TECHNISCHE UNIVERSITÄT WIEN

D I P L O M A R B E I T

Spatial Blind Source Separation for Soil Moisture Data

ausgeführt am

Institut für
Stochastik und Wirtschaftsmathematik
TU Wien

unter der Anleitung von

Associate Prof. Dipl.-Stat. PhD Klaus Nordhausen

und

Dipl.-Ing. Christoph Mühlmann

durch

Luzia Jorda

Matrikelnummer: 01026520

Wimbergergasse 9/18

1070 Wien

Wien, am 17.12.2020

Abstract

Spatial Blind Source Separation (SBSS) is a recent extension of Independent Component Analysis (ICA) for spatial data. Standard ICA ignores the spatial dependency structure of spatial data, while SBSS uses this information. The goal of this thesis is to evaluate the SBSS method in a new field of application. In cooperation with the Institute of Geodesy and Geoinformation of the Vienna University of Technology, Australian soil moisture data are investigated, consisting of gridded satellite observations from 1998 to 2018. Soil moisture is an essential factor in understanding climate processes and therefore weather extremes and climate change. Understanding space-time patterns of soil moisture facilitates insights in the fields of hydrology, agriculture, and socioeconomics. Principal Component Analysis (PCA), ICA and SBSS are all applied to the data and their results are contrasted with each other and the existing literature on Australian soil moisture data in the context of Blind Source Separation (BSS). Correlations between loadings of the results of PCA, ICA and SBSS and the most relevant climate modes for Australia are investigated via Spearman correlations for concurrent and time-lagged observations. One finding of this work is that the results of SBSS are consistent with existing studies, while ICA, when looking at anomalies, fails to provide new insights or even reproduce known results. Spatio-temporal dependencies of the observations are explicitly taken into account in the novel SBSS approach, while they are ignored in the context of standard ICA. The thesis aims to identify the advantages of SBSS over conventional PCA and ICA in the context of the presented analysis. Notable, higher correlations to the Climate Oscillation Indices (COIs) are obtained for SBSS, and new patterns of SBSS components complement existing knowledge. SBSS is a useful candidate for BSS of climate processes.

Danksagung

Für Mama. Für den ersten Schritt und alle anderen.

Danke, Klaus und Christoph, für das denkbar beste Betreuungsverhältnis, unzählige investierte Stunden und die große Geduld mit mir. Danke an die ESA und Wouter Dorigo für die Bereitstellung der Daten, die Kooperationsbereitschaft und vor allem die hochgeschätzte Expertise. Danke an Peter Filzmoser für das Zusammenführen und die Themenwahl. Danke, Laura, für die Datenaufbereitung, den fachlichen Rat, etliche E-Mails, Telefonkonferenzen und deine positive, stets aufbauende Art.

Danke an meine Eltern. Für alles. Danke an Anselm, der immer so realistisch war und danke an Agnes, die immer so unrealistisch war. Ohne euch beide hätte ich den Kopf verloren. Danke an Niki, dem das alles nie wichtig war.

David, ich hoffe, ich schaffe es, dir in Zukunft die Stütze und Hilfe zu sein, die du mir in den letzten Monaten (nicht nur, aber neben allem und vor allem bei dieser Arbeit) warst.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 17.12.2020

_____
Name der Autorin

Contents

i

1 Introduction

The expansion and continuous development of earth observations via satellites have led to dense data series over long time periods at a global resolution. One of the aims of remote sensing is to improve our understanding of climate and the underlying processes. This is important to improve predictions of climatic catastrophes, thus, preventing tremendous damage and to monitor climate change and identify its possible drivers.

The terms weather and climate are sometimes mistakenly used synonymously. They differ in their definition (we follow the definition of the "Zentralanstalt für Meteorologie und Geodynamik" (ZAMG) [7]) and consequently in their modeling. Weather refers to the simultaneous state of the atmosphere and consists of tangible phenomena such as low pressure, convection, or precipitation for some hours or a few weeks. On the other hand, climate describes a much longer process and can only be understood by looking at several decades or even longer time spans. The climate system consists of several components (atmosphere, hydrosphere, lithosphere, biosphere, and cryosphere), which are all influenced by the exchange of energy, mass, the water cycle, radiation and chemical transformations [7]. The climate system is correspondingly complex. The weather could also be interpreted as a realization of the climate at a particular time and location [9].

Fundamental factors to understand the climate are the Essential Climate Variables (ECV). One of them is soil moisture as a spatial-temporal variable. Studies indicate that it is useful to look at soil moisture to better understand global weather and climate phenomena (see e.g. [22]). Although soil moisture is only a small part of the hydrological cycle, it reacts strongly to events such as precipitation, humidity, wind and solar radiation [10]. ESA's CCI project provides a continuous dataset of soil moisture observations starting in 1979 with very dense global coverage. Nevertheless, not only the monthly and seasonal averages of these measurements can provide interesting information, but also the anomalies of the data should be analyzed for a better understanding of the processes.

In this work, we limit our study region to Australia, for several reasons. Australia, located south of the equator, is separated from other land surfaces. The continent is exposed to the influences of the Indian Ocean in the west and the Pacific Ocean in the east. In Australia we have just witnessed devastating fires towards the end of 2019, severely affecting the country and the local population. Bush fires, extreme droughts and floods are unfortunately not uncommon in Australia. This makes it an interesting site to study essential mechanisms between landmasses, oceans and the atmosphere that affect the climate. These factors include climate modes, for instance. Countless studies have found compelling arguments to show the influence of three specific climate modes on Australia, the Indian Dipole Mode (IOD), the Southern Annual Mode (SAM) and the El Niño Southern Oscil-

lation Mode (ENSO) (see for example [30], [19] and [14]). There are still many gaps in our understanding of the interrelationships of Earth's complex climate system. The exact influence of the climate modes on local and global climatic conditions, some of which are effective over long distances (so-called teleconnections) such as ENSO, still needs further investigation.

The CCI soil moisture data is a spatio-temporal dataset, which we interpret as multi-variate realizations in the context of SBSS by considering the vector of the different time points for each location. Principal Component Analysis (PCA) is a common tool for multivariate data, as it presents a way to conveniently reduce the dimension of the dataset and thus makes it easier to interpret. Especially in the field of remote sensing, the analysis of the influence of climate modes on soil moisture is often carried out with PCA. For in-depth PCA studies on soil moisture in our region of interest, Australia, see [9] and [10], respectively. This work aims at applying Spatial Blind Source Separation (SBSS), to gain further insights into the relationship between climate modes and soil moisture and to introduce SBSS as a suitable method for the analysis of other satellite observations at the global scale. While PCA offers an excellent way to deal with a large amount of data, it has its downsides in that it is restricted to second-order dependencies and no underlying dependence structure of the data, such as spatial correlations, is considered. It is often used for the purpose of dimension reduction to "compress" the data. SBSS is an extension of Independent Component Analysis (ICA) and was introduced in [28]. The reason why we prefer SBSS over classical ICA is that ICA does not take the spatial dependence structure of the data into account. Soil moisture observations usually portray spatial dependence. By accounting for this, SBSS could provide crucial new insights into the relationship between climate modes and soil moisture. We first want to investigate whether SBSS can affirm existing results of PCA for Australia and the three most important climate indices as influencing factors. For this, we follow the considerations of [9]. As an evaluation criterion, we use the robust Spearman Rank correlation in this thesis.

The structure of this thesis is a s follows. We present a brief overview of the CCI dataset and the three dominant climate modes for Australia, IOD, SAM and ENSO (Chapter 2). For a more detailed explanation we refer to [9] or the corresponding geological literature. Furthermore, the mathematical frameworks of PCA, ICA, and SBSS are discussed (Chapter 3). All three methods are applied and compared over the course of this work. The mean values of soil moisture observations and the corresponding anomalies are considered at a monthly resolution and at a seasonal level, i.e. observations for summer, autumn, winter and spring separately, at a monthly resolution. In Chapter 5, selected results of these analyses are presented and discussed. However, the complete list of all results can be found in the Appendix 7.4. In the Conclusion, we will summarize to what extent advantages of SBSS over PCA have been confirmed by the analysis, and what new insights have been established. Furthermore, we offer a short outlook on further research questions, which could be investigated in the future but would have exceeded the scope of this work.

2 Data

The Climate Change Initiative (CCI is a project of the European Space Agency (ESA). In cooperation with various scientific institutes (the Vienna University of Technology - TU Wien - being one of them), Essential Climate Variables (ECV) of the Global Climate Observing System (GCOS) are gathered. Soil Moisture (SM) is one of these ECVs [3].
SM is the water contained in the spaces between the soil particles and it was repeatedly shown that SM is an essential driver of the dynamics of water- and energy-fluxes on land and in the atmosphere [22]. Studies suggest that SM has a spatial dependence as well as a time stability [34]. It is challenging to introduce a generally valid definition because different scientific disciplines view SM differently and have different requirements. A general distinction is made between root zone soil moisture and surface soil moisture. While agriculture is interested in the root zone soil moisture, which can go down to two meters into the soil, satellite observations can only monitor information from the top few centimeters [5]. In the following, we use the terms of soil moisture and surface soil moisture interchangeably.

CCI provides a unique dataset that combines observations from multiple satellites and global coverage of the surface of the Earth. The data comes from active and passive microwave observations. Scatterometers provide active observations by measuring the scattered energy after an electromagnetic pulse. The TU Wien processes the active data. The passive dataset stems from measuring emissions from the surface of the Earth without sending pulses in the first place. The VU University of Amsterdam and the National Aeronautics and Space Administration (NASA) cooperate to process the data of these passive measurements. Adding passive measurements into the assembly of a combined dataset of active and passive measurements, it was shown in studies that the estimations did improve. As stated in [22], including passive measurements into the analysis can "reduce errors in forecasting soil moisture profile as a result of poor initialization and improve the resulting predictions of runoff and evapotranspiration". The CCI Soil Moisture dataset is, therefore, a solid tool for climate analysis. The CCI project was the first to provide a multi-decadal time series for soil moisture with global coverage starting from 1978. It provides data for long-term analysis of essential climate variables and their influence on the climate on a global scale. Upon request, the dataset is freely available and detailed manuals are provided [3].

The CCI SM dataset has a monthly temporal resolution and has the form of a longitude-latitude grid with a spatial resolution of 0.25° × 0.25°. Because of the curvature of the Earth this does not correspond to an equidistant grid in kilometers. The values are given

in $\frac{1}{100} m^3 m^{-3}$ as it is derived by on the basis of volumetric water content:

$$\frac{V_w}{V_w + V_s + V_a},$$

with the volume of soil V_s, the volume of water V_w, the volume of air V_a of the total soil composition. Note that SM values cannot be negative and the range of the timeframe we consider in this thesis spans from 0.02 to 0.46.

The corresponding anomalies of the dataset remove seasonal signals by subtracting climatologies $\bar{c}_{tl}$. For the determination of the climatologies $\bar{c}_{tl}$ the overall mean of each month t over the time span of the dataset and each location l was calculated:

$$\bar{c}_{tl} = \frac{1}{n_t} \sum_{i=1}^{n_t} x_{itl} \quad \text{with} \quad t \in \{\text{January, February, \ldots, December}\}. \tag{2.1}$$

Here the soil moisture observations are averaged over the total number of available observations $i = 1, \ldots, n_t$ for a given month t and location l. For example, when we look at 1998 till 2018, which is the timespan we consider in this thesis, the mean of all $n_t = 21$ January data at location l is used to calculate the January climatology at location l. The anomalies x' are calculated by subtracting the corresponding climatologies from the SM observations x_{itl} for each location and month:

$$x'_{itl} = x_{itl} - \bar{c}_{tl} \tag{2.2}$$

Note that the anomalies can be negative. The range of the considered timeframe spans from -0.2 to 0.28. The anomaly dataset is frequently used to smooth out seasonal effects. For the purpose of this thesis, we consider SM and the corresponding anomalies of Australia.

2.1 Climate Oscillation Indices (COIs)

Climate modes describe atmospheric and oceanic oscillation that propagates to meteorological variables such as rainfall and soil moisture. Climate modes vary substantially in their periodical length. Some modes will have repeating patterns with weekly phases. Others have multi-annual phases.

The Indian Ocean Dipole (IOD), the Southern Annual Mode (SAM) and the El Niño Southern Oscillation (ENSO) are three modes that are known to be major climate drivers for Australia are discussed in detail and used in further analysis of this work [10]. In Figure 2.1 a schematic illustration of major climate drivers for Australia is given.
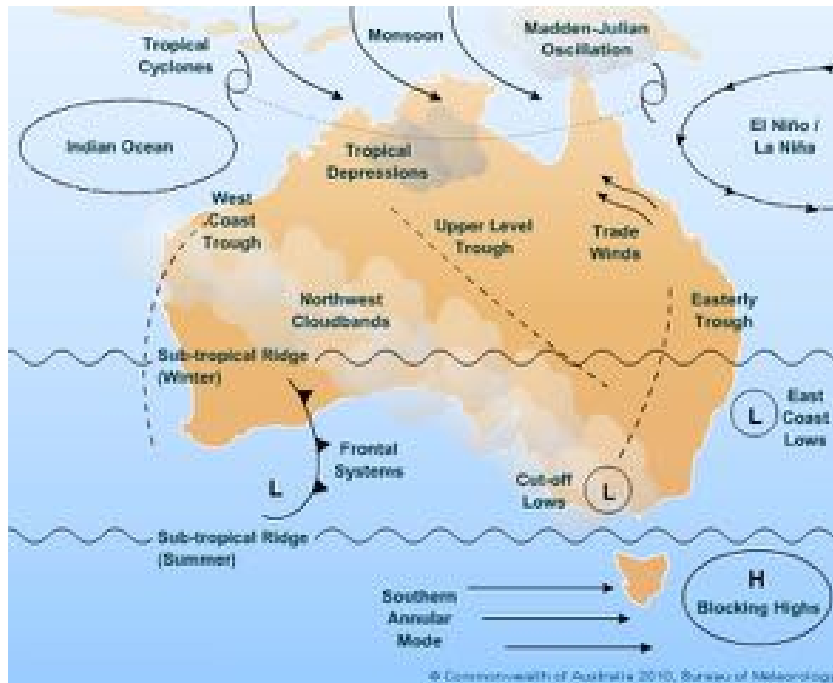
Figure 2.1: Climate drivers for Australia. Reprinted with permission of the Australian Bureau of Meteorology [2].

For each climate mode, a climate oscillation index (COI) is calculated to characterize the circulation patterns of the mode. This yields time series, that consist in many cases of a single value per entry. The data is usually in a monthly resolution and long-time records are available, which suits our analysis. Multiple sources grant open access to obtain the COI data online. The index can be derived by averaging data over a selected region or grid. Other indices are obtained by Principal Component Analysis (PCA) [4].

A remark on our notation: Since we restrict ourselves in this work to the continent of Australia, we stick to the local time calculation there. For this reason, the seasons are based on the Australian convention. When we speak of summer, we mean the months December to February. For that, we also write DJF. The same applies to autumn, which runs from March to May (MAM), to winter, which runs from June to August (JJA) and to spring, from September to November (SON).

2.1.1 Dipole Mode Index (DMI)

The Indian Ocean Dipole (IOD) is one of the major drivers of the Australian climate. It influences mainly the south of Australia during Australian winter (JJA) and spring (SON) [2]. The corresponding COI for the IOD is called Dipole Mode Index (DMI). The DMI is the sea surface temperature (SST) gradient of the western and eastern Indian Ocean at the equatorial level. It has a positive, negative and neutral phase. The term positive

(negative/neutral) IOD is used interchangeably for a positive (negative/neutral) DMI [6]. Each phase occurs on average every 3 to 5 years. Neutral phases of the DMI have little influence over the Australian climate other than a small rise of SST in the north-west of Australia. Positive and negative phases tend to begin in autumn or winter and become neutral at the end of spring when monsoonal periods begin in the north of Australia. A positive phase leads to a decrease of west winds over the Indian Ocean. The cold water in the north-west of Australia can rise and clouds dissolve. This dynamic leads to less rain during winter and spring in the area of impact (see the brown area in left-side panel of Figure 2.2).

During a negative phase, west winds over the Indian Ocean move warmer water to the north-west of Australia and prevent cold water from the depths to rise. Clouds form and this leads to an increase of rainfall during winter and spring over the area of impact (see the green area in the right-side panel of Figure 2.2). The DMI time series from 1998-2018 used in this analysis is retrieved from the National Oceanic and Atmospheric Administration (NOAA) [6].
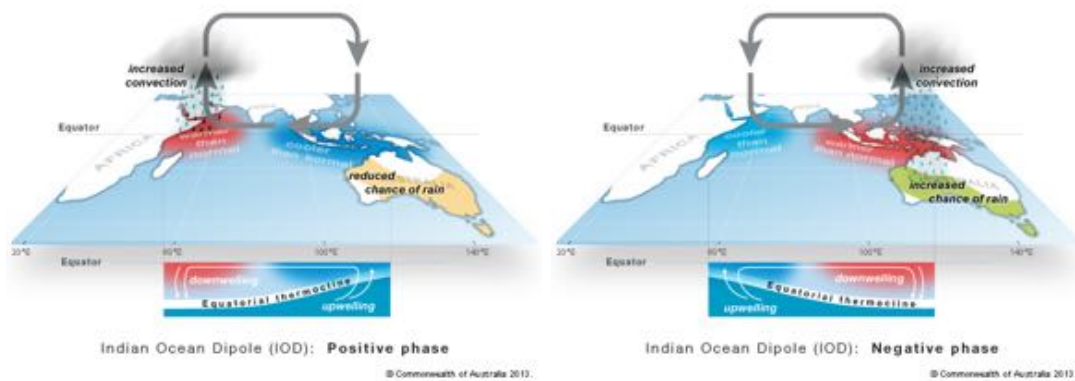

Figure 2.2: Schematic illustration of influence of positive (left panel) and negative (right panel) phases of the DMI on the Australian climate. Reprinted with permission of the Australian Bureau of Meteorology [2].

2.1.2 Southern Annual Mode Index (SAMI)

The Southern Annual Mode refers to the irregular north and south shifts of westerly winds south of a high-pressure belt of the southern hemisphere called the Sub-tropical Ridge. The corresponding index, the Southern Annual Mode Index (SAMI), is obtained by the zonal means of various stations at 40°S and 65°S [1]. The SAMI occurs in three phases (positive/negative/neutral) and influences the rain regime in South Australia. SAM seems to be related to ENSO, as a positive SOI tends to lead to a more negative SAMI and vice versa. There appears to be a general trend towards more positive phases. Each phase lasts up to two weeks and its impact on the Australian climate depends on the season [2]. SAM has its biggest influence over Australia during winter (JJA), when a negative SAMI phase caused by the high-pressure ring wandering north, brings cold fronts and rain to South Australia. An extreme event caused by the ring to be way up in the north and the

westerly winds blowing over Central Australia can lead to rain in the west and dryness in the east. The withdrawal of the high-pressure ring causes the positive phase of the SAMI (south shift of the ring) in Australians winter (JJA) and the cold fronts, which leads the rainclouds away from Australia and the south becomes drier while rain in the east may increase. These phenomena are illustrated in Figure 2.3.
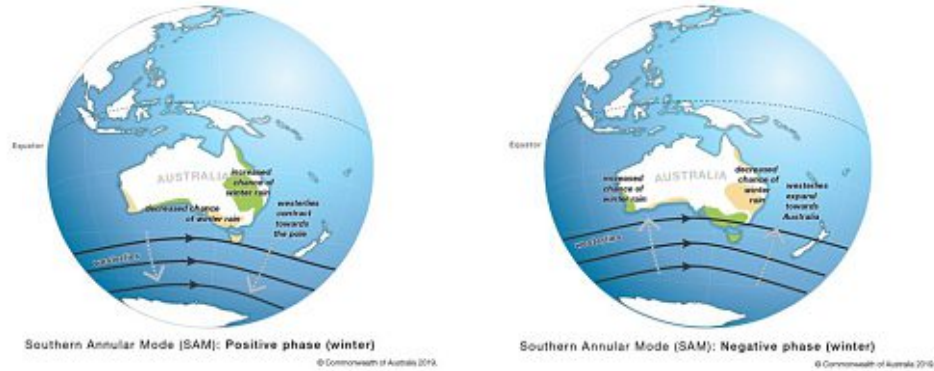

Figure 2.3: Schematic representation of the effect of positive (left panel) and negative (right panel) phases of the SAMI during winter (JJA) on the Australian climate. Reprinted with permission of the Australian Bureau of Meteorology [2].
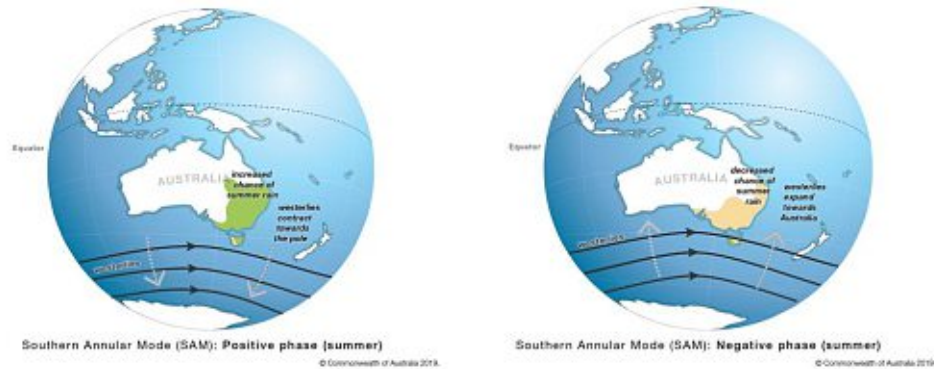

Figure 2.4: Schematic representation of the effect of positive (left panel) and negative (right panel) phases of the SAMI during summer (DJF) on the Australian climate. Reprinted with permission of the Australian Bureau of Meteorology [2].

During the summer (DJF), the impact of the SAMI is weaker and reversed. During summer, the Sub-tropical Ridge is far south and a northern movement during a negative phase will not necessarily reach South Australia (see Figure 2.1). A positive SAMI phase during summer can make way for tropical winds to rise in the north and moist winds in the east, leading to rain in the east. These phenomena are illustrated in Figure 2.4.

The climate in spring (SON) behaves similarly to the summer patterns, although in a weaker fashion. During fall (MAM), we expect no large impact of the the SAMI on the

Australian climate. A negative phase of the the SAMI during fall, however, can lead to warmer temperatures in North and Central Australia [2].

The SAMI time series from 1998-2018 used in this analysis is retrieved from the British Antarctic Survey (BAS) [1].

2.1.3 Southern Oscillation Index (SOI)

The El Niño Southern Oscillation is one of the primary climate drivers of Australia, comprising temperatures, winds and clouds over the Pacific Ocean at the equatorial level and, therefore the oscillation between an El Niño event to a La Niña event. Those events are often to blame for extreme weather phenomena. ENSO represents those extremes. The ENSO index is called Southern Oscillation Index (SOI) and is derived from measurements of the air pressure differences between stations in Tahiti in the South Pacific and Darwin, Australia. The SOI comes in three phases (positive/negative/neutral). A phase typically starts in the second half of the year and lasts till autumn of the following year. On average, the SOI needs four years to shift from a negative to a positive phase and back again and the same phases can repeat themselves in consecutive years [2]. Its main impact on the Australian climate happens during winter and spring. The neutral phase is characterized by constant trade winds blowing from east to west over the equatorial Pacific. Warm water is accumulated in the West Pacific (Australia) and gives way for cold water to rise to the sea surface in the East Pacific (South America). This dynamic is called the Walker Circulation. Half of the time, the SOI is in a neutral phase. During a negative phase (El Niño), trade winds decrease or even reverse and warm water drifts back east, causing the Walker Circulation to break down. This will reinforce the mechanisms of weakening the winds even more and bringing more warm water to the east. This leads to dryness in North and East Australia, with temperatures rising. See the left panel of Figure 2.5. A positive phase of the SOI (La Niña) is characterized by strong trade winds and cold water spreading in the western direction. The Walker Circulation gets reinforced and more rain or even floods are expected in Australia. Temperatures drop and the danger of cyclones emerges (see right panel of Figure 2.5).
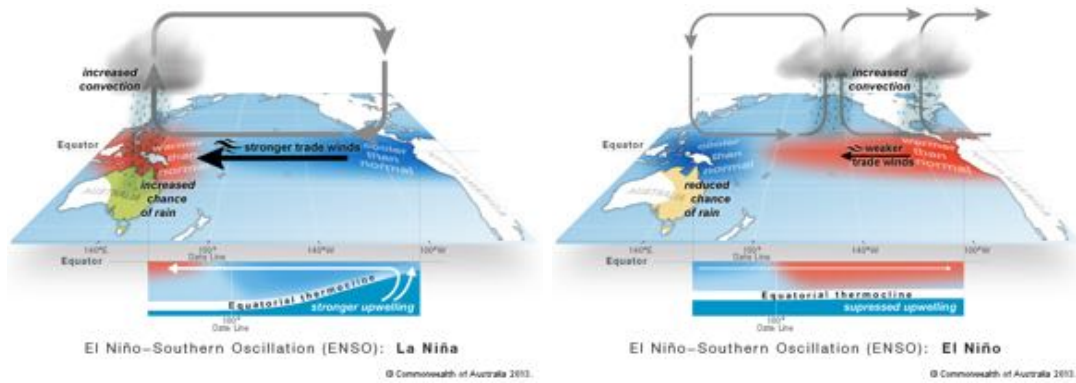
Figure 2.5: Schematic representation of the effect of positive (left panel) and negative phase (right panel) of the SOI on the Australian climate. Reprinted with permission of the Australian Bureau of Meteorology [2].

When a positive phase of the DMI coincides with a negative phase of the SOI (El Niño) the effects can add up and lead to even less rain and can be the cause of the absence of important winter and spring rainfalls in the south-east of Australia, followed by bushfires during summer. Negative phases of the DMI often pair with positive phases of the SOI (La Niña). This then leads to more rain and even floods, such as in the record-flood year of 2010 [2].

The SOI time series from 1998-2018 used in this analysis is retrieved from the Australians Bureau of Meteorology (BOM) [2].
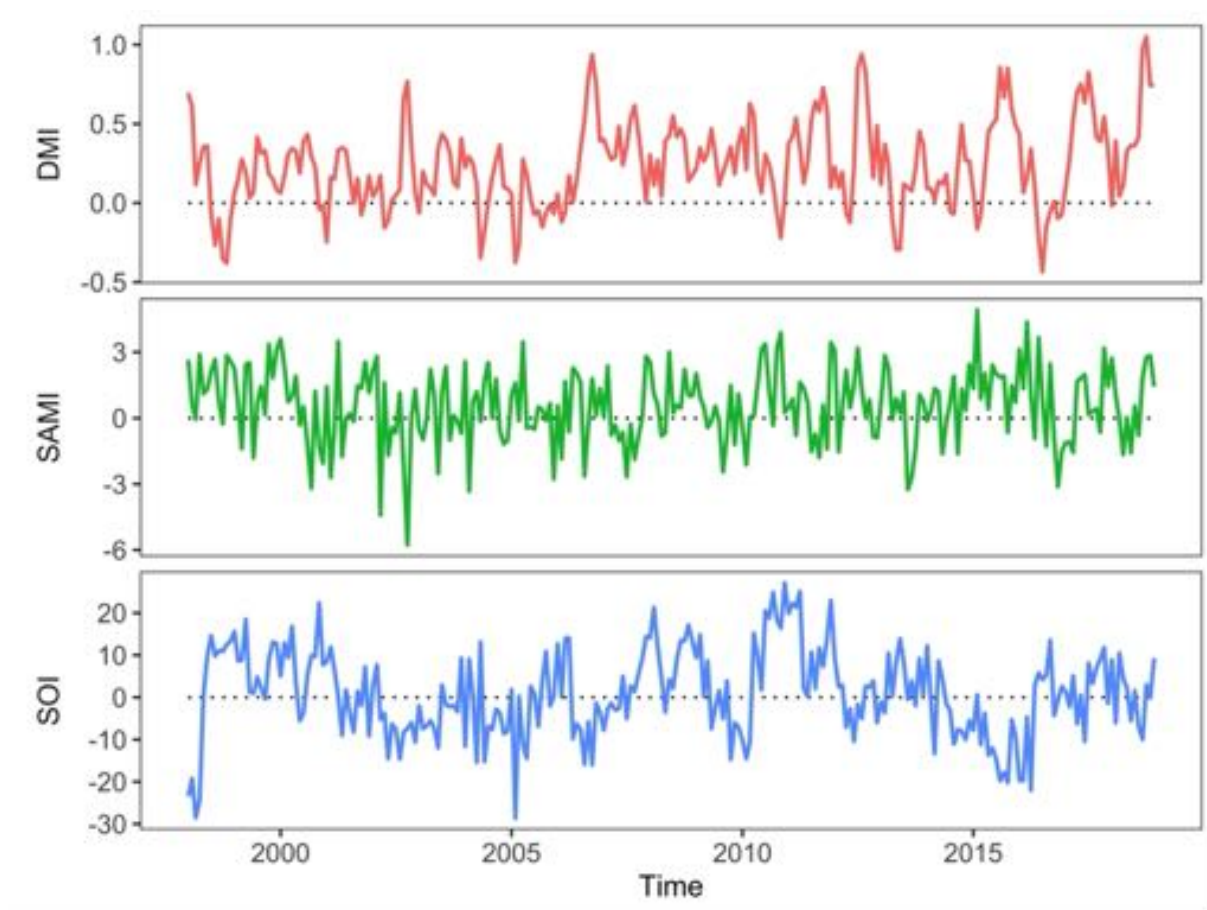
Figure 2.6: Time series of the Climate Oscillation Indizes (COIs) Dipole Mode Index (DMI), Southern Annual Mode Index (SAMI) and Southern Oscillation Index (SOI) from 1998-2018.

3 Methodology

We work with vector-valued observations, that we will interpret as realizations of random vectors. If not stated otherwise, we will assume, that we have iid realizations of these random vectors. However, we will also assume, that there are underlying latent sources (multivariate random vectors) that are directly related to our observations. This problem can be addressed in the context of Blind Source Separation (BSS). The goal of BSS is to reconstruct underlying, interpretable latent signals from observationable data. Most often this is achieved via a linear transformation of the observed data together with certain assumptions on the source signals.

Principal Component Analysis (PCA) can be used as a preliminary step to BSS for the purpose of dimension reduction. Although the idea behind PCA is rather similar to other methods of BSS, PCA does strictly speaking not belong to BSS. PCA is a simple tool for restructuring multivariate data. The idea of PCA is to find components via an orthogonal transformation matrix. These principal components are ranked by how much variance of the original data can be explained by the individual components. Not all principal components have to be employed in the subsequent modeling approaches of BSS and consequently it can serve as a tool for dimensionality reduction. This corresponds to only considering a subselection of principal components for further analysis and facilitates the analysis of high-dimensional data while capturing the most important aspects of the observations (linear combinations of data, that explain most of the variation). In the following we will shortly introduce PCA and subsequently two different methods for BSS, Independent Component Analysis (ICA) and Spatial Blind Source Separation (SBSS).

A remark on our notation: in the following $\mathbf{x}(i)$ will refer to random vectors and $\mathbf{x}_i$ to realizations (or observations) of said random vectors. In matrix notation $\mathbf{X}$ will refer to realizations of an array of random vectors unless explicitly stated otherwise.

3.1 Principal Component Analysis

PCA goes back to Hotelling in the 1930s, although Pearson already published a similar concept back in 1901 [12]. The goal of PCA is to find an orthogonal basis with a linear transformation (matrix) that transforms observations into uncorrelated principal components. The transformation is determined by the directions of the greatest variance of the original data, always with the constraint of orthogonality. Here orthogonality corresponds to un-correlatedness. We compare the results of PCA, which is a common tool in geological sciences, with the results of our further analysis to determine whether the follow-up methods offer any benefit for understanding the data.

3.1.1 PCA Framework

Our aim is to find an orthogonal transformation of the data such that $\mathbf{x}_i = \mathbf{A}\mathbf{s}_i$. Where $\mathbf{x}_i$ are vector-valued observations, $\mathbf{s}_i$ are the underlying principal components and $\mathbf{A}$ is an orthogonal matrix. We assume that $E\left(\mathbf{x}\left(i\right)\right) = \boldsymbol{\mu} = \mathbf{0}$ and therefore center the observed data $\mathbf{x}_i$ by their empirical means $\bar{\mathbf{x}}$. The mean adjustment is the first step of our PCA (or later on ICA) and is not explicitly stated every time. We refrain from including the parameter $\boldsymbol{\mu}$ in future equations and always assume the mean adjustment as prior step to PCA (and ICA). The goal of PCA is to find an orthogonal matrix $\mathbf{W}$ with $\mathbf{W} = \mathbf{A}^{-1}$. In matrix notation this corresponds to:

$$\mathbf{S} = \mathbf{W}\mathbf{X}, \tag{3.1}$$

with the realization of the random vectors $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}$ with n observations, each observation consists of a p-variate vector $\mathbf{x}_i$ for $i = 1, \ldots, n$, the principal components as vectors of $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_n) \in \mathbb{R}^{p \times n}$. The rows of $\mathbf{S}$ are called principal components. The matrix $\mathbf{W} \in \mathbf{O}(p)$, where $\mathbf{O}(p)$ is the set of all orthogonal matrices of dimension $p \times p$. The matrix $\mathbf{W}$ denotes the orthogonal loadings matrix that transforms $\mathbf{X}$ into $\mathbf{S}$.

PCA constructs an orthogonal transformation matrix $\mathbf{A}$, hence $\mathbf{W}$ exists. Furthermore, for the principal components we assume that $E\left(\mathbf{s}\left(i\right)\right) = \mathbf{0}, E\left(\mathbf{s}\left(i\right)\mathbf{s}\left(i\right)^T\right) = \boldsymbol{\Sigma}_{\mathbf{s}} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix. Thus the principal components are usually assumed to be un-correlated and their variances (diagonal elements of $\mathbf{D}$) differ. The component with the greatest variance is the first principal component. All components can be ranked based on their variance. If the variance of all principal components are equal, no meaningful dimensionality reduction can be achieved by PCA and $\mathbf{W}$ is not well-defined.

3.1.2 Derivation Of Un-Mixing Matrix

The transformation matrix $\mathbf{W}$ rotates the data $\mathbf{X}$. We could also interpret $\mathbf{W}$ as a set of vectors presenting $\mathbf{X}$ in a new orthogonal basis. The idea behind PCA is that the direction of the greatest variance of the data is the most important one and the structure of the data in said direction provides the most information. For the second direction we choose the one with the second highest variance of the data that is orthogonal to the first direction, and so on. The principal components of PCA (the rows of $\mathbf{S}$) are therefore in the order of the explained variance of the original data. Therefore, if it is a goal to reduce the dimension of the transformed data $\mathbf{S}$, we can simply select the first components and achieve the greatest variability of the original data for the selected dimension.

For the de-correlation of the data we look at the estimated covariance matrix $\mathbf{R}_{\mathbf{X}}$ of zero mean $\mathbf{X}$:

$$\mathbf{R}_{\mathbf{X}} = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T,$$

with the variances of each row of $\mathbf{X}$ as the diagonal elements and the covariances of the ith and jth row in the ijth entry. Since $\mathbf{R_X} \in \mathbb{R}^{p \times p}$ is a covariance matrix, it is symmetric. The principal components $\mathbf{S}$ are required to be un-correlated and to have zero mean. Their covariance matrix $\mathbf{\Sigma_s}$ is then a diagonal matrix and their empirical covariance matrix $\mathbf{R_S}$ can be expressed through the covariance matrix of the observed data $\mathbf{R_X}$:

$$\mathbf{R_S} = \frac{1}{n-1}\mathbf{S}\mathbf{S}^T = \frac{1}{n-1}\left(\mathbf{W}\mathbf{X}\right)\left(\mathbf{W}\mathbf{X}\right)^T = \frac{1}{n-1}\mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T = \mathbf{W}\mathbf{R_X}\mathbf{W}^T.$$

Amongst others, there are two common approaches to calculate $\mathbf{W}$. Firstly with the eigenvalue-decomposition. Applying the standard eigenvalue-decomposition, the covariance matrix of $\mathbf{X}$ can be expressed as the product of orthogonal matrix $\mathbf{B}$, containing the eigenvectors, and a diagonal matrix $\mathbf{D}$, containing the eigenvalues. We choose our transformation matrix $\mathbf{W} = \mathbf{B}$, so the covariance of our signals $\mathbf{S}$ becomes diagonal and we have found a feasible solution for the transformation matrix.

$$\mathbf{R_X} = \mathbf{B}^T\mathbf{D}\mathbf{B}$$
$$\mathbf{R_S} = \mathbf{W}\mathbf{R_X}\mathbf{W}^T = \mathbf{B}\mathbf{B}^T\mathbf{D}\mathbf{B}\mathbf{B}^T = \mathbf{D}.$$

The other way to calculate the transformation matrix $\mathbf{W}$ is through a generalization of the eigenvalue-decomposition, a so-called Singular Value Decomposition (SVD). We look at the SVD of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{3.2}$$

where $\mathbf{X} \in \mathbb{R}^{p \times n}$, $\mathbf{U} \in \mathbb{R}^{p \times p}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ and $\mathbf{\Sigma} \in \mathbb{R}^{p \times n}$ the diagonal matrix with the rank-ordered singular values, the square-roots of the eigenvalues, as its diagonal elements. The matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal. We choose $\mathbf{W} = \mathbf{U}^T$:

$$
\begin{aligned}
\mathbf{R_S} &= \mathbf{W}\mathbf{R_X}\mathbf{W}^T \\
&= \mathbf{U}^T\left(\frac{1}{n-1}\mathbf{X}\mathbf{X}^T\right)\mathbf{U} \\
&= \frac{1}{n-1}\mathbf{U}^T\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\right)\mathbf{U} \\
&= \frac{1}{n-1}\mathbf{U}^T\left(\mathbf{U}\mathbf{D}\mathbf{U}^T\right)\mathbf{U} = \frac{1}{n-1}\mathbf{D}.
\end{aligned}
$$

where $\mathbf{D} = \mathbf{\Sigma}^2$ is a diagonal matrix with the squared singular values of $\mathbf{X}$ (the eigenvalues of $\mathbf{X}\mathbf{X}^T$) in the diagonal entries. Those eigenvalues are the variances of the principal components $\mathbf{S}$. With this we have the desired diagonal structure of the variance-covariance matrix of the principal components (as described above). The components are ordered according to their variances (and therefore to their importance in explaining the observed data). PCA is a standard method to reduce the dimensions of a dataset. When we deal with high dimensional datasets, we face the difficulty of a major amount of computational effort and the results become hard to interpret and visualize [26]. Since the principal

components are ordered, the variability of the data is preserved as well as possible by choosing the first components for a given dimensionality q. A common criterion for q is a threshold requirement for the percentage of explained variance of the original data . A reasonable choice for the threshold is 80% [33, p. 4]. The scores $\mathbf{S}^{p \times n}$ reduce to $\mathbf{S}^{q \times n}$, which is ideally considerably smaller while still managing to explain 80% of the variability of the original observed data.

3.2 Blind Source Separation

In Blind Source Separation (BSS) we seek to uncover hidden source signals from observed data. The only assumption we make at this point is, that there is a linear mixing procedure that transforms the source signals into observable sensor signals. Further assumptions on the properties of the source signals (e.g. mutually uncorrelated, independent, ...) and the properties of the mixing model determine if we use second order statistics or higher order statistics (HOS) and which BSS method is the most suitable one for a specific application [33, p. 6 & 43].

