UNIVERSITETET I BERGEN

Inferring Vegetation Memory from Remote Sensing Data using novel Climate Reconstruction Products

Master's Thesis (M.Sc.)

30/05/2019

at Universitetet i Bergen

In Partial Fulfilment of the Requirements for the Degree Master of Science (M.Sc.)

Primary Supervisor: Dr. Alistair Seddon
Secondary Supervisor: Dr. Richard Davy

by
Erik Kusch
born 03/01/1995 in Eilenburg, Germany
Enrolment number: 239977

Abstract

The dependence of humans on stable, intact ecosystems has lead to an increasing appreciation of the need for resilience goals to be incorporated into environmental policy catalogues [1,2]. Remote sensing has played a key role in mapping vegetation sensitivity, resistance and resilience at a global scale. Within these studies, vegetation memory has been used as an important proxy for ecosystem recovery rates [3], potentially a key component of vegetation resilience [4]. In particular, strong vegetation memory effects have been identified in dryland regions [3,5,6] coinciding with decreased vegetation sensitivity towards climatological drivers [7].

However, vegetation memory has been identified to be more than straightforward assessments of recovery rates. A recent approach by Ogle et al. distinguishes intrinsic and extrinsic vegetation memory components [8]. Here, I aim to test the components and drivers of vegetation memory in dryland regions using state-of-the-art climate reanalysis data and refined approaches to identifying vegetation memory characteristics. This has lead to novel insights into spatial patterns of vegetation memory characteristics across three distinct predominantly dryland regions (Iberian Region, Caatinga, and Australia). My thesis shows that (1) dryland regions are characterised by strong intrinsic vegetation memory as well as strong extrinsic vegetation memory, (2) it is possible to distinguish intrinsic and extrinsic vegetation memory to a hitherto unachieved degree using climate reanalysis data sets, and (3) dryland vegetation does not react to bioclimatic forcing in the same way across the Earth.

In addition, this thesis aims to link the spatial patterns of vegetation memory with functional ecology datasets (i.e. TRY and COMPADRE). This is done to establish connections between plant life history strategies and vegetation memory characteristics thereby identifying causal pathways enabling vegetation memory. I based my hypotheses on the concept of information and material legacies by Johnstone et al. [9]. Regarding information legacies, I expected any life history strategy allowing plants to regenerate quickly to bioclimatic events to result in shorter recovery rates and lower vegetation memory. Material legacies, such as the ones identified via plant functional traits of leaf nitrogen content or vegetative height, I postulated to enable both resistance as well as recovery from perturbation events. My analyses did not find any support for the latter statement but did identify biological first time-step response of populations towards perturbations as well as plant reproductive strategy to be related to vegetation memory capabilities.

Thus, my study presents a synthesis of resilience thinking, vegetation memory modelling approaches, climate science, and functional ecology. My findings demonstrate novel observations of vegetation memory patterns across dryland regions such as regional differences of processes forming vegetation memory capabilities. Consequently, this study provides a helpful stepping stone for refining and combining already existing methodology which could, in turn, generate important knowledge of ecosystem functioning and resilience.

Acknowledgements

First and foremost I would like to express my utmost gratitude to my two supervisors: Dr. Alistair Seddon, and Dr. Richard Davy. Thank you for your valuable input and many lessons on research practices as well as your keeping up with my frequent emailing of revised manuscripts. I'd also like to thank Roberto Salguero-Gómez for supplying me with the data for the COMPADRE data base.

I never could've done the work presented in the following without the support of friends and family whom I owe massive thanks to. My family, for your unrelenting support in the form of goodie parcels shipped from Germany. Peter, for the many hikes and drinks we shared; our heart-to-heart conversations never fail to make me feel understood and supported. Rachel, for being a great friend, hosting me in your guest bed room and helping me bake awesome cake. Jenny, for calming the seas of life telling me to navigate them myself as best I can but never abandoning ship; it's great to have you on deck. Deborah, for genuinely outstanding advice on life, the universe, and everything; also for introducing me to two prime specimen of *Felis Devilis*.

From the bottom of my heart - Thank all of you! I'd be delighted to have you around for the next big project.

Contents

1 Introduction **1**

1.1 Vegetation Memory . . . 3

1.1.1 Terminology . . . 3

1.1.2 Ecological Relevance . . . 4

1.1.3 Quantifying Vegetation Memory . . . 4

1.1.3.1 Traditional Approaches . . . 4

1.1.3.2 Remote Sensing Approaches . . . 4

1.1.3.3 Synthesis . . . 7

1.2 Functional Aspects of Vegetation Memory . . . 9

1.2.1 Terminology . . . 9

1.2.1.1 Life History Traits . . . 9

1.2.1.2 Plant Functional Traits . . . 10

1.2.2 Ecological Relevance . . . 11

1.2.3 Unifying Information on Plant Functional Traits . . . 12

1.2.4 Putting Plant Function on the Map . . . 13

1.3 Thesis Outline . . . 14

2 Material & Methods **15**

2.1 Study Regions . . . 15

2.1.1 Iberian Region . . . 15

2.1.2 Caatinga, Brazil . . . 16

2.1.3 Australia . . . 17

2.2 Identifying Vegetation Memory . . . 18

2.2.1 Data . . . 18

2.2.1.1 Vegetation Indices . . . 18

2.2.1.2 European Centre for Medium-range Weather Forecasts ReAnalysis 5 (ERA5) . . . 20

2.2.1.3 Digital Elevation Models . . . 22

2.2.2 Analyses . . . 23

2.2.2.1 Data Preparation . . . 23

2.2.2.2 Memory Models . . . 25

2.2.2.3 Vegetation Memory Sensitivity . . . 30

2.3 Functional Aspects of Vegetation Memory . . . 31

2.3.1 Data . . . 31

2.3.1.1 COMPADRE . . . 31

2.3.1.2 TRY . . . 31

2.3.1.3 Floral Data . . . 32

2.4 Analyses . . . 33
2.4.1 Life History Traits . . . 33
2.4.2 Plant Functional Traits . . . 33

3 Results **34**
3.1 Identifying Vegetation Memory . . . 34
3.1.1 Vegetation-Memory Models . . . 34
3.1.1.1 Iberian Region . . . 34
3.1.1.2 Caatinga . . . 36
3.1.1.3 Australia . . . 38
3.1.2 Variance Partitioning . . . 40
3.1.3 Vegetation Memory Sensitivity . . . 41
3.1.3.1 $NDVI_{[t-1]}$ . . . 41
3.1.3.2 Air Temperature . . . 42
3.1.3.3 Soil Moisture (0-7cm) . . . 43
3.1.3.4 Memory Length . . . 44
3.1.3.5 $NDVI_{[t-1]}$ and Memory Length . . . 45
3.1.3.6 Soil Moisture and Memory Length . . . 46
3.1.4 Summary . . . 47
3.1.4.1 Vegetation Memory Models . . . 47
3.1.4.2 Vegetation Memory Sensitivity . . . 47
3.1.4.3 Regional Differences . . . 47
3.2 Functional Aspects of Vegetation Memory . . . 48
3.2.1 Life History Traits . . . 48
3.2.2 Plant Functional Traits . . . 51
3.2.3 Summary . . . 54
3.2.3.1 Life History Traits . . . 54
3.2.3.2 Plant Functional Traits . . . 54

4 Discussion **55**
4.1 Vegetation Memory . . . 55
4.1.1 Spatial Patterns of Vegetation Memory . . . 55
4.1.1.1 Intrinsic Vegetation Memory . . . 55
4.1.1.2 Extrinsic Vegetation Memory . . . 55
4.1.1.3 Regional Differences . . . 56
4.1.2 Intrinsic and Extrinsic Vegetation Memory . . . 56
4.1.3 Climate Reanalysis Data in Vegetation Memory Analyses . . . 57
4.1.4 Vegetation Memory Sensitivity . . . 57
4.1.4.1 Intrinsic Vegetation Memory . . . 57
4.1.4.2 Extrinsic Vegetation Memory . . . 57
4.2 Functional Aspects of Vegetation Memory . . . 58
4.2.1 Life History Traits . . . 58
4.2.2 Plant Functional Traits . . . 58

