending, and it is assumed that this ascent takes place without exchange of heat between the parcel and its environment (an adiabatic process), the first law of thermodynamics can be written in the meteorologically useful terms of measured variables

$$\frac{dT}{dz}\frac{1}{T} = \frac{AR}{C_p}\frac{dp}{dz}\frac{1}{p} \tag{2.4}$$

where A is the reciprocal of the mechanical equivalent of heat (= 2.39 x 10^{-8} cal/deg) and C_p is the specific heat for air at constant pressure (0.239 cal/gm deg). Combining (2.3) and (2.4) gives the rate of change of temperature with height in the free atmosphere

$$-\frac{dT}{dz} = \frac{Ag}{C_p}\frac{T}{\overline{T}} \tag{2.5}$$

Under most circumstances T is very nearly equal to $\overline{T}$ and the rate of decrease of temperature with height for an ascending parcel of dry air can be approximated by Ag/C_p which is 9.8°C/km or 5.4°F/1000ft. This is the dry adiabatic lapse rate (DALR) which is positive when temperature decreases with height. However, if the air parcel is saturated or becomes saturated during ascent, the DARL no longer applies, since heat will be gained by the parcel through the latent heat of condensation as water vapour condenses into droplets. The latent heat of condensation is given by:

$$L = 734 - 0.51T \tag{2.6}$$

where T is the temperature in K. Thus the saturated adiabatic lapse rate (SALR) is less than the dry adiabatic by an amount dependent on temperature and pressure. For example at 1000 hPa and 20°C the SALR is 4.4°C/km, whereas at 600 hPa and 20°C the SALR is 3.6°C/km. A standard atmosphere adopted by the International Civil Aviation Organisation is used in calibrating aneroid barometers for measuring the altitude of aircraft. This standard has a mean sea level pressure of 1013.25 hPa with an air temperature of 15°C and a lapse rate of -6.5°C/km to 11kms, then the temperature remains constant up to 20 kms, a lapse rate of -1.0 °C/km then applies up to 32 kms.

Linacre (1992) shows that if air is lifted one kilometre, the pressure of each of the air's components is reduced by about 13%. A fall of this magnitude in the saturation vapour pressure of water corresponds to a reduction of dew-point temperature by about 2°C. In other words, Linacre gives the adiabatic dew-point lapse rate as about 2 K/km (e.g. 1.9 at 17°C and 2.3 at 37°C).

Climatological data

The first consideration, when analysing any collection of data, must be the characteristics of the data themselves. Data for climate studies are generally acquired from government agencies, and the availability and characteristics of the data vary. Some linear variables have a natural lower bound of zero (pressure, rainfall, wind speed), others such as temperatures do not, and some such as wind direction and aspect are circular. Climatic data have seasonal patterns, and values tend to be higher or lower in certain seasons of the year. Most climatic variables are temporally and spatially autocorrelated, that is, consecutive readings or observations at stations nearby tend to be strongly correlated with each other. Climatic data tend to include outliers that can cause them to be skewed and not all climatic data are normally distributed. This is especially so for rainfall.

The management of climate data typically of megabyte dimensions requires a high level of organization. To describe and summarize those data in forms that convey their important characteristics is one of the most frequent tasks in climate data

analysis. With very large data sets an estimation of summary statistics is basic to understanding the data, and is the usual way to present long term climatic conditions. Characteristics often described include: a measure of the centre of the data, a measure of spread or variability, a measure of symmetry of the data distribution, and perhaps estimates of extremes such as large and small percentiles. Some can be well summarised by fitting appropriate statistical distributions (e.g. normal for temperature, gamma and power of normal for rainfall).

To draw inferences about a population, the data in the sample must be random, independent and homogeneous. By random it is meant that there is no choice or control in the selection of the sample. The degree of independence varies with the nature of the data being considered. Successive daily stream-flows are not independent, monthly stream-flows are much more independent. The serial correlation in time series reduces the effective length of the series and carrying out statistical tests of significance should be considered.

Homogeneity implies that the data belong to the same population. Observations, however, do not always reproduce natural conditions accurately, as there are many sources of error. The preparation of original observations for purposes of comparison, whether as single values or as a series, is one of the main tasks of climatological methods. An appropriate critique and treatment of the data should aim to eliminate both systematic and sporadic errors.

Four main factors affecting station homogeneity are changes in instrumentation, changes in station location, changes in observation times and changes in environment around the station. The effects of these four major factors have been discussed at length in Bradley and Jones (1985), Bradley *et al.* (1985) and Jones *et al.* (1985), and will be discussed further as they apply to the various elements and data sets.

The pressure and temperature data set

The Bureau of Meteorology three-hourly records

The Australian Bureau of Meteorology (BoM) maintains a national climate data bank, climate information is held in several unrelated data sets. The pressure and temperature data sets are derived from the BoM three-hourly records ("card 7") for all Australian stations. Data from the three-hourly data set used in this thesis were station number, date, station level pressure (recorded in 0.1 hPa), dry-bulb temperature (recorded in 0.1 °C), wet-bulb temperature (recorded in 0.1 °C) and dew-point temperature (recorded in °C). The three-hourly data set also includes other climatic data. Wind data are recorded at most of the pressure and temperature stations. Spatial interpolation of surface wind, which is highly dependent on local surface conditions, requires a denser station network than is provided by the three-

hourly data set. The cloud, soil and sea condition data are recorded at very few stations and are not sufficient to warrant their use for continental scale spatial interpolation.

The BoM three-hourly data set at the end of 1991 comprised more than one and a half gigabytes of data from the 1219 station records of varying time lengths. For this study the period from 1952 to 1990 was selected. The time limits are due to the few stations recording three-hourly data before 1952 (approximately six), and as the data set was acquired in late 1991, the months to December 1990 formed the most complete data from all states.

In late 1987 the BoM changed its processing system so that all observations made at the stations are recorded. This means that before late 1987 only two observations per day are available from the computer archive for most stations. These stations collected measurements at only 0900 and 1500 Local Standard Time (LST). Other gaps occur in the three-hourly record for many reasons, such as closure of a station or absence of the observer.

Each station is identified by a six-digit site number based on the rainfall districts, with the first three digits of the site number being the rainfall district number. All sites used in the data analysis are owned by the Bureau of Meteorology. The station identification number was used to extract the official Bureau station name, longitude and latitude of the station, and station and barometer height from the Bureau of Meteorology (1990b) dictionary.

Data set homogeneity

The data set suffers from large discontinuities in both temporal and spatial coverage, and contains errors. Some errors are simply positional. For example, it is possible to find errors in the station dictionary describing the longitude, latitude, and, especially, elevation. Occasionally, errors are found in the units, for example inches not correctly converted to millimeters. Other systematic errors stem from the mis-reading or adjustment of the recording devices.

Changes in station location are an important factor in determining station inhomogeneity. Any significant change in location is usually indicated by a change in station number. However the station dictionary gives the present location and does not include any previous changes in the barometer height. Although the Bureau of Meteorology does not change the station number in these cases, this does constitute a change in location. The most common height changes occur in the barometer heights at regional offices, as they relocate to different buildings or floors of buildings in the major cities. Jones (1991) in his assessment of the Southern Hemisphere pressure records states that in the majority of cases, documented site changes have an insignificant effect on the homogeneity of the data. This is discussed further in Chapter 4.

Individual stations do record changes in instrumentation, exposure and measurement techniques in the station logs, but again, these changes are not recorded in the station dictionary. However, all stations must have had temperature instrument changes in 1972 with the change from Fahrenheit to Celsius scales.

The continent of Australia covers over thirty degrees of longitude and three different time zones. The three-hourly records are collected at a specified time (e.g. 9 am, 12, 3 pm etc.), but this is local time so the readings are not taken simultaneously across the country. This does not constitute an inhomogeneity in an individual station's records; however, when the stations are considered as a spatial data set, temporal discontinuities occur across the timezone borders which tend to coincide with state borders.

This problem is further exacerbated by daylight saving time (DST) which is implemented in certain Australian States, usually between late October and early March. Data are collected at regular local clock times but recorded at local standard time (LST). The introduction of Daylight Saving Time in the 1970s in some states meant that, at some stations, data are a mixture of daylight and non-daylight saving observations, and were collected at different times during the period 1952 to 1990. For example, 9am clock time in Canberra on January 1 1960, would have been recorded as 0900 LST, while on January 1 1980, it would have been recorded as 0800 LST. Similarly, midday records are 1200 LST before the introduction of DST, but at 1100 LST in the period after the introduction of DST. Some stations have both the pre-daylight saving time and the daylight saving time recordings for a short change-over period. Figure 2.1 clearly shows the temporal discontinuity in the monthly average values for Canberra for the period 1970-1990.

Changes in the environment around a station, especially the growth of cities, has often been considered to have an important effect on station data homogeneity. As metropolitan areas have developed over recent times they have inadvertently produced a warming directly attributable to the human modification of the landscape (e.g. Landsberg 1981). For sites that have endured rapid growth, this warming is comparable in magnitude to that anticipated to have occurred so far as a result of the enhanced greenhouse effect (Chagnon 1992). Moreover, this effect has been detected for cities across the globe, ranging from the tropics to high latitudes and to a lesser extent even in relatively small communities. While the great majority of the Australian stations are not in the major cities, many of the three-hourly stations are at aerodromes, and environmental changes should be kept in mind when considering trends in the station data.

The derived data sets

To remove data inhomogeneities caused by different recording times from the three time zones and the introduction of DST, specific UTC (Coordinated Universal Time or Universel Temps Coordonné) times were used. UTC is ten hours behind local

standard time in the Eastern Standard Time Zone of Australia, and nine and a half and eight hours behind local standard time in the Central and Western Time Zones respectively. 0000 and 0600 UTC were selected, as they are the equivalent of 9 am and 3 pm for the centre of Australia.

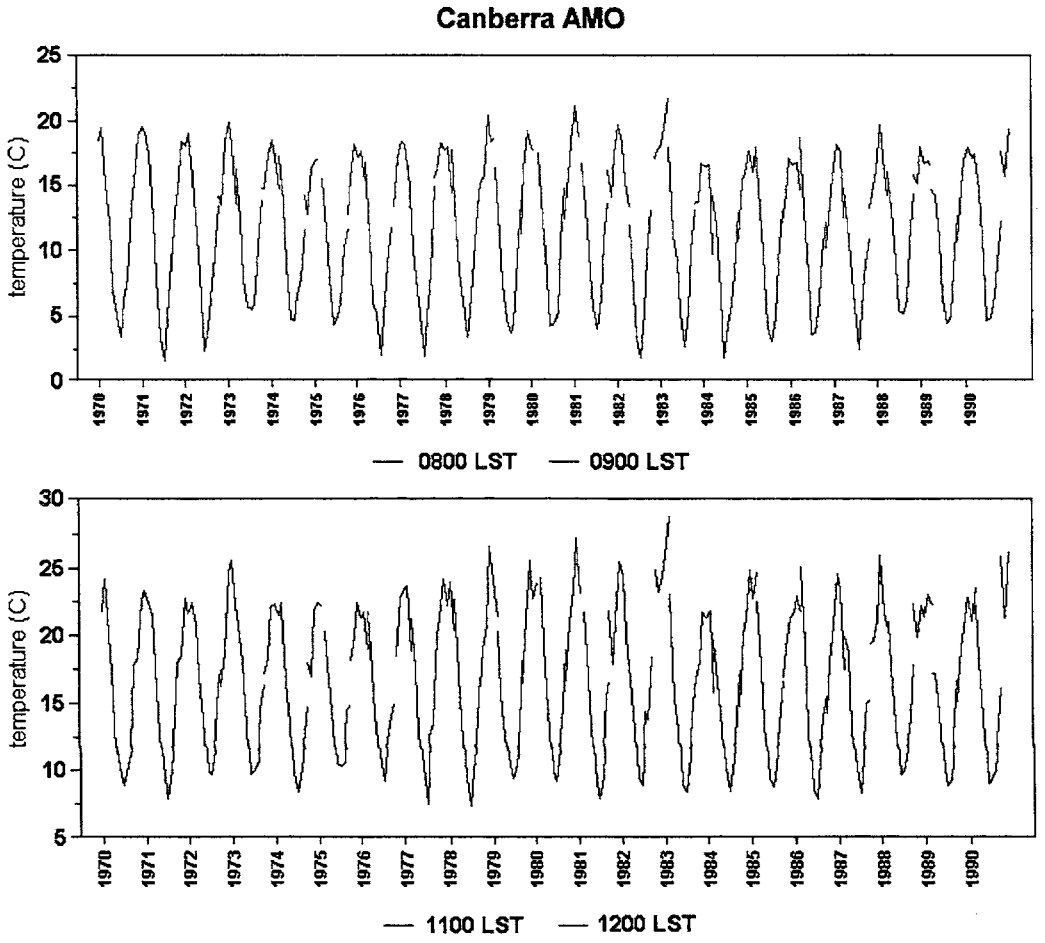


Figure 2.1: Canberra 0800 and 0900 LST, 1970 -1990, and Canberra 1100 and 1200 LST, 1970 -1990.

The 0000 UTC and 0600 UTC values were derived from the three-hourly time series for each day. In order to keep the time between the recorded data to less than six hours for any interpolated value, stations were selected by accepting a minimum of 2000 three-hourly records in each year. This meant that they were likely to be averaging seven records a day. This reduced the number of stations from 1219 to 131 for all months. The 0000 and 0600 UTC times were linearly interpolated from the three-hourly values. A spline function was considered for this purpose, as has since been done by Kong (1995), but in order to keep the process simple (and as can be seen in Figure 2.2) a linear interpolation was found to be satisfactory.

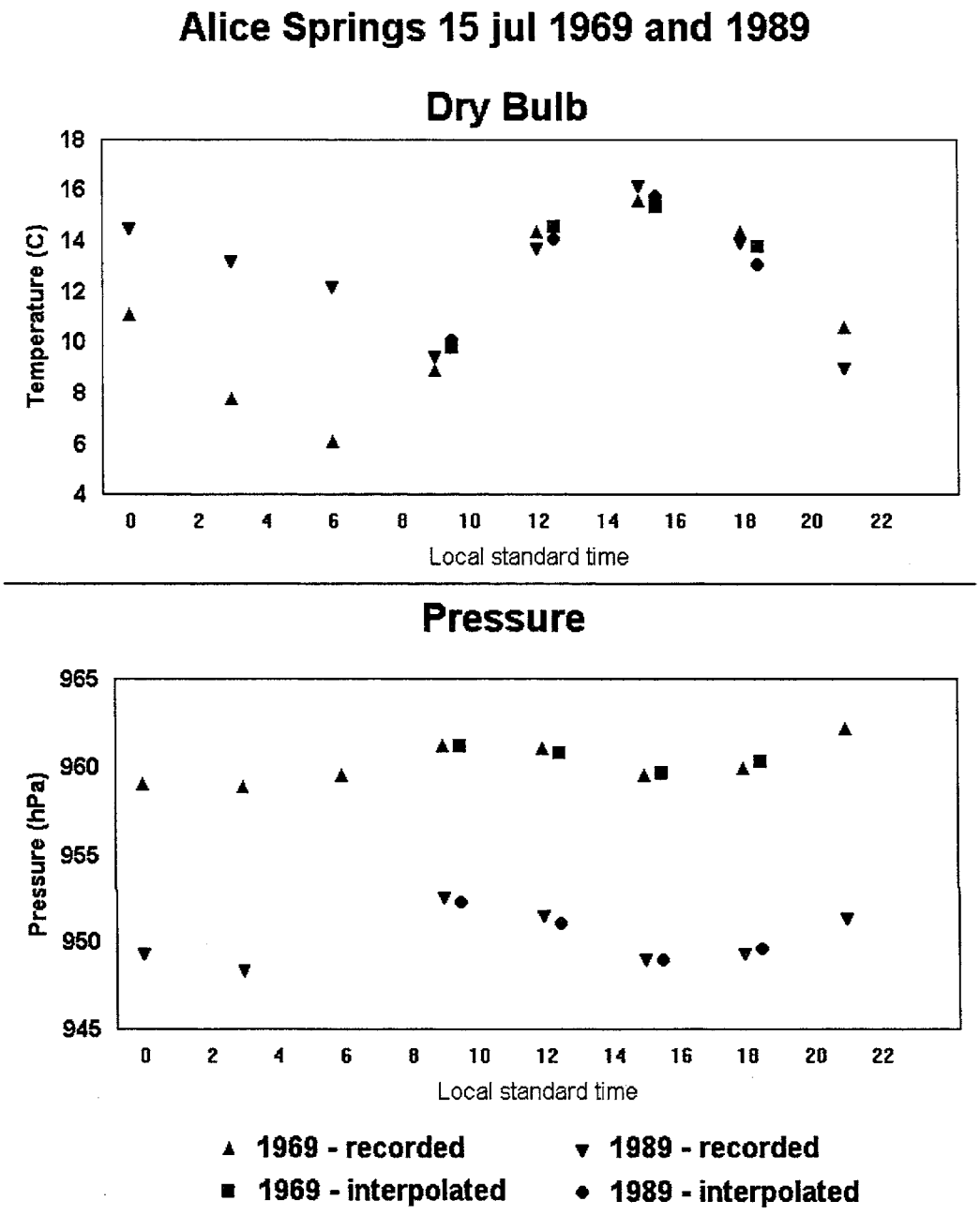


Figure 2.2: Actual and interpolated pressure and temperature values for Alice Springs for the 15th of July 1969 and 1989.

Graphs of data from all stations with greater than five years of record were examined. However, as the ANUSPLIN package allows for the detection of climate station inhomogeneities, and as an investigation of this ability of the ANUSPLIN package was part of the objectives of the study, no records were removed. The data homogeneity problems of the data set are discussed further in Chapter 4.

To model the spatial distribution of a climatological variable, the climatic series for each site must be summarized. Once the daily 0000 and 0600 UTC values had been

interpolated, unbiased monthly means for each station were obtained. Mean values for both the long-term monthly means and the individual monthly means were calculated, and formed the station data set. Long-term monthly averages (LTA) were calculated as the mean for all 0000 and 0600 UTC values for each month; for example the long term 0000 UTC January average is the average of all January 0000 UTC values for the period of record for the station. Individual monthly averages (AVG) were also calculated for all 468 months from January 1952 to December 1990. Any station with fewer than 20 daily interpolated UTC values for the month was rejected for that month.

Station locations, elevations, first year of record and number of months of record are listed in Appendix C. The minimum, maximum and variance were also calculated and are given with the LTA for each station on the CD. The CD also includes the monthly mean pressure and temperature records for the individual stations that were used to generate the set of surfaces. The LTA of the station pressure, dry-bulb, wet-bulb and dew-point temperatures for each station are also presented graphically on the CD. For each station with a complete record for the 468 months from January 1952 to December 1990 (shown in Figure 2.3) the CD contains graphics of the monthly time series and ten-year running monthly means for all the climatic variables.

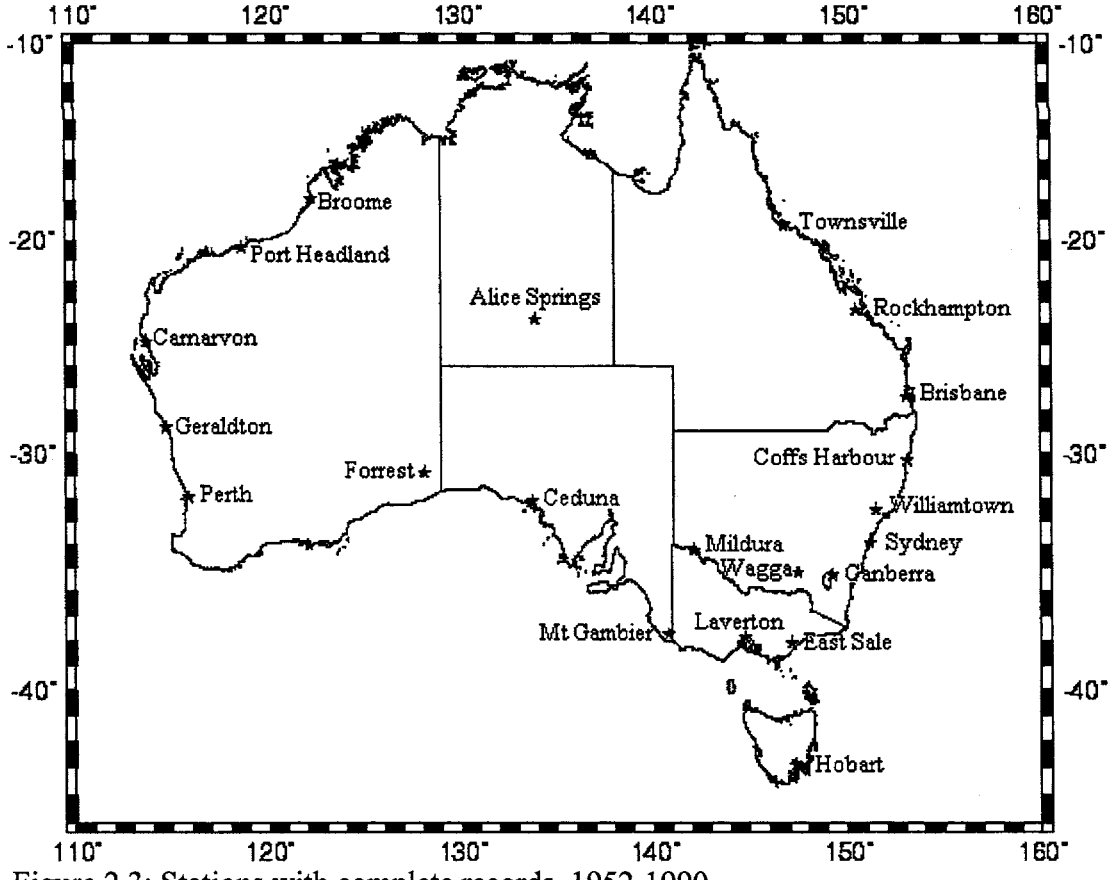


Figure 2.3: Stations with complete records, 1952-1990.

As can be seen from Figure 2.3, stations with complete three-hourly records tend to be located at major and regional centre airports. Major centres such as Adelaide and Darwin, while they do not have a complete record, are missing only two or three years and are almost complete.

Station data network

Temporal coverage

Although the period of record extends from 1952 to 1990, few stations have contributed data for the entire period of record. As shown in Figure 2.4, relatively few stations had complete or even near-complete records, and the number of stations for individual months varies between 34 and 113 stations, depending on the year. The implications of the differences in the length of records for different months and for assessing spatial long term averages are discussed further in Chapter 4.

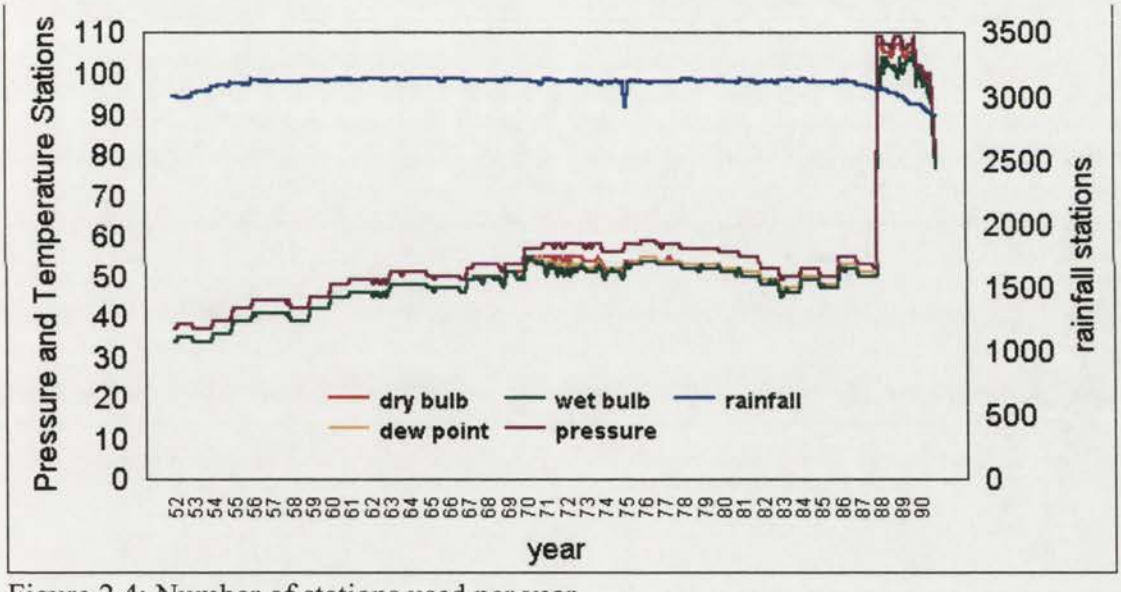


Figure 2.4: Number of stations used per year.

The most noticeable feature in Figure 2.4 is the greater number of rainfall stations compared to the three-hourly pressure and temperature stations. The rainfall stations are only read at 9am and require relatively simple and inexpensive equipment compared to the cost of maintaining three-hourly records stations. Many rainfall stations are farming homesteads where the primary meteorological variable of interest is the rainfall.

It is interesting to note the effect of the rise in public interest in climate change issues which are reflected in Figure 2.4 in the increase in number of stations recording data or becoming available to the public in 1988.

Spatial coverage

The spatial density of the data is heavily biased towards the eastern seaboard of Australia and is particularly poor in sparsely populated areas, but it is a good representation of all the station records from the Bureau of Meteorology. Figure 2.5 shows the station distribution for 1952, 1982 and 1989. A listing of the stations and the number of months of record is given in Appendix D.

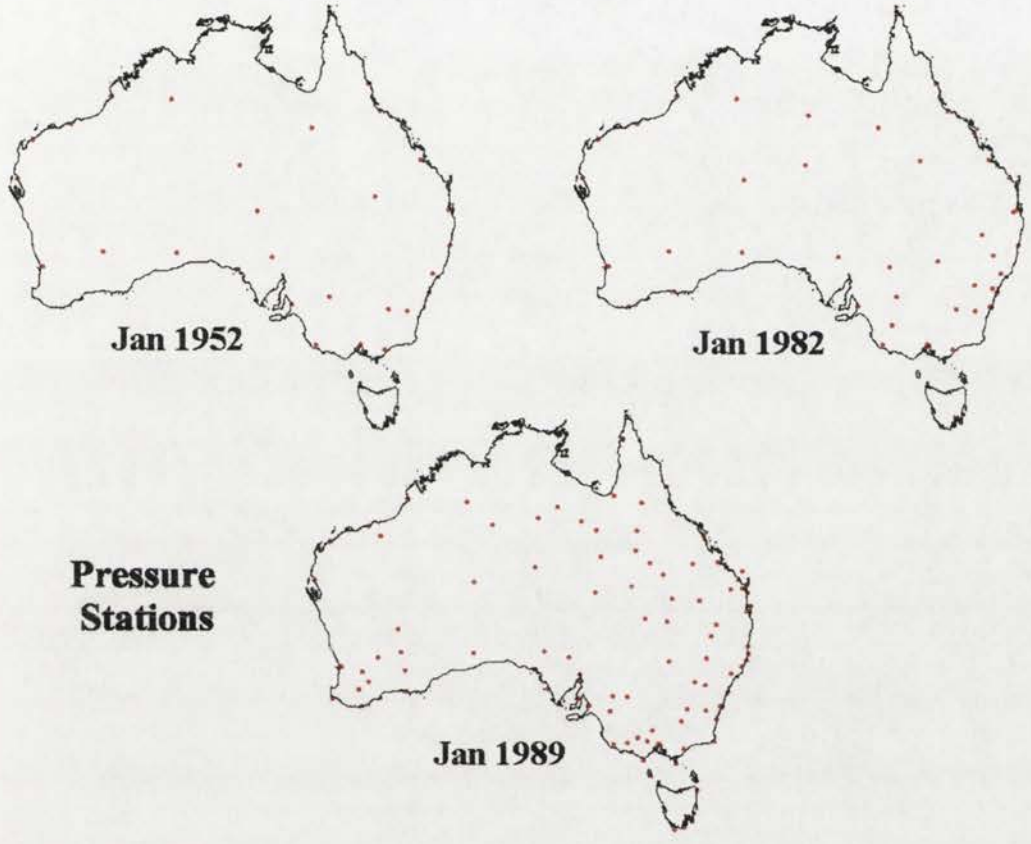


Figure 2.5: Station distributions for January 1952, 1982 and 1989.

Topographic coverage

While it is useful to have a description of the spatial distribution of the stations in the database, it is instructive to view the topographic representation that the stations provide. A 1/10th degree Australian Digital Elevation Model (DEM) was resampled from Hutchinson's (1994) 1/40th degree DEM and is shown in Figure 2.6d. The topography or DEM represented by the stations for the years 1952, 1982 and 1989 is shown in Figure 2.6 (a, b and c respectively). As can be seen in Figure 2.6, the 1989 DEM is closest to the Australian DEM; this shows the improvement that can be gained with the better representation given by a larger and better distributed station database. Figure 2.6 shows that the topographic representation is poorest for areas of high elevation. It is interesting that the topographic representation is better for the data sparse regions. The effect that the topographic and spatial coverages have on interpolation errors is discussed further in Chapter 4.

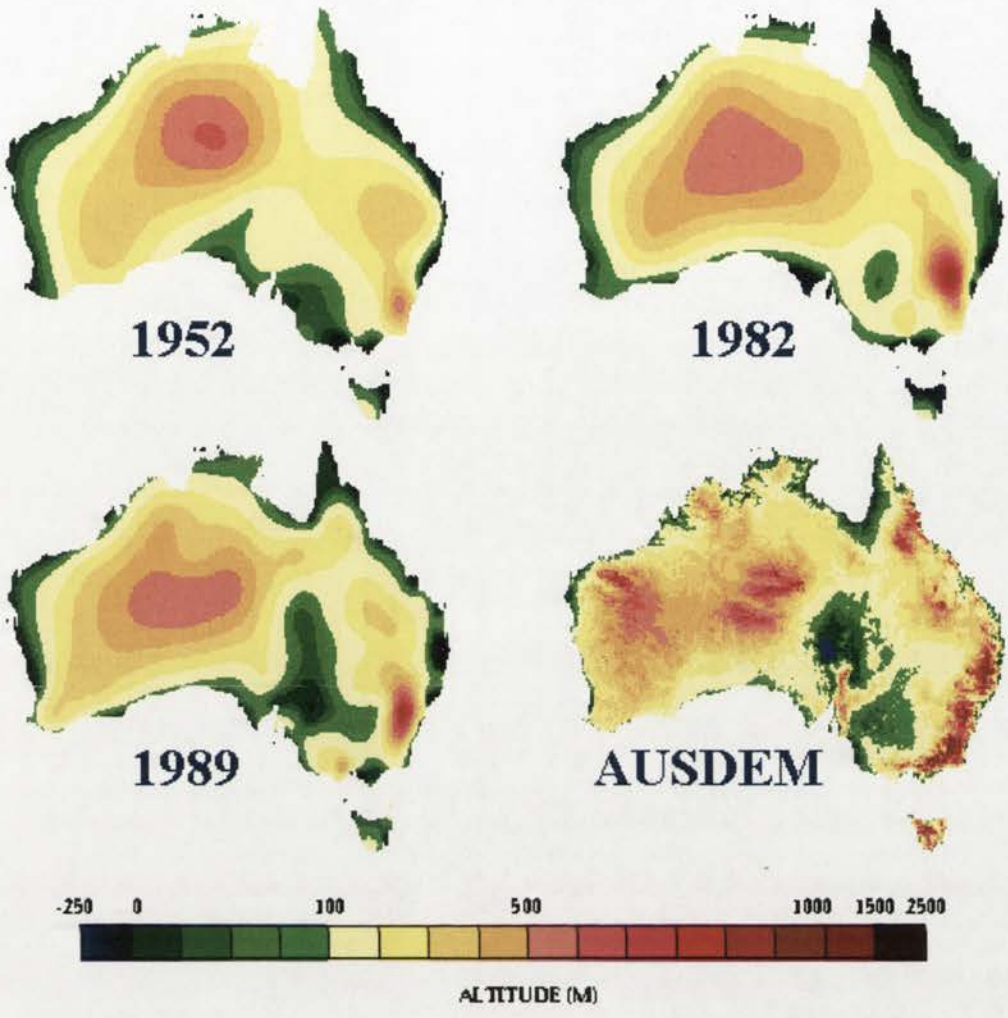


Figure 2.6: Station DEMs for 1952, 1982 and 1989, and Australian 0.1° DEM.

The rainfall percentile data set

The Bureau of Meteorology monthly rainfall records

Of the standard climate variables, rainfall displays the greatest variability in time and space and so makes the greatest demands on both temporal and spatial analysis techniques. For this reason, rainfall is also the most commonly measured climate variable (Hutchinson 1995a, 1995b). Rainfall measurements apply from 0900 LST to 0900 LST, and are recorded against the day of reading (Bureau of Meteorology, no date).

The Bureau of Meteorology, in 1994, held monthly rainfall information for 17,616 sites (open and closed) throughout Australia, of varying periods of record. From this network, those stations which had at least 35 of the years between 1952 and 1990 with complete data were accepted. This reduced the network to 3160 stations whose

locations are plotted in Figure 2.7. It is not practical to produce a list of the stations in the hard copy of the PhD, but a listing is given on the CD.

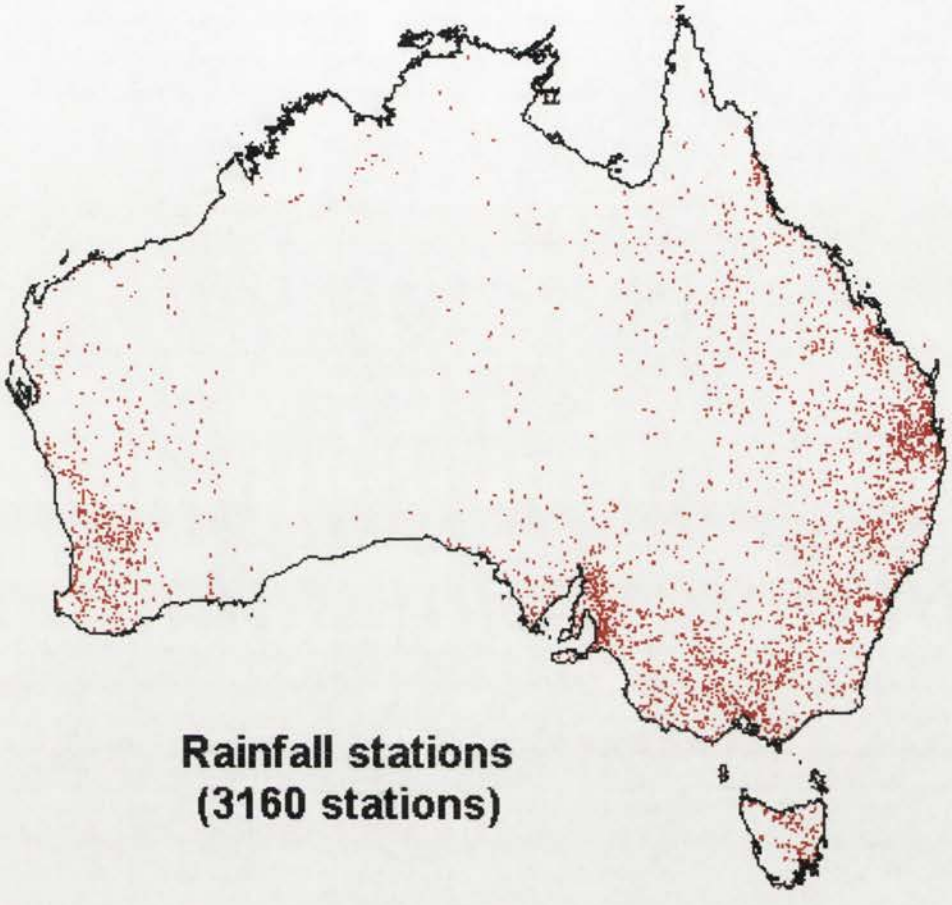


Figure 2.7: Locations of the 3160 Bureau of Meteorology stations with continuous monthly rainfall data, 1952-1990.

While the 3160 stations in the data set well represent the entire BoM station network, the BoM network is quite sparse in parts of central and northern Australia. Subsequent interpolated values should be viewed with caution in these areas.

Rainfall percentiles/deciles

Rainfall deciles for various periods ranging from several months to a year have been proposed by Gibbs and Maher (1967) and have been used by the Australian Bureau of Meteorology to gauge the severity of dry conditions across the continent (Gibbs 1975). Despite the lack of dependence on any water balance calculations, Smith *et al.* (1993) found a close connection between six-monthly rainfall deciles and the Palmer Drought Index. The maps of rainfall deciles displayed in Gibbs and Maher (1967) and Gibbs (1975) were produced manually but indicate that there is broad spatial coherence between monthly rainfall deciles and percentiles. This is distinctly not the case for

actual monthly rainfall totals which exhibit, even in their long term mean values, extremely complicated spatial patterns which are heavily dependent on topography.

Rainfall percentiles are really a measure of departure from normal conditions, independent of the actual monthly mean values. The major cause of such departures is broad scale changes in synoptic patterns which are essentially independent of topography. Thus, while two nearby sites may have very different monthly mean rainfalls, if the actual monthly rainfall at one site is half of the long term average, then it is likely to be approximately half of the long term average at the neighbouring site. Because the underlying causes for the rainfall anomalies are broad scale, it should then be possible to interpolate the rainfall percentiles from the network of rainfall stations used in this study. Moreover, rainfall percentiles provide, in a very direct way, a measure of conditions immediately comparable between different sites and times.

The rainfall decile for a particular period of the year is defined by ranking the total rainfall received for the nominated period with respect to the total received for the same period in all years of record. The percentile number $C(V)$ for a particular value in position m (the 'rank order') in a set of N values, arranged from smallest to largest, is commonly given by the following expression:

$$C(V) = \frac{100m}{(N+1)} \tag{2.7}$$

The use of $N + 1$ rather than N in equation 2.7 avoids the final value in a ranked set being regarded as the highest possible. The above expression is given by Gumbel (1954) and there have been several variations which attempt to achieve symmetry by making the last percentile differ from 100% as much as the first differs from 0%. Variations include those given by

$$C(V) = \frac{100(m-0.5)}{N} \tag{2.8a}$$

$$C(V) = \frac{100(m-0.44)}{(N+0.2)} \tag{2.8b}$$

The variation used by the United Kingdom Meteorological Office (Linacre 1992) and Smith *et al.* (1993) is the one used in this study and is given by

$$C(V) = \frac{100(m-0.31)}{(N+0.38)} \tag{2.9}$$

It should be noted that for sites where a significant proportion of the months receive no rainfall, the percentile value of zero rainfalls is not well defined. This is overcome by setting the percentile of zero rainfall to the mid-point of the relative frequency that the month is dry.

The first decile (or 10 percentile) is the rainfall amount not exceeded by 10% of all totals. The fifth decile or median is the amount not exceeded in 50% of all years. Decile

ranges are the ranges of values between deciles. Thus the first decile range is that below the first decile, and the eighth decile range lies between decile seven and decile eight. Monthly percentiles for the period 1952 to 1990 for Perth, Darwin, Alice Springs and Canberra are plotted in Figure 2.8.

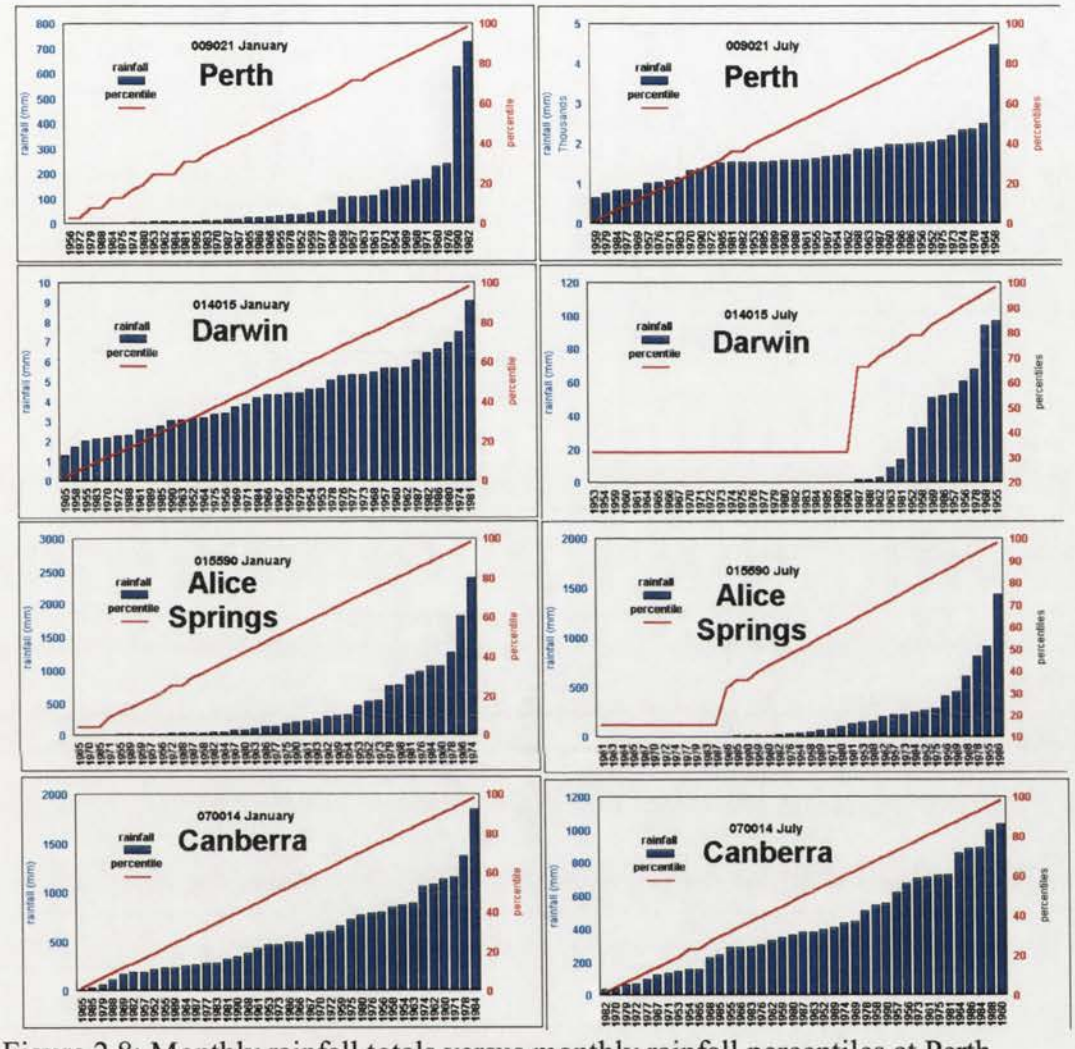


Figure 2.8: Monthly rainfall totals versus monthly rainfall percentiles at Perth, Darwin, Alice Springs and Canberra for January and July from 1952 to 1990.

Motivated by the apparently successful applications of rainfall deciles and percentiles, and the simplicity of their derivation, this study has used monthly rainfall percentiles as a spatially coherent measure of rainfall.

Conclusions

A monthly climate database has been constructed for pressure and temperature from the three-hourly records from the Australian Bureau of Meteorology, for all stations for the period 1952 to 1990, and for linearly interpolated specific UTC times (0000 and 0600). These times were selected to accommodate the different time zones across

the country (usually three) and to remove station inhomogeneity problems associated with the introduction of Daylight Saving Time (DST) in the early 1970s for some Australian states during the summer months.

A monthly climate database has also been constructed for rainfall percentiles. They are considered more spatially suited to interpolation than actual rainfall amounts. The distribution of rainfall stations which have full records available for the period 1952-1990 (3160 stations) was considered insufficient to represent the actual spatial variation of rainfall. The two most notable features of rainfall deciles are their simplicity in terms of calculation and interpolation and their lack of dependence on topographic features.

Chapter 3. The Spatial Interpolation of Climatic Variables.

Introduction

Maps have traditionally been the mechanism for storing information about the spatial distribution of climatic data. Visualization is useful for examining and validating observational data and the output of environmental models. Observations are spatially discrete, therefore mapping necessitates the estimation of the variable between the data points. Meteorological variables are continuous in space and are predictable within the bounds of stochastic fluctuations. Therefore it is reasonable to interpolate data and to derive a map from the surface produced. The interpolation can be done subjectively, by statistical modelling or by physical modelling. A map is obtained by calculating the values of the fitted function across a regular grid of the two independent variables denoting horizontal position.

Subjective interpolation is still widely used for meteorological purposes, as it allows for the incorporation of additional knowledge of the interpolator. This, however, makes it very difficult to reproduce and to store in digital form. Subjective interpolation is considerably more time consuming than computational methods. Physical modelling is also widely used for meteorological and climatological purposes. Physical models, however, require a large amount of additional meteorological data than are usually available, and are theoretically and computationally more complex than statistical interpolation methods.

Statistical interpolation techniques model a spatial distribution as a function of observational data across a region without prior knowledge of the spatial distribution or its underlying physical causes. Statistical interpolation of the mean values incorporates natural temporal variability. Various methods of spatially interpolating sparse point observations have been used, including partitioning into regions, moving averages, kriging and thin plate smoothing splines.

This chapter begins with a brief discussion of some of the techniques used for the spatial interpolation of climatic variables and some of the limitations of these techniques. Thin plate smoothing splines, the use of the ANUSPLIN (Hutchinson 1991a, 1994) set of programs, and the map-producing process are then described. The process of input option selection is discussed, and the daily pressure maps and seasonal rainfall decile maps are compared to hand-drawn Bureau of Meteorology maps.

Established methods of statistical interpolation

Various methods to spatially interpolate sparse observational data have been used, including partitioning into regions, local interpolation methods, moving averages, kriging and thin plate smoothing splines. Of these methods, statistical interpolation techniques appear to be best suited to the task. These techniques include the methods of thin plate smoothing splines (Wahba and Wendelberger 1980, Hutchinson 1991a) and

kriging (Delfiner and Delhomme 1975, Cressie 1991). Both methods attempt to achieve minimum error (optimum) interpolation, have the same underlying computational structure and have measures of their spatial accuracy. There are, however, some significant practical and theoretical differences between the two methods and these have been extensively examined by Hutchinson (1993a) and Hutchinson and Gessler (1994).

Local interpolation methods

A large number of local interpolation techniques have been developed (Akima 1978, Sibson 1981, Watson and Philip 1985). They divide the region containing the data points into elements to which simple functions are fitted so that the overall function is continuously differentiable. The resulting fitted surfaces suffer from somewhat arbitrary restrictions on their form and can be sensitive to the positions of the data points. This is especially true when the data points are irregularly spaced, and spurious effects may be generated away from the data points. These techniques become quite complex to implement in more than two dimensions and do not lend themselves easily to the smoothing of noisy data (Hutchinson 1991a).

Moving average methods

Moving average methods involve a subjective choice of weighting function, which is usually defined in terms of a radius of influence, R, beyond which the data points are ignored. For example a weight, $\alpha_j(x)$ may be defined by:

$$\alpha_j(x) = \frac{d_j^{-u}}{\sum_{d_j \le R} d_j^{-u}} \qquad \text{for all j with } d_j \le R \tag{3.1a}$$

$$\alpha_j(x) = 0 \qquad \text{for all j with } d_j > R \tag{3.1b}$$

so that $\Sigma\alpha_j = 1$, where d_j is the distance from x, the point to be interpolated to x_j for $j = 1,2,3,....M$. The vector x_j contains the coordinates of the sample point j, and u generally ranges from 2 to 3. The interpolated or estimated function, $g^*(x)$, is then defined by:

$$g^*(x) = \sum_{j=1}^{M} \alpha_j(x) g_j \tag{3.2}$$

where g_j is the j^{th} observation. The choice of an optimum radius can be a problem when points are irregularly spaced and it is difficult to impose objective data smoothing. These methods have been used to derive digital elevation models (Eastman 1995) and in a variety of meteorological applications (Cressman 1959 and Shepard 1968).

Kriging

The term "kriging" is derived from the name of D.G. Krige, who first introduced the use of moving averages to avoid systematic overestimation of reserves in the field of mining. The main idea is to consider a spatial variable as a realisation of a spatially autocorrelated random function that accounts for the complicated behaviour of natural spatial data:

$$z_i = f(x_i) + \varepsilon_i \qquad (i = 1,\ldots,n) \tag{3.3}$$

where f is a function to be estimated from the observations z_I, which include a zero mean, usually spatially discontinuous, error term ε_i. The x_i are commonly assumed to represent coordinates in two- or three-dimensional Euclidean space. While kriging is normally applied locally it can be used to generate surface discontinuities.

Kriging is a weighted moving average interpolation method based on equation 3.2, in which the weights α_j, are derived from an unbiased estimation procedure that minimises the variance. The variance is a function of the spatial covariance or semivariogram function and the separation of the data points, and of the point or block being estimated (Delfiner and Delhomme 1975). They give the experimental semivariogram, $\phi_{ij}(h_\eta)$ as

$$\phi(h_\eta) = \frac{1}{2M_n}\sum_{\eta=1}^{M_\eta}[g(x_\eta + h) - g(x_\eta)][g(x_\eta + h) - g(x_\eta)] \tag{3.4}$$

where η is an index for different lag classes, h is a separation scalar belonging to lag class h_η and M_η is the number of increment pairs for the lag class h_η. Kriging extends easily to higher dimensions and attempts to achieve minimum error (optimum) interpolation. Furthermore, kriging provides an easy criterion for the goodness-of-fit, the mean square error.

Commonly assumed functional forms for $\phi(h_\eta)$ include the linear, spherical or Matheron, exponential, Gaussian, cubic, periodic or hole-effect models, or combinations of these models. The method has been used extensively for spatial interpolation in hydrology and soil science and a comprehensive coverage of kriging is given in Cressie (1991).

Thin plate smoothing splines

The original thin plate (formerly Laplacian) surface fitting technique was described by Wahba (1979), with modifications for larger data sets due to Bates and Wahba (1982), Elden (1984) and Hutchinson (1984). Thin plate smoothing spline interpolation is a global method as it uses all the given data for prediction at each point (Laslett *et al.* 1987). The method is also non-parametric and so is relatively insensitive to the distribution of the parent population.

Thin plate smoothing splines are formally related to kriging methods and have a common computational formulation in relation to observed data when kriging is applied globally. In the case of kriging, the z_i in equation 3.3 are assumed to be values of a spatially autocorrelated random field. In the case of splines, the z_i are assumed to be values of a smooth unknown function.

The value of the variable over a region is interpolated by estimating the true underlying function. This function is estimated to be as smooth as possible without significantly distorting the given data values (Hutchinson 1984).The observational model for a thin plate spline with two independent spline variables is that the observed monthly mean z_i at the position x_i of the i^{th} station is given by equation 3.3 above, where f is an unknown smooth function and the ε_i are independent random errors with zero mean and variance $d_i\sigma^2$, where d_i is the (local) relative variance and σ^2 is the common (normally unknown) variance.

The unknown smooth function is estimated by finding the function f (a thin plate spline) which minimises

$$\sum_{i=1}^{n} \left[\frac{z_i - f(x_i, y_i, h_i)}{d_i}\right]^2 + \lambda J_m(f) \qquad (3.5)$$

where $J_m(f)$ is a measure of the roughness of f in terms of the m^{th} order derivatives of f (usually second order), and λ is a positive smoothing parameter.

In meteorological and climatological applications the weights, d_i, are used when the data do not have equal length record. They have been set equal to the reciprocal of the number of years of record at the i^{th} data point and to the variance divided by the number of years for rainfall interpolation (Hutchinson and Bischof 1983). This aspect of the spline functions is discussed further in Chapter 4.

The smoothing parameter λ determines a trade-off between data fidelity and surface roughness. It is usually calculated by minimising the generalised cross validation (GCV). This is a measure of the predictive error of the fitted surface which is calculated by removing each data point in turn and summing, with appropriate weighting, the square of the discrepancy of each omitted data point from a surface fitted to all the other points. The value of λ is chosen to minimise this sum, that is, to minimise the predictive error as ascertained by the performance of the fitted function in predicting omitted data points.

It is possible to calculate the GCV implicitly, and hence efficiently, because the fitted values depend linearly on the data. Intuitively the GCV is a good measure of the predictive power of the fitted surface, as has been verified both theoretically and in applications to real and simulated data (Wahba 1990 and Hutchinson 1991a).

Statistical interpretation proceeds by way of analogy with least squares regression analysis. Thin plate splines can be viewed as a non-parametric generalisation of linear regression analysis. The influence matrix is defined to be the matrix A which takes the vector of data values z_i to the vector of fitted values f_i. The GCV is calculated by

$$GCV = \frac{(1/n)\sum_{i=1}^{n}[(z_i - f_i)/d_i]^2}{[(1/n)\,trace(I - A)]^2} \qquad (3.6)$$

The number of parameters of the fitted model is taken to be the trace(A) or the sum of diagonal elements of A. The degrees of freedom of the residual from the data is trace(I-A). It should normally exceed n/2 when there are adequate data (Hutchinson and Gessler 1994). Since the data points each have a specified noise, using the data points to assess the predictive error gives an over large estimate of the true error of the fitted surface. Approximately we have

$$GCV = TMSE + \sigma^2 \qquad (3.7)$$

where TMSE is the true mean square error over the data points. Thus minimising GCV efficiently minimises the TMSE, without knowledge of σ^2. The method affords an estimate of σ^2 by analogy with linear regression. If σ^2 were known, an unbiased estimator of TMSE could be given by

$$(1/n)\sum_{i=1}^{n}[(z_i - f_i)/d_i]^2 - (2/n)trace(I - A)\sigma^2 + \sigma^2 \qquad (3.8)$$

This can be used to determine the smoothing parameter but it is sensitive to errors in the value of σ^2. Since σ^2 includes both measurement and interpolation error, *a priori* estimation of σ^2 is difficult if not impossible. It is preferable to determine the smoothing parameter by minimising the GCV, which does not require any variance estimate.

Partial thin plate smoothing splines

Bi-variate interpolation of mean monthly climate is rarely appropriate. Much of the perceived spatial variation in monthly mean weather parameters is directly attributable to corresponding variation in topography. There exist significant additional forcing variables, in particular elevation effects on pressure, temperature and rainfall. The temperature environmental lapse rate is about 6-7 °C per 1000 m increase in elevation (Linacre 1992). Less systematic 'lapse rates' also exist for rainfall. The inclusion of elevation can be used to greatly improve the accuracy of interpolation (Hutchinson 1991a , 1991b). Neither the underlying topography nor the overlying climate field can be similarly calculated simply as an analytic function of geographic position (i.e. latitude and longitude).

Since both pressure and temperature decline approximately linearly with elevation, it is natural to interpolate these climatic variables by a partial spline with two independent position variables (longitude and latitude) and a single linear

dependence on elevation. Partial spline interpolation of monthly mean temperature has been investigated by Hutchinson (1991a). Partial thin plate smoothing spline is advantageous when dealing with data which are limited, either in terms of accuracy or spatial density.

The extension to partial splines is based on Bates *et al.* (1987). This allows the incorporation of parametric linear sub-models (or covariates), in addition to the usual independent spline variables. They are a robust way of allowing for such dependencies provided a parametric form for this dependence can be determined. In the limiting case of no independent spline variables (not currently permitted in ANUSPLIN), the procedure would become a simple multi-variate linear regression procedure.

The observational model for a partial thin plate spline is that n data values z_i at positions x_i are given by

$$z_i = f(x_i) + \sum_{j=1}^{p} \beta_j \psi_j(x_i) + \varepsilon_i \qquad (i = 1,...,n; j = 1,...p) \qquad (3.9)$$

where f is an unknown smooth bi-variate function to be estimated, the ψ_j are a set of p known functions, and the β_j are a set of unknown parameters which have also to be estimated. As for the previous formulation for thin plate splines, x_i commonly represent coordinates in two or three dimensional Euclidean space, the ε_i are independent random errors with zero mean and variance $d_i\sigma^2$.

The function f and the parameters β_j are estimated by minimising :

$$\sum_{i=1}^{n} [(z_i - f(x_i) - \sum_{j=1}^{p} \beta_j \psi_j(x_i)) / d_i]^2 + \lambda J_m(f) \qquad (3.10)$$

where $J_m(f)$ is a measure of the roughness of the spline function f defined in terms of the m^{th} order derivatives of f and λ is a positive smoothing parameter as defined for thin plate splines. For both thin plate splines and partial thin plate splines the solution to the minimisation problem can be solved explicitly, with the estimate of f having an expansion in terms of a scalar function of distance from each data position. The form of the scalar function depends on the dimension of the x_i and the order of derivative m defining the roughness penalty (Wahba 1990). The solution reduces to an ordinary thin plate smoothing spline when there is no parametric sub-model.

A comprehensive introduction to the technique of thin plate smoothing splines, with various extensions, is given in Wahba (1990). A summary of the basic methodology, with climate interpolation principally in mind, can be found in Hutchinson (1991a). Further developments intended for various geographic applications have been demonstrated by Gu and Wahba (1993) and Mitasova and Mitas (1993). More comprehensive discussion of the algorithms and the associated statistical analyses are given in Hutchinson and Gessler (1994) and Hutchinson (1993a, 1995a). Recent

comparisons with geostatistical (kriging) methods have been presented by Hutchinson (1993a), Hutchinson and Gessler (1994) and Laslett (1994).

Computational techniques for optimising λ and solving for f and β_j have been optimised for speed and storage space (Elden 1984, Bates *et al.* 1987). However, the number of data points is limited in general by the requirement to perform a tri-diagonalization of a full matrix with dimension slightly less than n. If there are more than a few hundred data points, it is preferable to solve the above minimization problem using a thin plate function f defined in terms of a restricted number of data positions or knots (Bates and Wahba 1982, Luo and Wahba 1997). The computations can then be arranged so that only matrices of order equal to the number of knots have to be stored and decomposed (Hutchinson 1984). This permits data sets with up to a few thousand points to be easily processed on quite modest computers, provided that the fitted spline function can be well described by no more than the knot limit of the spline program. Implementations of these techniques have been described by Hutchinson (1984) and Bates *et al.* (1987).

Errors in the model

In applying data smoothing to climate means, two components of the covariance structure of the ε_i in the model given by equation 3.3 should be recognised. The first component allows for deficiencies in the model being fitted. These are essentially due to local effects, including measurement error, which are below the spatial resolution of the observed point data network. The magnitude of these effects depends on the adequacy of the model represented by the function z_i and on the spatial density of the data. These effects can be reasonably assumed to be independent between different locations.

The second component arises when using serially incomplete data to interpolate monthly means for a standard period. This component can be ignored for some variables such as temperature. Monthly means of temperature settle down after a few years (in the absence of climate change). However, this component is significant for rainfall which normally exhibits large variation from year to year. This variability gives rise to strong correlations between the error components of rainfall means at stations with common periods of record.

Hutchinson (1995a) shows how the resulting correlated error structure can be incorporated into the interpolation process. The model permits the rigorous estimation of this smooth function by a thin plate smoothing spline in two ways. The statistical error structure of the data can either be accommodated directly, by using the corresponding non-diagonal error covariance matrix, or observed means can first be standardised to long term estimates using linear regression. Both methods gave similar interpolation accuracy and error estimates of the fitted surfaces were in good agreement with residuals from withheld data (Hutchinson 1995a).

Advantages of thin plate smoothing splines

There is a formal inclusion of thin plate splines within kriging (Wahba 1990). Thin plate spline techniques and the method of kriging extend easily to higher dimensions, and both methods attempt to achieve minimum error (optimum) interpolation. The order of the drift for kriging corresponds to the order of the derivative in the roughness penalty for splines. A non-zero nugget effect with kriging corresponds to imposing data smoothing with splines. The semi-parametric models are the analogue of kriging with external drift, except that the local nature of kriging means the external drift is not a constant. The different formulations and underlying formal connections of splines and kriging are well documented (Wahba 1990, Seaman and Hutchinson 1983, Cressie 1991, Hutchinson 1993a, Laslett *et al.* 1987, Laslett 1994, and Hutchinson and Gessler 1994). They suggest that when the semivariogram model in kriging is well chosen, and the order of the polynomial for the drift in kriging and in the roughness penalty with splines are consistent, the two methods can produce equally good results.