A common example for BSS in the literature is the so-called "cocktail party problem". At a party there are multiple people talking, music is playing, and microphones are spread throughout the room. For simplicity, we suppose that the number of sound sources does coincide with the number of microphones. Each microphone records a linear mixture of the conversations and music. Depending on the distance between a microphone and each sound source, the recordings capture a different weighted mixture of the overlaying and concurrent chats, rendering the underlying conversations impossible to understand. The problem becomes: How can we un-mix the recordings in order to separate the individual speakers and the music (we will call those source signals) and thus be able to follow the actual conversations instead of getting lost in the superposition of speakers, sounds and noise (we will call those sensor signals)?

A simple form of a linear BSS model following the notation of [27] writes as follows:

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i + \boldsymbol{\mu}, \tag{3.3}$$

with p-variate vectors $\mathbf{x}_i$ for $i = 1, \ldots, n$ being observed signals - realizations of a random vector, with p-variate latent vectors $\mathbf{s}_i$ for $i = 1, \ldots, n$ being hidden source signals (or scores), with a p-variate location vector $\boldsymbol{\mu}$ and $\mathbf{A} \in \mathbb{R}^{p \times p}$ the full-rank mixing matrix that transforms the source signals into the sensor signals.

The only information we have are the realizations of the random vectors $\mathbf{X}$. This is not enough information to reasonably construct $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_n) \in \mathbb{R}^{p \times n}$ and $\mathbf{A} \in \mathbb{R}^{p \times p}$. There are a multitude of possible solutions and we have to make additional assumptions to make this model solvable [16, Chapter 1.2].

A common assumption is that the source signals are mutually independent. [16] states that this assumption is "realistic and fully justified in many problems" [33, p. 65]. BSS-

approaches based on this assumption are called Independent Component Analysis (ICA). In the corresponding literature on BSS we often see the terms BSS and ICA used interchangeably [33, p. 70]. While ICA is only one possible way to address a BSS-problem, it is among the best-known approaches [33, p. 47]. In this work we conduct an ICA, where we perform a prior dimensionality reduction in the form of a classical Principal Component Analysis (PCA). This is a standard procedure for filtering noise and reducing the complexity of high-dimensional data.

[16, Chapters 1.1, 1.1.2.2 & 1.1.2.3] discuss the origin of the BSS method. In the 1980s [18] formulated the BSS problem in the context of neural modeling. In the early beginnings only Gaussian data was considered. As Gaussian data is fully characterized by its first two moments, second order statistics were enough to solve the source separation problems. Classical PCA was able to decompose data into un-correlated source signals. For Gaussian data un-correlatedness implies independence. However, there was no known solution for non-Gaussian data. In fact, for many years the scientific community was sure that it was impossible to solve the BSS-problem for non-Gaussian data. ICA can be seen as an enhancement on PCA, where the condition of un-correlated source signal is replaced by independence of the source signals. ICA was first mentioned 1987 [21] and was formalized in 1992 [15].

3.3 Independent Component Analysis

In ICA the linear transformation in form of a full-rank, so-called mixing matrix, results in mutually independent source signals. At most one of these signals can be Gaussian and they are assumed to be independent and identically distributed (iid). The solution is unique up to signs and order.

3.3.1 Model Formalization

Looking at the model equation (3.3) in matrix notation and $\mathbf{X}$ already adjusted by its mean row-wise:

$$\mathbf{X} = \mathbf{AS}, \tag{3.4}$$

where $\mathbf{X} \in \mathbb{R}^{p \times n}$ and $\mathbf{S} \in \mathbb{R}^{p \times n}$. The rows of $\mathbf{S}$ are called independent components. The linear transformation from $\mathbf{S}$ to $\mathbf{X}$ is given by an unknown matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$. The range of $\mathbf{A}$ represents the sub-space of the source signals [16, Chapter 5.2.1].
Again with $\mathbf{W} = \mathbf{A}^{-1}$ we reformulate the model equation and arrive at equation (3.1):

$$\mathbf{S} = \mathbf{WX}.$$

At this point we remember that we have conveniently assumed that both $\mathbf{X}$ and $\mathbf{S}$ consist of p-variate data and therefore $\mathbf{A}$ must have the form of a $p \times p$-square matrix.
The goal of ICA is to estimate the un-mixing matrix $\mathbf{W}$. This model as stated is still not solvable in a unique way. First we need to make some additional assumptions.

3.3.2 Model Assumptions

(ICA 1) The mixing matrix $\mathbf{A}$ is invertible.

(ICA 2) $E\left(\mathbf{s}\left(i\right)\right) = \mathbf{0}, E\left(\mathbf{s}\left(i\right)\mathbf{s}\left(i\right)^{T}\right) = \mathbf{\Sigma_s} = \mathbf{I_p}$, hence $\mathbf{S}$ is whitened.

(ICA 3) The components of $\mathbf{s}(i)$ are mutually independent for $i = 1, \ldots, n$.

(ICA 4) At most one component of $\mathbf{s}\left(i\right)$ is Gaussian.

Some notes on and implications of these assumptions are:

(ICA 2) Signals satisfying these assumptions are called white signals. Whitening is a pre-processing step of ICA. White data is un-correlated, since the second-order dependencies (the off-diagonal elements of the covariance matrix $\mathbf{\Sigma_s}$) are zero. The variances (diagonal elements of the covariance matrix $\mathbf{\Sigma_s}$) are normalized. This is a convenient assumption to fix the scales of the components.

(ICA 3) Assuming independence of the source signals at any arbitrary sample index n provides a criterion for separability of the source signals with the application of Higher Order Statistics (HOS). In many fields this assumption is not overly restrictive [33, p. 65]. Formally put, when $f_{\mathbf{s}(i)}$ denotes the Probability Density Function (PDF) of the source signal $\mathbf{s}\left(i\right)$, then statistical independency of the p source signals means their joint density is composed of the product of the marginal densities of the components of the source signal. The additional assumption that the source signals are independent and identically distributed (iid) is commonly employed and hints at the fact that a possible dependence structure of $\mathbf{S}$ is not considered explicitly in ICA [16, Chapter 1.3.2].

(ICA 4) It was shown by Darmois that iid-processes that are additionally normally distributed have no ICA solution [16, Chapter 1.3]. During the pre-processing the data will be whitened, hence empirical correlations are removed and no second-order dependencies remain. Since for Gaussian data un-correlatedness implies independence, there is no useful information in the higher-order moments, we are not able to find a criterion for separation other than the classical PCA solution [33, p. 67].

3.3.3 Derivation Of Un-Mixing Matrix

Given the four assumptions above, the un-mixing matrix $\mathbf{W}$ can be estimated and the observable data $\mathbf{X}$ can be transformed into the desired unobservable signals $\mathbf{S}$ up to signs and order. ICA usually is a two-step procedure. The un-mixing matrix $\mathbf{W}$ is decomposed into a product of matrices via SVD. These matrices firstly whiten $\mathbf{X}$ and secondly rotate the standardized $\mathbf{X}$ in a way that the resulting source signals become independent.

We perform SVD on the mixing matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

and therefore $\mathbf{W} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T$.

with $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{V} \in \mathbb{R}^{p \times p}$, $\mathbf{U} \in \mathbb{R}^{p \times p}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ are per construction orthogonal rotation matrices, where the inverse equals the transposed matrix $\mathbf{U}^T = \mathbf{U}^{-1}$ and $\mathbf{V}^T = \mathbf{V}^{-1}$. The matrix $\boldsymbol{\Sigma}$ is diagonal, where the inverse exists because of our assumption that $\mathbf{W}$ exists.

First, we use second-order statistics to determine $\mathbf{U}$ and $\boldsymbol{\Sigma}$. This transformation will whiten $\mathbf{X}$.

Look at the SVD of the covariance matrix of zero mean $\mathbf{X}$ and the model equation (3.4):

$$
\begin{aligned}
\mathbf{R_X} &= \frac{1}{n-1}\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{n-1}\left(\mathbf{A}\mathbf{S}\right)\left(\mathbf{A}\mathbf{S}\right)^T \\
&= \frac{1}{n-1}\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right)\mathbf{S}\mathbf{S}^T\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right)^T \\
&= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{R_S}\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T \\
&= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T \\
&= \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T.
\end{aligned}
$$

Where we used the assumption that $\mathbf{R_S} = \mathbf{I_p}$ and $\mathbf{V}^{-1} = \mathbf{V}^T$. The result $\mathbf{R_X} = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T$ is now independent of $\mathbf{S}$ and $\mathbf{V}$ and corresponds to the eigen-decomposition of the covariance matrix $\mathbf{R_X} = \mathbf{B}\mathbf{D}\mathbf{B}^T$.

Therefore we choose $\mathbf{U} = \mathbf{B}$ to be the eigenvectors of the empirical covariance matrix $\mathbf{R_X}$ and $\boldsymbol{\Sigma} = \mathbf{D}^{1/2}$ to be the diagonal matrix with the square-root of the eigenvalues as its diagonal elements.

Going back to the SVD of $\mathbf{W}$ with $\mathbf{U} = \mathbf{B}$ and $\boldsymbol{\Sigma} = \mathbf{D}^{1/2}$:

$$\mathbf{W} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{B}$$

$$\mathbf{W}\mathbf{X} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{B}\mathbf{X} = \mathbf{S}.$$

The de-correlation in form of the diagonal covariance matrix $\mathbf{R_{BX}}$ corresponds to PCA described in the previous chapter in detail. The matrix $\mathbf{D}^{-1/2}$ normalizes the data to unit-variance for each component. Together we have whitened the data $\mathbf{X}$ and will write $\mathbf{X}_w = \mathbf{D}^{-1/2}\mathbf{B}\mathbf{X}$.

For $\mathbf{X}_w$ it holds that $\mathbf{R_{X_w}} = \frac{1}{n-1}\mathbf{X}_w\mathbf{X}_w^T = \mathbf{I_p}$ and our model equation reduces to:

$$\mathbf{S} = \mathbf{V}\mathbf{X}_w. \tag{3.5}$$

So far we transformed the problem of finding an unknown un-mixing matrix into finding a product of three matrices via SVD. These new matrices are either orthogonal or diagonal matrices, which have special characteristics that reduce the complexity of the problem. Two out of three matrices we obtain with second-order statistics [16, Chapter 1.4]. After the pre-whitening we only need to determine the orthogonal matrix $\mathbf{V}$. PCA is part of the pre-whitening and uses second-order estimations, where less estimation uncertainty is involved than for HOS estimations. On the other hand, since it is unlikely that the actual signals follow a Gaussian distribution, interesting information is often represented in the higher order information [33, p. 71]. Therefore, when including HOS information, as a trade-off we limit our estimation procedure to the 4th-order cumulant only [16, Chapter 5.1.2]. We then find a matrix $\mathbf{V}$ that separates the whitened data into independent source signals (in terms of product moments up to a certain order). There are different approaches to estimating $\mathbf{V}$. We focus on the "Joint Approximate Diagonalization Of Eigen-Matrices" (`JADE`) approach, which we will use in our analysis. The algorithm of `JADE` was introduced at the beginning of the 1990s [13]. We follow the structure and notation of [27] and [13].

Joint Approximate Diagonalization Of Eigen-Matrices (JADE)

We look at the 4th-order cumulant matrix of a whitened p-variate random vector $\mathbf{x}_w$, $\mathbf{C}(\mathbf{Q})$ for any $p \times p$ matrix $\mathbf{Q}$:

$$\mathbf{C}(\mathbf{Q}) = [(\mathbf{x}_w^T \mathbf{Q} \mathbf{x}_w) \mathbf{x}_w \mathbf{x}_w^T] - \mathbf{Q} - \mathbf{Q}^T - \operatorname{tr}(\mathbf{Q}) \mathbf{I_p}. \tag{3.6}$$

The cumulants of the random vectors can be put into "cumulant slices" via $\mathbf{C}(\mathbf{E}^{ij})$ where $\mathbf{E}^{ij} := \mathbf{e}_i \mathbf{e}_j^T, i, j = 1, \ldots, p$. The column-vector $\mathbf{e}_i$ has only zeros and a one in the ith entry and $\mathbf{E}^{ij}$ is a matrix of zeros and a one at the ijth element. [13]. Then $\mathbf{C}(\mathbf{E}^{ij})$ are the 4th order joint cumulants of $\mathbf{X}_w$ (the whitened observed signals from 3.5) and can be seen as linear combinations of "parallel cumulant slices" where the coefficients are entries of $\mathbf{E}^{ij}$.

Cumulants function as a measure of non-Gaussianity or respectively independence, where mutually independent variables result in a zero cumulant [16, Chapter 5.1.2]. Therefore, diagonalizing the cumulant matrices for all matrices $\mathbf{E}^{ij}$ corresponds to removing dependencies among components of the random vector in terms of higher-order mixed moments. Here, mixed moments correspond to centralized mixed moments, as the components of the random vector are assumed to have mean zero. With $\mathbf{x}_w^i$ denoting the i-th component of the random vector $\mathbf{x}_w$, diagonalizing the cumulant matrices mentioned above translates to setting $E\mathbf{x}_w^{i\,I}\mathbf{x}_w^{j\,J}\mathbf{x}_w^{k\,K}\mathbf{x}_w^{l\,L} = 0$, where $i, j, k, l = 1, \ldots, p$ and the set of exponents $(I, J, K, L) \in \mathcal{I} = \{(I, J, K, L) : I, J, K, L = 0, 1, 2, 3; I + J + K + L = 4\} \backslash \{(2, 2, 0, 0), (2, 0, 2, 0), (2, 0, 0, 2), (0, 2, 2, 0), (0, 2, 0, 2), (0, 0, 2, 2)\}$. This means that these mixed moments of the components of the random vector are equal to zero, with the exception of $E\mathbf{x}_w(i)^2\mathbf{x}_w(j)^2 = 1$, where $i, j = 1, \ldots, p$. This translates to, assuming the components are independent, and hence the product moment just being the product of the moments, their variance being equal to unity.

In order to obtain independent source signals $\mathbf{s}(i)$ we want to diagonalize the cumulant matrices: $\mathbf{C}(\mathbf{E}^{ij})$. The reason for joint diagonalization of the cumulant matrices is that

the cumulant matrices are dependent among themselves and the same elements (mixed moments) occur in multiple matrices. By changing one, elements in other cumulants matrices are changed simultaneously. Hence they need to be jointly diagonalized, not iteratively. Let $\mathbf{V}$ be the orthogonal matrix that minimizes the sum of the squared off-diagonal. First note that:

$$\sum_{i=1}^{p}\sum_{j=i}^{p}\|\text{diag}\left(\mathbf{V}\mathbf{C}\left(\mathbf{E}^{ij}\right)\mathbf{V}^{T}\right)\|^{2} + \sum_{i=1}^{p}\sum_{j=i}^{p}\|\text{off}\left(\mathbf{V}\mathbf{C}\left(\mathbf{E}^{ij}\right)\mathbf{V}^{T}\right)\|^{2} = \sum_{i=1}^{p}\sum_{j=i}^{p}\|\mathbf{C}\left(\mathbf{E}^{ij}\right)\|^{2}. \tag{3.7}$$

Instead of minimizing the off-diagonal elements of the fourth-order cumulants, we maximize the diagonal elements, see [16, Chapter 5.4] and [23, Section 2.2]. We choose $\mathbf{V}$ to be the orthogonal rotation matrix that maximizes the approximation criterion:

$$c\left(\mathbf{V}\right) = \sum_{i=1}^{p}\sum_{j=i}^{p}\|\text{diag}\left(\mathbf{V}\mathbf{C}\left(\mathbf{E}^{ij}\right)\mathbf{V}^{T}\right)\|^{2}. \tag{3.8}$$

Under the criterion (3.8) the matrices $\mathbf{C}\left(\mathbf{E}^{ij}\right)$ are approximately jointly diagonalized. If the model is true, the cumulant matrices can be fully diagonalized at the population level, but only approximately for a given realization. Hence we need the approximation criterion (3.8) [16, Chapter 5.2.5]. This line of action motivates the name of the JADE algorithm, "Joint Approximate Diagonalization Of Eigen-Matrices". JADE seeks to diagonalize all estimated fourth-order cumulants as well as possible to make the source signals as "independent" as possible. The joint diagonalizer then becomes the orthogonal matrix $\mathbf{V}$ that maximizes (3.8) and thereby jointly diagonalizes multiple cumulant matrices. The optimal $\mathbf{V}$ corresponds to the last missing part of the ICA source signal solution. Different algorithms exist to approximate the diagonalization of the matrices $\mathbf{C}\left(\mathbf{E}^{ij}\right)$, a popular one is the use of Jacobi iteration with the identity matrix $\mathbf{I_p}$ as starting point for $\mathbf{V}$ [27, Section 5.3] and [23, Section 2.2]. The matrix $\mathbf{V}$ then is part of our un-mixing matrix $\mathbf{W} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{B}$. While in PCA the principal source signals are ordered by their variance in decreasing order, JADE will use the 4th moments of the independent components to order them.

3.3.4 Uniqueness Of Solution

When we talk about a solution for ICA we have to keep in mind that by construction the ICA solution is not unique. Factors of ambiguity are the order of the source signals (permutation) and their signs. For a detailed explanation of this see [27, Section 3.1].

Remember, it is reasonable to make the assumption (ICA 2) that the covariance matrix $\mathbf{\Sigma_s} = \mathbf{I_p}$ to fix the scale beforehand.

Although we will not make an explicit mention of these "internal ambiguities" every time we talk about a solution, we take them into account when interpreting the results. The nature of this constructional ambiguities is presented in detail in [33, p. 40-41].

3.3.5 Applications

Blind Source Separation and especially ICA have been applied successfully in numerous scientific disciplines. [16] dedicates Chapter 1.5 to an overview of ICA applications and go further into detail at the end of the book. Three domains are mentioned as standard applications: biomedical, audio and communication applications. Other applications are image processing, monitoring and pre-processing for classification. Besides the wide field of application possibilities, ICA performs for various types of underlying data, such as time series, matrix and tensor valued data or functional data. For a more detailed discussion see [27].

3.4 Spatial Blind Source Separation

[28] (see also [8]) suggest an extension of BSS that makes use of spatial information. Therefore, from now on we will assume that the data we work with is always spatial data and we extend the concept of random vectors to random fields. When we look at spatial data this means our observations are associated with locations. It is reasonable to expect spatial dependencies and hence that observations close to each other are more similar than those apart. Contrary to PCA and ICA, Spatial Blind Source Separation (SBSS) explicitly considers this spatial dependence structure of the random field when estimating the latent components. To take the spatial structure into account, local covariance matrices are estimated and jointly diagonalized.

These local covariance matrices are estimated for observations within a fixed spatial distance. We use for example balls (circles in case of a two-dimensional spatial domain) and rings around local observations. One key aspect is to find a reasonable radius delta δ for these balls (circles) or rings, up to which spatial dependencies are taken into account.

3.4.1 Preliminary Definitions

Random Fields

In our analysis we assume that the observations (later denoted as $\mathbf{X}$) are realizations of a random field F. A random field F is a family of random vectors F_t defined on the same probability space [32]:

$$\{F_t : t \in T\}, \tag{3.9}$$

where T denotes the index set. Since it suffices for our purposes and out of convenience, we assume that $T \subseteq \mathbb{R}^2$ refers to a spatial domain. In our analysis the spatial coordinates are fixed and not random and for a given index $l \in T$, $\mathbf{x}(l)$ is a real-valued p-variate random vector. The p-variate realizations are used as multivariate spatial random vectors in the context of Blind Source Separation. In the context of Spatial Blind Source Separation we will assume second order-stationarity of the latent random field, meaning the mean and

variance are constant across locations:

$$E[\mathbf{x}(l)] = \boldsymbol{\mu} \in \mathbb{R}^p.$$

$$COV[\mathbf{x}(l)] = E[(\mathbf{x}(l) - \boldsymbol{\mu})(\mathbf{x}(l) - \boldsymbol{\mu})^T] = \boldsymbol{\Sigma}_{\mathbf{x}}.$$

And the autocovariance only depends on the spatial distance:

$$\gamma[\mathbf{x}(l_1), \mathbf{x}(l_2)] = E[(\mathbf{x}(l_1) - \boldsymbol{\mu})(\mathbf{x}(l_2) - \boldsymbol{\mu})^T] = \gamma(l_1 - l_2),$$

With this, the marginal properties of the random field are constant up to the second order, and the autocovariances are independent of shifts in location, only depending on distances.

For a detailed introduction of random fields see [32].

3.4.2 Model Formalization

We follow the structure of [28] and [8] and adopt the familiar BSS model, where we assume the observations to have zero mean ($E(\mathbf{x}(l)) = \mathbf{0}$) and the realizations of the random variables are observed at $l_i \in T, i = 1, \ldots, n$, where l_i correspond to spatial locations. The spatial locations usually, but not exclusively refer to longitude and latitude coordinates:

$$\mathbf{S} = \mathbf{W}\mathbf{X}, \tag{3.10}$$

with the realization of the random field $\mathbf{X} = (\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_n}) \in \mathbb{R}^{p \times n}$. Each observation $\mathbf{x}_{l_i}$ at the ith spatial point consists of a p-variate vector $(x_{1l_i} \ldots x_{pl_i})^T$. And with the unobservable source signal matrix $\mathbf{S} = (\mathbf{s}_{l_1}, \ldots, \mathbf{s}_{l_n}) \in \mathbb{R}^{p \times n}$, where $\mathbf{s}_{l_i}$ at the ith spatial point consists of a p-variate vector $(s_{1l_i} \ldots s_{pl_i})^T$. We call the rows of $\mathbf{S}$ (latent random field) the SBSS components. The matrix $\mathbf{W} \in \mathbb{R}^{p \times p}$ denotes the un-mixing matrix of the model which transforms $\mathbf{X}$ into $\mathbf{S}$.

3.4.3 Model Assumptions

(SBSS 1) $E(\mathbf{s}(l_i)) = \mathbf{0} \ \forall l_i \in T$.

(SBSS 2) $COV(\mathbf{s}(l_i), \mathbf{s}(l_i)) = E\left(\mathbf{s}(l_i)\mathbf{s}(l_i)^T\right) = \boldsymbol{\Sigma}_{\mathbf{s}} = \mathbf{I_p} \ \forall l_i \in T$.

(SBSS 3) $COV(\mathbf{s}(l_i), \mathbf{s}(l_j)) = E\left(\mathbf{s}(l_i)\mathbf{s}(l_j)^T\right) = \gamma(l_i, l_j) = \gamma(h), \ \forall l_i \in T$ with $\gamma(h)$ being a diagonal matrix and the diagonal elements only depends on the spatial distance $h = (l_i - l_j)$.

In (SBSS 1) and (SBSS 2) we repeat the assumptions of the general BSS model. The means of the source signals are assumed to be zero for convenience, the source signals are uncorrelated and the scales are fixed to unit-variance. Assumption (SBSS 3) says that the source signals have no cross-dependence, but there is correlation among individual components of the random vector at different locations. The process $\mathbf{s}(l_i)$ is therefore second-order stationary [28, Section 2], [8, Section 2].

3.4.4 Derivation Of Un-Mixing Matrix

As with the general BSS model we exploit the fact, that after whitening $\mathbf{X}$ we are confronted with the updated model equation:

$$\mathbf{S} = \mathbf{V}\mathbf{X}_w. \tag{3.11}$$

The whitening corresponds to second-order procedures. In contrast to ICA we estimate $\mathbf{V}$ not by jointly diagonalizing 4th-order cumulants but local covariance matrices of $\mathbf{X}$ which are a measure of local dependence. With the diagonalisation of the local covariance matrix we find a local neighborhood that maximizes the spatial correlation within [8, Section 1].

Joint Diagonalisation of Local Covariance Matrices

The local covariance matrix is given by:

$$\mathbf{M}\left(f_\delta\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} f_\delta\left(l_i - l_j\right)\mathbf{x}\left(l_i\right)\mathbf{x}\left(l_j\right)^T, \text{ for a given } \delta>0. \tag{3.12}$$

Where f_δ denotes a kernel function that determines to what extent neighboring realizations are taken into account for the estimation of the local covariance matrices. Possible kernels include Gaussian kernels, balls and rings around the center points. The kernels are given by:

$$f_{\delta(Ball)}(\|l_i - l_j\|; \delta) = \mathcal{I}(\|l_i - l_j\| \leq \delta) \tag{3.13}$$

$$f_{\delta(Ring)}(\|l_i - l_j\|; (\delta_u, \delta_{u+1})) = \mathcal{I}(\delta_u \leq \|l_i - l_j\| \leq \delta_{u+1}) \tag{3.14}$$

$$f_{\delta(Gauss)}(\|l_i - l_j\|; \delta) = \exp(-0.5(\Phi^{-1}(0.95)\|l_i - l_j\|/\delta)^2) \tag{3.15}$$

Note that in the case of the kernel function being a simple indicator function $\mathcal{I}$, which checks if the distances between points are below a thresholds δ (balls with radius δ), in the case of $\delta = 0$, the local covariance matrix reduces to the ordinary covariance matrix:

$$\mathbf{M}\left(f_0\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}\left(l_i\right)\mathbf{x}\left(l_i\right)^T. \tag{3.16}$$

In the case of ring kernels, the orthogonal matrix $\mathbf{V}$ from equation (3.11) is obtained similarly to `JADE` by jointly diagonalizing the local covariance matrices $\mathbf{M}\left(f_1\right), \ldots, \mathbf{M}\left(f_k\right)$ of the whitened data $\mathbf{X}_w$ for k kernels through maximizing

$$\sum_{u=1}^{k} \|\text{diag}\left(\mathbf{V}\mathbf{M}\left(f_u\right)\mathbf{V}^T\right)\|^2, \tag{3.17}$$

where the ring kernels $f_1, \ldots, f_k$ with $k + 1$ radii are defined as in equation (3.14). The optimization of the joint diagonalization is performed as with ICA in equation (3.8). For a detailed description of SBSS see [8].

4 Implementation

4.1 Data Preparation

For the preparation of the CCI SM dataset we follow [9]. [9] conducted their analysis for observations until spring (SON) 2010. We extend this timespan due to the availability of new data and include observations starting in summer (DJF) 1998 until spring (SON) 2018.

ESA [3] provides the soil moisture dataset in the form of a netcdf-file. The individual observations of mean soil moisture (soil moisture anomaly respectively) are associated with a spatial location (longitude, latitude) and a time and hence represent spatio-temporal data. Longitude and latitude dimensions span a grid, where the grid-points mark the locations where the soil moisture observation have been measured. The dimension time is given in days since the first measurement. The dataset has a monthly temporal resolution and for readability and further analysis we transform the day-count into the `POSIXct`-format. The code for reading a netcdf-file in `R` [29] and the conversion into a data frame with the respective date as a `POSIXct`-format can be found in the Appendix (7.1). We interpret the time series as multivariate realizations of spatial random vectors which facilitates the application of classical multivariate methods.

According to [9], a lot of spatial observations are necessary (high spatial and temporal resolution) for reliable PCA results. For this reason we discard all measurements from the years before 1998. The dataset before 1998 has low temporal and spatial resolution and we would have to mute single time periods with gaps. However, skipping time periods arbitrarily comes with implications on the results. We avoid this conflict at the price of a shorter observation period. From 1998 onwards, there is no need to mute out any time periods due to the availability of measurements. We work with complete time series of length p. Where p is either $p = 252$ months or $p = 63$ seasons. There is still a relatively large portion of missing observations in the data and we need to find a way to deal with those. The methods we apply and their respective `R`-Packages need complete datasets. We first take a look at possible causes for missing entries. There are spatial grid-points that consist of missing observations for all p observations. The subset of the CCI SM dataset we work with, covers Australia but actually includes fragments outside the main landmass. Grid points where every observation is missing are for example located in the ocean and will be removed entirely. Then there are also grid-points that have a large amount of NaNs throughout the n observations, for example locations of Australian salt lakes or national parks with exceptionally dense vegetation. We define a threshold criterion (at least 75% of possible observations at each site) and only include grid-points that meet the criterion in our analysis. We discard observations at other grid points due to the sparse temporal coverage. The threshold is fixed based on the consideration of an empirical threshold ratio

in [9] . The data from summer (DJF) 1998 till spring (SON) 2018 has $p = 252$ months. We include every grid-point with at least 189 observations and discard the rest.

The remaining missing entries for the time series included in the analysis will be replaced by the `R` function imputePCA from the package `missMDA`. It is designed to impute missing data as a preliminary step for PCA [20].

The resulting data $\mathbf{X} \in \mathbb{R}^{n \times p}$ has the following form:

```
        lon     lat  1998-01-31 1998-02-28 1998-03-31 ...
1  113.625 -24.625 0.13299665 0.12145019 0.13265938 ...
2  113.875 -24.875 0.09817864 0.10561483 0.10170237 ...
3  113.875 -24.625 0.09449693 0.09308374 0.10205750 ...
4  113.875 -24.375 0.09093077 0.09143615 0.09236076 ...
5  113.875 -24.125 0.09681950 0.09842435 0.09336428 ...
```

Note that in the Sections 3.1, 3.3 and 3.4 we operate with observed data in form of $\mathbf{X} \in \mathbb{R}^{p \times n}$ because this notation is consistent with the literature of the field of research. The `R`-Packages we conduct our analysis with, however, require the transposed of the data matrix $\mathbf{X}$. From now on, we will indicate whether we are looking at $\mathbf{X}$ as element of $\mathbb{R}^{n \times p}$ or as element of $\mathbb{R}^{p \times n}$, respectively. In either case with n being the number of observations (grid-points) and the p-sized feature-space (months or seasons).

For the purpose of this thesis we will look at three different versions of the soil moisture data, the Monthly Soil Moisture Means (mSMM) 1998-2018, the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 and the Seasonal Decomposition Soil Moisture Anomalies (sdSMA) 1998-2018. We perform PCA, ICA and SBSS on each of those versions separately and present selected results in Chapter 5. The versions are constructed as follows:

4.1.1 Monthly Soil Moisture Means 1998-2018

For the first analysis we look at SM Means of the CCI dataset from summer (DJF) 1998 till spring (SON) 2018. At each point in space at most $p = 252$ monthly observations are available. The threshold criterion (189 out of 252) for the muting of grid-points yields $n = 10,706$ observations in form of grid-points that are included. We employ PCA as a mean of dimension reduction. We select the first q principal components that jointly explain a certain percentage of variance of the original data. A common choice for such a percentage threshold for dimension reduction is 80% [33, p. 4]. An illustration of the percentage of explained variance of the first ten principal components for the Monthly Soil Moisture Mean data can be found in Figure 4.1:
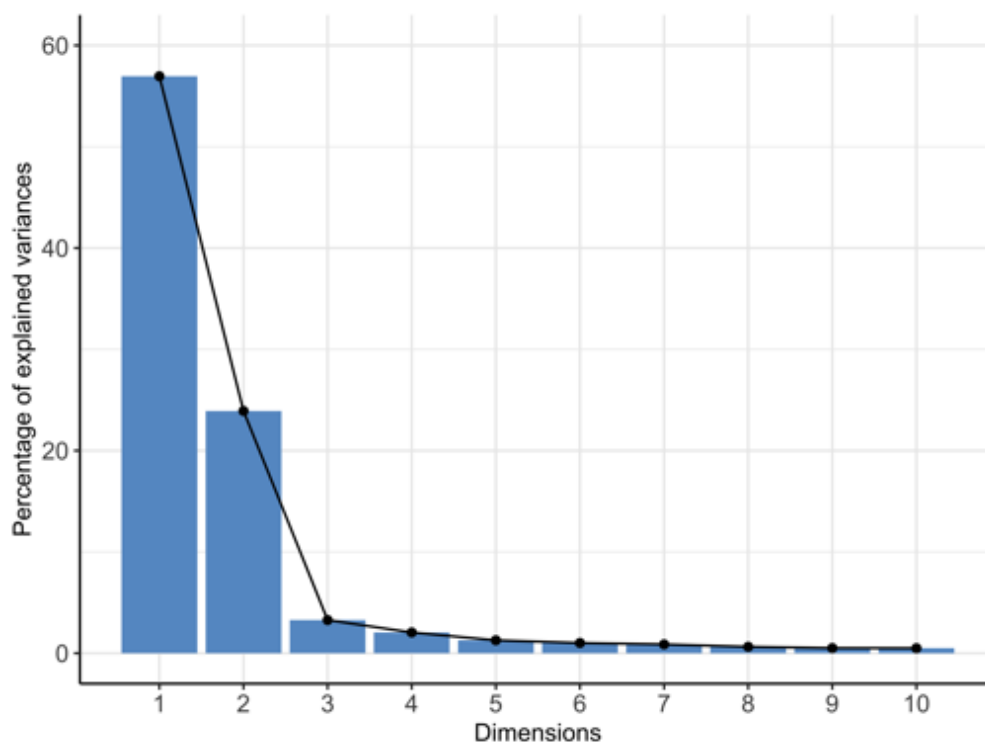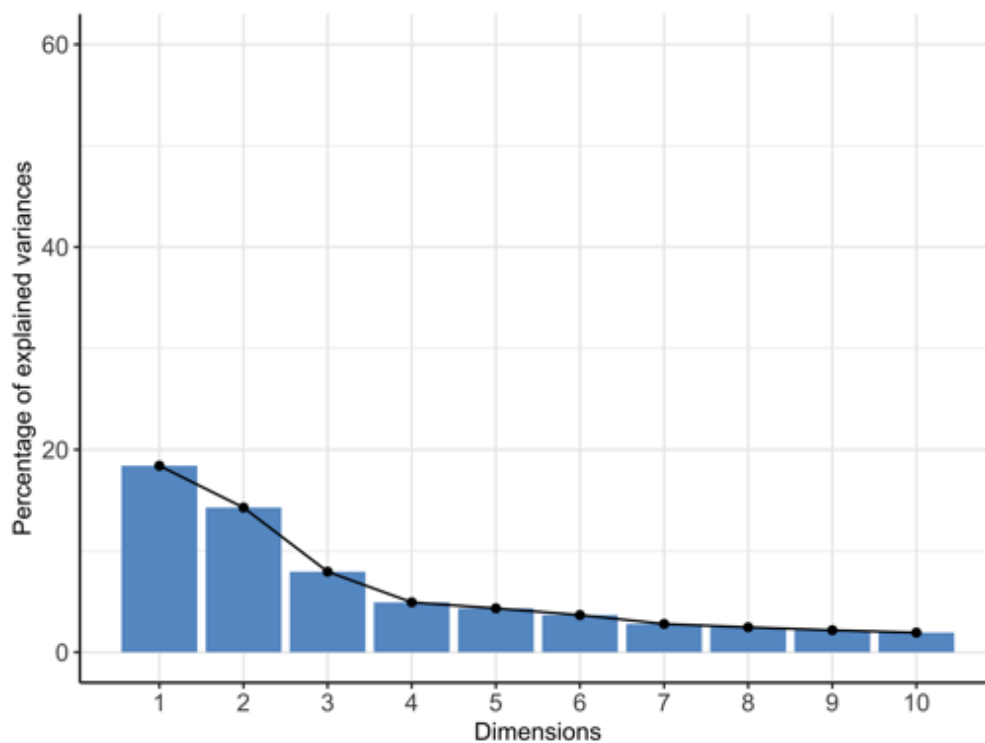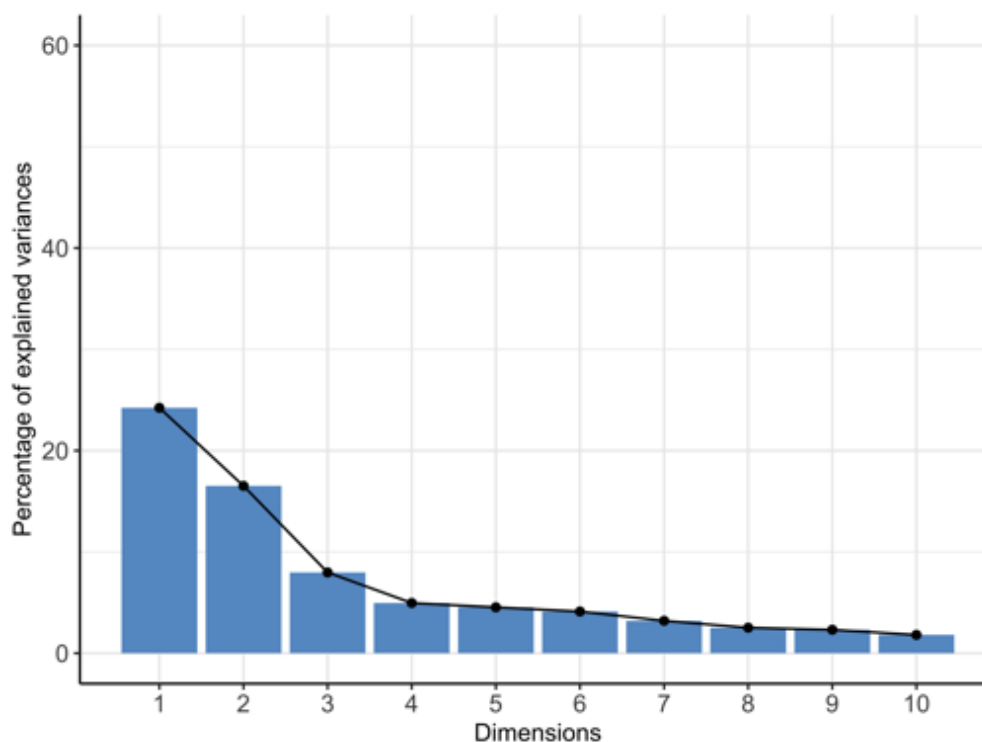
Figure 4.1: The scree plot of the Monthly Soil Moisture Means (mSMM) 1998-2018 shows the percentage of the explained variables of the original data for the first ten principal components.

With this particular dataset we obtain less than ten dimensions when using the unusually strict criterion of 90%. Therefore we will make an exception to the rule and choose to define the threshold at 90% only for this particular version, the Monthly Soil Moisture Means (mSMM) 1998-2018. All this information can be found in Table 4.1.

Table 4.1: Metadata of the Monthly Soil Moisture Means (mSMM) 1998-2018 of the CCI dataset.

Monthly Soil Moisture Means	**mSMM 1998-2018**
Timespan	1998-2018
Temporal Resolution	$p = 252$ Months
Grid Points	$n = 10,706$
Threshold for Pixel Removal	189 (252)
Dimension after PCA (90%)	$q = 8$ (252)

4.1.2 Monthly Soil Moisture Anomalies 1998-2018

In this version of the dataset we analyze the SM anomalies (as calculated in 2.2) of the CCI dataset from summer (DJF) 1998 till spring (SON) 2018. It proved to be beneficial to look at the anomalies of the CCI SM dataset as seasonal effects are eliminated, therefore the remaining structure of the data might be understood better [9]. The dataset consists of $p = 252$ months. And the threshold criterion (189 out of 252) to mute single spatial locations yields $n = 10,486$ grid-points. By employing PCA with a percentage threshold of 80%, we obtain $q = 30$ principal components for the Monthly Soil Moisture Anomalies (mSMA) 1998-2018, see Table 4.2 for a summary and Figure 4.2 for the respective scree plot.

Figure 4.2: The scree plot of the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 shows the percentage of the explained variables of the original data for the first ten principal components.

Table 4.2: Metadata of the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 of the CCI dataset.