5 Conclusion **59**
5.1 Vegetation Memory . . . 59
5.2 Functional Aspects of Vegetation Memory . . . 60

Bibliography **60**

A Appendix **69**
A.1 Project Requirements . . . 69
A.1.1 R Requirements . . . 69
A.1.2 ERA5 Data . . . 72
A.2 Data . . . 74
A.2.1 HWSD Data . . . 74
A.2.2 Study Regions . . . 76
A.2.2.1 Iberian Region . . . 76
A.2.2.2 Caatinga, Brazil . . . 78
A.2.2.3 Australia . . . 80
A.3 R Analyses Codes . . . 82
A.3.1 Master File . . . 82
A.3.2 Vegetation Memory . . . 87
A.3.2.1 GIMMs Data . . . 87
A.3.2.2 ERA5 . . . 88
A.3.2.3 Calculation of Vegetation Memory . . . 91
A.3.3 Plant Functional Traits . . . 97
A.3.4 COMPADRE . . . 102
A.4 Results . . . 103
A.4.1 Vegetation Memory Models . . . 103
A.4.1.1 Iberian Region . . . 103
A.4.1.2 Caatinga . . . 107
A.4.1.3 Australia . . . 111
A.4.2 Variance Partitioning . . . 115
A.4.2.1 Iberian Region . . . 115
A.4.2.2 Caatinga . . . 117
A.4.2.3 Australia . . . 119
A.5 R Data Visualisation Codes . . . 121
A.5.1 Preamble . . . 121
A.5.2 Data Overview . . . 121
A.5.2.1 NDVI . . . 121
A.5.2.2 ERA5 . . . 121
A.5.2.3 HWSD . . . 122
A.5.2.4 Study Regions . . . 123
A.5.3 Vegetation Memory Models . . . 125
A.5.4 Results . . . 127
A.5.4.1 Memory Models . . . 127
A.5.4.2 Variance Partitioning . . . 129
A.5.4.3 Soil Layer Comparison . . . 130
A.5.4.4 Vegetation Memory Sensitivity . . . 131
A.5.5 Plant Function . . . 133
A.5.5.1 Life History Traits . . . 133
A.5.5.2 Plant Functional Traits . . . 136
A.5.6 Miscellaneous Figures . . . 139
A.6 Declaration Of Authorship . . . 142

List of Figures

1.1 Drought Memory Length and Strength 5
1.2 Water Memory Length 6
1.3 Intrinsic Vegetation Memory 6
1.4 Components of Vegetation Memory 7
1.5 Concept of Combining Plant Functional Trait Information with Species-Distribution Maps 13
1.6 Study Outline 14

2.1 European Ombrotpyes 15
2.2 South-American Ombrotpyes 16
2.3 Australasian Ombrotpyes 17
2.4 Global Representation of the Normalised Difference Vegetation Index (NDVI) 19
2.5 Overview of air temperature data 20
2.6 Overview of soil moisture data 21
2.7 DEM elevation data 22
2.8 Kriging Concept 24
2.9 Vegetation Memory Model Flowchart 27
2.10 Memory Model NDVI Data Treatment 27
2.11 Memory Model Qsoil Data Treatment 28
2.12 exemplary Vegetation Memory PCA 28
2.13 PCA Regression Coefficients 29
2.14 Variance Partitioning 30
2.15 GBIF *Plantae* Occurrence Overview 32

3.1 Vegetation Response Coefficients (Iberian Region; Qsoil1) 35
3.2 Vegetation Response Coefficients (Caatinga; Qsoil1) 37
3.3 Vegetation Response Coefficients (Australia; Qsoil1) 39
3.4 Variance Partitioning Across Study Regions 40
3.5 Intrinsic Vegetation Memory Sensitivity 41
3.6 Tair Vegetation Memory Sensitivity 42
3.7 Qsoil1 Vegetation Memory Sensitivity 43
3.8 Qsoil1 Vegetation Memory Length Sensitivity 44
3.9 Intrinsic Vegetation Memory and Qsoil1 Memory Length 45
3.10 Soil Moisture Vegetation Memory and Qsoil1 Memory Length 46
3.11 Short 50
3.12 PFTs and Intrinsic Memory (Iberian Region) 51
3.13 PFTs and Tair Memory (Iberian Region) 52
3.14 PFTs and Qsoil1 Memory (Iberian Region) 53

A.1 Working Directories . . . 70
A.2 HWSD slope aspect data . . . 74
A.3 HWSD slope incline data . . . 75
A.4 Data Overview (Iberian Region) . . . 76
A.5 TRY/COMPADRE Data (Iberian Region) . . . 77
A.6 Data Overview (Caatinga, Brazil) . . . 78
A.7 TRY/COMPADRE Data (Caatinga, Brazil . . . 79
A.8 Data Overview (Australia) . . . 80
A.9 TRY/COMPADRE Data (Australia) . . . 81
A.10 Vegetation Response Coefficients (Iberian Region; Qsoil2) . . . 103
A.11 Vegetation Response Coefficients (Iberian Region; Qsoil3) . . . 104
A.12 Vegetation Response Coefficients (Iberian Region; Qsoil4) . . . 105
A.13 Vegetation Response Coefficients (Iberian Region, QSoil Layers) . . . 106
A.14 Vegetation Response Coefficients (Caatinga; Qsoil2) . . . 107
A.15 Vegetation Response Coefficients (Caatinga; Qsoil3) . . . 108
A.16 Vegetation Response Coefficients (Caatinga; Qsoil4) . . . 109
A.17 Vegetation Response Coefficients (Caatinga, QSoil Layers) . . . 110
A.18 Vegetation Response Coefficients (Australia; Qsoil2) . . . 111
A.19 Vegetation Response Coefficients (Australia; Qsoil3) . . . 112
A.20 Vegetation Response Coefficients (Australia; Qsoil4) . . . 113
A.21 Vegetation Response Coefficients (Australia, QSoil Layers) . . . 114
A.22 Variance Partitioning (Iberian Region; Qsoil1) . . . 115
A.23 Variance Partitioning (Iberian Region; Qsoil1) . . . 116
A.24 Variance Partitioning (Caatinga; Qsoil1) . . . 117
A.25 Variance Partitioning (Caatinga; Qsoil1) . . . 118
A.26 Variance Partitioning (Australia; Qsoil1) . . . 119
A.27 Variance Partitioning (Australia; Qsoil1) . . . 120

List of Tables

1.1 Basic Vocabulary of the Ecological Stability Debate . . . 2
1.2 Basic vocabulary of the vegetation memory framework . . . 3
1.3 Basic Vocabulary of Life History Trait Frameworks . . . 10
1.4 Basic Vocabulary of Plant Functional Trait Frameworks . . . 11

2.1 Core Information about NDVI data . . . 18
2.2 Core Information about ERA5 data . . . 20
2.3 Core Information about HWSD data . . . 22
2.4 Interpretation of Memory Model Coefficients . . . 25

3.1 Mann-Whitney U-Test (Iberian Region, Qsoil1 Model) . . . 34
3.2 Mann-Whitney U-Test (Caatinga, Qsoil1 Model) . . . 36
3.3 Mann-Whitney U-Test (Australia, Qsoil1 Model) . . . 38
3.4 Vegetation Memory Sensitivity (Mean NDVI and $NDVI_{[t-1]}$ Coefficients) . . . 41
3.5 Vegetation Memory Sensitivity (Mean Tair and Tair Coefficients) . . . 42
3.6 Vegetation Memory Sensitivity (Mean Qsoil1 and Qsoil1 Coefficients) . . . 43
3.7 Vegetation Memory Sensitivity (Mean Qsoil1 and Qsoil1 Memory Lags) . . . 44
3.8 Vegetation Memory Sensitivity (Qsoil1 Memory Lags and $NDVI_{[t-1]}$ Coefficients) . . . 45
3.9 Vegetation Memory Sensitivity (Qsoil1 Memory Lags and Qsoil1 Coefficients) . . . 46
3.10 COMPADRE FSC-1 and Vegetation Memory (Iberian Region) . . . 48
3.11 COMPADRE FSC-2 and Vegetation Memory (Iberian Region) . . . 48
3.12 COMPADRE Pi and Vegetation Memory (Iberian Region) . . . 49
3.13 COMPADRE Rho and Vegetation Memory (Iberian Region) . . . 49
3.14 COMPADRE Reactivity and Vegetation Memory (Iberian Region) . . . 49
3.15 PFTs and $NDVI_{[t-1]}$ Memory (Iberian Region) . . . 51
3.16 PFTs and Tair Memory (Iberian Region) . . . 52
3.17 PFTs and Qsoil1 Memory (Iberian Region) . . . 53

A.1 R Packages . . . 69
A.2 HWSD Variables and Explanations . . . 74
A.3 Mann-Whitney U-Test (Iberian Region, Qsoil2 Model) . . . 103
A.4 Mann-Whitney U-Test (Iberian Region, Qsoil3 Model) . . . 104
A.5 Mann-Whitney U-Test (Iberian Region, Qsoil4 Model) . . . 105
A.6 Mann-Whitney U-Test (Iberian Region, Qsoil Layers) . . . 106
A.7 Mann-Whitney U-Test (Caatinga, Qsoil2 Model) . . . 107
A.8 Mann-Whitney U-Test (Caatinga, Qsoil3 Model) . . . 108
A.9 Mann-Whitney U-Test (Caatinga, Qsoil4 Model) . . . 109

A.10 Mann-Whitney U-Test (Caatinga, Qsoil Layers) 110
A.11 Mann-Whitney U-Test (Australia, Qsoil2 Model) 111
A.12 Mann-Whitney U-Test (Australia, Qsoil3 Model) 112
A.13 Mann-Whitney U-Test (Australia, Qsoil4 Model) 113
A.14 Mann-Whitney U-Test (Australia, Qsoil Layers) 114