While Kriging and thin plate spline methods can give results of similar accuracy, there is a practical difference between them. Kriging requires the spatial covariance function or variogram be estimated first, which is critical to the process. This covariance structure can be difficult to estimate and validate. The method can be hampered by *ad hoc* assumptions about the form that the variogram should take, and there are computational difficulties in assessing the merit of the different functional forms (Hutchinson 1991a). Thin plate smoothing splines, on the other hand, require the estimation of a smoothing parameter that determines an optimal balance between fidelity to the data and smoothness of the fitted spline function. As this is normally done totally automatically by minimising the GCV, the thin plate smoothing splines are easier to use. Moreover the GCV offers a direct estimation of prediction error which is relatively free from parametric assumptions. This ease of operation does not sacrifice accuracy. The data smoothing paradigm embodied in splines permits efficient detection of data errors (Hutchinson 1993a, Hutchinson and Gessler 1994).

Thin plate splines display two advantages from the point of view of robust calibration. The first is that the underlying spatial covariance functions which give rise to thin plate splines have just two parameters. This is essentially a consequence of the fact that thin plate splines are unchanged by an arbitrary uniform scaling of the independent variables (Hutchinson 1993a). The scale dependent variogram models commonly used in kriging have an additional parameter defining the range of the variogram. This parameter is the least well determined by data (Dietrich and Osborne 1991). The second feature is that spline calibration methods do not depend directly on the adequacy of a stationary covariance function. In many applications such a model is a poor one and the validity and performance of standard covariance calibration methods are compromised. The method of minimising GCV directly addresses the predictive accuracy of the fitted surface and appears to be less dependent on the veracity of the underlying statistical model (Wahba 1990).

Lam (1983), in her comprehensive review of methods of spatial interpolation, has noted that no single method may be singled out as 'best'. Rather, the selection of an appropriate interpolation method depends largely on the type of data, the degree of accuracy desired, and the amounts of computational effort afforded.

Splines are more than just a smooth interpolator. They are an attractive procedure for interpolating spatial data and estimating the variance of the noise and expected true mean square error over the data points. When interpolating using thin plate smoothing splines, the precise form of the dependence of the data on physical factors is determined by the data themselves and not by pre-determined relationships as is the case with physical models (Hutchinson 1989). Interpolation using thin plate smoothing spline does not require prior standardization of unequal record lengths. The data can be weighted according to the amount of variation about the long term mean value of the variable at each station and the length of record at each station. This allows for a much larger network of stations to be used. Thin plate smoothing spline models may be validated individually or comparatively using several statistical techniques.

The ANUSPLIN computer package

The ANUSPLIN package (Hutchinson 1991a, 1994) is a suite of FORTRAN programs for the application of the thin plate smoothing spline technique. The package contains programs (SPLINA and SPLINB) for fitting surfaces to noisy data as functions of one or more independent variables (see Figure 3.1). The package also contains programs for interrogating the fitted surfaces in both point and grid form (LAPGRD and LAPPNT). It also includes programs for calculating standard error surfaces (ERRGRD and ERRPNT). ANUSPLIN also includes diagnostic programs (AVGCVA and AVGCVB) which calculate the GCV for each surface and the average GCV for a range of values of the smoothing parameter. These values are written to a file for inspection and for plotting.

The package can satisfactorily handle higher dimension splines, such as elevation or distance to the sea, in addition to the usual independent variables latitude and longitude. ANUSPLIN can allow for the incorporation of a linear parametric sub-model. An advantage of the partial spline formulation is that the coefficient of the linear sub-model (e.g. a lapse rate) is determined automatically from the data so that it does not have to be specified beforehand.

The surface fitting approach is described in the next section and examples are given in Chapter 4. Appendix B gives a description of all the components of ANUSPLIN used in the production of this thesis.

Fitting surfaces with SPLINA and SPLINB

SPLINA (ver3.1) and SPLINB (ver3.1) are programs which fit a partial thin plate smoothing spline to multi-variate noisy data. The package was primarily developed for fitting climate surfaces and applications to climate interpolation have been described by Hutchinson and Bischof (1983), Hutchinson *et al.*(1984), Hutchinson (1989, 1991a, 1991b).

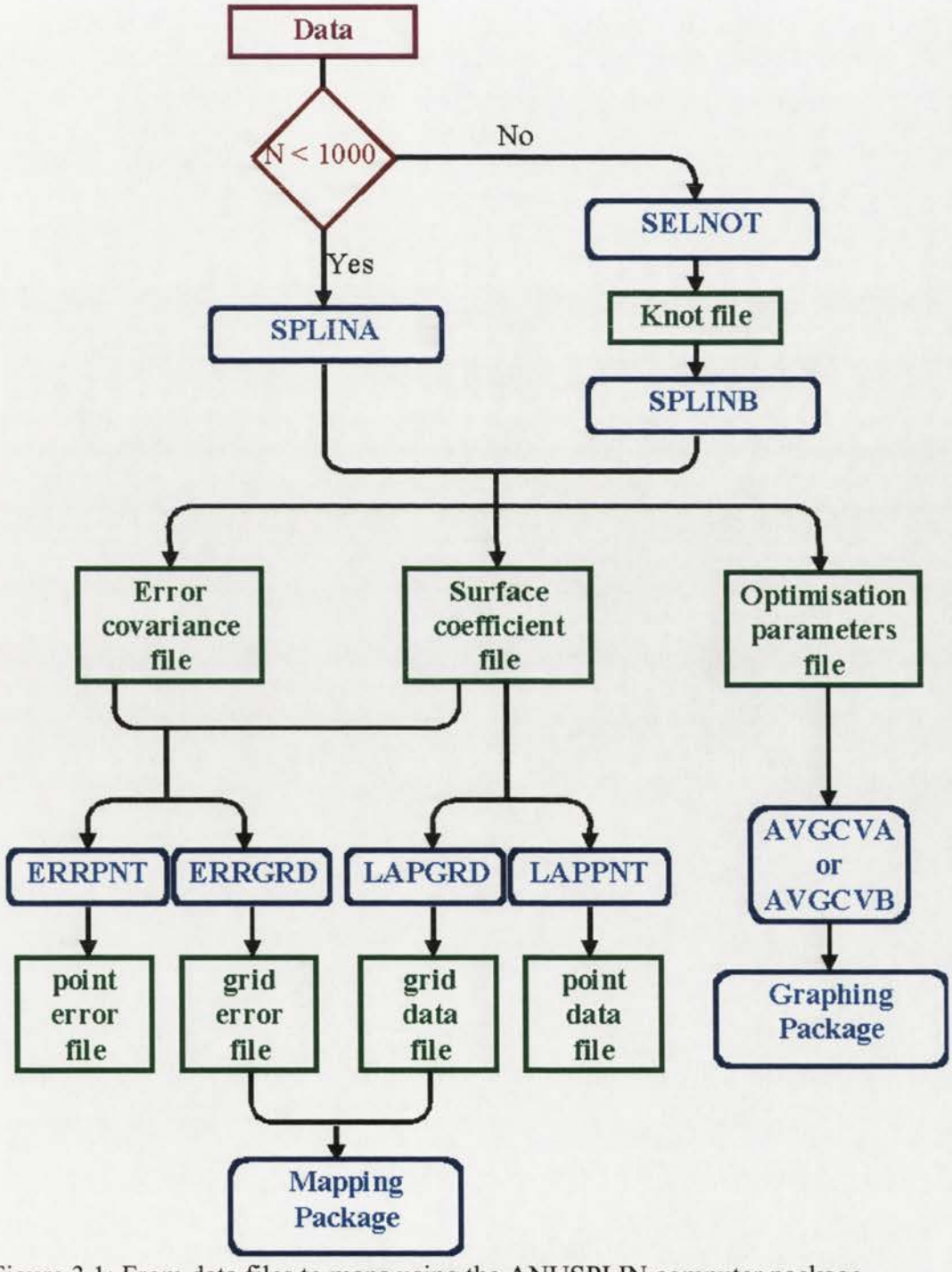


Figure 3.1: From data files to maps using the ANUSPLIN computer package.

For SPLINA, the main storage requirements of the program are proportional to the square of the number of data points and the processing time is proportional to cube of the number of data points. For SPLINB, storage requirements are proportional to the square of the number of knots and processing time is approximately proportional to the cube of the number of knots, since a singular value decomposition of a matrix of this order is required. Limits on numbers of data points for versions used were 1000 data points for SPLINA and 5000 data points and 1000 knot points for SPLINB, but larger limits are now possible.

The spline technique is an iterative procedure in climate surface construction which provides valuable information about the database and can be used to examine the effects of changing the station database. The fitting approach is illustrated by means of a flowchart in Figure 3.2. This is the simplest approach and is used as the first step in all time series interpolations in this study.

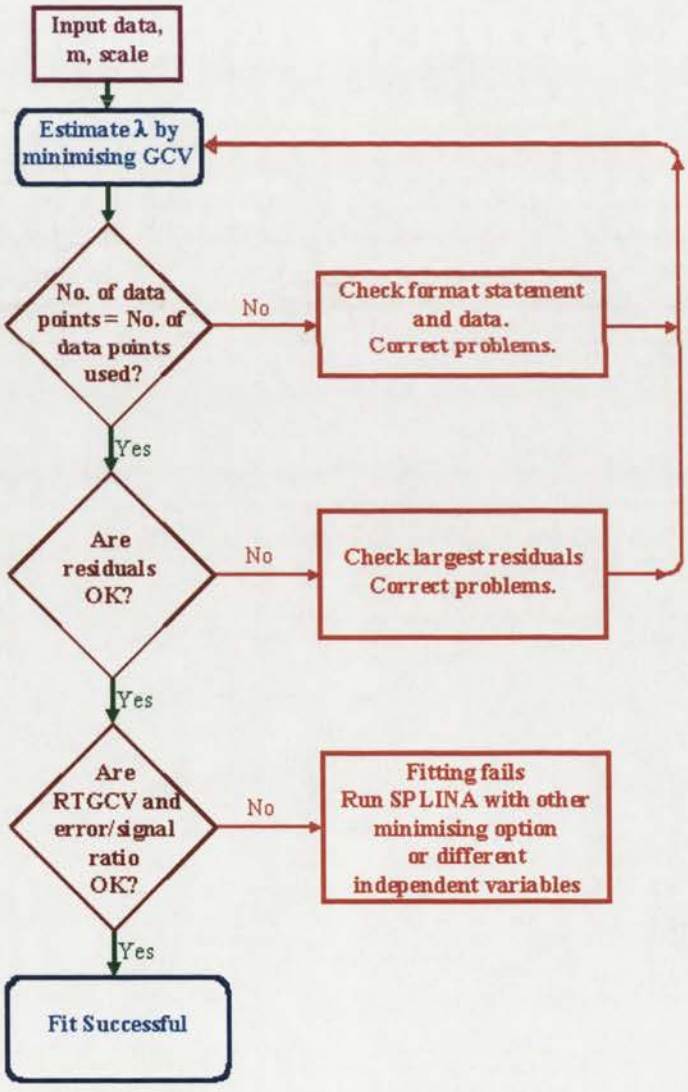


Figure 3.2: Flowchart of the SPLINA fitting approach.

The key features of SPLINA are best understood by examining the input data and options, and the output statistics and files.

Input data and options

To fit the surface, the values of the independent variables need to be known at the data points and should be accurately located in position and elevation. A single relative variance estimate for each data point (a uniform weighting of 1 for each data point for the monthly surfaces) is also required. Input options for SPLINA are shown in Table 3.1

Table 3.1: Input options and description for SPLINA and SPLINB.

Option	Notes
Number of independent spline variables	May not exceed specified limit (currently 10).
Number of independent covariates	Limit depends on the number of spline variables.
Independent variable lower and upper limits and transformation code	Data points outside these limits are ignored.
Transformation parameters	Required for non-zero transformation code and is useful for scaling elevation to appropriate units.
Order of derivative	Lower and upper limits specified by the program.
Number of surfaces	Make negative to stop list of data and fitted values, with estimated standard errors being written to output.
Common smoothing parameter option	For optionally using a common smoothing parameter when fitting multiple surfaces.
Optimization directive for each surface	0 - fixed smoothing parameter - supply value 1 - minimize GCV 2 - minimize true mean square error using supplied error standard deviation estimate 3 - fixed signal - supply value
Data file name	Must be supplied.
Data file format	Must be supplied.
Knot index file	File of indices of data points selected as knots (required only for SPLINB).
Optimization parameters file name	Blank if not required.
Surface coefficients file name	Normally required but may be blank if surface coefficients are not required.
Error covariance file name	Blank if not required. (NB only available in SPLINA ver 3.1)

There are normally at least two independent spline variables, longitude and latitude, which are usually in units of decimal degrees. Elevation above sea-level (in units of kilometres), is often appropriate when fitting surfaces to pressure, temperature or precipitation. This effectively exaggerates the effects of elevation by a factor of about 100. It may be included as a third independent variable or as a covariate. Other factors which influence the climatic variable may be included as additional covariates if appropriate parameterisations can be determined and the relevant data are available. These might include, for example, other topographic effects such as aspect (Hutchinson 1995a).

The lower and upper limits for each independent variable or covariate determine the range of the function. Coincident data points are permitted in SPLINB but they are not supported in SPLINA and terminate the program. Data points at positions which lie outside the supplied independent variable limits are rejected. This can be useful as a way of fitting a surface to a subset of the data without having to create a separate data file. It also gives a simple check on the specified data format and the order of the independent variables in the data file. An error in these specifications would be indicated if fewer than the expected number data points were selected. Transformation of independent variables available for SPLINA include: none, division, multiplication, log, exponential and tanh. Normally no transformations are performed on longitude and latitude. Minor improvements can sometimes be gained by slightly altering the scaling of elevation from kilometres when it is an independent spline variable.

The order of derivative to be minimized is usually two. Higher orders usually increase smoothness but are rarely appropriate when fitting climate surfaces. The number of surfaces options can be used to interpolate up to twelve surfaces at the same time. Running several surfaces at the same time allows for a common smoothing parameter obtained by minimising the average GCV over all surfaces. While the multiple surfaces option produces summary statistics for each surface, the ranked residuals list gives a single value for all monthly surfaces. Due to this factor, the interpolation for each month was done individually so as to produce a ranked residuals list for each data file.

In the normal use of ANUSPLIN, the input error variance is not needed and the fitted surface is uniquely determined. However if an error variance is supplied ANUSPLIN minimises the minimum mean square error to estimate the smoothing parameter.

Output statistics

The SPLIN programs always produce a diagnostics file which includes summary statistics (see Table 3.2a) and a list of the 100 largest residuals, ranked in descending order (see Table 3.3). Depending on values of the user directives, a list of the data and fitted values with estimated standard errors may be written to output (see Table

3.4). Table 3.2a shows summary statistics for the July 1974 pressure surface and an explanation of the statistics are given in Table 3.2b.

The signal is an estimate of the effective number of parameters in the fitted model and the error is an estimate of the effective degrees of freedom of the residual sum of squares of the fitted model. These together, add to the number of data points. The surface fitting procedure is normally considered to have failed to find a genuine optimum value of the smoothing parameter rho if either rho is very small and the signal is the maximum possible (equal to the number of knots) or rho is very large and the signal is the minimum possible (a number which depends on the number of independent variables and the order of the derivative). Both of these conditions are flagged by an asterisk in the output. The signal to error ratio is a good initial data check as flagged months were often found to have missing data values, which had been read as zeroes.

The generalized cross validation (GCV), mean square residual (MSR), the error variance estimate (VAR) and the mean square error (MSE) are given together with their square roots (RTGCV, RTMSR, RTVAR, RTMSE) which are in the units of the data. The RTGCV can be interpreted as the root mean square validation error, which includes measurement error in the data. The root mean square error (RTMSE) is an estimate of the true error of the fitted surface after the effects of measurement error have been removed.

The MSR is weighted according to the relative variance estimates d_i as provided in the data file. For the GCV calculation these relative variances are rescaled to have average value 1 in order to facilitate comparisons of GCV values across different models. If the relative variance estimates are actually estimates of the absolute value of the error variance (so that $\sigma^2 = 1$), then the variance estimate provided by the procedure should be approximately 1. The goodness of fit of the fitted model may be checked by comparing the residual sum of squares (n*MSR where n is the number of data points) with the critical points of a σ^2 times a chi-squared variable with *df* degrees of freedom, where *df* = the error degrees of freedom and σ^2 is an *a priori* estimate of σ^2.

When an estimate of σ^2 is available an alternative strategy is to provide the corresponding standard deviation estimate σ to the program. The program then minimises an unbiased estimate of the true mean square error instead of the GCV. This is not normally recommended since it depends on having a reasonably accurate estimate of σ^2. It is rare that a true *a priori* estimate can be made as it incorporates an unknown interpolation error. It is generally preferable to minimise GCV, since this appears to be more robust and does not depend on knowing σ^2. The *a priori* estimate of σ^2 can then be used to check the goodness of fit of the model as described above. Specifying the error variance may be preferable when fitting surfaces to very small data sets (less than 20-30 data points).

When fitting a series of surfaces it is possible to specify RTVAR as the standard deviation for a problem file, that is, the RTVAR from other files in the series are used.

Table 3.2a: Summary statistics from the diagnostics file, all stations, pressure, July 1974. All statistics except error and signal in hPa.

RHO	ERROR	SIGNAL	MEAN	STD DEV
0.272E+01	36.8	18.2	999.173	22.280

GCV	MSR	VAR	RTGCV	RTMSR	RTVAR
0.116E+01	0.517E+00	0.774E+00	0.108E+01	0.719E+00	0.880E+00

MSE	RTMSE	COVARIATES	COVARIATE ERROR
0.257E+00	0.507E+00	-0.114E+03	0.5

Table 3.2b: Explanation of summary statistics.

Output statistic	Quantity	Notes
RHO	smoothing parameter x n	
ERROR	degree of freedom for noise	gives the proportion of estimated error variance due to residuals
SIGNAL	degree of freedom for signal	characterises the degree of complexity of the fitted surface
MEAN	mean of Z	
STD DEV	standard deviation of Z	
GCV	general cross validation	mean square predictive error
MSR	mean-square residuals	
VAR	estimated error variance	
RTGCV	square root of GCV	
RTMSR	square root of MSR	
RTVAR	square root of VAR	
MSE	expectation of the predictive mean-square error [TMSE] see equation 3.8	Error variance contributed by the error of the fitted surface
RTMSE	square root of MSE	
COVARIATE	β	lapse rate
Actual and fitted values and std error		Tabulated Used to check data
Ranked rms residuals	root-mean square residuals	Ranked and tabulated Used to check data

SPLINA also outputs the coefficients of any covariates and standard errors of these coefficients as well as an estimate of the true mean square error of the smoothed data values. This estimate depends on the value of σ^2 as estimated by the procedure (VAR) or the input error standard deviation estimate when this has been provided by the user.

The ranked list of residuals (see Table 3.3) is particularly useful in detecting miscellaneous data errors. Errors in these locations can often be indicated by large values in the output ranked residual list. The list of data and fitted values (see Table 3.4) is also useful for this task, especially when fitting a surface to a new data set.

Table 3.3: Highest ten ranked root mean square residuals (from stations), all stations, pressure, July 1974.

Rank	Station No.	RTMSR (hPa)
1	34	3.30
2	9	1.84
3	11	1.76
4	33	1.10
5	37	1.07
6	25	1.05
7	47	1.04
8	35	0.95
9	4	0.70
10	24	0.64

Table 3.4: List of the data and fitted values and standard errors for the first and last three stations in the data file, all stations, pressure, July 1974.

station no.	longitude °E	latitude °S	elevation (km)	actual pressure (hPa)	fitted pressure (hPa)	standard error (hPa)
1	127.67	18.23	0.424	968.34	968.16	0.56
2	122.23	17.95	0.009	1014.64	1014.78	0.53
3	123.67	17.37	0.012	1014.31	1014.41	0.53
73	144.75	37.87	0.014	1010.60	1010.12	0.42
74	147.50	42.83	0.003	1006.40	1006.67	0.56
75	147.33	42.88	0.057	1000.19	1000.51	0.57

Output files

The production of all output files is optional. They are shown in Table 3.5.

Table 3.5: Output files from SPLINA and SPLINB.

Output files	**Notes**
Surface coefficients file	Fitted surface coefficients stored in a file.
Error covariance file	Used with ERRGRD and ERRPNT to calculate error values
Optimisation parameters file	For post processing by AVGCVA and AVGCVB to calculate GCV values as a function of different smoothing parameter values (see Figure 3.1)

The fitted surface is stored compactly as a set of surface coefficients describing the fitted function of the independent variables (usually longitude, latitude and elevation). To calculate a regular grid of fitted climate values, for mapping and other purposes, a regular grid of each independent variable is required. The surface coefficients are used to calculate values of the fitted surface by other programs in ANUSPLIN, including LAPGRD and LAPPNT. These fitted surfaces can be for a fixed elevation (e.g. MSL), or if values are required on the earth's surface an appropriate DEM must be used.

The covariance file contains the covariance matrix of the parameters of the fitted surface. It is used by ERRGRD and ERRPNT to calculate spatially distributed standard errors of the fitted surface. Examples of the output are given in Chapter 4.

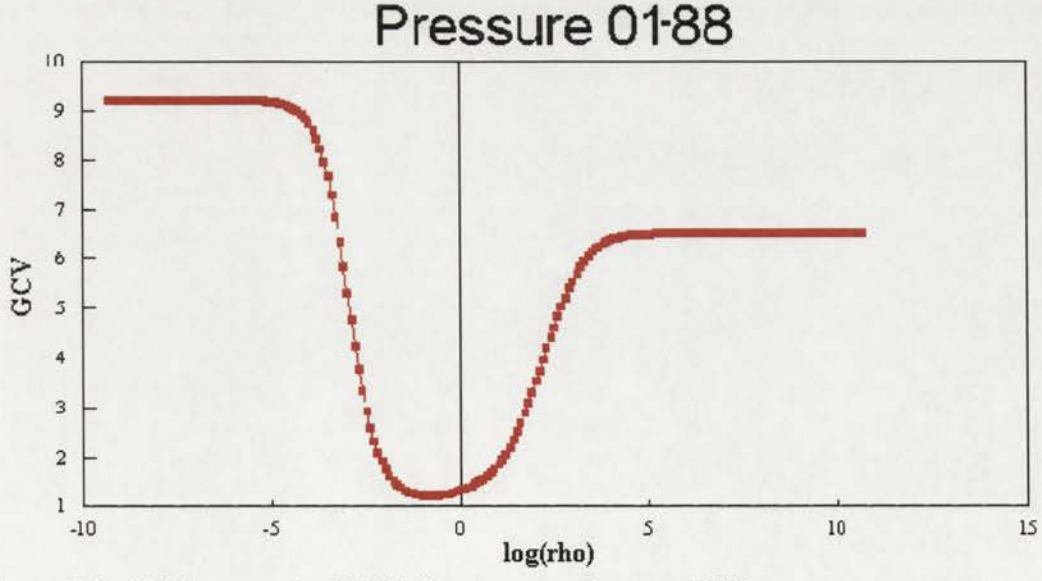


Figure 3.3: GCV versus log RHO for pressure, January 1988.

The file containing parameters used to calculate the optimum smoothing parameter is used by AVGCVA to graph GCV as a function of the smoothing parameter. Figure

3.3 shows an example of the output from AVGCVA. SPLINA normally finds the 'smoothest' local minimum (i.e. the largest value of rho if there are multiple minima).

From surface coefficient files to maps

The procedure, as shown in Figure 3.1, used to produce maps can be applied to any month within the period of acquired monthly data from 1952 to 1990. Coefficients of the surfaces fitted to the monthly values are stored by month. They can be used to calculate values of the climate variable for any month at any site using the LAPPNT program in the ANUSPLIN package. Alternatively, regular grids across Australia of monthly climate variables for any month can be calculated using the LAPGRD program in ANUSPLIN and displayed by any of a number of commonly available computer programs which can display raster images. These procedures could clearly be used to set up an operational procedure for interrogating and displaying the spatial distribution of monthly climate data. The system could be updated to incorporate later years as new data became available.

The maps were produced by calculating the corresponding fitted surfaces on a regular 0.1 degree longitude by 0.1 degree latitude grid and displaying the grid using the GMT plotting package (Wessel and Smith 1991).

ANUSPLIN assessment

Determination of the choice of various SPLINA and SPLINB input options

Input options that would effect the outcome of the process include the choice of the number of independent spline variables versus the independent covariates, the choice of limits and transformation code, the order of the derivative, and the optimisation directive. The advantages of running the spline function by minimising the GCV have been extensively covered and the minimum GCV option was considered as the most appropriate method of determining the surface coefficients. It was the first choice optimisation directive for all surfaces.

Longitude and latitude are normally independent spline variables, but elevation above sea level can either be a third independent variable or a covariate. The choice of limits was determined by the geographical extent of the Australian continent. The transformation code or scale was set to 0 (no transformation) for longitude and latitude. Preliminary computer runs for assessing the independent variable versus the covariate options for altitude were done for various scale transformations. The surfaces were run with the altitude data in metre transformations and examples are given in Table 3.6 for January 1952, 1968 and 1989 pressure surfaces for both the spline and partial spline models. The transformations have no effect on the partial spline models so the data are only entered on the kilometre scale (m/1000).

Apart from the one and ten metre scales for 1952 only minor improvements were obtained by altering the scaling of elevation when it was an independent variable. Values of the RTGCV are very similar for the range of 100 metres to two kilometres and very little change in the interpolation accuracy is achieved by changing the scale within this range. The partial spline gave very similar results to the spatially varying spline model for the later years, and performed marginally better for the earlier years which had a smaller station network. These results are most likely due to the greater vertical representation in the later years, and suggest that the station network of the earlier years did not justify a spatially varying lapse rate. It is interesting to note, however, that the RTGCV decreases for the partial spline as the station network density increases. The tri-variate spline may be superior with a denser station network or for shorter time scales (e.g. daily values) when there are broad latitudinal or other gradients in the temperature lapse rate.

Table 3.6: RTGCV for January 1952, 1968 and 1989 pressure surfaces for both the spline and partial spline models, for various scales of elevation.

	Jan 1952 (36 stations) RTGCV		Jan 1968 (51 stations) RTGCV		Jan 1989 (109 stations) RTGCV	
Scale	**covariate**	**indep. spline**	**covariate**	**indep. spline**	**covariate**	**indep. spline**
m		1.93		0.89		0.79
m/10		1.42		0.87		0.64
m/100		0.94		0.84		0.58
m/1000	0.91	0.95	0.82	0.84	0.57	0.53
m/2000		0.95		0.84		0.59
m/5000		0.95		0.84		0.59

An advantage of the partial spline model is that there is no need to determine the most appropriate scale. The incorporation of elevation as an independent spline requires a suitably dense network of stations as the lapse rates are local or spatially varying, whereas the partial spline model does not require as dense a network of stations and a single lapse rate is calculated automatically. The interpolation accuracy achieved with these models was better, or only slightly inferior to that obtained by incorporation of a spatially varying dependence on elevation. The results from the temperature surfaces showed similar results.

Many previous rainfall studies (e.g. Chua and Bras 1982 and Phillips *et al.* 1992), have shown that the incorporation of a constant linear dependence on elevation produces a more accurate result than having no elevation dependence. Hutchinson (1995a), however, has shown that errors were significantly reduced by using a tri-variate thin plate spline which incorporated a spatially varying dependence on elevation. These results are to be expected as spatial rainfall patterns are more complex than pressure or temperature patterns.

Similar tests were done for the order of the spline function with altitude scaled in kilometres. Second to fifth order examples are given in Table 3.7 for both the spline and partial spline models for January 1952, 1968 and 1989.

The order of derivative which minimized the GCV was usually two. Higher orders can sometimes increase smoothness, but as can be seen in Table 3.7, they only marginally improved the interpolation accuracy and on occasion significantly decreased interpolation accuracy. Order two was therefore chosen for all analyses.

Table 3.7: RTGCV for January 1952, 1968 and 1989 pressure surfaces for both the spline and partial spline models, for various order of spline models.

Order of spline	Jan 1952 (36 stations) RTGCV		Jan 1968 (51 stations) RTGCV		Jan 1989 (109 stations) RTGCV	
	covariate	indep. spline	covariate	indep. spline	covariate	indep. spline
2	0.91	0.95	0.82	0.84	0.57	0.59
3	1.05	0.88	0.79	1.90	0.56	0.48
4	1.02	0.96	0.81	2.14	0.55	0.48
5	1.01	1.69	0.91	3.46	0.55	0.49

As a result of these tests the pressure and temperature surfaces were generated by a second order partial spline function as they are superior for small data sets. More complex spline models did not produce a significant improvement in the interpolation accuracy with larger data networks.

Visual assessment

Visualization is a useful option for examining and validating observational data and the output of environmental models. Initial tests on the usefulness of the output from the ANUSPLIN programs were performed by visual comparison with Bureau of Meteorology maps .

Daily pressure

The first test was to assess the ability of ANUSPLIN to represent the pressure daily pattern in comparison to daily maps produced by the Bureau of Meteorology. Comparisons were undertaken on the 15th day of the mid-season months for 1968 and 1988.

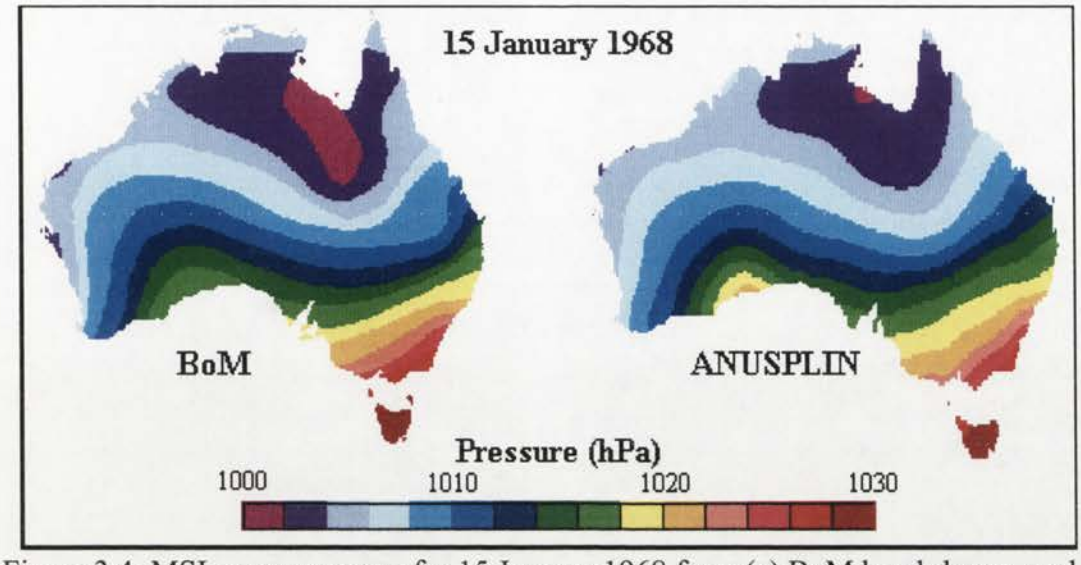


Figure 3.4: MSL pressure maps for 15 January 1968 from (a) BoM hand-drawn and (b) SPLINA.

Figure 3.4 shows that the ANUSPLIN program output for 15 January 1968 compared well with the Bureau of Meteorology maps. Differences were noted between the maps which were possibly due to the ANUSPLIN maps (234 stations) being produced with a slightly different set of stations than those used to produce the hand drawn Bureau of Meteorology maps (unknown number of stations). Overall there was very good agreement between the two sets of maps.

Bureau of Meteorology rainfall maps

The ability of the ANUSPLIN programs to reproduce rainfall decile maps was also tested with some seasonal data from the Bureau of Meteorology. The summer of 1993/4 rainfall deciles data and various output versions were provided by The Environment Resource Information Network (ERIN) and the Bureau of Meteorology. Figure 3.5 shows the location of the stations coloured in the decile values for the summer of 1993/4. It is interesting to note that while there is some spatial coherence, adjacent stations in the Northern Territory can have rainfall deciles at opposite ends of the scale.

Unfortunately there were some difficulties associated with the Bureau of Meteorology's decile values of 0 for a station with the driest summer on record and 11 for the wettest summer on record. This method of determining deciles produces a marked skew effect as there can be many years with no rainfall for a specific month and thus a zero decile, while few records are likely to be the highest on record or decile11. When these values are spatially interpolated they tend to lower or raise values for nearby stations and give false interpolated values.

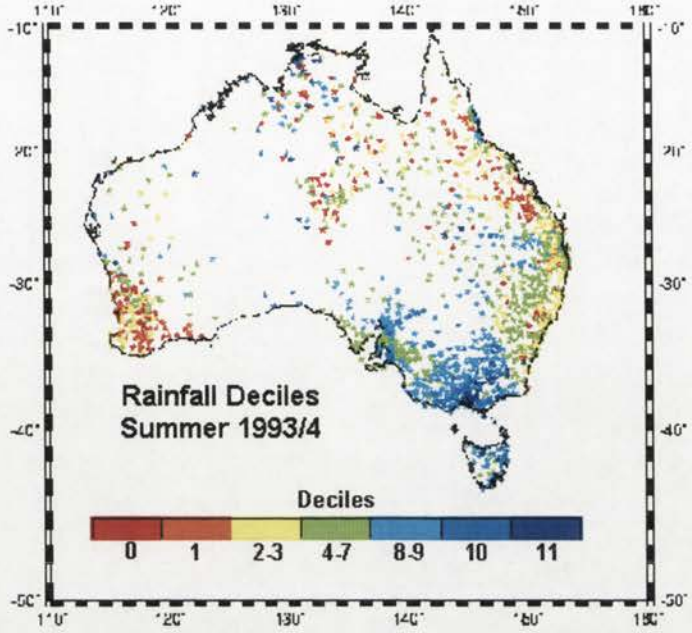


Figure 3.5: The location of the stations coloured in the decile values for the summer of 1993/4.

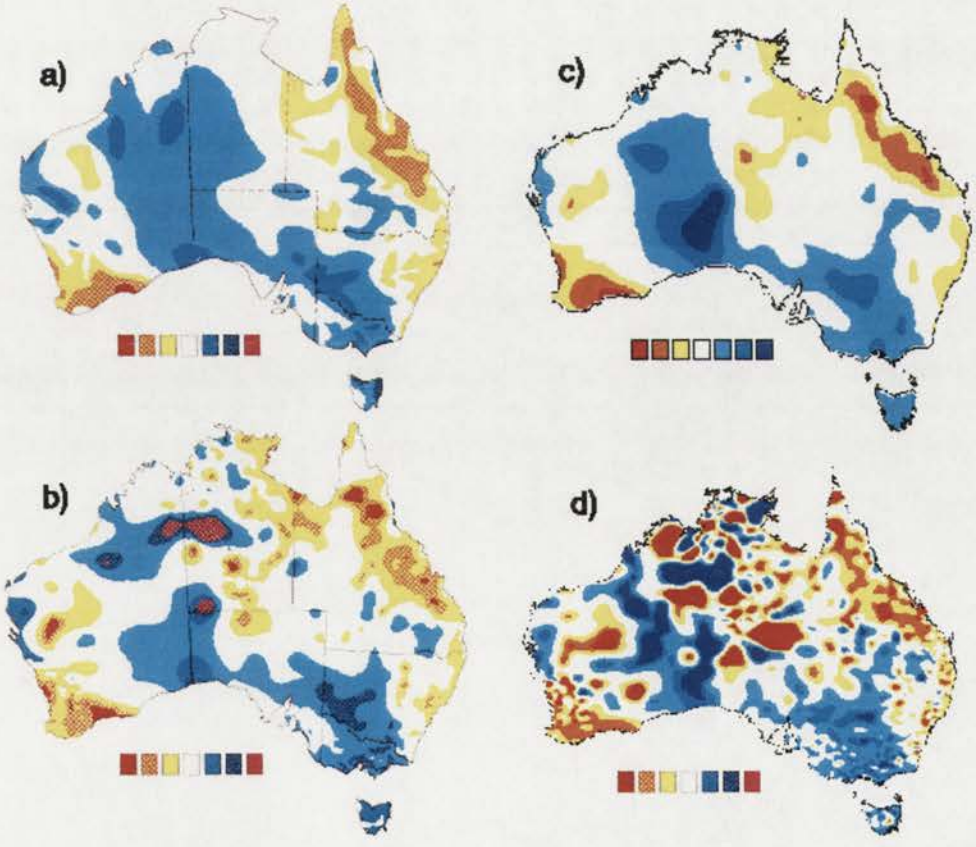


Figure 3.6: Decile values for the summer of 1993/4 (a) BoM manually analysed maps based on district averages and selected stations derived from telegraphic reports, (source: Bureau of Meteorology); (b) MAPGEN GIS package output, (source: Bureau of Meteorology); (c) ANUSPLIN package output, and (d) ARCINFO GIS package output, (source: ERIN).

Figure 3.6 shows the hand drawn, BoM-preferred map, and also the outputs from MAPGEN and ARCINFO spline functions. The ANUSPLIN output was produced with the default values with a second order spline. If the BoM hand drawn map is the test base, the ANUSPLIN output performs better than the other methods in estimating sensible regional values due to the data smoothing applied. The ARCINFO and MAPGEN outputs are more exact interpolation methods in that they have little data smoothing and station values are fitted more closely.

Actual versus interpolated values

A simple method to assess the value of the spline function is to evaluate how the actual station values compare to the interpolated values at the station. Figure 3.7 shows the actual values plotted against the interpolated values at the station height. While the above example shows that a close fit doesn't necessarily mean a good model, it is a useful test to show that the smoothing doesn't distort the station values too much.

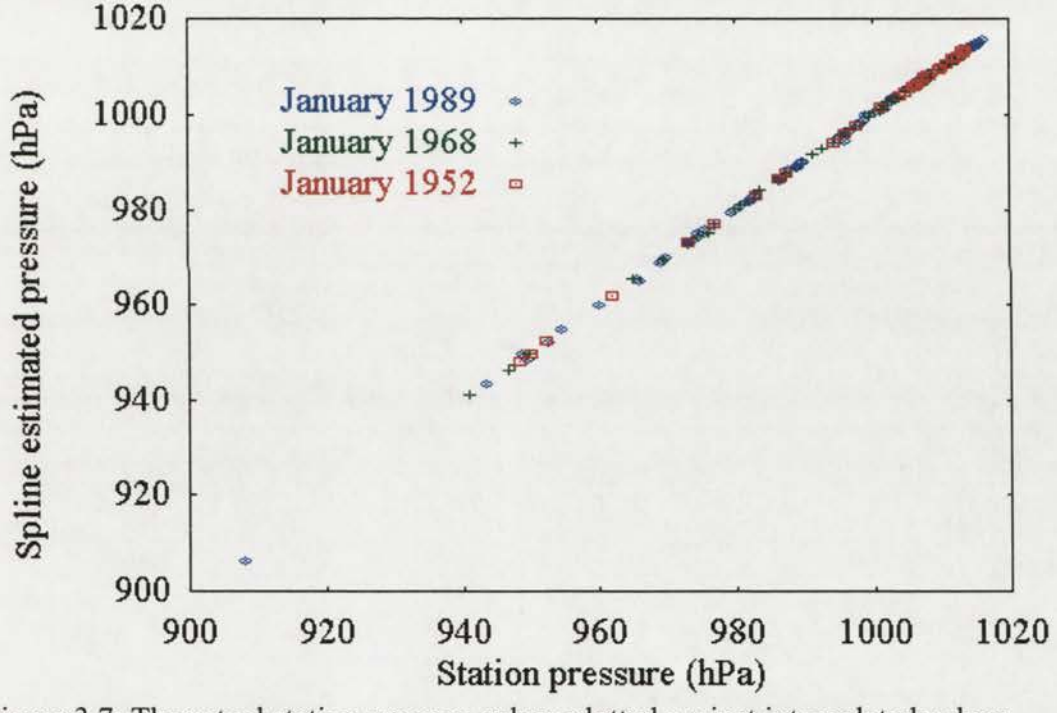


Figure 3.7: The actual station pressure values plotted against interpolated values.

Figure 3.8 shows the difference between the actual and interpolated values plotted against altitude. This figure shows that the spread is quite evenly distributed over all represented altitudes. That is, the spline function does not systematically overestimate or underestimate low or high altitudes.

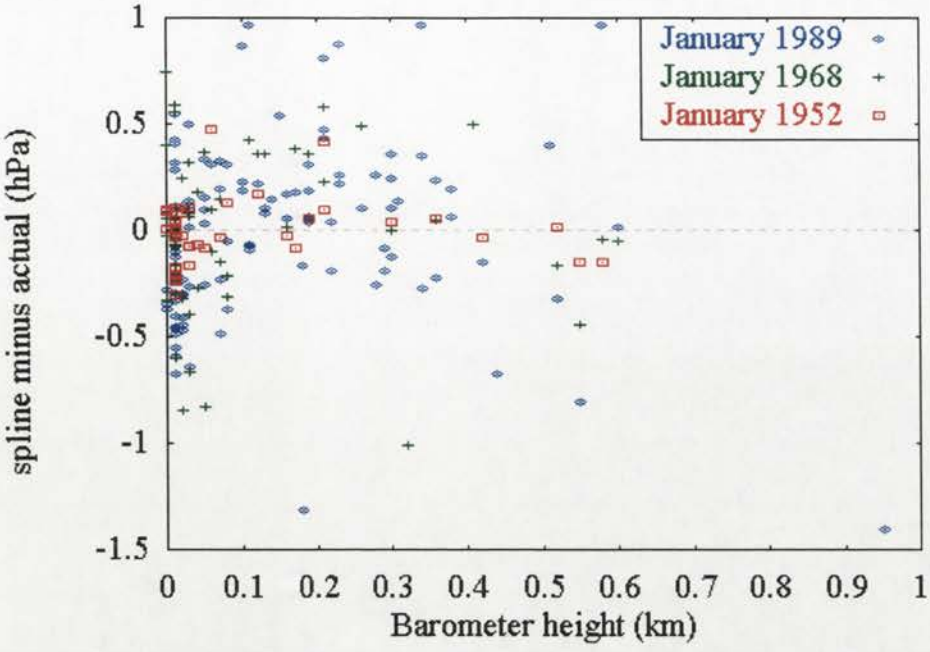


Figure 3.8: The difference between the actual and the interpolated pressure values plotted against barometer altitude.

For the temperature variables, the average monthly deviation of the fitted surface from the measured climate station data was never more than ± 2.5°C; at least 90% of all deviations were less than ±0.5°C. Table 3.8 shows the percentage of the deviations of the fitted temperature surfaces from the station data.

Table 3.8: Distribution of the deviations of the fitted temperature surfaces from the station data (% of total data points).

Deviation (°C)	**-2.5 →-1.0**	**-1→ -0.5**	**-0.5→0.5**	**0.5→1.0**	**1→2.5**
dry-bulb 24,381 points	0.9%	3.2%	91.8%	3.5%	0.7%
wet-bulb 23,969 points	0.1%	1.9%	95.5%	2.1%	0.2%
dew-point 22,846 points	0.4%	3.6%	92.5%	3.0%	0.4%

Conclusions

Partial thin plate splines appear to be well suited to the interpolation of meteorological fields. The dependence of the fitted data on the estimated function is determined from the data themselves and there are no unnecessary restrictions imposed on the form of the fitted function. The thin plate smoothing spline technique directly addresses two major aspects of the general interpolation problem, namely the statistics of the data errors and the roughness or complexity of the underlying

meteorological field. The data are smoothed according to objective error minimizing criteria, and automatic minimisation of the GCV makes the procedure easy to use and robust in the presence of minor misspecifications of the model.

Thin plate smoothing splines therefore provide an interpolation technique with an algorithmic structure that is objective and straightforward to use. The method is not difficult to implement in higher dimensions, and the inclusion of appropriate independent variables in addition to the usual position variables of longitude and latitude can add significantly to the accuracy of the analyses in data sparse areas. In particular, elevation has played a major role in producing accurate continent-wide interpolations of monthly mean temperature and rainfall. The ability of partial thin plate splines to blend empirical and physical process-based aspects of spatial variability of meteorological fields makes them particularly appealing.

The method works well with irregularly spaced data and does not generate spurious effects in areas away from data points. Standardization of records to a given period is not required and variance estimates may be used to check the validity of the model. The data smoothing process also permits the use of short period means, which allows for the densest possible network. Regular derivatives of all orders are available so that direct analysis of slope and aspect can be performed.

The method is shown to be statistically very sensitive to spatial trends. It also allows for the investigation of outliers in the output statistics and for an analysis of the error. This then enables sensitivity studies and allows for the assessment of station networks. These points are discussed further in Chapter 4.

Chapter 4. Fitting the Australian Climate Surfaces

Introduction

The estimation of meteorological variables where data are not measured is essential if efficient use is to be made of meteorological networks which are limited in extent by economic necessity. A surface is a function of two or more independent variables denoting geographic position. Monthly mean surfaces have been created using the spline techniques for all of Australia (Hutchinson 1991a), Africa (Hutchinson *et al.* 1993), northern Latin America, South America (Hutchinson and Corbett 1995), and parts of eastern Asia (Zuo *et al.* 1996). Systematic procedures for storing and interrogating these surfaces have been assembled in the ANUCLIM package (McMahon *et al.* 1995). These climate surface databases have served as the foundation for a range of bioclimatic analyses, from the analysis of Australian elapid snakes (Nix 1986) to the identification of tree-species adaptation zones (Mackey 1993).

The objective of the third step of the methodology introduced in Chapter 1 is to apply the spline technique for interpolating the spatial distribution of climate from ground-based point observations to further assess both the technique and the actual surfaces. Details are given of the methods used to generate estimates of continent-wide descriptions of monthly climatic variables as functions of geographic location and elevation. As the pressure and temperature data come from a different station network than the rainfall data network, the size of their respective networks require different ANUSPLIN options and are discussed separately.

The chapter begins with a description of the application of the surface fitting process discussed in the previous chapter. The 0000 UTC pressure surfaces are used to illustrate the iterative nature of the process, and to show the value of the spline analysis for assessing data homogeneity. This is followed by a discussion of the adequacy of sparse data networks and the errors associated with different station networks. The results from the standard error routines in ANUSPLIN are then discussed. The chapter also discusses problems associated with changes in spatial data coverage with reference to the computation of the "continental" average temperature through time.

Lapse rates in general, and the environmental lapse rate (ELR) produced by the surfaces are discussed. An examination of the results of the time series of lapse rates shows that there has been change in the temporal variability since the early 1970s, however this is not seen to be related to the change from Celsius to Fahrenheit scale, nor to the number of stations in the database.

The chapter concludes with a description of the procedure for fitting the rainfall percentile surfaces and the problems associated with these surfaces.

The monthly pressure and temperature surfaces

A limitation, at least for large area studies in Australia, is the lack of spatially distributed climate information sufficient to reliably estimate the spatial variation of pressure and temperature to some degree of accuracy. This is especially the case during the earlier years of the period. In keeping with other Australian studies (Hutchinson 1991a), a partial thin plate smoothing spline was fitted, with two independent position variables (longitude and latitude) and a single linear dependence on elevation for the pressure and temperature surfaces. The Bureau of Meteorology's 1990 station dictionary was used for station and station barometer elevation. The two independent spline variables allow for broad changes of the pressure and temperature regimes with respect to position, to be accommodated. Using altitude as a covariate takes into account temperature and pressure lapse rates. In each case the partial thin plate smoothing spline was optimised by minimising the GCV, and the coefficient of the linear sub-model (the lapse rate) was determined automatically.

Both the long term means (LTA) and the individual monthly means (AVG) were spatially interpolated, and a mean sea-level (MSL) surface extracted. These surfaces and the original station data formed the base data for the statistical analysis and for the models discussed in Chapters 5 and 6.

The monthly pressure and temperature surface construction

The surface construction began by using SPLINA to generate spline coefficients for each month for both 0000 and 0600 UTC from the database described in Chapter 2. Actual station elevations were used when fitting thin plate splines. An advantage of this is that access to a DEM was not required. The input options that were used for SPLINA are shown in Table 4.1.

Table 4.1: The input options that were used for SPLINA to produce the pressure surfaces.

NUMBER OF INDEPENDENT SPLINE VARIABLES:	2
NUMBER OF INDEPENDENT COVARIATES:	1
LIMITS AND TRANSFORMATION CODE FOR VARIABLE 1:	110 160 0
LIMITS AND TRANSFORMATION CODE FOR VARIABLE 2:	-50 -10 0
LIMITS AND TRANSFORMATION CODE FOR VARIABLE 3:	0 2500 1
ENTER 1 TRANSFORMATION COEFFICIENT(S):	1000
COMMON INPUT DIRECTIVE:	1 (minimise GCV)
ORDER OF SPLINE (AT LEAST 2):	2
NUMBER OF SURFACES (1 TO 12):	1
OPTIMIZATION DIRECTIVE FOR EACH SURFACE:	1

Input elevations were scaled to kilometres for consistency with other spline analyses. SPLINA does allow other transformations of the independent variables. Transformations possible at present are shown in Table 4.2

Table 4.2: Independent variable transformation codes for SPLINA.

code	transformation
0	none
1	x/a
2	x*a
3	a*log(x+b)
4	(x/b)**a
5	a*exp(x/b)
6	a*tanh(x/b)

Detecting data inhomogeneity and data errors

There are several useful output statistics (see Table 3.2) in the diagnostics file generated by the SPLINA program, which can be used to check for data homogeneity. The square root of the generalised cross validation (RTGCV) is a good measure of the predictive power of the fitted surface, especially when comparing the values from the same climate variable fitted for different times. In combination with the ranked residuals listing, which allows for each station to be assessed in relation to its deviation from the calculated surface, the RTGCV provides a powerful tool for investigation of station homogeneity. The example given in Figure 4.1 is a time series of the RTGCV for 0000 UTC MSL pressure derived from all stations in the database.

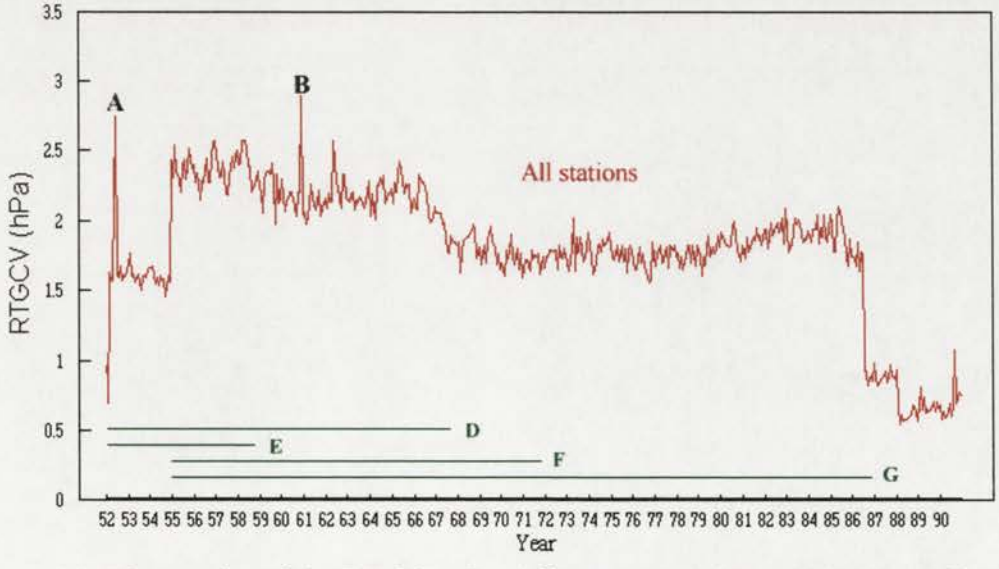


Figure 4.1: Time series of the RTGCV (hPa) for pressure 1952-1990, using all stations.

The five stations with the highest root mean square residuals in the diagnostics file for the analyses represented by points A (June 1952) and B (December 1960) and January 1982 in Figure 4.1 are given in Table 4.3

Table 4.3: Ranked root mean square residuals (hPa) for pressure June 1952, December 1960 and January 1982.

	June 1952		December 1960		January 1982	
rank	station	RTMSR	station	RTMSR	station	RTMSR
1	086038	9.58	086071	7.63	086071	8.11
2	009034	4.30	027022	5.00	040214	2.98
3	009021	2.84	009034	4.08	094029	2.25
4	017043	2.16	066062	3.91	094008	2.20
5	031011	2.12	031011	2.74	009021	2.19

Figure 4.2, below, shows a time series for the stations with root mean square residuals greater than 3.0 in Table 4.3. Some errors are simply positional. For example, the station dictionary describing the longitude, latitude, and elevation can contain errors and only gives the present station location. It is not uncommon for meteorological stations, especially regional offices, to be moved. The Graphs of stations 009034, 027022, 066062 and 086071 all show discontinuities in the time series, and all are consistently listed in the ten stations with the greatest residuals for the period of their discontinuities (lines D, E, F and G in Figures 4.2 and 4.3).

The graphs of stations 086038 and 027022 show that points A and B in Figure 4.2 are outliers, although the latter also shows a discontinuity in the time series. These outliers arose from reading or recording errors on single days. Closer inspection of the data files reveals that on the 22nd of June 1952 station 086038 had a 0900 LST pressure recording of 5.8 hPa (compared with its June average of 1009 hPa) and on December 25 1960, station 027022 had a 0900 LST pressure recording of 1.2 hPa (compared with its December average of 1003 hPa). When possible, errors are checked and corrected in the station database, it may however, be necessary to remove the station from the database. New coefficients are then generated. Figure 4.3 is a time series of the RTGCV for 0000 UTC MSL pressure with the above suspect data removed.

A comparison between Figures 4.1 and 4.3, shows the removal of the suspect data reduces the RTGCV from values around 2.0 hPa to values around 1.0 hPa. Figure 4.3 suggests that there are still two possible outliers at point C and possibly two stations with homogeneity problems for the periods marked by lines H and I. A comparison of the RTMSR shown in Tables 4.3 and 4.4 shows that some stations, such as 031011, have a much reduced error. Others, such as 040214, do not show the same improvement, and this suggests this station has data homogeneity problems and is possibly one of the stations suggested by H or I in Figure 4.3.

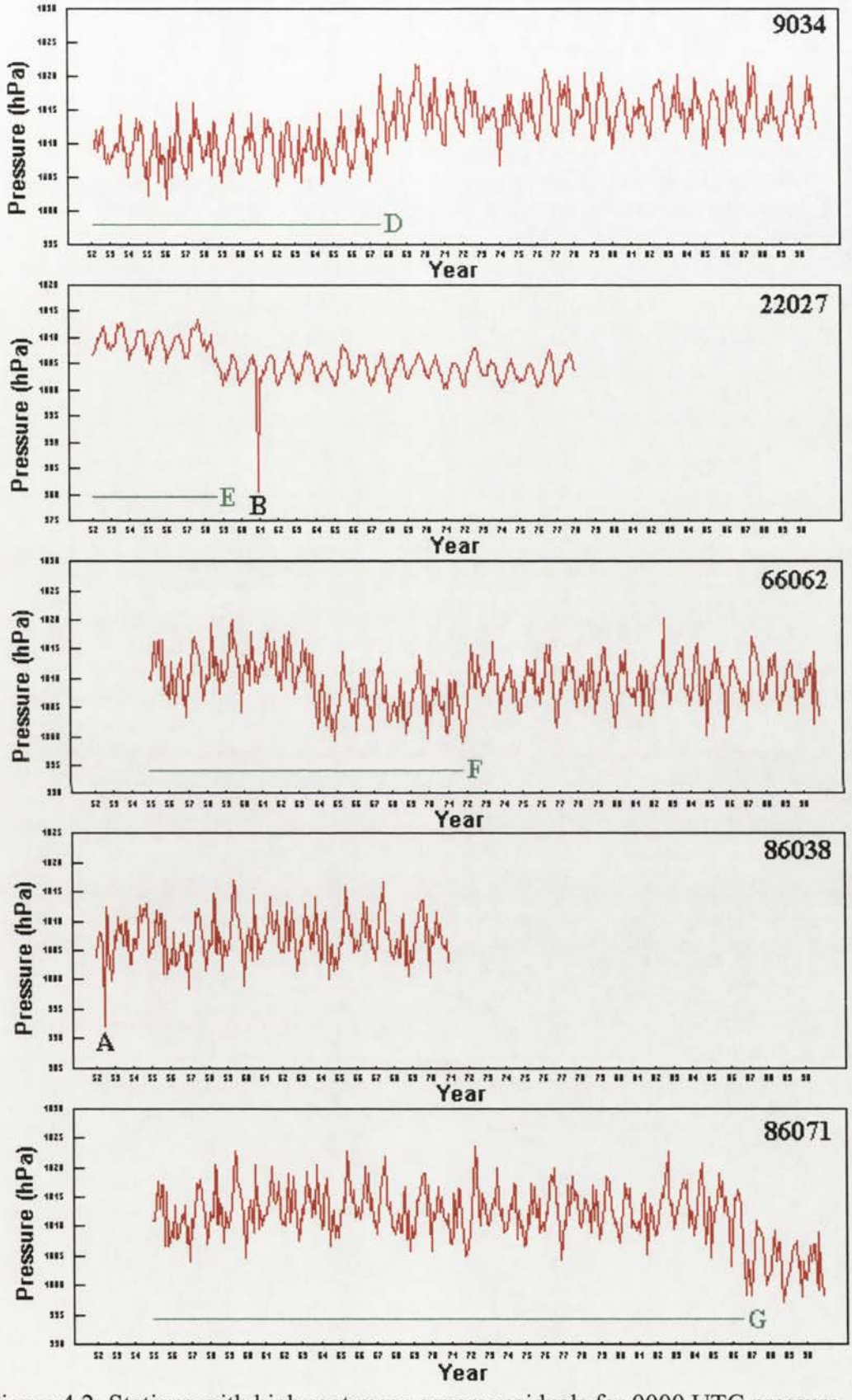


Figure 4.2: Stations with high root mean square residuals for 0000 UTC pressure 1952-1990.

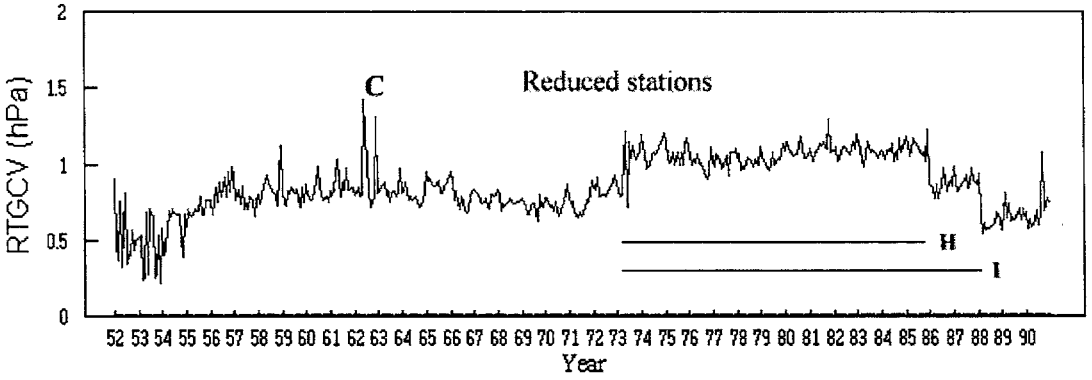


Figure 4.3: Time series of the RTGCV in hPa for pressure 1952-1990, for the reduced data set.

Removing the worst stations significantly reduced the RTGCV and the root mean squared residual. The signal did not change greatly, indicating robustness of the initial fit to the full data set. The iterative nature of this surface generation process permits careful examination of the effects of changing the station database.

Table 4.4: Ranked root mean square residuals (hPa) for pressure June 1952, December 1960 and January 1982 for surfaces created with the reduced number of stations.

	June 1952		December 1960		January 1982	
rank	station	RTMSR	station	RTMSR	station	RTMSR
1	061078	0.82	023000	1.16	040214	2.90
2	066037	0.73	009021	1.00	009021	1.86
3	094029	0.63	029009	0.93	063231	1.58
4	031011	0.57	085072	0.73	040223	1.19
5	017043	0.54	078031	0.70	059040	1.02

This process of assessing each station in relation to its deviation from the calculated surface and correcting errors may be repeated many times, steadily reducing overall surface error. Stations H (009021) and I (040214) in Figure 4.3 were removed from the data set and Figure 4.4 shows the time series of the RTGCV for the data set from the final iteration. This data set was used to produce the surfaces in the atlas (see Appendix A) and for the statistical analysis presented in this chapter and the modelling presented in Chapters 5 and 6.