Monthly Soil Moisture Anomalies	mSMA 1998-2018
Timespan	1998-2018
Temporal Resolution	$p = 252$ Months
Grid Points	$n = 10,486$
Threshold for Pixel Removal	189 (252)
Dimension after PCA (80%)	$q = 30$ (252)

4.1.3 Seasonal Decomposition of Soil Moisture Anomalies 1998-2018

Some Climate Oscillation Indices (COIs) are known to have an exceptionally large impact on single seasons throughout the year. It might be beneficial to look at a particular season individually to get a better understanding of the influence patterns of a COI on Australian climate. Therefore we look at the Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 (sdSMA), where we perform PCA, ICA and SBSS for each individual season (four times in total). Firstly only on summer data, which corresponds to the months of December, January and February in Australia. Then for the months of fall, then winter and spring. Each analysis uses $n = 10,486$ grid-points and a p-variate feature-space of $p = 63$ available months per season of the whole dataset at hand. The percentage threshold for PCA dimension reduction of 80%, yields $q = 16$ principal components for summer and fall (see scree plot of Figure 4.3), $q = 15$ principal components for winter and $q = 17$ principal components for spring, respectively. A summary of this information can be found in Table 4.3. In Chapter 5 we present selected results of the Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 for the months of summer. A complete list of all results can be found in the Appendix 7.4.

Figure 4.3: The scree plot of the Summer Decomposition Soil Moisture Anomalies (sdSMA) 1998-2018 shows the percentage of the explained variables of the original data for the first ten principal components.

Table 4.3: Metadata of the Seasonal Decomposition Soil Moisture Anomalies (sdSMA) 1998-2018 of the CCI dataset.

Seasonal Decomposition Soil Moisture Anomalies		**sdSMA 1998-2018**
Timespan		1998-2018
Temporal Resolution		$p = 63$ Months p. Season
Grid Points		$n = 10,486$
Threshold for Pixel Removal		189 (252)
Dimension after PCA (80%):	Summer (DJF)	$q = 16$ (63)
	Fall (MAM)	$q = 16$ (63)
	Winter (JJA)	$q = 15$ (63)
	Spring (SON)	$q = 17$ (63)

4.2 Lambert Azimuthal Equal Area Projection

For the entire analysis we make use of the Lambert Azimuthal Equal Area Projection which maps the longitude and latitude coordinates of the spheric Earth onto a disk and

is area-preserving. We choose Australia to be the center point of the projection (located approximately at latitude -25 and longitude 135) . A detailed explanation, further benefits of the projection and the mathematical framework of the projection can be found in [31]. For the implementation we use the R function `spTransform` of the package rgdal [11].

4.3 PCA, ICA and SBSS with corresponding R Packages

Our analysis is performed in R. A selection of function calls can be found in the Appendix 7.2.

We use the R function `prcomp` for PCA which is based on SVD (see Equation 3.2). The principal components are compared to the results of ICA and SBSS. Additionally, we use PCA as a method of dimension reduction. The percentage threshold of the desired explanatory power of the variance of the original data, leaves us with a selection of the first q principal components in the form of a $n \times q$ matrix with real entries. This dimension reduced dataset in form of principal components will become the input for further analysis (ICA and SBSS).

For ICA we use `JADE` which is implemented in the R-Package `JADE` [25]. The mathematical framework of the `JADE` algorithm is explained in Section 3.3.3. Recall that the multivariate realizations of soil moisture data have a spatial structure, this structure is not explicitly considered in `JADE`.

For SBSS we use the package `SpatialBSS` [24] to solve (3.17). We choose ring kernels for modeling the spatial structure of the data (see Equation 3.12). We use $k = 4$ disjoint rings with radii pairs of $\delta \in \{(0, 60), (60, 120), (120, 180), (180, 240)\}$ all in km. around each observation. The outer radius of 240 km is based on the consultation with experts on the field. SBSS explicitly considers the spatial structure of the data.
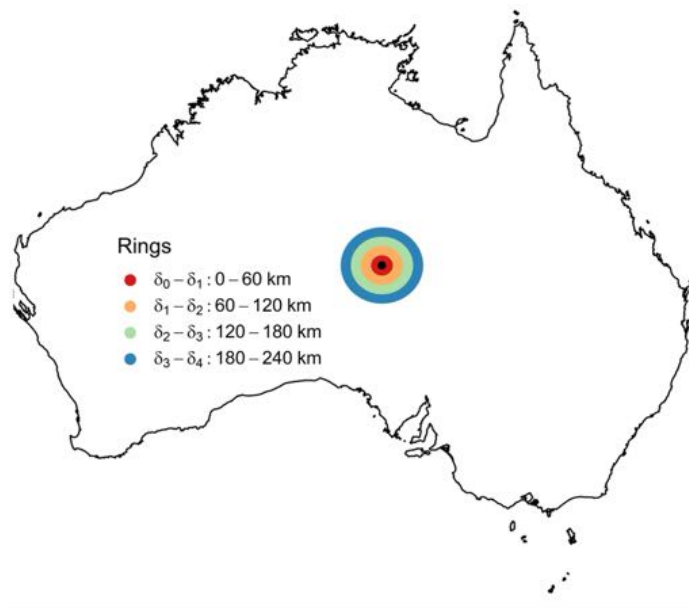
Figure 4.4: Schematic illustration of four disjoint rings around an exemplary starting point with radii pairs of $\delta \in \{(0, 60), (60, 120), (120, 180), (180, 240)\}$.

4.4 Fixing Signs of Components and Scaling

Since PCA will result in components that have arbitrary signs we decided on a criterion to fix the signs to make the results reproducible. We follow the convention used in the R function JADE. Whenever the sum of a column of the transformation matrix $\mathbf{W} \in \mathbb{R}^{p \times n}$ is negative the sign of each element is reversed:

$$w_{\cdot,j} = -w_{\cdot,j}, \text{ if } \sum_{j=1}^{p} w_{i,j} < 0,$$

where $w_{\cdot,j}$ are the elements of the the j-th column of the un-mixing matrix $\mathbf{W}$. We use this criterion to manually fix the signs of the transformation matrices of PCA and SBSS as well.

Note this notation refers to the notation used in Sections 3.1, 3.3 and 3.4, rather than the notation used in the programming environment R.

4.5 Correlation Analysis

In order to quantify the results of our analysis and validate our claims on advantages, disadvantages and the general effectiveness of explaining the Australian climate via PCA, ICA or SBSS, we conduct a Spearman correlation analysis to measure how strong the monotonic relationship between our results (loadings of principal components or independent components) and the Climate Oscillation Indices is. Spearman correlation is commonly

employed to assess the plausibility of a relationship between random variables [17].

The Spearman correlation coefficient is estimated as follows:

$$\hat{\rho}_s = \frac{\sum_{j=1}^{p}\left(R_j^{\mathbf{x}} - \bar{R}^{\mathbf{x}}\right)\left(R_j^{\mathbf{y}} - \bar{R}^{\mathbf{y}}\right)}{\sqrt{\sum_{j=1}^{p}\left(R_j^{\mathbf{x}} - \bar{R}^{\mathbf{x}}\right)^2}\sqrt{\left(R_j^{\mathbf{y}} - \bar{R}^{\mathbf{y}}\right)^2}}$$

where $R_j^{\mathbf{x}}$ is the rank of the j-th observation of the vector $\mathbf{x}$, with 1 being the smallest, the same applies to $R_j^{\mathbf{y}}$. $\bar{R}^{\mathbf{x}}$ is the average rank. After the selection of the first q components for a dimension reduction, the un-mixing matrix of PCA $\mathbf{W}_{PCA} \in \mathbb{R}^{p \times p}$ reduces to $\mathbf{W}_{PCA} \in \mathbb{R}^{q \times p}$. We separately correlate the q rows of $\mathbf{W}_{PCA}$ and the Climate Oscillation Index vectors $\mathbf{v}_{DMI} \in \mathbb{R}^{1 \times p}$, $\mathbf{v}_{SAMI} \in \mathbb{R}^{1 \times p}$ and $\mathbf{v}_{SOI} \in \mathbb{R}^{1 \times p}$ respectively (again, we refer to the notation used in Sections 3.1, 3.3 and 3.4). Here p will always stand for the 252 months from 1998-2018.

For ICA (SBSS) the combined loading matrix (the loading matrix of itself ICA (SBSS) is merely $\mathbf{W}_{ICA(SBSS)}$) consists of the product of the loading matrix of the ICA (SBSS) and the un-mixing matrix of the PCA on the principal components in question $\mathbf{W}_{ICA(SBSS)}\mathbf{W}_{PCA}$. Where $\mathbf{W}_{ICA(SBSS)} \in \mathbb{R}^{q \times q}$ and $\mathbf{W}_{PCA} \in \mathbb{R}^{q \times p}$. Each of the q rows of the new loading matrix $\mathbf{W}_{ICA(SBSS)}\mathbf{W}_{PCA} \in \mathbb{R}^{q \times p}$ is correlated with $\mathbf{v}_{DMI} \in \mathbb{R}^{1 \times p}$, $\mathbf{v}_{SAMI} \in \mathbb{R}^{1 \times p}$ and $\mathbf{v}_{SOI} \in \mathbb{R}^{1 \times p}$ respectively.

It is not unreasonable to think that the effect of COIs on the Australian Climate comes with a certain delay. We therefore consider the cross-correlation between the ranks of the loading vectors of the resulting components of PCA, ICA and SBSS and the ranks of the COIs for different monthly time lags, including 0, 1, 2, 3, 4, 6 and 12 (COIs are shifted into the past). The lags naturally have the same temporal structure as the data and therefore lag = 12 corresponds to a whole year of delay, while lag = 0 stands for an immediate relatedness without delay.

After consultation with subject experts, we consider correlations underneath 0.2 to be not meaningful. For the seasonal decomposition of the data we can expect much higher correlations. The benchmark we use in this case is 0.4.

We want to underline a discrepancy with the correlation analysis in connection with the uncertainty of signs of the PCA results. The uncertainty of the signs of the transformation matrix of PCA propagates to the loadings of the ICA and SBSS components. If one sign is changed the resulting Spearman correlations may differ drastically. To remain within the scope of this work we perform a Monte Carlo study to approximate the maximal correlation. Details on this problem and an instruction for the reconstruction of our presented correlation results can be found in the Appendix 7.3.

5 Results

In the following we present a small selection of the results of PCA, ICA and SBSS on the CCI SM dataset. The complete results in the form of plots and tables can be found in the Appendix 7.4. To reconstruct our findings, permutation indices are given in 7.4 as well. They can be used to determine the optimal signs for the principal components as described in the Appendix 7.3. For now we focus on some interesting findings amongst the principal and independent components. The resulting components are presented separately for each of the three different datasets (means, anomalies and seasonal decompositions) as described in 4.1.1, 4.1.2 and 4.1.3.

For ease of interpretation we choose to display the maps of the resulting components on a shared color scale for each method (PCA, ICA, SBSS) individually.

We start by looking at the quarterly means of the original observations in order to be able to understand and interpret the resulting components.

5.1 Soil Moisture Data Means

The maps of the seasonal CCI Soil Moisture data means in Figure 5.1 look consistent with our expectations. General rainfall regimes are modeled. To confirm the claim, we refer to the homepage of the BOM, where comparable graphics are presented [2]. We identify the summer and fall seasons to be similar to each other as well as the winter and spring seasons. From December till May, we identify high values, which correspond to especially wet regions at the north and east coast. From June till November, the high-value wet regions are found in the south-east coastline. In the winter season, from June till August, there is also a definite wet spot at the south-west edge of Australia. We observe that Central Australia and large parts of the west coast seem relatively dry throughout all seasons and do not follow any seasonal patterns. Seasonal influences to the data means seem to be most prominent to the north and south-east of Australia.
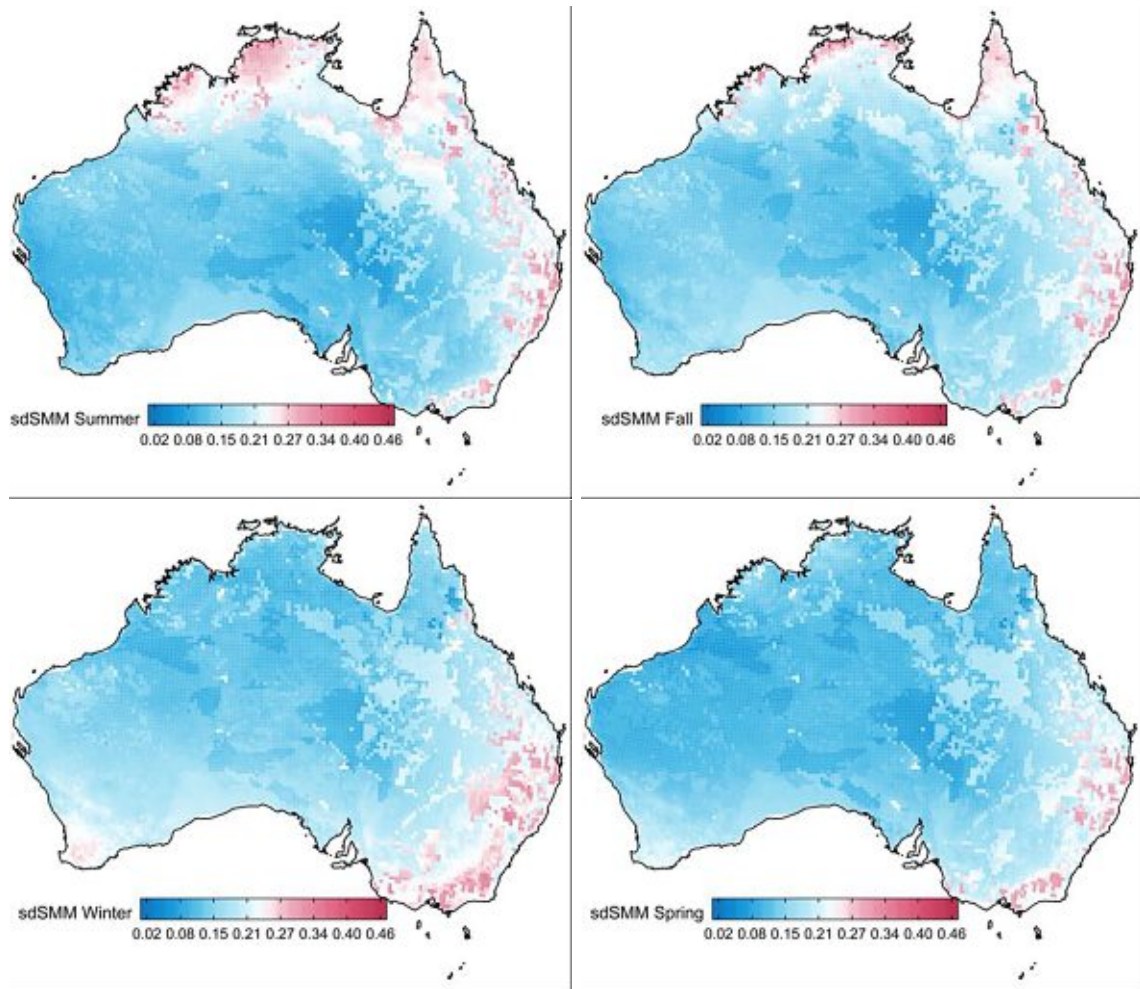
Figure 5.1: Maps of SM Means for Summer (DJF), Fall (MAM), Winter (JJA) and Spring (SON).

5.1.1 Monthly Soil Moisture Means (mSMM) 1998-2018

The PCA of the SM means from 1998 till 2018 with a monthly temporal solution yields eight principal components (see Table 4.1). Note that we use here a threshold criterion of 90% for the determination of the dimension when we usually apply the 80% threshold. For this dataset PCA manages to reproduce the overall data mean, the yearly latitudinal movement of the ITC-zone (Intertropical Convergence Zone) and an east-west division. ICA mirrors weather extrema from 2002 and 2012, which have been extraordinary years in the Australian weather records and the multi-annual cycles of ENSO. An expected east-west division is also reproduced with SBSS as well as notable new patterns.

PC1 (mSMM)

The first principal component (PC1) represents roughly 57% of the variance of the original data and seems to display the overall data mean. Compare therefore Figure 5.2 (left panel) and Figure 5.1 (all panels). We already observed that the data means display for all seasons high-valued wet spots in the east and a thin coastal belt with high values reaching over the entire north coast. In the winter seasons, an additional spot of high values is visible at the south-west coast. All these features reproduce in PC1 in Figure 5.2 (left panel). Please note that the legends in Figure 5.2 (left panel) and Figure 5.1 (all panels) are different as we adapt the scale of the legend to the data range. In this case, we look at the ranges of the SM Mean dataset and the principal components. The right panel of 5.2 affirms the claim of PC1 as the overall data mean. All seasons load positively onto the component and they all seem to contribute evenly. The summer season (yellow dots) seems to dominate slightly. The summer season is mainly associated with impacting the northern coastal belts (see Figure 5.1 (top left panel)). About the uncertainty of the "true sign" of each component (compare Section 4.4), we can strongly assume that the PC1 displays its "true sign". As the displayed pattern can be explained perfectly with the given sign and we know that PCA tends to return an overall mean in one component as well.
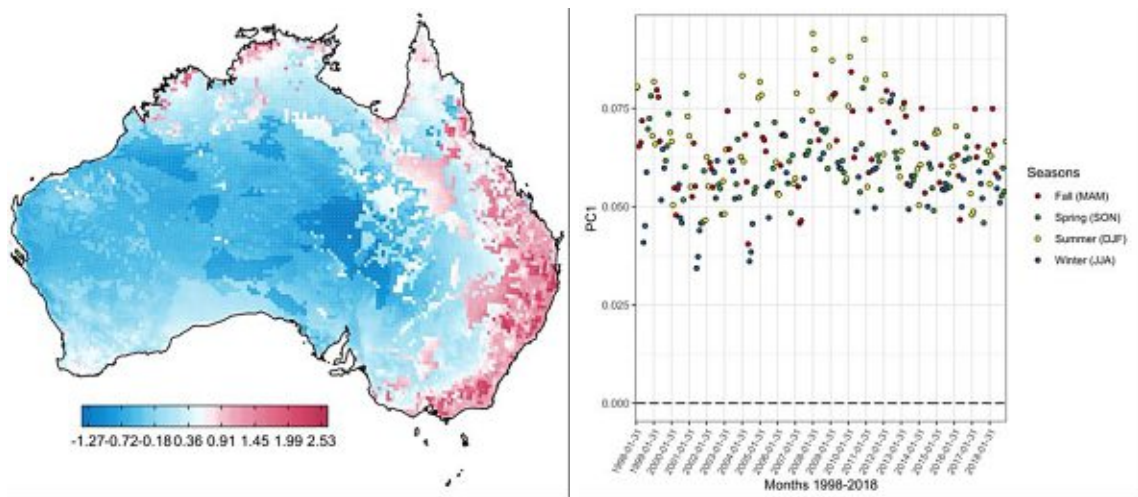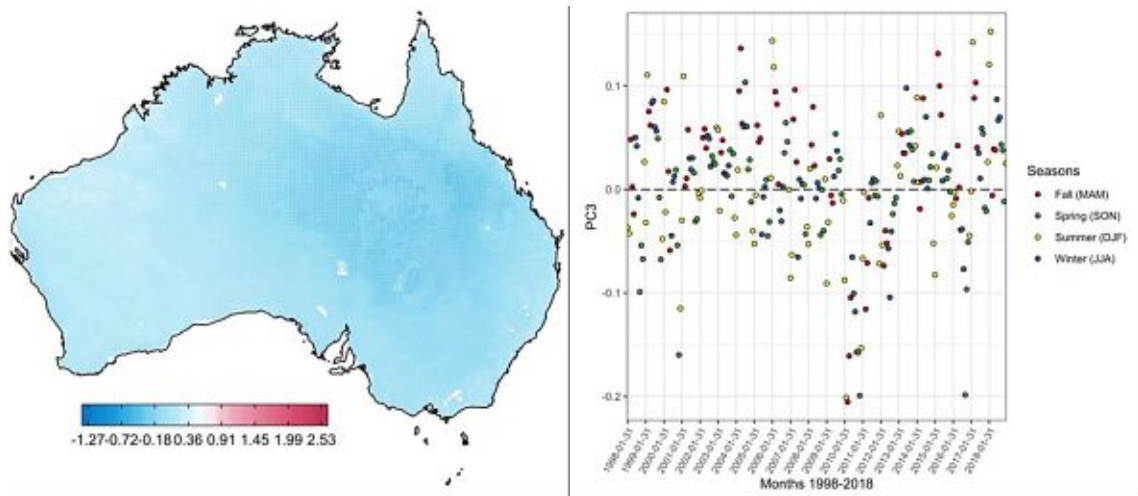

Figure 5.2: Left panel: Map of the first principal component (PC1) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the first principal component (PC1).

We expect the correlations to the COIs to be equally high since we assume that not a single COI will mainly influence the data means. However, we find a correlation of roughly 0.25 with the SOI at a two months delay (see Table 5.1). The area we expect the greatest impact of the SOI is precisely the east coast of Australia (see Figure 2.5 and [2]).

Table 5.1: **PC1** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.109	Lag 4
SAMI	0.198	Lag 0
SOI	0.248	Lag 2

PC2 (mSMM)

The second principal component (PC2) represents roughly 24% of the variance of the original data. We note in Figure 5.3 (left panel) that there is an extremely smooth north-south transition, which we will interpret as the yearly latitudinal movement of the ITC-zone (Intertropical Convergence Zone). This belt of converging north-east and south-east trade winds shifts north and south around the equator with the seasons. It brings heavy rainfalls and thunderstorms and influences the weather of affected regions substantially (compare [9]). We expected this pattern to appear in a principal component of this dataset. [9] described a similar result with the seasonal decomposition of the anomaly dataset from 1979-2010 [9, see Mode 2 p.64 & 66]. The loading plot (see Figure 5.3 (right panel)) however, seems insightful as it shows a definite division of the seasons. Fall and summer seasons load mainly negatively, and winter and spring seasons contribute positively.
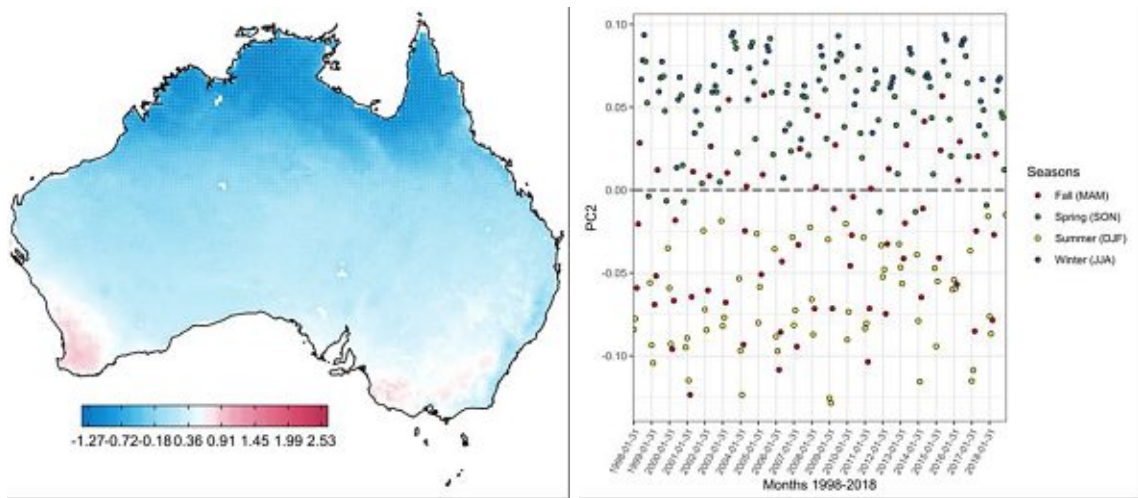

Figure 5.3: Left panel: Map of the second principal component (PC2) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the second principal component (PC2).

The absence of high correlation values with the COIs is expected and affirmed by our correlation analysis in Table 5.2. There is no correlation greater than 0.2. We disregard further considerations of impact areas of the COIs and see ourselves confirmed in the claim that PC2 could be associated with the movement of the ITC-zone.

Table 5.2: **PC2** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.145	Lag 12
SAMI	0.119	Lag 12
SOI	0.198	Lag 2

PC3 (mSMM)

The third principal component explains only around 3.2% of the variance of the original data. Figure 5.4 (left panel) exhibits a very weak east-west division that almost disappears at the shared scale of the coordinates. However, if we look at the map without adjusted scaling, this separation becomes very clear. We expect an influence of the SOI since the area of the biggest impact of the SOI is the east (see Figure 2.5). The SOI could be the explanation for this pattern. Although the correlation analysis attests the SOI the most prominent influence and therefore seems to agree with our considerations, the correlation value does not surpass the 0.2 mark (see Table 5.3). We withhold from further conclusions as the results are not convincing enough to contribute valuable information.

Figure 5.4: Left panel: Map of the third principal component (PC3) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the third principal component (PC3).

The SOI is assumed to have the biggest influence on the Australian climate in winter and spring. This is not deducible by the loading plot (see Figure 5.4 (right panel)). We do observe an interesting drop in the loading plot around 2010 when severe floods hit the east of Australia in spring [2].

Table 5.3: **PC3** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.173	Lag 4
SAMI	0.152	Lag 0
SOI	0.199	Lag 0

IC5 (mSMM)

The fifth independent component (IC5) shows a well-defined high-value spot in the south-east (see Figure 5.5 (left panel)). Remember, other than principal components, independent components are not ordered according to their explanatory power over the variance of the original data. Although there are criteria under which they are ordered, for our interpretation, independent components do not obey the strict hierarchy we see with principal components. In Figure 5.5 (right panel) we note two outliers in the loading plot. There is summer (DJF) 2002 with an extremely negative loading value and fall (MAM) 2012 with an extremely positive loading value. According to the BOM, the year 2002 was the 4th driest year since 1900. Extreme bushfires haunted Australia in the summer of 2002. This period is called "Black Christmas" since it brought tremendous damage on over half a million hectares of land [2]. The year 2012 was a year full of extrema and contrasts in the weather. It started at cooler temperatures than average and with increased rainfalls. It was the 7th wettest start in over 100 years. La Niña brought rainfalls and floods, and the tropical cyclone "Lua" passed over western Australia. From April on, it became drier and warmer [2].
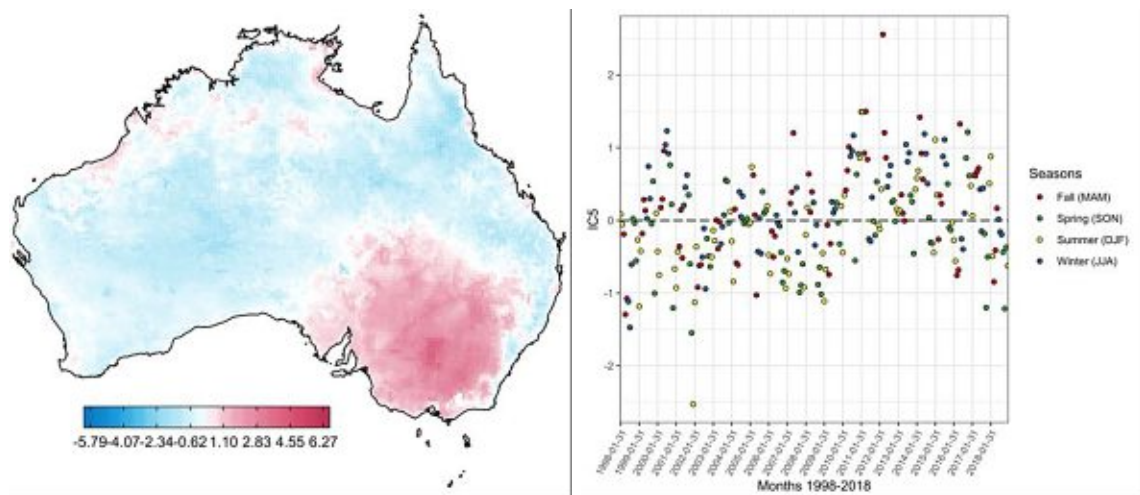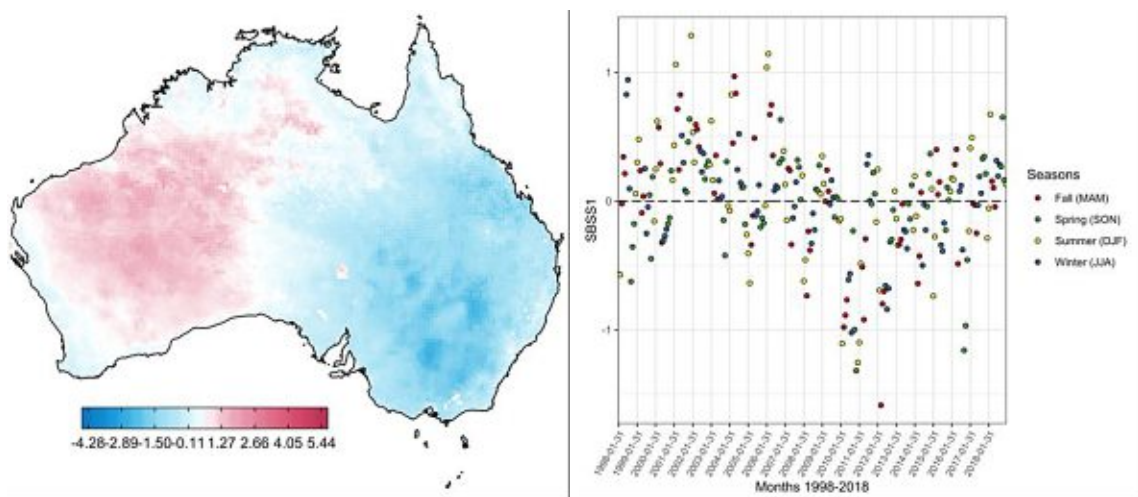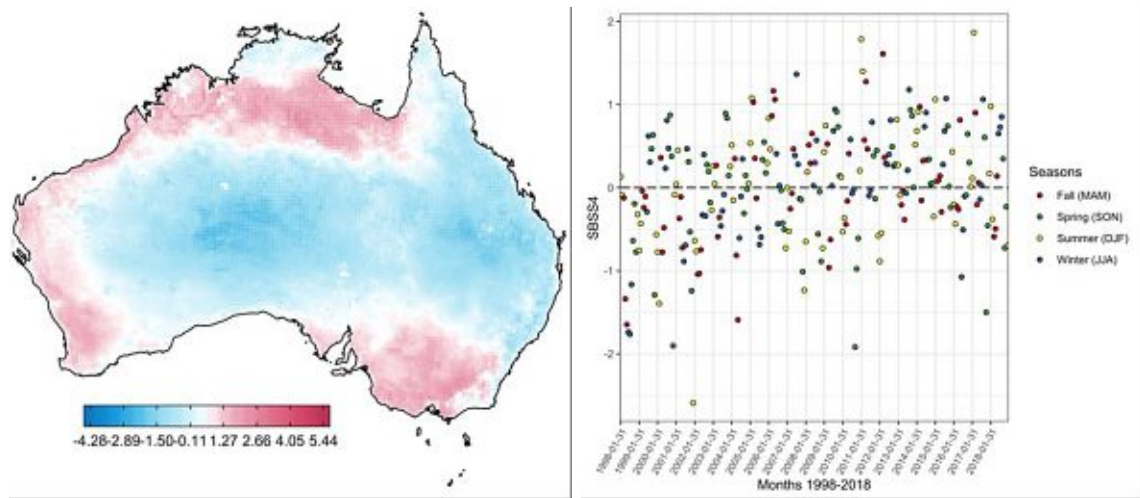

Figure 5.5: Left panel: Map of the fifth independent component (IC5) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the fifth independent component (IC5).

Table 5.4 exhibits a correlation of IC5 with the SOI of 0.24 at lag 3. A lag of 3 months is reasonable. A delay of 12 months, as with DMI and SAMI in Table 5.4, is more likely to be an annual pattern and it is preferable to look at the components without any delay (lag= 0) for further interpretations of the component in relation to COI.

Table 5.4: **IC5** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.171	Lag 12
SAMI	0.142	Lag 12
SOI	0.240	Lag 3

IC8 (mSMM)

The 8th independent component (IC8) in Figure 5.6 (left panel) shows a high-value band stretching from the north to the south in the western inland of Australia. This region is typically associated with the area of impact of the DMI (see Figure 2.2). However Table 5.5 reveals correlation values over 0.24 for all three COIs. We assume that all three indices are related to IC8 without temporal delay, since a lag of 12 months, as already explained, rather points to an annual pattern.The low values in the north-east of Australia could be linked to the SOI. The other low-value area in the form of a west coast belt, cannot be easily related to the SOI. ENSO has on average a four-year cycle from a negative (El Niño) to a positive (La Niña) phase [2]. The loading plot (see Figure 5.6 (right panel)) reveals a pattern that could be related to those four-year windows.
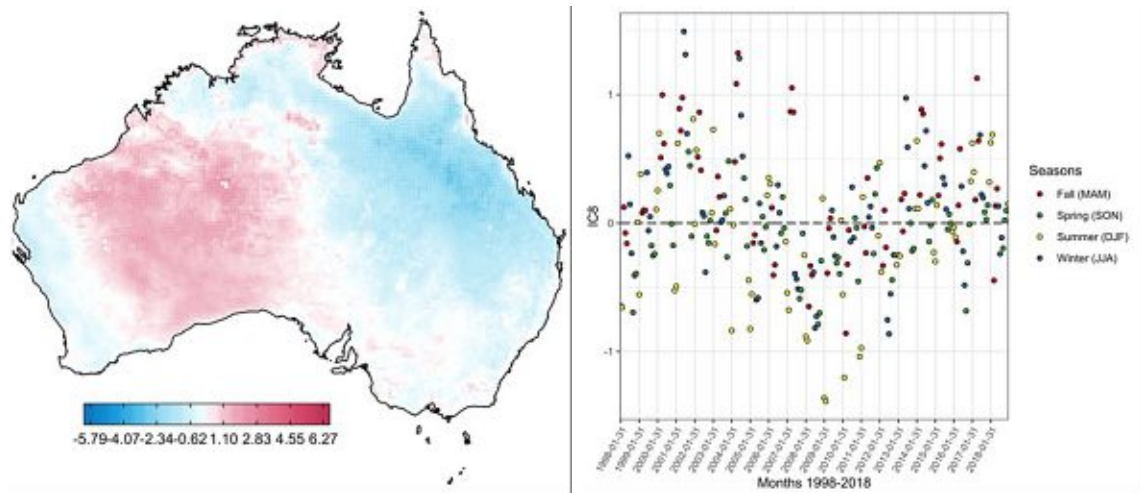
Figure 5.6: Left panel: Map of the eighth independent component (IC8) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the eighth independent component (IC8).

Table 5.5: **IC8** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.244	Lag 12
SAMI	0.289	Lag 0
SOI	0.284	Lag 0

SBSS-C1 (mSMM)

The first independent component of SBSS (SBSS-C1) shows an apparent east-west division (see Figure 5.7 (left panel)). Additionally, a narrow band over the west coast of Australia stands out without a clear explanation. However, the small spots with high values southeast of the center can be explained easily. It is the location of "Lake Eyre". A huge national park, which is responsible for outliers. As already noted with IC5 values for summer 2002 and fall 2012 stand out in the loading plot in Figure 5.7 (right panel). This time with reversed signs. Possible reasoning for these two seasons to stand out have been discussed already.

Figure 5.7: Left panel: Map of the first SBSS independent component (SBSS-C1) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the first SBSS independent component (SBSS-C1).

In Table 5.6, the SOI shows the highest correlations with SBSS-C1, particularly around 0.25 without a monthly delay. This result is not surprising when we consider the impact area of the SOI, the east of Australia (see Figure 2.5).

Table 5.6: **SBSS1** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.245	Lag 12
SAMI	0.216	Lag 0
SOI	0.257	Lag 0

SBSS-C4 (mSMM)

The 4th independent component of SBSS (SBSS-C4) shows an interesting new pattern. A broad low-value band stretching from east to west over the whole continent (see Figure 5.8 (left panel)). The loading plot in Figure 5.8 (right panel) seems to be following a very weak positive trend.

Figure 5.8: Left panel: Map of the fourth SBSS independent component (SBSS-C4) of the mSSM dataset. Right panel: Loading plot grouped by seasons of the fourth SBSS independent component (SBSS-C4).

The pattern of SBSS-C4 in Figure 5.8 (left panel) seems to be in accordance with the high correlation with the SAMI in Table 5.7. The SAMI especially influences the south of Australia in horizontal lines depending on the movement of the subtropical ridge, as discussed in Chapter 2.1.2.

Table 5.7: **SBSS4** of Monthly Soil Moisture Means 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.207	Lag 12
SAMI	0.246	Lag 0
SOI	0.194	Lag 6

5.2 Soil Moisture Data Anomalies

We perform our analysis on the anomalies of the CCI dataset (see 4.1.2 and 4.1.3). First of all we display the quarterly anomalies for a better understanding of the data (see Figure 5.9). The anomalies are mean-adjusted data, whereby the mean value is not calculated over the entire dataset, but separately for each month (see Equation (2.2)). As expected, the map scales are now centered around zero. We assume that, due to the nature of the anomalies, analyzing them provides additional insight into the structure of the data. Again we work with data at a monthly resolution. Additionally, we perform a seasonal decomposition. Apparently, the anomalies are subject to a stronger seasonal pattern than the mean values. The summer season will be especially relevant for us, as we will inspect it in greater detail later on. We see a well-defined spot of below-average, particularly dry conditions in the north-east, while the rest of the continent is wet above average.
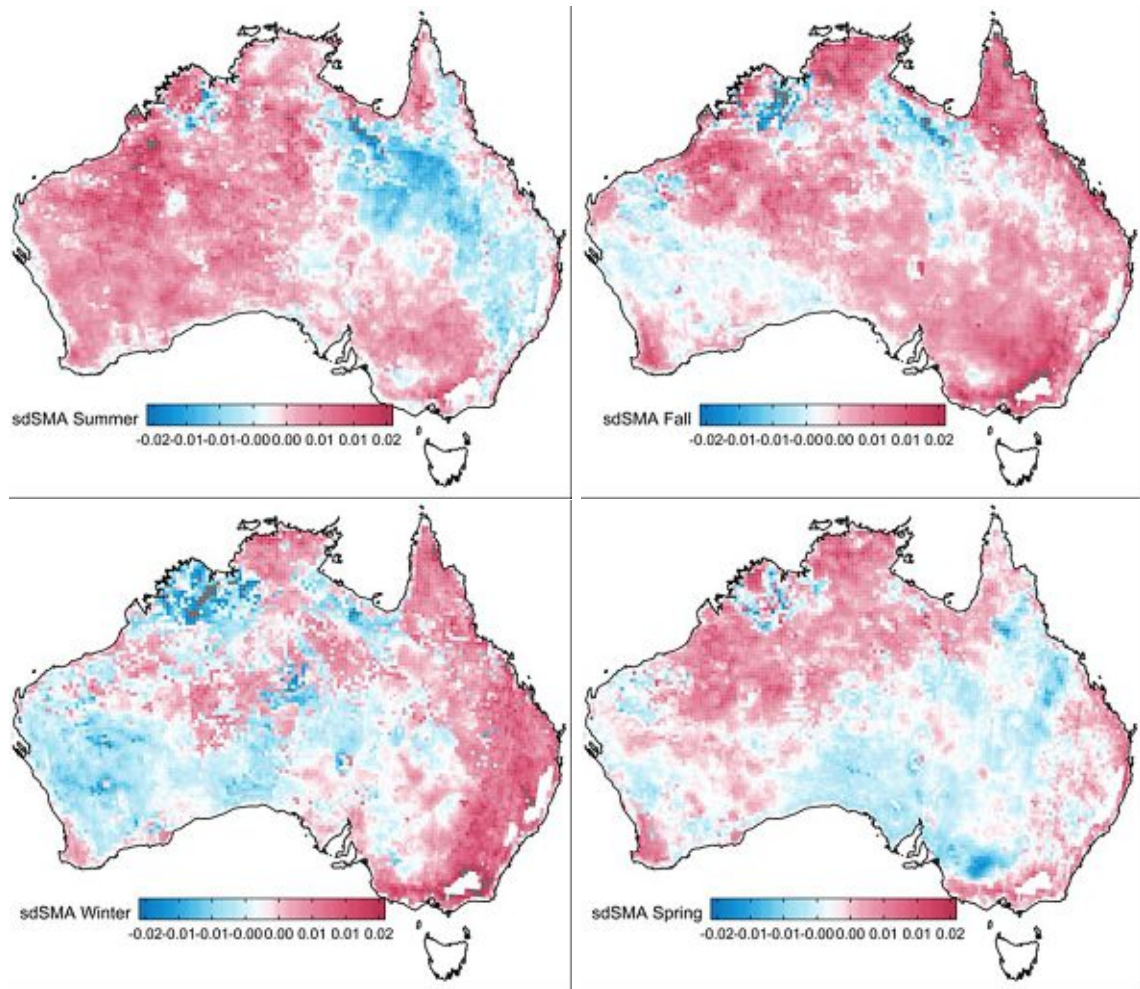
Figure 5.9: Maps of SM Anomalies for Summer (DJF), Fall (MAM), Winter (JJA) and Spring (SON).