List of Abbreviations and Acronyms

H Vegetative Height

$NDVI_{[t-1]}$ Autoregressive NDVI Coefficient

N_{mass} Leaf Nitrogen Content per Leaf Dry Mass

π Period of Oscillation

ρ Damping Ratio

AIC Akaike Information Criterion

AVHRR Advanced Very High Resolution Radiometer

BIEN Botanical Information and Ecology Network

CHELSA Climatologies at High Resolution for the Earth's Land Surface Areas

COMPADRE COMPADRE Plant Matrix Data Base

CRU Climatic Research Unit

DEM Digital Elevation Model

DSSs Decision Support Systems

ECMW European Centre for Medium-range Weather Forecasts

ERA5 European Centre for Medium-range Weather Forecasts ReAnalysis 5

EVI Enhanced Vegetation Index

FDD Functional Diversity Dispersion

FDR Functional Diversity Richness

FSC Fast-Slow Continuum

FSC-1 Life History Speed within the Fast-Slow Continuum

FSC-2 Reproductive Strategy within the Fast-Slow Continuum

GBIF Global Biodiversity Information Facility

GIMMS Global Inventory Modelling and Mapping Studies

HWSD Harmonized World Soil Database

iDiv Deutsches Zentrum für integrative Biodiversitätsforschung

LCCSs Land-Cover Classification Systems

LHT Life History Trait

MODIS Moderate-Resolution Imaging Spectroradiometer

NDVI Normalized Difference Vegetation Index

PCA Principal Component Analysis

PFT Plant Functional Trait

Qsoil Soil Moisture

Qsoil1 Soil Moisture (0-7cm)

Qsoil2 Soil Moisture (7-28cm)

Qsoil3 Soil Moisture (28-100cm)

Qsoil4 Soil Moisture (100-255cm)

SPEI Standardized Precipitation Evapotranspiration Index

SPI Standardized Precipitation Index

Tair Air Temperature (at 2m above ground)

TRY TRY Plant Trait Data Base

VI Vegetation Index

1. Introduction

Ecosystems are subject to disturbance regimes which can undergo substantial alterations, especially due to climate change[9]. These climatic changes include[10]: (1) increased frequency and intensity of climate extremes[11], and (2) increasing temperatures and levels of aridity in dryland regions[12] which lead to shifts in environmental and ecosystem processes[9,13,14]. This has raised growing concern that future disturbance dynamics will exceed ecological tipping points and cause non-linear shifts in ecosystem state in relation to **attractors**[15] (see table 1.1 for definitions of bold font terms relating to the resilience debate). Such a **regime shift**[16–18] can have major impacts on local as well as global welfare of mankind[19–21]. Therefore, policy makers have made it their mission statement to enforce actions which serve to maintain or boost **resilience**[2]; a biological property which can safeguard against regime shifts. The Aichi biodiversity targets, for example, include the following statement: *"By 2020, ecosystem resilience [...] has been enhanced, [...] thereby contributing to climate change mitigation and adaptation and to combating desertification"* as target 15[1]. Whilst the Aichi targets are set to encompass global action, national policy makers, such as the Australian Government, are also recognising the importance of achieving climate resilience: *"We achieve climate resilience when short, medium and long-term decision making considers current climate risks and a changing climate"*[22]. Conceiving such strategies and evaluating their respective efficacies requires a thorough understanding of resilience theory and a robust set of tools to measure the resilience of ecosystems.

Generally, resilience can be understood as the capacity of a system to absorb **disturbances** and respond to changing conditions so as to still retain the same function, structure, identity, and feedbacks[18,23]. First introduced to biological sciences in 1973 by Holling, resilience thinking encompasses two separate characteristics of ecosystems[23]:

1. *Stability* is defined as the ability of a system to return to its original state after a temporary disturbance. The more stable a system, the less it fluctuates in its characteristics and the faster it returns to its initial conditions after a perturbation. Contemporary literature usually refers to this property as **engineering resilience**[5,14,24].

2. *Resilience*, on the other hand, was characterised as the property of a system that allowed it to withstand disturbances without collapsing. The more resilient a system, the higher its chances of persisting under changing conditions. Today this is mostly referred to as **ecological resilience**[25–27].

Within this early context on resilience thinking, a system could be resilient whilst not being very stable and fluctuating in its properties a great deal[23]. Consider Andean grasslands as an example: wildfires lead to immense fluctuations in vegetation cover without making it impossible for the system to recover[28]. As such, resilience thinking often considers large spatial and temporal extents[29] which lend themselves well to remote sensing approaches[30,31].

Given the importance of assessing resistance in ecosystems with respect to global sustainability goals, baseline metrics on which to assess and compare resilience in different ecosystems are required[26]. This poses a **major research challenge** of choosing a measure by which to assess resilience and ecosystem components reliably through time and space. Consequently, recent studies have focussed on identifying and quantifying sub-components of ecosystem resilience[7,32–34]. This is especially the case in remote sensing research where data availability is largely determined through what data repositories are available.

One solution to establishing comparable metrics of resilience across ecosystems is to break down ecological resilience into different components. Hodgson et al.[4] propose representing resilience as two separate entities: (1) *Resistance* to perturbation and (2) *Recovery* from perturbation.

Table 1.1: Basic Vocabulary of the Ecological Stability Debate - A few definitions that are key to the understanding of the resilience framework.

Term	Definition
Attractor	A regime towards which a system moves asymptotically [18,20,35–39]. Also known as *stable state* [29], *(stable) equilibrium* [24] or *regime* [37].
Disturbance	Any impact that perturbs a systems trajectory from a given attractor [26,37,40–42].
Regime shift	The restructuring of a system from one set of dominant processes and feedbacks governed by one attractor to an alternate set governed by an alternate attractor [35,36,43–45]. Also known as *tipping point* [16] or *bifurcation* [18].
Ecological Resilience	'[...] measures the magnitude of disturbance that can be absorbed before the ecosystem's structure changes' [3]. Also known as *General resilience* [39]. Ecological resilience is a measure of how much a system can be changed until it shifts from one stable state/attractor to another [3,4,23,26,27,39,46,47].
Engineering Resilience	'[...] measures the speed of recovery after the disturbance' [3]. Also known as *Recovery* [26], *Stability* [23] or *Resiliency* [39]. Engineering resilience is a measure of how fast a system reverts to its pre-change state. It can be measured by assessing the return time [3,4,23,26,27,37,39,46,47].

Ecosystem resistance capacities are of vital importance to ecosystem integrity, particularly in agricultural landscapes, and are usually studied in terms of linear regression models [30]. Recovery rates have been used as proxies for resilience metrics in many study settings including (1) palaeoecological research (e.g. by linking recovery rates of tropical ecosystems and disturbances categories [48]), (2) observational studies (e.g. drought-impacts across different temporal and spatial scales [49]), and (3) experimental research (e.g. through applying small-scale disturbances and recording recovery rates [50]). Contemporary research often identifies recovery rates in terms of a fixed point of equilibrium, a perturbation event, and a measure of time to recover to the equilibrium. Another option is to use autoregressive models, effectively tracking how ecosystem properties drive themselves through time thus enabling ecosystem recovery [3].

The *resistance* and *recovery* aspects of the Hodgson et al. framework to resilience can be represented through **ecological stress memory** (also know as 'ecological memory', see table 1.2 for definitions of bold font terms relating to the ecological memory framework). Ecological memory is a compound metric and whilst it is often understood to be synonymous with **adaptation to disturbances** resulting in increased biological fitness in the face of repeated stress events [51], other studies employ **ecological memory as a proxy for recovery rates** [3].

Nyström & Folke reported a direct link between spatial resilience and ecological memory of coral reefs [52]. In forest ecosystems, a positive link between ecological memory and ecosystem resilience has been identified by Johnstone et al [9]. Additional support for this argument stems from remote sensing studies such as De Keersmaecker et al. [3] and Seddon et al. [7] who identified autoregressive coefficients of vegetation properties as proxies of engineering resilience (i.e. recovery rates).

However, even quantifying the ecological memory of an entire ecosystem is difficult as ecological memory is a compound metric unifying a multitude of sub-processes [51]. For example, the aforementioned Nyström & Folke recognised legacies, mobile links, and support areas as parts of ecological memory [52] whilst Jonstone et al. propose information and material legacies as the key factors of ecological memory [9]. Hence, other ecosystem-scale studies have refocussed their attention on identifying components of ecological stress memory such as **vegetation memory** [9,51].

1.1 Vegetation Memory

Ecological memory (hence also vegetation memory) effects can be understood as the influence of antecedent conditions on ecosystem properties given contemporary disturbance regimes [5,8]. The existence of such temporal effects emphasize the necessity of considering time in ecological frameworks by providing a link between antecedent environmental conditions and plant performance [7,51,53,54]. In fact, there is empirical evidence that past conditions inform contemporary vegetation morphology [51], phenology [5], primary productivity [8,55], species composition [51,56], and carbon cycles [5,54]. Studying vegetation memory sheds light on an important set of ecosystem processes which can provide information about the expected responses of ecosystems to climate change [51] and enhance our capabilities of accurately predicting ecosystem responses to a host of bioclimatic stressors.