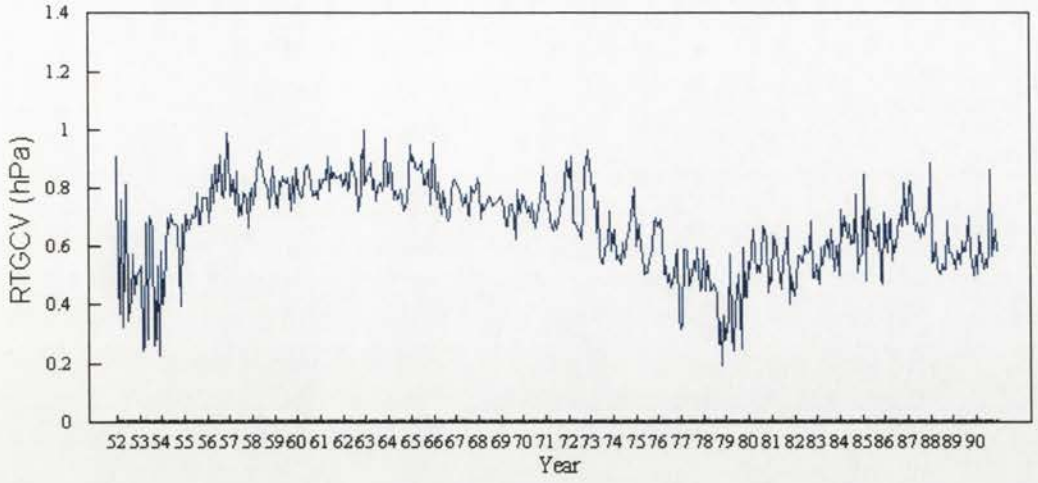


Figure 4.4: Time series of the RTGCV in hPa for pressure 1952-1990, for the final data set used to produce the surfaces.

Once the statistics reach a level where there is very little change with the removal of additional stations and residuals are within expected ranges, the fitted surface coefficients are used to create the MSL monthly surfaces. Coefficients of the surfaces fitted to the monthly values were stored by month. Values for any month at any site or for grids can be calculated using the LAPPNT or LAPGRD programs in the ANUSPLIN package. Maps were obtained using LAPGRD to calculate regular 0.1^{o} grids of the variables at sea-level and using POLGRD (Hutchinson, no date) to trim these grids to the polygon defined by the Australian mainland. A visual representations for each month were created using Generic Mapping Tools (GMT) (Wessel and Smith 1991) although they could be displayed by any of a number of commonly available computer programs which can display raster images. The grids were used as the base data for the statistical analysis and modelling in Chapters 5 and 6. Visual representations (see Figure 4.5 below) are presented on the CD and are further discussed in Chapter 5. These procedures have been used to establish an operational procedure for interrogating and displaying the spatial distribution of monthly climatic variables. The system can be updated to incorporate later years as new monthly values become available.

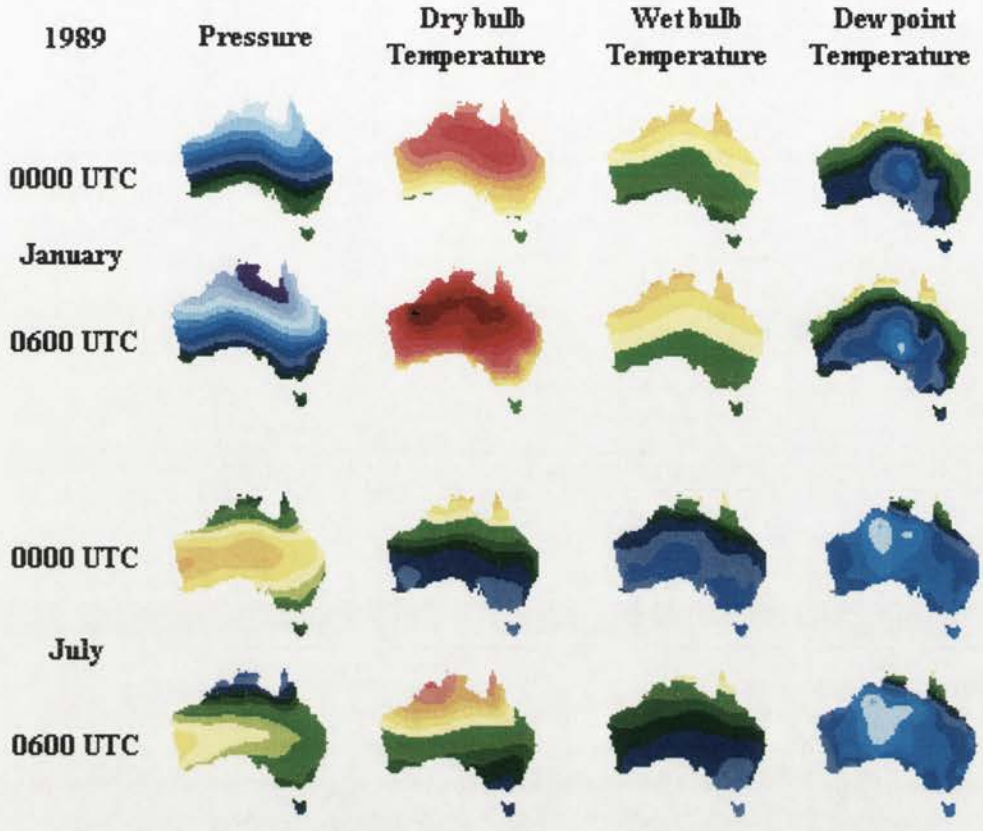


Figure 4.5: Examples of pressure and temperature surfaces.

Estimating true station or barometer elevation using SPLINA

It is possible to use the value of the covariate estimated by the spline function to approximate the true station barometer height. This is done by subtracting the surface MSL value from the station monthly average to give a pressure difference. Using the value of the lapse rate as estimated by SPLINA the pressure difference is converted to a height value.

For example station 086071's barometer is listed in the Bureau of Meteorology's dictionary (1990b) as being at 113m. The lapse rate for December 1960 was 114 hPa/1000m and for January 1982 it was 112 hPa/1000m. The pressure difference between the surface MSL value and the station monthly average was 5.39 hPa and 4.56 hPa respectively. Converting these differences to height suggests that station's height was 47 metres in 1960 and 41metres in 1982. The World Weather Records 1951-1960 lists the station's barometer height at 44 metres in 1960 and the Bureau of Meteorology's 1983 dictionary lists the station's barometer height at 38.1 metres. Considering the errors associated with fitting the function, a root mean square error of 1.2 hPa/1000m, these values are extraordinarily close to the actual values. It confirms the ability of SPLINA to accurately assess spatial relationships, as was

found in the previous chapter, and to facilitate detection and correction of errors in station locations.

Adequacy of the number of stations

The tests on the full spline *versus* the partial spline, which were described in Chapter 3, suggested that the early station networks were not adequate to justify the use of a spatially varying lapse rate. This then raises the question of whether the early station network is sufficient to represent the required surface. One way to test this, is to fit a surface to a later month using only the earlier station network. Figure 4.6 shows the stations in the January 1952 and January 1989 data sets. Of the 1952 stations all but 5 were still operating in 1989. The five missing stations were replaced with the closest in position and altitude from the additional stations available from the 1989 data.

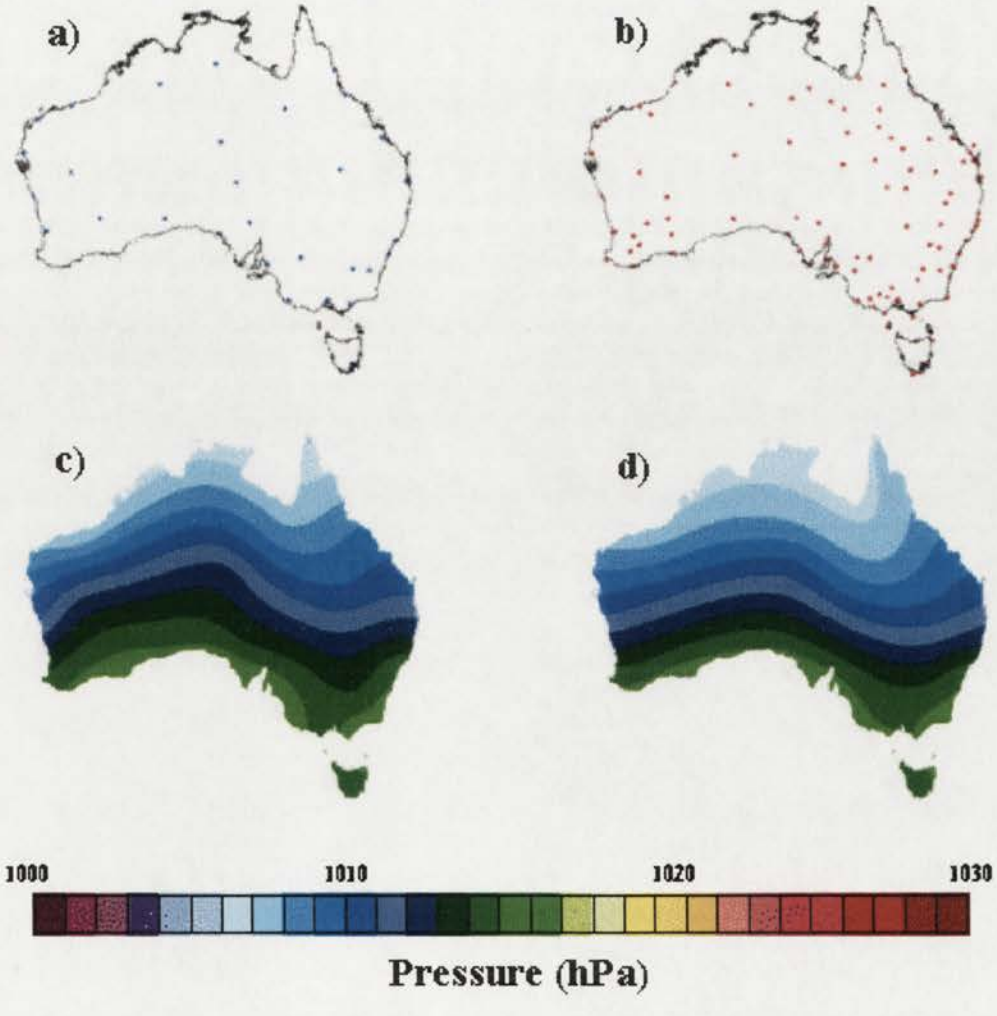


Figure 4.6: Location of stations in the January 1952 (a) and 1989 (b) data sets, and the January 1989 pressure surface as estimated from the equivalent of the January 1952 stations (c) and from the full January 1989 data set (d).

Figure 4.6c shows the 1989 January surfaces constructed with the equivalent of the January 1952 stations and Figure 4.6d shows the same surface constructed with the full January 1989 data set. Figure 4.6 shows that on the whole the smaller data set does not show much change for the southern part of the country. The smaller data set does, however, overestimate pressure in the northern half of the country. On the whole, the addition of the extra stations does very little to improve the prediction of broad-scale geographic patterns. This is confirmed by the output statistics shown in Table 4.5

Table 4.5: Output statistics for the January 1989 pressure surface as estimated from the 1989 equivalent of the January 1952 stations and as estimated from the full January 1989 data set.

January 1989 average pressure	1952 equivalent (37 stations)	all stations (109 stations)
ERROR	7.9	77.5
SIGNAL	29.1	31.5
MEAN	998.037	995.202
STD DEV	19.558	19.432
RTGCV	0.549	0.574
RTMSR	0.117	0.408
RTVAR	0.254	0.484
RTMSE	0.225	0.260
COVARIATE	-115	-113
COV. STD ERROR	0.6	0.4

There is very little difference between the two sets of output statistics. The addition of the extra stations does not change the degree of freedom or the signal, indicating that most of the additional stations only provide noise. This suggests that a threshold has been reached by the 1952 network so that additional stations' information would not have yielded improved predictions of spatial patterns. However, it must be remembered that these results are for MSL surfaces. Additional stations at high altitudes, provided by the 1989 data would improve the Australian surface maps, as is reflected by the lower surface covariate standard error in Table 4.5.

Errors

The issue of data quality and how errors in the spatial attributes propagate through GIS operations has received little attention in the past. To date, most work on spatial modelling with GIS has been concentrated on the derivation of computational models that operate on spatial data. This has included the building of large databases and the linking of computational models with a GIS. The large amount of effort spent on these issues clearly demonstrates the importance and value of spatial modelling and thus it is important to understand the spatial distribution of the errors.

The problem of spatial data quality is obvious because no map stored in a GIS is completely error free. Here 'error' is used in a wide sense to include not only measurement error but also variability and interpolation error. As discussed in Chapter 2, the data stored in the original database have been collected in the field and so a certain amount of error must be expected. The production of the climate surfaces involves the interpolation of irregularly sampled point observations and as a result methods of spatial analysis clearly play a role in determining the size and variability of estimated climatic variables and the associated errors. The prevailing errors are caused by both measurement and interpolation error.

Many techniques can be used to analyse how errors propagate through the interpolation process, but few if any of these techniques have been incorporated into commercial software packages. SPLINA incorporates a routine that produces files of coefficients, the error covariance matrix of the fitted surface coefficients, which can then be used by ERRPNT to calculate the standard model or standard prediction errors for single point observations or ERRGRD to calculate the standard model or standard prediction errors for the entire surface. The standard model error is an estimate of the errors associated with the fitting of the surface. The standard predictive error is as the standard model error and also includes a measure of the variability of the data.

Analysis of the errors resulting from spatial interpolation provides information about the strengths and weaknesses of the historical station networks. Moreover, the standard prediction error estimates associated with the fitted surfaces can indicate areas where the network is inadequate and allow strategies for efficient network design to be developed.

Figure 4.7 shows the estimated standard prediction error surfaces calculated for different underlying grids for January 1989 (for all stations and for the 1952 equivalent stations) and for January 1982. The underlying grids were the Australian DEM, a MSL grid and a station DEM which was created from the stations whose data were used to create the surface coefficient file (see Chapter 2). The error surfaces for the underlying Australian DEM show a clear dependence on elevation with the greatest errors in areas in which the difference between the Australian DEM and the station data DEM are greatest, such as in the eastern highlands. This elevation dependence is clearly shown by the station DEM error surfaces which had the lowest errors compared with the other extraction grids for any month. An estimation of the heights for which the interpolated pressure surfaces have the minimum standard error, are not surprisingly, very similar to the heights of the station DEM.

The MSL error surfaces for some months (see Figure 4.7) show that the greatest error is associated with elevation dependence. Other months, however, such as the January 1989 surfaces, which have a low RTGCV for the surface and small regional differences in the climatic variable of interest, the elevation dependence has little

effect on the output errors and the spatial deficiencies are prominent (e.g. in the central Australian deserts).

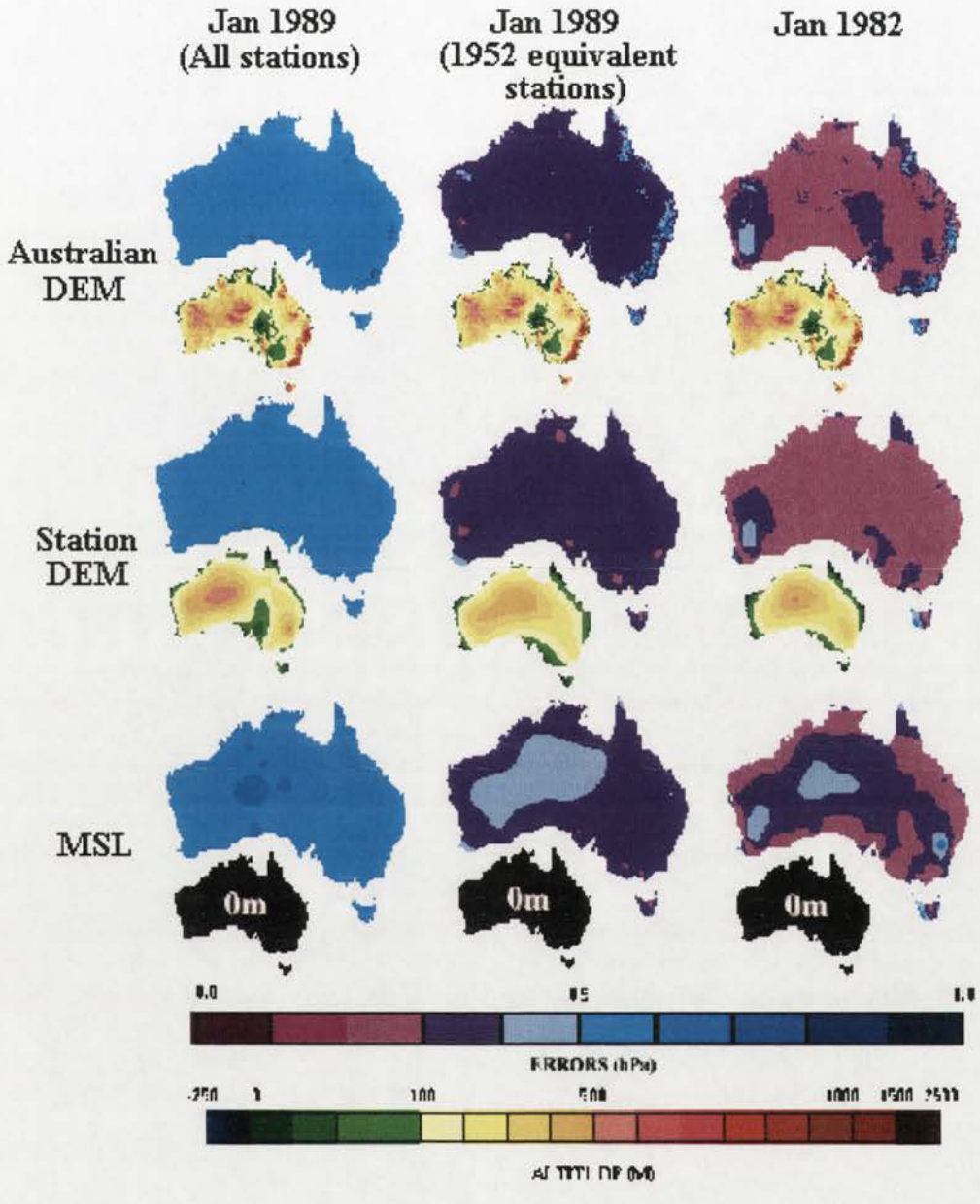


Figure 4.7: Estimated standard prediction errors for January 1989 and 1982 pressure surfaces and the extraction grids used.

If the existing data banks and coverage of the Australian mainland were to be improved, the error surfaces suggest that a greater number of stations in higher altitude areas would best improve the accuracy of the spatial interpolation process, especially when deriving climatic variables at station level.

It is important to note in Figure 4.7 that the error surfaces for any climatic variable are not the same from month to month. Systematic errors do not appear in any

specific location. The average value of the error surfaces varies from month to month and it is not proportional to the number of data points but rather due to the variability associated with the data. This is true not only for differences from month to month for a single climate variable, but also between the different climate variables for the same month, as can be seen in Figure 4.8.

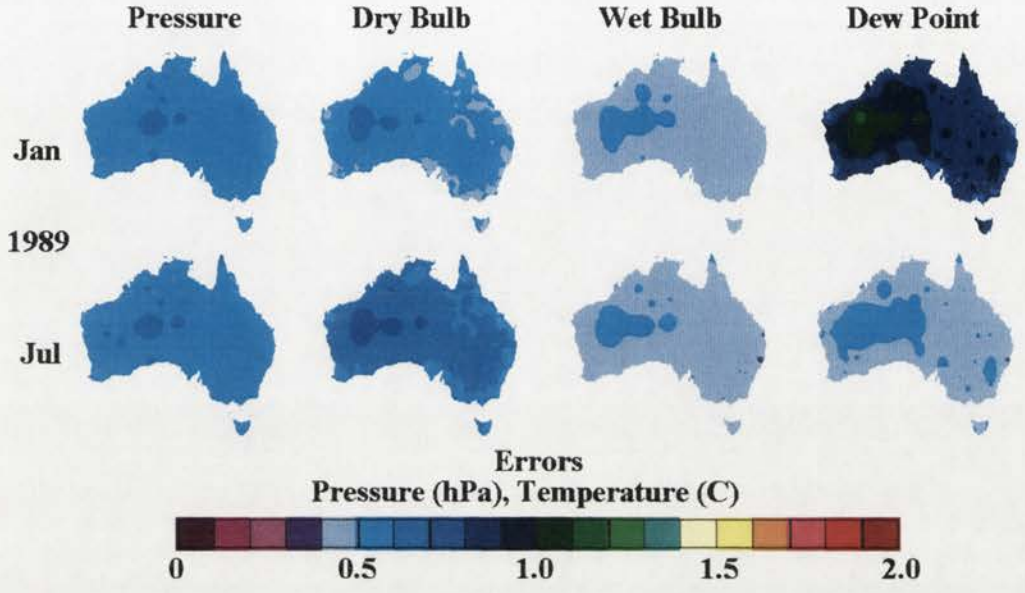


Figure 4.8: Estimated prediction errors for January and July 1989 MSL pressure, dry-bulb, wet-bulb and dew-point temperatures.

Figure 4.8 shows several interesting points. All surfaces show the predominance of the spatial deficiencies that were noticeable for the pressure surfaces. While the standard predictive error values are not really comparable it is interesting to note that the wet-bulb temperature surfaces had the lowest values and for the dew-point surfaces the highest values. The dew-point error surfaces are interesting in that they reflect, not unexpectedly, a combination of the dry-bulb and wet-bulb error surfaces. The dry-bulb predictive errors are lower in summer than in winter while the dew-point surfaces are lower in winter than in summer. Overall there appears to be little or no seasonal pattern to the error surfaces. These results are reflected in the spline output statistics given in Table 4.6, where higher STD DEV, RTMSE and covariate errors are associated with higher spatial predictive errors and RTGCV.

For the final resultant surfaces fitted to the temperature variables, the RTMSE was less than 0.6 °C for all surfaces. For the pressure surfaces created with the uncorrected data set the RTMSE was less than 1.5 hPa, with only 33 months having values greater than 1.0 hPa. For the pressure surfaces, created with the reduced data set, all 468 months had RTMSE values which were less than 0.7 hPa. This value was reduced to 0.6 hPa with the pressure surfaces generated with the final data set and which were used for the modelling in Chapters 5 and 6 and for the atlas on the CD.

Table 4.6: Output statistics for the January and July 1989 MSL pressure and temperature surfaces.

1989	pressure		dry-bulb		wet-bulb		dew-point	
	Jan	Jul	Jan	Jul	Jan	Jul*	Jan	Jul*
ERROR	77.5	67.5	35.2	39.2	52.5	34.5	28.8	20.4
SIGNAL	31.5	37.5	71.8	64.8	47.5	68.5	71.2	82.6
MEAN	995.2	1000.0	24.6	12.1	18.3	9.46	13.65	6.4
STD DEV	19.43	19.26	4.48	4.38	3.44	3.26	4.75	3.26
RTGCV	0.57	0.54	0.71	0.79	0.51	0.58	1.24	0.75
RTMSR	0.41	0.38	0.23	0.30	0.27	0.20	0.36	0.15
RTVAR	0.48	0.47	0.41	0.49	0.37	0.34	0.67	0.33
RTMSE	0.26	0.28	0.33	0.39	0.26	0.28	0.56	0.31
COVARIATE	-113	-116	-5.21	-5.74	-4.70	-4.85	-4.65	-3.68
COV. ERROR	0.4	0.5	0.5	0.6	0.4	0.4	0.9	0.5

* These surfaces were fitted by minimising the average true mean square error using an average of the error standard deviation rather than by minimising the GCV (see Chapter 3).

Environmental lapse rates

Monthly average temperature values combine saturated and unsaturated temperatures, and as elevation was incorporated via a parametric sub-model, the fitted lapse rate does not spatially vary and it is not a free atmosphere lapse rate, but rather an approximate, continent-wide environmental lapse rate (ELR). Figure 4.9 shows the calculated long term mean lapse rates for Australia, which do not agree with those found by Linacre (1992). Linacre states that there is a seasonal effect in the temperature lapse rate and it is usually less in winter than in summer. As can be seen in Figure 4.9 the lapse rates from the surfaces are greater in winter than in summer for both dry-bulb and wet-bulb temperature. In terms of an environmental lapse rate this result is not so surprising, and is due to the physical nature of the Australian continent. That is, as altitude increases, distance from the coast also increases resulting in higher temperatures in summer and colder temperatures in winter. Thus the continental environmental lapse rates are a combination of not only a rise in altitude but also an increasing distance from the coast, or continentality.

The monthly fitted lapse rates are shown in Figure 4.10. It can be seen that the pressure lapse rates vary very little for the entire period. Temperature lapse rates are more complicated with a large seasonal variation for the 1950s and 1960s followed with a period during the 1970s with little seasonal variation, and then a return of high variability again in the 1980s. This variation is most likely related to the rainfall (see Figure 5.14) rather than station numbers and distribution (see Figures 2.4 and 4.6).

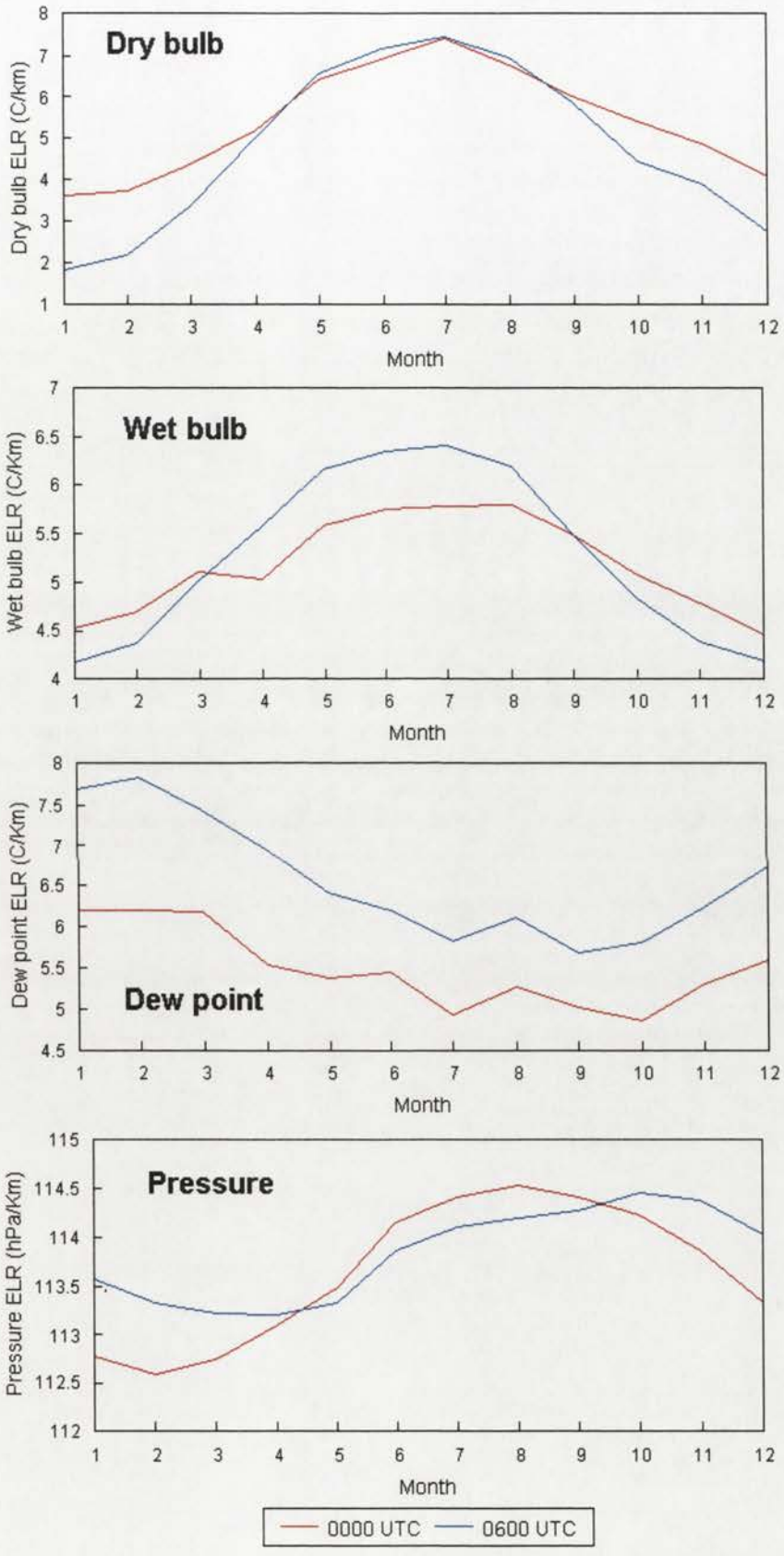


Figure 4.9: Long term Environmental Lapse Rates (ELR).

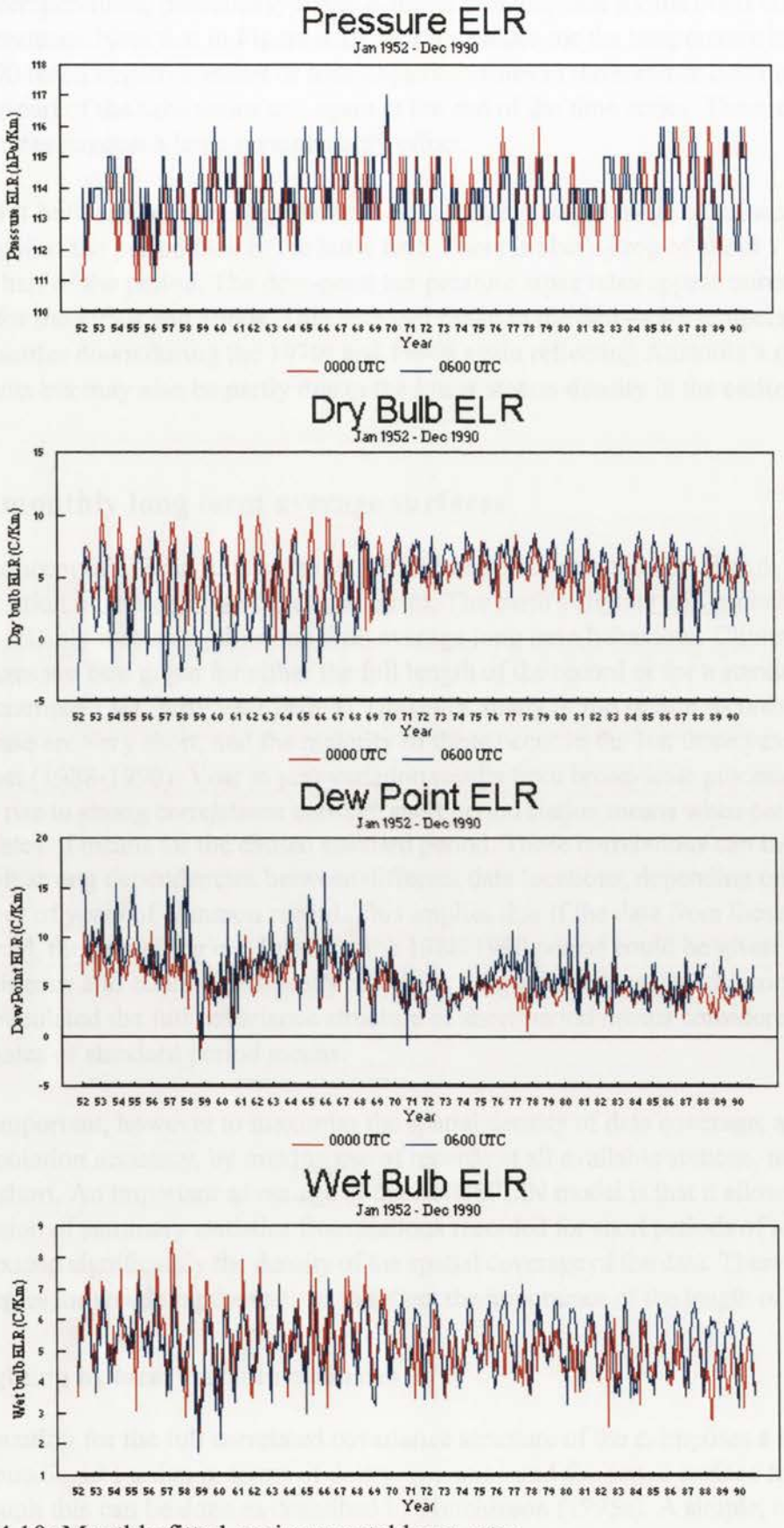


Figure 4.10: Monthly fitted environmental lapse rates.

The range and variation in the fitted ervironmental lapse rates are greater for 0600 UTC temperatures, particularly in the summer months, than for the 0000 UTC temperatures. Note that in Figure 4.10, that the values for the temperature lapse rates at 0600 reach negative values or near negative values in the summer months in the earlier part of the time series and again at the end of the time series. These negative lapse rates suggest a large continentality effect.

The wet-bulb temperatures lapse rates have a slightly larger range of values in the first half of the period than in the latter half. There is also a drop of about 1°C in the latter half of the period. The dew-point temperature lapse rates appear unrealistically high for the 1950s and 1960s. This seasonal range in the dew-point temperature lapse rates settles down during the 1970s and 1980s again reflecting Australia's rainfall amounts but may also be partly due to the lower station density in the earlier years.

The monthly long term average surfaces

The concept of a 'long term average' surface has problems associated with the assumption of stationarity of the time series. The earth's climate will probably never settle reliably into an equilibrium with average long term behaviour. Climatic averages are best given for either the full length of the record or for a standard period (for example a set thirty year period). However, many of the station records in the database are very short, and the majority of these occur in the last three years of the data set (1988-1990). Year to year variation results from broad-scale processes and gives rise to strong correlations between short period station means when considered as estimates of means for the chosen standard period. These correlations can therefore exhibit strong dependencies between different data locations, depending on the number of years of common record. This implies that if the data from these stations are used, the prevailing conditions in the 1988-1990 period could be given undue prominence and could significantly affect the long term means. Hutchinson (1995b) has calculated the full covariance structure of short period means considered as estimates of standard period means.

It is important, however to maximise the spatial density of data coverage, and hence interpolation accuracy, by making use of records at all available stations, no matter how short. An important advantage of the ANUSPLIN model is that it allows the inclusion of summary statistics from stations recorded for short periods of record. This can extend significantly the density of the spatial coverage of the data. There are several strategies for weighting the stations to reflect the importance of the length of record.

Weighting by local variance estimates

Accounting for the full correlated covariance structure of the ε_i imposes a significant computational burden in terms of data preparation and the actual surface fitting, although this can be done as described by Hutchinson (1995a). A simple, non-

correlated, approximation to this covariance structure was posed by Hutchinson and Bischof (1983), by setting the variance of the data mean at location *i* to

$$\text{var}(\varepsilon_i) = \frac{\sigma_i^2}{n_i} \tag{4.1}$$

where σ_i^2 is the local variance estimated directly from the monthly data at location *i* and n_i is the number of years of record. This recognises that stations with higher variances or shorter periods of record should be subjected to more data smoothing than stations with smaller variances or longer periods of record. Hutchinson (1995a) has shown that, when applied to annual mean rainfall, this model can perform similarly to the model which incorporates the complete correlated error structure. This leads to a very practical procedure for interpolating monthly means which requires just minimal summary data at each location. However, the method does ignore the strong correlations between means at locations with overlapping periods of record and some of this spatial correlation can be interpreted as signal by the spline fitting procedure.

Table 4.7 shows the output statistics from the different spline runs for the different weightings for the LTA. The Table shows the RTGCV and estimated root mean square errors of the surfaces were larger than the values obtained for uniform weighting. This local variance weighting method however, is more relevant to rainfall which has much more year to year variability than temperature and pressure.

Table 4.7a: Output statistics for the January LTA MSL pressure surfaces (all 125 stations using SPLINA).

SPLINA	Partial Spline			Spline		
January	weighting			weighting		
Pressure surfaces	$\frac{\sigma_i^2}{n_i}$	$\frac{1}{n_i}$	uniform	$\frac{\sigma_i^2}{n_i}$	$\frac{1}{n_i}$	uniform
ERROR	72.5	77.8	95.6	28.7	54.8	73.8
SIGNAL	52.5	47.2	29.4	96.3	70.2	51.2
RTGCV	1.12	1.15	0.781	1.02	1.06	0.787
RTMSR	0.649	0.714	0.597	0.235	0.467	0.464
RTVAR	2.46	11.2	0.683	1.41	8.71	0.604
RTMSE	0.553	0.556	0.331	0.430	0.528	0.387
COVARIATE	-113	-112	-113			
COV. ERROR	0.5	0.5	0.5			

Table 4.7b: Output statistics for the July LTA MSL pressure surfaces (all 125 stations using SPLINA).

SPLINA	Partial Spline			Spline		
January	weighting			weighting		
Pressure surfaces	$\frac{\sigma_i^2}{n_i}$	$\frac{1}{n_i}$	uniform	$\frac{\sigma_i^2}{n_i}$	$\frac{1}{n_i}$	uniform
ERROR	80.8	30.1	91.2	23.7	28.7	54.9
SIGNAL	42.2	92.9	31.8	99.3	94.3	68.1
RTGCV	1.46	1.30	1.06	1.33	1.09	1.02
RTMSR	0.957	0.319	0.783	0.256	0.255	0.456
RTVAR	2.27	2.89	0.909	1.12	2.37	0.683
RTMSE	0.691	0.560	0.462	0.524	0.462	0.508
COVARIATE	-116	-112	-116			
COV. ERROR	0.5	0.8	0.7			

Weighting by length of record

If the year to year variability of the climate variable is relatively large (e.g. rainfall), it may be reasonable to weight the observed means simply by setting

$$v_{ii} = \frac{1}{n_i} \tag{4.2}$$

This method has the significant practical advantage of giving the same weighting for each month when interpolating monthly means. This permits significant savings in computation and subsequent storage of the fitted surface coefficients, since the numerically expensive part of the calculation needs only to be done once in order to fit minimum GCV surfaces to monthly means for each month. Once again, however, Table 4.7 shows the RTGCV and estimated root mean square errors of the surfaces weighted in this way were larger than the values obtained for uniform weighting for pressure.

Uniform weighting

Uniform, diagonal weighting is often the default and sometimes only available option for interpolation methods. If variances do not vary greatly across a data network, uniform weighting can be the most appropriate form for it to take. Table 4.7 shows that the RTGCV and the root mean square error for the uniform weighting were smaller than the values obtained with the variable weighting models and demonstrates that the uniform weighting is the best method for pressure. While, this weighting is reasonable for climate variables such as temperature and pressure, for which data means "settle down" after only a few years (in the absence of longer term climate change), it is not generally recommended for interpolating rainfall means.

Figure 4.11 shows the output for the different SPLINA weighting methods for the January and July long term pressure averages. The differences between the methods are not significant. Those months with a higher signal to error ratio are more complex due to the reduced amount of smoothing that is applied. While there is little visual difference between the surfaces, the lower RTGCV values of the SPLINA uniform weighting make it the most attractive to use.

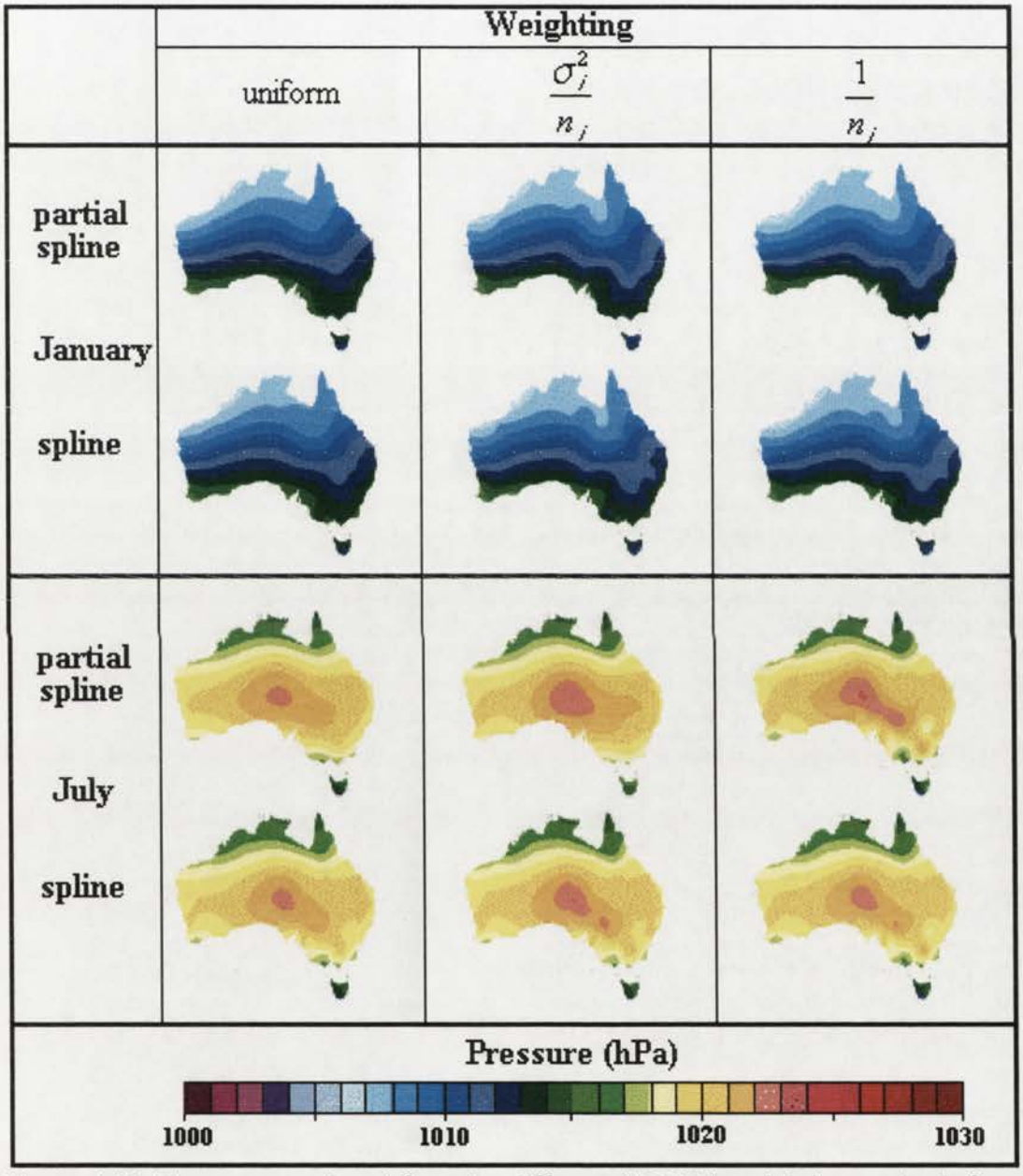


Figure 4.11: The maps produced from the different SPLINA weighting methods for the January and July long term pressure averages.

Equal weighting with SPLINB

Another possible model is to fit a second order spline to all the monthly data for each station. This requires the use of the SPLINB program as the data set has coincident points. The knots calculated by the SELNOT program are the set of unique station positions. This method allows for the incorporation of all data and places equal weighting on each monthly data value. This intuitively seems the most appropriate method of interpolating long term means. However this could be affected by biases in station occurrences over time, especially if the biases in time are also biased spatially. Table 4.8 shows the output statistics for the January LTA MSL pressure surfaces for all stations using SPLINB. The partial spline surface was produced for both the full data set and also for a reduced data set which had no station records from 1988 to 1990. This was done in order to assess if the denser station network of the 1988 to 1990 period had undue influence on the surfaces.

As can be seen in Table 4.8 the 1988-1990 data does not have an undue effect on the output statistics and Figure 4.12 shows there is little difference in the output maps although the full data set shows slightly higher pressure in winter and slightly lower pressure in summer. There is very little difference between the partial spline and the spline. A comparison of the SPLINA and the SPLINB surfaces shows very little visual difference, but the uniform weighting surfaces are less complex than the weighted SPLINA or the SPLINB surfaces. It is interesting to note, however, that the signal is fairly consistent amongst the different methods.

Table 4.8a: The spline and partial spline SPLINB output statistics for the January LTA MSL pressure surfaces.

SPLINB January pressure	partial spline		spline
	all months	no 1988-90	all months
Number of months	2074	1755	2074
Number of knots	130	74	130
ERROR	2020.2	1710.2	2010.9
SIGNAL	53.8	44.8	63.1
RTGCV	1.79	1.84	1.78
RTMSR	1.74	1.79	1.73
RTVAR	1.77	1.81	1.76
RTMSE	0.284	0.289	0.306
COVARIATE	-112	-112	

Table 4.8b: The spline and partial spline SPLINB output statistics for the July LTA MSL pressure surfaces.

SPLINB July pressure	partial spline all months	partial spline no 1988-90	spline all months
Number of months	2077	1765	2077
Number of knots	128	74	128
ERROR	2043.9	1737.6	2031.6
SIGNAL	33.1	27.4	45.4
RTGCV	2.56	2.64	2.56
RTMSR	2.52	2.60	2.51
RTVAR	2.54	2.62	2.53
RTMSE	0.321	0.326	0.375
COVARIATE	-116	-116	

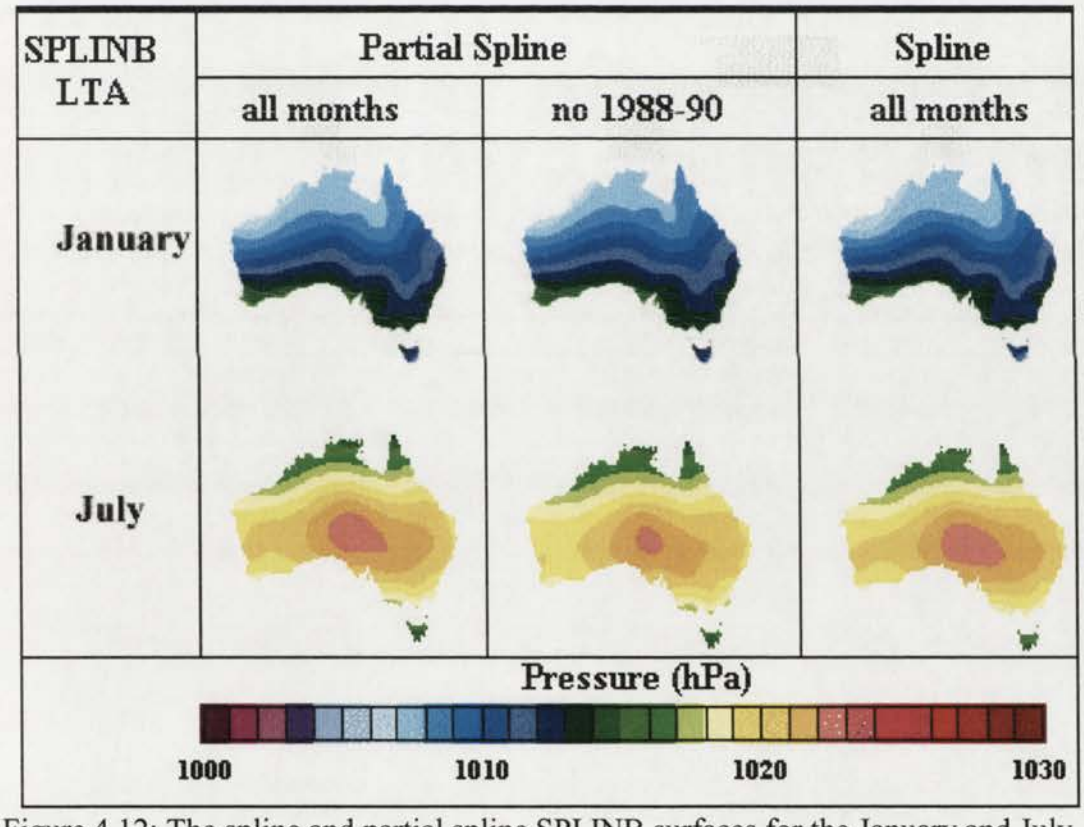


Figure 4.12: The spline and partial spline SPLINB surfaces for the January and July long term averages.

Other possible methods of determining the long term mean surfaces

Pre-standardising to a fixed period is another accepted method which is discussed in Hutchinson (1995b). This method, however, requires additional computing time. Hutchinson (1995b) shows that the method only marginally improves the surfaces provided by the SPLINB covariate method, hence it was not considered for this thesis.

Surface long term averages

Although ANUSPLIN allows the use of the above weighting options, the average of each of the 0.1° grid points of the primitive surfaces for the 39 years for each month was used as the long term average surface for assessing departures from the mean in Chapters 5 and 6. This method weights each month equally, but it is important to keep in mind that it is the average of smoothed data and incorporates the interpolation errors from the surfaces. However the use of the average of the primitives avoids regional interpolation errors which can create false anomaly values due to spatial data deficiencies and provides spatial anomalies which are equally spread around the LTA surface.

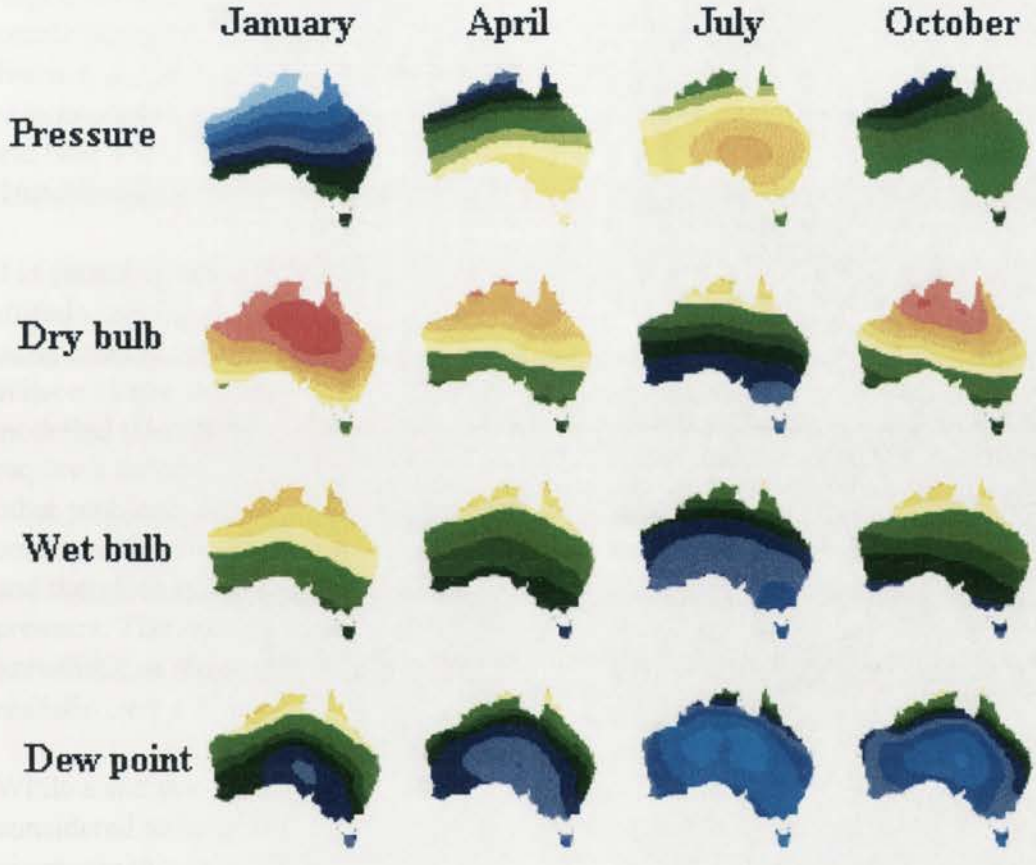


Figure 4.13: Surface long term averages (SFLTA) for mid season months.

Fitting the rainfall percentile surfaces

Spatial rainfall patterns on the Australian continent are complex due to the tropospheric movements discussed in Chapter 1, which bring rain to different parts of the country at different times of the year. The rains of Northern Australia come nearly entirely from summer monsoonal activity, the underlying cause of which is the instability of the equatorial and tropical maritime air spreading over continental

locations. The consequent thunderstorms are mostly localised and patchy. There are, however important instances of the southerly invasions of equatorial air with resulting widespread and plentiful rains. The trade winds bring some orographic rain to the North Queensland coast, but very little of this penetrates inland. The frontal rains from the mid-latitude depressions only skirt the Southern shores during the summer, but by March they advance northward and by June or July, the position of the frontal rains is consolidated. From August to September they slowly retreat and decrease at an almost constant rate. The amount of rain from these frontal systems tend to be increased even by a modest orographic uplift.

Methods for the interpolation of mean rainfall from ground-based point data have ranged from simple trend surface analyses (Edwards 1972, Hughes 1982) and local bi-variate interpolation techniques based on thiessen polygons (Thiessen 1911), inverse distance weighting (Shepard 1968) and delauney triangulations (Akima 1978) through to more sophisticated statistical methods. The last include geostatistical methods (Chua and Bras 1982, Hevesi *et al.* 1992, Phillips *et al.* 1992) and thin plate splines (Hutchinson and Bischof 1983, Hutchinson 1995b).

It is generally acknowledged that there are significant effects on rainfall related to altitude and aspect (see Chapter 2). Both of these effects were recently examined using local regression methods in the PRISM model (Daly *et al.* 1994, Stillman and Wilson 1996). While the effects of topography on received rainfall has been modelled (Hutchinson 1995a), the spatial interpolation of monthly rainfall amounts require a suitably dense station network and a similarly fine-scaled DEM. There are other problems associated with interpolating monthly rainfall amount at the continental scale. There are three quite different mechanisms that cause precipitation, and therefore rainfall is not nearly as spatially coherent or simple as temperature or pressure. The representation of altitude as a covariate at the continental scale is unrealistic as the relationship is not linear. Hutchinson (1995a) has shown it not to be realistic over a 1° box covering the Australian Capital Territory (ACT).

While a suitable DEM was available the station network (3160 stations) was not considered to be suitably dense. Hutchinson (1995a) gives a possible method to overcome this problem by anomalies, however, his approach was not explored in this thesis. As a result rainfall percentiles, rather than rainfall means, were chosen as an alternative way to represent the factors which influence the spatial and temporal variation of rainfall. An important feature of rainfall percentiles is that they exhibit a much higher degree of spatial coherence than actual monthly rainfall totals. The strong spatial coherence of the rainfall percentiles reflects the fact that they are essentially measuring normalised departures from average conditions. These departures reflect the broad-scale synoptic patterns operating for each month and are essentially independent of topography.

The SPLINB procedure in ANUSPLIN, which is especially designed to handle large data sets, was used to interpolate the monthly percentiles across Australia on a month

by month basis from the rainfall percentile data set described in Chapter 2. The rainfall percentiles were fitted with a second order thin plate smoothing spline as a function of longitude and latitude only, unlike existing interpolated monthly mean rainfall surfaces (Hutchinson and Bischof 1983, Hutchinson 1991a and 1991b) which depend critically on elevation. The percentiles were interpolated using a single SPLINB surface with 300 knots selected by SELNOT since the spatial patterns of percentiles are very broad. To illustrate the results of the surface fitting process, maps for a sequence of months from May to August for the relatively wet year of 1973 and for the severe drought year of 1982 which ravaged much of eastern Australia are presented. The maps are shown in Figure 4.14.

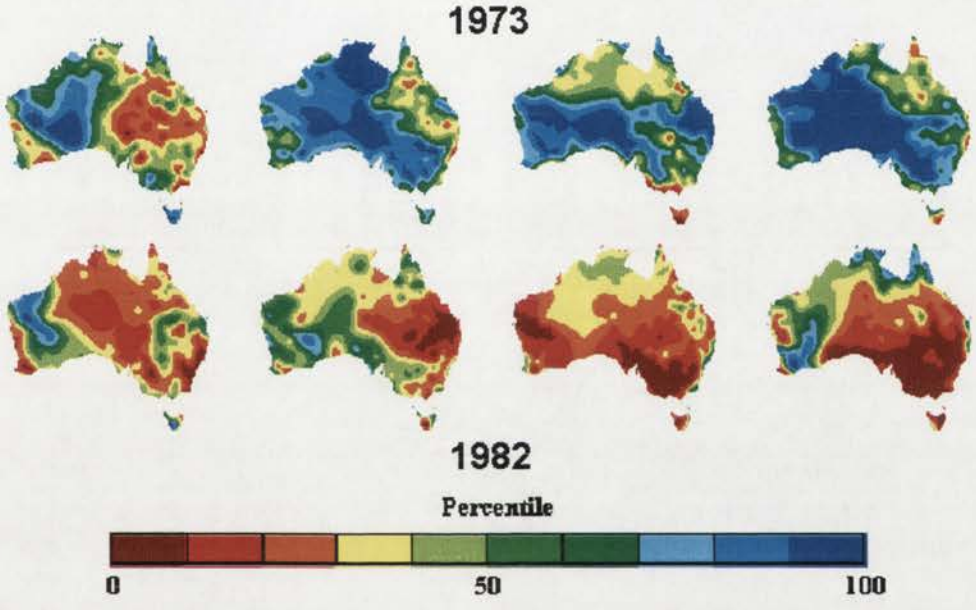


Figure 4.14: Maps of monthly percentiles across Australia from May to August for 1973 and 1982.

As with the pressure and temperature data, the procedure used to produce the maps of Figure 4.14 can be employed for any month within the period of acquired monthly rainfall data from 1952 to 1990. Coefficients of the surfaces fitted to the monthly percentiles were stored by month, and visual representations of the monthly percentiles are presented on the CD.

Conclusions

The task of spatial interpolation is made difficult by a number of factors. Climate data are often sparse, measured for a limited duration and of varying quality. More significantly, the density of the data may be much less than the resolution of the required interpolated grid. In spite of this, the size of the available data set may be sufficient to cause computational difficulties, particularly when dealing with continental or global data sets.

If space-time models are to be developed from ground-based data, which remain the principal source of surface weather data, then the parameters of these models need to be accurately interpolated across the landscape. This requires a sound understanding of the spatial and temporal scales of variation of these quantities and their dependencies on additional factors such as topography. Because it is impossible to achieve a weather station density of one station per 100 km^2, for climate modelling purposes it is important that the climatic data be interpolated between stations in a realistic way that is sensitive to elevation effects.

The technique of thin plate smoothing splines and its extensions can address these problems and provide insight into the nature of the spatial variation of climatological variables. The major factor in improving the accuracy and spatial resolution of the interpolated pressure and temperature surfaces has been the incorporation of elevation as the third independent variable.

The iterative process of "fitting" the statistical surface demonstrates a vitally important characteristic of the spline technique. Verification of climate data poses many challenges and one tool to assist in that verification is placing a climate station in its spatial context. The results of the spline interpolation provided feedback about the meteorological station database by providing summary statistics and the deviation for each station relative to the fitted surface. Examination of data residuals is a powerful aid for detecting and correcting the data errors. The example given of fitting the pressure surfaces has shown the usefulness of the summary statistics in that there is a reduction in surface error as indicated by the RTGCV after removal of a few stations. The robustness of the spline technique is demonstrated by the corresponding reduction in surface complexity or roughness with the removal of the station.

Analysis of the errors resulting from spatial interpolation provides information about the strengths and weaknesses of the historical station networks. Moreover, the standard error estimates associated with the fitted surfaces can indicate areas where the network is inadequate and allow strategies for efficient network design to be developed. If the existing data banks and coverage of the Australian mainland were to be improved, the error surfaces suggest that a greater number of stations in higher altitude areas would best improve the accuracy of the spatial interpolation process, especially when deriving climatic variables at station level.

The errors from the surface created with the 0.1° Australian DEM shows that the station network is poorest for the high altitude areas of the continent. The assessment of the error grids from the different underlying grids show the importance of good spatial coverage by the station network. The elevations generated from the surface of minimum error, not surprisingly, show that elevations of minimum errors are closely linked to the DEM generated by the stations. On the whole, however, the addition of the extra stations does very little to improve the prediction of broad-scale geographic patterns.

The techniques described here allow for the rapid revision of the surfaces as new climatological data are entered into the station database. In combination with a GIS, questions about the spatial extent of natural resource management problems can be efficiently addressed.

The long term monthly means of temperatures and pressure are also amenable to reliable spatial interpolation using thin plate smoothing splines. While it is possible to impose different station weighting when spatially interpolating these statistics it is interesting to note that the uniform weighting performed the best for the pressure and temperature long term averages. However climatic variables with high year to year variability, such as rainfall, usually require one of the alternative weighting strategies in order to include stations with incomplete records in the network. These alternative weighting strategies can increase the number of stations with data suitable for spatial interpolation by as much as an order of magnitude (Hutchinson 1995a).