5.2.1 Monthly Soil Moisture Anomalies (mSMA) 1998-2018

A threshold of 80% of the explained variance of the original data gives us 30 components (see Table 4.2). PCA results in expected east-west divisions, as well as the yearly latitudinal movement of the ITC-zone in form of a north-south division. In contrast to the north-south division we observe when looking at the CCI SM Means, we have no clear groups of positively loading seasons and negatively loading seasons. This can be explained by the way we calculate the anomalies (removing yearly cycles). For the anomaly dataset ICA fails in producing viable patterns. Small delimited spots are created, for which it is difficult to find reliable interpretations. SBSS on the other hand manages to display the east-west division and the north-south division whilst even resulting in higher correlations to the three key Climate Oscillation Indices.

PC1 (mSMA)

The first principal component (PC1) of the anomaly dataset with monthly temporal resolution represents 18% of the explained variance of the original data. The pattern in Figure 5.10 (left panel) looks promising and displays a well-defined low-value spot in the northeastern coastal region, which is associated with the area of impact of the SOI (see Figure 2.5). Unfortunately, the correlation analysis in Table 5.8 offers no satisfying results as all correlations are beneath the 0.2 mark. The loading plot in Figure 5.10 (right panel) shows a broader loading range for summer than for the other seasons and some clear outliers around the record flood year 2010, especially for fall season. All seasons contribute positively as well as negatively to the PC1.
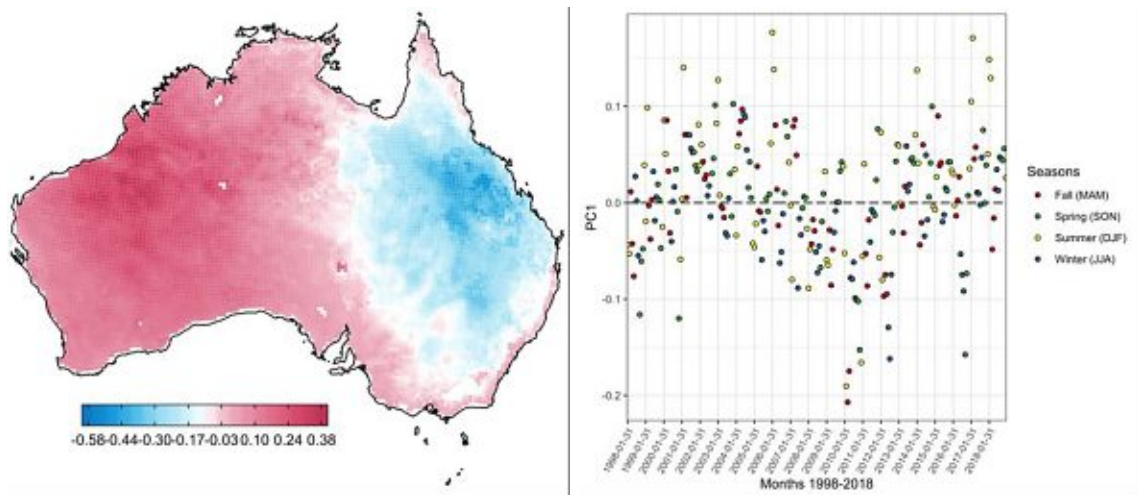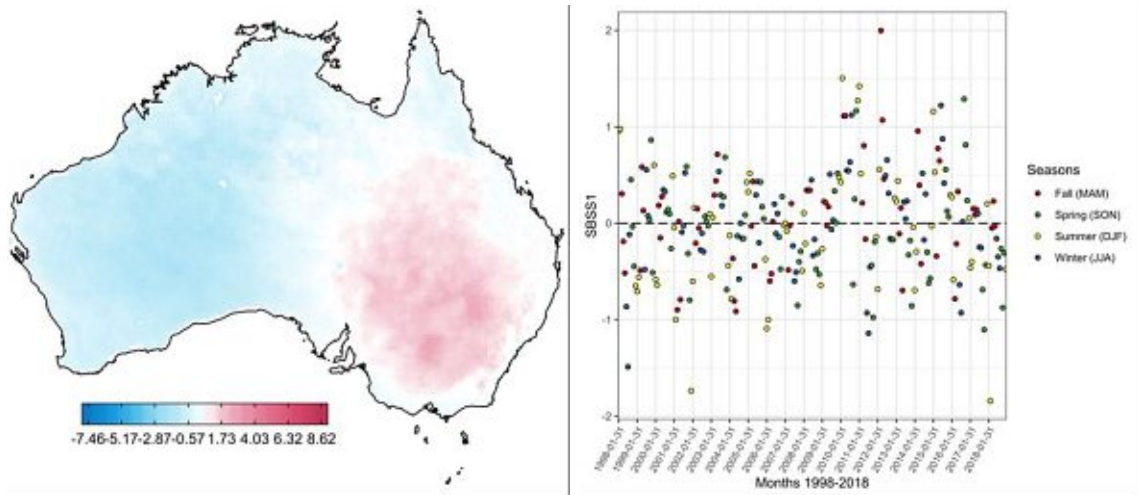

Figure 5.10: Left panel: Map of the first principal component (PC1) of the mSSA dataset. Right panel: Loading plot grouped by seasons of the first principal component (PC1).

Table 5.8: **PC1** of Monthly Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.138	Lag 12
SAMI	0.157	Lag 0
SOI	0.169	Lag 2

PC2 (mSMA)

Another interesting example is the second principal component (PC2). The variance of the original data explained by this component is 14%. The pattern in Figure 5.11 (left panel) shows a north-south division. We can again draw the connection to the ITC-zone and the

yearly overall rainfall patterns (compare [9]). Although we note that looking at the SM Anomaly dataset now, the pattern is by far not as smooth as the corresponding second component of the SM means. We want to point out that in contrast to the corresponding second principal component of the SM Means, the loading plot in Figure 5.11 (right panel) does not divide the seasons in positive and negative loaders. This information is lost when looking at SM Anomalies, which does not come as a surprise as by calculating the anomalies we remove a possible seasonal pattern within the data.
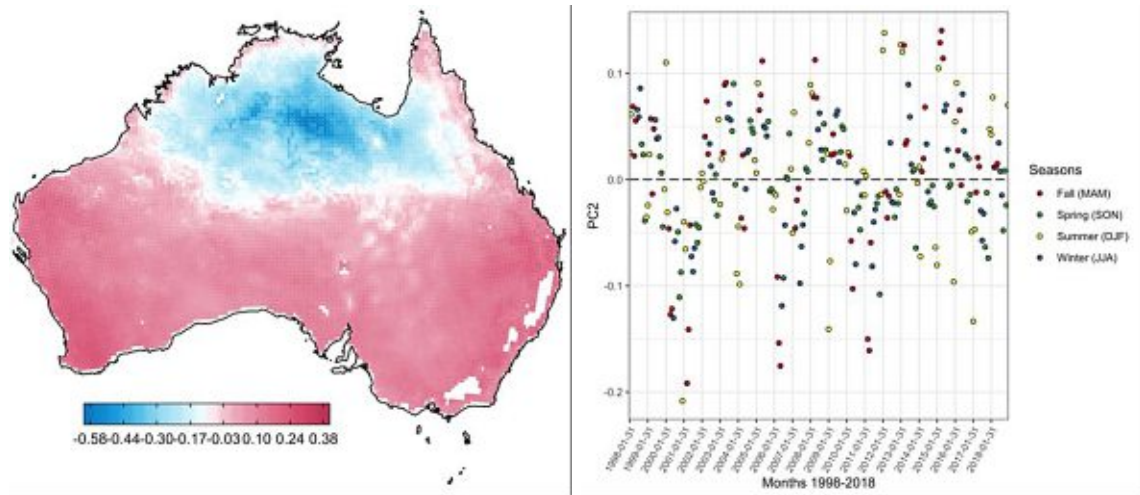
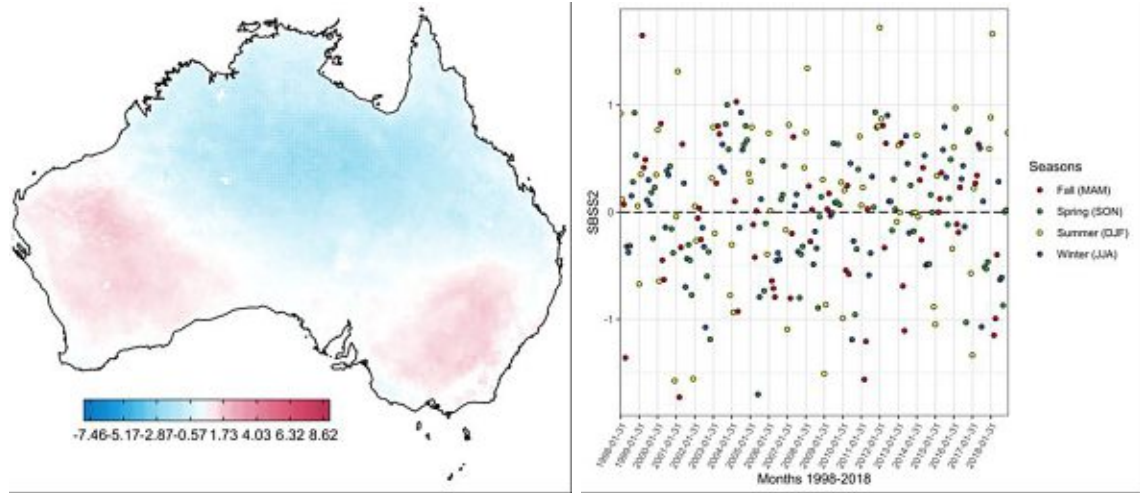Figure 5.11: Left panel: Map of the second principal component (PC2) of the mSSA dataset. Right panel: Loading plot grouped by seasons of the second principal component (PC2).

The correlation value in Table 5.9 with the PC2 and the SOI over 0.2 at a six-month delay is not expected. We do not expect the ITC-movement to be related to a particular COI. However, we also noted this tendency of an increased influence of the SOI with the corresponding SM dataset, although it was not as evident then.

Table 5.9: **PC2** of Monthly Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.160	Lag 6
SAMI	0.176	Lag 1
SOI	0.225	Lag 6

IC1 (mSMA)

The most striking observation for all 30 independent components of ICA is that there are no large area patterns. All ICs look noisy, and most of them have one small, well-defined spot at a seemingly random location. We shortly discuss the first independent component

(IC1) as an example. The other 29 ICs can be found in the Appendix 7.4. However, we see these results as a strong indication that ICA fails to offer insights to the SM Anomaly dataset. We will see in the next paragraph that SBSS does a better job.

In Figure 5.12 (left panel), we see a well-defined spot in central-north Australia that is found by ICA.
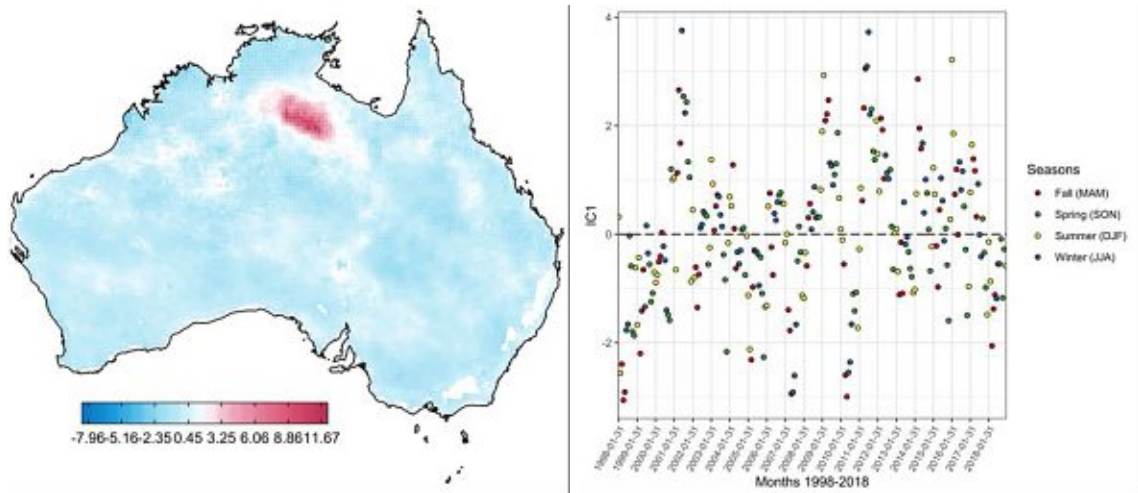
Figure 5.12: Left panel: Map of the first independent component (IC1) of the mSSA dataset. Right panel: Loading plot grouped by seasons of the first independent component (IC1).

Although Table 5.10 shows relatively high correlations with all COIs, especially the DMI and the SOI, without a monthly delay, we would expect a larger area of impact of those COIs. We cannot draw any conclusions from this spot and therefore argue that ICA is not recommendable for this dataset.

Table 5.10: **IC1** of Monthly Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.365	Lag 0
SAMI	0.297	Lag 0
SOI	0.330	Lag 0

SBSS-C1 (mSMA)

Overall we note with the SM Anomaly dataset that SBSS manages to produce some familiar patterns. We present selected examples here. A complete line-up can be found in the Appendix 7.4.

The first independent component of SBSS (SBSS -C1) displays a familiar east-west division shown in Figure 5.13 (left panel). The loading plot in Figure 5.13 (right panel) indicates that all seasons contribute equally. However, once again, we note that summer 2002 and fall 2012 are outliers. The summer of 2018 additionally catches our attention. According to the annual climate statement of the BOM the year 2018 was a record-breaking year in terms of heat and droughts across eastern Australia. It was one of the warmest years to be recorded, only with 2013 and 2005 being warmer [2].

Figure 5.13: Left panel: Map of the first SBSS independent component (SBSS-C1) of the mSSA dataset. Right panel: Loading plot grouped by seasons of the first SBSS independent component (SBSS-C1).

In Table 5.11, we observe relatively high values for the DMI at lag 12 and the SOI at lag 0. While the SOI could explain the high values in the south-east, the DMI could be responsible for the low values in the south to north band in western Australia (see Figures 2.2 & 2.5).

Table 5.11: **SBSS1** of Monthly Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.330	Lag 12
SAMI	0.298	Lag 2
SOI	0.317	Lag 0

SBSS-C2 (mSMA)

The second independent component of SBSS (SBSS-C2) likely represents the already discussed movement of the ITC-zone (see Figure 5.14 (left panel)). It is very promising that

SBSS reproduces this familiar pattern. The movement of the ITC-zone is known to have an essential role in the yearly overall rainfall regime (compare [9]).

Figure 5.14: Left panel: Map of the second SBSS independent component (SBSS-C2) of the mSSA dataset. Right panel: Loading plot grouped by seasons of the second SBSS independent component (SBSS-C2).

In agreement with these thoughts, all seasons seem to contribute equally in Figure 5.14 (right panel). It is again mentionable that we are not able to see a division of the seasons into positively and negatively loaders as seen with the PCA of the SM Mean dataset. The high correlation value of the SOI in Table 5.12 is not expected but concurs with our findings to date.

Table 5.12: **SBSS2** of Monthly Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.328	Lag 0
SAMI	0.291	Lag 1
SOI	0.411	Lag 4

5.2.2 Seasonal Decomposition of Soil Moisture Anomalies (sdSMA) 1998-2018

A threshold of 80% gives us 16 components for the summer season (DJF), 16 components for the fall season (MAM), 15 components for the winter season (JJA), and 17 components for the spring season (SON) (see Table 4.3). We examine some interesting results of the summer decomposition. ICA fails again to deliver useful components, while PCA and SBSS reproduce familiar and expected maps (east-west divisions and north-south divisions), but also some new patterns. The correlations with the COIs for a single season (summer) are

overall higher, as expected. SBSS results in notably higher correlations. The complete line-up of the results for every season can be found in the Appendix 7.4.

PC1 (sdSMA - Summer)

The first principal component (PC1) explains 24% of the variance of the original summer anomalies data. The DMI has, by far, the largest influence, according to our correlation analysis in Table 5.13. This influence of the DMI is not unexpected as it could explain the high values in the south-west (see Figures 2.2 & 5.15). As already mentioned, we expect higher correlation values overall for any seasonal decomposition. The results, however, are not particularly satisfying. We would have expected more values over 0.6 with PCA, ICA and SBSS.
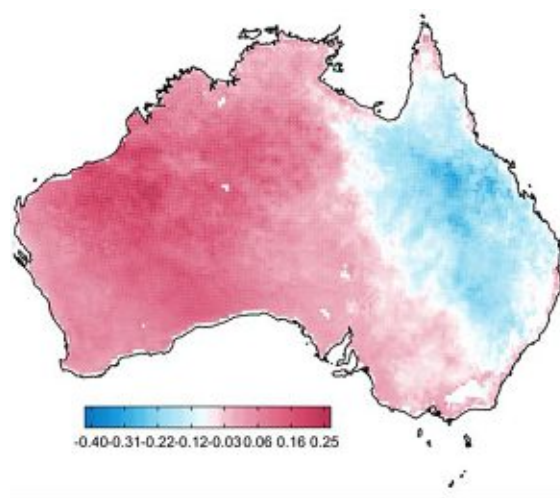
Figure 5.15: Map of the first principal component (PC1) of the sdSSA-Summer dataset.

Table 5.13: **Summer PC1** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.402	Lag 4
SAMI	0.269	Lag 2
SOI	0.248	Lag 12

PC2 (sdSMA - Summer)

The second principal component (PC2) explains 16% of the variance of summer anomalies data. In Figure 5.16, we see that there is a notable south-north division. However, it does look slightly modified from the usual components displaying the movement of the ITC-zone. Apart from the visual differences, we observe a difference in the results of the correlation

analysis in Table 5.14. The SAMI is the most prominent index, instead of the SOI, which we observed with maps displaying the movement of the ITC-zone so far. The SAMI could explain high values in the south of Australia (compare Figures 2.3 & 2.4).
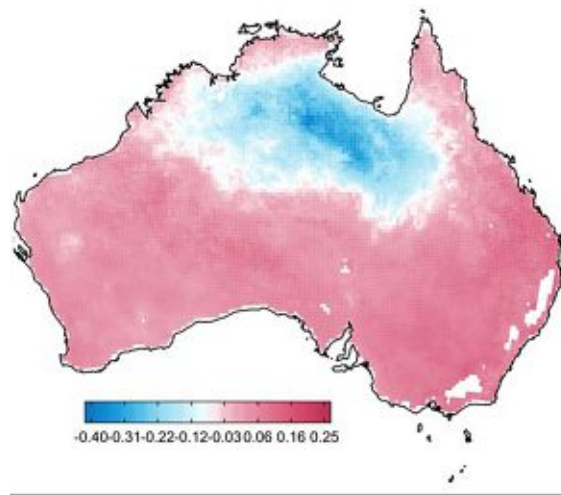
Figure 5.16: Map of the second principal component (PC2) of the sdSSA-Summer dataset.

Table 5.14: **Summer PC2** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.338	Lag 6
SAMI	0.425	Lag 0
SOI	0.107	Lag 12

PC3 (sdSMA - Summer)

The third principal component (PC3) explains roughly 8% of the variance of original summer anomalies data. Figure 5.17 shows a new pattern with a north-western band of low values and high values in Central Australia.
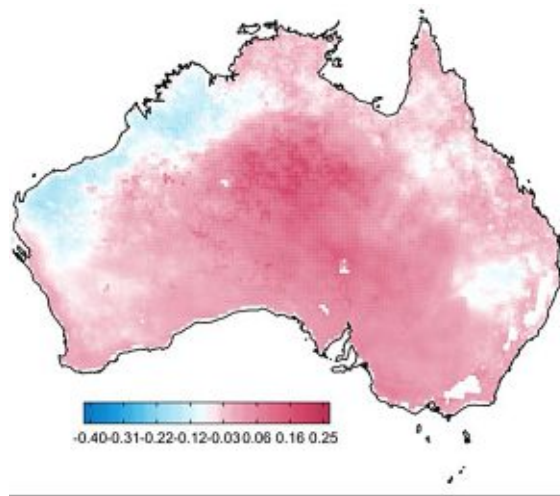
Figure 5.17: Map of the third principal component (PC3) of the sdSSA-Summer dataset.

Table 5.15 shows the highest correlation with the SOI at a one-month-delay. This result is not precisely following our straight-forward interpretation of the area of impact of the SOI. Instead of an influence in the east, we perceive a distinctive band of low values in the west of Australia as well as some unsharp low-valued spots in the east.

Table 5.15: **Summer PC3** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.254	Lag 12
SAMI	0.121	Lag 4
SOI	0.489	Lag 1

ICA (sdSMA - Summer)

The Seasonal Decomposition of the SM Anomaly dataset offers no insightful patterns or correlation structures whatsoever when ICA is applied. Again this points to the benefit of the SBSS method with this particular dataset.

SBSS-C1 (sdSMA - Summer)

The independent components of SBSS deliver familiar patterns. Thus we can confirm previous findings and identify new elements. Selected examples of the 16 independent components of the SBSS are:

The first independent component of SBSS (SBSS-C1) shows a high-value region in the east (see Figure 5.18). This pattern seems to agree with the correlation value in Table 5.16 of roughly 0.5 with the SOI at a delay of two months (compare Figure 2.5).
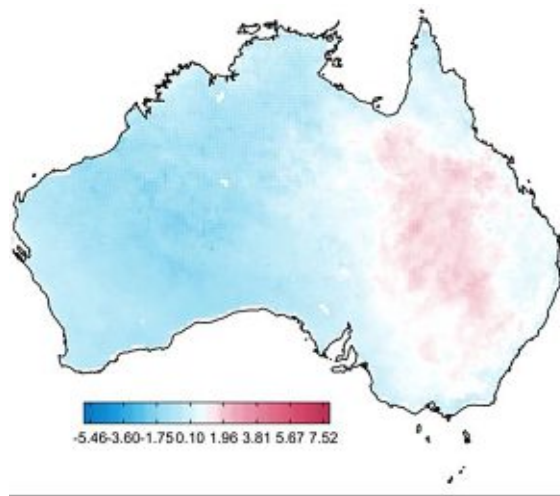
Figure 5.18: Map of the first SBSS independent component (SBSS-C1) of the sdSSA-Summer dataset.

Table 5.16: **Summer SBSS C1** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.461	Lag 1
SAMI	0.435	Lag 2
SOI	0.499	Lag 2

SBSS-C2 (sdSMA - Summer)

The second independent component of SBSS (SBSS-C2) displays a notable low-value region in the Central north (see Figure 5.19). In Table 5.17, a relatively high correlation with all COIs is eminent. We did not notice this pattern in [9] and a more in-depth study by subject experts of this component could bring new insights into the relationship between soil moisture data and COIs.
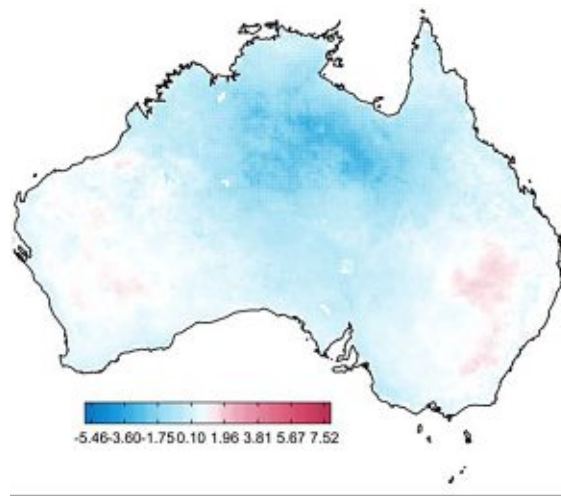
Figure 5.19: Map of the second SBSS independent component (SBSS-C2) of the sdSSA-Summer dataset.

Table 5.17: **Summer SBSS C2** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.514	Lag 4
SAMI	0.494	Lag 0
SOI	0.456	Lag 0

SBSS-C3 (sdSMA - Summer)

The third independent component of SBSS (SBSS-C3) shows an interesting high-value belt crossing Central Australia (see Figure 5.20). This region is associated with the extended area of impact of positive or negative DMI phases (see Figure 2.2). A positive phase of the DMI is related to reduced rainfall in the area and vice versa.
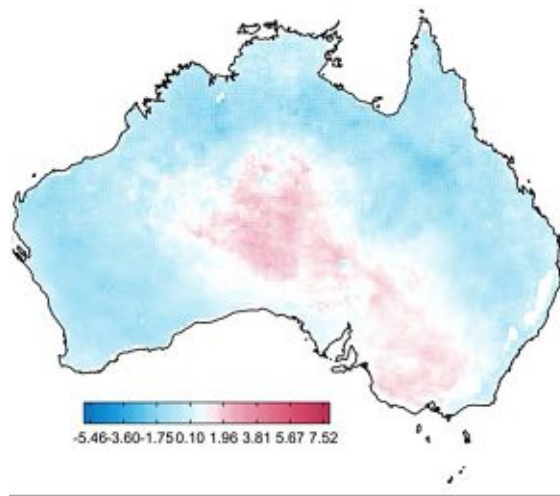
Figure 5.20: Map of the third SBSS independent component (SBSS-C3) of the sdSSA-Summer dataset.

The correlation results in Table 5.18 seem to be only partly consistent with our thoughts and considerations. A relatively high correlation with DMI (0.5 at a three-months lag), but also the SOI (0.68 at a four-months lag) is observable (compare Figures 2.2 & 2.5).

Table 5.18: **Summer SBSS C3** of Seasonal Decomposition of Soil Moisture Anomalies 1998-2018: Maximal Absolute Correlation and Monthly Lag.

	Abs. Correlation	Monthly Lag
DMI	0.515	Lag 3
SAMI	0.368	Lag 6
SOI	0.680	Lag 4

6 Conclusion

The aim of this thesis was to test the applicability of Spatial Blind Source Separation to satellite observations of soil moisture data. In order to benchmark SBSS against other more commonly used methods in this context, the CCI SM data was analyzed with PCA, as well as ICA in addition to SBSS. Besides the novel methodological approach, new insights into the behavior of soil moisture in relation to climate indices have been produced by the analysis. While PCA and ICA generally ignore the spatial dependence structure of soil moisture data, SBSS considers this information explicitly in its model formulation. The following conclusions can be drawn:

- SBSS and PCA are able to reproduce known results and patterns from previous analyses and are consistent with our assumptions about the dynamics of soil moisture and their relationship to the climate indices important for Australia. SBSS also produces some promising new patterns. In contrast, ICA is not able to reproduce known mechanisms.

- The CCI SM data was modified in three different ways and used as a basis for the analysis, to reflect different aspects of the underlying processes captured by the data. For all three versions higher correlations between loadings and COIs could be found for SBSS as compared to PCA. Subject experts point out that this indicates an advantage of the SBSS method.

- Although it is reasonable to assume that soil moisture reacts with a certain delay (lag) with respect to oscillations of the climate indices, the plausibility of the results must always be evaluated with the involvement of subject experts. Immediate effects (lag = 0) may occur, a delay of a whole year (lag = 12) could probably indicate an underlying seasonal pattern, the disregard of which may distort the interpretation.

COIs highly correlated among each other and should therefore be analyzed together and not separately as it was done in our case. Given the components of PCA are only unique up to sign, the "true" sign could be determined and fixed with the help of subject experts, but requires expertise and hinders the reproducibility of the results. Unfortunately, this will not always be possible, especially if the results are highly dimensional. The sign uncertainty is propagated to the correlation analysis and distorts possible inferences drawn from said correlations. Here we mitigated the problem by means of a small Monte Carlo study. In the future, a maximization criterion could be used to more objectively adress the issue of the correct sign.

This work has shown that SBSS is a promising candidate for the statistical analysis of soil moisture data and other climatic variables subject to spatial dependencies.

7 Appendix

Disclaimer: For the sake of avoiding excessively long, but uninformative sections, the entire code of the thesis is not given here, but can be obtained from luzia.jorda@gmx.at upon request.

7.1 Reading of NetCDF-Files in R

The main data of the thesis, satellite observations of soil moisture, are provided by ESA [3] in a NetCDF-format. Although the NetCDF-format is no standard format for data within R, the files can be imported to R via R-Packages, such as ncdf4.

```
library(ncdf4)            # package for netcdf manipulation
library(reshape2)         # package includes function ''melt''

sm_data <- nc_open("/.../ESACCI-SOILMOISTURE.nc")

lon  <- ncvar_get(sm_data, "lon")
lat  <- ncvar_get(sm_data, "lat")
my_t <- ncvar_get(sm_data, "time")

# soil moisture mean, output is an array with axis; lon, lat, time
sm_mean <- ncvar_get(sm_data, attributes(sm_data$var)$names[1])

# time of sm_data is given in days since start date
start_t <- as.POSIXct('1970-01-01 00:00:00', format = "%Y-%m-%d %H:%M:%S",
tz = "UTC")
t_new <- start_t + 60 * 60 * 24 * my_t
dimnames(sm_mean) <- list(as.character(t_new), lon, lat)
my_sm_mean <- melt(sm_mean)
names(my_sm_mean) <- c("time", "lon", "lat", "sm_mean")
```

7.2 Function Calls in R

In the following we present a short overview of the function calls of the PCA, ICA and SBSS analyses including the required packages.

```
library(missMDA)     # imputePCA
library(JADE)        # JADE
```

```
library(SpatialBSS) # SBSS

# impute missing data
temp_pca_sm <- imputePCA(my_sm_mean, scale=FALSE)

# PCA
res_pca_m2018 <- prcomp(temp_pca_sm$completeObs,
center = TRUE, scale = FALSE)

# ICA (via JADE) of the first 8 principal components (q = 8)
res_jade_m2018 <- JADE(res_pca_m2018$x[,c(1:8)])

# loading matrix is product of W_{pca}(pxq)W_{ica}^T(qxq)
loadings_ica_m2018 <- tcrossprod(res_pca_m2018$rotation[,c(1:8)],
res_jade_m2018$W)


# SBSS of the first 8 principal components (q = 8)
kernel_parameters <- matrix(c(0, 60, 60, 120, 120, 180, 180, 240))
kernel_matrix_ring <- spatial_kernel_matrix(coords,kernel_type = "ring",
kernel_parameters = kernel_parameters)

sbss_result_m2018 <- sbss(field, coords, kernel_type = 'ring',
kernel_parameters <- kernel_parameters, ordered=TRUE, kernel_list =
kernel_matrix_ring)

# loading matrix is product of W_{pca}(pxq)W_{sbss}^T(qxq)
loadings_sbss_m2018 <- tcrossprod(res_pca_m2018$rotation[,c(1:8)],
sbss_result_m2018$w)
```

7.3 Monte Carlo Study for Maximum Absolute Correlation

As explained in Section 4.5, we have chosen Spearman correlation to assess the plausibility of dependencies between our results, the resulting loadings from the different approaches of BSS, and COIs. For PCA we correlate the loadings in the form the rows of the un-mixing matrix $\mathbf{W}_{PCA}$ with time series of three COIs. For ICA and SBSS this is done with the COIs and the rows of the loading matrix in form of the matrix product $\mathbf{W}_{ICA}^{T}\mathbf{W}_{PCA}$ and $\mathbf{W}_{SBSS}^{T}\mathbf{W}_{PCA}$, respectively. These loading matrices transform the observations into the respective components.

However, the solution to PCA in the form of the un-mixing matrix $\mathbf{W}_{PCA}$ is only unique up to sign. This means multiplying a row of $\mathbf{W}_{PCA}$ by -1 corresponds to the same solution, but the direction of one of the basis vectors in the new coordinate system (of PCA) is flipped. For the correlation analysis between loadings of PCA and COIs, this translates

to the fact that the sign of the correlations are unique up to sign (but the magnitude of the corresponding correlation is not affected). For ICA and SBSS the problem is more complex. The sign uncertainty in the rows of $\mathbf{W}_{PCA}$ propagates to the loading product $\mathbf{W}_{ICA}^T\mathbf{W}_{PCA}$ and $\mathbf{W}_{SBSS}^T\mathbf{W}_{PCA}$, respectively.

The effects of the sign changes of the rows of $\mathbf{W}_{PCA}$ can lead to pronounced different correlations for the loading product matrices (because the product of the loading matrices can be vastly different, depending on the sign of the rows of $\mathbf{W}_{PCA}$). In order to address this issue, but at the same time not blow up the scope of the thesis, we decided to conduct a Monte Carlo study to find the "correct" (maximal) correlation. To make our results reproducible, we have set a random seed for the Monte Carlo study (17122019) and output a permutation index for all correlation results of ICA and SBSS (see Appendix 7.4). The permutation index displays the signs of the rows of the un-mixing matrix $\mathbf{W}_{PCA}$ for the respective correlation value. The sign of a row of the un-mixing matrix $\mathbf{W}_{PCA}$ is indicated as positive, if the corresponding row-sum is positive, and negative for negative row-sums. For $\mathbf{W}_{PCA} \in \mathbb{R}^{q\times q}$, there are 2^q possible configuration of row-signs. We provided the code for this Monte Carlo study for better understanding and reproducibility). We are aware that this procedure is not optimal. Future work on this method could focus on a sensible maximization criterion that would provide a universally applicable solution to this issue. For now, we are content with the following small Monte Carlo study:

```
my_corr <- function(loadings_pca, loadings_bss, k)
{
  set.seed(17122019); rnd_comb <- sample(c(-1,1), 16000, replace=TRUE)

  # the rnd_comb_matrix consists of 2000 combinations of arbitrary sign
  # transformations (-1 or 1)
  # for the q = 8 dimensions of the loading matrix

  rnd_comb_matrix <- matrix(data = rnd_comb, nrow = 2000, ncol = 8)
  coi <- coi_m2018[, c("dmi", "sam", "soi")]
  corr_sam <- matrix(data = NA, nrow = 2000, ncol = 7)
  corr_soi <- matrix(data = NA, nrow = 2000, ncol = 7)
  corr_dmi <- matrix(data = NA, nrow = 2000, ncol = 7)
  results <- matrix(data = NA, nrow = 3, ncol = 3)
  colnames(results) <- c("max_cor", "lag  ","index")
  row.names(results) <- c("DMI", "SAM", "SOI")
  temp_rotation <- loadings_pca[, c(1:8)]

  for (i in 1:2000){
    for (j in 1:8){
      temp_rotation[,j] <- loadings_pca[,j] * rnd_comb_matrix[i,j]
    }
```

```
    loadings_ica_m2018 <- tcrossprod(temp_rotation, loadings_bss)
    tmp_ccf_dmi <- ccf(rank(loadings_ica_m2018[, k]), rank(coi[, "dmi"]),
    12, "correlation")
    corr_dmi[i,] <- tmp_ccf_dmi[["acf"]][c(13, 14, 15, 16, 17, 19, 25)]

    tmp_ccf_sam <- ccf(rank(loadings_ica_m2018[, k]), rank(coi[, "sam"]),
    12, "correlation")
    corr_sam[i,] <- tmp_ccf_sam[["acf"]][c(13, 14, 15, 16, 17, 19, 25)]

    tmp_ccf_soi <- ccf(rank(loadings_ica_m2018[, k]), rank(coi[, "soi"]),
    12, "correlation")
    corr_soi[i,] <- tmp_ccf_soi[["acf"]][c(13, 14, 15, 16, 17, 19, 25)]
  }

# the matrix results contains the ''absolute maximal correlation value'',
# the monthly lag and the permutation index (index of sign combination
# of rnd_comb_matrix) for all COIs

  results[1, 1] <- max_dmi <- max(abs(corr_dmi))
  ind <- which(corr_dmi == max_dmi, arr.ind = TRUE)
  colnames(corr_dmi) <- c("Lag 0", "Lag 1", "Lag 2", "Lag 3", "Lag 4",
  "Lag 6", "Lag 12")
  results[1,2] <- colnames(corr_dmi)[ind[1,2]]
  results[1,3] <- ind[1,1]

  results[2,1] <- max_sam <- max(abs(corr_sam))
  ind <- which(corr_sam == max_sam, arr.ind = TRUE)
  colnames(corr_sam) <- c("Lag 0", "Lag 1", "Lag 2", "Lag 3", "Lag 4",
  "Lag 6", "Lag 12")
  results[2,2] <- colnames(corr_sam)[ind[1,2]]
  results[2,3] <- ind[1,1]

  results[3,1] <- max_soi <- max(abs(corr_soi))
  ind <- which(corr_soi == max_soi, arr.ind = TRUE)
  colnames(corr_soi) <- c("Lag 0", "Lag 1", "Lag 2", "Lag 3", "Lag 4",
  "Lag 6", "Lag 12")
  results[3,2] <- colnames(corr_soi)[ind[1,2]]
  results[3,3] <- ind[1,1]
}
```

7.4 Complete List of Results

In Section 7.4.1 PCA, ICA and SBSS results of the CCI Soils Moisture Mean dataset with monthly resolution are listed. Section 7.4.2 presents PCA, ICA and SBSS results for the

CCI Soil Moisture Anomaly data set with monthly resolution and the Sections 7.4.3, 7.4.4, 7.4.5 and 7.4.6 list the results of all three methods in form of a seasonal decomposition for summer, fall, winter and spring, respectively.