1.1.1 Terminology

Vegetation memory can - in the broadest of terms - be understood as:

The impacts of antecedent conditions on current vegetation productivity. [3,5,8,51]

In a comprehensive paper, Ogle et al. expanded on this definition by identifying the following three important characteristics of vegetation memory [8]:

1. *Length* - The extent of time through which past conditions significantly affect the current state of vegetation.
2. *Strength* - The magnitude of the effect that past events/conditions have on the current state of vegetation
3. *Temporal Patterns* - The variation in relative impacts of antecedent conditions at different points in time.

Tackling the characteristic of vegetation memory origin, Ogle et al. introduced the notion of **intrinsic** and **extrinsic** memory effects [8] (see table 1.2 and figure 1.4).

Due to the complexity of extrinsic memory components and drivers I find it useful to add another criterion to extrinsic memory characteristics: *Extrinsic Source*. This represents the nature of the antecedent extrinsic conditions that drive current states of vegetation (i.e. the origin/source of the extrinsic memory effect), including (but not limited to) herbivory pressures and climate events such as changes in temperature or water availability.

Table 1.2: Basic vocabulary of the vegetation memory framework - A few definitions that are key to the understanding of vegetation memory and its components.

Term	Definition
Ecological Memory	Ecological memory is synonymous with information and material legacies: adaptations, individuals, and materials that persist after a disturbance and drive the responses to future disturbances [9,52].
Vegetation Memory	Vegetation memory is defined as any response of a single plant or vegetation compounds following disturbance event that alters the response of the plant/plant community towards future stress events. This includes a modification of interaction with other ecological components [51].
Intrinsic Memory	Intrinsic memory refers to the influence of antecedent conditions of the focal system in determining contemporary conditions of the same system. Also known as *endogenous memory.* [8]
Extrinsic Memory	Extrinsic memory refers to the influence of antecedent conditions of the environment (usually abiotic, climate factors) in determining contemporary conditions of the focal system. Also known as *exogenous memory.* [8]

1.1.2 Ecological Relevance

Vegetation memory is identified as the effect of antecedent conditions on current vegetation patterns [8]. The stronger and longer the memory effect, the more vegetation will be influenced by antecedent anomalies of the ecosystem and abiotic processes thus leading to slower recovery rates in the case of intrinsic memory and signalling higher sensitivity to these drivers [3] concerning extrinsic memory. A higher sensitivity translates to lower resilience and so:

Intrinsic vegetation memory has been proposed as a proxy for engineering resilience. [5,57]

Patterns and characteristics of vegetation memory are a vital information criterion for Decision Support Systems (DSSs) aimed at conservation or management efforts due to the inverse relationship between vegetation memory and engineering resilience.

Additionally, vegetation memory is a compound metric which may be driven via a variety of forcing factors [7]. Thus making use of the differentiation of *intrinsic* and *extrinsic* components of vegetation memory by Ogle et al. [8] aids in gaining additional understanding of what local vegetation reacts to. Contemporary studies of vegetation memory often focus on intrinsic vegetation memory [3] or a handful of extrinsic drivers [5,6]. Identifying further extrinsic memory drivers may enhance our understanding of extrinsic vegetation memory processes.

1.1.3 Quantifying Vegetation Memory

Vegetation memory is influenced by a multitude of processes which range from abiotic characteristics [3,7,54] to biotic system components [3,9]. Comprehensive studies of vegetation response to biotic and abiotic forcing are numerous [5,7,7,8,48,54].

1.1.3.1 Traditional Approaches

Traditional studies of vegetation memory leverage data obtained via extensive sampling campaigns and are thus laborious to carry out. Nevertheless, a multitude of such studies exist which provide a solid foundation for the understanding of ecological memory to date. Some traditional approaches - like Johnstone et al. [9] - establish theoretical frameworks of vegetation memory (i.e. as material and information legacies) which require empirical assessment through further research. One such study identified vegetation memory through spectral reddening as a proxy of spatial correlation of ecosystem properties. Increased spatial correlation is regarded as a proxy of high memory and ecosystems approaching regime shifts [58]. Additionally, these largely field-based studies are able to make use of a host of information criteria including (but not limited to) seed banks, wood thickness, canopy structure, and root depth [59]. Such information is often invaluable as it highlights the functional property of vegetation memory which is often identified through vegetation responses to different disturbances but rarely explained in terms of plant function that enables memory.

Due to limited data availability, such traditional studies are - although vitally important for our understanding of vegetation memory pathways - limited in scope. This will often lead to an advanced causal understanding of vegetation memory but only at limited spatial and temporal extents.

1.1.3.2 Remote Sensing Approaches

Remote sensing studies forego tedious field work data collection efforts and rely solely on easily available data. This data may either be obtained via remote sensing methods (i.e. satellite or drone imagery) or be pulled from published data sets from the field.

Using these data sets, remote sensing studies of vegetation memory offer vast improvements in spatial and potentially temporal coverage when compared to traditional approaches. This, in turn, allows for landscape-scale patterns of vegetation memory characteristics to be identified. Unfortunately, to achieve these large spatial extents, one often has to sacrifice spatial and temporal resolution. Especially climate data can be unreliable in some regions of the world when field measurements are sparse.

Recent remote sensing studies of vegetation memory have established a multitude of approaches to delineating vegetation memory from a variety of data sets. Examples include the use of:

1. **Correlation coefficients** of different lags of drought indices.

Prominent examples of these approaches (e.g. Vicente-Serrano et al.[6], and Liu et al.[5]), identify vegetation characteristics through time via the Normalized Difference Vegetation Index (NDVI) (aggregated from bi-weekly records to monthly-maximum composites). Drought characteristics are usually assessed either via the Standardized Precipitation Evapotranspiration Index (SPEI) or Standardized Precipitation Index (SPI). Additionally, data is assessed for certain quality criteria (e.g. monthly NDVI < .1 are removed) and standardised. Subsequently, correlation coefficients assess how well NDVI and drought index anomalies coincide (strength of memory, A in figure 1.1) and at what temporal lag these coefficients are the most relevant (length of memory - an extrinsic memory characteristic, B in figure 1.1 and figure 1.2).

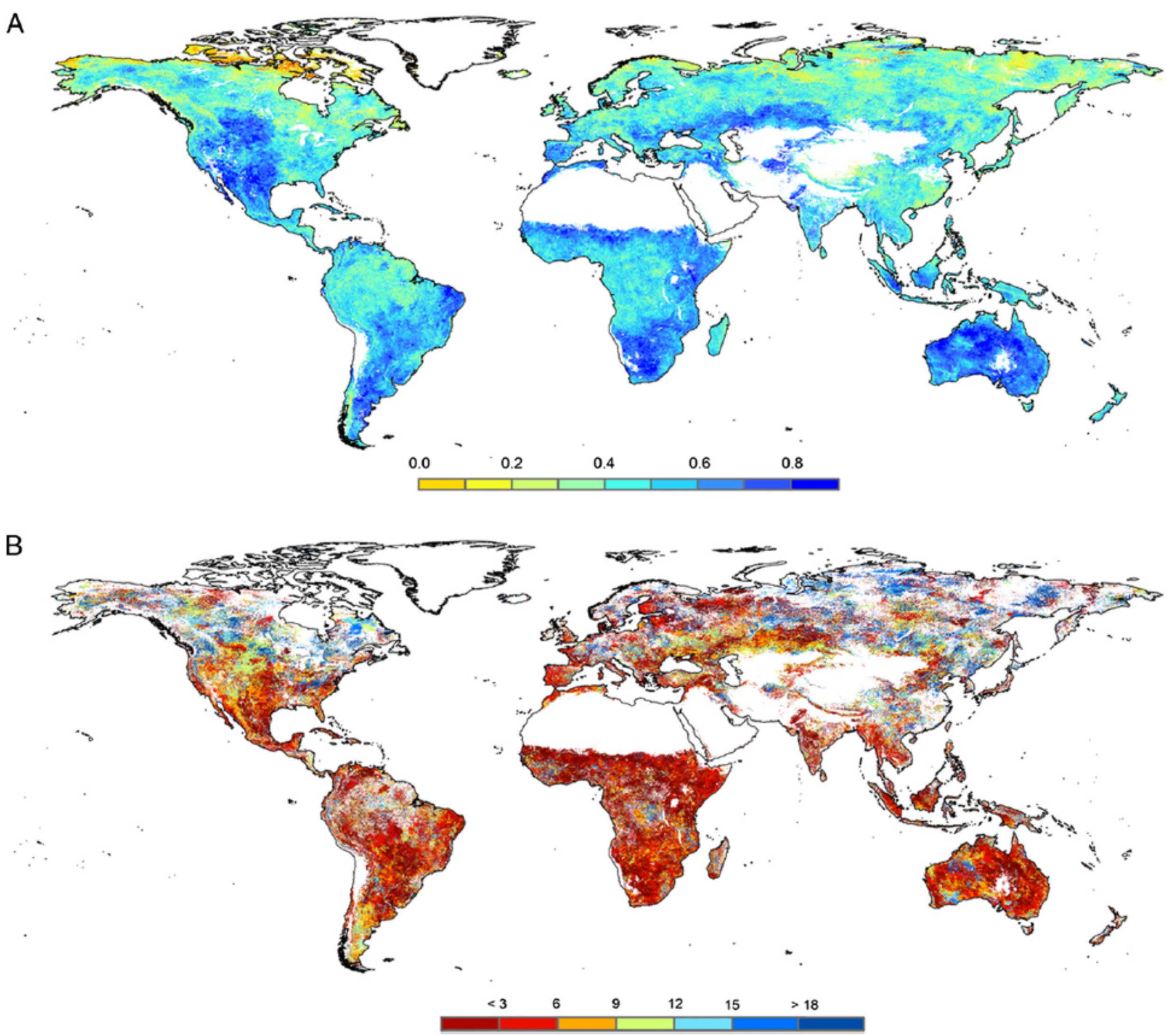


Figure 1.1: Drought Memory Length and Strength - Spatial patterns of drought memory length assessed via correlation coefficients. Figure lifted from Vicente-Serrano et al[6].