Chapter 5. The Australian Spatio-temporal Climate Database and Mapping Techniques Used to Describe the Database

Introduction

The methods described in the previous chapter were used to produce a gridded set of MSL long-term means (LTA) and monthly means (AVG) for pressure and temperature from January 1952 to December 1990, for both 0000 UTC and 0600 UTC. A gridded set of rainfall percentiles for the same period was also produced. These 4212 grids which were interpolated from the station climate database, form the set of 'primitives' in the gridded spatial database. This chapter begins with a description of the set of 'primitive' surfaces and the methods used to create them.

There are many ways in which information can be obtained from a spatial climate database. The most common is simple data retrieval. However, particularly in the earth and environmental sciences, new attributes are often derived from variables already held in the database. These new attributes can be derived from climate models which can be logically, statistically or physically based.

A logical model computes a new attribute by applying simple logical 'rules' to the input variables. These logical models produce surfaces which either represent a specific time or represent change over time. This chapter continues with a discussion and examples of types of spatial logical models used to generate the 'derived' attribute grids and the maps presented on the CD. Logical models were applied to the 'primitive' surfaces to produce an additional 16,000 'derived' surfaces, the graphics of which are presented on the CD, and are discussed further below. Several of the derived attributes are used in the spatial statistical modelling discussed in Chapter 6.

Statistically based models are discussed further in Chapter 6. Physically based models such as Global Climate Models (GCMs) require large computer systems. These models are briefly discussed but were considered beyond the scope of this thesis. While these modelling distinctions do exist, models often comprise combinations of these types.

This chapter continues with a discussion of the issues concerning the simultaneous support of both space and time in databases. The consideration of time and space is not an entirely new aspect of geographic information systems and spatial and temporal patterns are discernible in most physical and social systems. While these patterns can be very complex, it is essential to be able to examine change on the basis of time and to be able to examine overall patterns of temporal relationships. In a similar manner that a spatial distribution pattern can be random, uniform or clustered, a temporal distribution can be chaotic, steady state or cyclic. One can choose to focus on moments of change, or on overall data currency, and on individual versus aggregate change. Likewise, rather than seeking a simple snapshot of a past moment, one might prefer to view what was generally true during that period.

Communication is the transfer of information to the user and is both aspatial and spatial. Aspatial communication includes charts, graphs, and reports whereas spatial communication involves both static and dynamic mapping techniques. Static techniques include dimension shifting, which displays data on a temporal axis (e.g., accident data or a time slice map with space constant), and symbology, as when displaying migration and other flow data. Dynamic techniques include animation and sonification. Animation provides the illusion of movement to show change over time.

While graphics can be considered as mainly a display mechanism, they are actually a powerful visualization tool that allows one to determine new patterns in data. Since human cognition is visual, the tool is important and powerful. However, researchers constantly face the problem in attempting to visualize complex systems that they run out of visualization techniques since there are generally more parameters than there are distinct ways of collectively showing the data.

This chapter continues with a discussion of the mapping techniques used to describe spatio-temporal data. Several of these techniques are discussed and investigated. Static graphs are used to show the trends in the time series for Australian averages and for the 2.5° sub-grid (extracted from the 0.1° grid) and the results are discussed.

The use of visualization and animation to convey the information stored in a spatio-temporal database is examined. The use of snapshots to give an overall view of the temporal variation, the use of animation to convey a sense of change over time and the use of sound to enhance visualization applications are all considered. The chapter concludes with an investigation of the combination of animation and sound as a method of enhancing the presentation of information. Several of these animation and sound combinations are presented on the CD.

The 'primitive' surfaces

The methods described in the previous chapter were used to produce:

- Surface coefficient files and a gridded set of monthly means (AVG) for pressure and temperature (dry-bulb, wet-bulb and dew-point) from January 1952 to December 1990, for both 0000 UTC and 0600 UTC.
- The long term average (LTA) and surface long term average (SFLTA) methods were also used to produce surface coefficient files and a gridded set of MSL long-term means for both 0000 UTC and 0600 UTC for the period from January 1952 to December 1990.
- A gridded set of monthly rainfall percentiles for the period from January 1952 to December 1990.

Surfaces for the individual monthly means (AVG) were spatially interpolated using a second order partial spline function minimising the GCV. This method failed for 59 surfaces (7 for pressure, 18 for dry-bulb, 15 for wet-bulb and 19 for dew-point).

These surfaces were recalculated by using the average monthly error standard deviation, RTVAR (see Chapter 3), from the rest of the time series. The main output statistics from these 'primitive' surfaces are given on the CD.

It should be noted that the rainfall percentiles surfaces were created from the percentiles of monthly 9am to 9am station rainfall totals (and therefore are not actually 0000 UTC) and 0600 UTC surfaces are not appropriate (see Table 5.1). It should be noted, however, that rain, unlike temperature, is accumulated over time and the time of day is hardly relevant. LTA and SFLTA maps were not created for the rainfall percentiles as the average should be a constant 50% across the maps.

These 4212 grids which were interpolated from the station climate database and the SFLTA surfaces form the set of 'primitives' in the gridded spatial database. These 'primitive' surfaces are conveniently stored in surface coefficient files for use with LAPGRD and LAPPNT for grid or point data extraction. Both 0.1° grids and colour images were created and the images of each of these grids are presented on the CD.

Table 5.1: The number of surfaces in the sets which make up the 'primitive' surfaces.

	00 UTC			06 UTC		
	LTA	SFLTA	AVG	LTA	SFLTA	AVG
Pressure	12	12	468	12	12	468
Dry-bulb	12	12	468	12	12	468
Wet-bulb	12	12	468	12	12	468
Dew-point	12	12	468	12	12	468
Rainfall %	-	-	468	-	-	-

While no similar data sets of the average surfaces exist, the output statistics as described in the previous chapter and the graphics suggest that the surfaces are a faithful representation of the climatic conditions prevailing at the time.

The long term average pressure maps clearly show the seasonal migration of the high pressure belt with the winter maximum at approximately 28°S and the appearance of the northern heat lows in the summer (see Chapter 1). The long term average January and July pressure maps are similar to those shown in Figure 2.4 produced by Gentilli (1971). The differences between the two can be attributed to different climate stations in the databases and different methodologies for determining the long term averages. The average monthly long term averages (SFLTA) maps do not have as high a pressure in mid-winter or as low a pressure in summer as the interpolated long term average (LTA) maps. The northern heat lows, which are more apparent in the LTA 0000 UTC surfaces than the SFLTA surfaces, are very obvious in the both 0600 UTC maps.

While the monthly average pressure maps also show the seasonal cycles, both latitudinal and longitudinal variations associated with the different positions of the

long wave troughs are evident. Months which have the high pressure belt situated further north than normal tend to have overall lower pressure, tend overall to be wetter and have lower dew-point temperatures than those with their centres situated further south (see Figure 5.1).

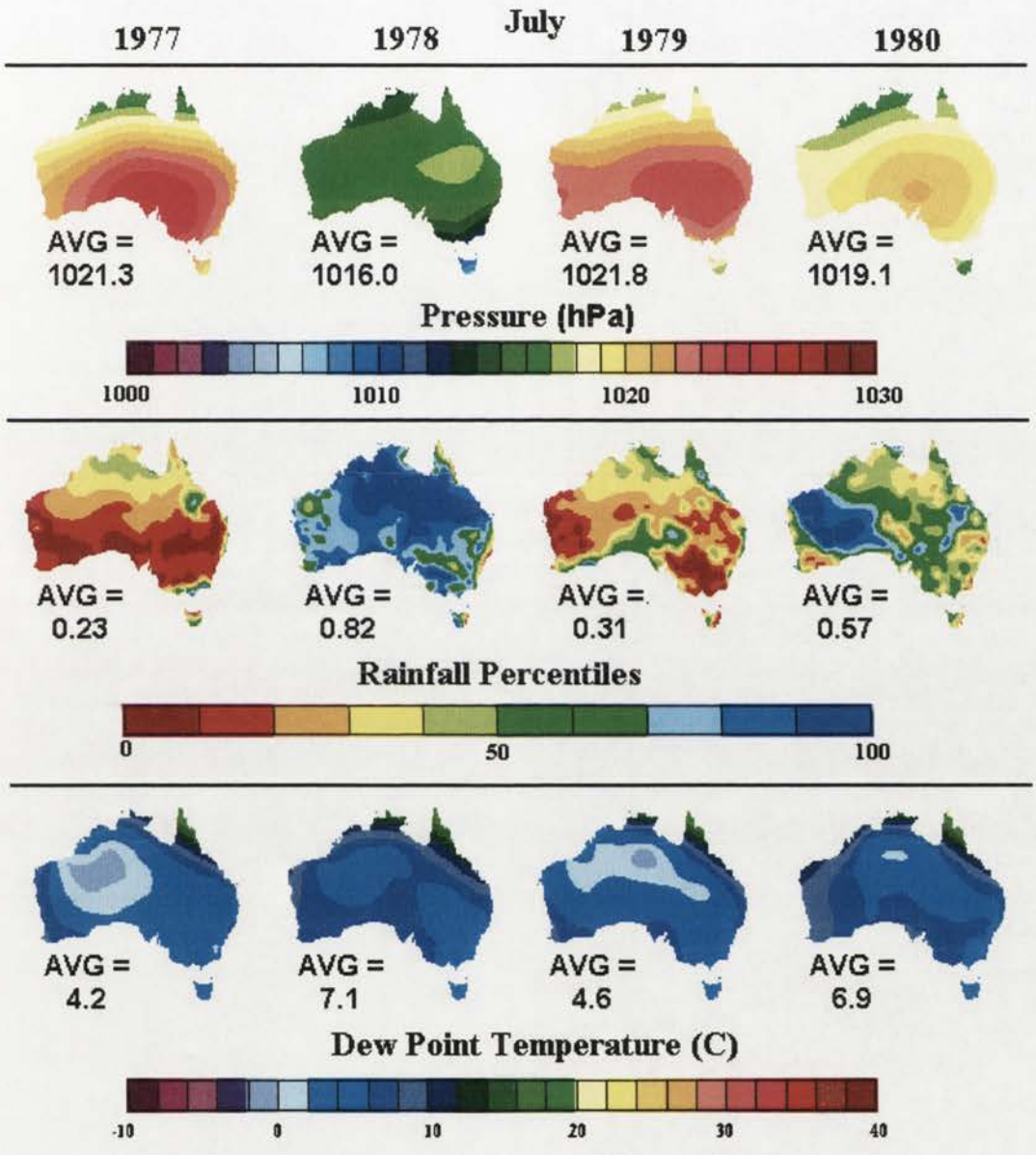


Figure 5.1: July MSL pressure, rainfall deciles and the MSL dew-point temperature for 1977 to 1980.

MSL dry-bulb, wet-bulb and dew-point temperature maps are not commonly produced. The temperature surfaces, however, clearly show the seasonal variations and the diversity across the continent due to the effects of latitude, as described in Chapter 1. The diurnal patterns and the differences in the length of day are very evident in the surfaces and can be seen in the January and July dry-bulb temperature

maps shown in Figure 5.2. The cooling effect of onshore breezes in the late afternoons is also very noticeable, especially along the western coast of Australia. The wet-bulb temperature maps generally show similar trends to the dry-bulb temperature maps. While the dew-point temperature maps show similar trends, the degree of continentality is evident and is most pronounced during extreme wet or dry months as is shown in Figure 5.1.

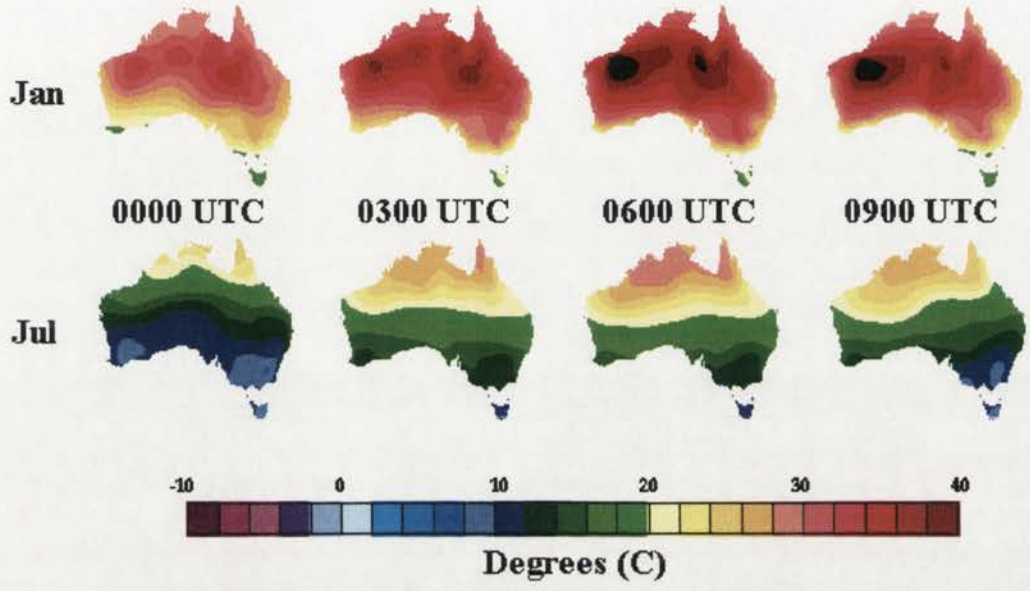


Figure 5.2: January and July SFLTA MSL dry-bulb temperature surfaces for 0000, 0300, 0600 and 0900 UTC.

The monthly rainfall percentile maps compare well with the hand drawn percentile maps for individual states from the BoM (1968-1989). The rainfall percentile maps clearly show the complex spatial patterns and high degree of variability associated with Australian rainfall. The particularly wet and dry periods associated with various ENSO events, for example the 1969 to 1972 drought in eastern Australia or the 1973 to 1974 wet, are evident.

Overall the set of primitives show that the spline interpolated surfaces are a faithful representation of the Australian climate. Some of the individual maps however can have relatively simple (low signal) or complex (high signal) patterns which are due to the nature of the daily variability within the month. As these monthly maps are an average of the daily variability they may not be representative of the individual daily patterns.

The derived surfaces

Logical models using the set of 'primitives' have been used to compute new attributes for:

- a specific time, such as the vector normal and gradient of the pressure surfaces or the estimation of the wet-bulb depression.
- change over time, either in relation to some long term average (e.g. monthly anomalies) or in relation to another specific time (e.g. the lag difference between successive months).

Logical models can be created for any specific time already represented in the database. In all these transformations the set of primitives is used to create new data from existing data and so extend the database. A proportion of these surfaces have been used in preliminary analysis for models discussed later in the chapter and a graphic of each of the surfaces is presented on the CD.

Logical models for specific times

Data extraction using DEMs

The simplest spatial logical model includes the extraction of climatic variables for various altitudes at specific times. The LAPPNT and LAPGRD programs were used to extract both point values and grids. Examples include the extraction of mean sea-level pressure values (as in Figure 4.1), or the extraction of surface temperature values using elevation data from the Australian DEM (Hutchinson 1993b) as given in Figure 5.3.

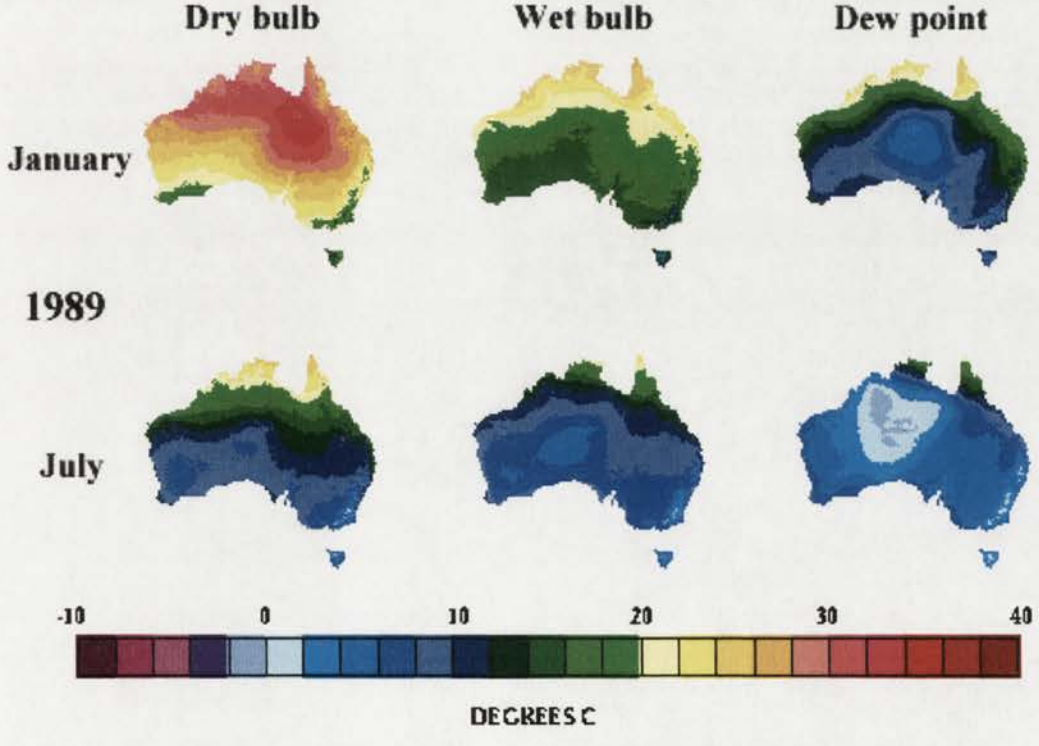


Figure 5.3: January and July 1989 MSL temperature surfaces used in combination with the Australian DEM to give the surface values.

Aspect and slope maps

The use of ANUSPLIN in conjunction with the SLPGRD and ASPGRD computer programs extends simple data interpolation and retrieval to more complex queries, such as deriving the slope and aspect of the surfaces. The programs normally used to extract slope (SLPGRD) and aspect (ASPGRD) from a DEM and can be applied to the climate surfaces by assuming that the climate surface is the DEM. This can be extremely useful for determining the vector quantities which represent geostrophic wind speed and direction. Given an isobaric map, geostrophic wind can be calculated and the geostrophic wind field can be considered as equivalent to a forcing function. The geostrophic wind direction blows perpendicular to the horizontal pressure gradient with a speed proportional to the gradient, with the low pressure system to the right of an observer facing downwind in the southern hemisphere. That is, the aspect or vector normal of the pressure surface is equivalent to the geostrophic wind direction less 90°, and the slope or gradient of the pressure surface is analogous to the geostrophic wind speed. The January and July 1989 pressure, aspect and slope maps are given as an example in Figure 5.4. The dominant red and purple of the easterlies in the summer synoptic patterns (see Chapter 2), and the intrusion of the blue and greens of the westerlies over the southern areas in winter when the ridge of high pressure extends across the continent are demonstrated very clearly in the aspect maps.

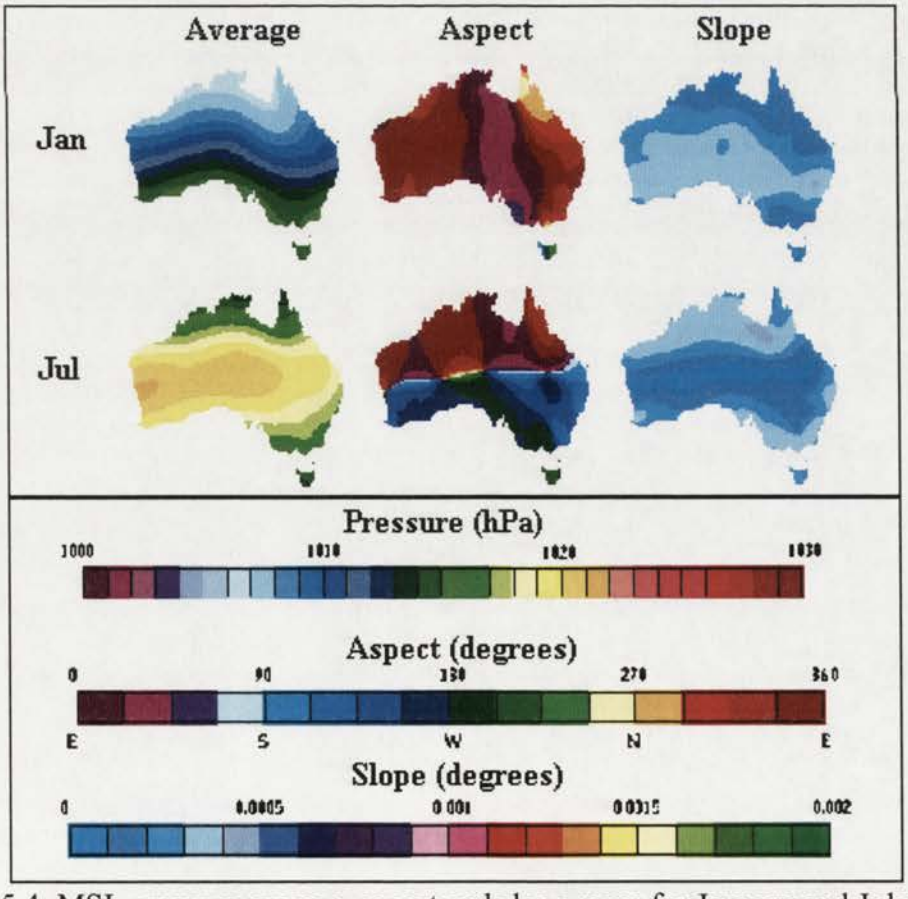


Figure 5.4: MSL pressure average, aspect and slope maps for January and July 1989.

The slope and aspect surfaces for the 0000 UTC temperature surfaces are presented on the CD and provide a number of points of interest. The winter temperature aspect maps should normally display a southerly aspect as the temperatures in Australia get progressively colder towards the south of the continent. The dry-bulb aspect surfaces for several winter months show this trend (e.g. the July 1986 surface in Figure 5.5) with the blue and purple of the southerly aspect apparent. Other monthly surfaces, however, show that the coldest temperatures (shown as a trough or reversal of aspect in the July 1989 surface in Figure 5.5) are some distance inland, which suggests that warmer than usual ocean temperatures may have maintained higher air temperatures over the southern coastal areas. Similarly the line of maximum heating features in the summer maps could be investigated, although some consideration must be given to the position of the solar heat maximum when viewing the maps.

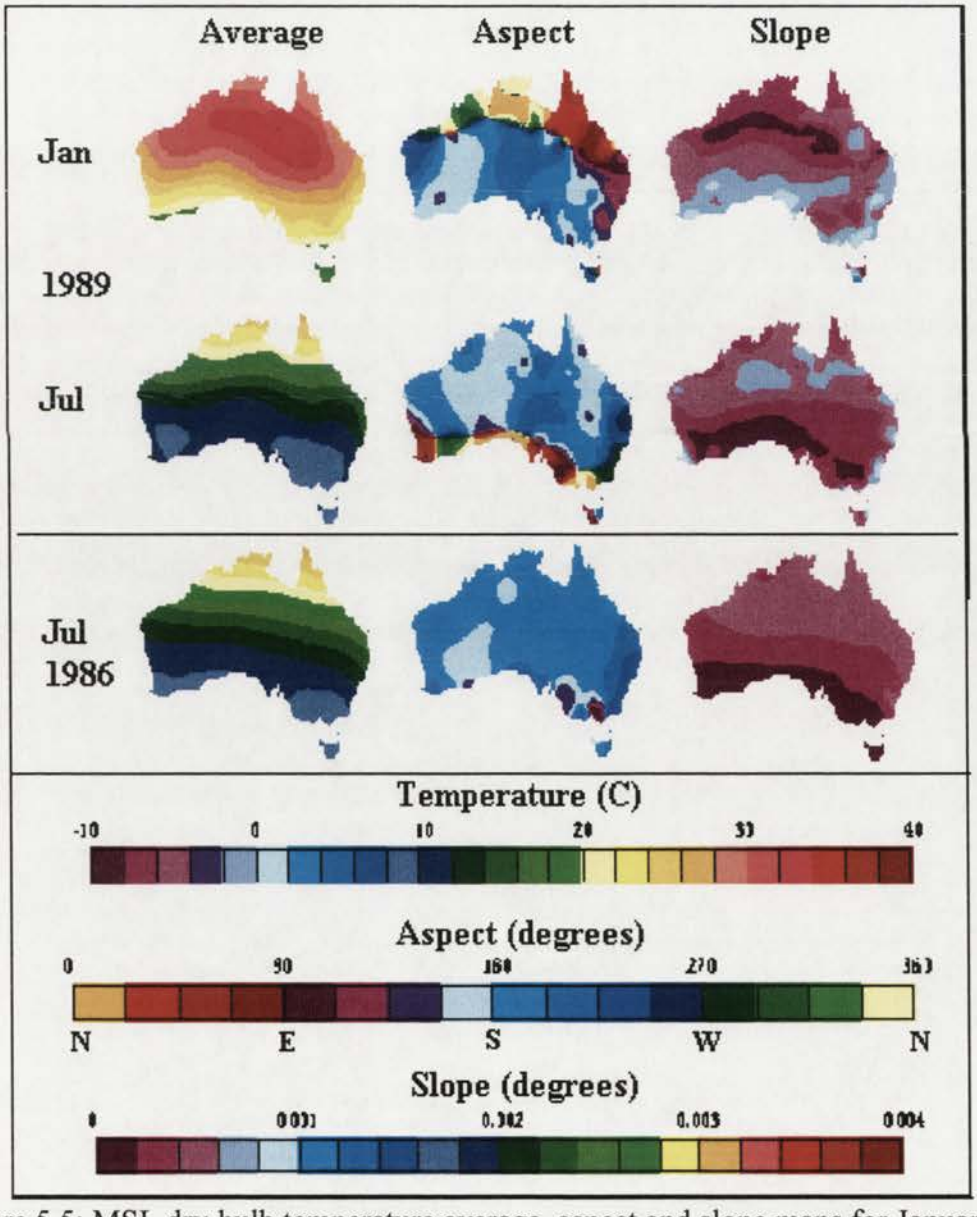


Figure 5.5: MSL dry-bulb temperature average, aspect and slope maps for January and July 1989.

The maps of the slope of the temperature surfaces show areas of greatest temperature gradient; these areas may be the regions of greatest sensitivity to climate change. This possible sensitivity, however, requires further investigation which might be better achieved at a scale finer than the continental scale. The wet-bulb temperature and dew-point temperature aspect and slope maps are also included in the atlas on the CD, but require further analysis.

Temperature difference surfaces

Maps which show the difference between the various temperature values are also given on the CD. Temperature difference maps (calculated by subtracting one temperature grid from another) were created for both 0000 and 0600 UTC for $T - T_w$, $T - T_d$ and $T_w - T_d$. The January and July wet-bulb depression, derived by subtracting the wet-bulb temperature from the dry-bulb temperature ($T - T_w$), are given as an example in Figure 5.6.

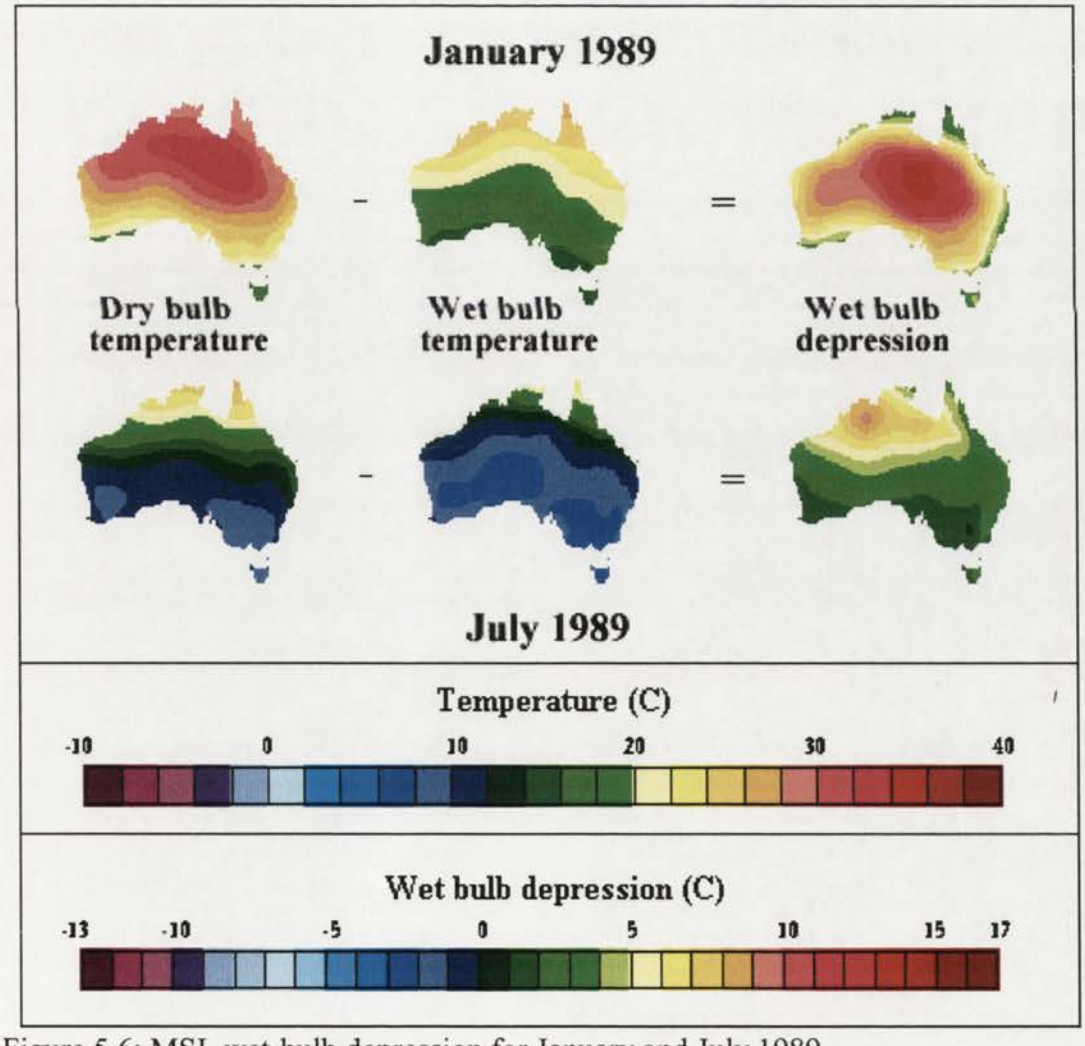


Figure 5.6: MSL wet-bulb depression for January and July 1989.

Logical models representing change over time

In order to represent change over time, it is important to recognise what is meant by 'change over time'. The anomaly surfaces represent the difference at the time of interest relative to the long-term mean. The lag surfaces, on the other hand, represent the change in values over time compared with the values at another specific time (e.g. the previous month). These lag surfaces could be considered to represent change at the monthly time scale. Diurnal variations at the monthly time scale can be shown by surfaces produced by modelling the difference between the 0600 and 0000 UTC surfaces, which are close to the maximum and minimum values respectively.

Anomalies

The anomaly surfaces represent a relative change over time. Generally anomalies are used when a distribution which is independent of topography is required. Certain monthly weather anomalies, for example, display broad spatial patterns which are independent of topography (Walsh *et al.* 1982, Seaman and Hutchinson 1983, Smith *et al.* 1993), unlike the corresponding non-normalised quantities. The anomaly maps, presented in Figure 5.7 and on the CD are independent of altitude as they are derived for the MSL values.

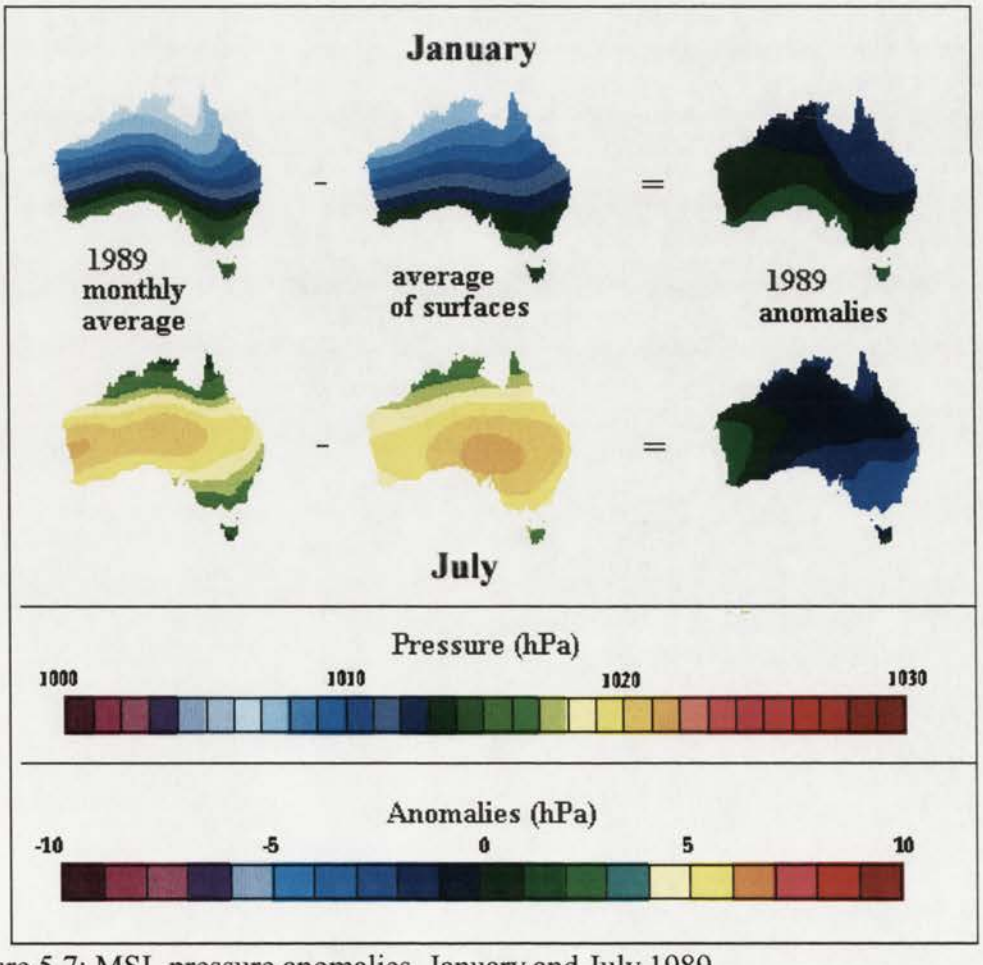


Figure 5.7: MSL pressure anomalies, January and July 1989.

The anomaly maps presented here are the difference between the individual monthly surface (AVG) and the average of the surfaces (SFLTA) (see Chapter 4). It should be noted, however, that these surfaces are not the same as the surfaces produced by interpolating the monthly anomalies for each station. The difference between the two surfaces is attributable to the fact that the SFLTA is the addition of smoothed surfaces rather than interpolated actual monthly anomalies. The difference is similar to those shown in Chapter 4 associated with the different methods of interpolating the LTA. The anomaly grid sets form part of the database for the space-time models considered in Chapter 7.

Monthly lag surfaces

Change at the monthly time scale can be shown by monthly lag maps, as shown in Figure 5.8, which represent the change in the climatic variable between the month of interest and the previous month. The maps do not have the seasonal patterns removed, and thus the normal seasonal variations are apparent. The one-month lag surfaces were produced for all variables for both 0000 and 0600 UTC. Similar methods were used to produce two- and three-monthly lag grid sets which were used as the database for the lag regression model discussed in Chapter 6.

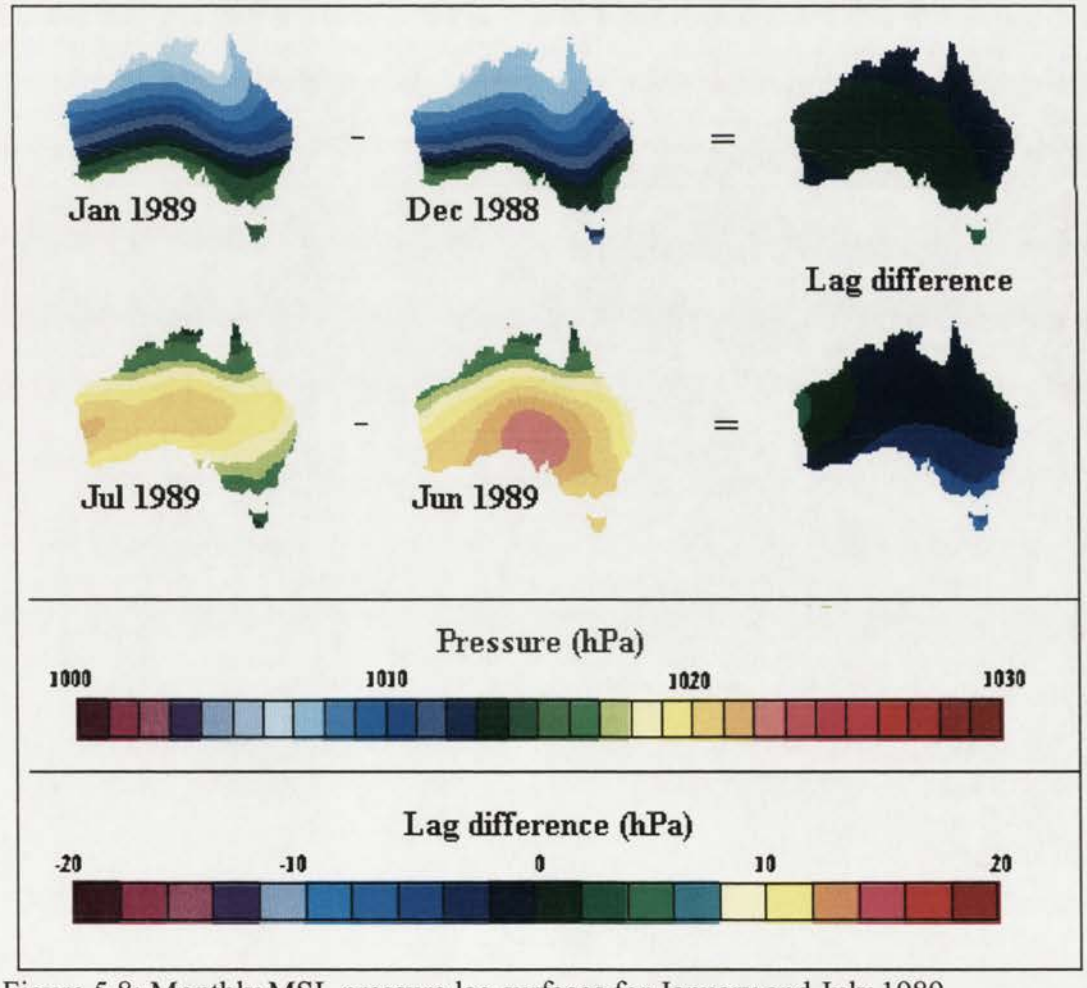


Figure 5.8: Monthly MSL pressure lag surfaces for January and July 1989.

Diurnal difference surfaces

The diurnal variation at the monthly time scale can be shown by the diurnal difference maps, as illustrated in Figure 5.9, which represent the change in the climatic variable between 0600 UTC and 0000 UTC for the month of interest. It should be noted, however, that although the diurnal difference surfaces represent change in a climatic variable at the daily time scale the surfaces are for monthly averages and are at the monthly time scale.

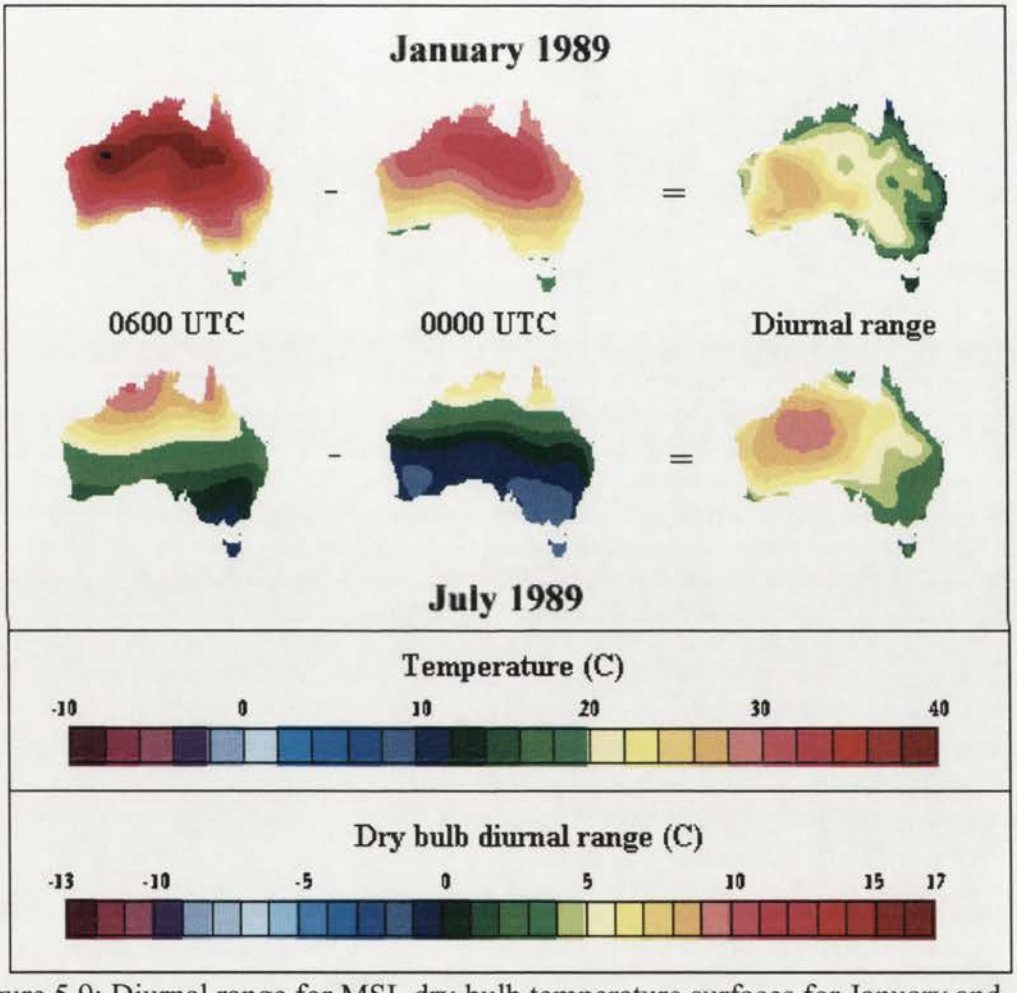


Figure 5.9: Diurnal range for MSL dry-bulb temperature surfaces for January and July 1989.

The diurnal differences surfaces have been produced for pressure and all temperature variables. It should also be noted that the difference between 0600 and 0000 UTC is not exactly the 'diurnal range', which is normally the difference between the minimum and maximum values for a day. The minimum and maximum temperatures, however, do not occur at a specific time over the continent nor do they occur at the same time at any station at different times of the year. A closer representation of the 'diurnal range' could have been produced if the set of primitives had included times closer to dawn than 0000 UTC and closer to the time of maximum heating in the afternoon than 0600 UTC.

Spatio-temporal databases

It is useful at this point to investigate issues concerning the simultaneous support of both space and time in databases. One would expect different geographical applications to have somewhat common requirements for handling temporal data, just as the requirements for handling spatial data. But as Langran (1993) states, fundamental decisions concerning why and how temporal data are to be used have a striking impact on the implementation of a spatio-temporal data system. No single correct choice exists, rather, options exist with advantages and disadvantages.

The grid data model, which uses a sequence of spatially registered grids, is the most common and the only model available within existing GIS. This model stores a complete snapshot of the world 'state' at a known point in time. This approach is straightforward, and the world 'state' for any given point in time within the recorded temporal interval can be easily retrieved or interpolated. The actual changes that occured at locations between given points in time are not explicitly stored, and can only be derived by comparing the pixel value between successive snapshots.

On the other hand, in the event-based approach, the time associated with each change is stored in increasing order from the initial 'world state' at time t_0 to the latest recorded time at t_n. These may be recorded at any desired temporal resolution. For most phenomena, the length of any temporal interval (ie temporal distance between t_{i-1} and t_i), and any other such interval will be unequal. Associated with each t_i are the changes which occurred between t_{i-1} and t_i. The only exception to this that the entire 'base map' or starting world state must be stored with the first recorded time t_0. The changes associated with any t_i may either be extensive, affecting a large geographical area and many features within it, or may be only at a single location or feature. Recording only times when change occurs as opposed to a complete time-line that contains an entry for every 'tick of the clock' at a given temporal resolution can also be viewed as temporal run-length-encoding (Peuquet and Duan 1995).

Several spatio-temporal database storage models have recently been proposed that explicitly record spatial changes through time as they relate to specific geographical/cartographic entities instead of locations (Langran 1992). The first of these models was proposed by Langran (1989) and relies on what she described as 'amendment vectors'. The sequence of events through time representing the spatio-temporal manifestation of some process, is noted via a time-line or temporal vector, i.e. a line through the single dimension of time instead of a two dimensional surface over space. This time line, then, represents an ordered progression through time of known changes from some known starting date or moment to some other known, later, point in time. Each point in time along the time-line can have associated with it a particular set of locations and features in space-time that changed at that particular time and the new value to which they changed. Because this construct provides a terse description of change it is storage efficient if there is little change from month to month. While this method is appropriate for events which do not change frequently such as land use maps, it is not as appropriate with climate data which undergo cyclical change with the seasonal fluctuations.

The extended grid and vector models described above incorporate the temporal dimension in some way while still retaining their fundamental spatial organisational basis. All vector models are feature-based in the sense that all locational and temporal information, as well as other types of attribute information, is stored relative to specific geographical/cartographic features, and/or the topologically-defined elements (lines and nodes) which make up those features (Peuquet 1984). Conversely the grid mode, or more generally any tessellation model, can be said to be location based, since all information is stored relative to specific locations and features are not explicitly encoded (Peuquet and Duan 1995).

Because of these two fundamentally different types of stored information, a vector model can be used most effectively to store information and perform tasks relative to spatial features, including topological relationships between them over space, but not nearly as effective to store information and perform tasks relating to a specific location or set of locations. Conversely a grid or tessellation model can much more effectively perform tasks relative to specific locations and locational overlay.

To provide sophisticated temporal analysis capabilities and to efficiently answer specific types of time based queries, it is necessary to utilise a type of representation which is specifically suited to that type of application. The spatio-temporal data base used in this thesis is maintained in several forms. The 'primitive' set is held as a series of surface coefficient files which can be interrogated for any given month in time within the recorded temporal interval. The advantage of the surface coefficient files is that they are compact. Both the 'primitive' and 'derived' data sets are held in both grid and graphic form. While the grid form of the surfaces is easier to interrogate if held in a GIS, it was not practical to include these on the accompanying CD due to the large amount of computer storage space required (approximately 20 gigabytes). The graphic database requires only one tenth of the computer storage space required by the grids and is presented on the CD.

Mapping techniques used to describe the spatio-temporal database

There are several different mapping techniques that are available to describe spatio-temporal data. The techniques which were used in this thesis include:

- graphs which show a time series at a point or aggregated for an area.
- static maps which show the spatial distribution of the variable of interest at a fixed time.
- time sequences of maps, e.g. multiple editions or composites.
- animation, where both space and time are scaled to depict the metamorphosis of a study area.
- the use of sound
- combining animation and sound

Each technique has certain advantages which are discussed below.

Graphs showing long term averages or trends in the data set

Interest in the statistical nature of climate change has heightened in recent years (e.g. Pittock 1988, Houghton *et al.* 1990). Human induced climate change such as the greenhouse effect and urban heat islands are at the fore. A necessary prerequisite for studies of the causes of climatic change and variability is a comprehensive set of long-term climatic data for as large an area as possible. The regular grid format is virtually mandated by the requirement to interface with visual representation packages and for climate modelling studies. The time series developed in this thesis represent a new compilation of pressure, temperature and precipitation data for the Australian continent for January 1952 to December 1990.

The long term Australian monthly average temperature and pressure values

Figure 5.10 shows the Australian monthly average temperature and pressure values which were calculated by averaging the 0.1° grid cells for all months from 1952 to 1990. These values are used in the spatio-temporal models discussed in Chapter 7.

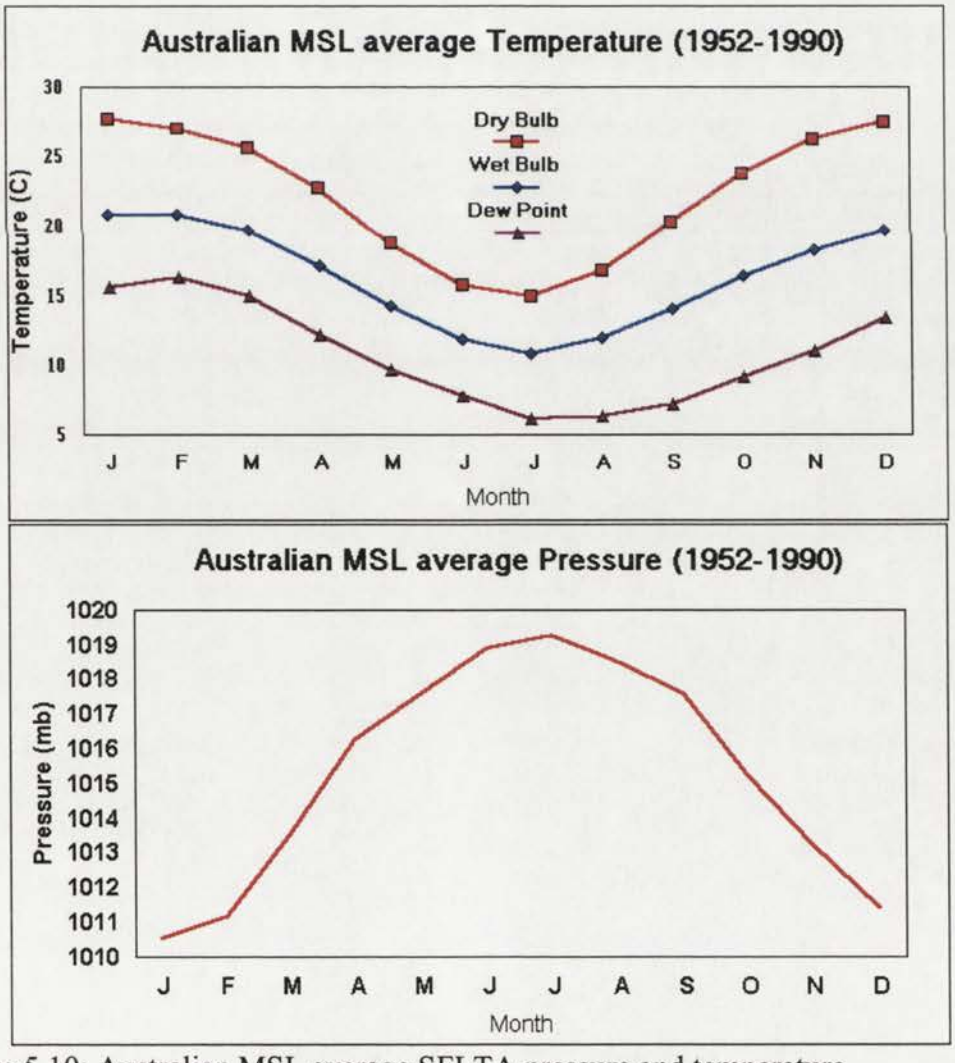


Figure 5.10: Australian MSL average SFLTA pressure and temperature.

Previous studies of trends in temperature and pressure over Australia

The investigation of pressure in the Australian region has mainly been concerned with the relationship between synoptic features and regional weather patterns (see Chapter 1). Few studies have investigated time series of Australian MSL pressure. Jones P.D. (1991) investigated Southern Hemisphere MSL pressure data on a 5° latitude by 10° longitude grid. Jones P.D. (1991) gives seasonal and annual time series for four indices which are based on the average of grid points in the region of the major oceanic anticyclones. Jones P.D. (1991) found statistically significant changes in the South Pacific anticyclones (-1.0 hPa) and the South Atlantic anticyclone (+0.6 hPa). Karoly et al (1996) used the Jones P.D. data set and found that the leading mode for interannual and interdecadal variations is associated with the El Niño/Southern Oscillation, but there are periods when it has reduced variance and somewhat different structure (e.g. 1916-1935).

Several studies have investigated changes in surface air temperature over Australia. In earlier studies data inhomogeneities have made it difficult to compile an accurate data set, although warming trends in Australian temperature have been identified by using uncorrected data (e.g. Tucker 1975 and Coughlan 1979). Positive trends were also identified in shorter or smaller datasets that took into account urbanisation (Coughlan et al 1990) and discontinuities (Plummer et al 1995).

Jones et al (1986) and Jones (1994) overcame the temporal and spatial discontinuities by using gridded global air temperature anomaly data for the Southern Hemisphere. The gridded data set was generated from the station anomalies by interpolating average station data that were weighted according to the inverse of the distance of the station to the nearest grid point. They found it necessary to reduce all station data to anomalies because of different station elevations and, to a lesser extent, different observation times. In the study done by Jones et al (1986) more than 50% of gridded anomalies used data from a single station in 1990, and the best period, in 1960, still had 40% of gridded anomalies using data from a single station.

Jones P.A. (1991) used 316 stations from the BoM to look at trends of cloud cover and climate. Jones P.A. (1991) found cloud cover and mean temperature have increased while diurnal temperature range has decreased since the beginning of the century. More recently Torok and Nicholls (1996) and Plummer (1996) have detected warming trends using the BoM's high quality historical air temperature dataset for mean annual temperatures. The BoM's high quality historical air temperature dataset has been adjusted for inhomogeneities caused by station relocations changes and other discontinuities. As the dataset is comprised of minimum and maximum annual temperatures, temporal discontinuities caused by the introduction of daylight savings time are unlikely to lead to artificial discontinuities in the series. Torok and Nicholls (1996) calculated spatial averages using a technique based on Theissen polygons (see Chapter 3), while Plummer (1996) combined unweighted individual seasonal and annual station times series into regional averages.

This analysis of Australian continental sea-level pressure and temperature avoids some of the space and time biases of previous studies. The use of the specific UTC times was specifically implemented to overcome the different observation times (see Chapter 2). The ability of the spline function to detect data inhomogeneity and the data smoothing to help overcome many of the dataset problems encountered in the past. The use of interpolation techniques applied to a smaller study area avoids the problems associated with sparse data encountered by Jones et al (1986). A majority of previous Australian temperature trend studies have focused on maximum and minimum station values. This analysis presents areal average gridded MSL values for specific UTC times for temperature and pressure. It should also be noted that most previous studies have concentrated on annual values while this study is based on monthly values. A graphical presentation of individual monthly trends are included on the CD.

Trends in the spatial temperature and pressure datasets

Each of the pressure and temperature surface coefficient files have been used to produce a 0.1° grid of sea-level monthly means over the Australian continent from January 1952 to December 1990 for both 0000 UTC and 0600 UTC. Spatial averages were calculated for each month using each point of the 0.1° grid on the Australian continent to give a single figure representative of an Australian mean. The purpose of this was to review the behavior of Australian continental sea-level pressure and temperature from January 1952 to December 1990 based on the sea-level surfaces. Figure 5.11 shows 120 month (10 year) moving average for pressure and temperature for 0000 UTC and 0600 UTC.

The main problem associated with assessing an overall mean in this fashion is that large areas whose interpolated values are based on few stations, such as in the centre of Australia, could unduly influence the final mean. It should be noted that the mean values are based on the continental 0.1° grid points. A rectangular grid was not considered a viable option as the errors associated with extrapolated offshore points are large due to increasing error away from data points.

The increase found in the dry-bulb temperature (~0.7°C) over the period 1952 to 1990, if a true indication of the temperature trend, is alarming. An investigation of regional values (taken from a 10° subgrid extracted from the 0.1° grids) of the long term dry-bulb temperature showed increases in every region although some areas show no warming trend until the 1970s (see Figure 5.12). This late increase in the period, although it is considerably larger, is substantiated by Jones et al (1986) and Jones (1994) who found the Southern Hemisphere area average showed little or no overall trend between about 1945 and 1970, but that since 1970 a strong warming trend has set in. Jones et al (1986) and Jones (1994) found that the overall warming since 1900 is about 0.5°C, of which roughly 0.3°C occurred between 1900 and 1945 and 0.2°C since 1970.

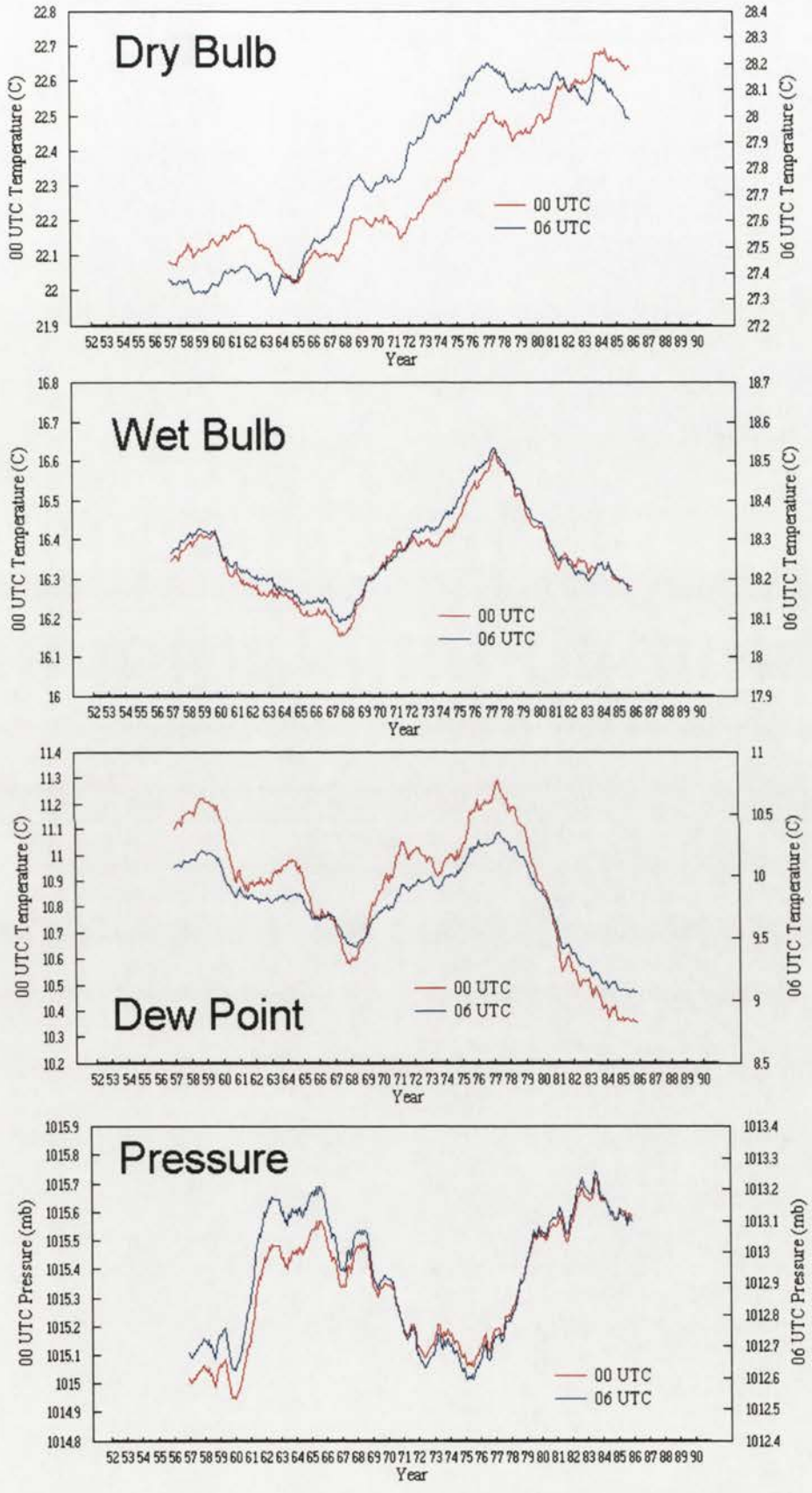


Figure 5.11: 120 month (10 year) moving average of the Australian MSL average for temperature and pressure, 1952-1990.