7.4.1 Monthly Soil Moisture Means 1998-2018

PCA
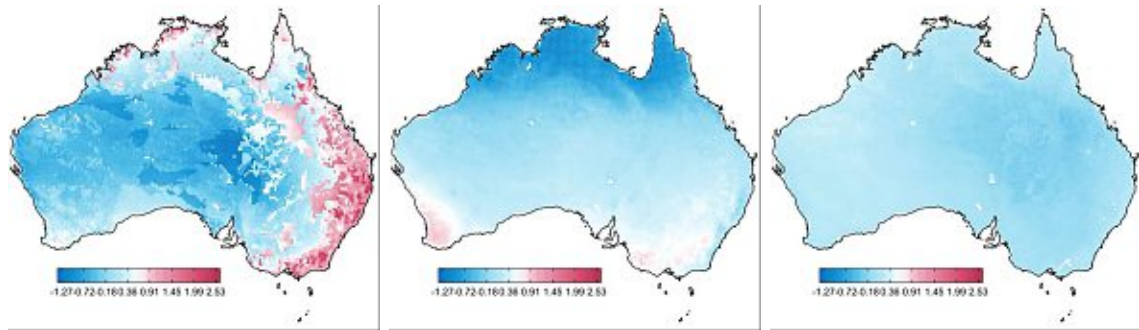
Figure 7.1: Monthly Soil Moisture Means: 1st - 3rd principal components (left to right)
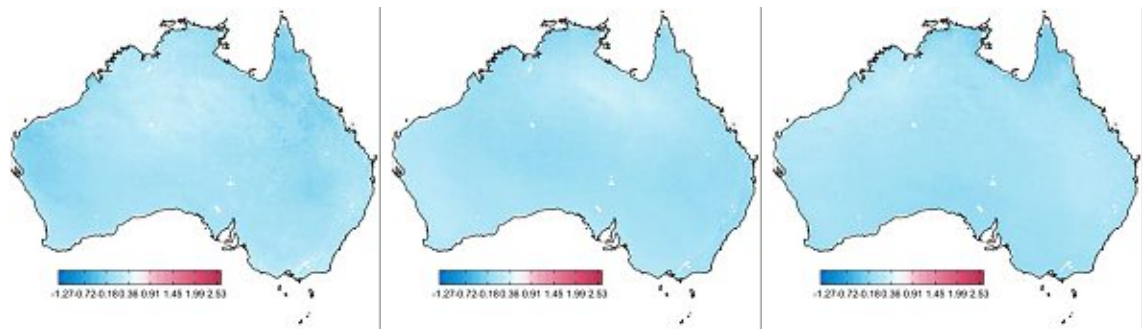
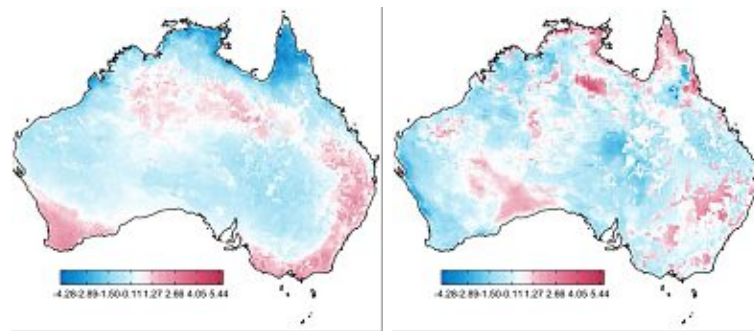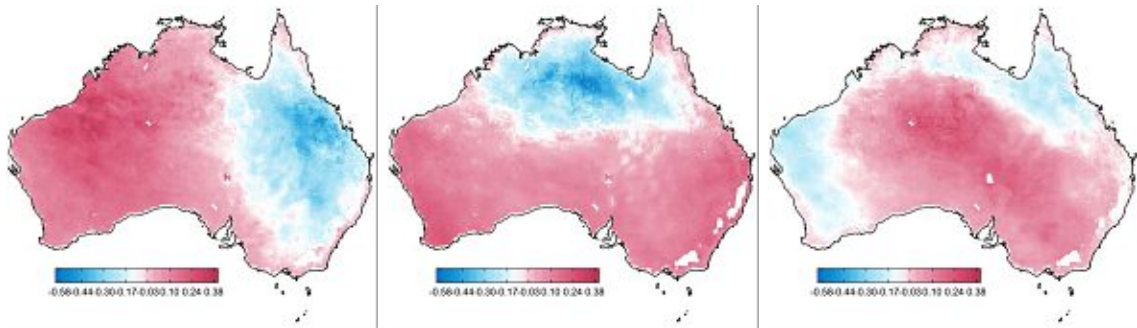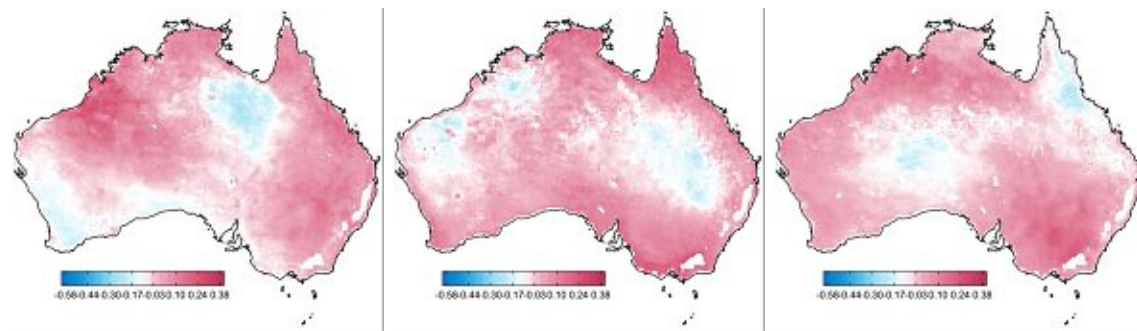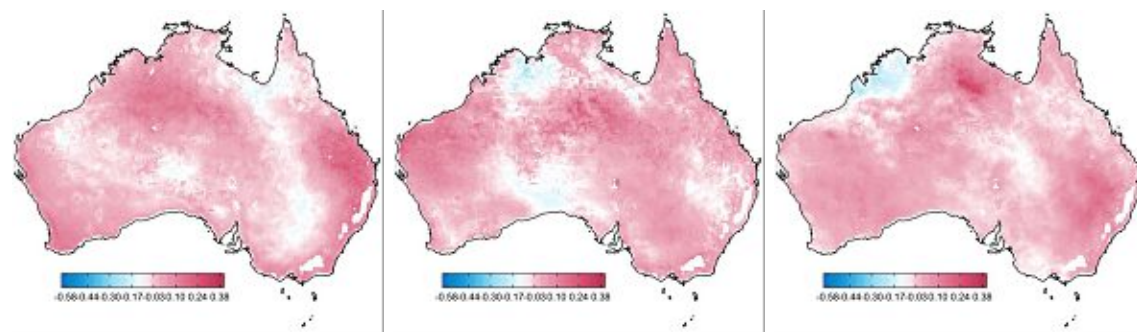Figure 7.2: Monthly Soil Moisture Means: 4th - 6th principal components (left to right)
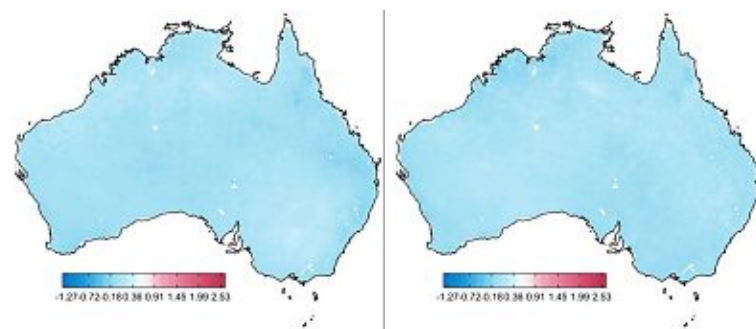
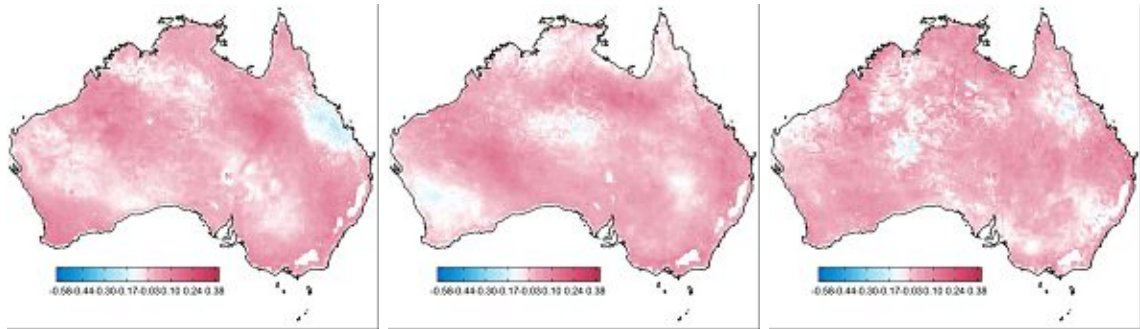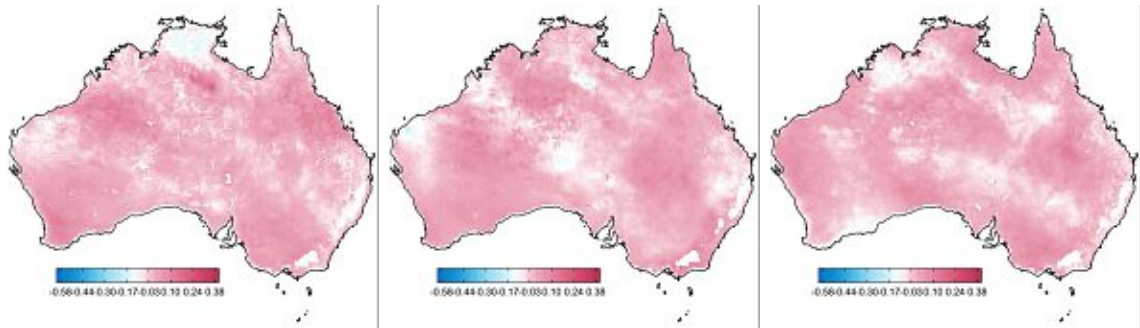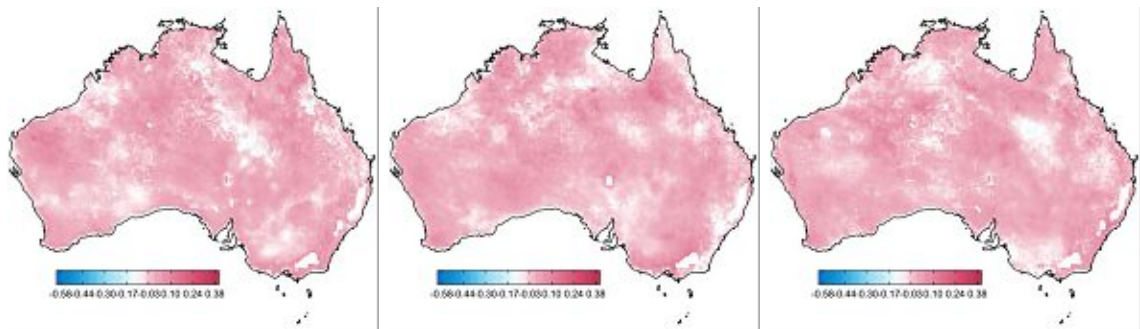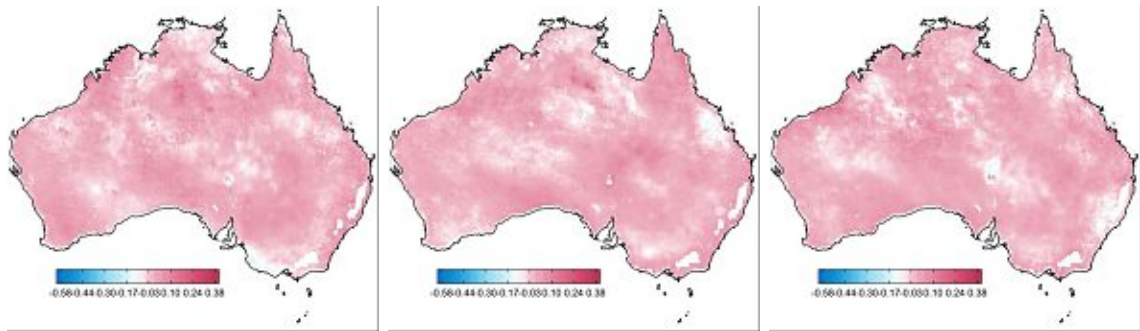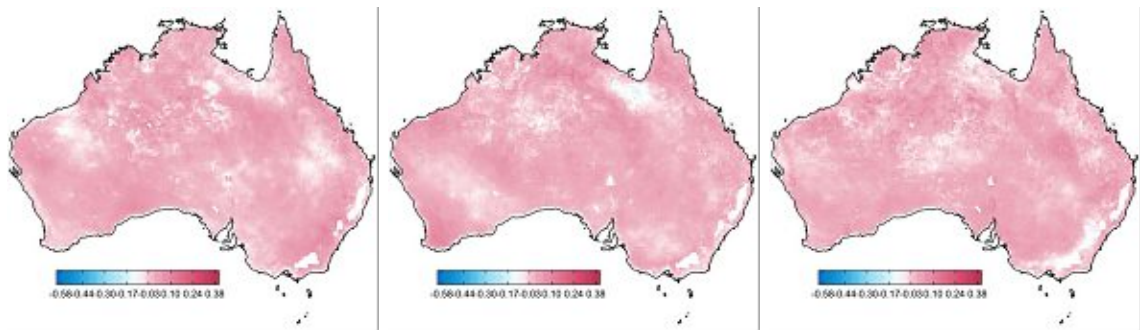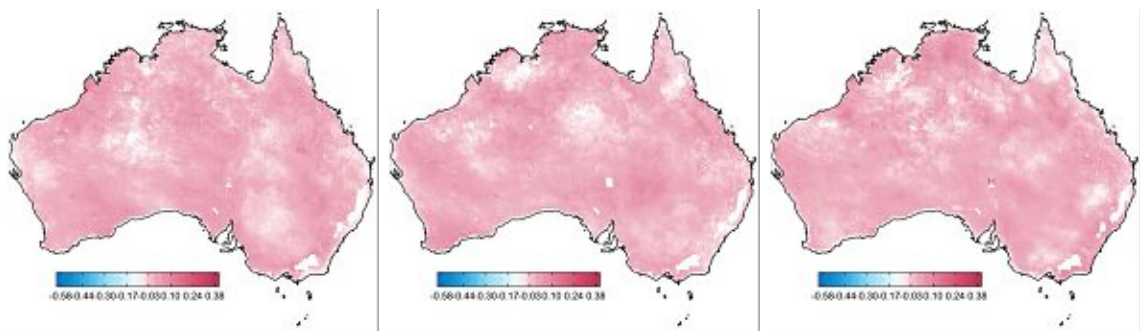Figure 7.3: Monthly Soil Moisture Means: 7th - 8th principal components (left to right)

Table 7.1: Correlation analysis of the PCA loadings of the Monthly Soil Moisture Means (mSMM) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.109	4	0.198	0	0.248	2
PC2	0.145	12	0.119	12	0.198	2
PC3	0.173	4	0.152	0	0.119	0
PC4	0.158	2	0.187	3	0.137	4
PC5	0.209	0	0.299	0	0.117	0
PC6	0.101	6	0.060	6	0.218	0
PC7	0.145	3	0.049	1	0.234	3
PC8	0.204	12	0.155	12	0.064	0

ICA
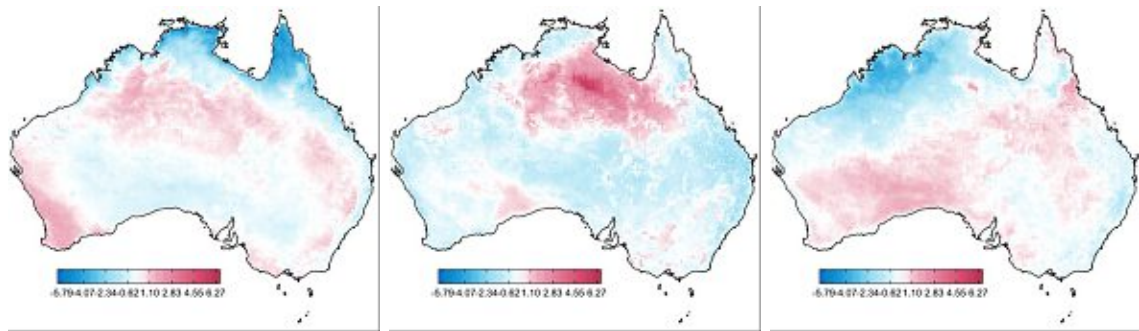
Figure 7.4: Monthly Soil Moisture Means: 1st - 3rd independent components (left to right)
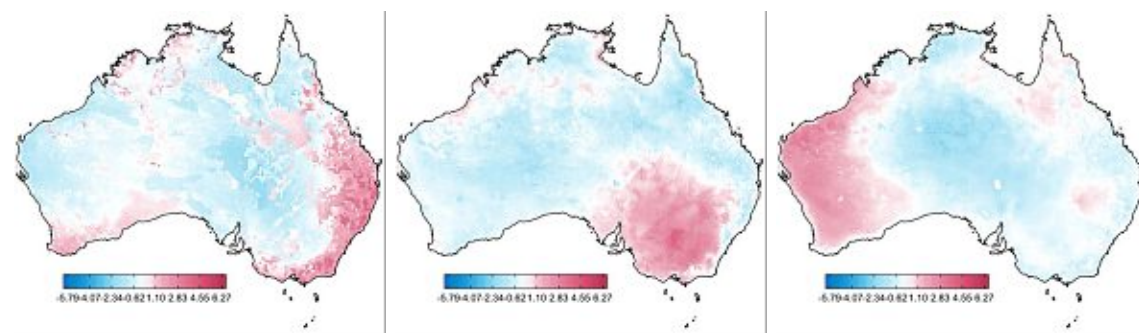
Figure 7.5: Monthly Soil Moisture Means: 4th - 6th independent components (left to right)
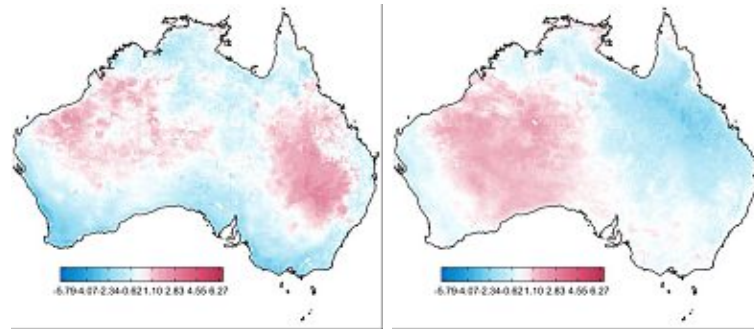
Figure 7.6: Monthly Soil Moisture Means: 7th - 8th independent components (left to right)

Table 7.2: Correlation analysis of the ICA loadings of the Monthly Soil Moisture Means (mSMM) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.177	6	44	0.177	3	305	0.279	0	102
IC2	0.217	0	39	0.266	1	233	0.200	3	191
IC3	0.250	12	29	0.204	4	14	0.234	0	25
IC4	0.233	12	50	0.237	1	225	0.277	3	27
IC5	0.171	12	191	0.142	12	20	0.240	3	113
IC6	0.223	6	112	0.304	0	165	0.264	0	23
IC7	0.239	12	37	0.222	0	147	0.235	0	203
IC8	0.244	12	287	0.289	0	287	0.284	0	87

SBSS
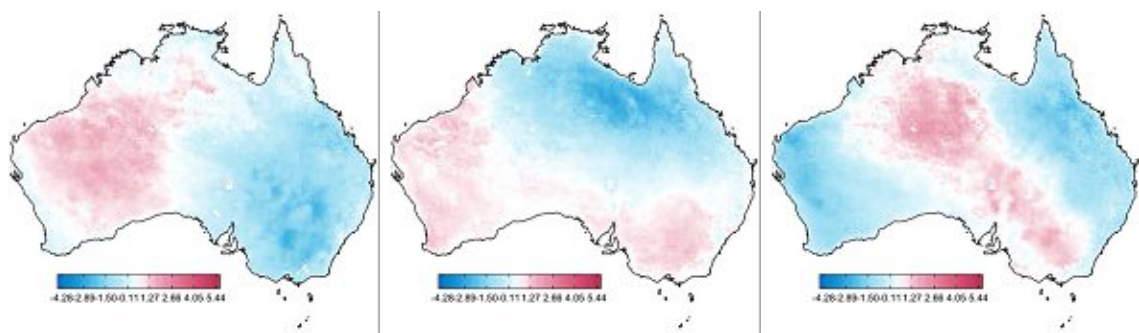
Figure 7.7: Monthly Soil Moisture Means: 1st - 3rd SBSS components (left to right)
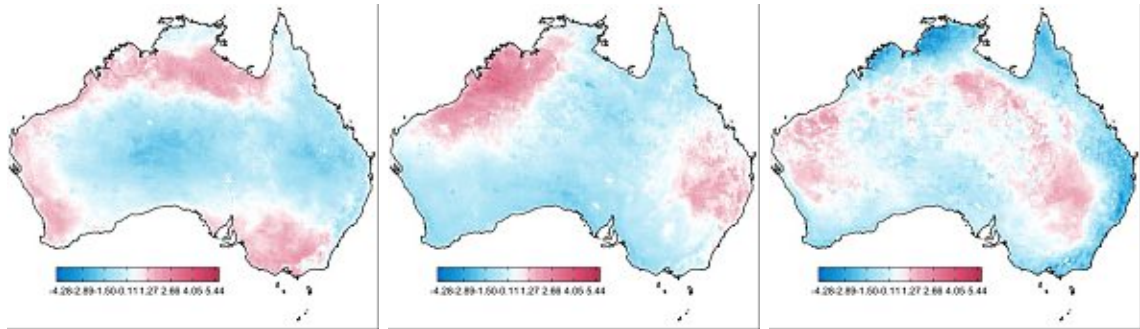
Figure 7.8: Monthly Soil Moisture Means: 4th - 6th SBSS components (left to right)

Figure 7.9: Monthly Soil Moisture Means: 7th - 8th SBSS components (left to right)

Table 7.3: Correlation analysis of the SBSS loadings of the Monthly Soil Moisture Means (mSMM) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.245	12	8	0.216	0	19	0.257	0	17
SBSS-C2	0.230	12	16	0.314	0	191	0.267	0	184
SBSS-C3	0.200	6	107	0.253	3	75	0.256	3	28
SBSS-C4	0.207	12	87	0.346	0	133	0.194	6	85
SBSS-C5	0.208	6	37	0.185	0	250	0.268	0	105
SBSS-C6	0.237	12	71	0.196	12	313	0.244	0	250
SBSS-C7	0.200	6	229	0.245	3	233	0.270	3	155
SBSS-C8	0.234	12	2	0.206	12	199	0.170	0	102

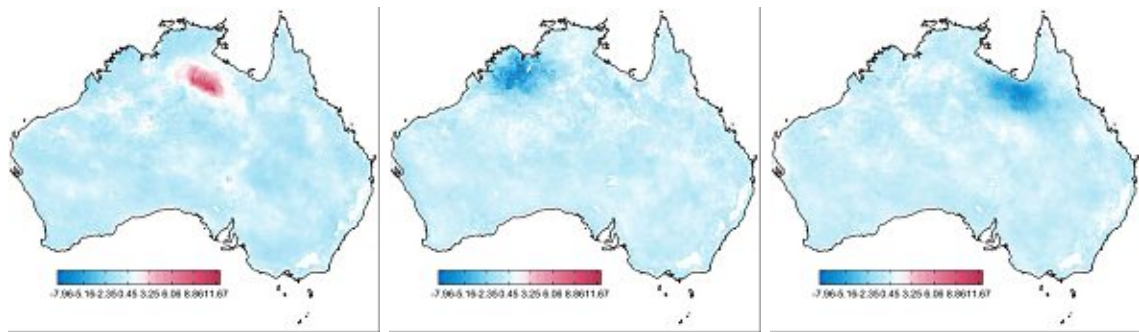### 7.4.2 Monthly Soil Moisture Anomalies 1998-2018

PCA

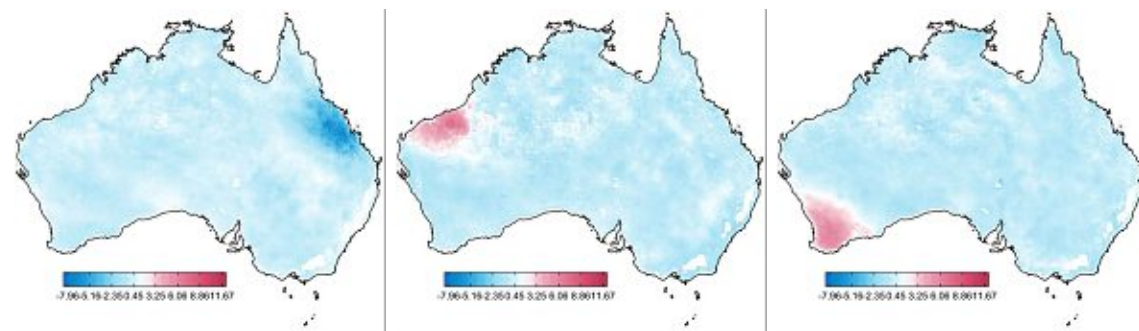Figure 7.10: Monthly Soil Moisture Anomalies: 1st - 3rd principal components (left to right)

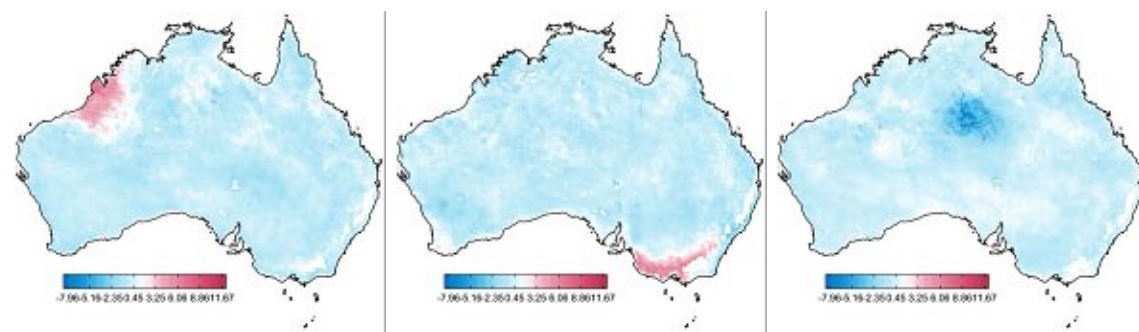Figure 7.11: Monthly Soil Moisture Anomalies: 4th - 6th principal components (left to right)

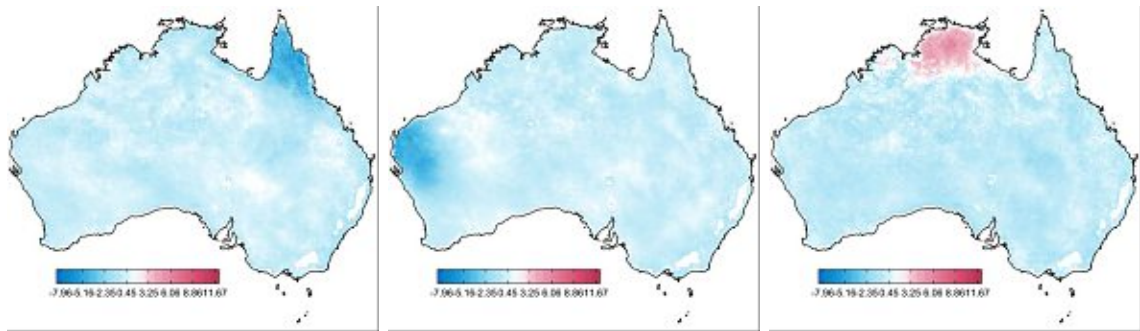Figure 7.12: Monthly Soil Moisture Anomalies: 7th - 9th principal components (left to right)

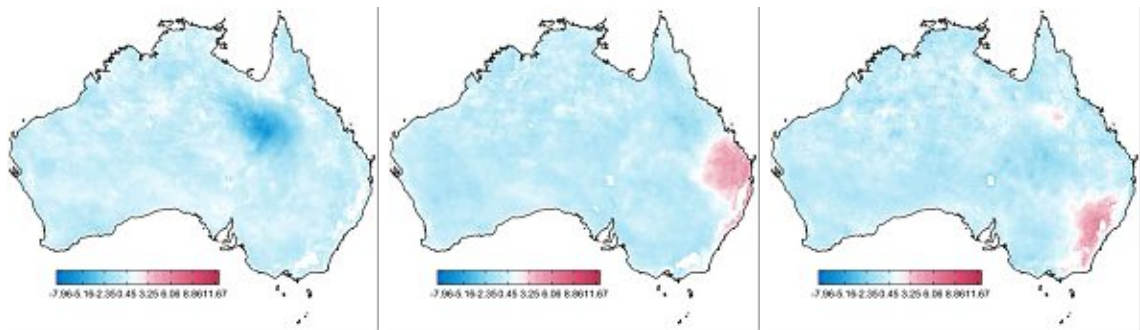Figure 7.13: Monthly Soil Moisture Anomalies: 10th - 12th principal components (left to right)

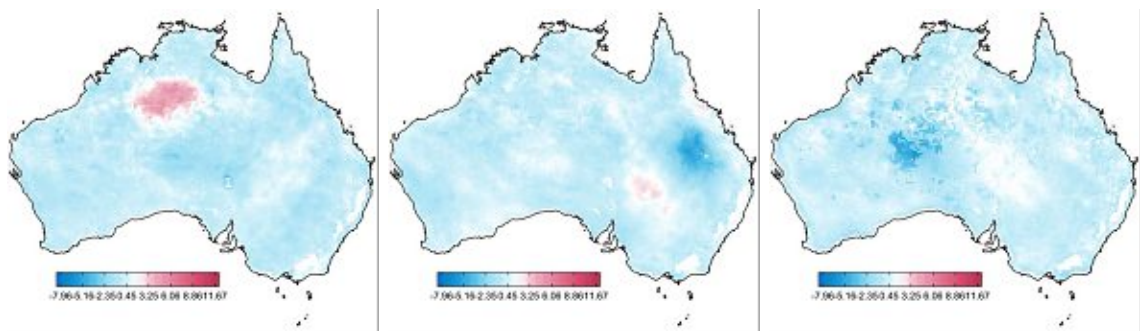Figure 7.14: Monthly Soil Moisture Anomalies: 13th - 15th principal components (left to right)

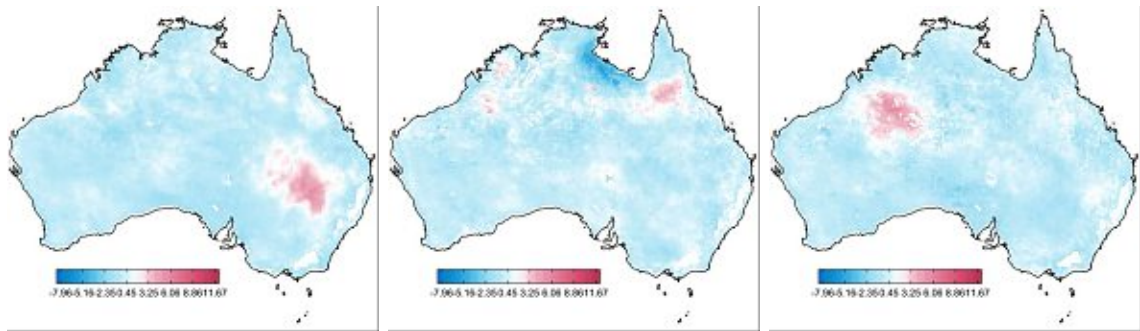Figure 7.15: Monthly Soil Moisture Anomalies: 16th - 18th principal components (left to right)

Figure 7.16: Monthly Soil Moisture Anomalies: 19th - 21st principal components (left to right)

Figure 7.17: Monthly Soil Moisture Anomalies: 22nd - 24th principal components (left to right)

Figure 7.18: Monthly Soil Moisture Anomalies: 25th - 27th principal components (left to right)
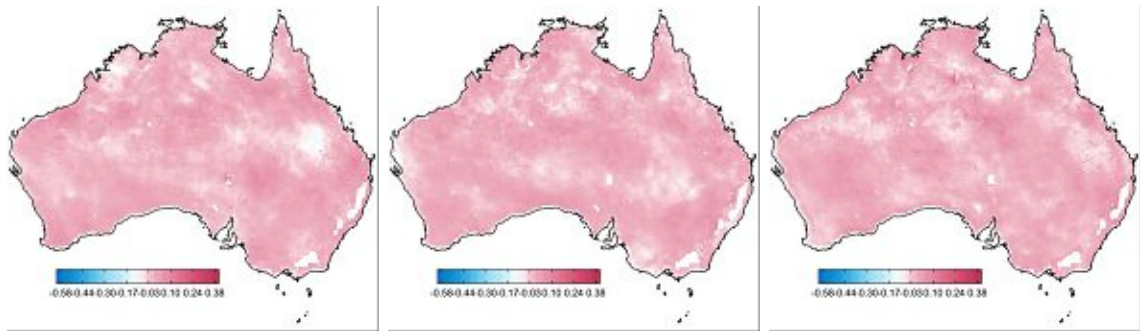
Figure 7.19: Monthly Soil Moisture Anomalies: 28th - 30th principal components (left to right)

Table 7.4: Correlation analysis of the PCA loadings of the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.138	12	0.157	0	0.169	2
PC2	0.160	6	0.176	1	0.225	6
PC3	0.174	0	0.110	2	0.242	6
PC4	0.068	6	0.172	0	0.324	3
PC5	0.101	3	0.109	3	0.184	0
PC6	0.104	12	0.117	1	0.098	3
PC7	0.080	0	0.164	4	0.102	4
PC8	0.166	4	0.178	4	0.061	2
PC9	0.210	12	0.083	12	0.079	1
PC10	0.099	3	0.185	1	0.172	4
PC11	0.164	0	0.122	2	0.115	12
PC12	0.337	1	0.133	4	0.243	3
PC13	0.058	0	0140	0	0.089	6
PC14	0.158	2	0.093	3	0.110	4
PC15	0.082	0	0.156	0	0.166	2
PC16	0.081	1	0.091	6	0.223	3
PC17	0.169	12	0.180	4	0.217	1
PC18	0.295	3	0.069	4	0.286	12
PC19	0.239	12	0.112	2	0.109	12
PC20	0.102	0	0.110	6	0.192	0
PC21	0.074	0	0.053	0	0.228	6
PC22	0.196	6	0.160	4	0.214	12
PC23	0.102	12	0.164	4	0.074	1
PC24	0.219	6	0.147	6	0.161	12
PC25	0.132	6	0.119	4	0.069	1
PC26	0.095	1	0.176	6	0.104	6
PC27	0.092	4	0.109	1	0.156	0
PC28	0.155	0	0.125	0	0.031	6
PC29	0.248	0	0.057	12	0.199	6
PC30	0.160	2	0.155	6	0.115	12

ICA

Figure 7.20: Monthly Soil Moisture Anomalies: 1st - 3rd independent components (left to right)

Figure 7.21: Monthly Soil Moisture Anomalies: 4th - 6th independent components (left to right)

Figure 7.22: Monthly Soil Moisture Anomalies: 7th - 9th independent components (left to right)

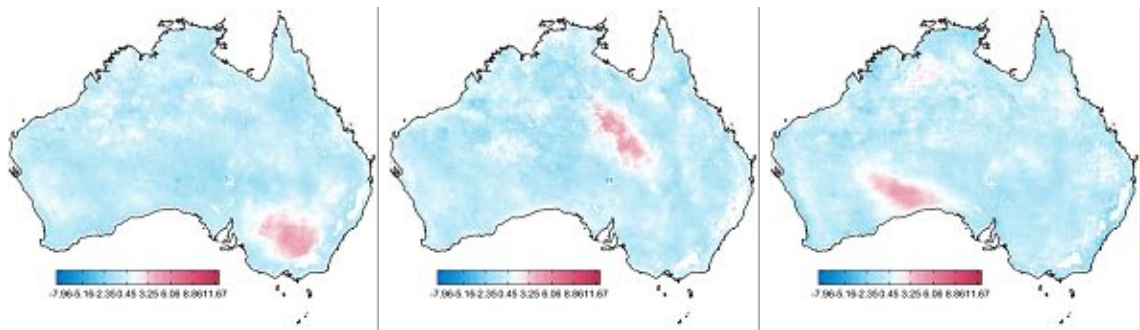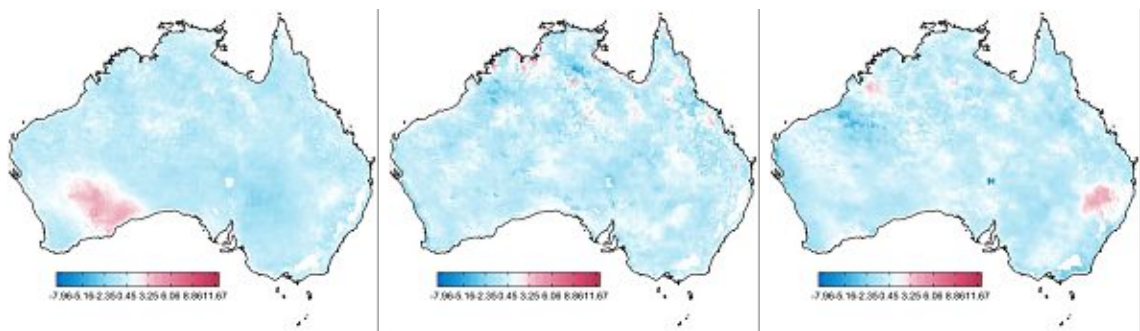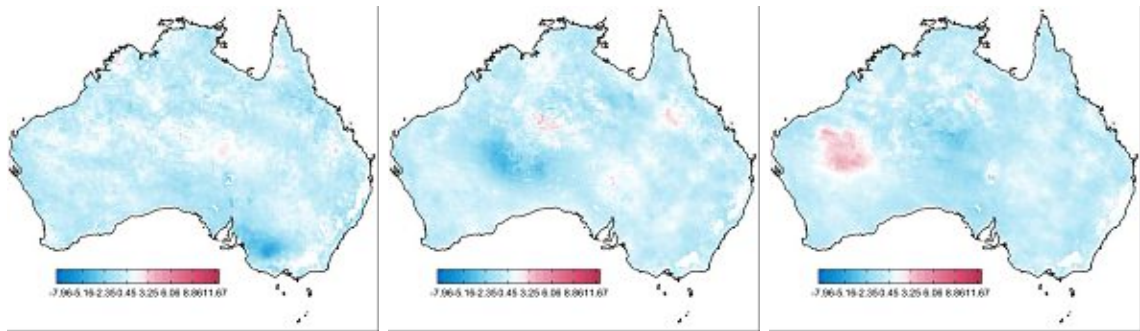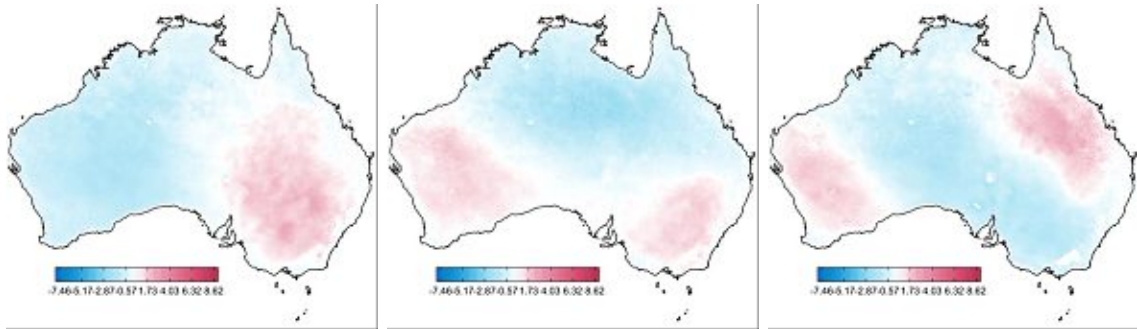Figure 7.23: Monthly Soil Moisture Anomalies: 10th - 12th independent components (left to right)