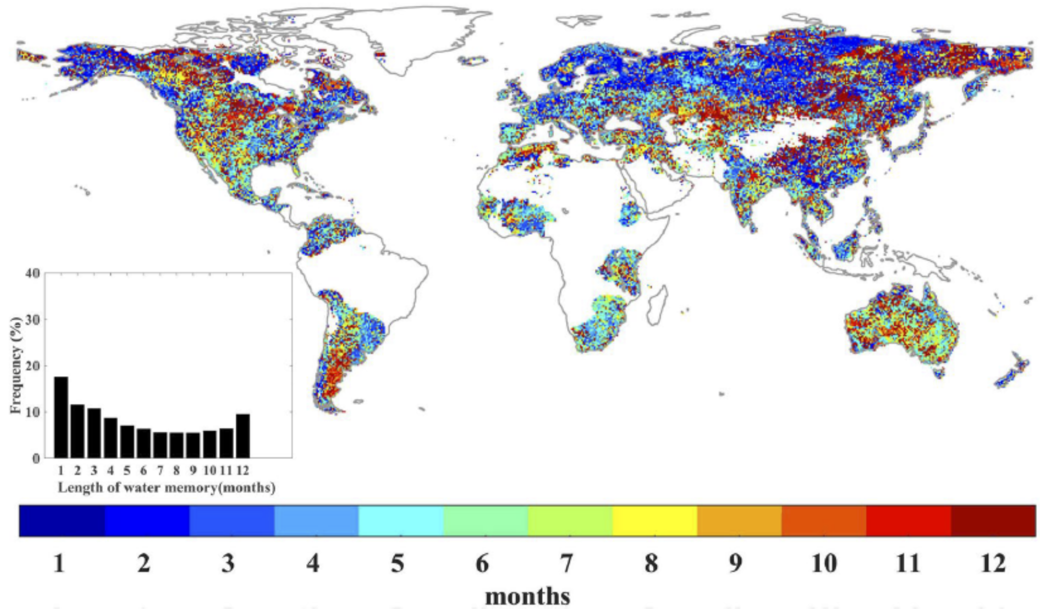


Figure 1.2: Water Memory Length - Spatial patterns of water memory length assessed via correlation coefficients. Figure lifted from Liu et al [5].

2. **Linear models** of antecedent vegetation and climate characteristics and subsequent model selection.

In their 2015 paper, De Keersmaecker et al. introduced a comprehensive vegetation memory modelling approach using NDVI, temperature, and SPEI data in a linear model. Again, certain data preparation steps such as removal of low-quality data, calculation of monthly-maximum composites of NDVI and subsequent standardisation were carried out. Climate data which did not meet the same spatial resolution as the NDVI data was resampled using the nearest neighbour method. Study results of the **first autoregressive model** (NDVI data regressed against NDVI data from the previous time step, also referred to as *intrinsic memory*) are presented in figure 1.3. Temperature records where implemented at no time lag whilst SPEI data was used as three-month lagged data.

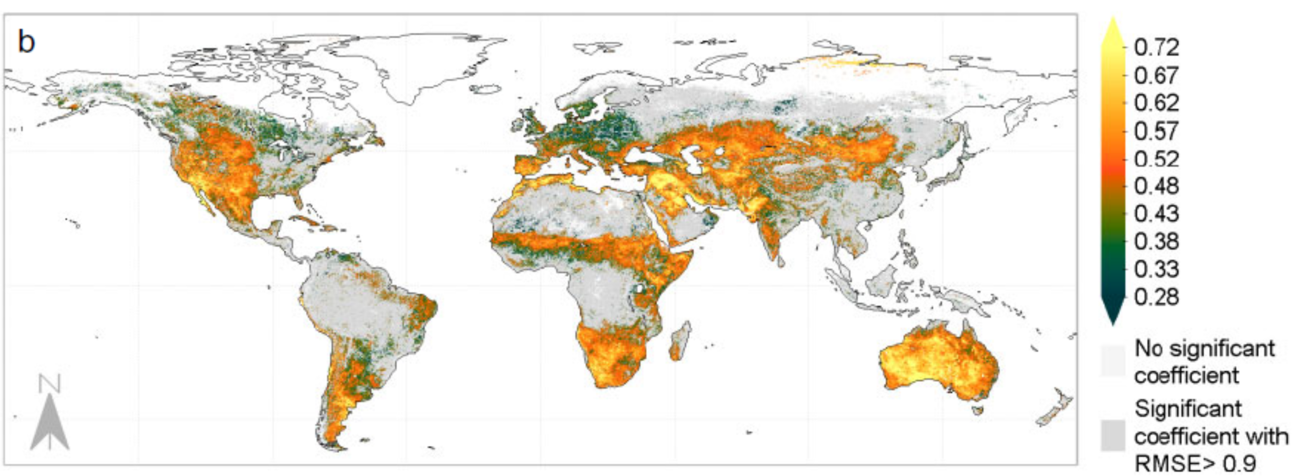


Figure 1.3: Intrinsic Vegetation Memory - Autoregressive coefficients of intrinsic vegetation memory depicting engineering resilience (high memory scores identify low resilience). Figured lifted from De Keersmaecker et al [3].

Seddon et al.[7] delineated autoregressive Enhanced Vegetation Index (EVI) patterns very similar to those presented in figure 1.3.

1.1.3.3 Synthesis

Models

Some common modelling practices have emerged from contemporary remote sensing vegetation memory studies:

1. **Intrinsic Memory** is usually identified as the relationship between vegetation properties at time step t and prior time $t-1$ (usually at monthly intervals)[3]. See figure 1.4 for a visual representation.
2. **Extrinsic Memory** is attributed to combinations of temperature, precipitation, and/or drought records[3,5,6,8]. See figure 1.4 for a visual representation.
3. **Model Selection** is used to determine the length of extrinsic memory effects by identifying the length of lag during which extrinsic forcing holds the most explanatory power of vegetation characteristics[3,5,6].
4. **Anomalies** (i.e. standardised records) of both response and predictor variables are used as these are metrics which can be linked directly to perturbation events[3].

Vegetation memory models of this type can be summarised as:

$$Y_t = \alpha * Y_{t-1} + \sum_{e=1}^{E} (\beta_e * X_e) \tag{1.1}$$

with Y_t and Y_{t-1} denoting anomalies of vegetation properties at times t and $t-1$ respectively, X_e indexing anomalies of the e'th extrinsic force (out of a total E extrinsic drivers), and α and β representing the **coeffcients** of **intrinsic** and **extrinsic** memory effects respectively. Rerunning such a model with independent, extrinsic inputs (X_e's) at different time lags and subsequent model selection can then be employed to identify the length of the extrinsic vegetation memory related to the e'th extrinsic force.