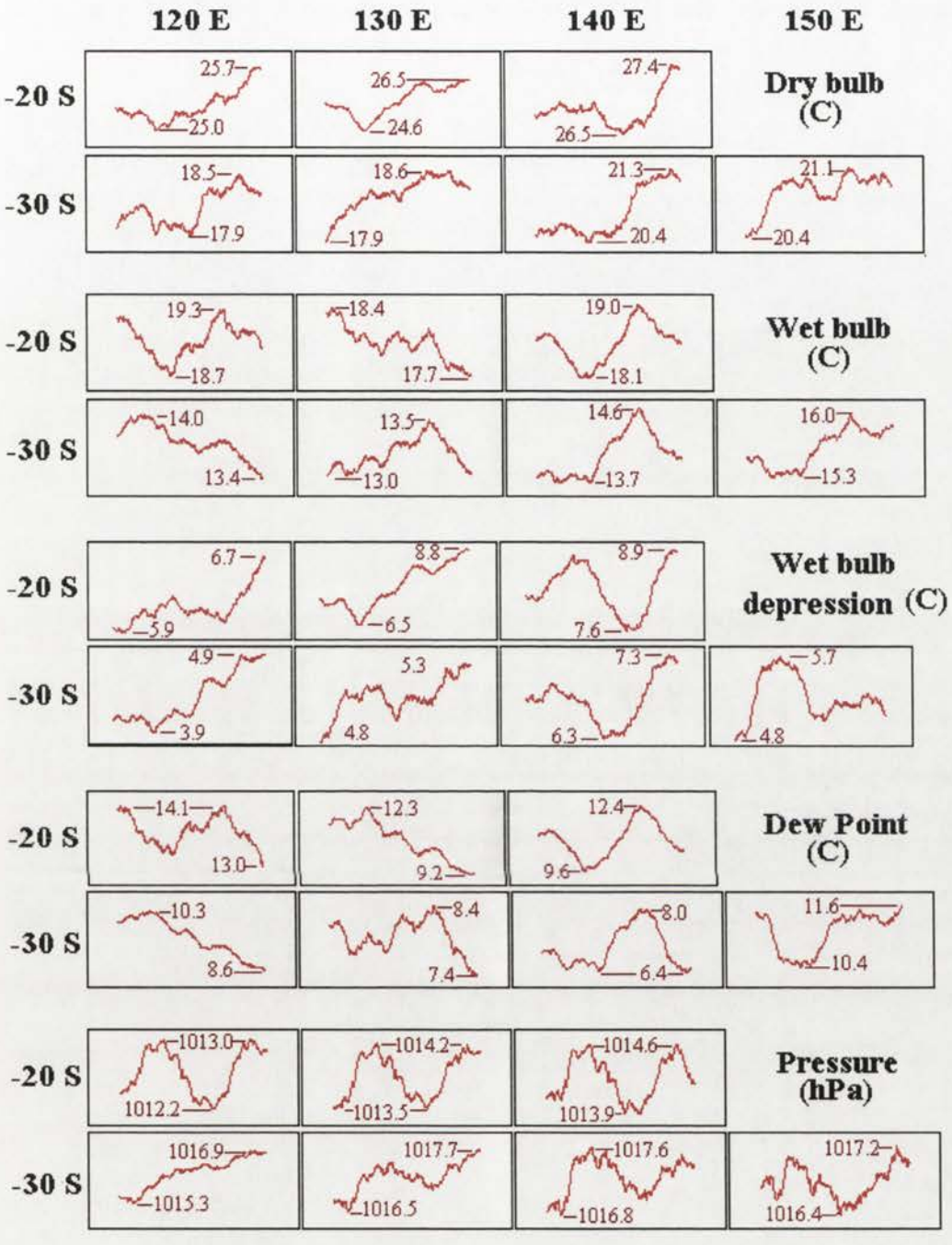


Figure 5.12: 10° grid point 120 month (10 year) moving average for pressure and temperature (dry-bulb, wet-bulb, wet-bulb depression, and dew-point), 1952-1990.

The 0.7°C increase in the dry-bulb temperature is also larger than that found by Torok and Nicholls (1996) for minimum (0.98 °C/100 yrs) and maximum (0.54 °C/100 yrs) temperature from 1910. They do, however, note that the trend has been strongest since the middle of the century. The increase in the dry-bulb temperature is within Plummer's (1996) assessment that mean temperatures have increased at a rate of 0.1 to 0.2 °C per decade over most of Australia since 1951. Plummer (1996) also

found that larger increases have occurred in some areas, especially in the northeast. This spatial variability in the temperature trends can also seen in Figure 5.12.

The average pressure, wet-bulb temperature and dew-point temperature graphs do not show similar rising trends. The wet-bulb and pressure graphs are almost mirror reflections. The pressure peaks and troughs, however, seem to preceed the wet-bulb temperature troughs and peaks. While all the variables shown in Figure 5.12 have similar patterns, the shortness of the length of record (39 years) may easily hide longer term trends and patterns.
The dew-point temperature graph, not surprisingly, is a reflection of the dry-bulb and wet-bulb temperature graphs. That is, the smaller wet-bulb depression the greater the dew-point temperature (see Figure 5.13). It is interesting to note in Figure 5.13 that there are two distinct trend lines in the correlation between the wet bulb depression and the dew-point temperature. These two distinct periods are separated by the very wet 1973-1974 period which suggest that a switch in states occurred in this period (see also "Trends in the rainfall percentiles" below).

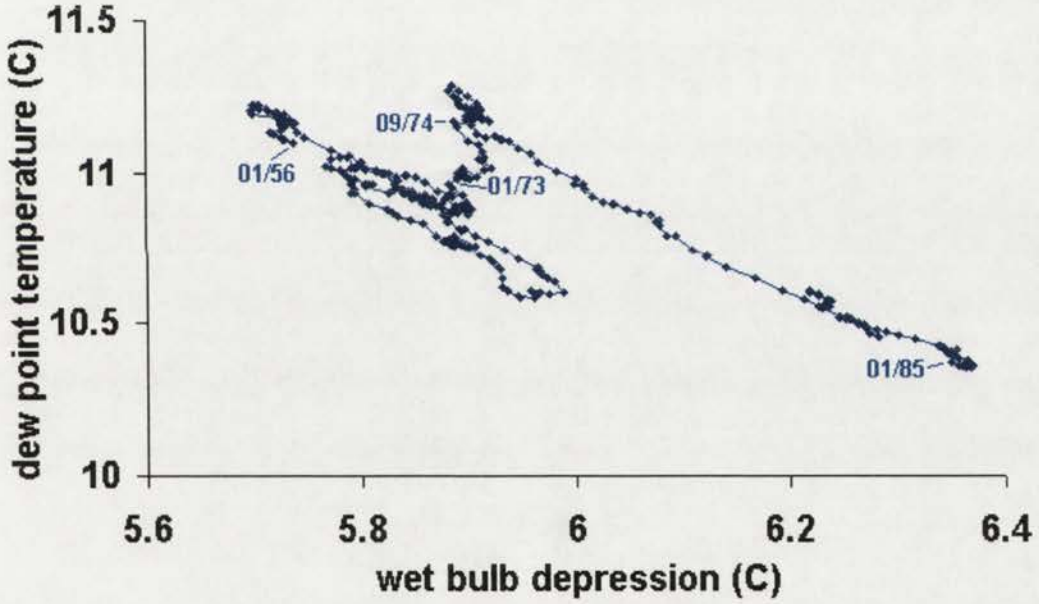


Figure 5.13: 120 month moving average wet bulb depression versus the 120 month moving average dew point temperature average 1952 to 1990.

There is a possibility that the data are biased toward an increasing trend due to the stations being mainly located within urban areas or at aerodromes. Other attractive climatological explanations for variations in these parameters exist. For example, observed change in sea surface temperature (SST), if representative of reasonable changes in vertical thermal structure, could result in the observed temperature changes. Changes in the moisture regime could be due to changes to larger circulation features such as the Southern Oscillation.

Previous studies of trends in rainfall over Australia

The seasonal cycles and long term trends in Australian rainfall have been studied by several workers. Srikanthan and Stewart (1991) and Lavery et al (1997) and references therin give an extensive review. Many authors have studied El Niño/Southern Oscillation relationships with Australian rainfall. For a comprehensive coverage see for example Allan *et al* (1996), Kane (1997) and references therin.

Trends in the rainfall percentiles

Figure 5.14 shows the Australian mean percentiles and the SOI for January 1952 to December 1990. The Australian mean percentiles were drier (below 0.5) in the first half of the period than in the second half. This trend is also evident in work done by other authors (see for example, Lavery *et al* 1997).

There appears to have been a change in the relationship between rainfall percentiles and the SOI. From 1950 to 1970 a SOI value of zero corresponds to a drier continent, while a negative SOI in the 1970 to 1990 period occurred while the continent was wetter. As was the case for the wet-bulb depression and dewpoint correlation, the switch coincides with the very wet period experienced over Australia in 1973-1974. It should be noted, however, that this is for 10 year moving averages and they may hide some things. It is also interesting to note that the variability of the Australian dry-bulb and dew-point environmental lapse rates (Figure 4.10) have similar patterns to the Australian average rainfall percentiles. This may be due to the reduced diurnal temperature ranges in cloudier periods which dampen the continental effect.

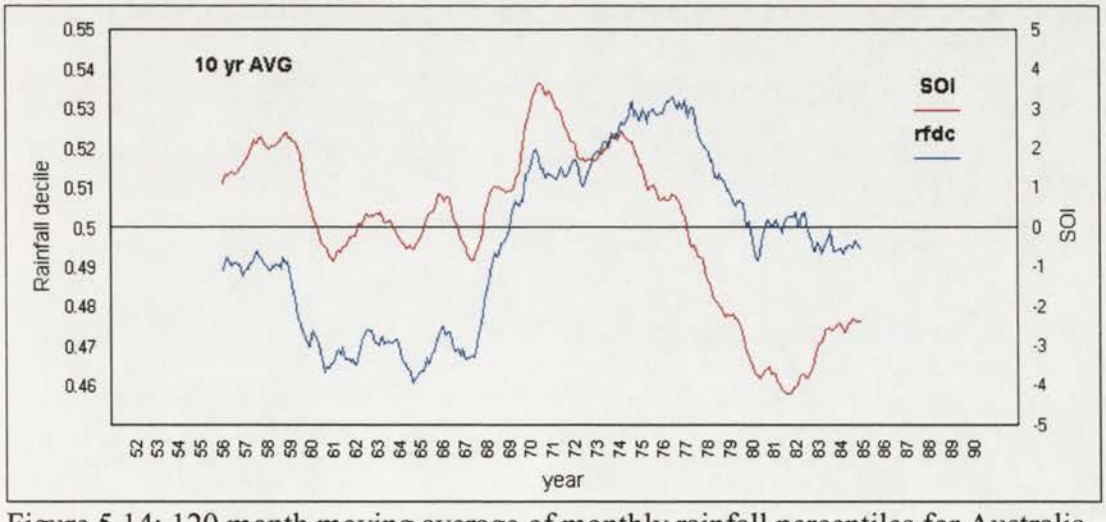


Figure 5.14: 120 month moving average of monthly rainfall percentiles for Australia and the SOI from January 1952 to December 1990.

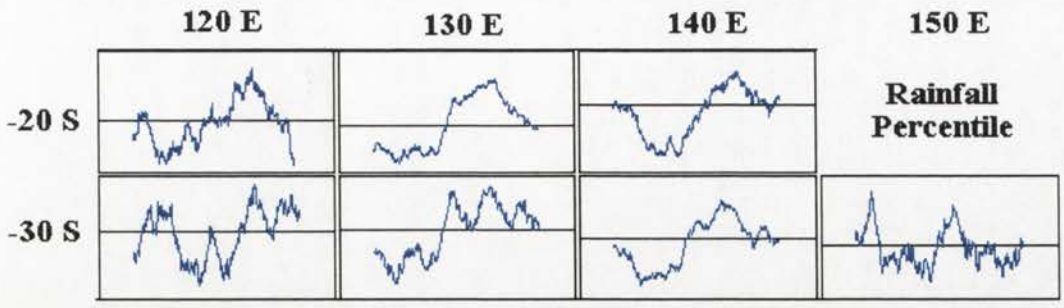


Figure 5.15: 120 month moving average of monthly rainfall percentiles for Australian 10° grid points from January 1952 to December 1990.

It may have been better to only compare the SOI to an eastern Australian average as it is generally acknowledged that this region is under a greater influence of the SOI (Allan 1988, Allan and Haylock 1993, Allan *et al.* 1996). However the regional 10° grid point long term means in Figure 5.15 shows that most of the points reflect the Australian mean rainfall. It is interesting to note, however, that only the point whose rainfall pattern closely resembles the SOI long term trends is the point 150°E 30°S.

Map images

The power of the image far exceeds that of the printed word or, for that matter, the matrix of numbers that compose the data set from which the image is generated. However, the simple notion that an image of something will tell us all, or nearly all that we need to know is far from true, since the fundamental nature of an image is that it is dependent on the object being observed and the technology used to interpolate and display the observed data.

Scientists generally need to perform one of two tasks with an image. Either to determine, quantitatively or qualitatively, the gross features of a dominant structure of phenomena or to find small-scale meaningful structures. On the whole, contours are used to display f(x,y) and vectors are be used for displaying magnitude and direction (e.g. wind). Colours can enhance specific morphological features of a given image, although gray-scale can provide better structural detail to the eye (e.g. visual meteorological photographs). Wolff and Yaeger (1993) suggest that the intent should be to use colour intelligently and consistently with the data analysis, presentations and publications. Images of scientific data are almost always tied to a specific numerical database: a matrix of numbers, either integers or floating-point. These numbers are precise representations of data and are immutable, although not necessarily error free (see Chapter 4). There is a need to ensure that the colour range, as defined by the mapping of the data values to colour space, visually represents the qualitative features in the data.

Another problem that scientists face is the effect of the eye's nonlinearity on the perception of the relationship between data points (or pixels) in an image of a data

set. Thus, for example, a highly nonlinear function, or a small region of large numerical fluctuations in a data set, could get mapped onto a colour range (e.g. violet) that appears narrow to the eye. Conversely, a relatively uninteresting region in a data set could get mapped onto a widely varying region of colour space (e.g. orange). Hence, choosing the proper mapping is critical to the efficacy of the data visualization.

The colour scales used in the graphics in this thesis are intended to be intuitive where possible. For example, the colour scale used for all the temperature maps progresses from purple and blue (cold) through green and yellow to red (hot). Similarly the rainfall percentile scale ranges from blue (wet) to green (normal) to red and brown (dry). Although pressure maps are not usually coloured (done here to make multiple or animated viewing easier as numbers on isobars in smaller graphics are hard to see), the scale ranges from blue (low) to red (high).

Snapshots and composite images

As we continue to add parameters the visual representation of the information can become a little cluttered. The most common approach is to pile on more visual information until the viewer gets confused. The next most common approach is to split the data into two or more images, each with a different set of parameters. This approach solves some problems but can make it difficult to determine the relationships between the variables.

The graphic for each of the surfaces in the database is a snapshot representing the variable of interest at a known point in time. Each cell within a separate snapshot contains the value for the corresponding location at that time. This approach is conceptually straightforward, and the 'world state' for any given point in time within the recorded temporal interval can be easily retrieved or interpolated. The actual changes that occurred at locations between given points in time are also stored in the database (see 'derived' surfaces above). At times, however, rather than view the amount of change, it may be better to compare successive snapshots.

Multiple time-slice snapshots are an intuitively appealing space-time model. The nature of each time slice captures the spatial distribution of the phenomena for the month of interest. The multiple images in a composite picture can convey a sense of the change that occurs over a period of time. For example, a composite picture could include all of a specific month's values for all years to show the annual variation (as in Figure 5.16) or it could show all the monthly values for a specific year to show the variation within a year (see the CD for these graphics).

The composite pictures, however, are not necessarily storage efficient for some environmental data as a complete snapshot is produced at each time slice, which duplicates all the unchanged data. However, when considering climatic phenomena which change continuously a large database is required. While the technique is useful

for examining shorter periods of time, there is a limit to the number of pictures that can be included in a composite.

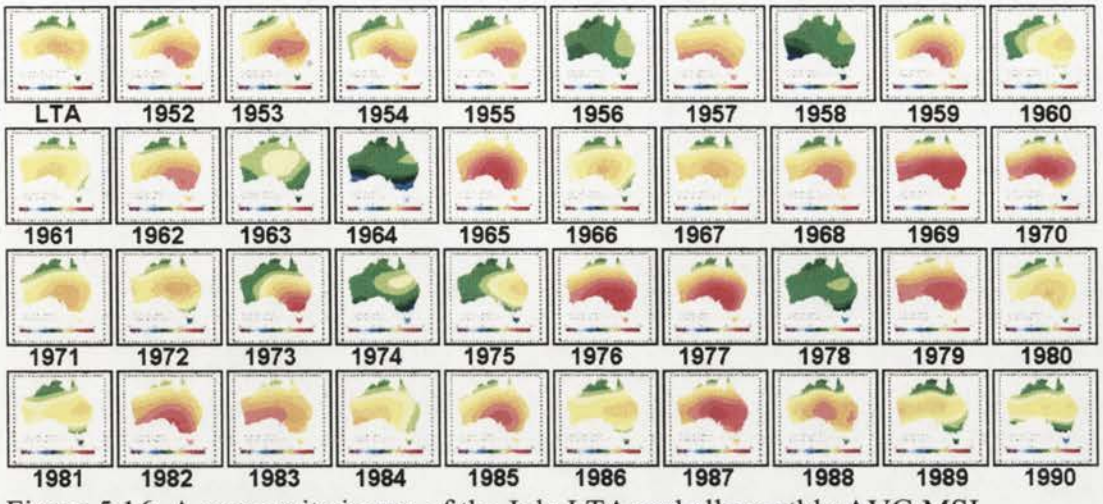


Figure 5.16: A composite image of the July LTA and all monthly AVG MSL Pressure snapshots 1952-1990.

Animation

Until recently film or videos of sequences of images were considered an extravagant luxury for researchers or, at best, a tool used for public lectures and program reviews. Today, however it is fairly easy for researchers to create animated sequences of images and, as animation is a basic medium of exchange of time-dependent, visual information, they are being used more regularly. As a result, it is now widely accepted that sequences of images of phenomena in any field can clearly be of value when one is interested in time-dependent as well as spatially varying problems.

Animated climatological time series were created for the long term averages (12 images) and for all the 44 primitive and derived variables for the monthly values from January 1952 to December 1990 (468 images). To view these animations requires a desktop animation package (a shareware animation package is provided on the CD typical of the type used to view the animations).

While the animation of the spatial time series is a useful communication tool and an interesting analogy to time-movement in space, it does have some inherent problems. Critical parameters which need to be determined include:

- number of frames per second
- resolution or image size
- colour depth

The speed or number of frames per second of the animation can greatly affect the interpretation of the viewer. To make sense, there has to be some temporal continuity in the spatial patterns. Different time compressions show different things in the time series. Wolff and Yaeger (1993) suggest that 10-15 frames per second is optimal as

the maximum time interval that the eye-brain system can perceive two sequential images as continuous is around 0.1 second. They therefore suggest that 5 frames per second are too jerky and disconnected to easily be perceived as a continuous animation. When viewing a movie we do not normally only look at one frame, but rather view each frame in relation to the others and from that extract the plot. On the whole the climate surfaces have continuity but at speed greater than 10 frames per second the speed is too quick to perceive monthly features and the seasonal plot becomes the dominating theme. Obviously there is an upper limit to the speed of the animation, where the pictures just become a blur. Equally a speed too slow fails to give the illusion of the passage of time. Clearly then, there is no 'best' speed for viewing the animation. Most animations programs allow for the user to change the speed of the animation.

The resolution or image size for the animation was chosen to best make use of available storage space. However it is possible to animate the individual snapshots which are twice the size of the animations. The depth of colours used was also predetermined by the choice of colours used to produce the maps.

The inherent problem of animated spatial time series is that our spatial knowledge has its limits and seasonal variation predominates when viewing the time series. Our spatial memory does not allow us to easily compare months in successive years, such as how one January compares with the next.

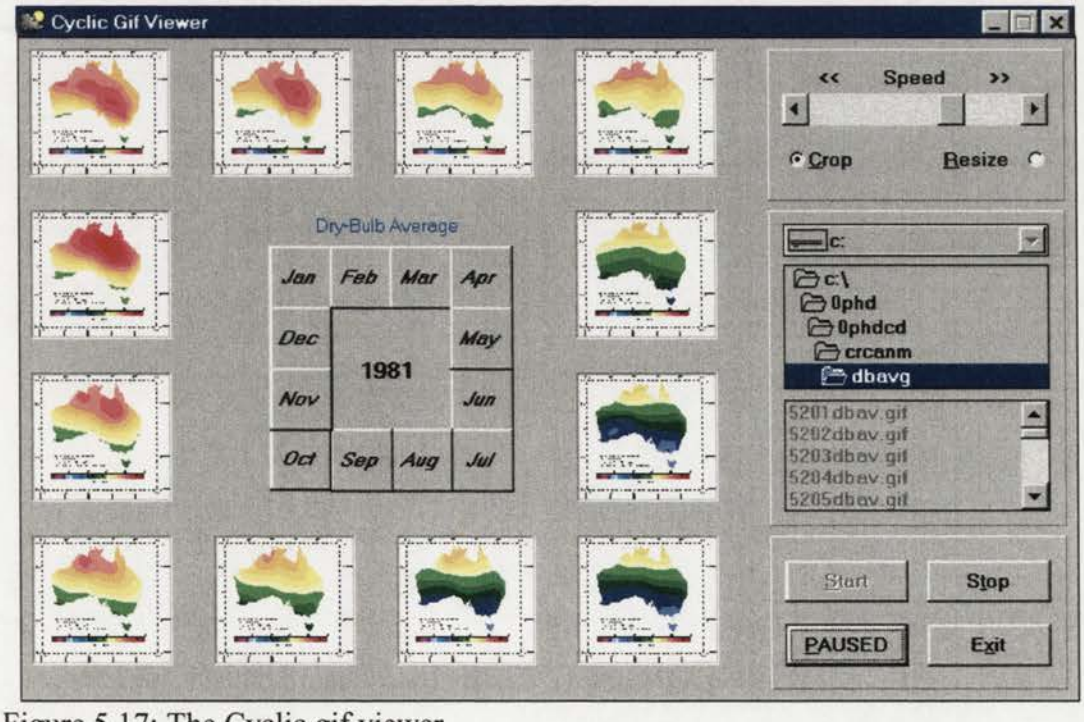


Figure 5.17: The Cyclic gif viewer.

A different method of animation was developed in conjunction with Jim Throssell in order to give a different perspective of a time series animation. The cyclic gif viewer

(see Appendix A) allows the viewer to see all 12 months with the replacement of each month in a clockwise animation (see Figure 5.17). This allows the viewer to see the individual months in context of the annual pattern. The program also allows for selected months to be turned off, so that one can animate just mid-season months or individual months.

There are other kinds of change which can be visualized using map animation. Spatial change, often called a "fly-by", is visualized by changing the observers viewpoint of some static object. Computerised flight simulators provide an excellent example of visualized spatial change. Possibilities could include a fly-by of a three-dimensional orthographic surface Fly-throughs are good for elevation, thus for displaying terrain data. Some of these techniques may give a better spatial understanding of a single month, however as the MSL surfaces are relatively simple or smooth, these were not created.

Sound

The traditional view of cartography (two-dimensional space) and visual variables may not be adequate for the needs of researchers to examine complex dynamic and multi-variate phenomena. The current generation of computer hardware and software allows access to a broad range of design options such as three-dimensionality, animation and sound. One problem that researchers constantly face in attempting to visualize complex systems is that they run out of visualization techniques since there are generally more parameters than there are distinct ways of collectively showing the data. To display more parameters one could use height or transparency, but the visual image is generally too cluttered for the eye to pick out specific features. Any more visual information would do more to confuse than to elucidate the problem. The use of sound as an exploratory "visualization" tool, has only evolved to the point where it could be considered a useful tool for scientific data analysis within the last few years. Sound used alone or in tandem with two- or three-dimensional space, provides a means of expanding the representational repertoire of cartography and visualization.

Our sense of vision is often perceived to be much more dominant than our sense of hearing. However, the sonic aspects of space have been undervalued compared to the visual (Ackerman 1990, Tuan 1993). Human communication is mainly carried out via speech and we commonly use audio cues in our day to day lives. The visually impaired use sound to aid in their understanding of the environment.

Sound is inherently a temporal phenomena. As a result it is particularly suited for use with map animation. Sound can be closely linked to the dynamic variables and their applications and may be used to enhance the comprehension of the information presented in a dynamic display. In addition, potential uses of the dynamic variables may be suggested by examining temporal issues in sound and music.

MacEachren (1995) and Krygier (1994) suggest that sound patterns may be more easily distinguished than visual patterns and are especially valuable for dealing with cyclic temporal data. It is also possible that we are better able to perceive repeated patterns in a time series aurally (due to our chorus history) than visually. The application of sound becomes more important as the amount of data being visualized increases while we seek to isolate the few interesting patterns from the noise.

Uses of sound in geographic visualization include sound as:

- a vocal narration,
- a mimetic signal,
- an additional variable,
- a means of detecting anomalies,
- a means of reducing visual distraction,
- an alternative to visual patterns,
- an alarm or monitor,
- a means of adding non-visual data dimensions to interactive displays
- and for representing locations in a sound space.

Vocal narration is an obvious and important use of realistic sound. Voice over has been used with fly-by applications to provide an explanation of what is being seen (Jet propulsion Laboratory 1987). Vocal narration is an important way of using sound to enhance dynamic geographical visualizations. Gaver (1989) gives another use of sound as a mimetic sound icon, for example, the sound the computer makes when it has successfully opened windows. Abstract sounds can be used as cues to alert or direct the attention of users or can be mapped to actual data.

The ability to locate sounds in a two-dimensional "sound space" analogous to the two and three dimensions of the map, is an important aspect of sound which is particularly applicable to the display of spatial data. The location of the sound can be used in an abstract manner, as a cue to direct attention to a specific area of a visual display. Such applications of sound have been investigated but not in terms of geo-referenced data. Questions concerning hardware and software requirements (for two- or three-dimensional sound generation) and issues of the human ability to adequately locate sounds in a sound space need to be investigated.

The use of sound displays have been explored in the context of communicating scientific data to visually impaired students. Mansur et al (1985) compared tactile graphs to sound graphs (created with a continuously varying pitch) and evaluated subjects based upon judgments of line slopes, curve classification, monotonicity, convergence and symmetry. They found comparable accuracy of information communication capabilities between tactile and sound graphs, but the sound graphs were a quicker way of communicating information.

The sound sequences

The Musical Instrument Digital Interface (MIDI) specification was developed in the mid-1980s to allow music and audio production data to be transferred between different manufacturers' instruments. The MIDI allows a postscript-like code to represent certain attributes of a particular sound as it is played on a MIDI-equipped musical instrument.

The MIDI concept seemed to have potential for interactive data analysis, especially since personal computers can interface to MIDI instruments. Moreover, in contrast to the sampled data problem, where significant computation is required to digitize and play a data set, MIDI allows many sounds to be played at once. Within this context one could imagine using MIDI-based sound as an "exploratory" tool to study multi-spectral data. For example, one could use a mouse to roam over a multi-spectral image-based data set that had mineralogical data embedded under an image or a terrain-rendering of a particular region, much the same as a doctor uses a stethoscope to listen for abnormal sounds in a patients body.

MacEachren (1995) and Krygier (1994) show that there are many sonic attributes including pitch, timbre, register, rate of change, order, volume, and depth of frequency modulation. In addition to pitch and amplitude, other sound attributes that can be manipulated independently are the attack and decay rates. The human ear is particularly sensitive to attack rates, and short-attack-rate instruments such as the piano are perceived entirely differently than long-attack-rate instruments such as the violin.

There are several MIDI sound sequences included in the CD. The 0.1° grid values discussed earlier in this Chapter and several point location time series were used to create the sound series. Each sound sequence is created by transforming the anomalies to a musical scale. Figure 5.18 shows the dry-bulb and wet-bulb temperature anomalies for the first three years of the time series. The temperature scale is on the left hand y axis and the note to which it is transformed on the right hand y axis. The temperature anomalies were converted by representing values between -0.25 and 0.25 by high C, and each 0.5 degree Celsius away from these values represented by a semitone.

Once the anomalies have been converted to a musical note value they were inserted into a MIDI file. Three-four time was used with each month being represented by a quarter note and thus each bar represents a year. If an anomaly persists for longer than a month then the length of the note is extended (i.e. not all months are represented by an individual note). Thus the natural variations in the variable of interest produce the rhythm of the sound sequence.

The C major scale was originally chosen, although this also raises questions as to whether the final sound sequence is a product of the transformation into a scale

sympathetic to "Western" ears (the C major scale), and whether a person used to a different scale would interpret the sound sequence in the same way. Figure 5.19 shows the converted musical notation for dry and wet-bulb for January 1952 to December 1954 (see Figure 5.18 below).

Musical sequences were created for dry-bulb, wet-bulb and dew-point temperatures. All temperature values were transformed with each semitone covering 0.5°C of anomaly. The pressure anomalies were transformed with each semitone covering 1 hPa of anomaly. The SOI (5 per semitone) and the rainfall deciles (10% per semitone) were also transformed. A mirror image of the later two variables (i.e. converting low-to-high to high-to-low) were also created as some climatic variables vary inversely to these variables. For example high pressure is associated with low rainfall and vice versa, in order to hear the associations between them the sounds are easier to compare if they change pitch in the same direction.

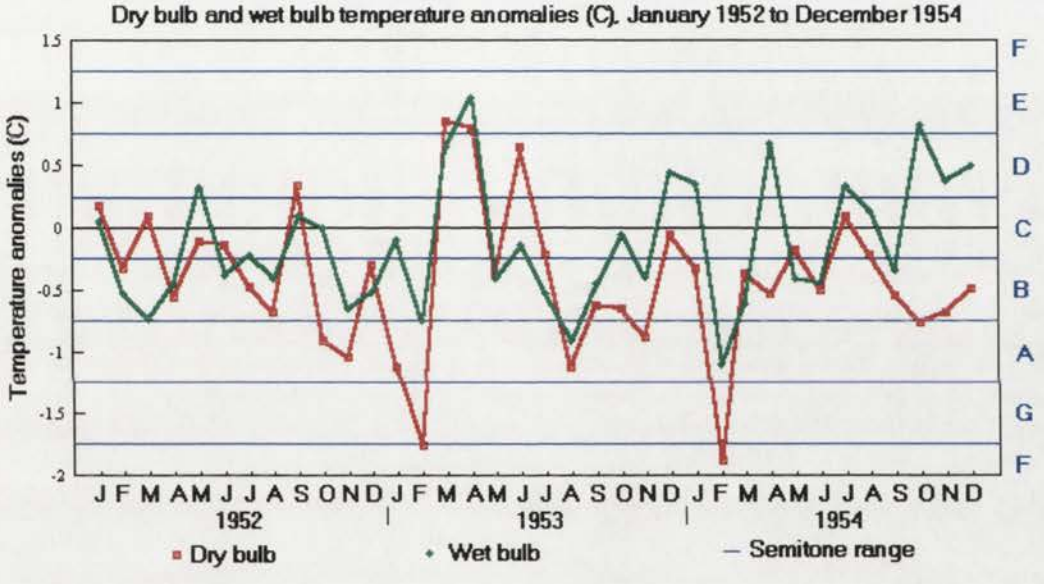


Figure 5.18: The dry and wet-bulb temperature anomalies January 1952 to December 1990.

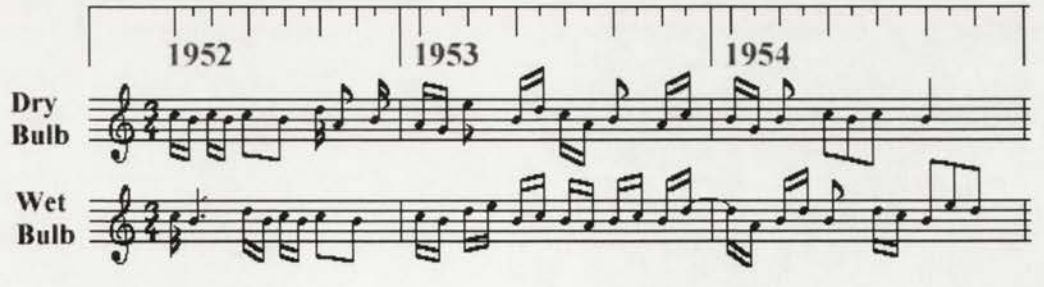


Figure 5.19: The converted musical notation for dry and wet-bulb for January 1952 to December 1954.

Just as the speed of viewing an animation can be shown to produce different interpretations to the viewer, so too can the speed of the sound files can give different

impressions of the time series. The sound files which were recorded at 30, 60, 90 and 120 beats per minute which are the equivalent of 2, 4, 6 and 8 months per second, are presented on the CD.

Combining sound and animation

Animation programs, such as the one on the CD, allow for a sound file to be played in conjunction with the animations. For example, one can view the animation of the dry-bulb temperature in conjunction with the dry-bulb music which gives the viewer a relative sound which represents how hot or cold that month is relative to the long term average. One can watch the dry-bulb temperature animation while listening to a different variable such as the SOI or a number of other variables. Alternatively, one can watch an animation while listening to a sound sequence for a specific place which gives an overall impression from the graphic while receiving specific site information from the sound file.

In order to synchronise the timing of the combination, the sound files which were recorded at 90 beats per minute require the animation to be run at 6 frames per second. Similarly those at 30, 60 and 120 beats per minute require frame speeds of 2, 4 and 8 per second respectively. While these speeds are lower than those suggested by MacEachren (1995), they allow for more of the change between frames or notes to be perceived. At faster speeds the music tends to blur and the seasonal cycle becomes the dominant visual image.

Conclusions

A set of 'primitive' surface coefficient files and associated MSL grids were interpolated from the station climate database. Both 0.1° grids and colour images were created and the images of each of these grids are presented on the CD. While this is a fairly coarse resolution, climatic representation at a finer scale would have required the inclusion of additional stations. This could have been achieved by the adjustment to UTC of the 9am and 3pm data from stations which only have two readings a day using a regression method but this was considered beyond the scope of this thesis.

The output statistics and the graphics suggest that the surfaces are a faithful representation of the climatic conditions prevailing at the time and are similar to previously produced maps (see Chapter 1). The differences between the 'primitives' and previously produced maps can be attributed to different climate databases (both temporally and spatially) and different methodologies for determining the long term averages.

Logical models have been used with the set of 'primitives' to compute attributes for the derived surfaces. Logical models can be created for any specific time already

represented in the database and the set of primitives can be used to create new values from existing data and so extend the database. A number of these surfaces have been used in the models discussed in Chapters 6 and 7.

The use of the SLPGRD and ASPGRD computer programs has been shown to be extremely useful for determining the vector quantities of the surfaces such as the slope and aspect. The pressure aspect maps clearly show the synoptic patterns and features such as the ridge of high pressure which extends across the continent in winter. The maps of the slope of the temperature surfaces show areas of greatest temperature gradient; these areas may be the regions of greatest sensitivity to climate change. This possible sensitivity, however, requires further investigation which might be achieved better at a finer scale than at the continental scale. The wet-bulb temperature and dew-point temperature aspect and slope require further work to determine the value and optimum utilisation of the information.

The thin plate smoothing spline models and the maps produced represent the result of a complex series of interactions between physical processes at many different space and time scales. Several different mapping techniques are used in this work to explore and describe the spatio-temporal data.

The grid point time series show some interesting trends. The increase found in the dry-bulb temperature (~0.7°C) over the period 1952 to 1990, if a true indication of the temperature trend, is alarming. This upward trend in the dry-bulb temperatures is however, not reflected in the pressure, wet-bulb temperature, dew-point temperature, rainfall percentiles or the SOI, all of which have cyclic patterns. There are two distinct trend lines in the correlation between the wet bulb depression and the dew-point temperature. These two distinct periods are separated by the very wet 1973-1974 period which suggest that a switch in states occurred in this period.

The Australian mean rainfall percentiles were drier (below 0.5) in the first half of the period than in the second half and this appears to be related to the very wet years in 1973 and 1974. There also appears to have been a change in the relationship between the rainfall percentiles and the SOI. This change is also apparent in the relationship between the Australian wet-bulb depression and the dew-point temperature.

An investigation of methods used to dealing with multiple spatial representations found that graphics, multiple editions and animation were all invaluable tools for analyzing large amounts of data. There is however a need for a better understanding of the theories that apply to geographic space and time.

This chapter concluded with a discussion of the issues concerning sound and its incorporation as a design variable for geographic visualization. Sound was suggested as a way to expand the limited possibilities of representing multi-variate data using graphics and to offer a way to enhance the comprehension of pattern in time. Sound in other words can provide more choices for representing and communicating ideas and can serve as a valuable addition to visual displays.

Chapter 6. The spatial Models

Introduction

There are many ways in which information can be obtained from a spatial climate database. The most common are simple data retrieval and determining new attributes by applying simple logical models as described in the previous chapter. Information can also be obtained by using models to simulate spatially and temporally distributed values of climatic variables. These models may be statistically or physically based.

Statistical or empirical models are based on an empirical understanding of the links between model input and intended output. Statistical climate modelling and forecasting are based on the principle of making predictive inferences about present or future climate variables from statistics of past climate. These models rely on statistical methods, such as regression equations with past and present observations, to obtain the most probable extrapolation. The development of the methods, such as those described in the previous chapters, for the spatial and temporal analysis of these climate data are required in order to construct stochastic climate models which respect relevant observed statistics.

Multiple linear regression has been particularly useful in climatic studies in the past and has been used primarily for point analyses. These point statistical analyses are useful if the parameters of such point models can be reliably interpolated to the particular points of interest. Usually, the mathematical form of the relationships between the climatic variables is not known, and such relationships must be established empirically. Whenever possible, the predictors should be selected on the basis of physically based structures which can be expected to retain their validity in changed climates. The models should be simply parameterised, so that they can be robustly calibrated from minimal data, and there should be enough data so that the number of predictors is smaller than the number of usable observations of each predictor (Hutchinson 1995a).

Physically-based models are based on first-principles equations which are believed to represent the physical, chemical and biological processes governing the climate system for the scales of interest. The physically based models are usually mathematically more complex than the empirical models and thus are more expensive to design, run and interpret. Due to these factors, the physically-based models were not considered for this thesis. In practice, many of the models used in the environmental sciences contain both physical and empirical elements.

Although the climate system is complex, in principle we can describe the known physical laws mathematically. In, practice, however, solving these equations in full, explicit detail is impossible (Schneider 1992). Meteorologists have found from theoretical considerations and from experience that useful detailed weather prediction of beyond about 10 days is virtually impossible using current observations (e.g. Sommerville, 1987).

Several reasons, however, make prediction of climate, in contrast to weather, feasible for comparatively long periods. For one thing, although day-to-day weather is not predictable far in advance, some success can be obtained in predicting average conditions for an extended period. For example, predicting the mean state of a gas rather than the state of a molecule of gas. Another reason is that the climate system is subject to forcing processes that may be of overriding importance for some time or space scales. For example, the annual variation in solar radiation causes seasons to follow each other predictably. While some forcing mechanisms are predictable others such as volcanic activity, are still largely unpredictable.

This chapter discusses the methods of estimation of the parameters of two types of multiple linear regression models derived from grid points taken from the MSL surface coefficient files. The first of these types are the spatial regression models which directly address the nature of spatial climatic relationships. Spline functions are used to extend these relationships to a multi-site framework. Several dew-point temperature regression models are presented. While some of these models appear to be similar, they become progressively simpler and more attractive due to their simpler parameterisations. The reliability of the surfaces and the spatial equations was tested on a set of observations withheld from the regression estimation procedures.

The incorporation of dependencies of model parameters on large scale circulation patterns is also briefly discussed and the output statistics from the spatial regression models for several climatic relationships are presented. A number of rainfall percentile model statistics are presented and possible improvements are discussed.

The second type of model presented is a forecast/hindcast multiple lag linear regression model in which spline functions are used to spatially extend lag regression parameters to obtain the future/past spatial dew-point temperature distribution. As for the spatial regression models the reliability of the surfaces and the spatial equations were tested with the withheld data set.

The spatial statistical models

Just as in the logical models presented in the previous chapter, in which time distinctions are apparent between those models whose output represents a specific time and those whose output represents a change over time, so to is there a time distinction to be made for the spatial statistical models. The spatial regression models were developed to determine a response variable in terms of other explanatory variables for times when the explanatory variables are known. The spatial lag regressions models were developed to model the value of a variable for unknown future times in terms of the first three monthly lag values of the same variable.

Several spatial dew-point temperature regression models are discussed here and although models were also produced for the other variables in the database, the dew-point temperature is described in detail.

Estimating the dew-point temperature

The dew-point temperature (T_d) has been discussed briefly in Chapter 2; it is an index of atmospheric water content. At T_d the mixing ratio of air becomes its saturation mixing ratio, and the mixing ratio lines on any thermodynamic diagram can be used to determine dew-point temperatures graphically. The usual procedure is to derive the vapour pressure e under the conditions of observation from the semi-empirical formula:

$$e = e_w - Ap(T - T_w) \tag{6.1}$$

where e_w is the saturation vapour pressure, p is the pressure of the air, T the temperature (in K) of the dry bulb, T_w temperature (in K) of the wet bulb and A is the psychrometer coefficient (see for example Sturman and Tapper 1996). In general the coefficient A depends on the design of the psychrometer, the rate of ventilation, and the temperature and humidity of the air. To obtain the dew point temperature it is necessary to find the temperature which has a saturation pressure equal to that at the temperature of interest. Because of the complexity of the thermodynamics involved, no simple relation exists between numerical values of T, T_w and T_d, and thermodynamic tables or diagrams have to be used (McIlveen 1992). However the various tables are based on different assumptions with regard to the values of the psychrometric coefficient in formulae. As the psychrometric formulae involve the atmospheric pressure (p), different tables are used for different altitudes. Different tables should also be used when the wet bulb temperature is below freezing point due to the fact that saturation vapour pressure differs for water and ice.

Figure 6.1 shows the actual average monthly T_d values and the T_d values estimated from T and T_w using psychometric tables for selected stations. The figure shows that the psychrometric tables do not perform as well for monthly average values as they do for daily observations. They tend to overestimate the values.

Linacre (1992) suggests that a first approximation of dew-point temperature can be obtained by equating it to the daily minimum temperature (T_{min}), since nocturnal cooling is often arrested at the dew-point owing to the release of latent heat when condensation occurs. However, this applies only in humid climates. There is no dew in a dry climate and T_d may be substantially less than T_{min}. Linacre (1992) also suggests that a very rough indication of the T_d in °C, except for extreme conditions, can be obtained from:

$$3T_d = 5T_w - 2T \tag{6.2}$$

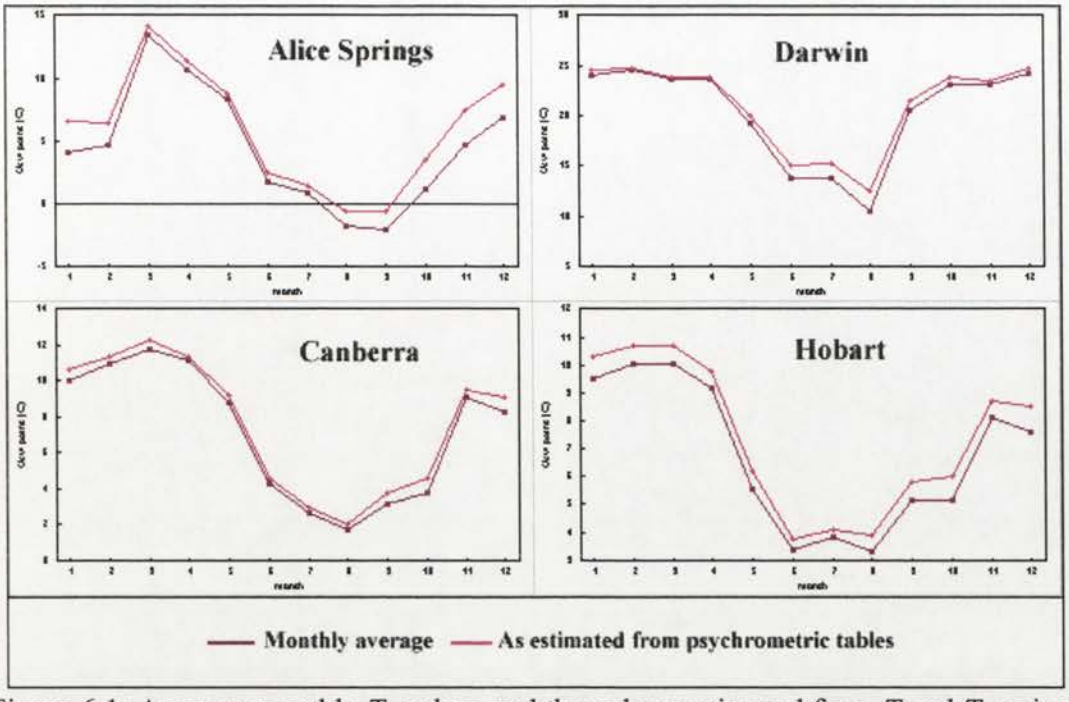


Figure 6.1: Average monthly T_d values and the values estimated from T and T_w using psychometric tables for selected stations for 1989.

Dew-point temperature surfaces produced for mid-season months using equation 6.2 have been compared with the surfaces from the spatial regression dew-point temperature models.

The spatial regression dew point temperature models

The dew point temperature regression model relates the monthly dew-point temperature to monthly dry-bulb and wet-bulb temperatures. Multivariate regression techniques have been applied to 119 points on the 2.5° x 2.5° grid (see Figure 6.2). The parameters of the regression equation are then extended spatially by fitting thin plate smoothing splines to the parameter values at the grid points. As the relationship is well understood the model was designed to examine the spatial integrity of the spline coefficient surfaces, to examine the effectiveness of models generated from the 2.5° grid, and to investigate the use of the spline technique to spatially interpolate the regression parameters.

As with any multivariate regression technique, the reliability of the regression equation cannot be determined solely from its performance over the calibration period. The only way to assess the true value of the reconstructions is to verify the models over an independent period for which climate data are available. To assess the accuracy of the various interpolation strategies, two years of data (1989 and 1990) were withheld from the regression fitting analyses. The two years of withheld data were the years with the greatest number of stations per month (see Figure 2.4), and therefore, the best to assess the validity of the method. This selection procedure imposed an exacting test on the proposed interpolation methods, since most of the withheld months had

more than 125 data points and would normally have been regarded as the most desirable months to be used to fit the surfaces.

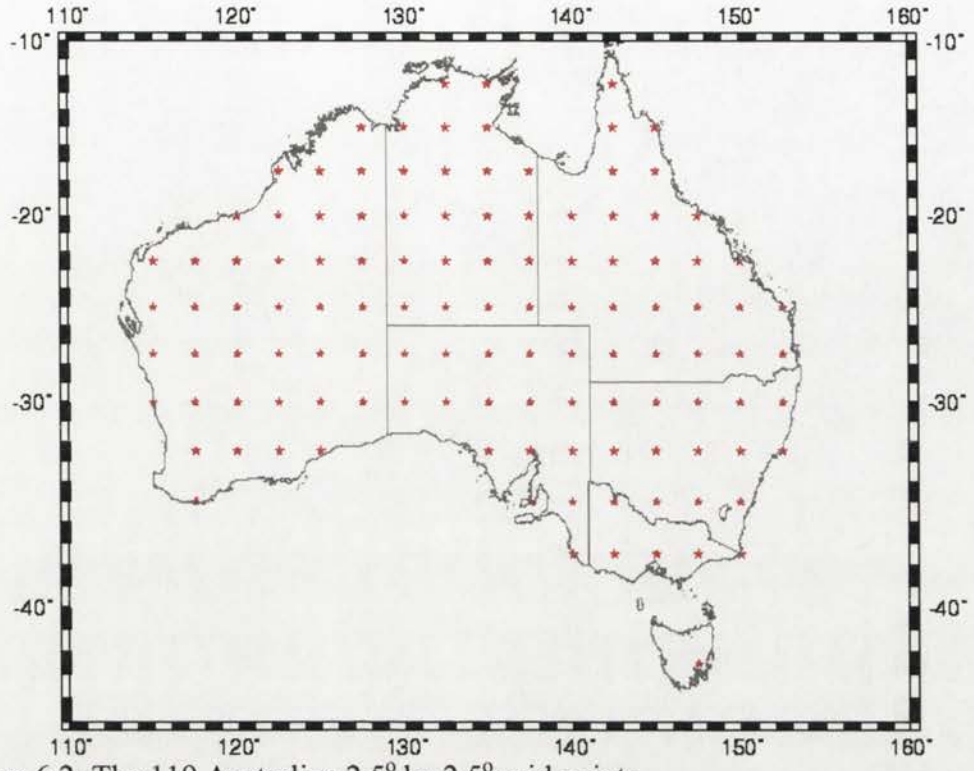


Figure 6.2: The 119 Australian 2.5° by 2.5° grid points.

Due to the seasonal variation of climatic variables, separate equations were developed for each month. A simple regression model relating dew-point temperature to dry bulb and wet bulb temperature for each month at a single point is given by

$$T_d = \alpha T + \beta T_w + \gamma + \varepsilon \qquad (6.3)$$

where T_d is the dependent or response variable (dew-point temperature), T and T_w are independent or explanatory variables (dry-bulb and wet-bulb temperature), α, β and γ are unknown parameters, and ε is the error term. This was extended to a spatial context by fitting equation 6.3 to the dry bulb, wet bulb and dew point temperatures at each one of a regular 2.5° x 2.5° grid of 119 points (see Figure 6.2) across Australia from the 'primitive' surfaces for the years 1952 to 1988. Theoretically the equation could have been applied to every point on a fine grid. The 2.5° scale was chosen to reduce the computation time and to give a number of grid points used in the regression equations similar to the number of actual stations in the present network. A preliminary investigation of the surfaces indicated that the surfaces for the parameters were relatively smooth and the 2.5° grid was considered to be sufficient.

The Genstat statistical package was used to determine values for α, β and γ at each of the 119 grid points for each month. The ANUSPLIN package was then used to spatially extend the values for α, β and γ by constructing a surface for each of these

parameters. Spatial extension of equation (6.3) then yields the dew-point temperature predictive model A as illustrated in Figure 6.3 for April 1989.

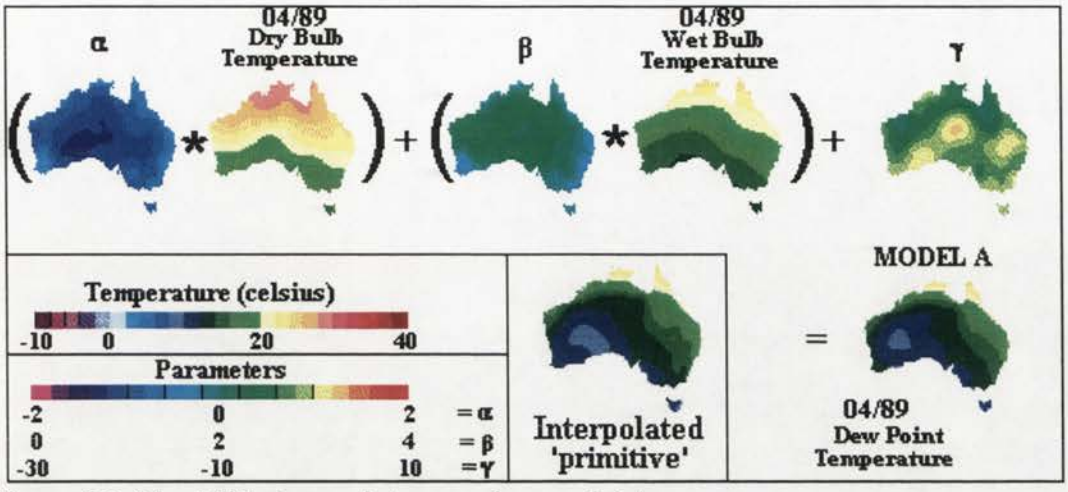


Figure 6.3: The MSL dew-point regression model A.

It is important to assess the statistical significance of the model; and therefore Genstat output files, with a description of the model including response and explanatory variates and graphical displays, were generated. The summary statistics include an analysis of variance, which includes the degrees of freedom (d.f.), the residual sum of the squares (s.s.), the mean squares (m.s.), and the variance ratio (v.r.). The probability that the variance ratio would be as large as that observed under the null hypothesis of no relationship (based on the F-distribution) was also included. The percentage variance (%var) accounted for is an adjusted R^2 statistic expressed as a percentage:

$$\%var = 100 \times (1 - (\text{Residual m.s.} / \text{Total m.s.})) \qquad (6.4)$$

Seber (1977) suggests that the adjusted R^2 statistic given in equation 6.4 is usually a better guide to the fit of the model than the unadjusted version, but neither version is an absolute measure of fit, and both depend on the range of response and explanatory values as well as on the goodness of fit.

Several checks on the adequacy of the model, and reports which concern any apparently extreme observations in the data, were also generated. The generated report gives any very large standardised residuals, that is residuals greater than that value c corresponding to probability $1/d$ of being exceeded by a standard normal deviate, where d is the number of the residual degrees of freedom. Of the 55,692 points made up of the points from the $2.5^\circ \times 2.5^\circ$ grids for each of the monthly surfaces, 426 or less than 1% had large residuals. An investigation of these residuals showed that they were fairly evenly distributed both spatially and temporally, although the months January to March 1980 contributed 29 of these large residuals. An investigation of the spline output files showed that these residuals were from months which had the highest standard deviations in the input wet bulb station data.

A check for particularly large values of leverage was also done. This check uses the criterion ck/N, where k and N are the number of parameters and number of units used in the regression model, and c is as used to check on the residuals (Genstat 1993). The sum of the leverages is always k, so this criterion highlights those observations with more than the average influence. Unlike the other checks, this one does not indicate a potential violation of assumptions, but rather that the analysis may be affected by some observations. Of the total data points (55,692) from all months 353 points had high leverage. Points 120°E, 22.5°S and 120°E, 25°S (see Figure 6.2) appear more frequently than others, suggesting either that the region is quite different from surrounding areas, or a sparse data network may have created distortions in the climate surfaces for some months.

Two Genstat checks were made on the constancy of the variance of the response variable. The fitted values were ordered into three roughly equal groups and Levene tests (Snedecor and Cochran 1989) were carried out to compare the variance of the standardised residuals in the bottom group with those in the top group, and then the middle group was compared with the other two groups combined. The Genstat program indicates whether the assumption of constant variance is tenable. A 'runs' test is carried out on the standardised residuals, ordered according to the fitted values, and a message is generated if the sign of successive residuals does not change often enough (using a 2.5% significance level), indicating whether some systematic pattern remains in the residuals (Genstat 1993). None of the monthly regression models failed either of these checks.

Table 6.1: Percentage variance for the dew point regression models for mid season months.

	MODEL A (d.f. =356) % variance	MODEL B (d.f. =120) % variance
JAN:	99.2	99.0
APR:	97.4	97.3
JUL:	95.7	95.4
OCT:	98.3	98.0

The model gives a good approximation of the sea-level values for dew point temperature, and the percentage variance explained by the model is high (see Table 6.1). However, the diagnostics from the statistical procedures suggest that while the spatial variation in the dry bulb and wet bulb temperature parameters are significant the variance ratios are not large. This indicates that the dew point model could be simplified by reducing the degrees of freedom in the model, that is the spatially varying parameters for dry bulb and wet bulb temperature could be reduced to a constant. The small reduction in the percentage variance explained (shown in Table 6.1) shows that the effect of reducing the degrees of freedom is small. This is also evident in Figures 6.5 which shows that there is little spatial variation in the α and β surfaces.

The Genstat procedure was then rerun to estimate the values for α, β and a spatially varying γ. The surfaces for the spatially varying γ were then produced to give the dew-point regression model B illustrated in Figure 6.4.

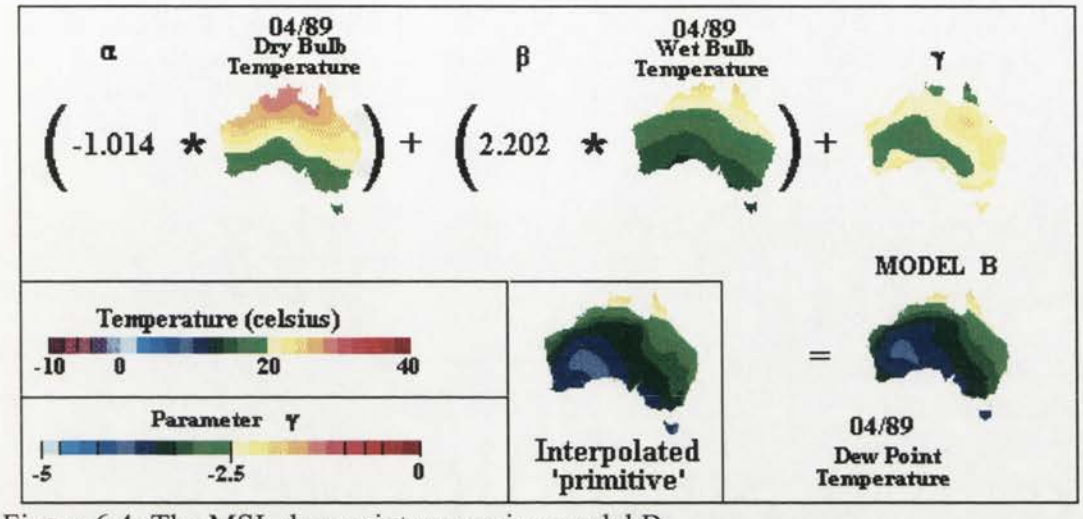


Figure 6.4: The MSL dew point regression model B.

It should be noted that the withheld data were not used to optimise smoothing parameters, but to validate the output surfaces. The model parameters were estimated in each case by minimising the GCV. In normal applications all years of data would be included in the regression model procedure and in the surface-fitting process. The crucial test of model performance is the level of agreement between the actual and reconstructed values over the verification period.

The regression models were compared with the surfaces interpolated from the observed dew-point temperature data ('primitives') and surfaces created by applying equation 6.2 (from Linacre 1992) to the 0.1° grids of dry-bulb and wet bulb temperature (see Figure 6.5). Although there was little difference in the explained variance between models A and B, and a graphical comparison with the surface interpolated from the observed shows little difference. Over all months model A performed slightly better than model B but at some expense in terms of its complexity. Both regression models A and B performed significantly better than the Linacre model in matching observed values, especially in July and October (Figure 6.5). Both models, however, show that the surfaces produced by the spline function display spatial coherence and that the regression model offers a simple and accurate alternative to existing dew-point temperature calculation methods for the monthly time scale.

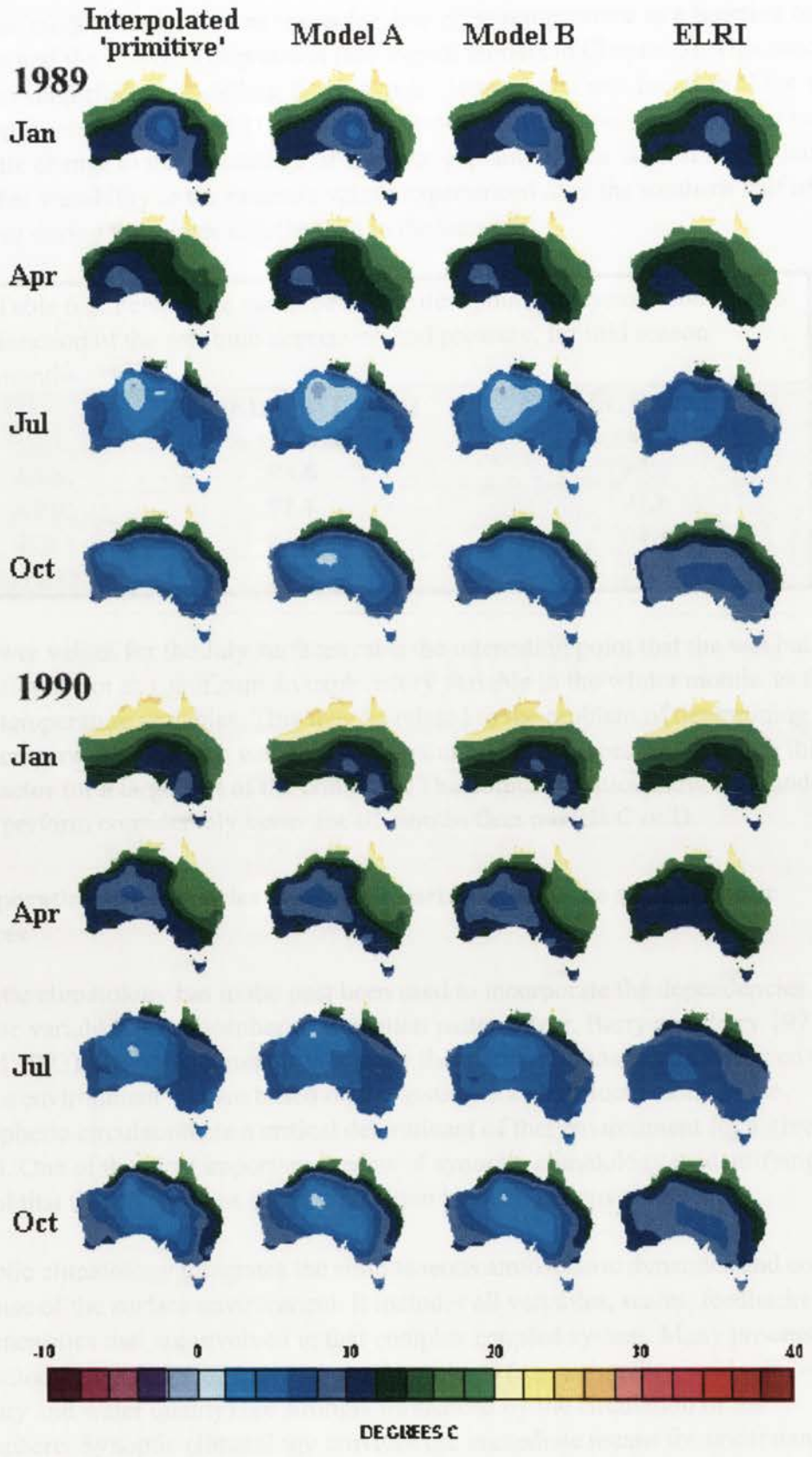


Figure 6.5: The output from the dew-point temperature interpolation, the dew-point regression models A and B, and the Linacre model (ELRI) for mid-season months in 1989 and 1990.