Figure 7.24: Monthly Soil Moisture Anomalies: 13th - 15th independent components (left to right)

Figure 7.25: Monthly Soil Moisture Anomalies: 16th - 18th independent components (left to right)

Figure 7.26: Monthly Soil Moisture Anomalies: 19th - 21st independent components (left to right)

Figure 7.27: Monthly Soil Moisture Anomalies: 22nd - 24th independent components (left to right)
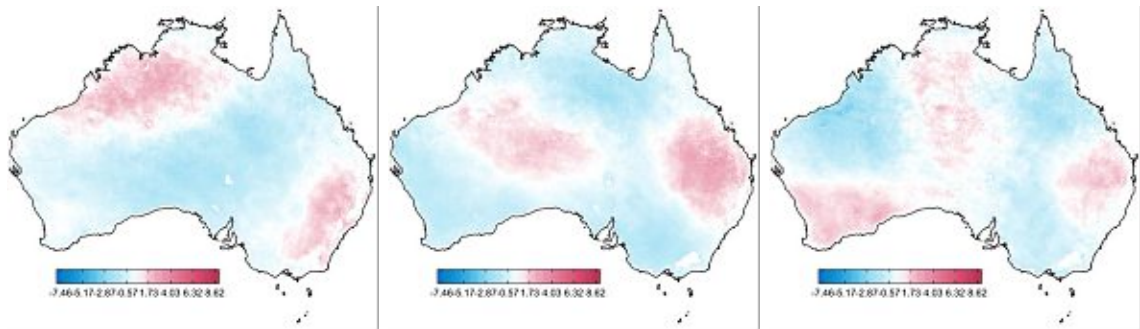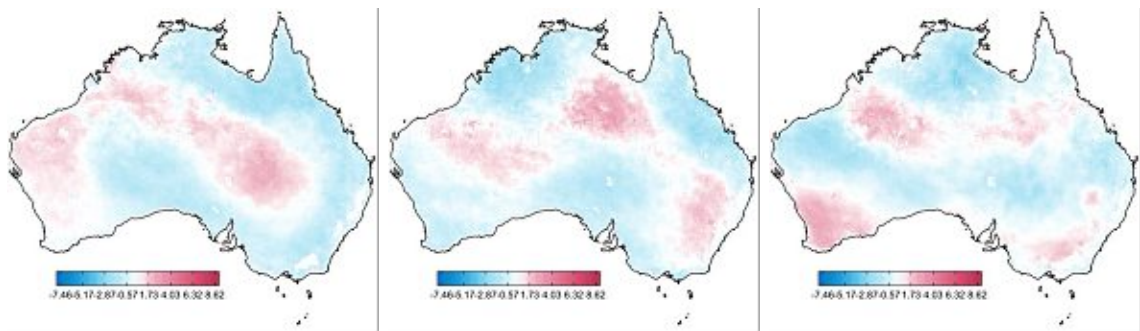

Figure 7.28: Monthly Soil Moisture Anomalies: 25th - 27th independent components (left to right)

Figure 7.29: Monthly Soil Moisture Anomalies: 28th - 30th independent components (left to right)

Table 7.5: Correlation analysis of the ICA loadings of the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.365	0	1581	0.297	0	1719	0.330	0	1401
IC2	0.262	6	1109	0.338	6	1190	0.284	0	49
IC3	0.336	0	1529	0.275	4	245	0.332	1	127
IC4	0.268	12	1758	0.324	0	1832	0.362	0	1695
IC5	0.381	0	968	0.305	4	537	0.357	12	1472
IC6	0.291	60	1952	0.307	0	899	0.330	0	387
IC7	0.354	12	1666	0.285	6	1328	0.357	12	327
IC8	0.374	12	805	0.315	6	106	0.397	12	1618
IC9	0.251	6	883	0.311	6	556	0.321	0	1000
IC10	0.233	0	1419	0.274	4	943	0.415	0	1655
IC11	0.372	0	904	0.330	4	1526	0.368	1	1282
IC12	0.264	12	1774	0.281	6	1763	0.405	3	1979
IC13	0.372	6	83	0.299	4	1293	0.440	12	1641
IC14	0.373	0	798	0.281	6	1967	0.283	3	328
IC15	0.368	0	1924	0.285	4	1224	0.443	12	1349
IC16	0.336	0	281	0.309	6	772	0.388	12	1079
IC17	0.374	0	1646	0.260	6	1985	0.292	12	1487
IC18	0.393	1	1352	0.257	4	1977	0.391	12	1436
IC19	0.394	0	1916	0.240	4	1501	0.335	6	1363
IC20	0.315	12	1266	0.338	6	1625	0.366	0	1817
IC21	0.322	0	318	0.323	0	749	0.382	12	1261
IC22	0.322	6	1088	0.326	6	342	0.359	12	44
IC23	0.383	1	716	0.271	6	957	0.420	12	124
IC24	0.382	0	717	0.334	4	197	0.398	12	1042
IC25	0.327	12	1249	0.282	6	1145	0.367	12	1663
IC26	0.275	1	1537	0.332	6	1586	0.329	0	1419
IC27	0.322	1	991	0.309	6	1786	0.318	12	657
IC28	0.400	0	1638	0.278	4	1837	0.430	12	1988
IC29	0.371	0	1745	0.244	4	860	0.357	12	1913
IC30	0.358	1	1939	0.256	3	467	0.374	12	293

SBSS

Figure 7.30: Monthly Soil Moisture Anomalies: 1st - 3rd SBSS components (left to right)

Figure 7.31: Monthly Soil Moisture Anomalies: 4th - 6th SBSS components (left to right)

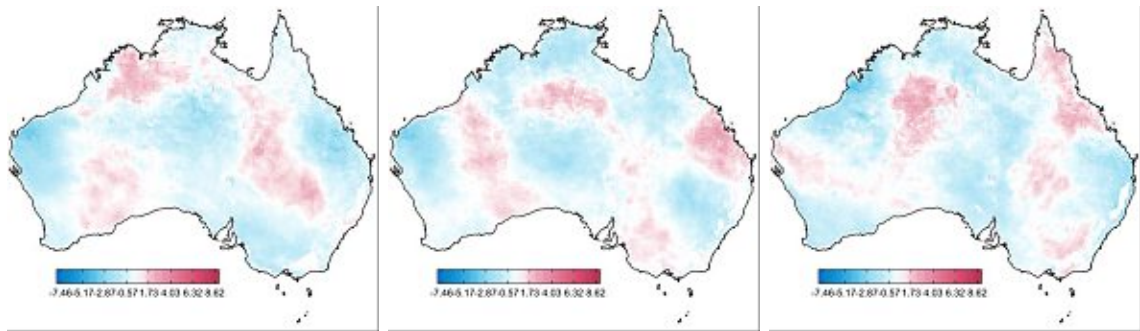Figure 7.32: Monthly Soil Moisture Anomalies: 7th - 9th SBSS components (left to right)

Figure 7.33: Monthly Soil Moisture Anomalies: 10th - 12th SBSS components (left to right)
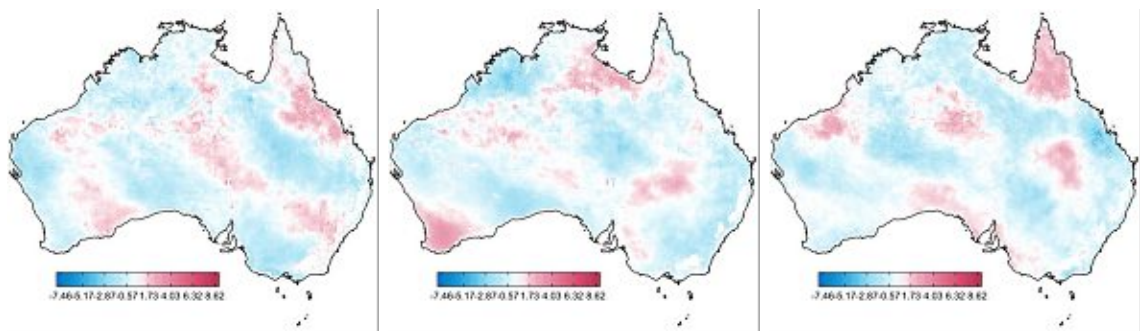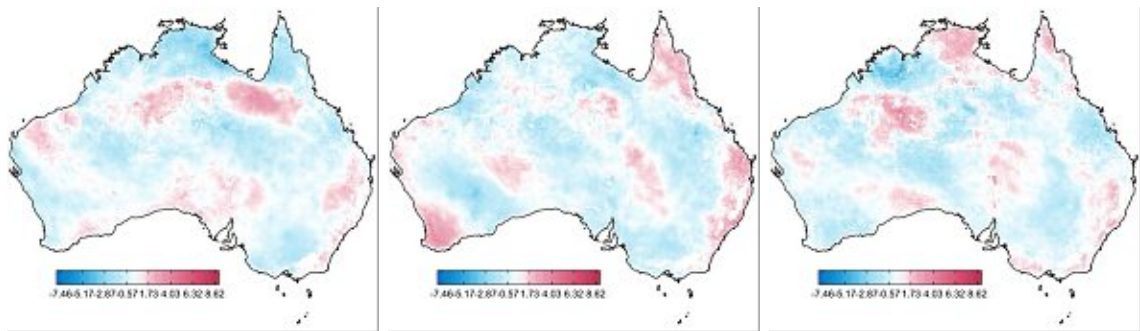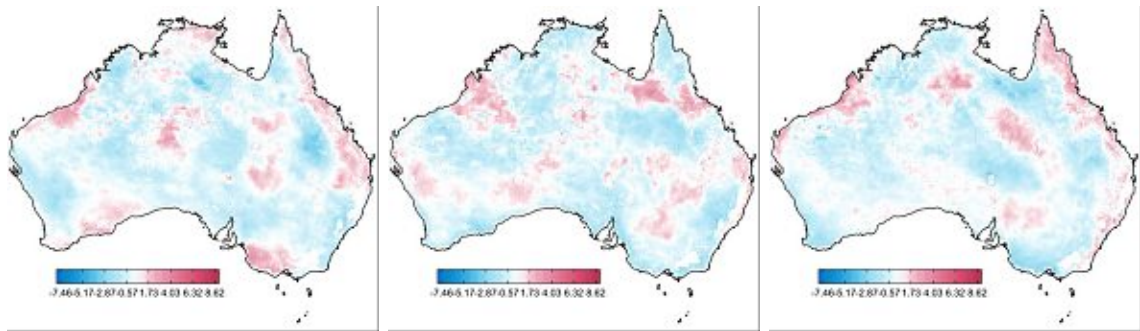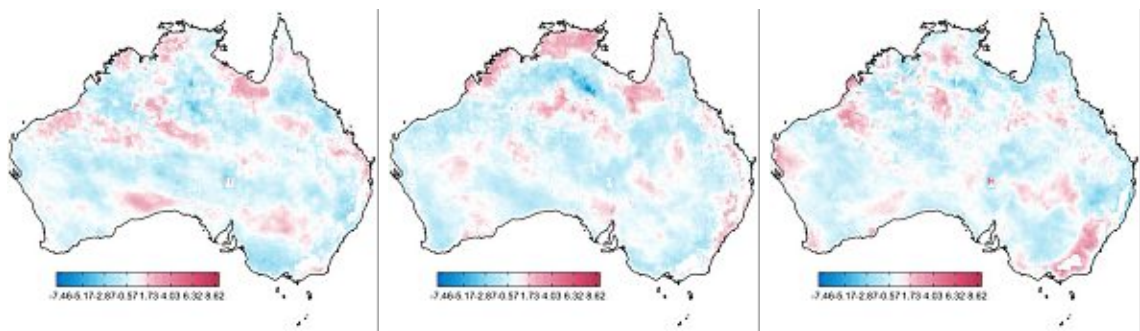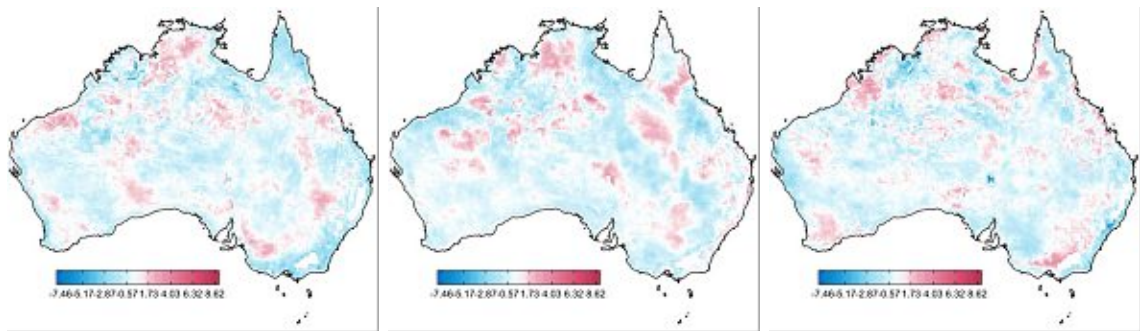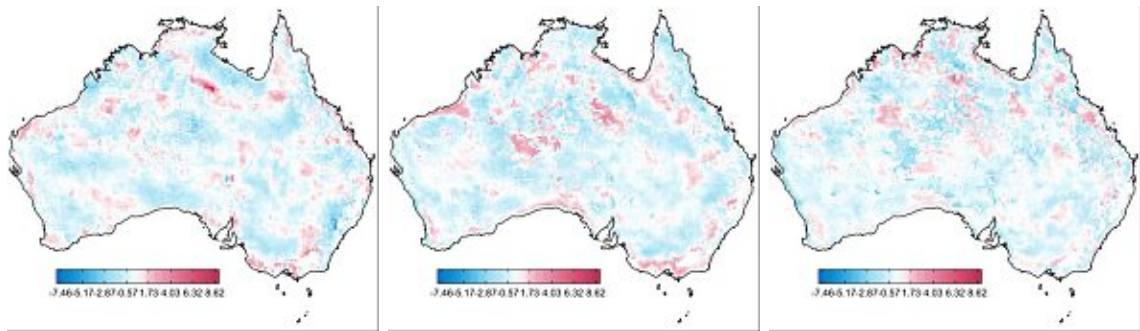

Figure 7.34: Monthly Soil Moisture Anomalies: 13th - 15th SBSS components (left to right)

Figure 7.35: Monthly Soil Moisture Anomalies: 16th - 18th SBSS components (left to right)

Figure 7.36: Monthly Soil Moisture Anomalies: 19th - 21st SBSS components (left to right)

Figure 7.37: Monthly Soil Moisture Anomalies: 22nd - 24th SBSS components (left to right)

Figure 7.38: Monthly Soil Moisture Anomalies: 25th - 27th SBSS components (left to right)

Figure 7.39: Monthly Soil Moisture Anomalies: 28th - 30th SBSS components (left to right)

Table 7.6: Correlation analysis of the SBSS loadings of the Monthly Soil Moisture Anomalies (mSMA) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.330	12	289	0.298	2	553	0.317	0	1337
SBSS-C2	0.328	0	1727	0.291	1	428	0.411	4	1226
SBSS-C3	0.387	0	1519	0.293	3	888	0.350	1	668
SBSS-C4	0.293	0	1076	0.314	0	1048	0.371	3	173
SBSS-C5	0.263	6	1947	0.318	3	1473	0.462	3	1465
SBSS-C6	0.360	0	277	0.300	3	523	0.367	6	1678
SBSS-C7	0.330	0	1305	0.307	4	1046	0.356	6	165
SBSS-C8	0.356	1	366	0.320	6	1959	0.325	12	1080
SBSS-C9	0.334	0	788	0.285	0	492	0.317	1	838
SBSS-C10	0.337	1	305	0.299	6	1756	0.298	0	736
SBSS-C11	0.360	1	611	0.317	6	322	0.352	12	1059
SBSS-C12	0.339	0	518	0.294	4	1305	0.353	12	733
SBSS-C13	0.332	0	952	0.286	6	821	0.377	12	1087
SBSS-C14	0.358	0	370	0.310	6	1793	0.349	12	264
SBSS-C15	0.273	6	1695	0.324	6	714	0.266	12	438
SBSS-C16	0.312	0	614	0.279	4	301	0.316	0	74
SBSS-C17	0.293	12	1756	0.256	4	1610	0.273	1	1132
SBSS-C18	0.264	1	318	0.304	6	1495	0.382	0	867
SBSS-C19	0.356	0	512	0.224	0	1442	0.315	12	1904
SBSS-C20	0.399	0	1300	0.298	6	1941	0.361	0	650
SBSS-C21	0.298	0	1634	0.282	4	60	0.421	0	203
SBSS-C22	0.353	0	1149	0.229	6	2	0.405	12	1344
SBSS-C23	0.330	1	25	0.289	6	1157	0.311	0	1660
SBSS-C24	0.287	0	457	0.266	6	126	0.302	12	1791
SBSS-C25	0.308	0	461	0.271	6	1881	0.350	1	28
SBSS-C26	0.355	0	1177	0.287	6	1219	0.355	12	391
SBSS-C27	0.200	6	107	0.253	3	75	0.256	3	28
SBSS-C28	0.368	0	396	0.438	6	357	0.438	12	1449
SBSS-C29	0.347	1	1913	0.269	6	1937	0.384	12	115
SBSS-C30	0.366	1	662	0.335	6	563	0.341	12	1363

7.4.3 Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 - Summer
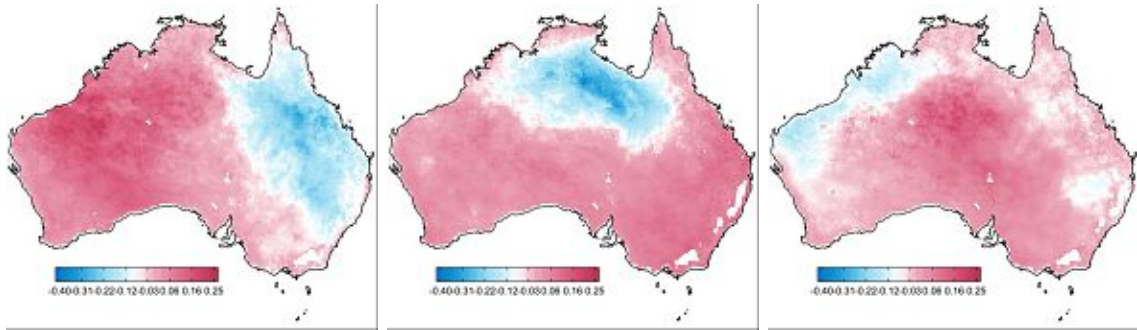
PCA

Figure 7.40: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 1st - 3rd principal components (left to right)
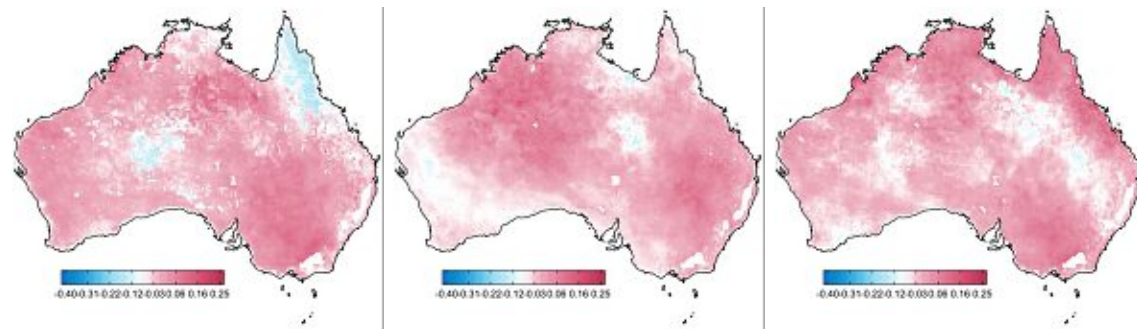

Figure 7.41: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 4th - 6th principal components (left to right)
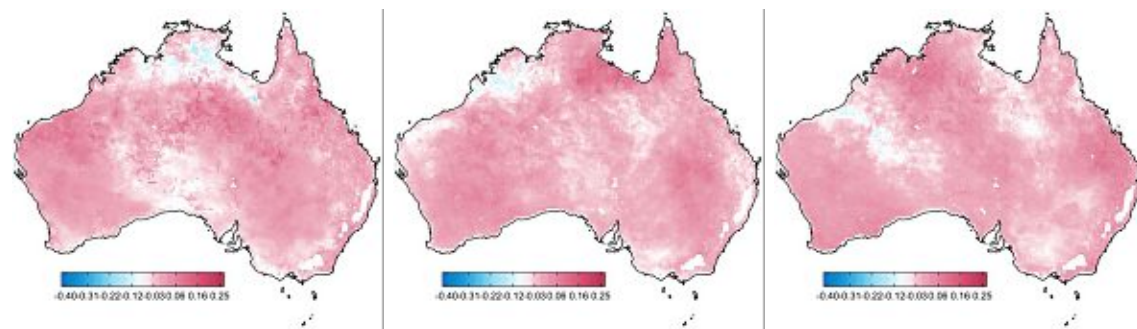

Figure 7.42: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 7th - 9th principal components (left to right)
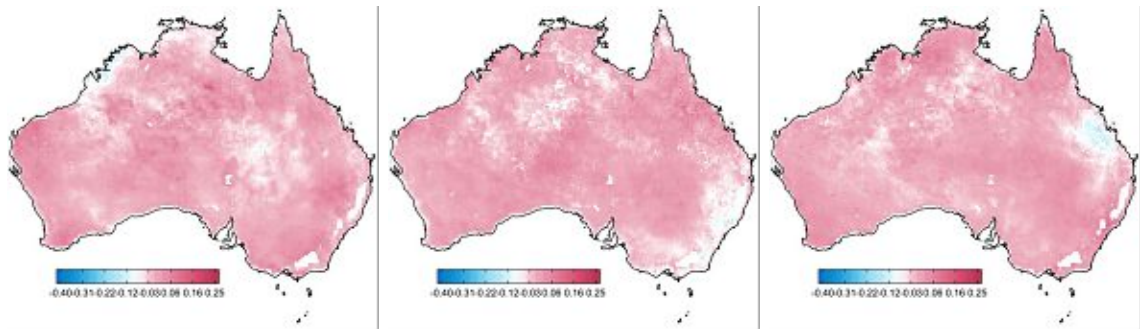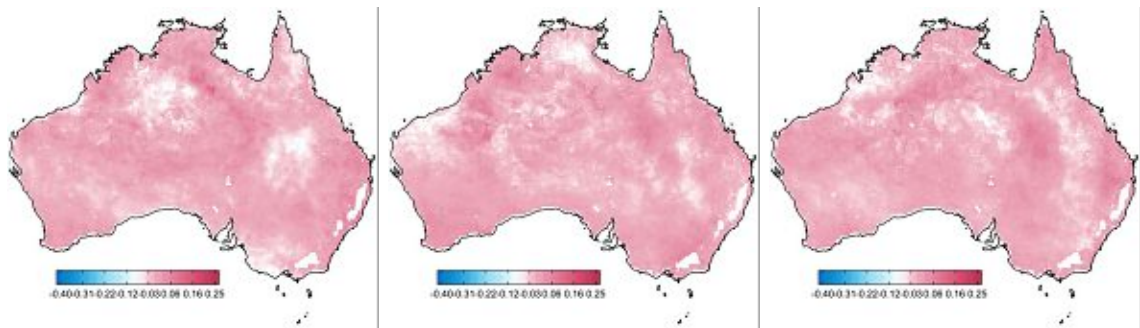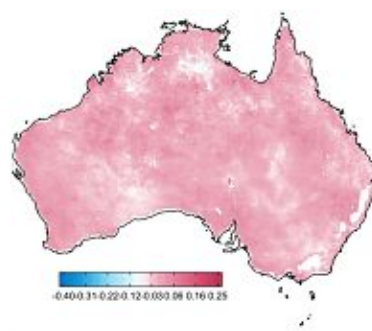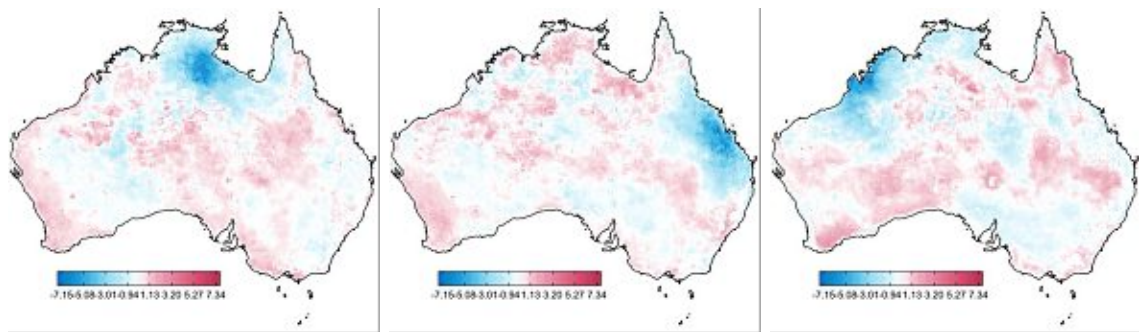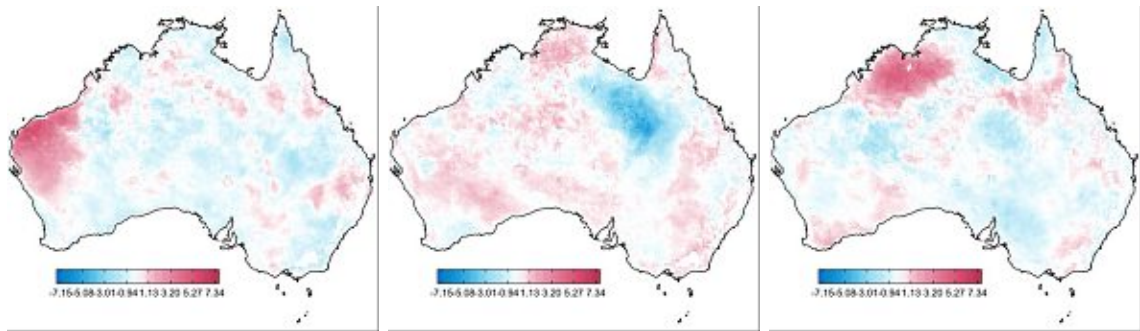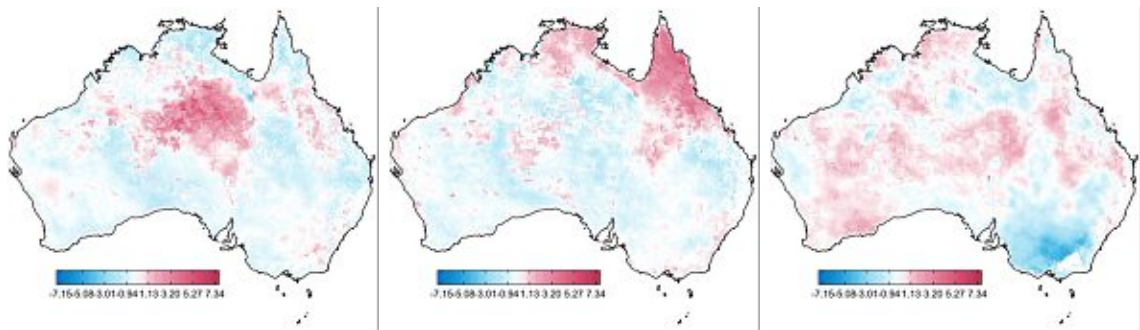
Figure 7.43: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 10th - 12th principal components (left to right)

Figure 7.44: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 13th - 15th principal components (left to right)
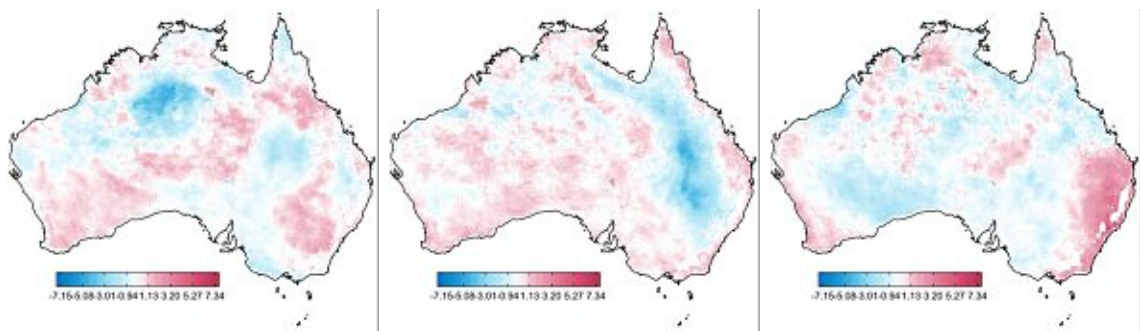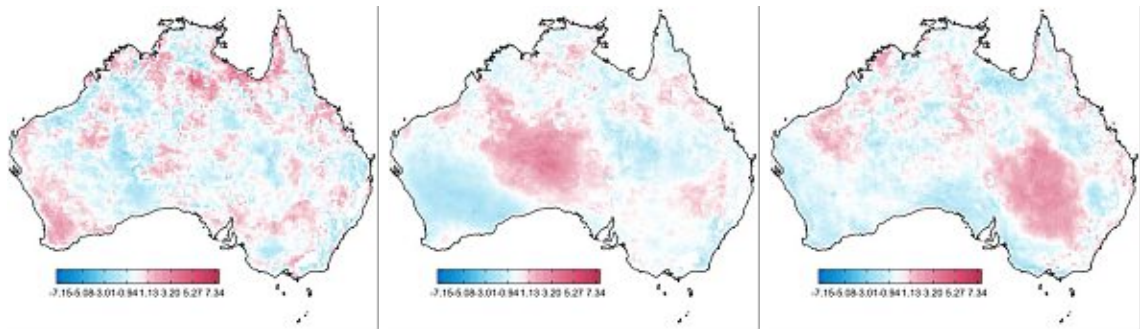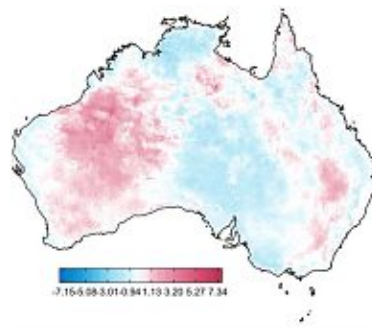

Figure 7.45: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 16th principal component

Table 7.7: Correlation analysis of the PCA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Summer) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.402	4	0.269	2	0.248	12
PC2	0.338	6	0.425	0	0.107	12
PC3	0.254	12	0.121	4	0.489	1
PC4	0.183	1	0.326	1	0.329	2
PC5	0.277	12	0.239	1	0.297	0
PC6	0.316	12	0.144	3	0.369	4
PC7	0.057	12	0.136	2	0.246	12
PC8	0.406	3	0.369	6	0.236	6
PC9	0.163	12	0.190	6	0.279	4
PC10	0.393	2	0.137	4	0.270	0
PC11	0.368	0	0.289	3	0.195	6
PC12	0.344	1	0.163	12	0.185	0
PC13	0.424	3	0.283	0	0.256	12
PC14	0.250	0	0.180	6	0.264	6
PC15	0.177	1	0.178	0	0.212	3
PC16	0.105	6	0.329	2	0.213	0

ICA

Figure 7.46: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 1st - 3rd independent components (left to right)

Figure 7.47: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 4th - 6th independent components (left to right)
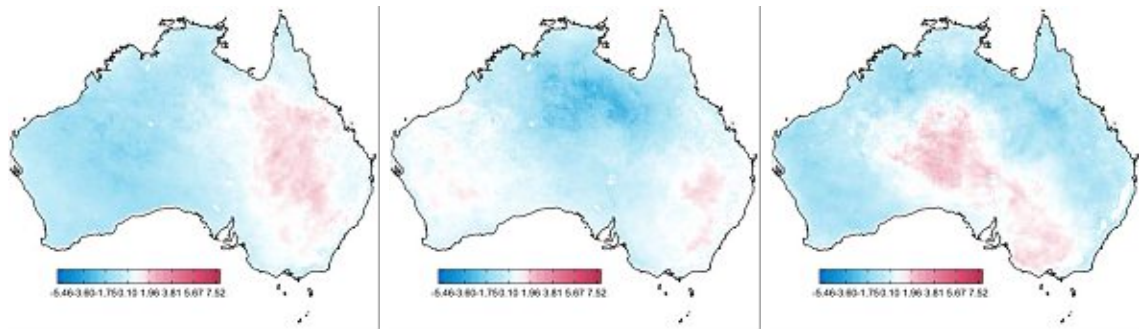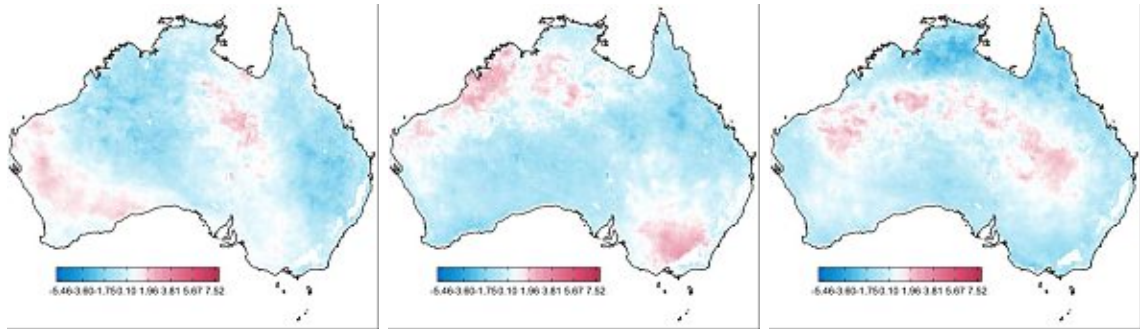

Figure 7.48: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 7th - 9th independent components (left to right)

Figure 7.49: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 10th - 12th independent components (left to right)

Decomposition of Soil Moisture Anomalies (Summer): 13th - 15th independent components (left to right)

Figure 7.50: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 16th independent component

Table 7.8: Correlation analysis of the ICA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Summer) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.586	3	222	0.509	3	411	0.497	6	678
IC2	0.504	1	165	0.359	0	831	0.423	4	636
IC3	0.544	3	479	0.421	3	313	0.511	3	1078
IC4	0.568	3	827	0.418	3	1300	0.561	4	611
IC5	0.542	3	1940	0.453	6	1107	0.501	4	1374
IC6	0.552	3	655	0.430	3	1763	0.572	4	1557
IC7	0.420	3	1871	0.332	0	1773	0.494	0	1477
IC8	0.476	1	984	0.470	0	877	0.552	4	984
IC9	0.551	2	238	0.532	3	1643	0.550	4	608
IC10	0.524	3	1172	0.431	0	1472	0.384	12	1224
IC11	0.592	3	413	0.412	6	1435	0.501	6	1935
IC12	0.575	1	1677	0.430	0	1562	0.373	6	1181
IC13	0.456	3	16	0.513	3	36	0.448	6	1777
IC14	0.591	3	662	0.411	3	1278	0.591	4	794
IC15	0.55	1	760	0.469	3	244	0.517	0	851
IC16	0.566	3	1601	0.432	3	1398	0.647	4	17

SBSS

Figure 7.51: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 1st - 3rd SBSS components (left to right)

Figure 7.52: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 4th - 6th SBSS components (left to right)
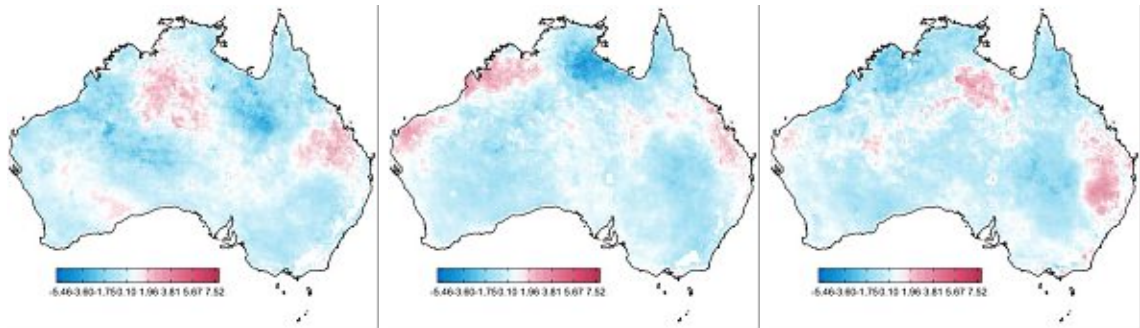

Figure 7.53: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 7th - 9th SBSS components (left to right)
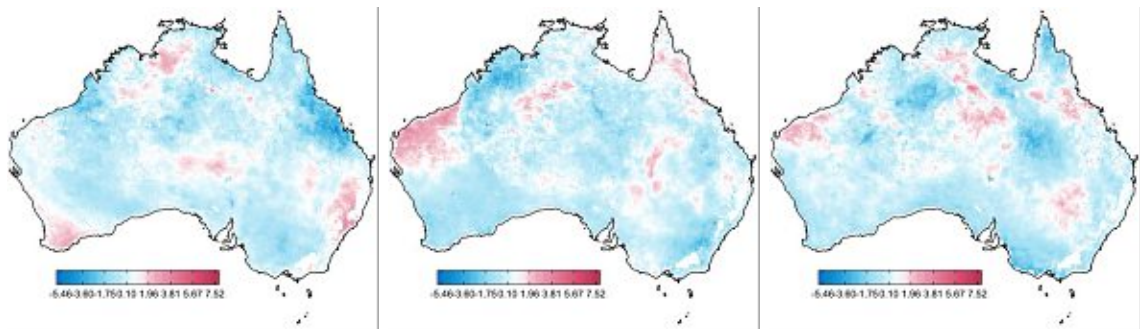