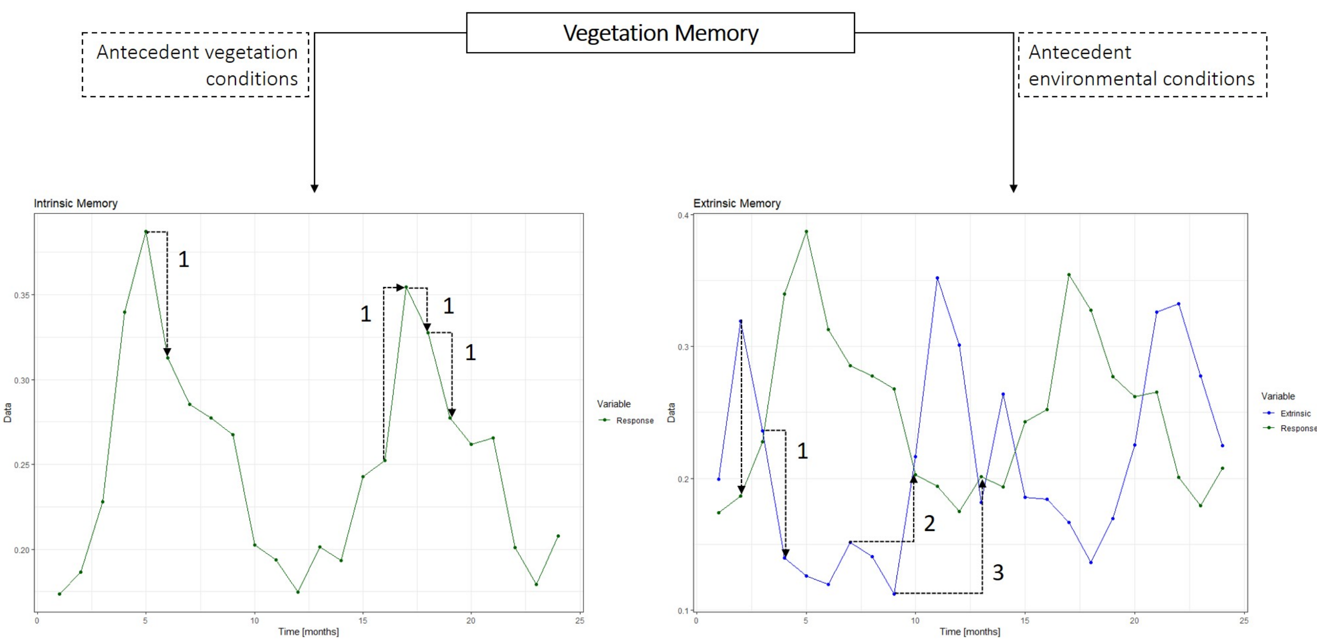


Figure 1.4: Components of Vegetation Memory - Conceptual layout of how intrinsic and extrinsic vegetation memory components can be identified. The concept recognises two important variables: (1) *Response* - contemporary vegetation properties at time t, and (2) *Extrinsic* forcing - antecedent environmental data. Within this framework, vegetation memory - both intrinsic and extrinsic - is identified as the predictability of the response variable through a regression onto itself and extrinsic variables respectively. Regressions of data onto each other are visualised as dashed arrows, lag effects are identified by numbers. Some parts of this figure have been generated via Chunk 35.

Spatial Memory Patterns

Results of the studies above clearly indicate some areas of special interest:

1. *High Autocorrelation Coefficients* indicating potentially strong vegetation memory, which has been interpreted as low engineering resilience, can be found largely in *dryland regions* (e.g.: North-East Brazil, Australia, and Spain; figures 1.1 and 1.3).
2. *Trends in Memory Length* signalling a change in causal patterns of memory effects can be found across the Iberian Peninsula, and the contiguous United States of America. See figures 1.1 and 1.2.
3. *Intrinsic and Extrinsic Memory* have often been assessed separately through recovery rates (intrinsic memory) and vegetation sensitivity (extrinsic memory). It is noteworthy that areas characterised by strong intrinsic memory such as the Sahel region have been identified to exert low vegetation sensitivity towards climatic drivers. This suggests a trade-off which requires further assessment in a unified approach of both memory components.

Furthermore, all of these regions are characterised by gradients in vegetation memory intensity (figures 1.1 and 1.3). A unified model, taking into account of intrinsic and extrinsic components is needed to understanding vegetation responses, and the links to resilience, in dryland regions.

Data

Vegetation properties are usually assessed using NDVI data. Additionally, indices of climate metrics are used for the calculation of extrinsic memory components.

Contemporary studies of vegetation memory largely rely on *precipitation* or *drought* data obtained via the SPEI drought index [3,57], and the Climatic Research Unit (CRU) data set [5]. These data sets are *observational* data sets as they are built from observational data often involving simple interpolation methods [60]. Climate science has developed novel data sets which are more advanced than purely observational data sets. These are known as climate reanalysis products.

Climate reanalysis data sets are *self-consistent*, and *gap-less* (in time and space) and thus **superior** to many observational data sets for ecological application. The most advanced reanalysis data set is the **European Centre for Medium-range Weather Forecasts ReAnalysis 5 (ERA5)** [61]. The superiority of ERA5 is largely due to:

1. The **volume of observational data** used to create the ERA5 product with data types ranging from satellite data to weather station/independent institute data collection efforts, and station data from a wide variety of data providers [62]. Traditional observational data sets are often characterised by their individual biases in sampling, coverage, and choice of interpolation and homogenization to create a gridded product.
2. Its **sophisticated nature** building upon data assimilation procedures [63], and complex models [64] to shed light on physical processes [65]. As such, ERA5 is currently the *state-of-the-art* climate reanalysis product in climate science and benefits from the developments in data assimilation methodology as well as understanding of physics of the climate system and their interactions, which have come about in recent decades.

Finally, ERA5 improves on other prominent climate data sets used in ecological studies like the aforementioned CRU and SPEI, WorldClim (used by Garris et al. [66]), or the Climatologies at High Resolution for the Earth's Land Surface Areas (CHELSA) data set [67] in a multitude of ways since it offers:

1. Superior *Temporal Resolution* (superior resolving of climate extremes)
2. Superior *Spatial Resolution* (advanced resolving of local/topographical features)

Therefore, one can reasonably expect to improve on contemporary vegetation memory studies by substituting observational climate data sets with the highly advanced ERA5 data set.

1.2 Functional Aspects of Vegetation Memory

Contemporary studies of vegetation memory largely describe memory effects in terms of spatial patterns and effect sizes, but forego analyses of plant physiology or morphology to **explain causal pathways** leading to vegetation memory. Addressing these could enable more refined DSSs. Expressions of plant morphology and physiology are manifold and so one may wish to enlist multiple different proxies of these biological properties to rationalise vegetation memory effects. I hypothesise that expressions of vegetation memory can be explained through the lens of **functional ecology**. A functional approach to ecosystem processes can potentially explain what functional expressions allow said processes to persist despite perturbations[68–71] thus informing resilience metrics such as vegetation memory. One of the most prominent tools of such studies is the *functional trait framework*. Arguably, the most important of these frameworks when understanding vegetation processes - such as vegetation memory - are those of **Life History Trait (LHT)** and **Plant Functional Trait (PFT)** frameworks.

1.2.1 Terminology

LHTs and PFTs both capture crucial information of vegetation capabilities to cope with stressors. Despite recording vastly different aspects of plant function in respect to time, they should be regarded as *complementary* due to their demonstrated capabilities in explaining natural phenomena.

1.2.1.1 Life History Traits

LHT frameworks (see table 1.3 for definitions of bold font terms relating to LHTs) assess plant/population characteristics through time thus mirroring the temporal aspect of vegetation memory leading up to, during, and following perturbations. Expressions of LHTs explain temporal community processes and may thus be major drivers of local vegetation memory expressions. Prominent examples of LHTs include measures relating to (1) **System Turnover**, (2) **Longevity**, (3) **Growth**, and (4) **Reproduction** all of which retain information of how individual biological processes shape vegetation communities and have been found to explain a majority of variance in plant life history strategies[72].

Specifically, I hypothesise that vegetation memory effects are at least partially linked to:

1. The **Fast-Slow Continuum (FSC)** is a Principal Component Analysis (PCA)-based approach whose two main axes capture over 60% of the variation in plant life history strategies[73] and are largely made up of the LHTs contained in table 1.3[72]. I hypothesize that expressions of the Fast-Slow Continuum (FSC) will strongly interact with vegetation memory effects due to the holistic nature of the FSC and the temporal nature of both LHTs and vegetation memory effects. The two axes read as follows:

   (a) *Life History Speed within the Fast-Slow Continuum (FSC-1)* explains around 35% of global LHT variation and contrasts species of low generation times, early maturity, and fast growth with species of high generation times, delayed maturity, and slow growth[73]. Fast species rank low on this axis[74] and possess an innate potential to recover fast from perturbations through re-population thus strengthening engineering resilience and lowering vegetation memory.

   (b) *Reproductive Strategy within the Fast-Slow Continuum (FSC-2)* captures approximately 25% of global LHT variation and contrasts species of low reproductive output with species that reproduce much, often, and for a long time[73]. Species of low reproductive output rank low on this axis[74] and are expected to struggle to return to a pre-disturbance state thus intensifying vegetation memory effects.

2. **Reactivity** is a *first time-step* information criterion and a proxy for instantaneous biological responses[75] and may thus be an important criterion in areas of high vegetation memory. I expect lower engineering resilience in these regions as high reactivity enables fast responses thus strengthening engineering resilience.

3. The **Damping Ratio** (ρ) can be regarded as a measure of intrinsic resilience as it indicates how fast transient dynamics of a perturbation event decay; the larger ρ, the faster the population will converge[75] and the higher the engineering resilience of the focal population.

4. The **Period of Oscillation** (π) is a measure of population life cycles in periodic environments with more rapidly growing populations usually being classified by higher π values[76] and theoretically higher engineering resilience.

Table 1.3: Basic Vocabulary of Life History Trait Frameworks - A few definitions that are key to the understanding of Life History Trait frameworks. Largely influenced by Salguero-Gómez et al[74].

Term	Definition
Life History Traits (LHTs)	LHTs characterise life history processes of populations thus identifying temporal aspects of their life cycles.
System Turnover	
Generation Time	The time needed to fully replace a population by a new cohort.
Longevity	
Survivorship Curve	Shape of age-specific survivorship curve type.
Age at Sexual Maturity	Time it takes for an average individual of a population to become sexually reproductive.
Growth	
Progressive	Mean probability to advance in life stages.
Retrogressive	Mean probability to regress in life stages.
Reproduction	
Net Reproductive Rate	Mean number of recruits produced by each individual in the population.
Degree of Iteroparity	Spread of reproduction during the lifespan of an individual.
Mature Life Expectancy	Time between age of sexual maturity and life expectancy.