A similar model (Model C) was tested for dew-point temperature as a function of pressure and the wet-bulb depression (see logical models in Chapter 5). This model was then simplified to modelling the dew-point temperature as a function of the wet-bulb depression only (Model D). Table 6.2 shows that the removal of pressure makes very little change to the percentage of variance explained. This is most likely due to the higher variability in the pressure values experienced over the southern half of the continent during the winter months than in the summer.

Table 6.2: Percentage variance for the dew point regression models as a function of the wet-bulb depression and pressure, for mid season months.

	MODEL C (d.f. =356) % variance	MODEL D (d.f. =237) % variance
JAN:	93.6	93.0
APR:	92.1	91.0
JUL:	86.2	84.1
OCT:	93.3	92.3

The lower values for the July surfaces raise the interesting point that the wet bulb depression is not as significant an explanatory variable in the winter months as the actual temperature variables. This may be related to the problem of determining dew point temperature when the wet bulb temperature is below freezing, although this is not a factor for a large part of the continent. The output statistics show that models A and B perform considerably better for all months than models C or D.

Incorporating dependencies of climatic variables on large scale synoptic features

Synoptic climatology has in the past been used to incorporate the dependencies of climatic variables on atmospheric circulation patterns (e.g. Barry and Perry 1973, Yarnal 1993). Synoptic climatologies study the effect of climatic variability on the surface environment and are based on the assumption that fluctuations in the atmospheric circulation are a critical determinant of that environment for a given period. One of the most important aspects of synoptic climatology is identifying the control that these variations in the circulation have on the environment.

Synoptic climatology integrates the simultaneous atmospheric dynamics and coupled response of the surface environment. It includes all variables, scales, feedbacks, and non-linearities that are involved in that complex coupled system. Many processes associated with a range of environmental problems (e.g. air quality, acid rain, water quantity and water quality) are strongly influenced by the circulation of the atmosphere. Synoptic climatology provides the immediate means for understanding how environmental processes and the atmospheric circulation are linked.

Barry and Perry (1973) concluded that every synoptic climatology has two stages: the first is the classification of the atmospheric circulation; and the second is the assessment of the relationships between those categories and the weather elements of a region. The atmospheric circulation is classified in order to identify the essentials of the different states of the atmospheric circulation and to communicate this essence by means of an understandable, descriptive class name. There are, however, problems associated with the determination of categories as no two daily or monthly pressure maps are identical. Apart from the almost infinite variety of synoptic patterns, synoptic systems differ markedly in size and intensity and there are seasonal variations in synoptic type characteristics. The classification of the circulation assumes that the atmosphere can be partitioned into discrete non-overlapping intervals and that the classification identifies all important map patterns or synoptic types (Yarnal 1993).

The assessment of the relationships between the categories in a classification and the weather elements in a region is usually undertaken deductively, relying on statistics to link the circulation and the surface environment. Several studies have been done for areas of Australia using this approach (e.g. Karelsky 1956, Evans and Allan 1992, Leighton 1994).

The regression models which are presented below aim to incorporate both steps of synoptic climatological analysis in a single step. That is, rather than extracting particular patterns to represent the mean atmospheric circulation, the relationship was established for all points between the variable of interest (e.g. temperature, rainfall percentiles) and the circulation pattern as represented by pressure and derived components. The components derived from the pressure maps are unit vectors (representing wind direction) and actual vectors (representing wind speed and direction) separated into their meridional (north-south) and zonal (east-west) components.

Models were derived for dry-bulb temperature (model E), wet-bulb temperature (Model F) and dew-point temperature (model G). Models were also derived, using the methods described above, to relate the temperature to pressure and wind speed and direction (models E1, F1 and G1), to the pressure and wind direction (models E2, F2 and G2), and to the pressure and wind speed (models E3, F3 and G3). All models were created both with spatially varying regression parameters (Jan(a) and Jul (a) with 475 d.f.) and without spatially varying regression parameters (Jan(b) and Jul (b) with 121 d.f.). Table 6.3 gives the percentage variance for the temperature regression models as a function of various combinations of components of the pressure, for January and July. As with the dew-point temperature models the relationships were derived overthe period 1952-1988

The dependencies of the temperature values on the larger-scale atmospheric conditions show some interesting seasonal differences. The output statistics show very good results for the wet-bulb temperature for both January and July, and for the

dry-bulb temperature in July. Table 6.3 suggests that the relationship between the dew-point and the wet and dry bulb temperature is not reflected in the relationship between the dew-point temperature and the pressure components. Overall the model performs for wet- and dry-bulb temperatures better in July than in January while the reverse is true for the dew-point temperature.

Table 6.3: Percentage variance for the temperature regression models (as a function of components of the pressure) for January and July.

	% variance			
DRY BULB (Model E)	**Jan (a)**	**Jan (b)**	**Jul (a)**	**Jul (b)**
Model E1 (pr + wd +ws)	**87.5**	**85.3**	**96.0**	**94.9**
Model E2 (pr + wd)	**87.4**	**85.2**	**95.7**	**95.0**
Model E3 (pr +ws)	**85.3**	**85.1**	**95.0**	**94.4**
WET BULB (Model F)	**Jan (a)**	**Jan (b)**	**Jul (a)**	**Jul (b)**
Model F1 (pr + wd +ws)	**94.5**	**93.5**	**92.2**	**88.6**
Model F2 (pr + wd)	**94.4**	**93.6**	**91.1**	**88.5**
Model F3 (pr +ws)	**92.4**	**91.6**	**90.2**	**87.8**
DEW POINT (Model G)	**Jan (a)**	**Jan (b)**	**Jul (a)**	**Jul (b)**
Model G1 (pr + wd +ws)	**85.6**	**83.7**	**74.6**	**66.6**
Model G2 (pr + wd)	**85.7**	**84.0**	**71.8**	**66.5**
Model G3 (pr +ws)	**82.4**	**80.0**	**71.7**	**66.2**

Where pr = pressure, wd = wind direction, ws = wind speed
(a) = model with spatially varying regression parameters, (b) = model with spatially varying constant

The difference between the spatially varying (a) and simpler (b) models is relatively small in summer and there appears to be no advantage in the more complex models. In winter, however, and especially for the dew-point temperature, the spatial variation in the regression parameters is of much greater importance. This suggests that the moisture regimes are more spatially complex in winter than in the summer. The wet- and dry-bulb models show merit as they explain a high percentage of the variance and while the dew-point temperature models do not perform as well there is some physical explanation of the results.

Rainfall models

The interpolation and space-time simulation of monthly rainfall depends on identifying the components of the spatial variation of monthly rainfall. Given the difficulties in determining the spatial distribution of monthly rainfall amounts, it is important to consider the efficacy of space-time analyses of rainfall at a monthly scale. Hydro-ecological responses are not always required or observed on a daily time scale, especially when assessing the responses over larger areas. Hamlin (1983) has observed that integrated hydrological responses of larger catchments are relatively insensitive to fine-scale variations in hydrological inputs, and the monthly time scale is known to be sufficient to resolve the occurrence and severity of drought (Bonacci

1993, Smith *et al.* 1993). As was discussed in Chapter 2, the monthly time scale should be sufficiently broad for both real-time interpolation and simulation of rainfall to be feasible, and has been examined by Walsh *et al.* (1982), Klein and Bloom (1987) and Lyons (1990).

The temporal and spatial correlations in the distribution of rainfall differ markedly due to the differences in occurrence and intensity produced by different rain producing mechanisms. The specification of a model for estimating rainfall percentiles may therefore be problematic, as the relationship between rainfall and the various climatic parameters may reflect different processes. Further problems may arise in areas in which there are a high proportion of months which are totally dry. The significance of this problem should be limited for monthly data, since for most parts of Australia the proportion of months which are totally dry is quite small.

If the surface rainfall environment is dependent on the dominant synoptic patterns for a particular month, then the monthly percentiles should reflect these patterns. However, this raises a problem in that the pressure and temperature surfaces represent averages for the month of interest, while the rainfall surfaces are presented as relative historical values (percentiles) resulting from cumulative processes. Thus if a large amount of rain in a location is the result of an extraordinary single event (i.e. the passage of a vigorous storm) the rainfall amount (and therefore the percentile value) will reflect this event, while the average 0000 UTC pressure and temperature values will smooth out the event and it will not be reflected in the monthly average values. Nevertheless, several relationships between the rainfall percentiles and the pressure and temperature components have been investigated.

The rainfall percentile models

Several models, using the methods described for models A and B (see regression models above) have been applied to characterise the spatial and temporal variations in rainfall percentiles as a function of combinations of temperature and pressure values. Table 5.4 shows the percentage variance for the rainfall percentiles as a function of T_d and T_w (Model H), T and T_d (Model I), P and T_d (Model J), P and T (Model K) and P and T_w (Model L). Again both the more complex models (a) with spatially varying parameters, and the simpler models (b) with only the spatially varying constant, were constructed.

While the amount of variance explained by the different models is not high, Table 6.4 does highlight some interesting points. The most notable point is the importance of the dew-point temperature, a measure of the moisture content. Models H, I and J (which include T_d) perform considerably better than Models K and L (which exclude T_d). The second point is the seasonal difference in the success of the models including the dew-point temperature. The two temperature models (H and I) perform better in spring and summer while the dew-point temperature and pressure (Model J) performs better in the autumn and winter.

Table 6.4: Percentage variance explained for rainfall percentile regression models for mid-season months.

Percentile Model	% variance							
	Jan (a)	Jan (b)	Apr (a)	Apr (b)	Jul (a)	Jul (b)	Oct (a)	Oct (b)
H (T_d and T_w)	40.2	36.2	27.1	23.5	31.8	26.7	34.0	22.9
I (T and T_d)	39.3	34.0	27.8	22.9	32.2	25.0	33.4	24.0
J (P and T_d)	32.8	32.1	31.0	28.6	36.6	31.1	29.3	21.8
K (P and T)	19.0	13.2	20.1	16.7	17.7	14.7	21.5	16.3
L (P and T_w)	19.6	18.4	24.0	22.5	28.8	25.5	8.0	5.7

Where T = dry-bulb temperature, T_w= wet-bulb temperature, T_d = dew-point temperature, P = pressure,
(a) = model with spatially varying regression parameters, (b) = model with spatially varying constant

The seasonal variation in the results of the different rainfall percentiles models is again due to the different types of moisture bearing systems which dominate the seasonal synoptic regimes. During the summer the convectional rainfall predominates and the temperature values are a better predictor of rainfall than the pressure values. In winter frontal and orographic rainfall predominate and the pressure value becomes a better indicator of potential rainfall. All rain producing mechanisms, however, can occur at any time of the year. Due to these different rain-producing mechanisms and the variability of their occurrence, the spatially varying models (a) perform better and in some cases significantly better than the simpler, non-varying models (b). The relative failure of some models in some seasons, especially of Model L for October compared to the rest of the mid-season months, raises further questions about the components of the spatial variation of monthly rainfall.

Incorporating dependencies of rainfall on synoptic patterns.

Correlations between monthly anomalies of precipitation and atmospheric circulation fields using empirical orthogonal function analyses have been examined by a number of authors (Stidd 1954, Walsh *et al.* 1982, Lyons 1990, Klein and Bloom 1987). The correlations are often statistically significant and have process-based explanations, but are generally not found to be large (Hutchinson 1995a). Correlations between daily surface climate anomalies and broad-scale surface weather anomalies have been investigated by Wilks (1989). Wilks found such correlations sufficient to identify actual spatial patterns in daily temperature anomalies, but insufficient to identify spatial patterns of daily precipitation anomalies, for which a stochastic approach is necessary. Methods for conditioning the parameters of daily precipitation models on classified large-scale circulation patterns have been developed by Hay *et al.* (1991) and Bardossy and Plate (1992). These models have been found to match various observed mean and extreme rainfall statistics.

Table 6.5 shows the percentage variance explained by different rainfall percentile regression models for January and July. As with models E, F and G (for temperature)

the derived components of the MSL pressure surfaces that were used were unit vectors (representing wind direction) and actual vectors (representing wind speed and direction) separated into their north-south and the east-west components. Again, models were created for both the complex (a) and for the simpler (b) versions. As with the dew-point temperature models the relationships were derived over a calibration period (1952-1988) and the results were tested with independent data from another period (1989-1990). The results in Table 6.5 show that Models M1, M2 and M3, using pressure and components of pressure, do not perform as well as those which are based on the dew-point temperature (Models H, I and J). However, it is interesting to note that models M1, M2 and M3 perform considerably better in the winter months than in the summer; this suggests that the differences in the rain-producing mechanisms play a major role in determining the rainfall percentile patterns. This is further substantiated by the improved performance of the more complex (a) models over the simpler (b) models.

Table 6.5: Percentage variance for rainfall percentile regression models (as a function of components of the pressure) for January and July.

Rainfall Percentile models	% variance			
	Jan (a)	Jan (b)	Jul (a)	Jul (b)
Model M1 (pr+wd+ws)	10.8	8.0	25.2	16.9
Model M2 (pr+wd)	10.9	8.4	21.9	15.3
Model M3 (pr+ws)	4.7	4.0	19.5	14.3

Where pr = pressure, , wd = wind direction, ws = wind speed (a) = model with spatially varying regression parameters, (b) = model with spatially varying constant

All the rainfall percentile models were found to provide only reasonable fits to monthly rainfall percentile distributions. This is mainly due to the difficulties in identifying the components of the spatial variation of monthly rainfall percentiles (i.e. the wrong choices or the model is too simple), and the difficulties in comparing the mean monthly pressure components with the percentile values.

It may also be due to a deficiency in the representation of rainfall anomalies by the rainfall percentiles. There are areas of Australia which have highly seasonal rainfall, and have a large proportion of 'dry' months with no recorded rainfall. As the percentile of zero rainfall was set to the mid-point of the relative frequency that the month is dry these areas have relatively high percentile values for these months. For example, Darwin had no July rainfall for 25 of the 39 months recorded and this results in a percentile value of 32 (see Figure 2.8). There may also be greater climatologial differences between values for some stations than is reflected by the differences in their percentile values. This can result in very dissimilar synoptic conditions being represented by similar percentile values. For example, Perth had two Januarys with more than 80 mm of rainfall with percentiles of 99 and 96 while the next highest rainfall of 30 mm has a percentile value of value of 93 (see Figure 2.8). Thus a number of different climatic conditions can produce similar percentile

values. The significance of these areas in the spatial interpolation of the regression parameters requires further investigation.

Further investigation is required to resolve the observed disparity between the intensity and duration of different rain-producing mechanisms. This then, raises the question of whether there are other possible methods of representing the rainfall which adequately describe the correlation structure in space and time of the rainfall anomalies, since strong correlations in both space and time are apparent. A possibility would be to use total monthly rainfall or rainfall anomalies rather than percentiles, although the initial reasons for choosing the percentiles over the rainfall also apply to the anomalies and that most of the problems that arose in the percentile models would also be apparent in the anomaly models.

An alternative possibility is a truncated power of a normal distribution, which has been posed as a suitable model for rainfall at a point over time steps ranging from hourly to monthly (Hutchinson 1995b, Hutchinson *et al.* 1993). Particular studies have examined the cube root and square root transformations of observed rainfall totals (Stidd 1973, Richardson 1977). Using graphical fitting methods, Stidd (1973) maintained that the cube root of observed rainfall totals, at time steps ranging from seasonal to hourly, were well matched by a truncated normal distribution. By keeping the power at the cube root, he found that the variance of the underlying normal distribution could be assumed to be constant for different time steps, thereby allowing the probability distribution of rainfall at one time step to be inferred from the distribution at any other time step. He suggested that such large scale order could be related to the -5/3 power law describing the spectra of atmospheric turbulence, which Panofsky and Dutton (1984) suggest holds from hourly through to seasonal time scales. This parameterisation has also been used by Bardossy and Plate (1992) to construct a space-time daily rainfall model with parameters which are conditioned on large scale atmospheric circulation patterns.

The truncated power of normal model has been examined in detail by Hutchinson *et al.* (1993). Their analysis of monthly and daily precipitation data does not support Stidd's claim of the universal applicability of the cube root transformation. However, by allowing the power of the truncated normal distribution to vary from station to station, the model was found to provide excellent fits to both monthly and daily rainfall data, particularly when the observation period was restricted to a standard thirty-year period. This method of representing the rainfall may provide the basis for a model which adequately describe the correlation structure in space and time of rainfall amounts as discussed by Hutchinson (1995b).

Spatial statistical forecasting

Climate prediction in Australia

The Bureau of Meteorology has been developing methods for seasonal climate prediction since early this century (Nicholls 1997). From 1954 the Bureau prepared monthly weather forecasts for internal assessment. These forecasts were based on an extension of the methods used in four-day forecasting; a zonal index cycle, blocking patterns, persistence and the movement of large scale anomalies. The resulting rainfall forecasts were no more accurate than would be expected from chance. The temperature forecasts were more accurate but not sufficiently accurate to be useful and the experiment was discontinued in 1971 (Nicholls 1997).

During the 1982/83 the drought on the Australian east coast, attention turned to the use of the El Niño/Southern Oscillation in climate forecasting. The 'Australian RAINMAN' computer package (see http://www.dpi.qld.gov.au/rainman/) was developed to allow users to investigate the likely consequences of particular phases or trends of the SOI on rainfall at a large number of locations.

The Bureau of Meteorology's Seasonal Climate Outlooks have been based on simple linear lagged regression, and variants on this, between the SOI and subsequent rainfall. In 1995 the Seasonal Climate Outlooks were modified to include trends and phases of the SOI and some patterns of sea surface temperature have been included since 1996. The Bureau of Meteorology also now use sophisticated time series methods to project the SOI into the future. These projections are used to find analogues which can be used, for example, in crop yield models (Nicholls 1997). More recently the Bureau of Meteorology, CSIRO Atmospheric Research and the Queensland Department of Natural Resources are investigating the use of dynamic methods, using coupled ocean-atmosphere models, to provide predictions for the El Niño/Southern Oscillation.

At present the Bureau of Meteorology only forecasts seasonal rainfall. Nicholls (1997) suggests there is a need for temperature forecasts and for more specific indices (e.g. severity of the frost season). The spatial lag linear dew-point temperature regression model is presented below as a simple temperature forecasting method.

Spatial lag linear regression model

Even within the class of statistical forecasting techniques there is a wide range of available methods. The discussion here is restricted to the regression models, but comprehensive coverage of the other models and their advantages and disadvantages are given in for example Bennett (1981) and Chatfield (1989).

The simplest spatial regression forecasting model is a straightforward extension of the linear regression equation:

$$Y_t = \alpha + \beta x_t + \varepsilon_t \qquad (6.5)$$

where x_t is an independent variable, Y_t is the variable to be forecast and ε_t is the error. The α and β are coefficients and the data are sampled over time instants t. This form of the linear regression can be recast as a forecasting equation by merely shifting the base of the time index t and can be extended to a whole set of lags of the leading indicator:

$$Y_{t+k} = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots\ldots + \beta_q x_{t-q} + \varepsilon_t \qquad (6.6)$$

This is a multiple regression equation where the independent variables are in fact lags of the same independent variable. Least squares equations can still be applied to estimate the coefficients (Chatfield 1989, Weisberg 1985).

The methodology discussed above has been used to produce a lag regression model for monthly dew point temperature. The parameter-efficient model for the spatial distribution of dew-point temperature is examined with the particular aim of assessing its utility for space-time simulation. It involves fitting a spatial lag distribution by linear regression. The regression equations were estimated for the required month as a function of the lag correlations with the dew-point temperature values for the preceding three months, and the spatially-varying constant calculated using the Genstat statistical package. By allowing the coefficients to vary gradually with geographic location, observed and predicted distributions at the monthly time step can be well matched using a surface for just one statistical parameter. As has been shown this parameter can be easily spatially interpolated from existing temperature station network. Values for this correlation are in keeping with observed lag correlations of between about 0.3 and 0.7 for a lag of one month, such as those found for temperature by Richardson (1981) and those obtained by Hutchinson *et al.* (1993) for normalised rainfall.

Again, as for the spatial regression the April 1989 dew point temperature is used as an example model and the spatial equation is shown in Figure 6.6.

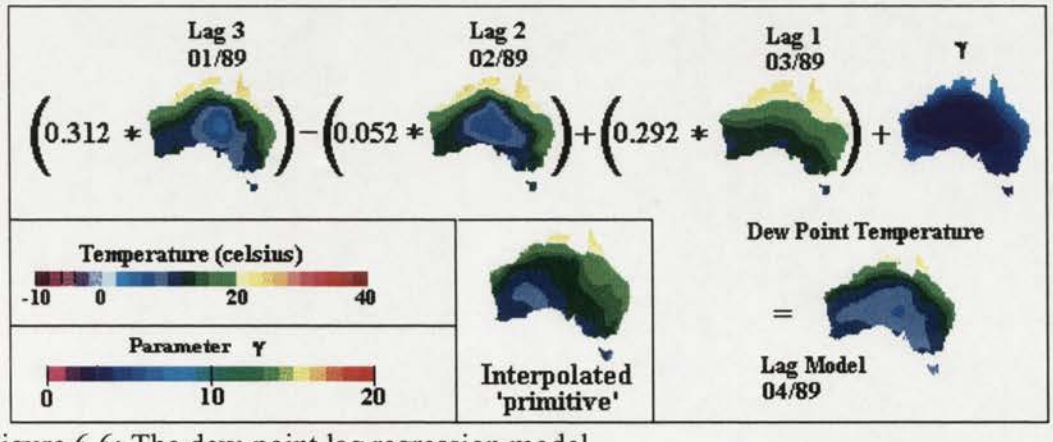


Figure 6.6: The dew-point lag regression model.

Table 6.6 shows that the regression model explains a high percentage of the variance and Figure 6.6 shows that the model performs reasonably well for April 1989. Similar or better results were obtained for other months from the validation set. Both these models suggest that the spline function is well suited to the spatial extension of not only climate variables but also time-series model parameters. The output statistics from the Genstat program (see Table 6.6) show that Model A performed slightly better than Model B but at some expense in terms of complexity.

Table 6.6: Percentage variance for the dew point lag regression models as a function of lags 1, 2 and 3.

MODEL N	Complex (A) (d.f = 475) % variance	Simple (B) (d.f = 121) % variance
JAN:	80.8	80.6
FEB	83.7	83.6
MAR	81.0	80.9
APR:	80.4	79.9
MAY	72.4	72.2
JUN	65.3	65.1
JUL:	73.0	72.9
AUG	79.1	78.7
SEP	80.3	80.1
OCT:	82.4	81.9
NOV	87.2	87.0
DEC	86.0	85.9

Due to the small difference in the explanation of variance in the A and B Models, the B Models were used to create output grids for all months in 1989 and 1990. Again the seasonal nature of the performance of the models is noticeable. While the statistics from the Genstat output were encouraging, there were withheld data to check the output surfaces for 1989 and 1990 against the actual surfaces. Figure 6.7 shows the validation rms for the B lag regression models.

For each month of 1989 and 1990 an anomaly average was calculated by averaging the difference between the predicted month and the long term average dew-point temperature on each point of the grid. This is the equivalent of using the long term average dew-point temperature to predict the monthly values. Figure 6.8 shows the validation rms as shown in Figure 6.7 and the anomaly average. As can be seen the value of the validation rms is dependent on the anomaly average. It is interesting to note however that the lag model performs better than would be expected for some months (eg. May 1989, Feb 1989 and 1990) and not so well in others (e.g. March 1989, August 1990). This suggests that there is some temporal and spatial correlation in the distribution of the anomalies.

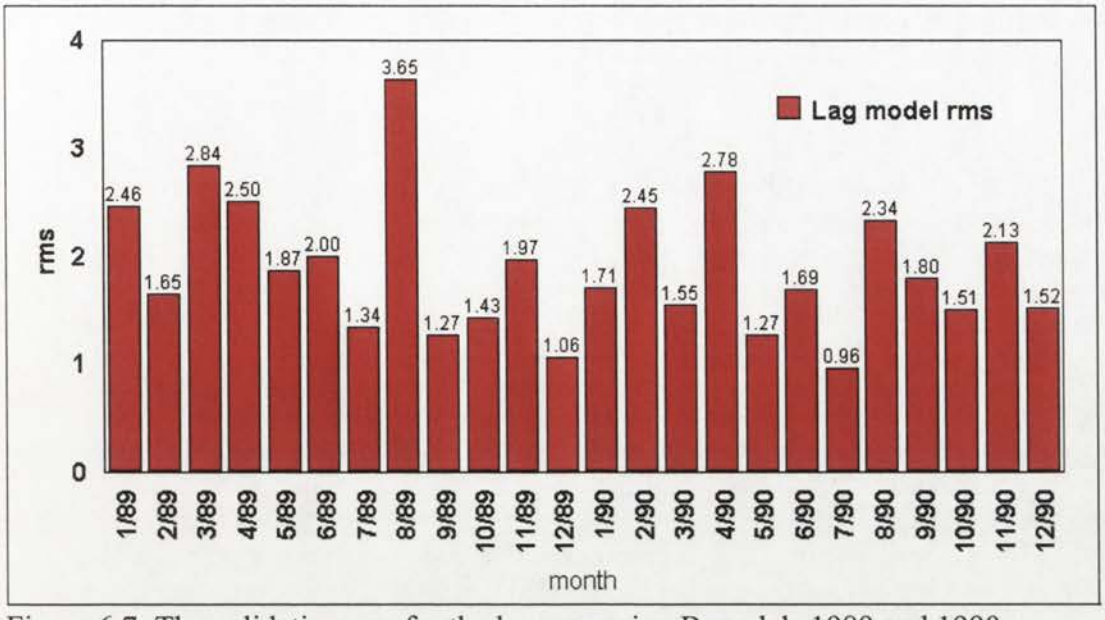


Figure 6.7: The validation rms for the lag regression B models 1989 and 1990.

As can be seen the validation rms range from 0.96 °C for July 1990 to 3.65 °C for August 1989. No seasonal trend is apparent in the performance of the models as measured by the validation rms for the two years of validation data. This suggests that the performance of the lag regression model as measured by the validation rms is dependent on the variance of the month to be predicted. That is the performance of the model is dependent on how close to the long term mean the predicted month values are.

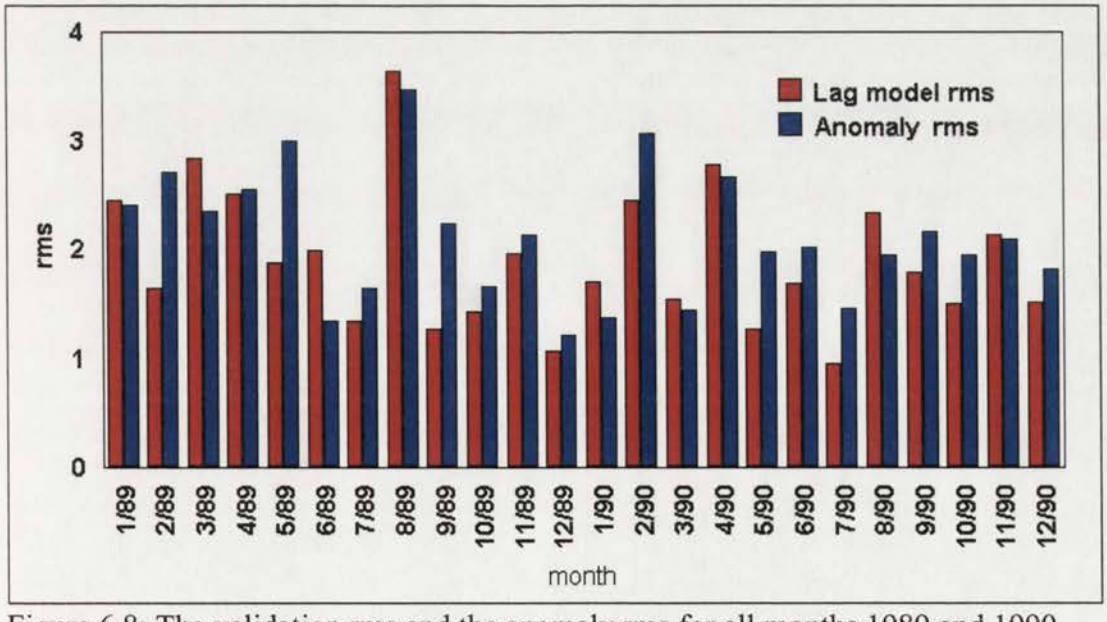


Figure 6.8: The validation rms and the anomaly rms for all months 1989 and 1990.

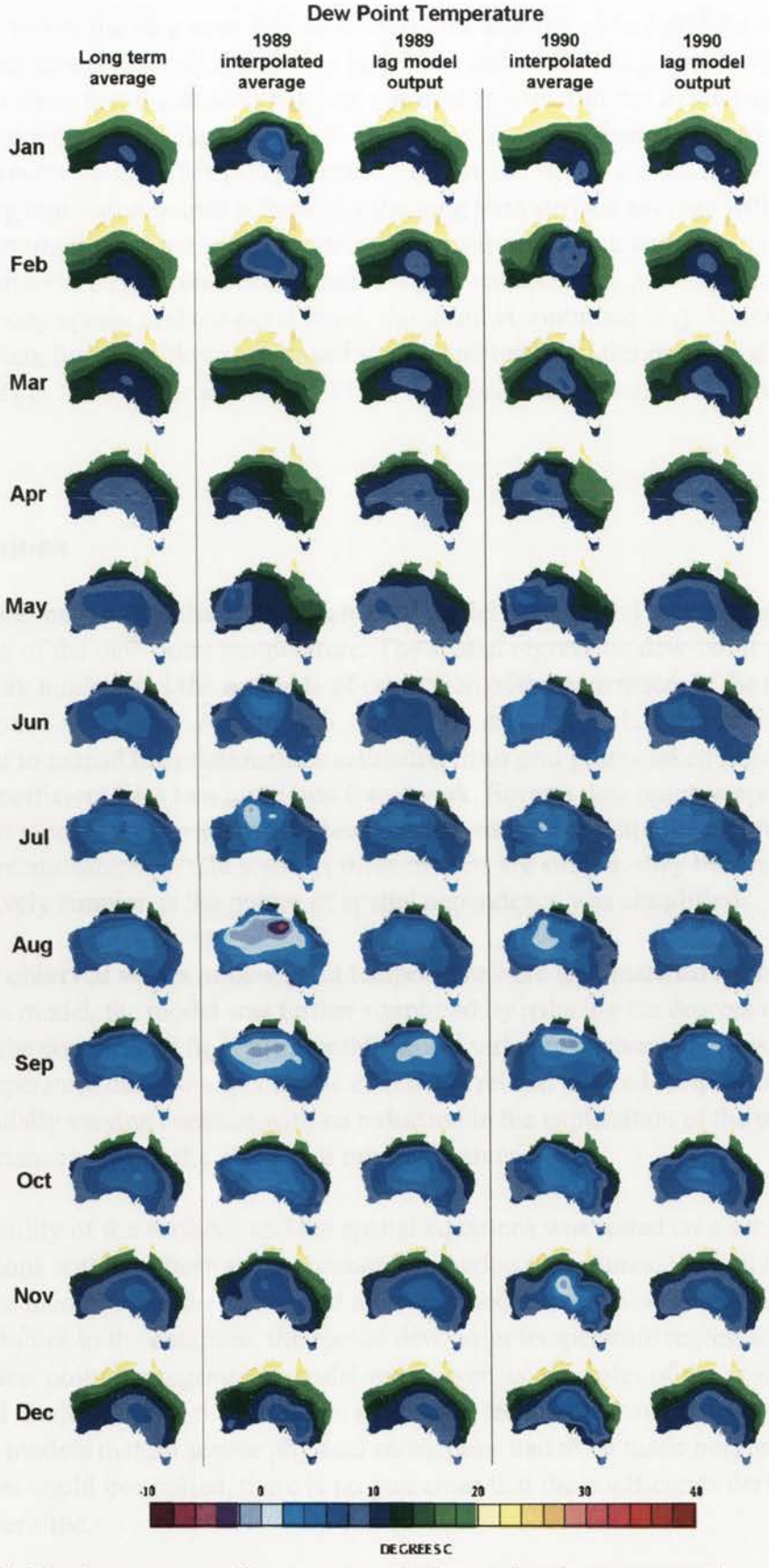


Figure 6.9: The long term surface average, the interpolated average surface and the output from the lag model B for all months for both 1989 and 1990.

Figure 6.9 shows the long term surface average, the average surface and the output from the lag model B for all months for both 1989 and 1990. The graphics shown in Figure 6.9 show that the models perform reasonably well, but not as well as the Genstat output statistics suggest. The high % of variance explained is in part due to spatial autocorrelation. This is supported by Figure 6.9 which shows, not surprisingly that the lag regression output is basically the long term surface average with some spatial perturbations. That is, the models tend towards the long term mean if the previous months have been close to their average values. If the previous months show a strong spatial and temporal trend, this trend is continued (e.g. March to July 1990). Often, however, there are mixed spatial and temporal trends and the model again tends to the long term average. These models are discussed further in Chapter 7.

Conclusions

This chapter began with the spatial statistical models and a brief discussion of the estimation of the dew-point temperature. The spatial regression dew-point temperature models and the methods of estimation of the parameters of the multiple linear regression equations are then discussed. Bivariate thin plate smoothing splines were used to extend the relationships estimated from grid points taken from the MSL surface coefficient files to a multi-site framework. Several dew point temperature regression models were presented. These models address directly the nature of spatial climatic relationships. While some of these models are similar, they become progressively simpler as the nature of spatial dependence was simplified.

Since the observed values of dew point temperature were well matched by the regression model, the model was further simplified by reducing the degrees of freedom. This left the goodness of fit of the monthly model virtually unchanged. Thus, the dew point temperature distribution could be effectively related to fixed temperature values and a spatially varying constant with no reduction in the explanation of the percentage of the variance between the actual and modeled values.

The reliability of the surfaces and the spatial equations was tested on a set of observations withheld from the regression estimation procedures. Several spatial regression models were discussed and although models were also produced for the other variables in the database, the spatial dew-point temperature regression model and the dew point lag regression model were given as examples of the spatial statistical models. While the statistical regression techniques have the advantage over dynamic models that no severe physical restrictions had to be made before the techniques could be applied, there is no guarantee that the coefficients derived will be stable over time.

The incorporation of dependencies of model parameters on large scale circulation patterns was also briefly discussed and the output statistics from the spatial

regression models for several climatic relationships were presented. The dependencies of the temperature values on the larger scale atmospheric conditions show some interesting seasonal differences. On the whole, however, it appears that there is considerable merit in the models and that there is some validity in the assumption that fluctuations in the atmospheric circulation are a critical determinant of the surface environment for a given period.

Several rainfall percentile model statistics were presented and possible improvements are discussed. The preliminary model results were given for spatial rainfall percentile models and models which incorporate dependencies on large-scale synoptic features. All the rainfall percentile models were found to provide only reasonable fits to monthly rainfall percentile distributions. This is mainly due to the difficulties in identifying the components of the dependency of the spatial variation of monthly rainfall percentiles.

The relative failure of Model L for October (using pressure and the wet-bulb temperature) compared to the rest of the mid-season months also raises further questions about the components of the dependency of the spatial variation of monthly rainfall. Not surprisingly, however, due to the different rain-producing mechanisms effecting the different parts of the continent, the spatially varying models (a) perform better and in some cases significantly better than the simpler (b) models. Other possible methods of representing the rainfall which may better describe the correlation structure in space and time of the rainfall are also suggested.

The different types of statistical forecasting models and the spatial statistical forecast models are presented. These models spatial extend lag regression parameters to obtain the future/past spatial dew point temperature distribution. On the whole the lag regression models perform reasonably well, but not as well as suggested by the percentage variance explained in the Genstat procedure. Not surprisingly, if the previous three months are "average" or give mixed signals the models tend towards the long term average. If there are strong trends in the same direction during the previous three months the model does perform well. Overall, the output results are encouraging.

As climate accounts for a significant percentage of the variability in most ecological data, it is therefore useful to produce a climate space-time model that generates monthly climate data. Importantly, the models can be calibrated in terms of observed monthly summary statistics whose spatial distribution can be well determined from the standard meteorological network using existing thin plate smoothing spline techniques. Once interpolated and tested these models are then in a convenient form for future use. However, since the statistical approach depends on historical data, it is obviously limited to predicting climates that have been observed or are not caused by new processes.

Chapter 7. The Spatio-temporal Spline Models

Introduction

Development of the previous models are guided by the requirements of this final step. That is to account for the persistence in space and time of climatic anomalies by the development of a spline space-time models which integrate the space and time domains. Development of complete space time models depends on taking proper account of the temporal and spatial correlations implied by this persistence.

The central theme underlying the spline cube model presented in this chapter is that the relative locations, in both time and space, of the points or areas to which the data refer can provide some information about the spatial and temporal pattern of variation in these data. Without spatial autocorrelation the surface of the earth would appear entirely random. As was seen in Chapter 4 the data do exhibit spatial correlation. Without temporal autocorrelation we would have highly variable weather. Spatial and temporal correlations occur simultaneously but do they act in, or can they be modelled in the same space?

The simplest models of covariance structure are the isotropic models used in kriging and optimal interpolation which do not incorporate any time dependence between successive realisations (see Chapter 3). The simultaneous identification and estimation of the spatial and temporal autocorrelation structures in space-time models is still in its infancy, even without the complications of long memory dependence.

Methods of representing time in spatio-temporal data models have been discussed by Langran (1989, 1992). Langran assumed that data are embedded in Euclidean space, with the time dimension, being continuous, linearly ordered and orthogonal to the spatial dimensions. It is possible to go forwards or backwards in time, but not to branch. Mason et al (1994) suggest that although all dimensions are treated in the same way in the data model, this does not mean that, for example the same algorithm must be applied for interpolation in time as for interpolation in space. Different ways of treating space and time may sometimes have to co-exist. But nothing appears to be gained by hardwiring the human psychological view of a separate space and time into the data model. It is possible that allowing free specification of dimensions for all functions supported could well reveal beneficial new ways of relating to multi-dimensional data.

Models which incorporate dependencies in both space and time are described by Matalas (1967), Jenkins and Watts (1968), Rodríguez-Iturbe and Mejía (1974) and Rouhani and Wackernagel (1990). Most available data sets exhibit time-rich/space-poor characteristics as well as some form of temporal periodicity and spatial non-stationarity. To resolve this difference an operational solution is to split the space-time correlation either into a product (Rodríguez-Iturbe and Mejía 1974) or a sum (Rouhani and Hall 1989) of space and time components. Goodall and Mardia (1994) produced a spatio-temporal state space model using a spatio-temporal Kalman filter.

Rouhani and Wackernagel (1990) took the time series at each measurement point as separate, but correlated. Thus they focused on the dimension richest in information, which allows backwards and forwards forecasting, but not spatial prediction. When

considering environmental monitoring data, one can remove the effect of time and then view the data as repeated measurements in space (see Loader and Switzer 1992, Mardia and Goodall 1993).

This chapter discusses the development of new statistical methods for describing the temporal variations in the climate in ways which respect spatial relationships. A need to identify a concept that treats the components of cartographic time most effectively, and to develop that view of time as the base for a spatio-temporal model was required. The existing methods for interpolating climate are a first step in this process, and lead to the design of a variation of the spline model to link the spatial and temporal dimensions. The space-time spline model depicts processes of two-dimensional space (longitude and latitude) that are played out along a third (temporal) dimension. The representation of the third (temporal) face of a changing two dimensional object can appear complex, due to the seasonal nature of climate.

This chapter begins with a brief discussion of the issues that arose in the implementation of the space-time models. These include both computational and space-time issues. Two different representations of 'time' in space-time spline models are then examined. In the first model 'absolute' time, in which months are represented as equidistant and sequential series, is used. In the second model, a mapping of time to the long term monthly average to give an 'event' time is used. Several versions of this model were produced to better determine the relationships between the different dimensions. Both fixed and spatially varying long term means were used to give the 'event' time. Spline and partial spline versions of this model, which incorporated varying degrees of dependence on a transformed 'time' dimension, were tested.

The scaling of the 'time' axis used in the both spline models was shown to be critical to its performance and several methods for the selection of the scale were considered. The chapter concludes with a comparison of the results from the spline models and the results from the lag regression model presented in Chapter 6.

Implementation Issues for the space-time spline model

Space-time studies are an area of current interest in the field of Geographical Information Systems (GIS). These studies are typically beset by significant theoretical problems in deriving practical models and by difficulties in obtaining sufficient data to adequately test those models. Several critical implementation issues had to be considered in the development of the space time spline model. A number of implementation issues are raised that were considered either computing issues or space-time issues.

Computation issues

Computer issues include database size and program limits. The size of the data set can be a major limiting factor in the implementation of the spline space-time model. A data set which is relatively small in 2-D may become very large with the addition

of an extra dimension. This can lead to the requirement for very large amounts of computer memory and disk space. The limits of the ANUSPLIN (SPLINA) programs also set a restriction on the maximum number of data points which could be incorpoated in the model at the time (600 data points).

A further related point is that the data may be sparse in some dimensions compared with others. This highlights a serious difficulty in modelling higher dimensional spaces, namely that it is extremely difficult to provide sufficient input data for all points in the space. Mason *et al* (1994) suggest that when dealing with the sparse data sets characteristic of higher-dimensional spaces, it is particularly important to use an interpolation technique which allows the production of an error on each interpolated value, in order to prevent the user drawing inaccurate inferences from sparse data. While the grids used in the models have complete continental coverage, in some areas the data were interpolated from very sparse station network. The errors which are associated with these areas were shown in Chapter 4.

Space-time issues

The major implementation issue for the space-time spline model is whether or not space and time can be logically linked in a mathematical function. If they are logically linked then ways of establishing temporal intervals, defining the dynamics of dimensions, and defining temporal metrics become important.

Processes that operate in the world are embedded in time and can be directly or indirectly observed and measured. How then, does one represent the third (temporal) face of a changing two dimensional object? Time distribution can be summarized in terms of 'absolute' time or in terms of the process to be modelled to give an 'event' time. Annual cycles, which are very apparent in climate data, are a property of the processes not time. That is, cyclical time does not exist, but rather cyclical phenomena exist in linear time. A mapping of the climate variable to an 'event' time was used to incorporate the seasonal nature of the climate.

If it is possible to model space and time simultaneously what is the appropriate scaling of the time dimension? The scaling of elevation from metres to kilometres is discussed in Chapter 3. If one uses 'absolute' time is one unit of time in the time dimension equal to one degree of latitude or longitude in the space dimension? Is the change in the temporal dimension gradual and evenly spaced or is the change irregular and unequal? If so do processes have natural characteristic temporal scale at a specific times of the year?

The proposed model is a spline model which explicitly incorporates time and whose inputs consist of data from the mean sea level surfaces of monthly climate data arranged within a time dimension. The basis of the space-time spline model is to treat time as an independent variable when fitting the spline function. This implies, that the simultaneous space and time correlations can be modelled in the same space.

Both 'absolute' and 'event' time models were fitted. The output from the model is a monthly forecast or hindcast for the variable in question. The spatial and temporal coherence of at least some aspects of the model can be examined. A particular attraction of the model is the ease with which a mean sea level distribution can be put into a multi-site framework suitable for space-time studies.

As with the regression models, discussed in Chapter 6, the spatio-temporal spline models were created for several climatic variables but the dew-point temperature models are presented. The previous three months were used to predict the month of interest. While all months for 1989 and 1990 were fitted, the April 1989 models are used in the examples shown below. Figure 7.1 shows the monthly average and monthly dew-point temperature anomaly surfaces for January to April 1989. As can be seen in Figure 7.1 January and February were colder than normal, while the March and April were warmer than normal for most of the continent.

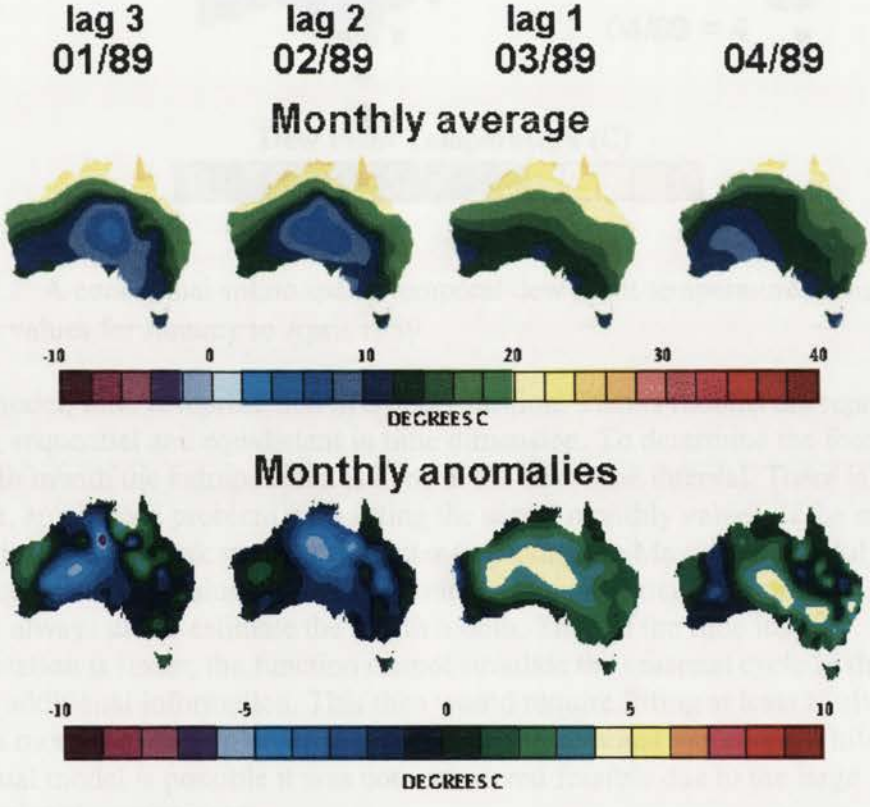


Figure 7.1: The monthly average dew-point temperature and monthly anomaly dew-point temperature surfaces for January to April 1989.

The 'absolute' time spline Model 1

Langran (1992) suggests that a useful starting point for a conceptual space-time data model is a set of successive images of space over time. Figure 7.2 shows a conceptual dew point spline model which has the successive monthly images evenly spaced and mapped in a linear fashion. That is, that the dew point temperature (T_d)

for the fourth month can be extrapolated from the spatial values from the previous three months and can be represented by

$$T_d(x, y, i) = f(x, y, i) + \varepsilon_{xyi} \tag{7.1}$$

where x and y are longitude and latitude and i is the month. In this case the input i values are January 1989 at $i = 1$, February 1989 at $i = 2$ and March at $i = 3$ to extrapolate April at $i = 4$.

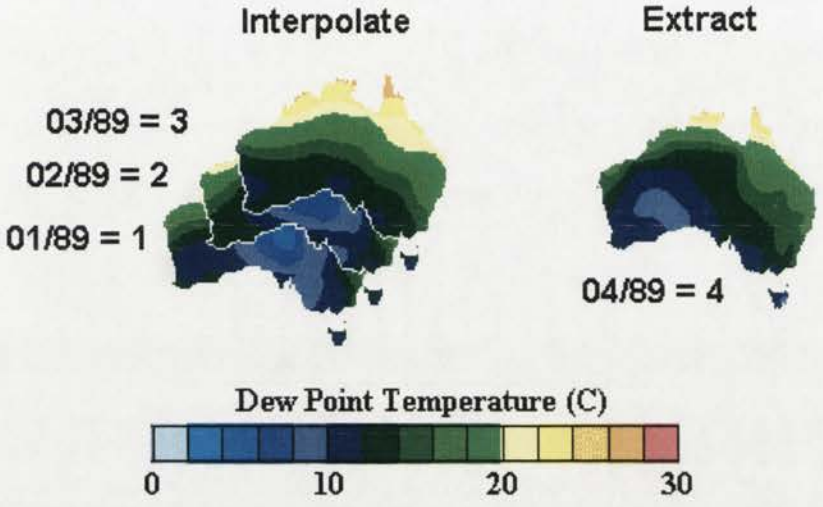


Figure 7.2: A conceptual spline spatio-temporal dew point temperature using monthly values for January to April 1989.

In this model, time is represented in a linear fashion. That is months are represented as being sequential and equidistant in time dimension. To determine the forecast for the fourth month the extrapolation is done at the next time interval. There is, however, an obvious problem with fitting the actual monthly values. If the month of interest is just after peak summer or winter (e.g.using the May, June and July values to extract the August values) then the extrapolation will extend the seasonal trend and will always under-estimate the fourth month. Thus, if the time interval representation is linear, the function cannot simulate the seasonal cycle in the values without additional information. This then would require fitting at least twelve to eighteen months of data in order to incorporate the seasonal variation. While this conceptual model is possible it was not considered feasible due to the large computational requirements.

In order to remove the seasonal trend from the 'absolute' time model, the monthly anomalies rather than the actual values were fitted (see Figure 7.3). That is, that the dew point temperature anomalies for the fourth month can be extrapolated from the anomaly values from the previous three months and can be represented by

$$\overleftrightarrow{T}_d(x, y, i) = f(x, y, i) + \varepsilon_{xyi} \tag{7.2}$$

where $\overleftrightarrow{T}_d$ are the monthly dew-point temperature anomalies.

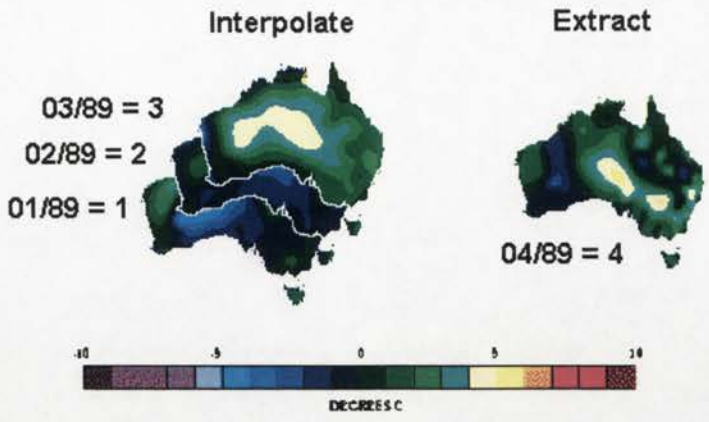


Figure 7.3: The 'absolute' time spline dew point temperature anomaly Model 1 for January to April 1989.

As the thin plate smoothing splines permitted straightforward incorporation of varying degrees of spatial/temporal dependence, seven different scales were used in the time domain. Second order spline functions were fitted to the dew point temperature anomaly data from the January, February and March 1989 'primitive' 2.5° grids. The scaling of 'time' was varied from *i /10* to *i*20* (where *i* are the time units defined above) in order to assess the relationship between the space and the time dimensions data. The grid from the April 'primitive' monthly surface was used as a validation surface.

Table 7.1 shows the output statistics from the spline Model 1 for April 1989. Table 7.1 shows that the validation rms of the 'absolute' time spline Model 1 is dependent on the scaling of the time dimension. The most noticeable point is that, for all scales greater than *i* scale the spline fits failed (all signal). The *i*15* was the best of the spline models for April 1989 with a validation rms of 1.053. A significant change in the validation rms occurs at the *i*5* scales and the reasons for this are not clear, although, it coincides with the initial failure of the spline function (i.e 100% SIGNAL).

Table 7.1: The validation rms, maximum difference, error, signal and RTGCV for the 'absolute' time dew-point temperature spline Model 1 for April 1989.

scale	Validation rms	Maximum Difference	ERROR	SIGNAL	RTGCV
i / 10	1.372	3.501	341.5	15.5	1.83
i / 5	1.373	3.506	340.9	16.1	1.83
i	1.237	2.931	263.0	94.0	1.71
i * 5	1.603	5.303	0	357.0 *	0.447
i *10	1.130	2.817	0	357.0 *	0.447
i *15	1.053	2.659	0	357.0 *	0.449
i *20	1.210	3.603	0	357.0 *	0.450

Figure 7.4 shows the resulting output grids which have been added to the long term average. It is interesting to note that as the scale gets smaller the extrapolation tends towards the average of the previous three months. The variation in the output grids shows the importance of using the appropriate scale when fitting the model.

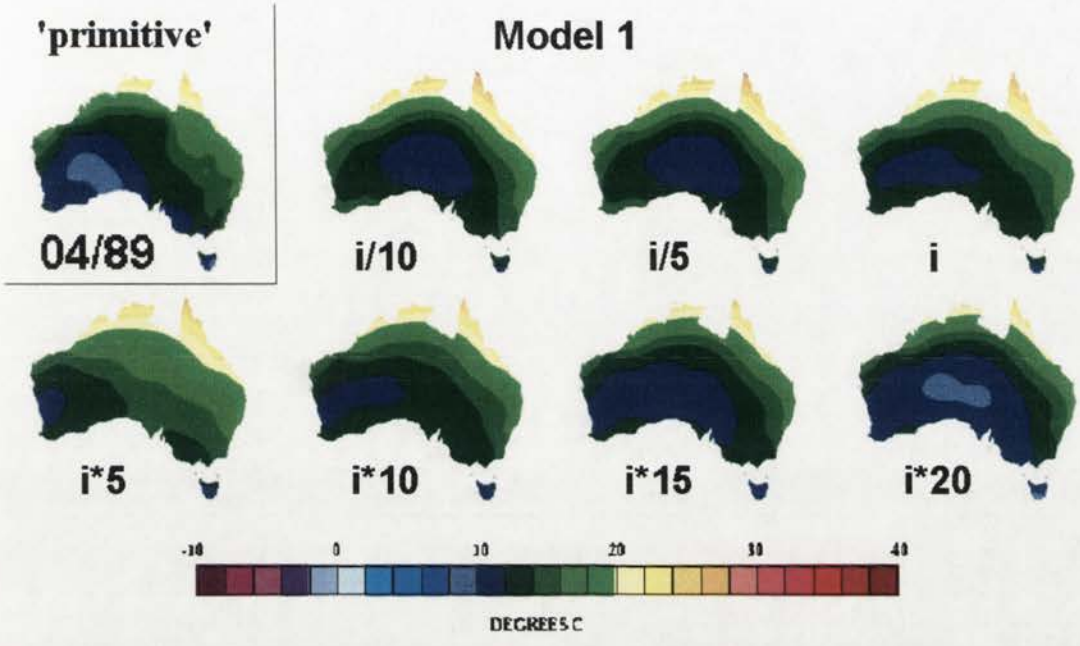


Figure 7.4: The output grids for the spline Model 1 for different time dimension scale options.

Second order spline models were fitted for all the months for1989 and 1990 with three independent variables (longitude, latitude and time). Figure 7.5 shows the validation rms for all months from January 1989 to December 1990. As can be seen from Figure 7.5 the best 'time' scale is not well defined. Overall the *i*15* gave the best results with an average validation rms of 1.455 (see Table 7.2) for all months in 1989 and 1990.

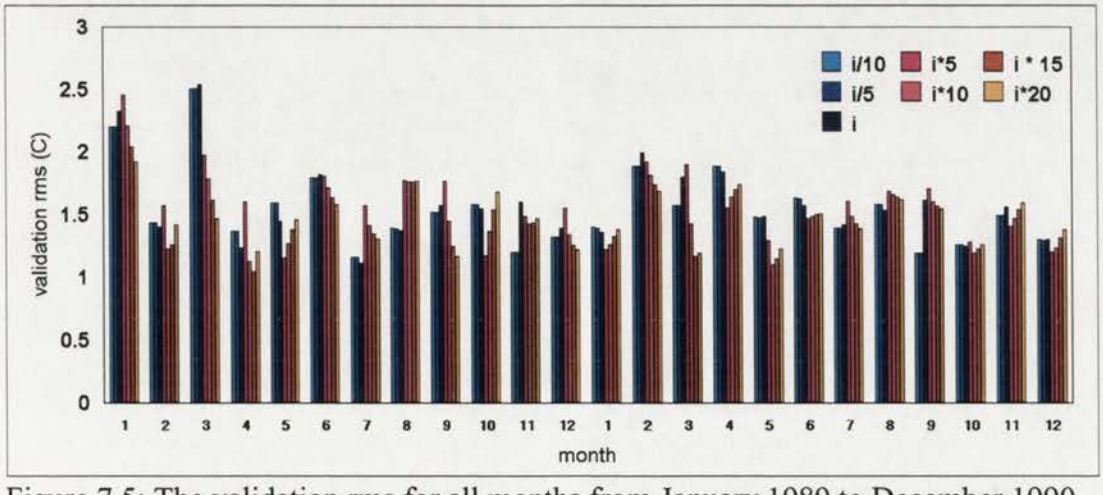


Figure 7.5: The validation rms for all months from January 1989 to December 1990 for the different scales fitted.

The choice of scale appears to be related to the seasonal cycle although this is not consistent. Overall, the summer months are fitted better at the *i*10* to *i*20* scales while the winter months are fitted better at the *i* to *i*5* scale. This, however, is not necessarily true for all months, for example, November and December 1990 give better results when fitted at the *i*5* scale). A seasonal scale which used

- *i*20* for January, February
- *i*15* for March, April, December
- *i*10* for May, October, November
- *i*5* for June and September
- *i* for July and August

was assessed and gave an average validation rms of 1.441. Table 7.2 shows this is a better result than is achieved by choosing any one particular scale. The 'best' scale is the average validation rms of the surfaces chosen by selecting the scale which has the minimum validation rms for each month.

Table 7.2: The average validation rms for the different scale options for the 'absolute' time dew-point temperature spline Model 1 for 1989 and 1990.

Scale	best	Season	*i*	*i*5*	*i*10*	*i*15*	*i*20*
Avg rms	1.350	1.441	1.591	1.591	1.480	1.455	1.469

The model, however, does show some merit and performs considerably better for most months than the lag regression model (see Chapter 6). The output from Model 1 is compared with the output from the dew-point lag regression model and the spline Model 2 later in this chapter.

The event 'time' spline Model 2

The apparent seasonal nature of the appropriate scaling for the temporal dimension in Model 1, suggests that a model based on 'event' time may overcome the problem of choosing the appropriate scale. To incorporate the seasonal nature of the climate a mapping of the climate variable to an 'event' time was used.

Model 2A - 'event' time as a mapping of the long term average values

In Model 2A, 'event' time is achieved by the mapping of time to the Australian long term average values. Conceptually this model is very similar to the 'absolute' time Model 1, but rather than having a linear sequential time dimension each month is plotted to a height which is equal to the Australian long term average surface dew-point temperature values (see Figure 5.10). This transformation makes 'time' annually periodic. It should be noted that the Australian long term average surface dew-point temperature values are averaged over both space and time. Thus, for example, the January values are fitted or extrapolated at an 'event' time height of 15.61, February at 16.35 and so on (see Figure 7.6).