Figure 7.54: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 10th - 12th SBSS components (left to right)
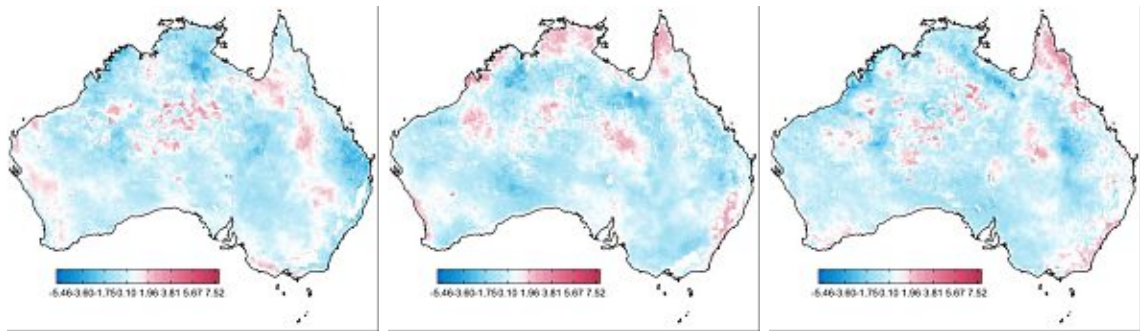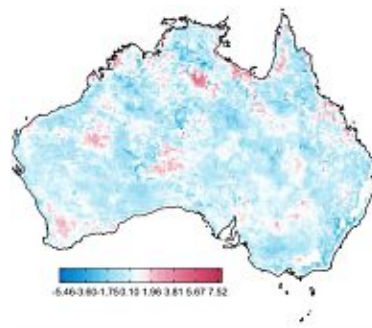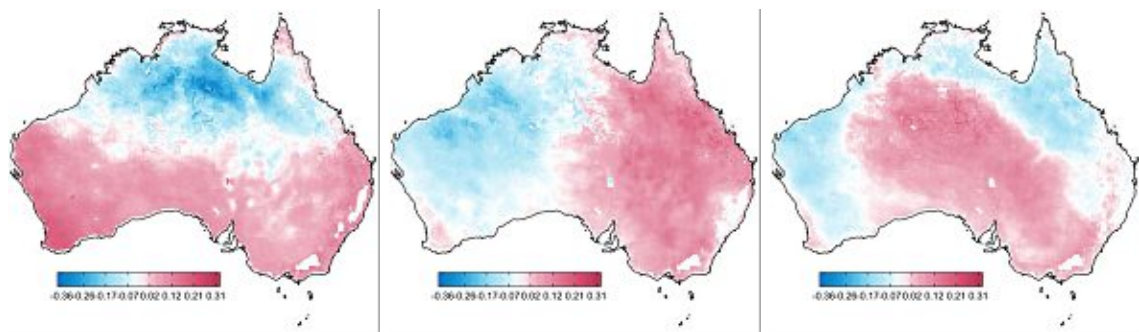
Figure 7.55: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 13th - 15th SBSS components (left to right)

Figure 7.56: Seasonal Decomposition of Soil Moisture Anomalies (Summer): 16th SBSS component

Table 7.9: Correlation analysis of the SBSS loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Summer) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.461	1	958	0.435	2	648	0.499	2	1278
SBSS-C2	0.514	4	1173	0.494	0	854	0.456	0	1261
SBSS-C3	0.515	3	1054	0.368	6	1899	0.680	4	431
SBSS-C4	0.438	1	701	0.381	1	1422	0.494	6	21
SBSS-C5	0.531	1	816	0.49	3	1382	0.571	6	1974
SBSS-C6	0.488	2	487	0.380	3	850	0.579	2	790
SBSS-C7	0.413	1	946	0.382	6	1700	0.501	4	64
SBSS-C8	0.497	3	660	0.497	3	1180	0.433	3	1971
SBSS-C9	0.575	1	953	0.347	0	859	0.403	0	1567
SBSS-C10	0.503	1	1432	0.397	0	1511	0.458	4	1631
SBSS-C11	0.369	3	1566	0.325	6	893	0.362	3	1371
SBSS-C12	0.540	3	6	0.395	0	1354	0.435	12	1917
SBSS-C13	0.498	3	1609	0.1353	3	159	0.377	12	1645
SBSS-C14	0.516	1	252	0.456	3	247	0.486	0	1680
SBSS-C15	0.465	1	15	0.373	0	896	0.463	4	1398
SBSS-C16	0.469	3	491	0.509	3	597	0.419	6	1413

7.4.4 Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 - Fall PCA

Figure 7.57: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 1st - 3rd principal components (left to right)
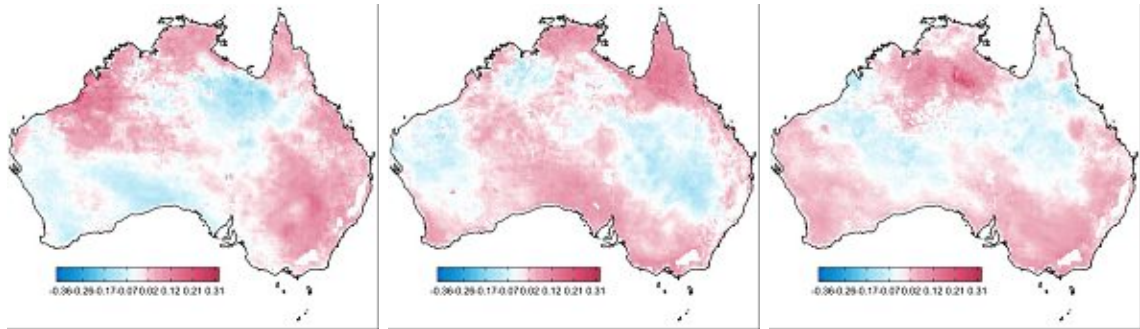
Figure 7.58: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 4th - 6th principal components (left to right)
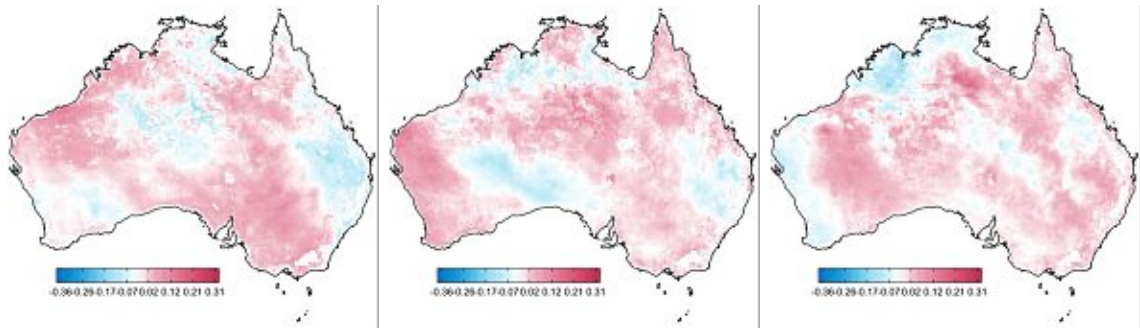

Figure 7.59: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 7th - 9th principal components (left to right)
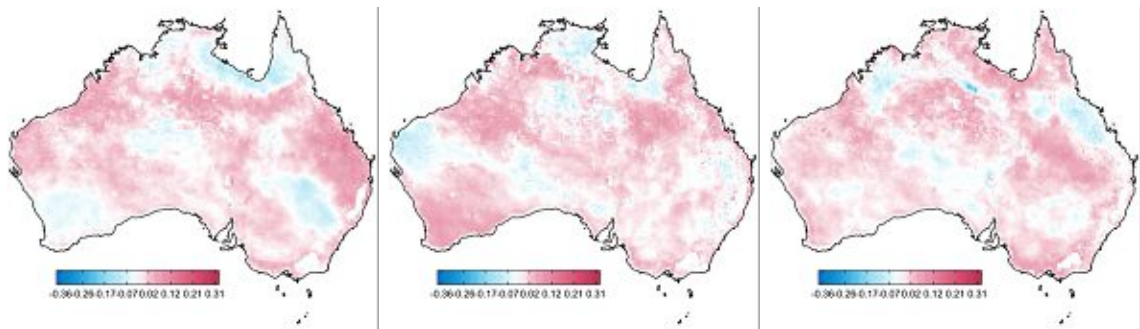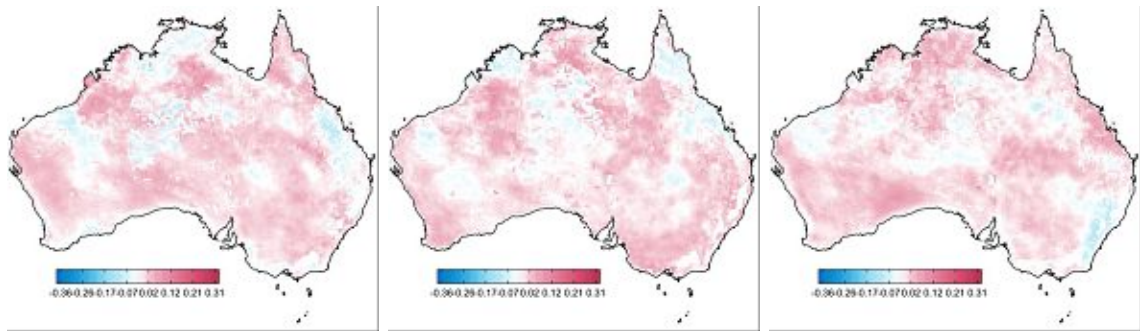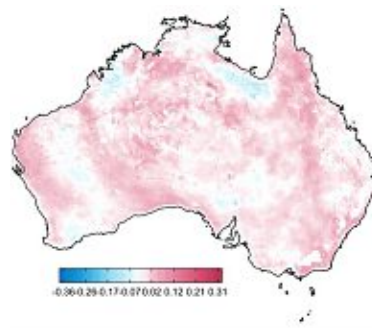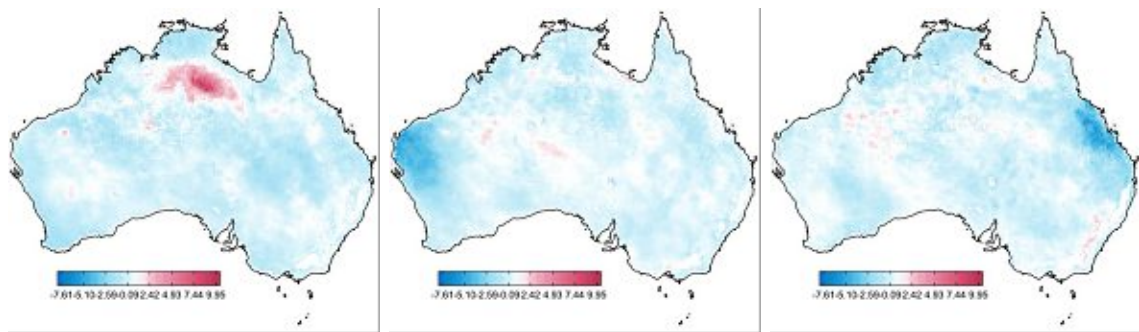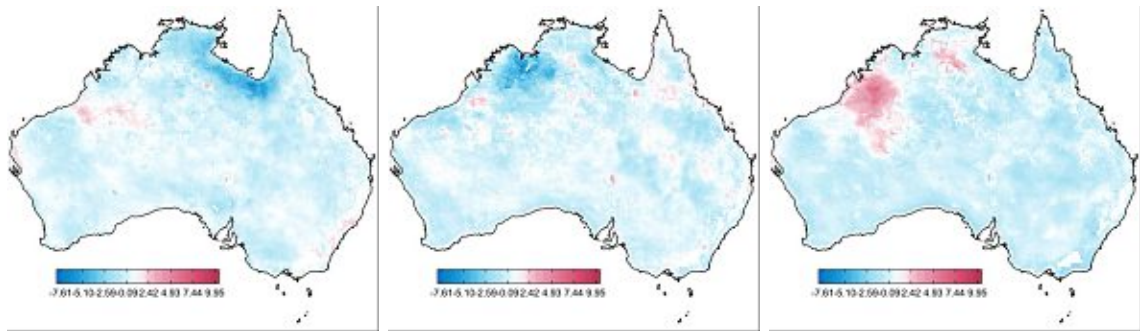

Figure 7.60: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 10th - 12th principal components (left to right)

Figure 7.61: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 13th - 15th principal components (left to right)

Figure 7.62: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 16th principal component

Table 7.10: Correlation analysis of the PCA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Fall) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.319	1	0.318	1	0.445	0
PC2	0.340	4	0.183	0	0.233	6
PC3	0.190	6	0.123	3	0.362	4
PC4	0.135	2	0.178	0	0.308	0
PC5	0.171	12	0.260	6	0.200	6
PC6	0.345	3	0.540	4	0.130	6
PC7	0.160	6	0.169	3	0.177	1
PC8	0.405	2	0.329	2	0.155	3
PC9	0.248	6	0.137	6	0.155	0
PC10	0.107	4	0.096	2	0.170	6
PC11	0.169	12	0.193	6	0.333	12
PC12	0.199	0	0.276	6	0.373	6
PC13	0.253	12	0.301	4	0.337	4
PC14	0.288	4	0.219	12	0.132	12
PC15	0.286	1	0.216	1	0.284	12
PC16	0.182	6	0.272	0	0.381	2

ICA

Figure 7.63: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 1st - 3rd independent components (left to right)

Figure 7.64: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 4th - 6th independent components (left to right)
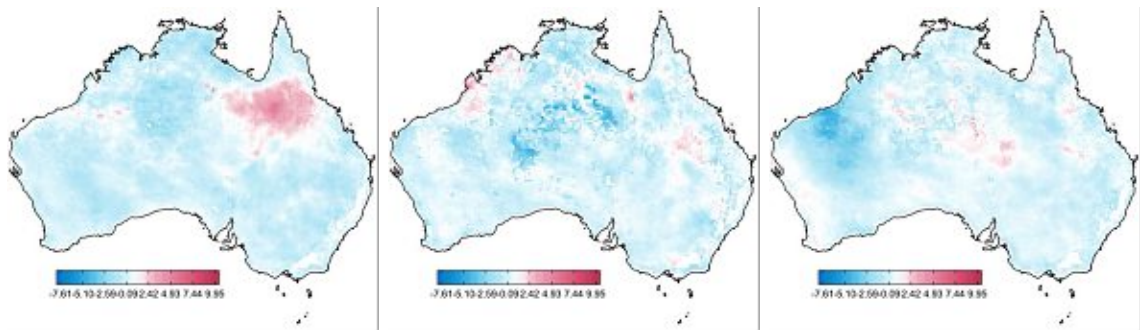

Figure 7.65: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 7th - 9th independent components (left to right)
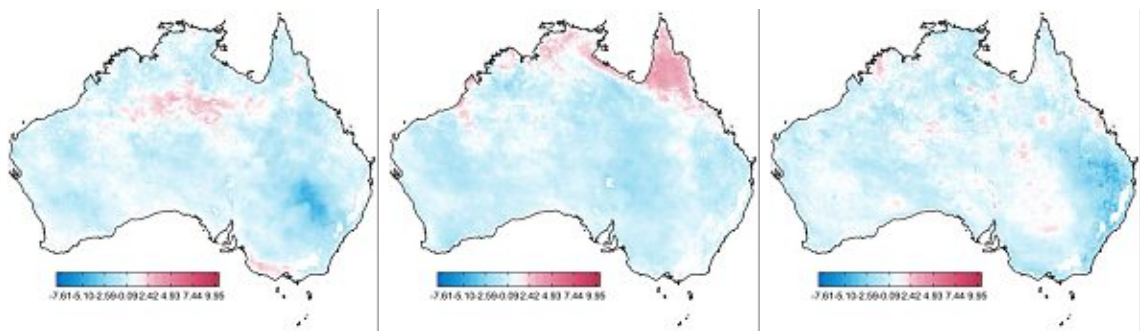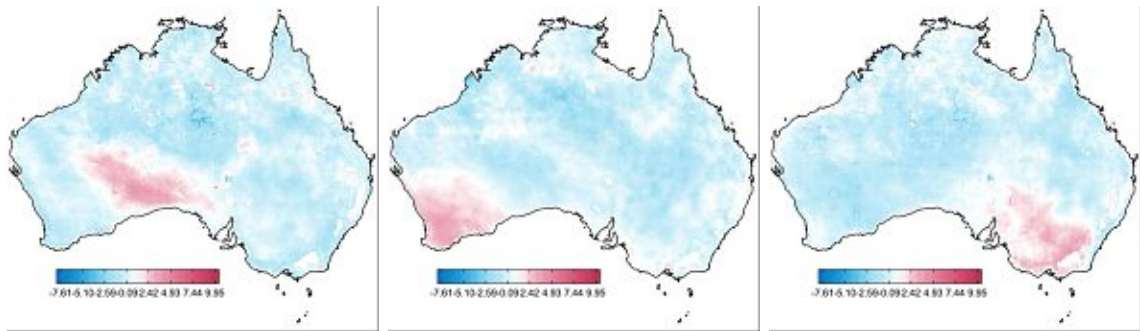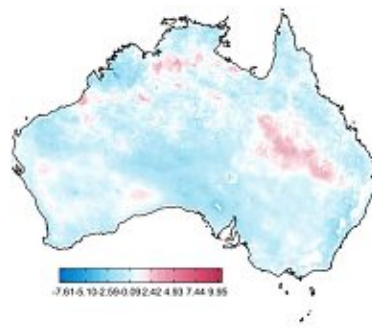

Figure 7.66: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 10th - 12th independent components (left to right)

Figure 7.67: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 13th - 15th independent components (left to right)

Figure 7.68: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 16th independent component

Table 7.11: Correlation analysis of the ICA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Fall) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

|	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.604	6	956	0.579	4	1871	0.449	6	516
IC2	0.488	6	171	0.469	3	329	0.522	4	1468
IC3	0.524	6	1367	0.365	3	1987	0.496	3	63
IC4	0.395	6	1831	0.339	3	1014	0.559	2	1312
IC5	0.572	6	11625	0.473	4	1519	0.409	4	1199
IC6	0.488	6	1325	0.393	4	1522	0.539	3	1083
IC7	0.529	6	229	0.472	4	1320	0.533	12	601
IC8	0.530	6	1230	0.432	4	79	0.539	4	1687
IC9	0.592	6	1818	0.472	4	1525	0.484	3	1787
IC10	0.457	2	103	0.377	6	1317	0.571	6	1176
IC11	0.393	6	1609	0.384	3	1954	0.525	2	743
IC12	0.486	2	556	0.470	3	1281	0.442	12	1368
IC13	0.610	2	727	0.456	3	351	0.483	12	529
IC14	0.471	6	80	0.473	4	1509	0.537	12	916
IC15	0.549	4	1269	0.474	4	1241	0.423	6	1023
IC16	0.490	6	1022	0.373	1	1876	0.537	12	656

SBSS
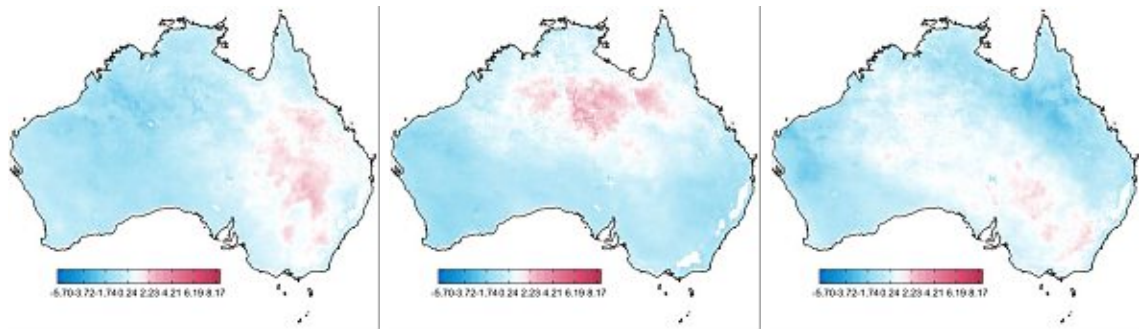
Figure 7.69: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 1st - 3rd SBSS components (left to right)
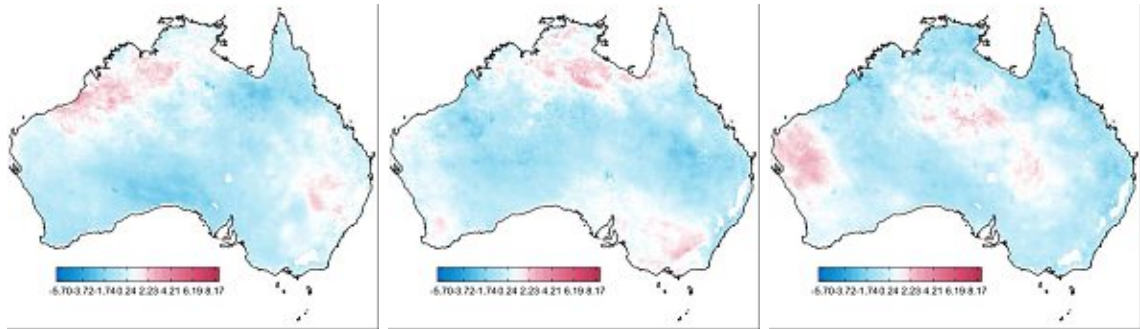
Figure 7.70: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 4th - 6th SBSS components (left to right)
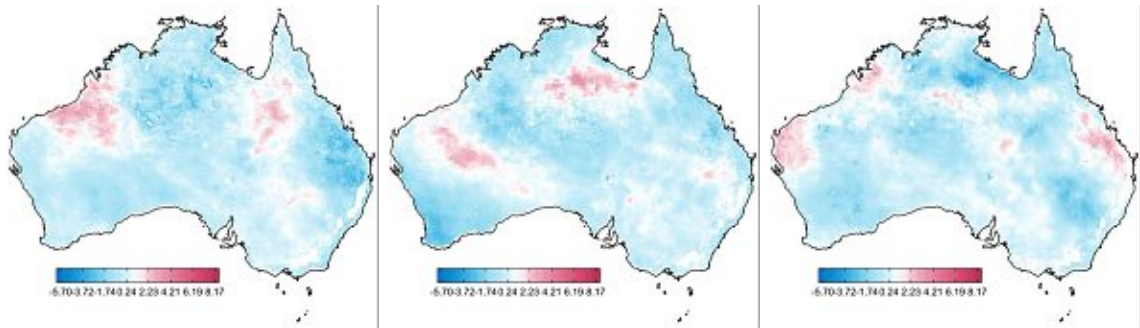

Figure 7.71: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 7th - 9th SBSS components (left to right)
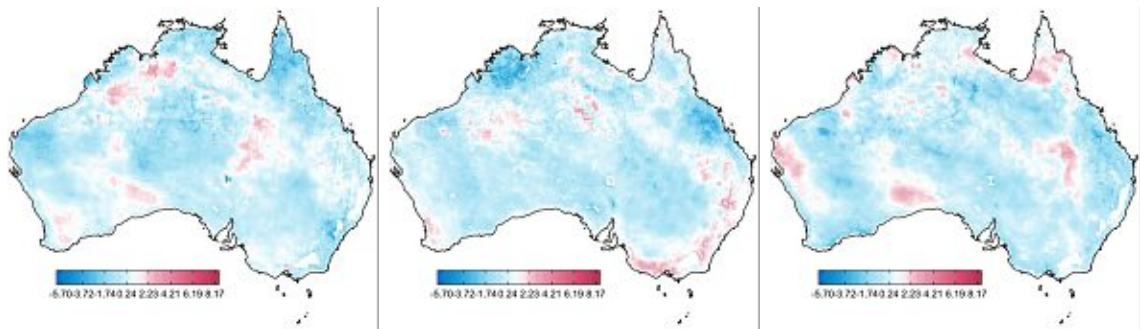

Figure 7.72: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 10th - 12th SBSS components (left to right)
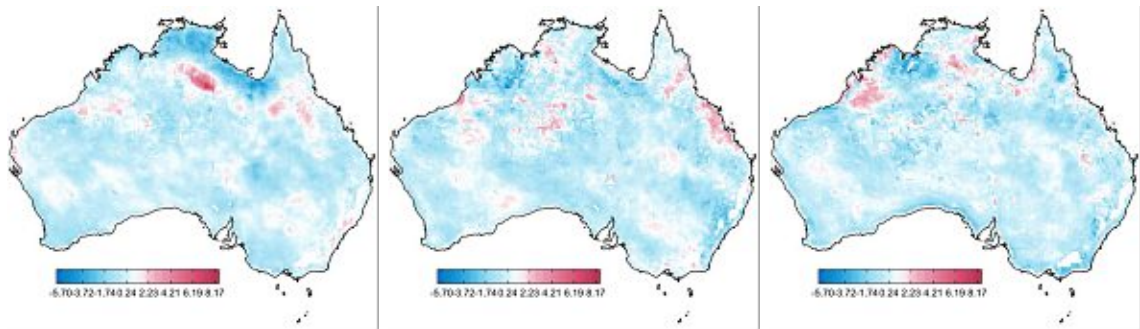
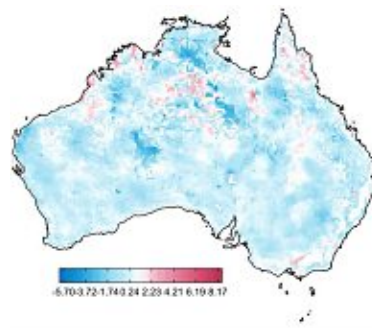Figure 7.73: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 13th - 15th SBSS components (left to right)

Figure 7.74: Seasonal Decomposition of Soil Moisture Anomalies (Fall): 16th SBSS component

Table 7.12: Correlation analysis of the SBSS loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Fall) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.528	2	549	0.353	4	1233	0.415	6	1730
SBSS-C2	0.631	2	1723	0.603	3	990	0.701	1	1792
SBSS-C3	0.442	2	1734	0.413	3	1850	0.554	1	1060
SBSS-C4	0.531	2	1583	0.428	6	769	0.607	6	363
SBSS-C5	0.543	2	134	0.639	4	49	0.408	6	354
SBSS-C6	0.584	2	996	0.463	3	140	0.486	4	1529
SBSS-C7	0.506	6	870	0.493	4	763	0.415	12	825
SBSS-C8	0.417	2	1994	0.432	3	1083	0.395	6	786
SBSS-C9	0.329	4	956	0.281	2	1440	0.343	6	894
SBSS-C10	0.462	2	901	0.394	3	876	0.496	12	957
SBSS-C11	0.468	6	1700	0.412	3	1865	0.424	12	710
SBSS-C12	0.474	6	1641	0.335	6	657	0.455	3	1348
SBSS-C13	0.444	6	1917	0.389	4	220	0.519	3	602
SBSS-C14	0.556	6	127	0.350	2	694	0.422	12	1124
SBSS-C15	0.459	6	1656	0.304	4	969	0.471	3	267
SBSS-C16	0.395	6	658	0.403	4	79	0.489	12	332

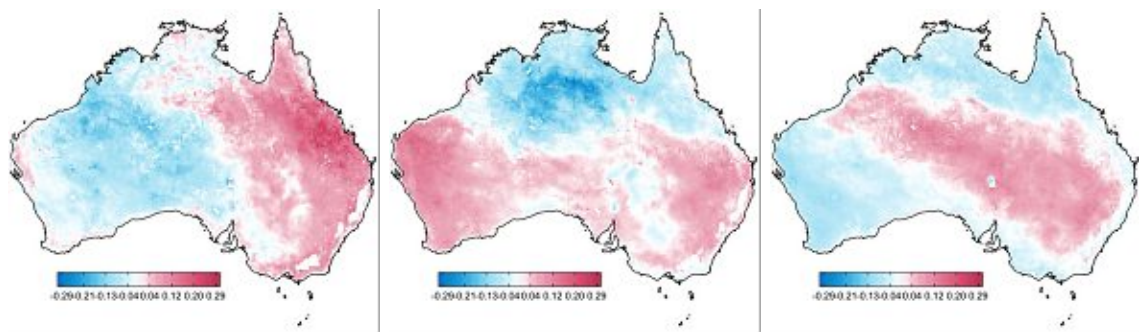## 7.4.5 Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 - Winter PCA

Figure 7.75: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 1st - 3rd principal components (left to right)
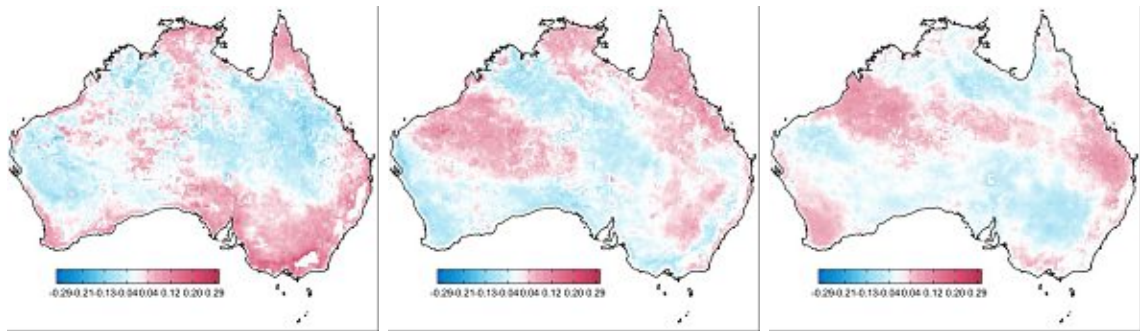
Figure 7.76: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 4th - 6th principal components (left to right)
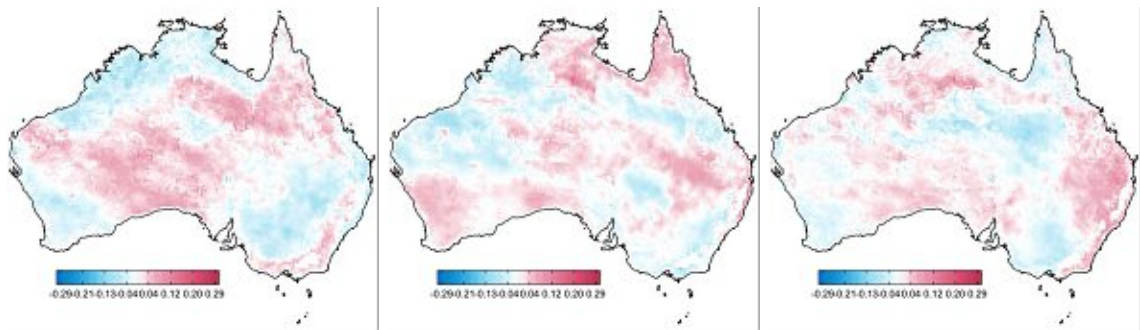

Figure 7.77: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 7th - 9th principal components (left to right)
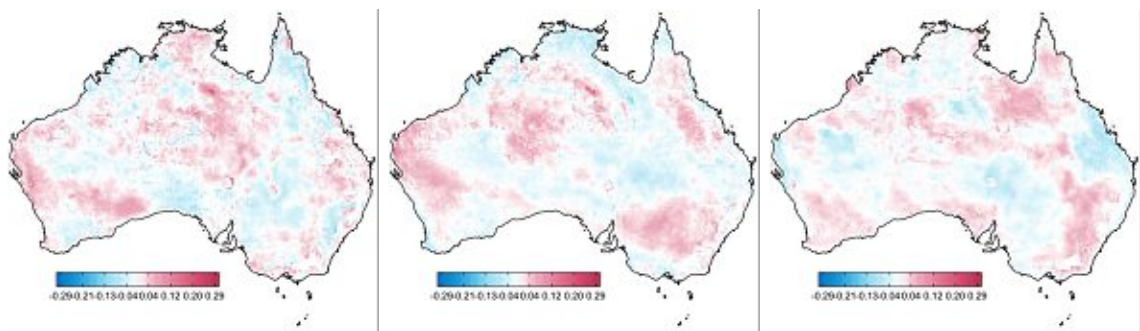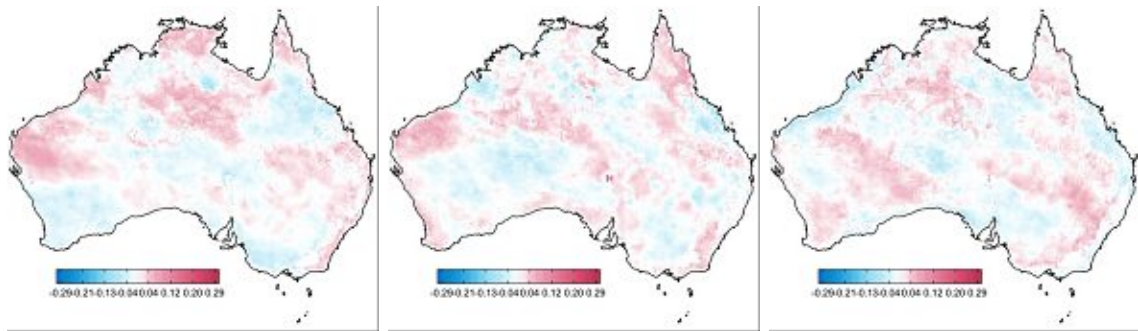

Figure 7.78: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 10th - 12th principal components (left to right)

Figure 7.79: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 13th - 15th principal components (left to right)

Table 7.13: Correlation analysis of the PCA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Winter) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

|	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.146	0	0.209	12	0.189	2
PC2	0.259	3	0.333	12	0.351	4
PC3	0.317	0	0.361	0	0.306	6
PC4	0.158	3	0.243	6	0.306	0
PC5	0.246	6	0.170	3	0.223	12
PC6	0.054	4	0.238	0	0.113	2
PC7	0.152	0	0.166	12	0.295	6
PC8	0.365	12	0.325	6	0.298	1
PC9	0.158	1	0.176	1	0.277	4
PC10	0.471	2	0.258	6	0.312	12
PC11	0.238	0	0.291	0	0.213	4
PC12	0.288	6	0.177	12	0.251	2
PC13	0.198	12	0.272	0	0.218	6
PC14	0.410	4	0.404	12	0.358	1
PC15	0.231	1	0.175	12	0.105	12

ICA
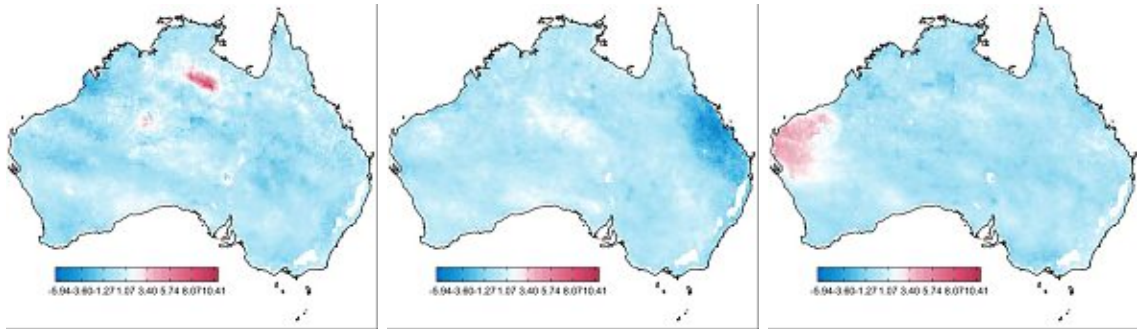
Figure 7.80: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 1st - 3rd independent components (left to right)
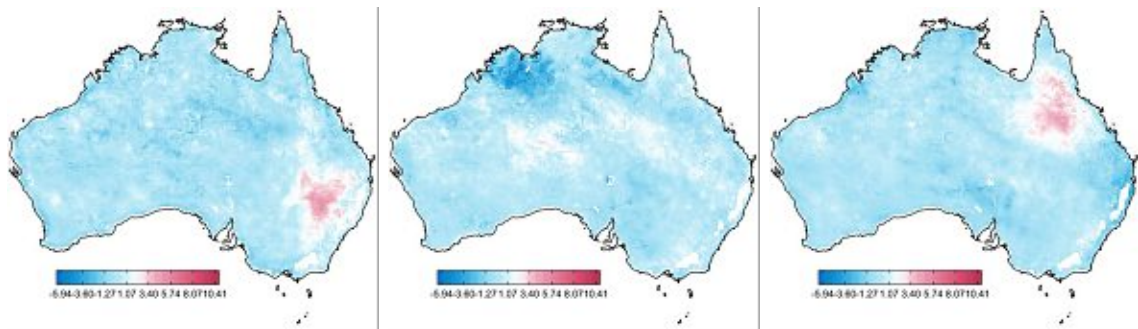

Figure 7.81: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 4th - 6th independent components (left to right)
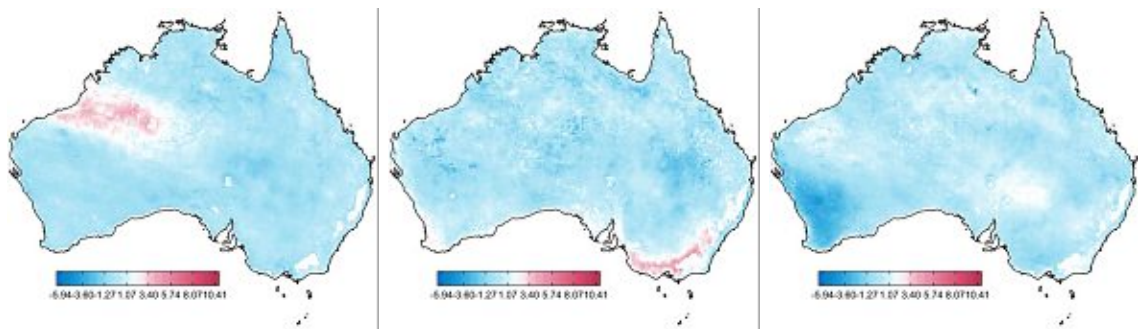

Figure 7.82: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 7th - 9th independent components (left to right)
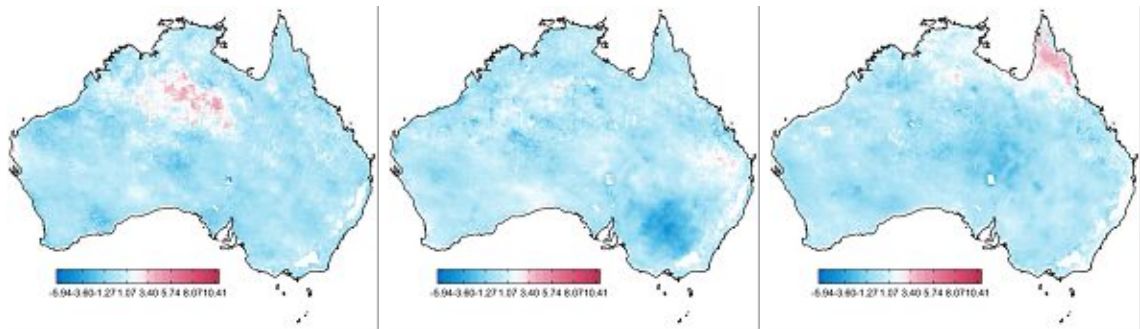
Figure 7.83: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 10th - 12th independent components (left to right)
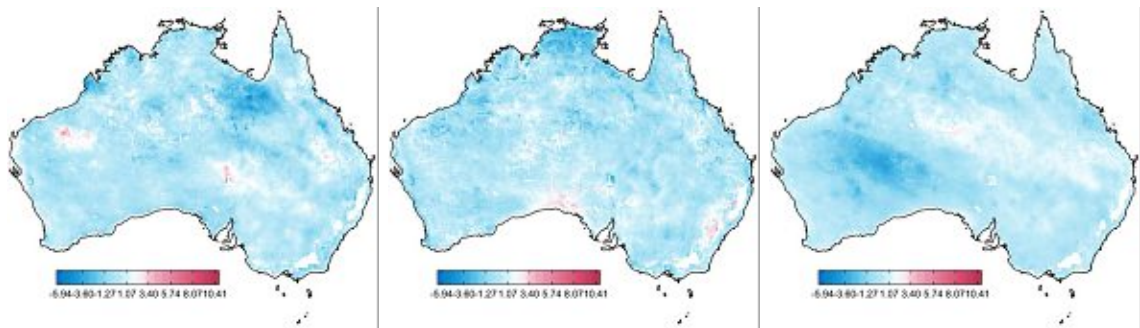