1.2.1.2 Plant Functional Traits

PFT frameworks (see table 1.4 for definitions of bold font terms relating to PFTs) convey information as *snapshots* of reality as their data sets usually do not contain repeated measures of the same individuals over a period of time. These data sets have been linked to evolution[77] and community composition patterns[71,78]. A change in their expression has been found to be correlated with climate change patterns[79] potentially suggesting a causal link between extrinsic vegetation memory and PFT expressions. To classify the way in which PFTs are related to community functions, Nock et al. present the concept of (1) **Effect Traits**, and (2) **Response Traits**[80]. Classifying PFTs according to this framework can profoundly aid the understanding of their influence on spatio-temporal patterns of vegetation performance[44,71]. Effect traits determine how PFTs alter ecosystem processes. Response traits affect how vegetation reacts to perturbations. Whilst effect traits predominantly drive post-disturbance recovery, response traits signify potential resistance or recovery along a continuum of individual PFT-driven responses either influencing resistance or recovery.

Recent studies suggest that vegetation community functions are governed by integrated phenotypes, which can be regarded as combinations of PFT ranges[80,81] thus calling for a scientific approach focussed on **Functional Diversity Richness (FDR)** which can be identified via an approach in which PFT values of each individual are placed within a multidimensional system with each axis representing the range of values for one specific PFT[82].

Additionally, one may also want to use a measure of evenness which, in a functional biology setting, can be referred to as **Functional Diversity Dispersion (FDD)** which is representative of how evenly dispersed data records are in multidimensional space.

Identifying FDR and FDD requires a lot of individual PFT data and so contemporary studies in PFT research have focussed heavily on methods of dimensionality reduction by identifying the PFTs which capture a vast part of the global variation of plant function [78,83]. One such framework has been established by Westoby & Wright in which three important PFT dimensions are recognised: (1) the leaf economic spectrum, (2) the seed size/mass spectrum, and (3) the height of the canopy at maturity [84].

Table 1.4: Basic Vocabulary of Plant Functional Trait Frameworks - A few definitions that are key to the understanding of Plant Functional Trait frameworks.

Term	Definition
Plant Functional Traits (PFTs)	Plant Functional traits characterise morphological, biochemical, physiological, structural, phenological and/or behavioural aspects of organisms which influence the fitness of said organisms [80].
Effect Traits	These determine a species influence on ecosystem processes [80].
Response Traits	These determine a species ability to colonise a certain habitat and persist despite environmental pressures [80].
Functional Diversity Richness (FDR)	Functional diversity richness measures the range of the PFT spectrum of a given study region/species assembly in multivariate trait space via a multivariate convex hull effectively measuring niche space [82].
Functional Diversity Dispersion (FDD)	Functional diversity dispersion measure the dispersion of PFT values in multivariate trait space with respect to the local functional space centre effectively measuring niche segregation [82].

Using an empirical approach similar to the aforementioned FSC, Díaz et al. identified:

1. *Vegetative height* (H) - an important criterion to accessing light resources
2. *Stem specific density* (SSD) - reflecting a trade-off between growth potential and mortality risk
3. *Leaf area* (LA) - vital consequences for leaf water balance
4. *Leaf mass per area* (LMA) - a proxy for the trade-off between carbon gain and leaf longevity
5. *Leaf Nitrogen Content per Leaf Dry Mass* (N_{mass}) - reflects a trade-off between photosynthetic potential and acquiring nitrogen
6. *Diaspore mass* ($Dmass$) - reflects a trade-off between seedling survival and colonization ability

as the most important PFTs to understanding the global spectrum of plant functioning [83]. Recovery rates of vegetation post-disturbance (i.e. vegetation memory) is influenced by energy and biomass availability. Resource allocation and hence recovery potential using the above six traits is best understood employing H, LMA, and N_{mass} due to their direct link to realising photosynthetic potential. Resistance potential, on the other hand, is influenced heavily by the physical toughness of plant material, presence of storage organs, and many other aspects of plant function captured by PFTs and so SSD, LMA, and N_{mass} are especially promising as drivers of ecosystem resistance.

1.2.2 Ecological Relevance

The diverse range in which measures of plant function can be expressed are representative of the diverse strategies for establishing, growing and reproducing of plants [78]. Hence, understanding how plant functioning enables plants to persist despite (a)biotic forcing (i.e. vegetation memory) is a promising step to understanding ecosystem-scale responses to climate change. This is due to the positive effect of biodiversity on ecological stability which has long been recorded [68] and even led to the formulation of prominent concepts like Yacchi's insurance hypothesis which postulates that the loss of a subset of species can be compensated for by the presence of other species given a sufficient overlap in functional aspects of both, thus leading to a stabilisation of ecosystem processes [85].

PFTs and LHTs capture information about a variety of ecosystem processes. [86]

Most studies to date have focussed on species diversity [18,19,47]. Some made use of the concept of functional species [27] which presents researchers with a species classification system that is vague at best due to loose definitions of where to draw the lines between functional species. Functional trait biology, which captures the functional diversity of biological systems, is often employed as an improvement to the functional species/type concept [51,53] as it eliminates the necessity of phylogenetic analyses [80] and enables studies to capture intra-specific variation [87–89] thus making landscape-scale studies (such as remote sensing approaches) of vegetation memory patterns and their links to functional diversity possible.

PFTs and LHTs can be captured irrespective of species relationships. [88]

Thus, PFTs and LHTs could make for *highly informative* proxies of *ecosystem functioning* which can be recorded across *large spatial and temporal scales* since studies aren't constrained to a handful of species. Additionally, LHTs and PFTs represent an improved understanding of functional expressions of plant communities by *expressing information at sub-species level* [87]. Recent research based upon LHT data, for example, has revealed robust trade-off axes in plant [73,74] as well as animal life histories [90,91]. Information about LHTs has been demonstrated to be of use in trying to understand biological responses to temporal phenomena [92] and should thus prove crucial in delineating causal, biological pathways to establishing vegetation memory.

1.2.3 Unifying Information on Plant Functional Traits

PFTs are usually recorded through individual trait campaigns or for single-study purposes. Combining PFT information from these individual data sets can prove challenging due to different measurement practices and understandings of PFTs. A comprehensive handbook on PFT data collection practices by Pérez-Harguindeguy et al. [93] presents an elegant solution to making PFT data sets more comparable. Additionally, LHT and PFT data sets have been aggregated into large data bases within online repositories such as:

- The **TRY Plant Trait Data Base (TRY)** [89] is a data base of PFT records connecting vegetation scientists across the globe and hosted by the Deutsches Zentrum für integrative Biodiversitätsforschung (iDiv). Within its current release version (version 4) TRY contains almost seven million PFT records for 1,800 traits with about half of all records being geo-referenced [94].
- The **Botanical Information and Ecology Network (BIEN)** [95] is a repository of global plant diversity, function, and distribution. Its current release (version 4.1) contains over 200 million records at global coverage [96].
- The **COMPADRE Plant Matrix Data Base (COMPADRE)** [72] contains LHT-driven matrix population models which produce important information about community processes through time based on records of different LHTs [74]. In its current release (version 5.0.0) it contains over 7,000 matrix models [97].

Combining information from these data bases should result in a holistic view of global plant function as such an approach would unify PFTs and LHTs to further a spatio-temporal understanding of global plant function.

1.2.4 Putting Plant Function on the Map

Vegetation memory patterns are inherently spatial phenomena [3,5,6,8]. So are patterns of plant function [82]. Whilst the above mentioned data bases (TRY, BIEN, and COMPADRE) contain vast amounts of geo-referenced PFT and LHT records, almost no comprehensive, gapless, and internally consistent global products of plant function are available [67,86]. The method for extrapolating PFT records to local/global map products used here is inspired by the approach layed out by Ordonez & Svenning [82] but foregoes some of their more advanced data handling techniques. Species-specific PFT means are combined with geo-referenced species occurrences (a conceptual depiction of this can be seen in figure 1.5) to gridded products.