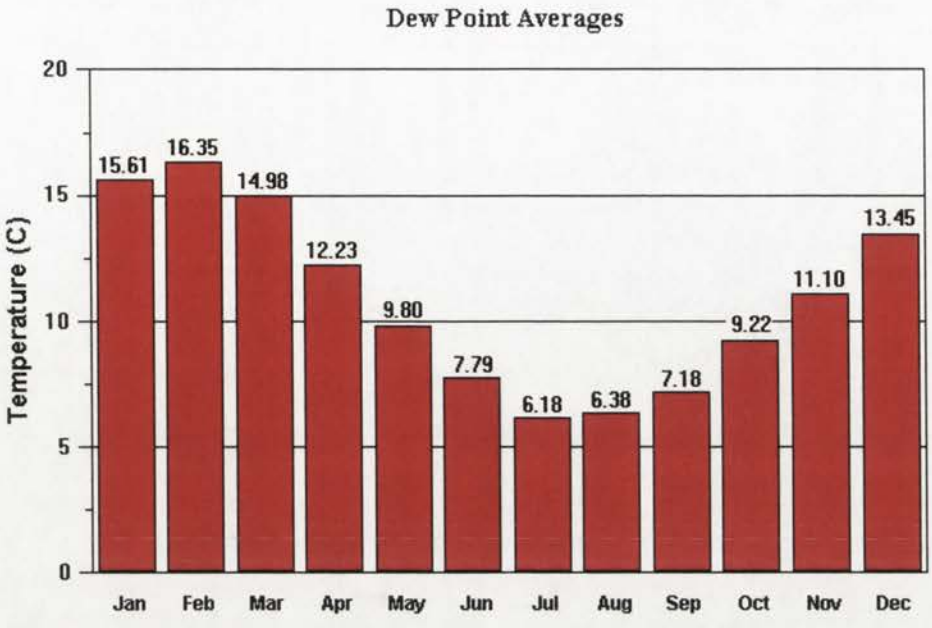


Figure 7.6: The average Australian SFLTA dew point temperatures.

That is, that the dew point temperature for the fourth month can be extrapolated from the spatial values from the previous three months and can be represented by

$$\overleftrightarrow{T}_{\mathrm{d}}(x, y, i) = f(x, y, \overline{\overleftrightarrow{T}_{d}(i)}) + \varepsilon_{xyi} \tag{7.3}$$

where $\overleftrightarrow{T}_{d}$ are the dew-point temperature anomalies, x and y are longitude and latitude, $\overline{\overleftrightarrow{T}_{d}(i)}$ are the average Australian SFLTA dew point temperatures and i is the month. For example, the model for April 1989 would have input i values of: 15.61 for January 1989, 16.35 for February 1989 and 14.98 for March 1989. To extrapolate the dew-point anomalies for April 1989 an i value of 12.23 would be used (see Figure 7.7).

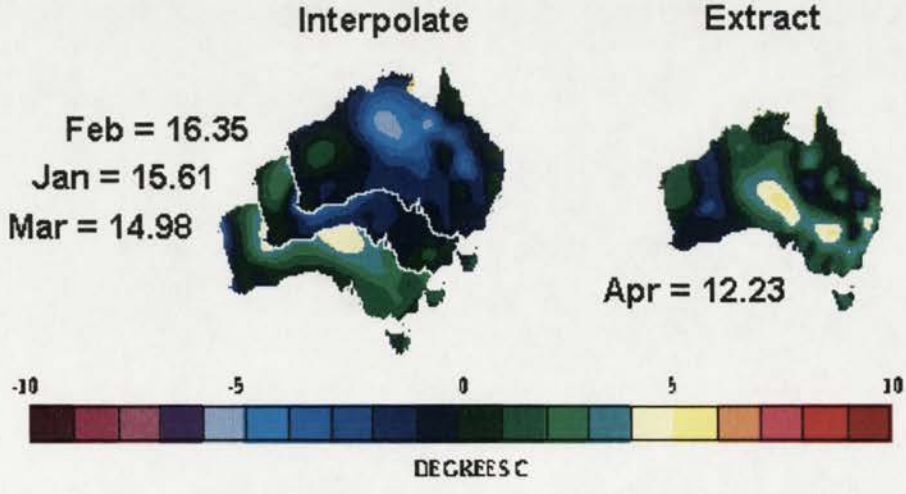


Figure 7.7: The 'event' time spline dew point temperature Model 2A for January to April 1989.

Table 7.3 shows the output statistics for the Model 2A output for April 1989. A comparison of the figures in Table 7.3 and Table 7.1, shows the use of the Australian long term average gave similar results to Model 1. Again, the higher *i* factor models failed (100% signal) but the validation rms associated with these models tends to be lower, with the minimum rms (1.219 °C) occurring at the *i*15* scale. The *i*5* surface shows the same unusual high validation rms that was evident in Model 1.

Table 7.3: The rms, maximum difference, RTGCV for the long term average 'event' dew-point temperature spline Model 2A for April 1989.

scale	Validation rms	Maximum Difference	ERROR	SIGNAL	RTGCV
i / 10	2.559	8.868	341.4	15.6	1.82
i / 5	2.559	8.867	341.0	16.0	1.82
i	2.556	8.433	264.6	92.4	1.76
i*5	2.565	9.025	0	357	0.477
i * 10	1.696	4.928	0	357	0.443
i*15	1.219	3.019	0	357	0.447
i*20	1.295	4.07	0	357	0.449

Figure 7.8. shows the interpolated April 1989 surface and the output from the spline model 2A for various scale factors. The output appears to be more sensitive to the appropriate scaling than the spline Model 1. That is, there are greater differences between the different scaling options in Model 2A than between the different scaling options in Model 1.

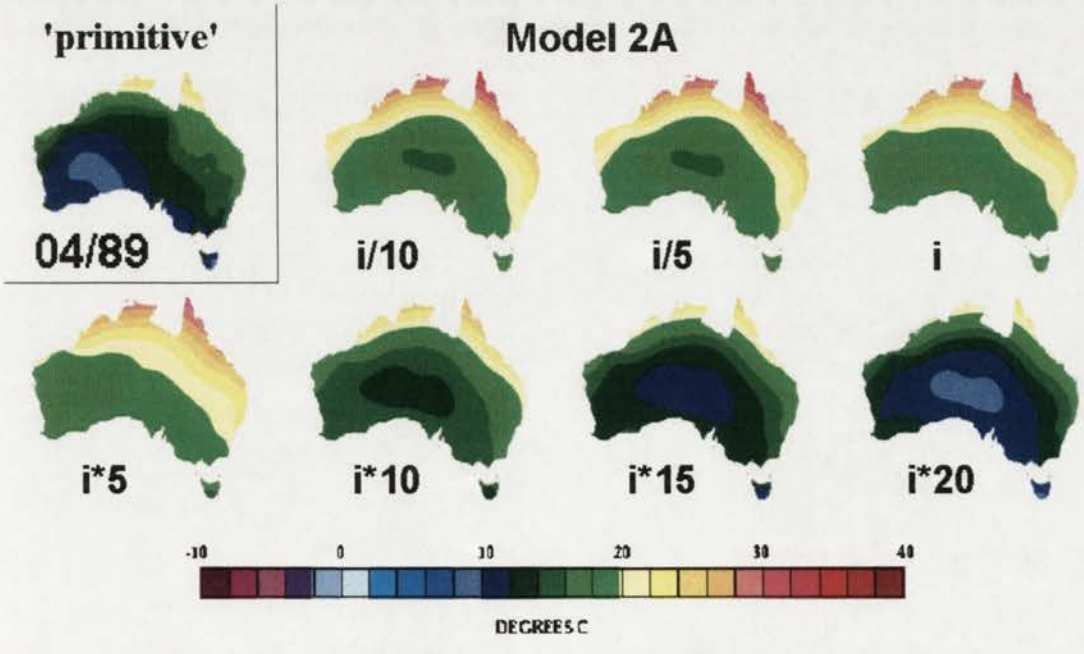


Figure 7.8: The output grids for the spline Model 2A for different time dimension scale options.

Second order spline models were fitted for all the months for 1989 and 1990 with three independent variables (longitude, latitude and time). Figure 7.9 shows the validation rms for all months from January 1989 to December 1990. As can be seen from Figure 7.9 the best 'time' scale is not well defined. Overall the *i*15* gave the best results with an average validation rms of 1.515 (see Table 7.4) for all months in 1989 and 1990.

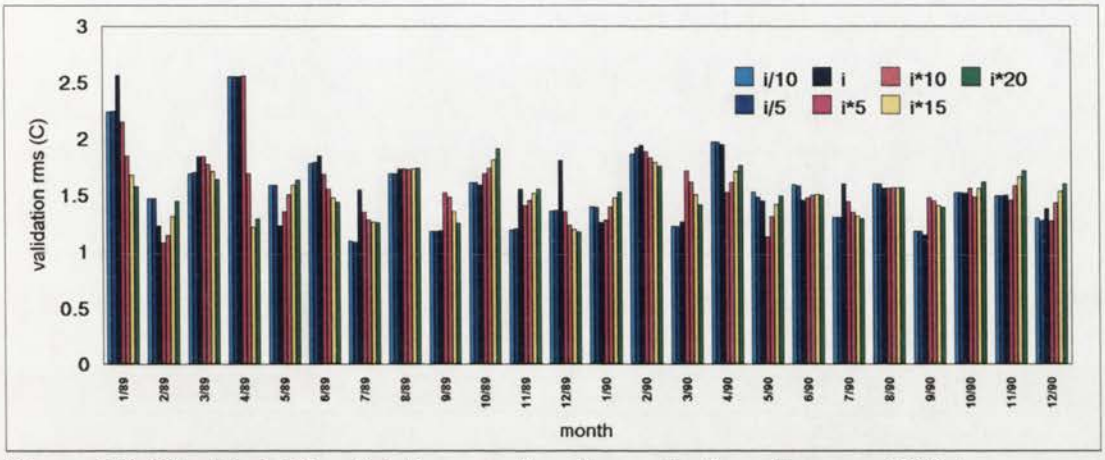


Figure 7.9: The Model 2 validation rms for all months from January 1989 to December 1990 for the different scales fitted.

Table 7.4: The average validation rms for the different scale options for the event time dew-point temperature spline Model 2A for 1989 and 1990.

Scale	best	*i/10*	*i/5*	*i*	*i*5*	*i*10*	*i*15*	*i*20*
Avg rms	1.362	1.563	1.562	1.615	1.565	1.526	1.515	1.525

Transformation of the time variable to the Australian long term monthly average dew point temperature performed better than Model 1 for some months but gave no improvement in the average rms. This suggests that the representation of the time dimension by a fixed Australian long term averages is not an advantageous mapping of time.

An investigation of the spatial distribution of the long term monthly averages as shown in Figure 7.10 shows that neither the Australian long term average values, nor its seasonal pattern are representative of the seasonal values or patterns across the country. For example, the long term values at the fixed value of 12.23 for April for the whole continent are only valid for the centre of the continent at 20 °S, and a small section of the east Coast at 30 °S. Figure 7.10 also shows that the seasonal pattern of the long term average at different points across the country varies. For example, the minimum temperature occurs in August for the centre of Australia at 20 °S, while it occurs in July on the east and west coasts at 20 °S.

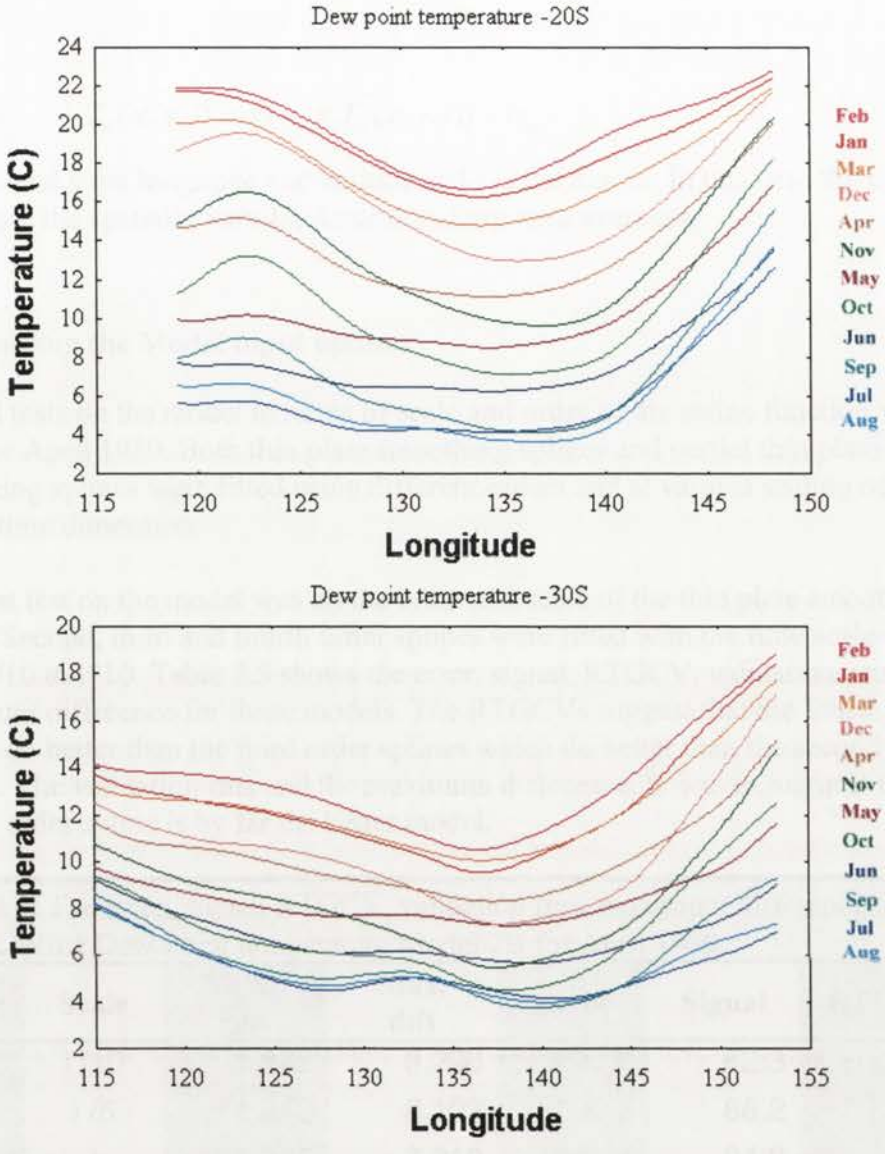


Figure 7.10: The long term average dew-point temperature transects through 20°S and 30°S.

Model 2B - 'event' time as a mapping of the local long term average values

Model 2A suggests that an improved representation of the event time intervals would be achieved by using spatially varying long term average values. That is, the monthly anomalies are plotted on an underlying grid that is the long term average surface for that month. While this model appears more complicated than the spline Model 1 or 2a, it is actually computationally simpler. Considering that the long term monthly average plus the monthly anomaly is the monthly value, Model 2B essentially fits a spline function to three consecutive months dew-point temperature values. An extraction grid of the spatially varying long term monthly averages is required to interpolate/extrapolate the fourth month.

The dew-point temperature for the fourth month can be interpolate/extrapolated from the spatial values from the previous three months and can be represented by

$$T_d(x,y,i) = f(x,y,\overline{T_d}(x,y,i)) + \varepsilon_{xyi} \qquad (7.4)$$

where x and y are longitude and latitude and i is the month. In this case the input i values are the spatially varying Australian long term averages.

Determining the Model input options

Several tests on the model in terms of scale and order of the spline function were done for April 1989. Both thin plate smoothing splines and partial thin plate smoothing splines were fitted using different orders and at various scaling of the 'event' time dimension.

The first test on the model was on the order and scale of the thin plate smoothing spline. Second, third and fourth order splines were fitted with the time scale varying from i/10 to i*10. Table 7.5 shows the error, signal, RTGCV, validation rms and maximum difference for these models. The RTGCVs suggest that the fourth order splines do better than the third order splines which do better than the second order splines. The validation rms and the maximum difference, however, suggest that the second order spline is by far the better model.

Table 7.5: The error, signal, RTGCV, validation rms, maximum difference for the full spline Dew Point temperature Model 2B for April 1989.

Order	Scale	Valid. rms	Max. diff	Error	Signal	RTGCV
2	i /10	1.842	8.290	294.7	62.3	2.11
2	i /5	1.850	8.102	290.8	66.2	2.10
2	i	1.875	7.319	262.1	94.9	2.00
2	i * 5	1.475	4.311	218.7	138.3	2.18
2	i * 10	1.306	3.916	225.4	131.6	2.34
2	i * 15	1.750	5.832	349.8	7.2	2.38
2	i * 20	1.755	5.846	350.2	6.8	2.38
3	i /5	2.99	21.197	309.0	48.0	1.64
3	i	3.16	23.113	296.2	60.8	1.64
3	i * 5	2.76	17.955	269.1	87.9	1.81
3	i * 10	2.47	12.914	255.3	101.7	1.99
4	i /5	3.52	30.218	304.3	52.7	1.58
4	i	3.84	37.057	297.8	59.2	1.55
4	i * 5	3.68	32.411	272.2	84.8	1.71
4	i * 10	3.49	27.164	261.2	95.8	1.92

Figure 7.11 shows the varying scale and order space-time spline Model 2B outputs for January to April 1989. The figure shows that the second order splines perform much better than the third and fourth order splines. The second order Model 2B with a scale of *i*10* has the lowest validation rms (1.306 °C), the smallest maximum difference (3.916 °C) and figure 7.11 shows that it gives the best representation of the April 1989 dew-point temperature surface. The second order spline models show that the scaling of the time domain for April 1989 is important in determining the output values. This is discussed further, below.

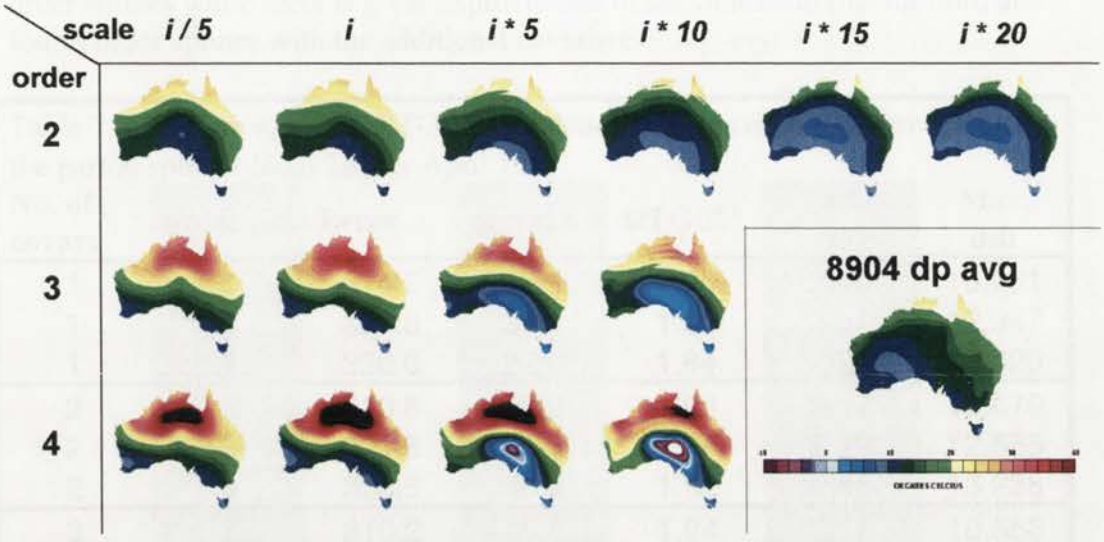


Figure 7.11: Varying scale and order space-time spline Model 2B outputs for April 1989.

The partial spline models

It is most likely that the higher order splines may have been failing due to the lack of data in the time domain (ie 3 months), but the lower GCV is a result of a better fit in the space domain (119 points). In order to assess whether or not the higher order splines were failing due to the lack of data, additional covariates were added to separate out a more flexible dependence on the time domain. Second, third and fourth order partial splines were fitted with one, two and three covariates.

The single covariate model is the stable partial spline, discussed in chapters 3 and 4, which gives linear interpolation in the time (i) domain. The function can be represented by:

$$f(x,y)+\beta * i \tag{7.5}$$

By the addition of a second covariate, in this case the i^2 value divided by the mean i^2 value, allows a quadratic dependence on i. The function can be represented by:

$$f(x,y)+b_1 * i + b_2 * \frac{i^2}{(\overline{i^2})} \tag{7.6}$$

By the addition of a second covariate, in this case the i^3 value divided by the mean i^3 value, allows a cubic dependence on i. The function can be represented by:

$$f(x,y) + b_1 * i + b_2 * \frac{i^2}{(i^2)} + b_3 * \frac{i^3}{(i^3)} \qquad (7.7)$$

Table 7.6 shows the output statistics for the partial spline models. A comparison with Table 7.5 shows that the RTGCV improves for the second order derivative splines but there is no improvement in the RTGCV for the third and fourth order derivative splines. The validation rms, however, do not show any improvement for the second order splines while there is great improvement in the validation rms for third and fourth order splines with the additional covariates.

Table 7.6: The error, signal, RTGCV, validation rms, maximum difference for the partial spline Model 2B, for April 1989.

No. of covars	order	Error	signal	RTGCV	Valid. rms	Max. diff
1	2	314.2	42.8	1.95	2.10	10.761
1	3	325.8	31.2	1.86	2.31	12.947
1	4	330.0	27.0	1.84	2.37	13.629
2	2	310.8	46.2	1.93	2.17	10.679
2	3	321.8	35.2	1.84	2.39	12.585
2	4	326.5	30.5	1.81	2.46	13.238
3	2	310.2	46.8	1.94	2.18	10.558
3	3	321.0	36.0	1.84	2.40	12.514
3	4	325.5	31.5	1.82	2.46	13.304

Figure 7.12 shows the output grids for the partial spline models and confirms that the outputs for the partial splines with the additional covariates are much better for the higher order derivative spline models than for the comparable full spline models. The higher order derivative partial splines overestimate the dew-point temperatures but are within a reasonable range, while the full spline higher order derivative models gave very high and very low values. The over-estimation of the dew-point temperature suggests that the time dimension is still not adequately represented in the higher order splines.

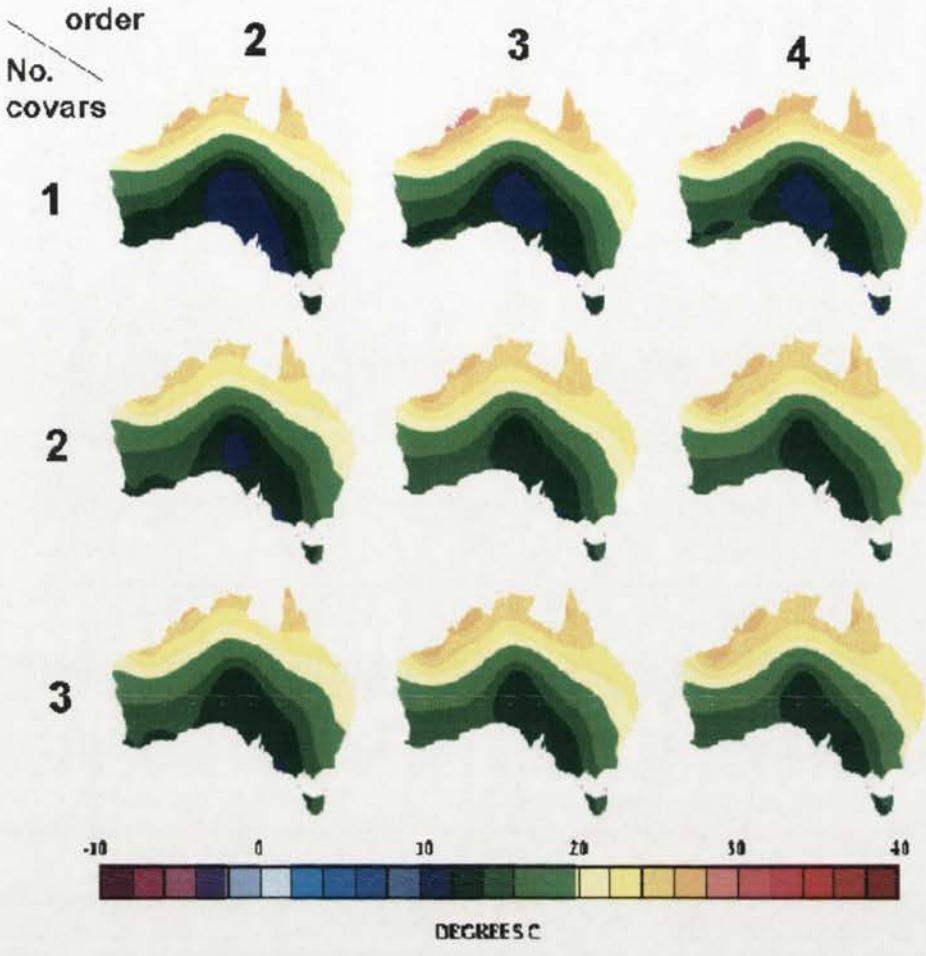


Figure 7.12: Outputs for varying order and number of covariates for the space-time partial spline Model 2B, for January to April 1989.

The second order partial splines on the whole perform reasonably well. The model fitted to two independent variables (longitude and latitude) and only one covariate (dew-point temperature) performed better than the other second order partial splines. This model, however, did not perform as well as the full spline model as measured by the validation rms, which suggests there is both spatial and temporal dependence.

The spline models

Second order spline models were fitted for all the months for 1989 and 1990 with three independent variables (longitude, latitude and 'event' time). As the thin plate smoothing splines also permitted straightforward incorporation of varying degrees of temporal dependence, five different scales were used in the time domain (see Table 7.5). Figure 7.13 shows the validation rms for all months from January 1989 to December 1990. As can be seen from Figure 7.13 the best pseudo-time scale is not well defined, with all pseudo-time scales performing better than the others for at least two months. The *i/10* (8 months) and the *i* scale (7 months) give the best results. The choice of scale appears not to be related to the seasonal cycle nor is it consistent for specific months (e.g. January 1989 is best at *i/10* while January 1990 is best at the *i*

scale). The determination of the 'best' scale does not appear to be related to whether the extraction grid is an interpolation (extracting data for psuedo-times that lie inside the input grids) or an extrapolation (extracting data for psuedo-times that lie outside the input grids).

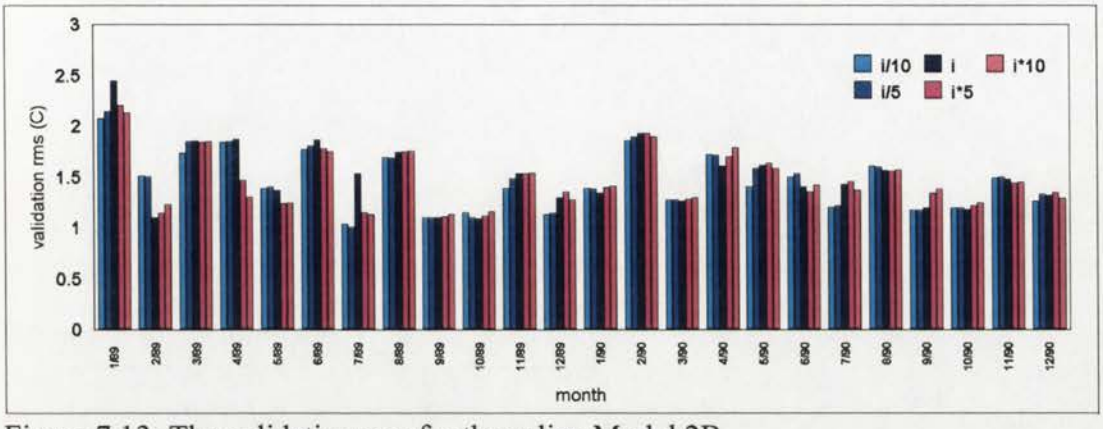


Figure 7.13: The validation rms for the spline Model 2B.

As can be seen in Figure 7.13, the "best" scale of the 'event' time is not obvious, but for most months the difference between the different scales is minor. For some months however the correct scaling can have as much as a degree difference in the validation rms (e.g. January and July 1989). Figure 7.14 shows the output for the different scaling options from the spline Model 2B for mid-season months for 1989 and 1990. For most months the different scaling of the 'event' time dimension makes little difference to the output grid. Other months, such as January and July 1989, show that a difference of 1 degree in the validation rms can have a significant impact on the output grid.

In Chapter 3 the fitting of the surface was refined by minimising the RTGCV. In the case of the space-time spline model, however, a reduced RTGCV did not necessarily show improved results in the extrapolated surfaces. An investigation of the output statistics from the SPLINA program in Table 7.5, shows the minimum validation rms does not occur with the minimum RTGCV nor is it related to the ratio of signal to error statistics for April 1989.

On the whole, however, all the spatially varing spline Model 2B (Table 7.5) performs better than either Model 1 (Table 7.1) or Model 2A (Table 7.3) and significantly better than the partial spline models (see Table 7.6).

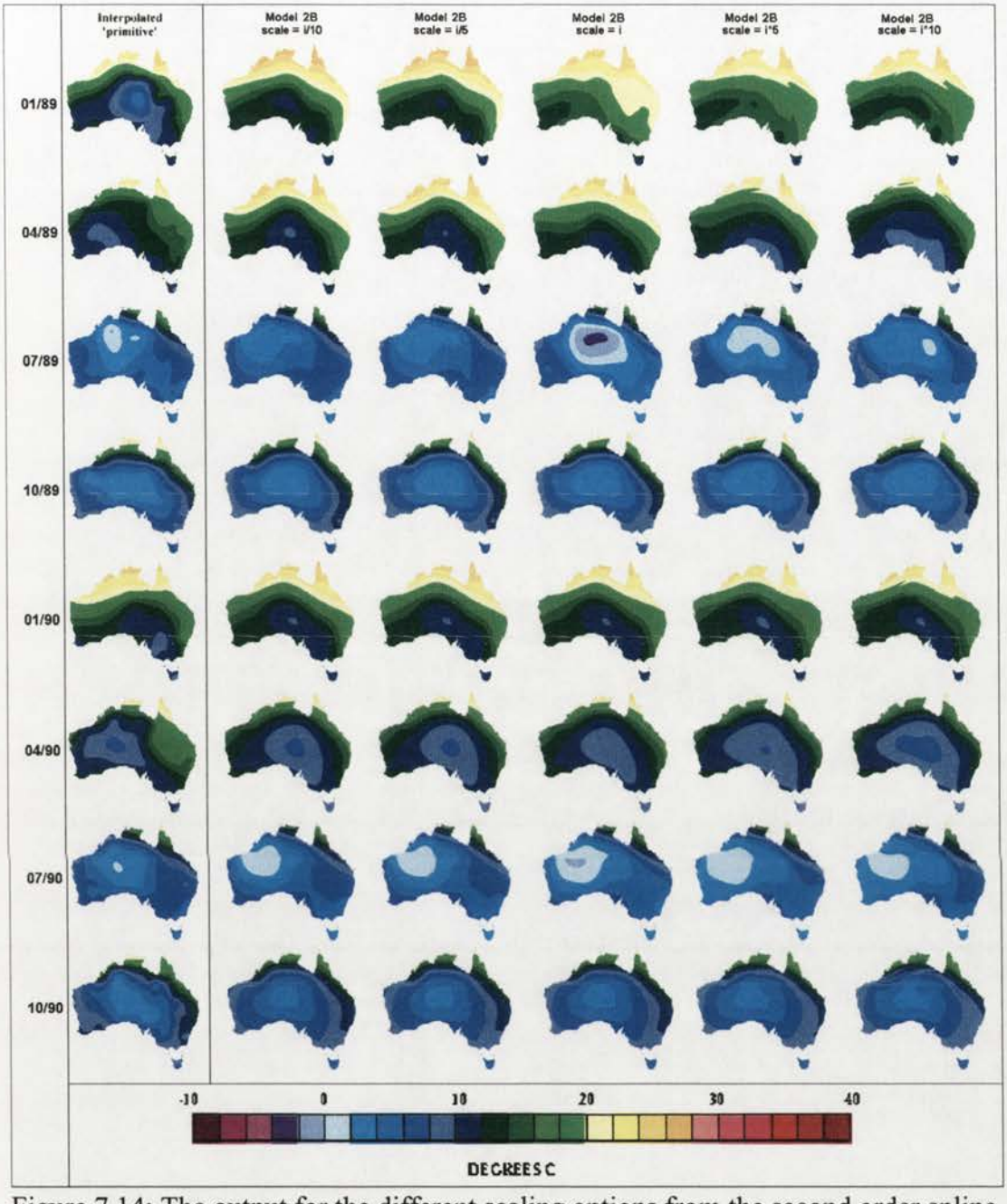


Figure 7.14: The output for the different scaling options from the second order spline Model 2B for mid-season months for 1989 and 1990.

Selecting the pseudo-time scale by hindcasting

To circumvent the scaling problem and to test whether or not the same scaling occurs through the entire space, hindcast extractions were done for all months from January 1989 to September 1990. That is, for each spline model that was fitted for a forecast month, the same model was used to extract the month previous to the input 'primitive' grids. For example, the December 1988 surface was extracted as a hindcast to select the appropriate scale for the April 1989 model.

Figure 7.15 shows the 'best' (minimum) validation rms and the validation rms as selected by the hindcast model. For 11 of the 24 months the 'best' scale factor was the same as the scale selected by the hindcast model. For another 6 months the scale factor selected by the hindcast model was one off the 'best' scale (e.g. the i scale rather than the i/5 scale), suggesting that the correct scale was somewhere in-between the two scales (e.g. i/3). The greatest difference between the 'best' validation rms and the validation rms as selected by the hindcast model was 0.25 °C.

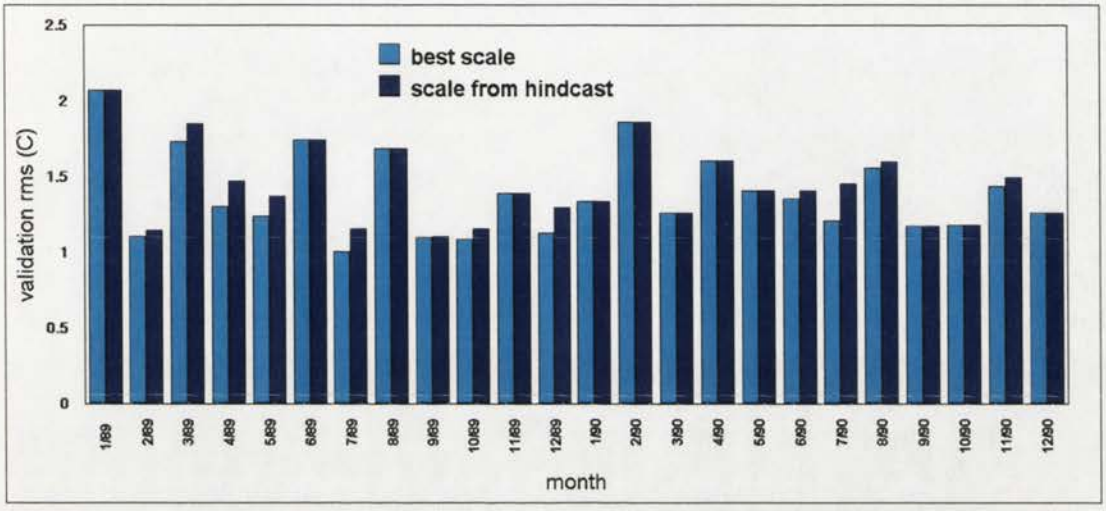


Figure 7.15: The minimum validation rms and the validation rms as selected by the hindcast model.

Table 7.7 shows the average validation rms for the different scale options for the 'event' time dew-point temperature spline Model 2B for 1989 and 1990. As can be seen the hindcast option performs better than any fixed scale, although not as well as would be possible if the 'best' scale could be pre-determined.

Table 7.7: The average validation rms for the different scale options for the 'event' time dew-point temperature spline Model 2B for 1989 and 1990.

Scale	'best'	*Hind-cast*	*i/10*	*i/5*	*I*	*i*5*	*i*10*
Avg valid. rms	1.388	1.440	1.457	1.480	1.508	1.475	1.469

As the selection of the scale factor by the hindcast model was better than selecting any one particular scale, this suggests that the hindcast option for selecting the scale factor is a viable way to circumvent the selection of the 'time' scale. The premise that the same scale operates when extracting data from the same three months also appears to have some validity. This then suggests that the spatial and temporal distributions of some climatic variables can be modelled simultaneously. It appears that there is merit in the spatially varying spline models and that there is some validity in the assumption that the incorporation of a spatially varying dependence on appropriately scaled time dimension contributes to surface accuracy.

Comparison of the output from the lag regression model and the space-time spline models

As the spatio-temporal spline models were fitted for the same months as the dew-point temperature lag regression model (see Chapter 6), some comparisons can be made as to the usefulness of the models. In Chapter 6, the validation rms from the lag regression models was compared to the validation rms using the average monthly surfaces (see Figure 6.8). Figure 7.16 below shows these figures and the validation rms for the 'best' scale spline models. Figure 7.17 is similar, but shows the validation rms for the seasonal option for Model 1, the fixed *i*15* option for Model 2A and the hindcast option for Model 2B.

As can be seen in Figures 7.16 and 7.17 the validation rms for the spatio-temporal models is not dependent on how different the forecast month is from the long term average. This is not true for the dew-point lag regression model. This then suggests that the regression model, while it incorporates some temporal history, tends to forecast the long term average.

All the spatio-temporal spline models performed better than the lag regression Model in all but six of the 24 months. If the 'best' scales could be pre-determined the spatio-temporal spline models out-performed the lag regression model in all but three of the 24 months. For most of the months in which the lag regression Model does better (e.g. December 1989 and July 1990) the differences in validation rms for the two models are small. For some months, such as August 1989 and April 1990, all the spline models do markedly better than the lag regression Model. Table 7.8 shows the average validation rms for all months for the long term average, the lag regression model and both the 'best' and 'appropriate' scale factors for all the spatio-temporal spline models. As can be seen in Table 7.8, all the spatio-temporal models perform considerably better than the lag regression model which performs slightly better than using the surface long term average to predict the monthly values.

Table 7.8: The average validation rms for the long term average, the lag regression model and the spatio-temporal spline models for 1989 and 1990.

Model	SFLTA	*Lag regression*	*Spline Model 1*	*Spline Model 2A*	*Spline Model 2B*
'best'	2.114	1.906	1.362	1.362	1.388
'appropriate'			1.441	1.515	1.440

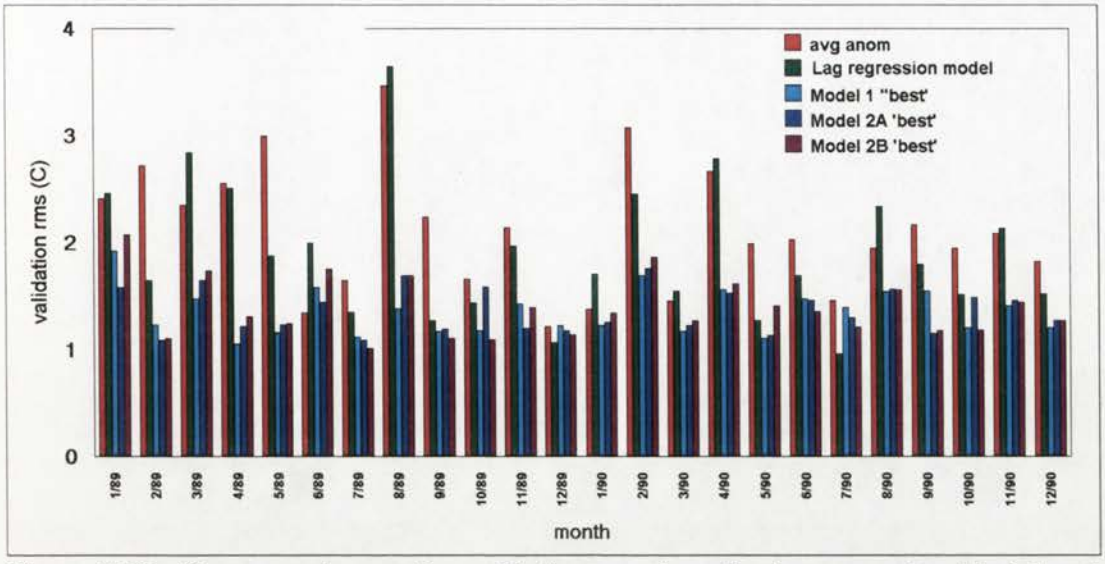


Figure 7.16: The anomaly rms, the validation rms from the lag regression Model and the 'best' validation rms for the forecast spline Model 1, 2A and 2B.

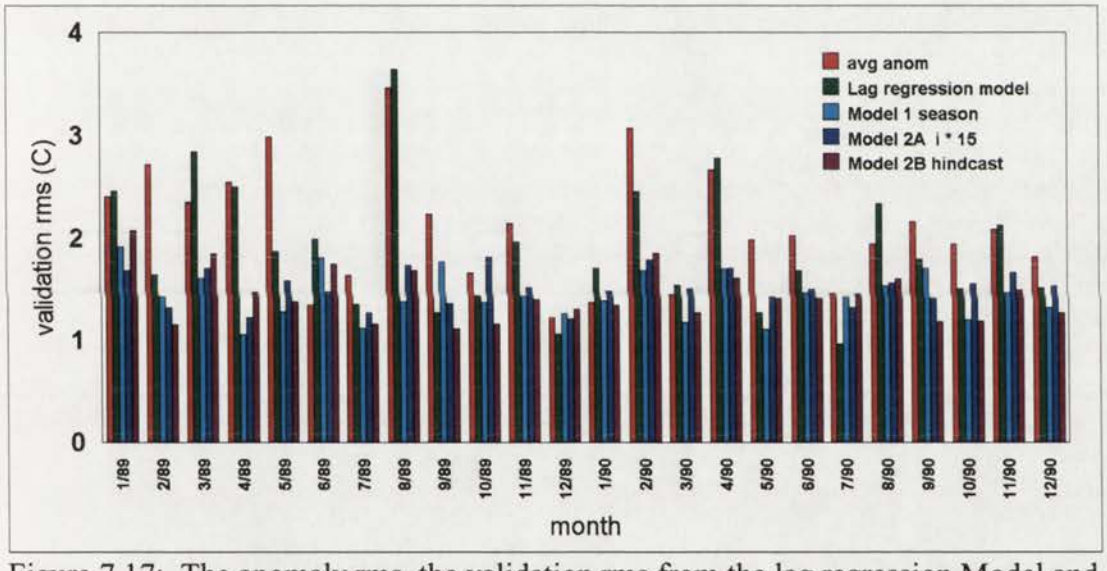


Figure 7.17: The anomaly rms, the validation rms from the lag regression Model and the validation rms for the season scale Model 1, the fixed *i*15* scale Model 2A and hindcast scale Model 2B.

Figures 7.18 and 7.19 show the monthly average dew-point temperature (i.e. the expected 'primitive' output), the output from the dew-point temperature lag regression model and the output for the spatio-temporal spline models for 1989 and 1990. Figures 7.18 and 7.19 show that much of the improved validation rms in the spatio-temporal spline models over the lag regression model is due to the spatio-temporal spline models forecasting the actual values better (i.e. overall colour matching is better). Only small improvements are due to the spatio-temporal spline models forecasting the spatial distribution of the values better (i.e. pattern matching is better).

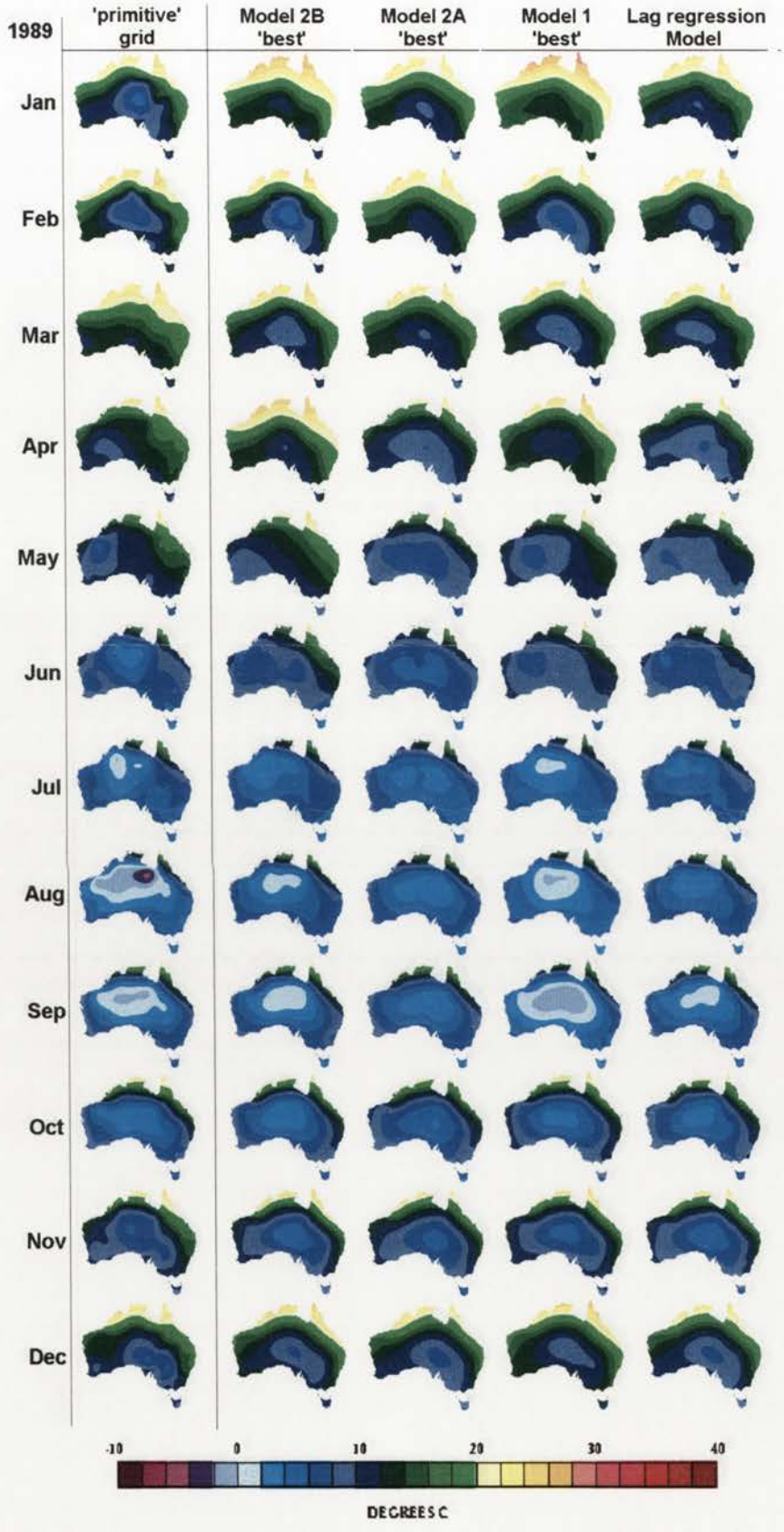


Figure 7.18: The monthly average dew-point temperature ('primitive'), the output from the dew-point temperature lag regression model and the output for the spatio-temporal spline models for 1989.

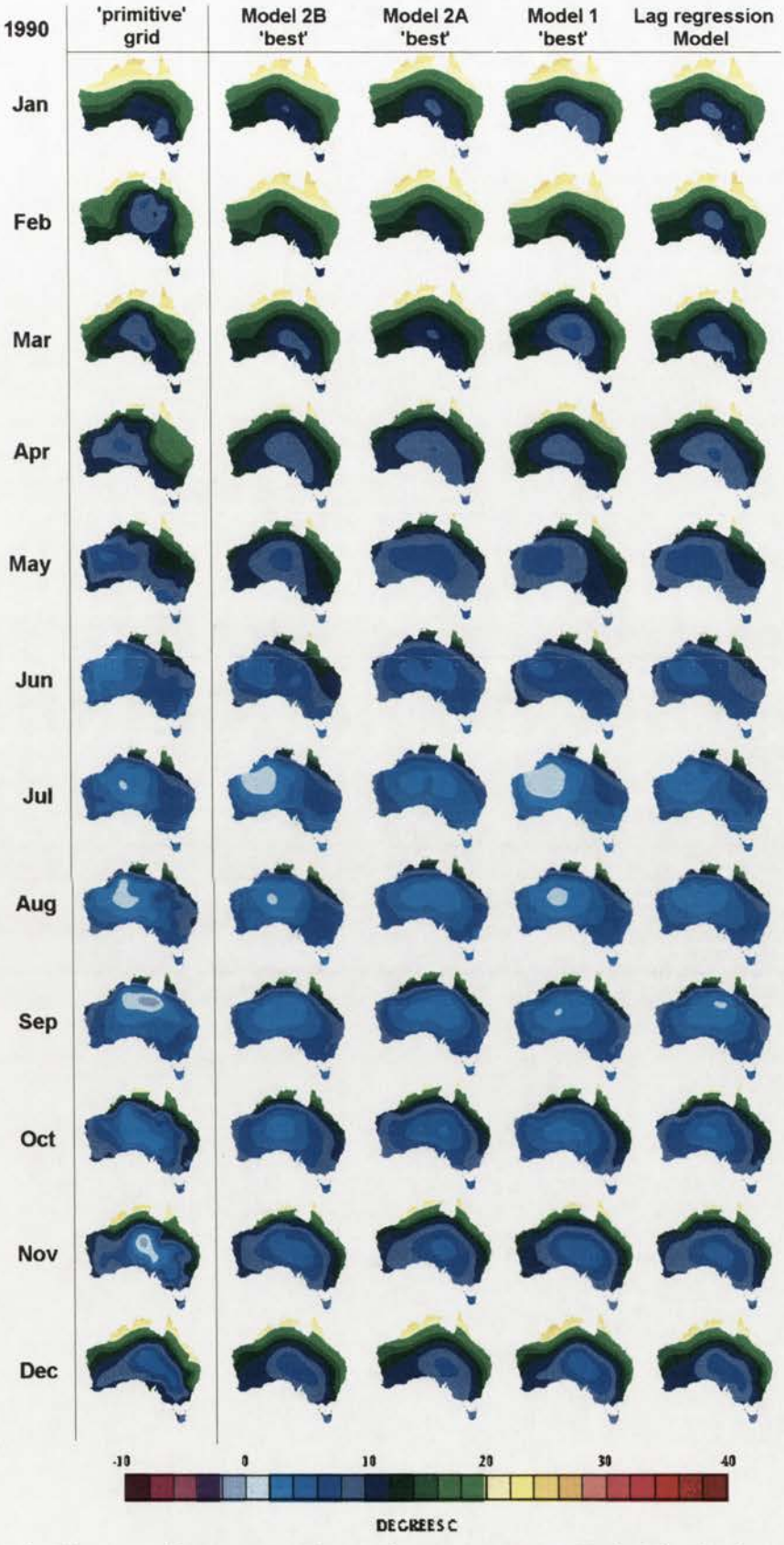


Figure 7.19: The monthly average dew-point temperature ('primitive'), the output from the dew-point temperature lag regression model and the output for the spatio-temporal spline models for 1990.

Conclusions

This chapter introduced a new type of spatio-temporal model for the analysis of climate variables which takes account of the natural correlations in time and space. The spatio-temporal spline model depicts processes of two-dimensional space that are played out along a third temporal dimension.

Two different representations of the time ('absolute' and 'event') dimension in the spatio-temporal spline models were examined. A particular attraction of the spline models is the ease with which a mean sea level distribution can be put into a multi-site framework suitable for space-time studies.

In Model 1, 'absolute' time is represented as an equispaced linear series. The monthly anomalies from the 'primitive' surfaces were fitted at unit intervals. The scale factor for the unit intervals was shown to be critical to the performance of the model. A seasonal scaling was shown to be the best option for the validation data, but only marginally better than any particular fixed scale factor.

Model 2A, which used a mapping of the long term average to represent an 'event' time, allows for the incorporation of the cyclic nature of climate. While Model 2A performed better than Model 1 for some months it gave no improvement in the average rms. Model 2A had similar problems to Model 1 with regard to the choice of the 'appropriate' scale factor for the 'event' time dimension. While the $i*15$ scale factor model showed the lowest average validation rms (1.515) for the fixed scales, the validation rms was considerably higher than could be achieved by the optimum ('best') scale factor for each month which gave an average rms of 1.362. This suggests that the representation of the time dimension by a fixed continental Australian long term average is not an advantageous option for mapping of the time dimension.

Model 2B, which used a mapping of the local long term average to represent an 'event' time, allows for the incorporation of spatial variations in the underlying seasonal cycles. Model 2B uses a spatially varying long term average transformation to represent a 'event' time. Several tests were done for this model and it was shown that a second order spline fitted to three independent variables gave the most realistic results.

Model 2B, however, produced some unexpected results in the models fitted with higher order derivatives, which suggested that the space-time spline model was not well defined in the time domain. That is, the function was modelling very little of the temporal richness, whilst heavily representing the spatial distribution. The incorporation of additional covariates (i^2 and i^3) did contribute significantly to interpolation accuracy of the 3rd and 4th order fits.

It was shown that the incorporation of a spatially varying dependence on appropriately scaled time made a dominant contribution to surface accuracy. The results of the spline and partial spline showed that better performances were achieved with time as an

independent variable. This suggests there is both spatial and temporal dependence in the distribution of the climate variables.

Thin plate smoothing splines also permitted straightforward incorporation of varying degrees of topographic/temporal dependence. This scaling of the dimensions can normally be slightly refined by choosing the elevation scaling to minimise the GCV of the fitted surface. In the case of all the models, however, a reduced GCV did not necessarily show improved results in the extrapolated surfaces.

The selection of the psuedo-time scale by use of the hindcast extraction was shown to be a possible way of circumventing the temporal scale factor problem. A comparison of the output of spatio-temporal Models 1 and 2 with the output of the lag regression model showed that the spatio-temporal models performed much better than the lag regression model. The approach of fitting space and time simultaneously appears to hold promise of some useful outputs for monthly forecasting from limited data.

Chapter 8 - Summary of Findings and Conclusions, Including Research Directions and Potential Applications of Methodology.

Introduction

The purpose of climatic data analysis is the characterization of the state of the atmosphere in a manner that is physically consistent and spatially and temporally coherent. The raw material for this process consists of the countless individual observations, irregularly distributed in space and time and often erroneous and conflicting. The final product is a regular space/time representation of the atmospheric state in which it is hoped that the maximum signal has been extracted from the observations and that the effect of the observational noise has been minimised.

At first sight, spatial analysis seems to be little more than classical interpolation. The observations are not simply evaluations of an analytic function, and the spatial analysis of atmospheric variables is much more than classical interpolation. First, the observations have errors that can be temporally and spatially correlated with each other or with the signal. Second the atmospheric state variables are related to each other by the governing laws; they are not independent. Third, there are large data voids where no direct observations exist.

This final chapter presents a summary of the major results and findings, followed by a discussion of research directions. The thesis concludes with comments on possible applications of the methodology to climate forecasting at the monthly time scale, monitoring climate change and environmental analysis.

Summary of findings and conclusions

This analysis of the Australian climate fields presented a framework for the spatial and temporal extension of climatic variables. The temporal and spatial dimensions of the problem were considered equally important and were given equal weight in this thesis. The spatial estimation and prediction of climatic variable site data requires the development and integration of a range of tasks, including:

- the production of a temporally homogeneous station climate database;
- an assessment of the appropriateness of the spline interpolation method for the spatial extension of climatic variables;
- the development of a spatial database of gridded estimates of primary climate elements for the period 1952-1990. This included an assessment of the validity of the surfaces, statistical checks on the homogeneity of the database and an assessment of the errors associated with the grids.
- The production of other gridded estimates of derived climatic variables by applying logical models to the primary climate variables.
- The investigation of the points of a 2.5 x 2.5 degree grid of estimates of climatic variables for an assessment of climate variability and change.

- The application of advanced graphic systems to better communicate the results of complex spatial and temporal analysis.
- The use of the 2.5 x 2.5 degree grid points to establish the correlation parameters for the regression and lag regression models of climatic variables. The spatial interpolation of those parameters to produce the grids required for the regression models. The assessment of the model performance based on a withheld data set.
- The production of the space-time spline model for dew-point temperature which requires the suitable transformation of the time dimension so that the climatic variable can be modelled simultaneously in two-dimensional space and one-dimensional time.

Each of these tasks presented several findings and conclusions.

The production of a temporally homogeneous station climate database

- Specific UTC times were required and selected to accommodate the different time zones across the country and to remove station inhomogeneity problems associated with the introduction of Daylight Saving Time in the early 1970s. Linear interpolation was found to be a satisfactory method of producing specific UTC times.
- The three-hourly data set from the Australian Bureau of Meteorology was found to suffer from large discontinuities in both temporal and spatial coverage. The data set was also found to contain errors from several sources.
- The number of stations for individual months varied between 34 and 113 stations. This was found to have implications for the spatial interpolation and for assessing spatial long term averages.
- The spatial distribution of the stations was found to be heavily biased towards the eastern seaboard of Australia and was particularly poor in sparsely populated areas, but it was a good representation of all the station records from the BoM.
- The station DEMs showed the distribution of stations to be poorest for areas of high elevation especially along the the eastern seaboard of Australia. The topographic representation was found to be better for the data sparse regions.
- Rainfall percentiles were considered more spatially suited to interpolation than actual rainfall amounts for the number and distribution of rainfall stations which had full records available for the period 1952-1990 (3160 stations).
- While the 3160 stations in the data set well represent the entire BoM station network, it was found to be quite sparse in parts of central and northern Australia. This had implications for the subsequent spatial interpolation and meant that interpolated values should be viewed with caution in these areas.

An assessment of the appropriateness of the spline interpolation method for the spatial extension of climatic variables

- Partial thin plate splines were found to be well suited to the interpolation of pressure and temperature fields. The dependence of the fitted data on the estimated function is determined from the data themselves and there are no

unnecessary restrictions imposed on the form of the fitted function. The thin plate smoothing spline technique directly addresses two major aspects of the general interpolation problem, namely the statistics of the data errors and the roughness or complexity of the underlying meteorological field.

- The advantages of running the spline function by minimising the GCV option was found to be the most appropriate method of determining the surface coefficients and was used as the first choice optimisation directive for all surfaces.
- An assessment of the options available in the ANUSPLIN package, including the incorporation of altitude in the climate surfaces, showed that a partial spline function which incorporated elevation gave the best results. The partial spline was found to have an added advantage in that there is no need to determine the most appropriate scale for elevation and a single lapse rate is calculated automatically.
- The interpolation accuracy for the pressure surfaces achieved with the partial spline models was found to be similar to that obtained by incorporation of a spatially varying dependence on elevation. The results from the temperature surfaces showed similar results.
- While it is possible to impose different weighting strategies when spatially interpolating climate data it was found that the uniform weighting performed the best for the pressure and temperature long term averages.
- An evaluation of the ANUSPLIN computer program's ability to represent meteorologically realistic daily pressure patterns and rainfall decile maps showed that the output was realistic. The method was shown to be statistically very sensitive to spatial trends. It also allowed for the investigation of outliers in the output statistics and for an analysis of the error. This then enables sensitivity studies and allows for the assessment of station networks.

The development of a monthly 'primitive' surfaces for Australian climate

- The spline technique was found to be an iterative procedure in climate surface construction which provided valuable feedback about the meteorological station database. Examination of the summary statistics and the data residuals was a powerful aid for detecting and correcting the data errors.
- The technique of thin plate smoothing splines and its extensions was found to provide insight into the nature of the spatial variation of climatological variables.
- The major factor in improving the accuracy and spatial resolution of the interpolated pressure and temperature surfaces was the incorporation of elevation as the third independent variable.
- The error grids associated with the climate surfaces and of the output statistics generated in the production of the surfaces contributed significantly to a better understanding of interpolation and network deficiencies.
- If the existing data banks and coverage of the Australian mainland were to be improved, the error surfaces suggest that a greater number of stations in higher

altitude areas would best improve the accuracy of the spatial interpolation process, especially when deriving climatic variables at station level.

- The addition of the extra stations would do very little to improve the prediction of broad-scale geographic patterns, but could significantly improve the ability to extract the surface conditions and this was reflected in the lower covariate standard error.
- An assessment of the environmental lapse rates determined from the spline functions showed that the incorporation of elevation gave reasonable results, although the environmental lapse rates can differ markedly from adiabatic lapse rates.
- The long-term monthly averages and the problems associated with incomplete time series and ways of overcoming these problems were investigated. Uniform weighting was shown to be the most appropriate method for fitting SPLINA to long term means of pressure and temperature.
- The output statistics and the graphics of the surfaces suggest that the 'primitive' surfaces are a faithful representation of the climatic conditions prevailing at the time and are similar to previously produced maps.