Figure 7.84: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 13th - 15th independent components (left to right)

Table 7.14: Correlation analysis of the ICA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Winter) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

|	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.572	12	80	0.533	0	635	0.535	6	1752
IC2	0.504	6	209	0.503	12	1501	0.406	1	209
IC3	0.522	12	954	0.547	12	1276	0.516	0	841
IC4	0.509	0	981	0.457	12	1525	0.503	12	1077
IC5	0.536	6	1733	0.513	6	409	0.652	6	1411
IC6	0.565	12	409	0.522	12	455	0.517	6	1482
IC7	0.520	6	220	0.492	0	1920	0.454	12	989
IC8	0.585	12	1779	0.509	6	311	0.483	0	869
IC9	0.508	12	420	0.557	0	297	0.609	6	1707
IC10	0.483	0	1422	0.545	12	1215	0.393	12	1631
IC11	0.456	0	103	0.488	12	855	0.463	12	1830
IC12	0.512	12	1853	0.522	12	1198	0.627	12	1271
IC13	0.488	6	828	0.499	12	133	0.425	1	828
IC14	0.486	1	171	0.450	0	1751	0.593	12	419
IC15	0.556	6	519	0.546	12	1957	0.538	12	328

SBSS
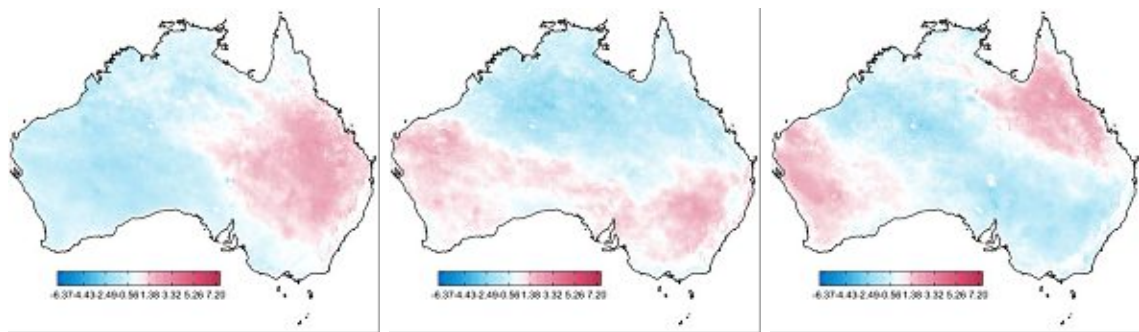
Figure 7.85: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 1st - 3rd SBSS components (left to right)
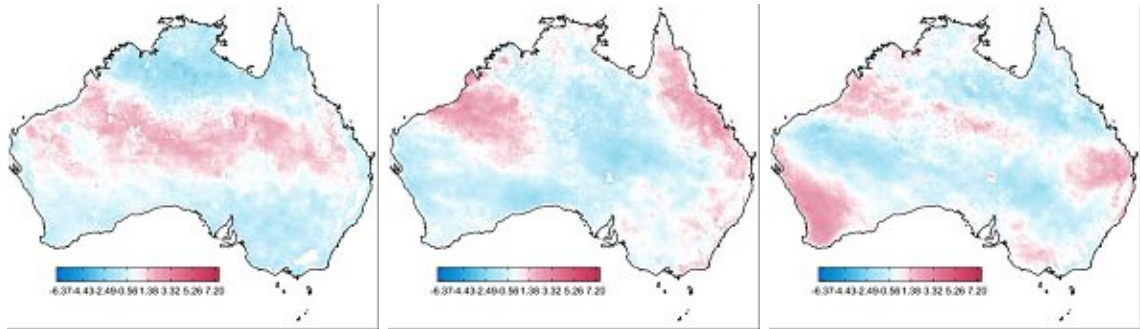
Figure 7.86: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 4th - 6th SBSS components (left to right)
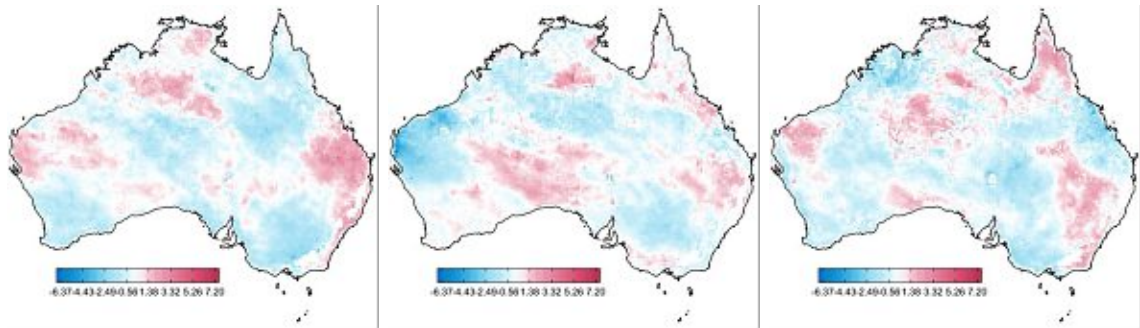

Figure 7.87: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 7th - 9th SBSS components (left to right)
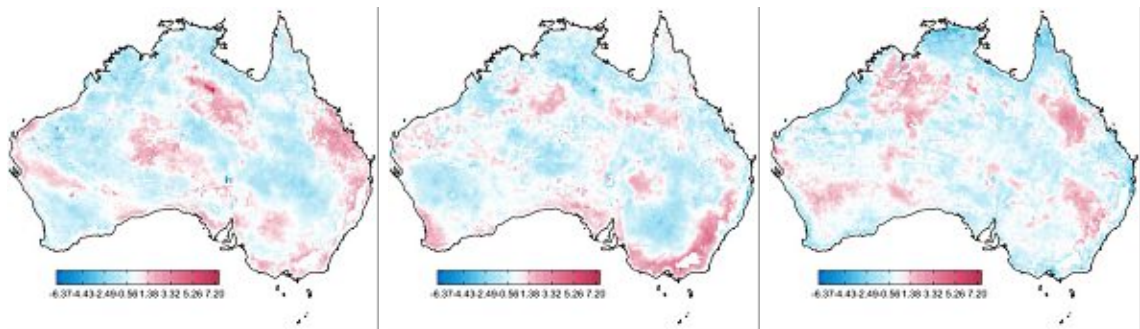

Figure 7.88: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 10th - 12th SBSS components (left to right)
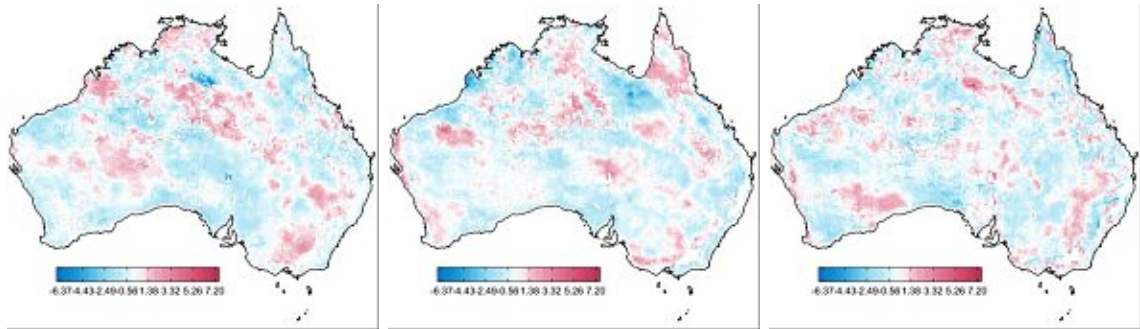
Figure 7.89: Seasonal Decomposition of Soil Moisture Anomalies (Winter): 13th - 15th SBSS components (left to right)

begintable[H] Correlation analysis of the SBSS loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Winter) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.513	0	1881	0.483	6	331	0.538	0	813
SBSS-C2	0.462	3	259	0.616	12	486	0.501	0	472
SBSS-C3	0.495	0	38	0.538	6	1644	0.583	12	1782
SBSS-C4	0.489	0	375	0.567	6	737	0.585	12	1717
SBSS-C5	0.450	6	73	0.463	0	695	0.524	6	1685
SBSS-C6	0.500	6	1279	0.412	12	480	0.512	6	1089
SBSS-C7	0.402	12	589	0.433	0	1132	0.383	6	1099
SBSS-C8	0.493	12	650	0.502	12	1762	0.485	12	78
SBSS-C9	0.547	6	1920	0.458	12	344	0.502	12	1175
SBSS-C10	0.539	0	1486	0.557	12	963	0.531	12	71
SBSS-C11	0.656	6	1705	0.561	12	1698	0.548	12	1906
SBSS-C12	0.468	12	150	0.509	0	673	0.428	0	360
SBSS-C13	0.478	4	922	0.545	12	795	0.469	0	1074
SBSS-C14	0.477	6	2	0.496	12	834	0.454	12	1909
SBSS-C15	0.524	2	1587	0.430	0	1579	0.523	12	485

7.4.6 Seasonal Decomposition of Soil Moisture Anomalies 1998-2018 - Spring PCA
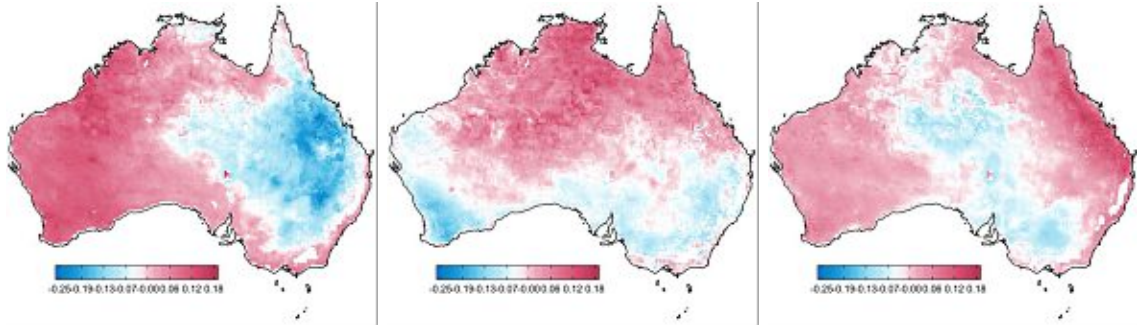

Figure 7.90: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 1st - 3rd principal components (left to right)
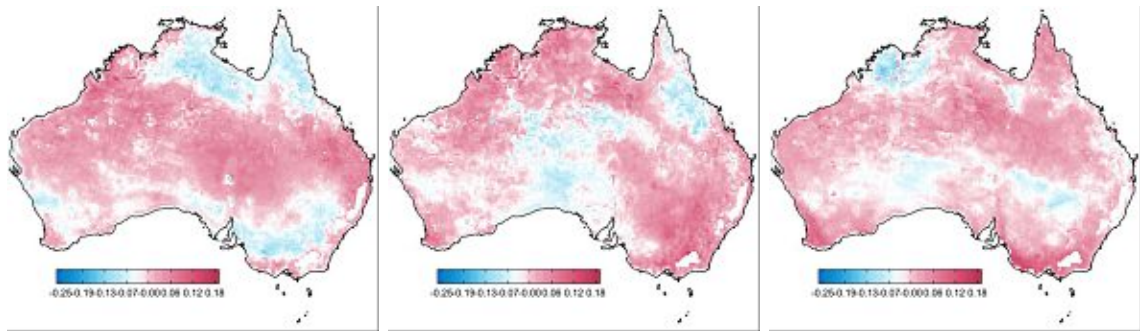

Figure 7.91: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 4th - 6th principal components (left to right)
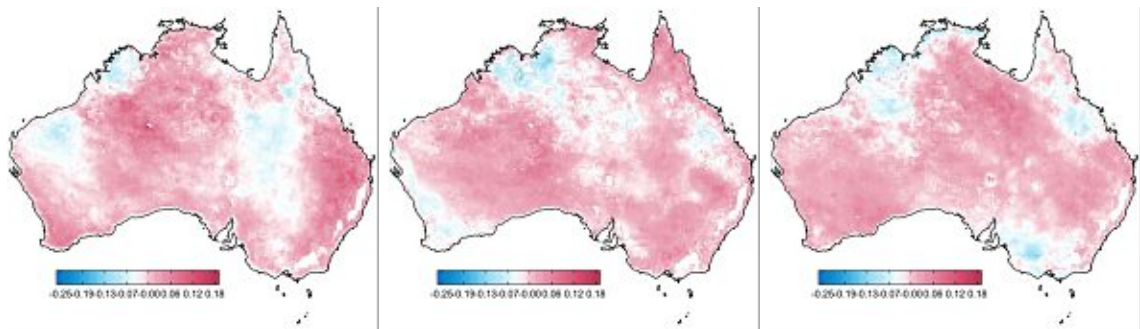

Figure 7.92: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 7th - 9th principal components (left to right)
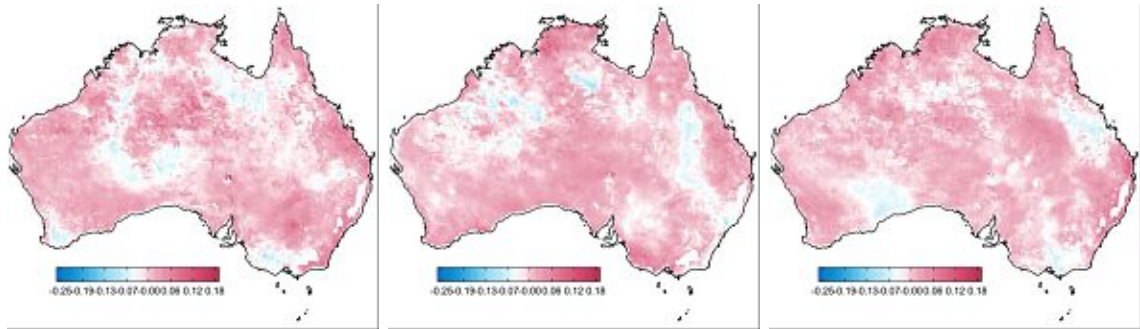
Figure 7.93: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 10th - 12th principal components (left to right)
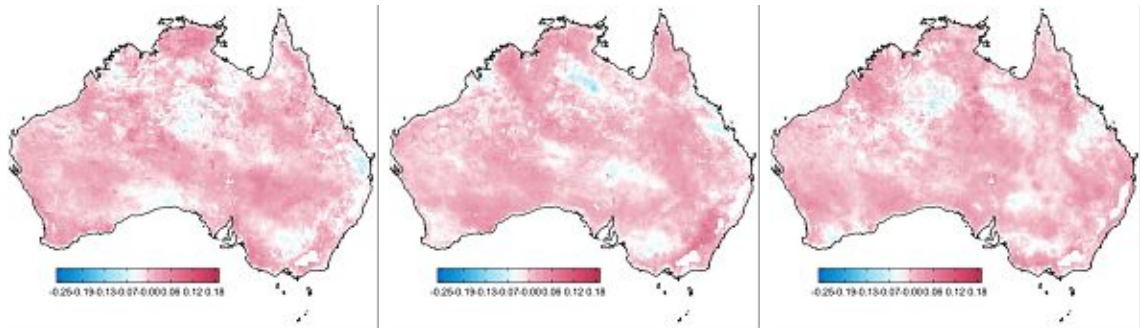

Figure 7.94: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 13th - 15th principal components (left to right)
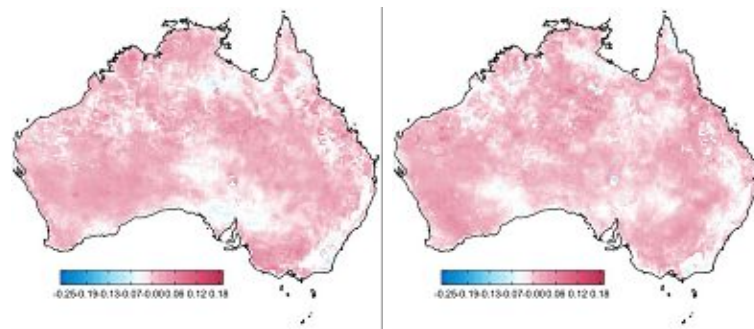

Figure 7.95: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 16th - 17th principal components (left to right)

Table 7.15: Correlation analysis of the PCA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Spring) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag.

|	DMI	Lag	SAMI	Lag	SOI	Lag
PC1	0.495	0	0.296	2	0.587	0
PC2	0.272	12	0.166	2	0.307	6
PC3	0.356	0	0.122	4	0.4270	12
PC4	0.389	1	0.247	3	0.239	1
PC5	0.242	12	0.210	6	0.218	12
PC6	0.243	3	0.356	0	0.298	12
PC7	0.322	6	0.294	12	0.233	0
PC8	0.185	12	0.268	4	0.289	12
PC9	0.339	4	0.394	1	0.192	6
PC10	0.153	0	0.323	2	0.312	1
PC11	0.209	12	0.177	2	0.215	6
PC12	0.315	12	0.269	0	0.263	12
PC13	0.344	2	0.172	12	0.244	4
PC14	0.174	4	0.297	0	0.159	12
PC15	0.211	2	0.205	6	0.341	12
PC16	0.278	4	0.330	4	0.158	1
PC17	0.192	6	0.289	4	0.201	1

ICA
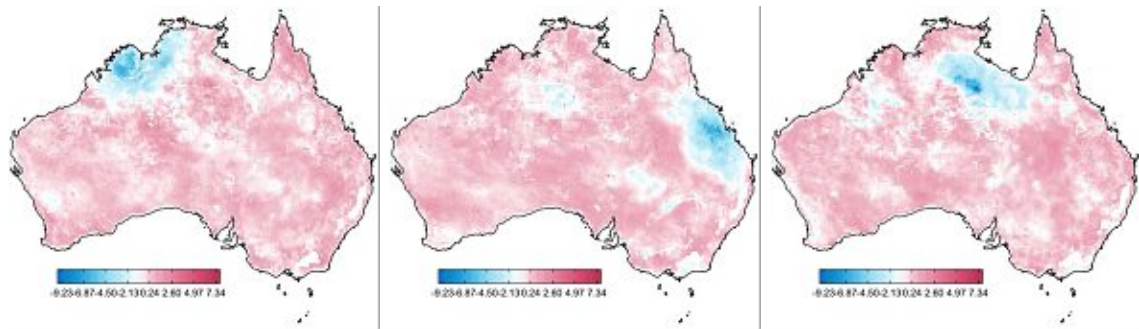
Figure 7.96: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 1st - 3rd independent components (left to right)
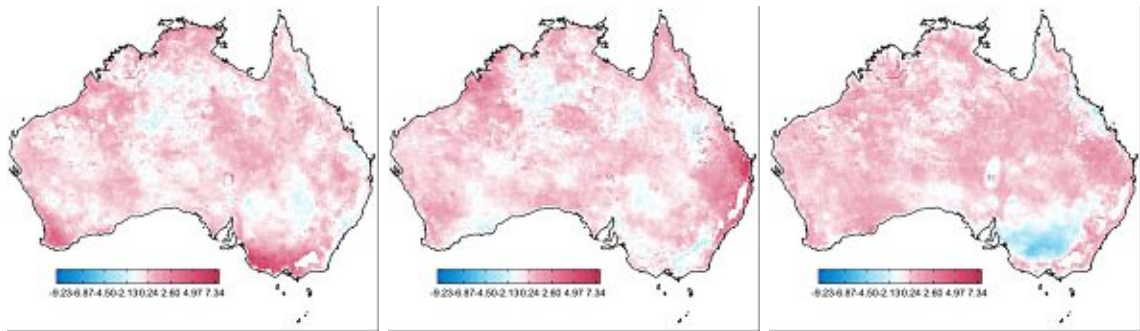
Figure 7.97: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 4th - 6th independent components (left to right)
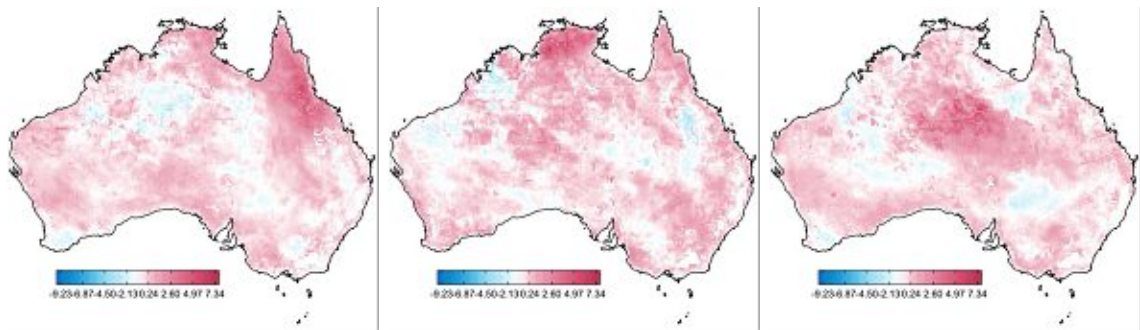

Figure 7.98: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 7th - 9th independent components (left to right)
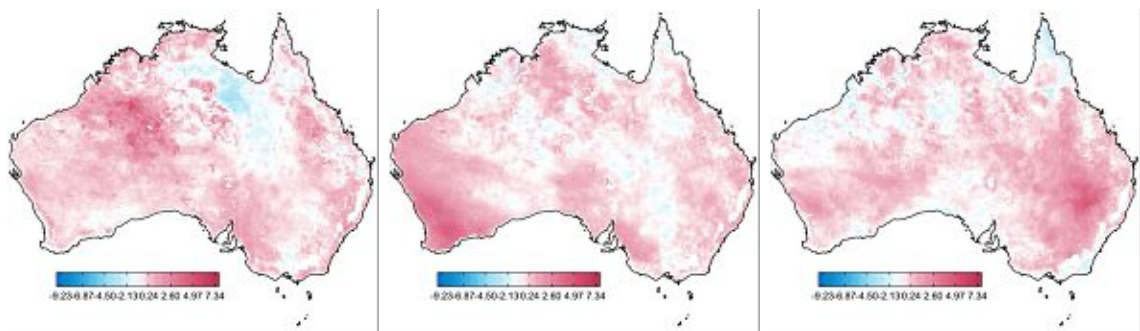

Figure 7.99: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 10th - 12th independent components (left to right)
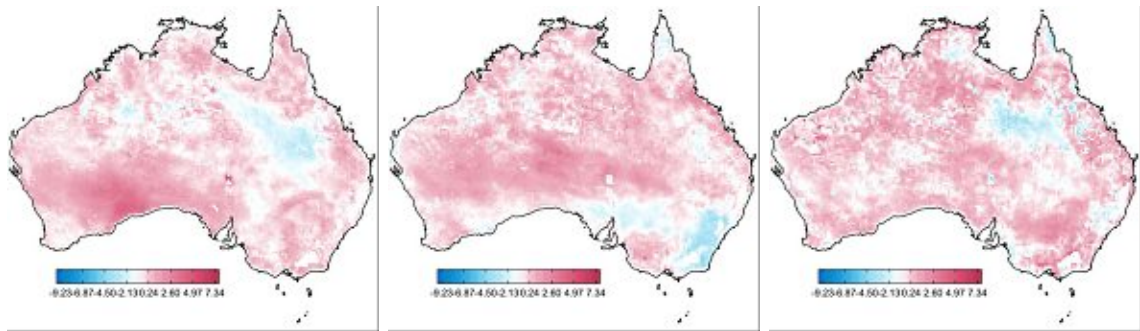
Figure 7.100: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 13th - 15th independent components (left to right)
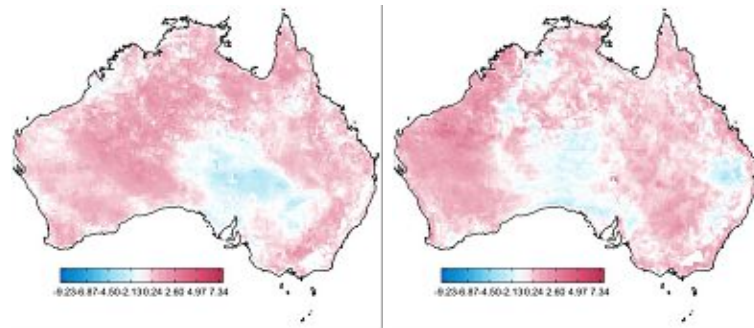

Figure 7.101: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 16th - 17th independent components (left to right)

Table 7.16: Correlation analysis of the ICA loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Spring) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

|	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
IC1	0.464	4	1489	0.579	4	1636	0.636	12	442
IC2	0.584	12	861	0.516	4	492	0.590	12	1405
IC3	0.456	12	1274	0.512	0	469	0.530	12	760
IC4	0.446	3	519	0.538	0	1431	0.549	12	975
IC5	0.513	12	111	0.489	0	1154	0.586	12	129
IC6	0.560	0	927	0.447	4	1882	0.620	12	1891
IC7	0.493	0	171	0.567	4	894	0.594	12	1990
IC8	0.547	12	1309	0.434	0	751	0.503	12	768
IC9	0.474	2	1672	0.563	4	479	0.641	12	1671
IC10	0.540	1	44	0.426	4	954	0.481	12	1482
IC11	0.511	12	1042	0.531	4	1350	0.501	12	1150
IC12	0.463	0	437	0.562	4	428	0.540	1	1627
IC13	0.587	12	176	0.564	4	171	0.582	12	670
IC14	0.450	4	105	0.525	4	377	0.470	12	122
IC15	0.425	12	1234	0.559	3	1868	0.436	1	1144
IC16	0.503	2	681	0.519	4	1515	0.571	12	863
IC17	0.478	0	100	0.561	4	133	0.582	1	1027

SBSS
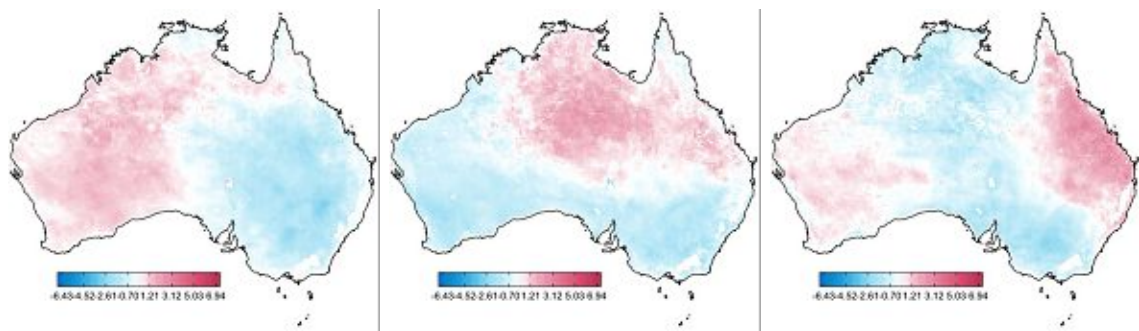
Figure 7.102: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 1st - 3rd SBSS components (left to right)

Figure 7.103: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 4th - 6th SBSS components (left to right)

Figure 7.104: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 7th - 9th SBSS components (left to right)
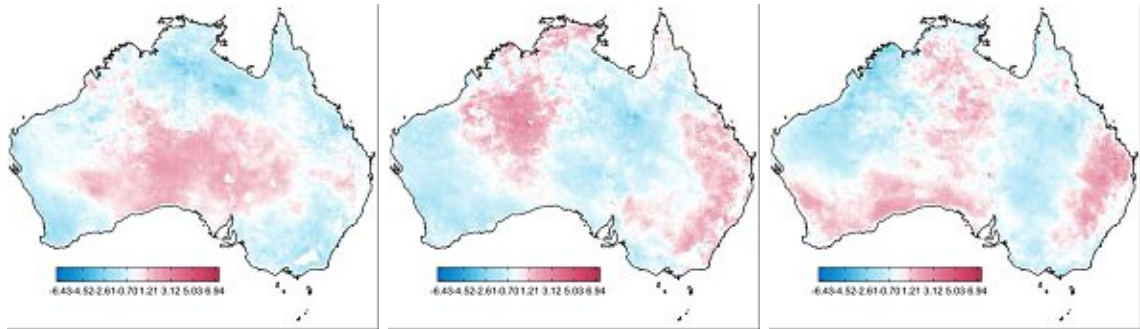

Figure 7.105: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 10th - 12th SBSS components (left to right)

Figure 7.106: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 13th - 15th SBSS components (left to right)

Figure 7.107: Seasonal Decomposition of Soil Moisture Anomalies (Spring): 16th - 17th SBSS components (left to right)
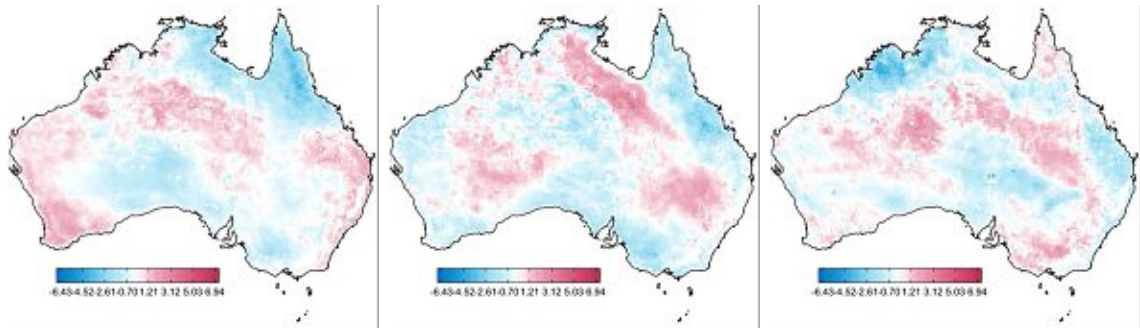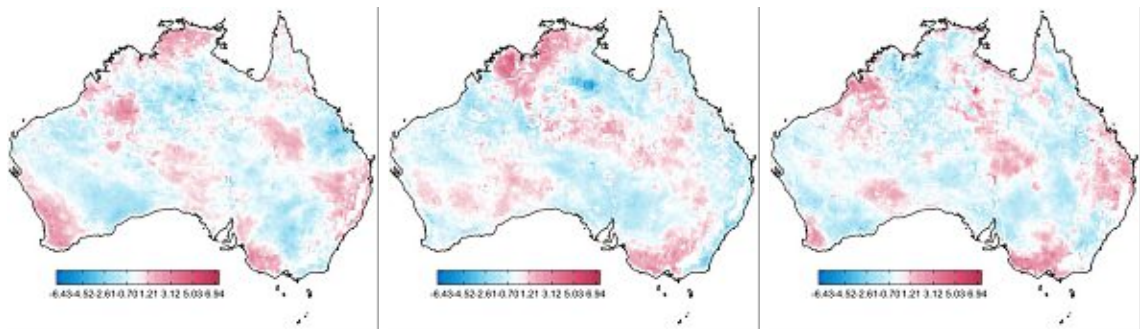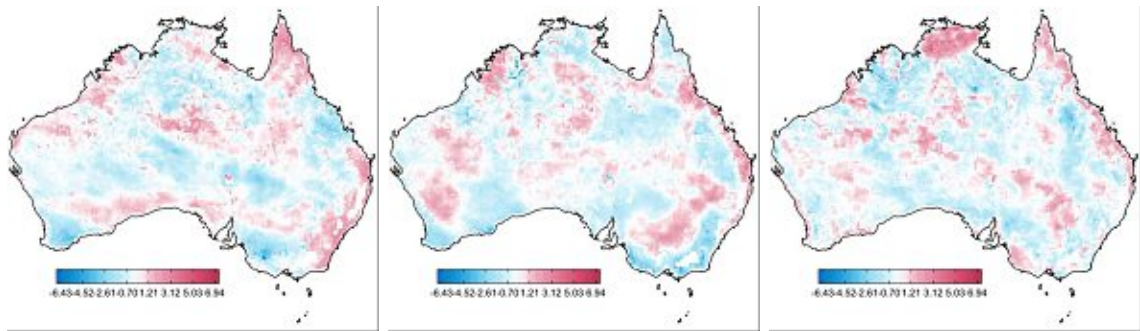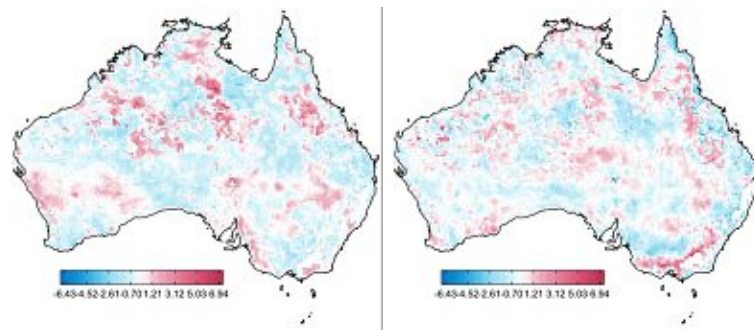
Table 7.17: Correlation analysis of the SBSS loadings of the Seasonal Decomposition of Soil Moisture Anomalies (sdSMA-Spring) 1998-2018 of the CCI dataset and the Climate Oscillation Indices with the corresponding optimal monthly lag and permutation index (see Section 7.3).

	DMI	Lag	Index	SAMI	Lag	Index	SOI	Lag	Index
SBSS-C1	0.658	0	539	0.483	0	958	0.752	1	600
SBSS-C2	0.619	12	28	0.458	0	863	0.792	12	798
SBSS-C3	0.624	0	417	0.394	4	1201	0.484	12	1605
SBSS-C4	0.470	12	1650	0.564	4	1105	0.664	12	1161
SBSS-C5	0.452	4	839	0.516	4	1352	0.598	12	1985
SBSS-C6	0.521	1	1508	0.428	12	1118	0.477	6	383
SBSS-C7	0.530	12	1800	0.546	4	1843	0.591	12	1710
SBSS-C8	0.474	12	239	0.533	4	176	0.553	1	1662
SBSS-C9	0.437	3	1942	0.578	4	1153	0.592	12	1744
SBSS-C10	0.470	12	1523	0.384	4	1030	0.515	12	647
SBSS-C11	0.414	4	501	0.492	0	1890	0.368	1	1687
SBSS-C12	0.433	12	1228	0.509	0	417	0.552	12	845
SBSS-C13	0.478	12	348	0.497	4	475	0.518	12	1611
SBSS-C14	0.412	12	1868	0.467	4	1970	0.415	12	1516
SBSS-C15	0.498	12	1496	0.446	4	1149	0.420	12	1055
SBSS-C16	0.545	2	899	0.431	0	342	0.495	12	1622
SBSS-C17	0.470	12	1874	0.538	4	1390	0.476	12	15

Bibliography

[1] BAS British Antarctic Survey. https://www.bas.ac.uk/. Accessed: 2020-09-24.

[2] BOM Australian Government Bureau of Meteorology. http://www.bom.gov.au/. Accessed: 2020-09-24.

[3] ESA CCI Soil Moisture Data. https://www.esa-soilmoisture-cci.org/. Accessed: 2020-09-24.

[4] NCAR National Center for Atmospheric Research Staff. https://climatedataguide.ucar.edu/. Accessed: 2020-09-24.

[5] NIDIS National Integrated Drought Information System. https://www.drought.gov/drought/. Accessed: 2020-09-24.

[6] NOAA Physical Sciences Laboratory. https://psl.noaa.gov/gcos_wgsp/. Accessed: 2020-09-24.

[7] ZAMG Zentralanstalt für Meteorologie und Geodynamik. https://www.zamg.ac.at/. Accessed: 2020-09-24.

[8] F. Bachoc, M. G. Genton, K. Nordhausen, A. Ruiz-Gazen, and J. Virta. Spatial blind source separation. *Biometrika*, 107(3):627–646, 2020.

[9] B. Bauer-Marschallinger, W. Dorigo, W. Wagner, M. Doubkova, and D. Chung. Spatiotemporal analyses of remotely sensed soil moisture with respect to regional climate modes and solar activity in Australia. *EGUAGA*, 14:11488–11587, 2012.

[10] B. Bauer-Marschallinger, W. A. Dorigo, W. Wagner, and A. Van Dijk. How oceanic oscillation drives soil moisture variations over mainland Australia: An analysis of 32 years of satellite observations. *Journal of Climate*, 26(24):10159–10173, 2013.

[11] R. Bivand, B. Rowlingson, and T. Keitt. *rgdal: Bindings for the "Geospatial" Data Abstraction Library*, 2020. R package version 1.5-16.

[12] R. Bro and A. K. Smilde. Principal component analysis. *Analytical Methods*, 6(9):2812–2831, 2014.

[13] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370. IET, 1993.

[14] F. Chiew, T. Piechota, J. Dracup, and T. McMahon. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, 204(1-4):138–149, 1998.

[15] P. Comon. Independent component analysis. In *Higher Order Statistics*, pages 29–38. Elsevier, 1992. Editor: J.L. Lacoume.

[16] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications.* Academic Press, 2010.

[17] J.D. Gibbons and S. Chakraborti. *Statistical Inference: Revised and Expanded.* CRC Press, 2014.

[18] J. Hérault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *10 Colloque sur le traitement du signal et des images.* GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1985.

[19] M. Ho, A.S. Kiem, and D.C. Verdon-Kidd. The southern annular mode: a comparison of indices. *Hydrology and Earth System Sciences*, 16(3):967–982, 2012.

[20] J. Josse and F. Husson. missMDA: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, 70(1):1–31, 2016.

[21] C. Jutten. *Calcul neuromimétique et traitement du signal: analyse en composantes indépendantes.* PhD thesis, Grenoble INPG, 1987.

[22] Y. Liu, A. van Dijk, R. de Jeu, and T. Holmes. An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resources Research*, 45(7):1–12, 2009.

[23] J. Miettinen, K. Nordhausen, and S. Taskinen. Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. *Journal of Statistical Software*, 76:1–31, 2017.

[24] C. Muehlmann, K. Nordhausen, and J. Virta. *SpatialBSS: Blind Source Separation for Multivariate Spatial Data*, 2020. R package version 0.8.

[25] K. Nordhausen, J.-F. Cardoso, J. Miettinen, H. Oja, E. Ollila, and S. Taskinen. *JADE: Blind Source Separation Methods Based on Joint Diagonalization and Some BSS Performance Criteria*, 2020. R package version 2.0-3.

[26] K. Nordhausen, M. Matilainen, J. Miettinen, J. Virta, and S. Taskinen. Dimension reduction for time series in a BSS context using R. *Journal of Statistical Software*, to appear.

[27] K. Nordhausen and H. Oja. Independent component analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(5):1440–1463, 2018.

[28] K. Nordhausen, H. Oja, P. Filzmoser, and C. Reimann. Blind source separation for spatial compositional data. *Mathematical Geosciences*, 47(7):753–770, 2015.

[29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2017.

[30] N.H. Saji, B.N. Goswami, P.N. Vinayachandran, and T. Yamagata. A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751):360–363, 1999.

[31] J. P. Snyder. *Map Projections - A Working Manual*. US Government Printing Office, 1987.

[32] E. Vanmarcke. *Random Fields*. MIT Press, 1983.

[33] X. Yu, D. Hu, and J. Xu. *Blind Source Separation: Theory and Applications*. John Wiley & Sons, 2013.

[34] Y. Zhao, J. Tang, C. Graham, Q. Zhu, K. Takagi, and H. Lin. Hydropedology in the ridge and valley: Soil moisture patterns and preferential flow dynamics in two contrasting landscapes. In *Hydropedology: Synergistic Integration of Soil Science and Hydrology; Lin, H., Ed*, pages 381–411. Academic Press, 2012.