The production of gridded estimates of derived climatic variables by applying logical models to the primary climate variables.

- Logical models were used with the set of 'primitives' to compute attributes for the derived surfaces. Logical models can be created for any specific time already represented in the database and the set of primitives can be used to create new values from existing data and so extend the database.
- The use of the SLPGRD and ASPGRD computer programs was shown to be extremely useful for determining the vector quantities of the surfaces such as the slope and aspect. The pressure aspect maps clearly show the synoptic patterns and features such as the ridge of high pressure which extends across the continent in winter. The maps of the aspect of the dry-bulb temperature surfaces show that the winter limit of ocean influence can vary greatly.

The investigation of the points of a 2.5 x 2.5 degree grid of estimates of climatic variables for an assessment of climate variability and change.

- The grid point time series displayed some interesting trends. The increase found in the 0000 UTC dry-bulb temperature (~0.7°C) over the period 1952 to 1990, if a true indication of the temperature trend, is alarming. This upward trend in the dry-bulb temperatures is however, not reflected in the pressure, wet-bulb temperature, dew-point temperature, rainfall percentiles or the SOI, all of which have cyclic patterns.
- The correlation between the wet bulb depression and the dew-point temperature showed two distinct trend lines. These two distinct periods are separated by the very wet 1973-1974 period which suggest that a switch in atmospheric states may have occurred in this period.

- The Australian mean rainfall percentiles were drier (below 0.5) in the first half of the period than in the second half and this also appears to be related to the very wet years in 1973 and 1974. This period also coincides with a change in the relationship between the rainfall percentiles and the SOI.

The investigation of mapping techniques for displaying the spatio-temporal database.

- An investigation of methods used to dealing with multiple spatial representations found that graphics, multiple editions and animation were all invaluable tools for analyzing large amounts of data.
- Sound was suggested as a way to expand the limited possibilities of representing multi-variate data using graphics and to offer a way to enhance the comprehension of patterns in time. Sound can provide more choices for representing and communicating ideas and can serve as a valuable addition to visual displays.

The spatial interpolation of the correlation parameters and the regression models

- Bivariate thin plate smoothing splines were used to extend the regression relationships estimated from grid points taken from the MSL surface coefficient files to a multi-site framework. The dew point temperature regression Model A showed that the extension of the correlation parameters directly address the spatial nature of climatic relationships.
- The simpler dew-point regression Model B showed that the dew point temperature distribution could be effectively related to fixed temperature values and a spatially varying constant with very little reduction in the explanation of the percentage of the variance between the actual and modeled values.
- The dependencies of the temperature values on the larger scale atmospheric conditions showed some interesting seasonal differences. The models appear to have merit and that there is some validity in the assumption that fluctuations in the atmospheric circulation are a critical determinant of the surface environment for a given period.
- All the rainfall percentile models were found to provide only reasonable fits to monthly rainfall percentile distributions. This is mainly due to the difficulties in identifying the components of the dependency of the spatial variation of monthly rainfall percentiles. Due to the different rain-producing mechanisms effecting the different parts of the continent, the spatially varying models (a) perform better and in some cases significantly better than the simpler (b) models.
- The lag regression models perform reasonably well, but not as well as suggested by the percentage variance explained in the Genstat procedure. The lag regression models tend towards the long term average if the previous three months were "average" or give mixed signals. If there are strong trends in the same direction during the previous three months the model does perform well. Overall, the

output results are encouraging. The reliability of the surfaces and the spatial equations was tested on a set of observations withheld from the regression estimation procedures.

The production of the dew-point temperature space-time spline model

- A new type of spatio-temporal model for the analysis of climate variables which takes into account the natural physically-based correlations in time and space was proposed. The spatio-temporal spline model depicts processes of two-dimensional space that are played out along a third temporal dimension. Two different representations of the time ('actual' and 'event') dimension in the spatio-temporal spline models were examined. A particular attraction of the model is the ease with which a mean sea level distribution can be put into a multi-site framework suitable for space-time studies.
- In Model 1, 'actual' time is represented as an equispaced linear series in which the monthly anomalies from the 'primitive' surfaces were fitted at unit intervals. The scale factor for the unit intervals was shown to be critical to the performance of the model. A seasonal scaling was shown to be the best option for the validation data, but only marginally better than any particular fixed scale factor.
- Model 2A, which used a mapping of the long term average to represent an 'event' time, allows for the incorporation of the cyclic nature of climate. While Model 2A performed better than Model 1 for some months it gave no improvement in the average rms. Model 2A had similar problems to Model 1 with regard to the choice of the 'appropriate' scale factor for the 'event' time dimension. While the I*15 scale factor model showed the lowest validation rms, it was considerably higher than could be achieved by the optimum ('best') scale factor for each month. This suggested that the representation of the time dimension by a fixed Australian long term averages is not a realistic option for mapping of the time dimension.
- Model 2B, which used a mapping of the local long term average to represent an 'event' time, allows for the incorporation of spatially variations in the natural seasonal cycles. Model 2B uses a spatially varying long term average transformation to represent a 'event' time. Several tests were done for this model and it was shown that a second order spline fitted to three independent variables gave the most realistic results.
- Model 2B, however, produced some unexpected results in the models fitted with higher order derivatives, which suggested that the space-time spline model was not well defined in the time domain. That is, the function was modelling very little of the temporal richness, whilst heavily representing the spatial distribution. The incorporation of additional covariates (i^2 and i^3) did contribute significantly to interpolation accuracy of the 3rd and 4th order fits.
- It was shown that the incorporation of a spatially varying dependence on appropriately scaled time made a dominant contribution to surface accuracy. The results of fitting the spline and partial spline showed that better performances were achieved with time as an independent variable. This suggested that there is both spatial and temporal dependence in the distribution of the climate variables.

- Thin plate smoothing splines also permitted straightforward incorporation of varying degrees of topographic/temporal dependence. This scaling of the dimensions can normally be slightly refined by choosing the elevation scaling to minimise the GCV of the fitted surface. In the case of the all the space-time models, however, a reduced GCV did not necessarily show improved results in the extrapolated surfaces.
- The selection of the pseudo-time sale by use of the hindcast extraction was shown to be a possible way of circumventing the temporal scale factor problem for model 2B. A comparison of the output of spatio-temporal Models 1 and 2 with the output of the lag regression model showed that the spatio-temporal models performed better than the lag regression model. The approach of fitting space and time simultaneously appears to hold promise of some useful outputs for monthly forecasting from limited data.

Future research directions

Aspects of this methodology and issues that warrant further research are discussed below.

The data sets

Changes in technology have meant that many organisations are now undertaking some form of spatial interpolation of climate data. These organisations still rely on the availability of quality controlled network of stations. Karl et al (1993) and Nicholls (1997) observe that virtually every monitoring system and data set require better data quality, continuity, and homogeneity if we expect to conclusively answer questions of interest to both scientists and policy makers. The continued maintenance and upgrade of conventional surface-based observing systems is vital to sustaining present capabilities into the future.

Analysis of the errors resulting from spatial interpolation provides information about the strengths and weaknesses of the historical station networks. Moreover, the standard error estimates associated with the fitted surfaces can indicate areas where the network is inadequate and allow strategies for efficient network design to be developed.

Interpolation issues

Central to the development of space-time models is the accurate spatial interpolation of climate data from standard meteorological networks. By incorporating a spatially varying dependence on elevation, the thin plate smoothing splines offer an elegant solution to the problem for the temperature and pressure data. The incorporation of additional variables (more sophisticated topographic effects or proximity to large

water bodies) in the interpolation could well improve spatial accuracy in regions with sparse data.

The rainfall percentile maps were shown to be faithful representations, but were not particularly useful for modelling purposes. Possibilities exist for different transformations of the rainfall data rather than deciles. The incorporation of additional variables (aspect in relation to prevailing winds and distance from the coast) may also be useful.

Climate Change and variability

Public and political interest in climate has never been higher. Due to this heightened public interest in climate, there is much research which is being conducted on climate change and variability in the country. However, the distinct shift that appears to have occurred in the moisture regime during the 1973/1974 period warrants further investigation.

The El Niño/Southern Oscillation phenomenon accounts for much of the interannual variability of the Australian climate. Relationships with this phenomenon have largely been ignored in this study but it is likely to have some influence on the regional time series.

The application of advanced graphical techniques

One problem that researchers constantly face in attempting to visualize complex systems is that they run out of visualization techniques since there are generally more parameters than there are distinct ways of collectively showing the data. The concept of a paper map is static and limited-due to physical limitations. In a GIS, one can interact with the data-zoom in for more detail if it exists, shift (or pan), or change scales if the resolution of the data allows. There are no currently available commercial programs which allow for easy access to spatio-temporal data.

Graphics present the entire picture simultaneously (sometimes making it difficult to discover hidden patterns), where as sound presents the picture linearly-similar to text and navigation. The musical sequences were found to be effective for enhancing the display of spatial data. The fact that the music sounds reasonable may be more due to the transformation used rather than the sequence itself. Other sound types or scales or transformations could be more effective in conveying the intended message.

A solid body of knowledge detailing the sound perception capabilities of the human physiological system presently exists. This knowledge can be used to underpin our understanding of the possibilities and the limitations of the sound variables as visual design elements. More difficult issues of identification , problem solving , judgment, remembering and understanding of the sound displays await attention. The sequential nature of sound raises questions of knowledge acquisition and memory.

We must also be aware of the problem of "sonic overload", of barraging the user with too many different variables and dimensions of sound. There are also question of how much information people can deal with. The combination of the visual and sonic may be the way to go, however there are few precedents and may be more confusing than enlightening. Like the visual there is only so much information that we can take in, and a discordant orchestra is as confusing as spatial images.

The animations with sound were found to be quite different at different speeds. Further investigation is required in order to determine the most appropriate speeds for both the pictures and the sound. This also raises the question of which factors distort space-time judgments and how does one determine this?

Climatic Models

The overlay of two registered images to produce a third image which is a logical or arithmetic function of these is a basic operation in a GIS, and is much employed in the empirical modelling of environmental processes. The wet-bulb temperature and dew-point temperature aspect and slope require further work to determine the value and optimum utilisation of the information. The maps of the slope of the temperature surfaces show areas of greatest temperature gradient; these areas may be the regions of greatest sensitivity to climate change. This possible sensitivity, however, requires further investigation which might be achieved better at a finer scale than at the continental scale.

The spatial coherence of the regression coefficients is a promising aspect in terms of spatial modelling. The further development of quantitative links between the climatic variables may provide further insight on the relative influences of local and synoptic-scale processes. Since these models have been produced at specific scales for the Australian continent, are these models scale or domain dependent? More research on the sensitivity to longer term climate change of the parameters of the various stochastic models is required.

Space-time spline model

The objectives of creating the space-time spline model were to build a computational framework, within which geographic phenomena and processes, and their temporal changes, can be simulated. This included extending the methods for dealing with the temporal dimension and to explore alternative mathematical formalisations which represent spatial and temporal reasoning processes better. Further studies related to this work may be worth pursuing.

The spatial spline model is based on the premise that the spatial and temporal dimensions operate in the same modelling space. Some links between spatial and

temporal scales are clear, while other links require the same process to be studied at different spatial and temporal scales.

A shortcoming of the monthly model is that the monthly time scale ignores important shorter time scale dynamics. The models were run at specific spatial and temporal scales and take no account of whether processes have natural/characteristic spatial and temporal scales or whether processes examined at one temporal or spatial scale correlate or describe processes at another temporal or spatial scale. To answer these questions, studies that examine data at different granularities (i.e., scales) in space and in time or using different climate variables (e.g. pressure, wet-bulb) are needed.

It is recognised that the state of the atmosphere is effected by processes and perturbations operating at a range of spatial and temporal scales. The observation window defined by this study examines only a subset of all possible perturbations to which the climate responds. Perhaps a better pseudo time could be achieved by using shorter long term means. The analysis should be repeated using shorter period long term averages especially considering the shift in atmospheric states observed between the pre- ad post- 73-74 wet period. over a full range of temporal scales, under prevailing and changing climatic conditions.

As higher-dimensional data are often sparse in one or more dimensions, there is a need to investigate the use of various effects to show the quality and nature of the data. What error models should be considered for the space-time model, and how should error, uncertainty, and imprecision about time and space be conveyed to a user? How do people resolve incomplete or insufficient spatial and temporal knowledge and how can missing information be supplied and flagged to the user?

Potential applications of methodology to climatological and environmental issues

The methodology is sufficiently developed to be applied to other space and time scales, even though benefits would flow from further research and development. While the results of this study are based on empirical analysis (i.e. statistical relations between climatic variables) the methodology is generic to the extent that the basic steps can be applied to any region. The spatial modelling techniques are entirely portable, and a new relations and predictive functions could be derived as long as suitable data were available. The question does arise of the generality of the modelling technique per se in areas of extreme climate or of predominantly different rain producing mechanisms or areas with a much greater variation in elevation.

Comparison with data from other climate studies is difficult primarily because climatic records are not usually dealt with on a spatial scale and less frequently are mean sea level temperature surfaces produced. Adiabatic lapse rates rather than environmental lapse rates are usually calculated. Environmental lapse over the entire

continent may be associated by false hills and valleys created in the temperature surfaces, or may be due to a continentality factor.

Given the above limitations, the following discussion addresses some potential applications of the methodology.

Climate Impact Assessment

Existing and predicted levels of greenhouse gases are expected to result in significant and relatively rapid changes to global climatic regimes. Likely scenarios for Australia, presented by Pittock (1988) and Whetton et al (1996), suggest temperature increases of 2-4 degrees Celsius, and changes in precipitation intensity and seasonality. The results of the temperature surface averages suggest that the temperature increase is already underway.

The methodology presented here could make a contribution at a number of levels to the general problems of assessing climate change. A useful starting point would be to maintain the surface database for continual monitoring of the continental climatic variables. A valuable contribution may be achieved with the inclusion of offshore data.

The climate surfaces could be used as a validation set for the GCMs. A related approach is to perturb the stochastic model parameters in line with predicted changes in monthly means and variances of temperature and rainfall. However, unless the stochastic models are simply parameterised, such approaches have to make a large number of unverifiable assumptions about which aspects of the model are to remain constant.

Long term climatic predictions.

The simple parameterisations of the dew-point temperature regression models make these models an attractive choice. The usefulness of generating realistic spatially detailed climate scenarios depends on the observed links between the model parameters remaining constant in changed climates.

Provided that the problems in accounting for the complicated structure of the transformed time variable can be overcome, it is possible that the space time cube could be useful for forecasting spatial and temporal distributions of climate variables at the monthly time step. These models could be perturbed according to broad scale climate change scenarios produced by general circulation models. This also depends on the adequacy of the long term mean used in the model in describing the underlying trends.

Conservation of climatically sensitive environmental regions.

Natural environments are facing increasing pressures from economic, social and environmental factors. The techniques described here allow for the rapid revision of

the surfaces as new climatological data are entered into the station database. In combination with a GIS, questions about the spatial extent of natural resource management problems can be efficiently addressed.

The computer-based methods used in this study significantly improve the accuracy with which spatial estimates of the climatic variables can be made. The capacity to estimate long term monthly mean climate in a landscape is particularly important as climate will account for a significant percentage of the variability in most ecological data.

Conclusions

Central to the development of space-time models is the accurate spatial interpolation of model parameters from standard meteorological networks. By incorporating a parametric sub-model for a linear dependence on elevation, partial thin plate smoothing splines offer an objective flexible tool and an elegant and simple solution. The iterative process of fitting the spline surfaces offers an excellent tool to evaluate the integrity of the station database. Monthly temperature and pressure surfaces were created using the spline techniques for Australia. This database was then used as a base for the mean sea level dew point temperature models.

The dew-point temperature regression model was shown to be useful and could be used as an alternative to the psychometric tables. While the lag regression models showed encouraging statistics they did not perform as well as the space-time spline models. It is recognised that the models are based on processes and perturbations operating at a specific range of spatial and temporal scales. The observation window defined by this study examines only a subset of all possible perturbations to which the climate responds. While the problems of the scale of the temporal dimension remain the results show much promise.

The focus of this thesis has been concerned with the development of space-time models of dew point temperature. Variables such as pressure, dry-bulb and wet-bulb temperature appear to have similar simple space-time structures, and this suggests that the models would work equally well for these variables. The approach of describing time in terms of local long term mean distributions appears to hold promise, provided significant complexities in the temporal domain can be accommodated. The space-time models and the techniques used allow for the rapid revision of the surfaces as new data are entered into the station database. The models could be perturbed according to broad scale climate change scenarios produced by general circulation models. Monthly climate model structures can also be used to guide the development of models at finer time scales. Overall, however, the space-time space model is shown to be a promising candidate for space-time simulation of climatological variables at the monthly time scale if the

As climate accounts for a significant percentage of the variability in most ecological data, it is therefore useful to produce a climate space-time model that generates monthly climate data. Importantly, the models can be calibrated in terms of observed monthly summary statistics whose spatial distribution can be well determined from the standard meteorological network using existing thin plate smoothing spline techniques. Once interpolated and tested these models are then in a convenient form for future use. However, since the statistical approach depends on historical data, it is obviously limited to predicting climates that have been observed or are not caused by new processes.

The techniques covered in this thesis provide additional means of addressing the problems of the spatial extension of limited site observations, an understanding of the physical determinants of several climatic variables, and an extension of the spatial model to a predictive model.

Bibliography

and Appendices

Bibliography

Akerman D. (ed) 1990. *A Natural History of the senses*, Random House, New York.

Akima H. 1978. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Transactions of Mathematical Software* 4:148-159.

Allan R.J. 1988. El Nino southern oscillation influences in the Australiasian region, *Progress in Physical Geography* 12:4-40.

Allan R.J. and Haylock M.R. 1993. Circulation features associated with the winter rainfall decrease in southwestern Australia, in *Journal of Climate* 6:1356-1367.

Allan R.J., Lindesay J. and Parker D. 1996. El Niño, Southern Oscillation and climatic variability. CSIRO Publishing, Collingwood, Vic.

Bardossy A. and Plate E.J. 1992. Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.* 28:1247-1259.

Barry R.G. and Chorley R.J. 1987. *Atmosphere, Weather and Climate*, (5^{th} edition), Routledge, London.

Barry R.G. and Perry A.H. 1973. *Synoptic climatology: methods and applications*, Methuen, London.

Bates D. and Wahba G. 1982. Computational methods for generalised cross-validation with large data sets. In C.Baker and G.Miller (eds), *Treatment of Integral Equations by Numerical Methods*, Academic Press, London.

Bates D., Lindstrom M., Wahba G. and Yandell B. 1987. GCVPACK - routines for generalized cross validation. *Communications in Statistics B - Simulation and Computation* 16, 263-297.

Bennett R.J. 1981. Statistical Forecating. *Concepts and techniques in modern geography CATMOG 28.*

Bonacci O. 1993. Hydrological identification of drought. *Hydrological Processes* 7:249-262.

Booth T.H., Nix H.A., Hutchinson M.F. and Jovanovic T. 1987. Niche analysis and tree species introduction. *Forest Ecology and Management* 23:47-59.

Bradley R.S. and Jones P.D. 1985. Data bases for detecting $C0_2$ -induced climate change. In *US Department of Energy State of the Art Report on the detection of climatic change*, US Department of Energy Carbon Dioxide Research Division, Washington DC.

Bradley R.S., Kelly P.M., Jones P.D., Diaz H.F. and Goodess C. 1985. A climatic data bank for the Northern Hemisphere land areas, 1851-1980. *DoE Technical Report No TR017*, US Department of Energy Carbon Dioxide Research Division, Washington DC.

Brook R.R. 1982. A study of the subtropical jet stream in the Australian region. *Australian Meteorological Magazine*. 30:223-239.

Bureau of Meteorology no date. *Data held in computer compatible form.* Australian Government Department of Supply , Maribyrnong, Victoria, 103 pp.

Bureau of Meteorology (Australia). 1968-1989. *Monthly Weather Review, NSW.* Various dates between Jan 1968 and Dec 1989. AGPS. Canberra.

Bureau of Meteorology (Australia) 1977. *Manual of Meteorology, Part 1. General Meteorology*. AGPS. Canberra.

Bureau of Meteorology (Australia) 1988. *Climatic Averages of Australia*. AGPS. Canberra.

Bureau of Meteorology (NCC) 1990a. *Summary of climate data held in the National Climate Centre computer archive.* National Climate Centre, Melbourne.

Bureau of Meteorology (NCC) 1990b. *Station dictionary,* National Climate Centre, Melbourne, (microfiche).

Chagnon S.A. 1992. Inadvertent weather modifications in urban areas; lessons for global climate, *Bulletin of the American Meteorological Society* 73:619-627.

Chatfield C. 1989. *The analysis of time series, an introduction* (4th edition). Chapman and Hall, London.

Chua S.-H. and Bras R.L. 1982. Optimal estimators of mean areal precipitation in regions of orographic influence. *Journal of Hydrology* 57:23-48.

Clark W.C. 1985. Scales of climatic impacts, *Climatic change* 7:5-27.

Colls K. and Whitaker R. 1990. *The Australian Weather Book*, Child and Associates, Sydney.

Coughlan M.J. 1979. Recent variations in annual-mean maximum temperatures over Australia, in *Quart. J. R. Meteorol. Soc.* 105:707-719.

Coughlan M.J. 1983. A comparative climatology of blocking action in the two hemispheres. *Australian Meteorological Magazine* 31:3-13.

Coughlan M.J., Tapp R. and Kininmonth W.R. 1990. Trends in Australian temperature records, in *Observed climate variations and change: Contributions in support of Section 7 of the 1990 IPCC Scientific Assessment*, edited by D.E. Parker, IPCC/WMO/UNEP, Chap.3, 1-28.

Cressman G. 1959. An operational objective analysis system, in *Monthly Weather Review* 87:367-374.

Cressie N.A.C. 1991. *Statistics for spatial data.* John Wiley & Sons, New York.

Crowder B. 1995. *The wonders of the weather* AGPS, Canberra.

Daly C., Neilson R.P. and Phillips D.L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140-158.

Delfiner P. and Delhomme J.P. 1975. Optimum interpolation by kriging. In J.D.Davis and M.J.McCullagh (eds), *Display and Analysis of Spatial Data.* John Wiley & Sons, New York, pp 96-114.

Dietrich C.R. and Osborne M.R. 1991. Estimation of covariance parameters in kriging via restricted maximum likelihood. *Mathematical Geology* 23, 119-135.

Eastman J.R. 1995. *IDRISI for windows, reference manual*, Version 1.0, Clark University, Worcester, Massachusetts.

Edwards K.A. 1972. Estimating areal rainfall by fitting surfaces to irregularly spaced data. *Proc. International Symposium on the Distribution of Precipitation in Mountainous Areas 2,* World Meteorological Organisation, No. 326:565-587.

Eldén L. 1984. A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems. *BIT* 24:467-472.

Evans J.L. and Allan R.J. 1992. El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region. *J. Clim. Appl. Climatology.* 12:611-623.

Flohn H. 1981. Climatic variability and coherence in time and space, in W. Bach, J. Pankrath and S.H. Schneider (eds) *Food-climate interactions,* D.Reidel Publishing Company, Dordrecht, The Netherlands.

Folland C.K. 1988. Numerical models of the rain gauge exposure problem, field experiments and an improved collector design, in *Quart. J. R. Meteorol. Soc.* 114:1485-1516.

Gates W.L., Mitchell J.F.B., Boer G.J., Cubasch U., Meleshko V.P. 1992. Climate modelling, climate prediction and model validation. In Houghton,J.T., Callander B.A. and Varney S.K. (eds) *Climate Change 1992, The Supplementary Report to the IPCC Scientific Assessment.* Cambridge University Press, 97-134.

Gaver W. 1989. The sonic finder: an interface that uses auditory icons, in *Human-Computer Interaction* 4:4:67-94.

Gedzelman S.D. 1985. Atmospheric circulation systems, in D.D. Houghton (ed) *Handbook of applied meteorology*, John Wiley and Sons, New York.

Genstat 5 Committee of the Statistics Department, Rothamsted Eperimental Station 1993. *Genstat 5 Release 3 Reference Manual.* Clarendon Press, Oxford.

Gentilli J.(ed.) 1971. *Climates of Australia and New Zealand*, World Survey of Climatology Vol.13. Elsevier Publishing Co., Amsterdam.

Gibbs W.J. 1975. Drought - its definition, delineation and effects. WMO Special Environmental Report No.5, World Meteorological Organization, Geneva, 1-39.

Gibbs W.J. and Maher J.V. 1967. *Rainfall deciles as drought indicators.* Bulletin 48. Australian Bureau of Meteorology, Melbourne.

Goodall C. and Mardia K.V. 1994. Challenges in Multivariate Spatio-Temporal Modelling, in *Proceedings for International Biometrics Conference*, Hamilton, Canada.

Gu C. and Wahba G. 1993. Semiparametric analysis of variance with tensor product thin plate splines. *Journal of the Royal Statistical Society, Series B* 55:353-368.

Gumbel E.J. 1954. *Statistical theory of extreme values and some practical applications*, Applied mathematics series 33 (US National Bureau Standards).

Hamlin M.J. 1983. The significance of rainfall in the study of hydrological processes at basin scale. *J. Hydrol.* 65:73-94.

Hammond World Atlas 1978. *Hammond World Atlas*, Hammond Press, USA.

Harvey D. 1969. *Explanation in Geography*. St Martin's Press, New York.

Hay L.E., McCabe Jr.,G.J., Wolock D.M. and Ayers M.A. 1991. Simulation of precipitation by weather type analysis. *Water Resour. Res.* 27:493-501.

Hevesi J.A., Istok J.D. and Flint A.L. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: Structural analysis. *Journal of Applied Meteorology* 31:661-676.

Houghton J.T., Jenkins G.J., Ephraums J.J (eds) 1990. *Climate Change: the IPCC scientific assessment*. Cambridge University Press, Cambridge.

Hughes D.A. 1982. The relationship between mean annual rainfall and physiographic variables applied to the coastal region of southern Africa. *South African Geographic Journal*, 64:41-50.

Hungerford R.D., Nemani R.R., Running S.W. and Coughlan J.C. 1989. MTCLIM: a mountain microclimate simulation model. Res. Pap. INT-414. Ogden, UT, U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 52pp.

Hutchinson M.F. 1984. A summary of some surface fitting and contouring programs for noisy data, CSIRO Division of Land Use Research and Mathematics and Statistics, *Consulting Report No. ACT 84/6*, CSIRO, Canberra.

Hutchinson M.F. 1987. Methods of generating weather variables. In A.H.Bunting (ed), *Agricultural Environments: Characterisation, Classification and Mapping*, Wallingford, CAB International, pp 149-157.

Hutchinson M.F. 1989. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Statist. -Simul.* 18:1059-1076.

Hutchinson M.F. 1991a. Continent-wide data assimilation using thin plate smoothing splines. In J.D.Jasper (ed), *Data Assimilation Systems*. BMRC Research Report No.27, Melbourne, Bureau of Meteorology, pp 104-113.

Hutchinson M.F. 1991b. Climatic analyses in data sparse regions. In R.C.Muchow and J.A.Bellamy (eds), "Climatic Risk in Crop Production", CAB International, Wallingford, 55-71.

Hutchinson M.F. 1993a. On thin plate splines and kriging. In Tarter M.E. and Lock M.D. (eds), *Computing and Science in Statistics, Vol. 25*. Interface Foundation of North America, University of California, Berkeley, 55-62.

Hutchinson M.F. 1993b. Development of a continent-wide DEM with applications to terrain and climate analysis. In M.F.Goodchild, B.O.Parks and L.T.Steyaert (eds), *Environmental Modeling with GIS*. Oxford University Press, pp 392-399.

Hutchinson M.F. 1994. Interpolating rainfall means - getting the temporal statistics correct. *Proc. Second Inter. Conference/Workshop on GIS and Environmental Modelling*. Breckenridge, Colorado, Sept 1993.

Hutchinson M.F. 1995a. Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geog. Inf. Sys.* 9:4:385-403.

Hutchinson M.F. 1995b. Stochastic space-time weather models from ground-based data, in *Agricultural and Forest Meteorology* 73:237-264.

Hutchinson M.F. and Bischof R.J. 1983. A new method for estimating mean seasonal and annual rainfall for the Hunter Valley, New South Wales. *Australian Meteorological Magazine* 31:179-184.

Hutchinson M.F. and Corbett J.D. 1995. Spatial interpolation of climate data using thin plate smoothing splines. In co-ordination and harmonisaton of databases and software for agroclimatic applications, *Agrometeorology Series Working Paper, No 13*. Proceedings of an expert consultation held in FAO Rome, Italy, from 29 November - 3 December, 1993.

Hutchinson M.F. and Gessler P.E. 1994. Splines - more than just a smooth interpolator, in *Geoderma* 62:45-67.

Hutchinson M.F., Kalma J.D. and Johnson M.E. 1984. Monthly estimates of wind speed and wind run for Australia, *Journal of Climatology*, 4:311-324.

Hutchinson M.F., Nix H.A. and McMahon J.P. 1992 Climate constraints on cropping systems. In C.J.Pearson (ed), *Ecosystems of the World: Field Crop Ecosystems*, Amsterdam, Elsevier, pp 37-58.

Hutchinson M.F., Nix H.A., McMahon J.P. and Ord K.D. 1993. Geographic information systems for Africa. Centre for Resource and Environmental Studies Annual Report, Australian National University, Canberra, pp 2-3.

Jenkins G.M. and Watts D.G. 1968. *Spectral analysis and its applications*. Holden-Day, San Francisco.

Jet Propulsion Laboratories 1987. LA the movie, Computer animated videotape. Jet propulsion laboratories, Pasedena.

Johnson K. 1992. *The AUSMAP Atlas of Australia*, Cambridge University Press, Victoria.

Jones P.A. 1991. Historical records of cloud cover and climate for Australia, in *Australian Meteorological Magazine* 39:181-189.

Jones P.D. 1991. Southern Hemisphere sea-level pressure data: an analysis and reconstructions back to 1951 and 1911, in *International Journal of Climatology* 11:585-607.

Jones P.D. 1994. Hemispheric surface air temperature variations: A Reanalysis and an update to 1993, in *Journal of Climate* 7:1794-1802.

Jones P.D., Raper S.C.B., Santer B., Cherry B.S.G., Goodess C., Kelly P.M., Wigley T.M.L., Bradley R.S., and Diaz H.F. 1985. A grid point surface air temperature data set for the Northern Hemisphere. *DoE Technical Report No 22*, US Department of Energy Carbon Dioxide Research Division, Washington DC, 251pp.

Jones P.D., Raper S.C.B. and Wigley T.M.L. 1986. Southern Hemisphere surface air temperature variations 1851-1984, in *Jnl Clim. Appl. Met.* 25:1213-1230.

Kane R.P. 1997. On the relationship of ENSO with rainfall over different parts of Australia, in *Australian Meteorological Magazine* 46:39-49.

Karelsky S. 1954. *Surface circulation in the Australian region.* Meteorology Study No.3, Bureau of Meteorology (Australia), Melbourne.

Karelsky S. 1956. *Classification of the surface circulation in the Australasian region.* Meteorology Study No.8, Bureau of Meteorology (Australia), Melbourne.

Karelsky S. 1961. *Monthly and seasonal anticyclonicity and cyclonicity in the Australian region.* Meteorology Study No.13, Bureau of Meteorology (Australia), Melbourne.

Karelsky S. 1965. *Monthly geographical distributions of central pressures in surface highs and lows in the Australian region, 1952-1963.* Bureau of Meteorology (Australia), Melbourne.

Karl T.R., Jones P.D., Knight R.W., Kukla G., Plummer N., Razuvayev V., Gallo K.P., Lindesay J., Charlson R.J. and Peterson T.C. 1993. A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature, in *Bulletin of the American Meteorological Society* 74:1007-1023.

Karoly D.J., Hope P. and Jones P.D. 1996. Decadal variations of the southern hemisphere circulation, in *International Journal of Climatology* 16:723-738.

Kesteven J.L. and Hutchinson M.F. 1996. Spatial Modelling of Climatic Variables on a Continental Scale, a paper presented to the Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling, Jan 21-25, 1996, Santa Fe, New Mexico, USA.

Kidson J.W. 1988. Indices of the Southern Hemisphere zonal wind, in *Journal of Climate*, 1:183-194.

Kidson J.W. 1991. Intraseasonal variations in the Southern Hemisphere circulation, in *Journal of Climate*, 4:939-953.

Klein W.H. and Bloom H.J. 1987. Specification of monthly precipitation over the United States from the surrounding 700 mb height field. *Mon. Weath. Rev.* 115:2118-2132.

Kong C.-W. 1995. Diurnal pressure variations over continental Australia, in *Aust. Met. Mag.* 44:165-175.

Krygier J.B. 1994. Sound and geographic Visualization, in MacEachren and Taylor (eds) *Visualization in modern cartography*. Elsevier Science Ltd., Oxford.

Lam N.S. 1983. Spatial interpolation methods: a review. *The American Cartographer* 10:129-149.

Landsberg H.E. 1981. *The urban climate*. Academic Press, New York.

Langran G. 1989. A review of temporal database research and its use in GIS applications, in *International Journal of Geographical Information Systems* 3:215-232.

Langran G. 1992. *Time in Geographic Information Systems*. Taylor and Francis, London.

Langran G. 1993. Issues of implementing spatiotemporal system, in *International Journal of Geographical Information Systems* 7:305-314.

Larson J.W., Taylor J.A., Kesteven J.L, Hutchinson M.F, Bates G.T. and Oglesby R.J. 1994. Regional-scale climate studies of Australia using RegCM2, a paper presented to the International Congress on Modelling and Simulation, Newcastle, November 27-30, 1995.

Laslett G.M. 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association* 89:391-409.

Laslett G.M., McBratney A.B., Pahl P.J. and Hutchinson M.F. 1987. Comparison of several spatial prediction methods for soil pH. J. Soil Science 38, 325-341.

Laughlin G. 1997. *The user's guide to the Australia coast*, New Holland, Sydney.

Lavery B., Joung G. and Nicholls N. 1997. An extended high-quality historical rainfall data set for Australia, in *Australian Meteorological Magazine* 46:27-38.

Leighton R.M. 1994. Relationship of Anomalies of (Anti)cyclonicity to some Significant Weather Events over the Australian Region, in *Australian Meteorological Magazine* 43:255-261.

Leighton R.M. and Deslandes R. 1991a. Monthly anticyclonicity and cyclonicity in the Australian region: 23 year (1965-1987) averages. *Technical Report No 64*, Bureau of Meteorology, Australia, 29pp.

Leighton R.M. and Deslandes R. 1991b. Monthly anticyclonicity and cyclonicity in the Australian region: averages for January, April, July and October. *Australian meteorological magazine* 39:159-154.

Lejenas H. 1984. Characteristics of Southern Hemisphere blocking as determined from a time series of observational data. *Quarterly Journal Royal Meteorological Society* 110:967-979.

Lewis A. 1995. Scale and scale problems in spatial environmental databases, unpublished PhD Thesis, CRES, ANU, Canberra.

Linacre E. 1992. *Climate data and resources: a reference and guide*, Routledge, London.

Linacre E. and Hobbs J. 1977. *The Australian climatic environment.* John Wiley and Sons, Brisbane, 354pp.

Loader C. and Switzer P. 1992. Spatial covariance estimation for monitoring data. In P. Guttorp and A.Walden (eds), *Statistics in environmental and earth sciences*, pp52-69. Charles W. Griffin, London.

Luo Z. and Wahba G. 1997. Hybrid adaptive splines, in Journal of the American Statistical Association, 92:437:107-116.

Lyons S.W. 1990. Spatial and temporal variability of monthly precipitation in Texas. *Mon. Weath. Rev.* 118:2634-2648.

MacEachren A.M. 1995. *How maps work, Representation, visualization and design.* The Guilford Press, NY.

McIlveen R. 1992. *Fundementals of weather and climate*, Chapman and Hill, London.

Mackey B.G. 1993. A Spatial analysis of the environmental relations of rainforest structural types. *Journal of Biogeography* 20:303-336.

Mackey B.G., Nix H.A., Stein J.A. and Cork S.E. 1989. Assessing the representativeness of the wet tropics of Queensland World heritage property. *Biological Conservation* 50:279-303.

McMahon J.P., Hutchinson M.F., Nix H.A. and Ord K.D. 1995. ANUCLIM users guide, unpublished, Centre for Resource and Environmental Studies, Australian National University, Canberra.

Mansur D., Blattner M. and Jot K. 1985. Sound graphs: a numerical data analysis method for the blind, in *Journal of Medical Systems*, 9:3:163-174.

Mardia K.V. and Goodall C.R. 1993. Spatial-temporal analysis of multivariate environmental monitoring data. In G.P.Patil and C.R.Rao (eds), *Multivariate environmental statistics.* Elsevier.

Mason D.C., O'Conaill M.A. and Bell S.B.M. 1994. Handling four-dimensional geo-referenced data in environmental GIS, in *Int. J. Geographical Information Systems*, 8:2:191-215.

Matalas N.C. 1967. Mathematical assessment of synthetic hydrology. *Water Resour. Res.* 3:937-945.

Meentemeyer V. 1989. Geographical Perspectives of Space, Time, and Scale, in *Landscape Ecology* 3:3/4:163-173.

Mitasova H. and Mitas L. 1993. Interpolation by regularized spline with tension: I. Theory and implementation. *Mathematical Geology* 25:641-655.

Nicholls N. 1997. Developments in climatology in Australia: 1946-1996, in *Australian Meteorological Magazine* 46:127-135.

Nix H.A. 1986. A biogeographic analysis of Australian elapid snakes, in R. Longmore (ed) *Atlas of Australian Snakes*, Bureau of Flora and Fauna, Canberra.

Orlanski I. 1975. A subdivision of scales for atmospheric processes, *Bulletin of the American Meteorological Society*, 56:527-530.

Panofsky H.A. and Dutton J.A. 1984. *Atmospheric Turbulence: models and methods for engineering applications*. John Wiley and Sons, New York.

Peuquet D.J. 1984. A conceptual framework and comparison of spatial data models. *Cartographica*, 21:66-113.

Peuquet D.J. and Duan N. 1995. An event based spatiotemporal data model (ESTDM) for temporal analysis of geographical data, in *International Journal of Geographical Information Systems* 9:1:7-24.

Phillips D.L., Dolph J. and Marks D. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* 58:119-141.

Pittock A.B. 1988. Actual and anticipated changes in Australian climate. In *Greenhouse, Planning for Climatic Change*, Pearman (ed), CSIRO, Melbourne. pp. 35-51.

Plummer N. 1996. Temperature variability and extremes over Australia: part 1 - recent observed changes, in *Australian Meteorological Magazine* 45:233-250.

Plummer N., Lin Z. and Torok S. 1995. Trends in the diurnal temperature range over Australia since 1951, in *Atmospheric research* 37:79-86.

Reiter E.R. 1961. *Jet-stream meteorology*. Uni. of Chicago Press. Chicago.

Reiter E.R. 1967. *Jet streams: How do they affect our weather*. Doubleday. New York.

Richardson C.W. 1977. A model of stochastic structure of daily precipitation over an area. *Colorado State University Hydrology Paper*, No. 91.

Richardson C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resour. Res.* 17:182-190.

Ripley B.D. 1984. Present position and future developments: Some personal views - Statistics in the natural sciences, in *J. Royal Stat. Soc.* (A), 147:2:340-348.

Rodríguez-Iturbe I. and Mejía J.M. 1974. The design of rainfall networks in time and space. *Water Resour. Res.* 10:713-728.

Rouhani S. and Hall T.J. 1989. Space-time kriging of groundwater data, in M. Armstrong (ed), *Geostatistics* vol. 2 pp639-650, Kluwer Academic, Dordrecht.

Rouhani S. and Myers D.E. 1990. Problems in space-time kriging of geohydrological data, *Math Geology* 22:611-623.

Rouhani S. and Wackernagel H. 1990. Multivariate geostatistical approach to space-time data analysis. *Water Resour. Res.* 26:585-591.

Schneider S.H. 1992. Introduction to climate modelling, in K.E. Trenberth (ed), *Climate System Modelling*, Cambridge University Press, Cambridge.

Seaman R.S. and Hutchinson M.F. 1983. Comparative real data tests of some objective analysis methods by withholding observations. *Austral. Meteorol. Mag.* 33:37-46.

Seber G.A.F. 1977. *Linear regression analysis*, Wiley, New York.

Shepard D.L. 1968. A two dimensional interpolation function for irregularly spaced data. *Proceedings of the 23rd National Conference, Association of Computing Machinery,* pp. 517-524.

Sibson R. 1981. A brief description of natural neighbour interpolation, in V.Barnett (ed) *Interpreting multivariate data*, Wiley and sons, Chichester.

Smith D.I., Hutchinson M.F. and McArthur R.J. 1993. *Climatic and agricultural drought: payments and policy.* Research Report RES7, Centre for Resource and Environmental Science.

Snedecor and Cochran W.G. 1989. *Statistical methods* (8th edition). Iowa State University Press, Ames.

Somerville R. 1987. The predictability of weather and climate. *Climate Change* 11:239-246.

Srikanthan R. and Stewart B.J. 1991. Analysis of Australian rainfall data with respect to climate variability and change, in *Australian Meteorological Magazine* 39:11-20.

Steyne D.G., Oke T.R., Hay J.E. and Knox J.L. 1981. On scales in meteorology and climatology, *Climate Bulletin* 39:1-8.

Stidd C.K. 1954. The use of correlation fields in relating precipitation to circulation. *J. Meteorol.* 11:202-213.

Stidd C.K. 1973. Estimating the precipitation climate. *Water Resources Research* 9, 1235-1241.

Stillman S.T. and Wilson J. 1996. Comparison of ANUSPLIN, MT-CLIM-3D and PRISM precipitation estimates, a paper presented to the Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling, Jan 21-25, 1996, Santa Fe, New Mexico, USA.

Streten N.A. 1982. Exploring Southern Hemisphere climates, in *Australian Meteorological Magazine.* 30:143-153.

Sturman A.P. and Tapper N.J. 1996. *The weather and climate of Australia and New Zealand*, Oxford University Press Australia, Melbourne.

Thiessen A.H. 1911. Precipitation averages for large areas. *Monthly Weather Review*, 39:1082-1143.

Torok S.J. and Nicholls N. 1996. A histroical temperature data set for Australia, in *Australian Meteorological Magazine* 45:251-260.

Troupe A.J. 1967. Opposition of anomalies of upper tropospheric winds at Singapore and Canton Island, *Australian meteorological magazine*, 50:35-43.

Tuan Y. 1993. Voices, sound and heavenly music, in *Passing strange and wonderful aesthetics, nature and culture*, Island press, Washington.

Tucker G.B. 1975. Climate: Is Australia's changing?, in *Search* 6:323-328.

Wahba G. 1979. How to smooth curves and surfaces with splines and cross-validation. *Tech. Report 555*. University of Wisconsin-Madison, Statistics Department.

Wahba G. 1990. *Spline Models for Observational Data.* CBMS-NSF Regional Conference Series in Mathematics 59. SIAM, Philadelphia.: Society for Industrial and Applied Mathematics, 169pp.

Wahba G. and Wendelberger J. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review* 108:1122-1143.

Walsh J.E., Richman M.B. and Allen D.W. 1982. Spatial coherence of monthly precipitation in the United States. *Mon. Weath. Rev.* 110:272-286.

Watson D.F. and Philip G.M. 1985. A refinement of inverse distance weighted interpolation, in *Geoprocessing* 2:315-327.

Weisberg S. 1985. *Applied Linear Regression*, 2nd Edition, John Wiley and Sons, New York.

Wessel P. and Smith W.H.F. 1991. *The GMT-SYSTEM.* School of Ocean and Earth Science Technology, Hawaii.

Whetton P., Mullan A.B. and Pittock B. 1996. Climate-change scenarios for Australia and New Zealand, in *Greenhouse: Coping with climate change*, Bouma W.J., Pearman G.I. and Manning M.R. (eds.), CSIRO Publishing, Melbourne, pp. 145-168.

Wilks D.S. 1989. Statistical specification of local surface weather elements from large-scale information. *Theor. Appl. Climatol.* 40:119-134.

Wolff R.S. and Yaeger L. 1993. *Visualization of natural Phenomena.* Springer-Verlag, New York.

World Meteorological Organisation 1983. *Guide to meteorological instruments and methods of observation.* WMO, Geneva.

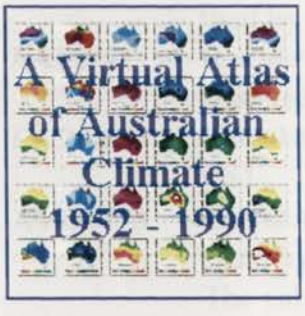
Yarnal B. 1993. *Synoptic Climatology in Environmental Analysis.* Belhaven Press, London.

Zuo H, Hutchinson M.F., McMahon J.P. and Nix H.A. 1996. Developing a mean monthly climatic database for China and Southeast Asia, in T.H. Booth (ed) *Matching trees and sites.* Proceedings of an international workshop held in Bangkok 27-30 March 1995, ACIAR Proceedings No. 63, 126p.

Appendix A - The Accompanying Computer Disk

The accompanying computer disk contains an electronic version of the PhD thesis and a virtual atlas of Australian climate, 1952-1990. The CD was designed to run on a PC with at least windows 95 and a sound card as the sound and animation programs are designed to work on a PC platform. Most of the files can also be accessed on a UNIX based machine.

The thesis can be accessed by using an HTML browser, such as Mosaic or Netscape. After mounting the CD-ROM, you can start your HTML browser and use the "open file" (for Netscape) command to open the CD-ROM Home Page, which is main.html. The Home Page contains links to all of the various files comprising the thesis. All of the chapters are in HTML, and most contain GIF images. A few chapters have links to animation and sound files, and many contain links to the World Wide Web. Your HTML browser must be configured to read these files (see the introduction and help files on the CD). All hypertext links are colored and/or underlined. To traverse a link, move the mouse to the colored/underlined text. The destination of the link will be displayed at the bottom of the browser window. Click on the highlighted text to go to the destination file.



The virtual atlas is the result of data analysis done for the PhD thesis and can also be accessed from main.html. The virtual atlas comprises:

- Station dictionaries and data.
- Grid point data and statistics extracted from the surfaces on a 2.5° by 2.5° grid.
- Over 20,000 graphic representations of both long term monthly and monthly climate variables.
- The Circular GIF viewer.
- Animations of both long term monthly and monthly climate variables.
- Several sound sequences which can be played in combination with the animation files.

Appendix B - The ANUSPLIN Programs

ANUSPLIN is a package of FORTRAN programs for fitting surfaces to noisy data as functions of one or more independent variables. The package also contains programs for interrogating the fitted surfaces in both point and grid form. A brief summary of the programs are described here. For further information contact the Centre for Resource and Environmental Studies, ANU or visit the CRES web site on http://cres.anu.edu.au/software/index.html.

Brief program summary

SPLINA A FORTRAN program which fits up to 12 different (partial) thin plate smoothing spline functions of one or more independent variables. Suitable for data sets with up to several hundred points.

SPLINB An approximate version of SPLINA designed for larger data sets. It fits a partial thin plate smoothing spline, with user defined knots, which are initially selected by SELNOT to multi-variate noisy data. Suitable for data sets with up to a few thousand data points.

SELNOT Selects an initial set of knots for use by SPLINB.

ADDNOT Updates knot index file when additional knots are selected from the data file used by SPLINB or SPLINBB.

DELNOT Adjusts knot index file when points are removed from the data file used by SPLINB or SPLINBB.

LAPPNT Calculates values of partial thin plate spline surface(s) at points supplied on a file.

LAPGRD Calculates values of partial thin plate spline surface(s) on a regular grid.

AVGCVA Calculates the GCV for each surface and the average GCV for surfaces fitted by SPLINA for a range of values of the smoothing parameter. The values are written to a user specified file for inspection and for plotting.

AVGCVB Calculates the GCV values as in AVGCVA for surfaces fitted by SPLINB.

APPENDIX C - List of Stations Used in the Pressure and Temperature Surfaces

Station No.	Long	Lat	Sev m	Bev m	Start	End	No. mths	Station Name
002012	127.67	-18.23	422	424	1/52	12/90	252	Halls Creek AMO.
003003	122.23	-17.95	7	9	1/52	12/90	468	Broome AMO
003032	123.67	-17.37	7	12	1/73	12/80	96	Derby Aero
004020	119.75	-21.18	189	189	1/88	12/89	24	Marble Bar
004032	118.62	-20.37	9	6	1/52	12/90	468	Port Hedland AMO
005007	114.08	-22.23	5	5	3/75	12/90	190	Learmonth AMO
005016	115.12	-21.63	4	5	3/75	12/90	129	Onslow PO
005017	115.12	-21.67	3	5	1/52	12/74	276	Onslow AMO
006011	113.67	-24.88	4	7	1/52	12/90	468	Carnarvon AMO
007045	118.55	-26.62	522	519	1/52	12/89	432	Meekatharra AMO
008051	114.70	-28.80	38	35	1/52	12/90	468	Geraldton AMO
009021	115.97	-31.93	20	29	1/52	12/90	466	Perth Airport MO (Composite)
009034	115.87	-31.95	19	20	1/43	12/90	526	Perth RO
009500	117.88	-35.02	18		1/60	12/63	48	Albany Town
009518	115.13	-34.37	14	14	1/88	12/90	35	Augusta (Cape Leeuwin) AWS
009741	117.80	-34.95	68	69	1/67	12/90	288	Albany AMO.
009789	121.90	-33.83	25	26	1/70	12/90	252	Esperance MO
010536	117.87	-32.33	295	296	1/88	12/89	24	Corrigin PO
010579	117.55	-33.68	310	311	1/88	12/90	35	Katanning PO
010592	118.47	-33.10	286	287	1/88	10/90	34	Lake Grace PO
011004	128.12	-30.83	159	160	1/52	12/90	468	Forrest AMO
012038	121.45	-30.78	365	370	1/52	12/90	466	Kalgoorlie AMO
012046	121.32	-28.88	376	380	1/88	12/90	35	Leonora PO
012065	121.78	-32.20	277	278	1/88	10/90	34	Norseman PO
012074	119.33	-31.23	355	356	1/88	11/90	35	Southern Cross PO
013017	128.30	-25.03	580	599	1/57	12/90	408	Giles MO
014015	130.88	-12.43	31	30	1/52	12/90	467	Darwin Airport AWS
014161	130.83	-12.47	27	40	5/67	12/73	79	Darwin RO
014508	136.82	-12.28	52	53	1/86	12/90	60	Gove Airport

Long - longitude, Lat - Latitude, Sev - Station elevation, Bev - Barometer elevation
LH - lighthouse, RO - regional office, PO - Post Office
AMO - airport met office, MO - Met office, AWS - automatic weather Station

Station No.	Long	Lat	Sev m	Bev m	Start	End	No. mths	Station Name
014626	133.38	-16.27	210	214	1/52	12/68	203	Daly Waters AMO
015085	135.95	-18.63	218	219	1/88	12/90	36	Brunette Downs Stn
015135	134.18	-19.63	375	376	1/70	12/90	252	Tennant Creek MO
015548	130.02	-20.18	340	340	1/88	12/90	36	Rabbit Flat Stn
015590	133.88	-23.82	546	547	1/52	12/90	468	Alice Springs AMO
016001	136.82	-31.15	165	167	1/52	12/90	456	Woomera AMO
016044	134.57	-30.72	120	120	1/88	12/90	36	Tarcoola PO
017043	135.45	-27.57	113	117	1/52	12/71	226	Oodnadatta AMO
018012	133.72	-32.13	15	16	1/52	12/90	468	Ceduna AMO
018069	134.88	-33.65	4	5	1/88	12/90	18	Elliston PO
018070	135.85	-34.72	4	5	1/88	12/90	36	Port Lincoln PO
018103	137.53	-33.03	13	15	1/88	12/90	36	Whyalla (Norrie) PO
019066	137.78	-32.53	4	8	1/88	12/90	36	Port Augusta Power Station
022008	137.68	-34.38	185	186	1/88	12/90	36	Maitland PO
023000	138.58	-34.93	40	49	1/55	12/76	264	Adelaide (West Terrace)
023013	138.63	-34.80	14	16	1/52	12/54	36	Parafield Airport
023034	138.53	-34.95	6	4	2/55	12/90	429	Adelaide Airport Composite
023090	138.62	-34.92	47	51	2/77	12/90	167	Adelaide RO
024016	140.75	-34.17	20	21	1/88	12/90	36	Renmark PO
025509	140.52	-35.33	99	100	1/88	11/90	35	Lameroo PO
026021	140.78	-37.75	63	69	1/42	12/90	476	Mt. Gambier AMO
027022	142.22	-10.58	58	62	1/52	12/77	312	Thursday Island MO
027042	141.88	-12.63	11	12	1/88	12/90	36	Weipa Composite
029009	140.52	-20.67	189	190	1/52	12/74	276	Cloncurry AMO
029041	141.07	-17.67	8	10	1/88	11/90	35	Normanton PO
029127	139.48	-20.67	343	344	1/76	12/90	180	Mount Isa AMO
030018	143.55	-18.30	292	295	1/88	11/90	35	Georgetown PO
030045	143.13	-20.73	211	215	1/88	10/90	33	Richmond PO
031011	145.75	-16.88	3	7	1/52	12/90	456	Cairns AMO
032040	146.77	-19.23	4	6	1/52	12/90	468	Townsville AMO
033119	149.22	-21.12	30	32	1/60	12/90	372	Mackay MO
035027	148.17	-23.53	179	180	1/88	12/90	36	Emerald PO
036031	144.28	-23.43	192	193	1/72	12/90	120	Longreach AMO
036143	145.47	-24.42	283	284	1/88	11/90	35	Blackall PO

Long - longitude, Lat - Latitude, Sev - Station elevation, Bev - Barometer elevation
LH - lighthouse, RO - regional office, PO - Post Office
AMO - airport met office, MO - Met office, AWS - automatic weather Station

Station No.	Long	Lat	Sev m	Bev m	Start	End	No. mths	Station Name
037010	138.12	-19.92	234	235	1/88	11/90	35	Camooweal PO
037051	143.03	-22.40	182	184	1/88	12/89	24	Winton PO
038002	139.37	-25.92	47	48	1/88	12/90	36	Birdsville Police Station
038003	139.90	-22.92	157	158	1/88	11/90	35	Boulia PO
038024	142.65	-25.43	126	129	1/88	11/90	35	Windorah PO
039015	152.35	-24.87	14	14	1/60	12/89	211	Bundaberg PO
039039	151.62	-25.63	106	108	1/88	11/90	35	Gayndah PO
039059	152.72	-24.12	4	6	1/88	12/90	16	Lady Elliot Is. LH
039083	150.48	-23.38	10	14	1/52	12/90	468	Rockhampton AMO
039123	151.27	-23.85	75	76	1/58	12/90	396	Gladstone MO
040004	152.72	-27.65	27	31	1/53	12/81	348	Amberley Aero
040043	153.47	-27.03	99	102	1/88	12/89	24	Cape MOreton LH
040068	153.20	-25.93	77	78	1/88	12/90	36	Double Is. Point LH
040126	152.72	-25.52	11	11	1/88	10/90	34	Maryborough Composite
040214	153.03	-27.48	38		1/51	12/85	419	Brisbane RO Roof
040223	153.12	-27.38	4	5	1/52	12/90	468	Brisbane AMO
041038	150.30	-28.55	217	219	1/88	10/90	34	Goondiwindi PO
041359	151.73	-27.42	406	407	1/76	12/81	72	Oakey AMO
043030	148.78	-26.57	299	301	1/88	12/90	36	Roma PO
044021	146.27	-26.42	303	303	1/52	12/90	466	Charleville AMO
044026	145.68	-28.07	189	191	1/88	12/90	35	Cunnamulla PO
045017	143.82	-28.00	129	130	1/88	12/89	24	Thargomindah PO
047007	141.47	-31.98	315	316	1/63	12/75	156	Broken Hill (Patton Street)
048027	145.83	-31.48	221	264	1/63	12/90	336	Cobar MO
053048	149.85	-29.47	212	212	1/69	12/90	264	Moree MO
055054	150.85	-31.08	404	410	1/61	12/79	204	Tamworth (Airport)
059030	153.08	-30.92	117	114	1/88	12/90	36	Smoky Cape LH
059040	153.12	-30.32	5	6	1/52	12/90	468	Coffs Harbour MO
060026	152.92	-31.45	7	10	1/88	12/90	35	Port Macquarie (Hill Street)
060030	152.48	-31.90	5	7	1/88	11/90	35	Taree Radio Station
061078	151.83	-32.80	9	8	1/52	12/90	468	Williamtown AMO
063231	149.12	-33.38	948	951	1/70	12/90	252	Orange Airport
064008	149.28	-31.28	509	511	1/88	12/90	36	Coonabarabran PO

Long - longitude, Lat - Latitude, Sev - Station elevation, Bev - Barometer elevation
LH - lighthouse, RO - regional office, PO - Post Office
AMO - airport met office, MO - Met office, AWS - automatic weather Station

Station No.	Long	Lat	Sev m	Bev m	Start	End	No. mths	Station Name
065026	148.15	-33.15	324	323	1/89	12/90	24	Parkes (Macarther Street)
066037	151.17	-33.93	6	4	1/52	12/90	468	Sydney Airport Obs.Office MO
066062	151.20	-33.87	42	69	1/55	12/90	432	Sydney RO
067033	150.78	-33.60	19	21	1/54	1986	264	Richmond AMO/MO
068076	150.53	-34.95	109	110	1/56	11/90	419	Nowra RAN Air Station
069018	150.15	-35.92	17	17	1/88	11/90	35	Moruya Heads Pilot Station
070014	149.20	-35.32	571	577	1/52	12/90	468	Canberra AMO
070282	149.12	-35.27	564	579	1/75	1987	156	Canberra City
072146	146.92	-36.07	165	171	1/86	12/90	59	Albury Airport
072150	147.47	-35.17	221	213	1/52	12/90	468	Wagga AMO
076031	142.08	-34.23	51	52	1/52	12/90	468	Mildura Airport AMO
078031	141.63	-36.33	129	130	1/59	1986	336	Nhill Composite /MO
081003	144.28	-36.75	225	226	1/88	12/90	36	Bendigo Prison
082002	145.98	-36.55	170	170	1/88	1988	12	Benalla (Shad.St.)
084016	149.92	-37.57	15	15	1/88	12/90	36	Gabo Island
085072	147.13	-38.12	5	8	1/52	12/90	468	East Sale AMO
085096	146.42	-39.13	89	97	1/88	11/90	35	Wilsons Promontory LH
086038	144.90	-37.73	86	81	1/52	12/70	228	Essendon Airport AMO
086071	144.97	-37.82	35	113	1/55	12/90	432	Melbourne RO
086282	144.85	-37.68	132	141	1/71	12/90	240	Melbourne Airport
087031	144.75	-37.87	18	21	1/52	12/90	468	Laverton Aero AMO
089002	143.78	-37.52	441	442	1/88	12/90	36	Ballarat Composite
089085	142.98	-37.27	295	296	1/88	12/90	36	Ararat Prison
090015	143.52	-38.87	82	83	1/88	12/90	36	Cape Otway LH
090172	142.43	-38.30	74	76	3/83	12/90	94	Warrnambool Airport
090173	142.07	-37.65	233	234	1/84	11/90	81	Hamilton (Wilken)
091057	146.78	-41.05	28	30	1/88	11/90	35	Low Head LH

Long - longitude, Lat - Latitude, Sev - Station elevation, Bev - Barometer elevation
LH - lighthouse, RO - regional office, PO - Post Office
AMO - airport met office, MO - Met office, AWS - automatic weather Station

Station No.	Long	Lat	Sev m	Bev m	Start	End	No. mths	Station Name
092045	148.35	-41.00	13	14	1/88	11/90	35	Eddystone Point LH
094007	147.48	-42.83	10	15	1/51	12/57	84	Cambridge Aero AMO
094008	147.50	-42.83	4	27	1/59	12/90	384	Hobart Airport AMO
094010	147.15	-43.48	55	53	1/88	12/90	36	Cape Bruny LH
094029	147.33	-42.88	55	57	1/51	12/90	480	Hobart R O
094041	146.27	-43.65	147	147	1/88	12/90	36	Maatsuyker Is LH
098001	143.85	-39.93	24	24	1/88	12/90	36	King I. (Currie PO)

Long - longitude, Lat - Latitude, Sev - Station elevation, Bev - Barometer elevation
LH - lighthouse, RO - regional office, PO - Post Office
AMO - airport met office, MO - Met office, AWS - automatic weather Station