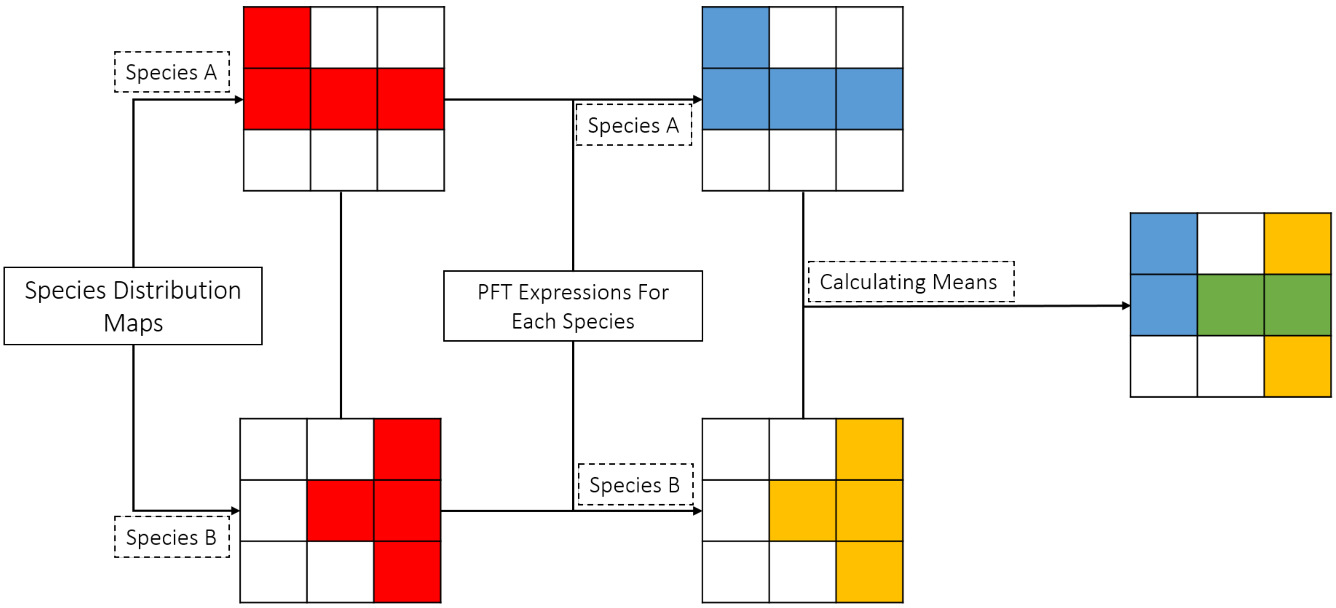


Figure 1.5: Concept of Combining Plant Functional Trait Information with Species-Distribution Maps - Conceptual framework of mapping plant function from PFTs and species distribution maps into gridded map products of mean PFT expressions. Inspired by Ordonez & Svenning [82]. Red cells indicate presence of a species, blue and yellow cells depict the average PFT of species A and B respectively. Green cells represent the average of PFT means of species A and B whilst white cells denote the absence of species and thus PFT data.

Species occurrence data - referred to as **floral data** in the case of plant occurrences - can be obtained via the **Global Biodiversity Information Facility (GBIF)** data base which currently contains over 1 billion geo-referenced occurrence records of species across all kingdoms of life [98].

Unfortunately, this approach is heavily reliant on geo-referenced occurrence records which are subject to sampling bias [99] and renders intra-specific expressions of PFTs virtually non-existent as species are expressed via single mean values for each PFT. Zoological studies have resolved these issues by recording relative abundances of species in addition to occurrence records [100] to calculate more accurate representations of community-weighted mean PFT records.

Recent approaches in PFT ecology have seen the development of mapping procedures of PFTs according to abiotic conditions through linear models with spatial components [67] and complex Bayesian methods [86]. No such method is available for extrapolating LHT records to broader coverage as these come with an added dimension of time.

Despite the obvious short-comings of this mapping method, linking PFT and LHT data to remote sensing vegetation memory studies is a novel approach in itself. This is built around the notion of vegetation memory being driven not only by antecedent conditions but also by functional characteristics of the local vegetation. Making use of the above PFT extrapolation method thus serves as an exploratory analyses of how well it performs when trying to extract causal pathways leading to vegetation memory.

1.3 Thesis Outline

Remote sensing analyses have been the key to identifying vegetation memory patterns through time and space in a multitude of settings [3,5,6]. Refining such approaches will facilitate important knowledge for DSSs on environmental processes in the face of climate change as well as changes in land use by mankind [101]. Considering the direct link between vegetation memory and ecosystem resilience, this will garner an advanced comprehension of resilience and, in turn, enable maintenance and enhancement of ecosystem resilience thus stabilising natural systems. It is therefore the **first main goal** (see figure 1.6) of this study to:

I. Identify vegetation memory patterns while improving on contemporary approaches.

To do so, I am going to answer the following research questions while focussing on dryland regions and ERA5 data:

1. *To what extent can extrinsic and intrinsic vegetation memory be identified in dryland regions?*
2. *How well can we distinguish between intrinsic and extrinsic memory in dryland regions?*

A multitude of recent studies have identified vegetation memory characteristics of a diverse cast of focal ecosystems [5,7,33,51,54,57,102,103]. However, only a few of these have also delineated the ecological processes and causal pathways which have led to the observed patterns of vegetation memory. Achieving an understanding of how functional aspects of vegetation may alter vegetation memory capabilities should serve to bolster our understanding of how resilience comes about and can be maintained. Therefore, the **second main goal** of my thesis is to:

II. Determine how vegetation memory and plant function are linked.

Achieving this goal is possible by answering the following research questions:

1. *Which traits of biological function (PFT and LHT) are related to vegetation memory characteristics?*
2. *What biological traits cause areas to exert intrinsic and extrinsic memory respectively?*

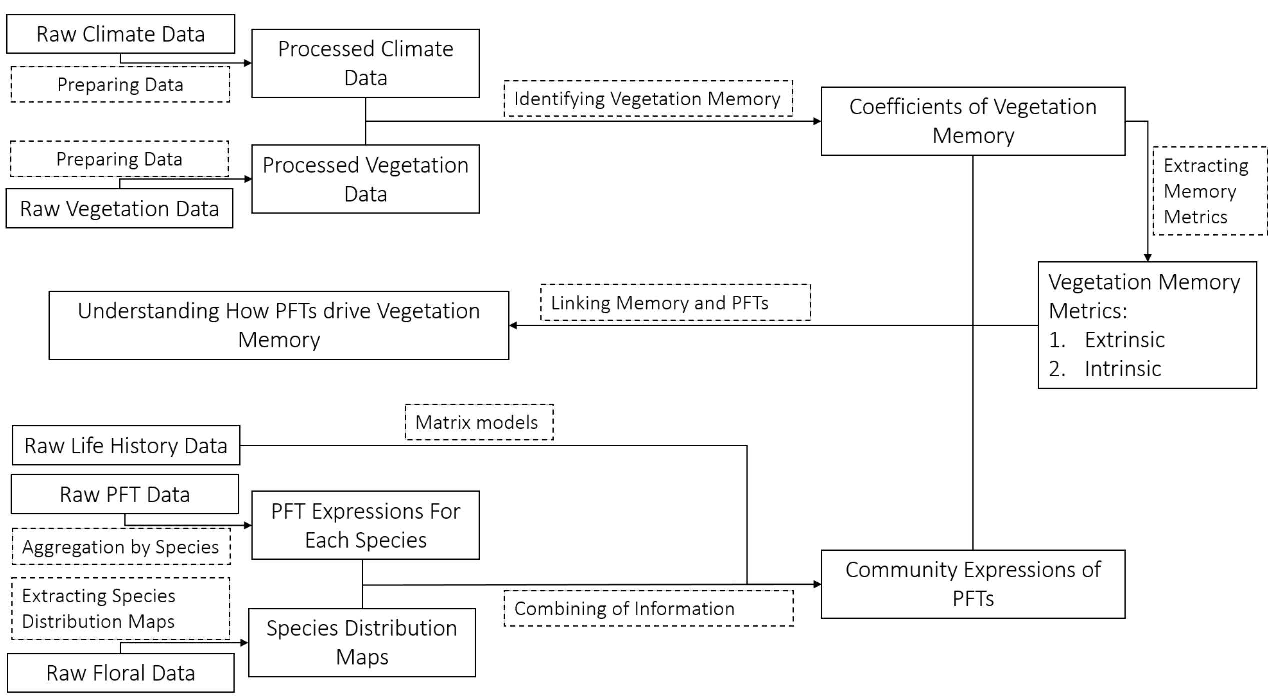


Figure 1.6: Study Outline - Conceptual study flow-chart. Data sets and products are shown in boxes with solid outlines whilst data handling and computational procedures are depicted via dashed-outline boxes.

2. Material & Methods

2.1 Study Regions

The spatial extent of study sites for assessing vegetation memory effects was limited to regions which were (1) suitable for remote sensing studies (i.e. receive satellite coverage year-round), (2) relatively well-sampled in terms of PFTs, LHTs and floral data, and (3) contain large areas classified as **drylands** (as classified via ombrotypes within Rivas-Martínez et al.[104]) which previous studies identified as regions of strong memory effects given autoregressive approaches. Within `R`, limiting to study regions is done using shapefiles (http://www.naturalearthdata.com/downloads/10m-cultural-vectors/) and the `rgdal`[105] package (see Chunk 3).

2.1.1 Iberian Region

The Iberian region - encompassing Portugal, Spain, Andorra and France within this study - has been selected as a study region due to a clear gradient in ombrotypes (see figure 2.1), its high density of TRY PFT and COMPADRE LHT data availability, and previously reported clear patterns of vegetation memory effects (i.e. stronger vegetation memory in Spain when compared to France)[3] and vegetation sensitivity (i.e. temperature/cloudiness sensitivity across the Pyrenees and water-driven vegetation sensitivity in all other areas)[7]. See figures A.4 and A.5 for an overview of Global Inventory Modelling and Mapping Studies (GIMMS) NDVI, ERA5 data, TRY PFT (both raw and extrapolated), and COMPADRE LHT data across the Iberian region.

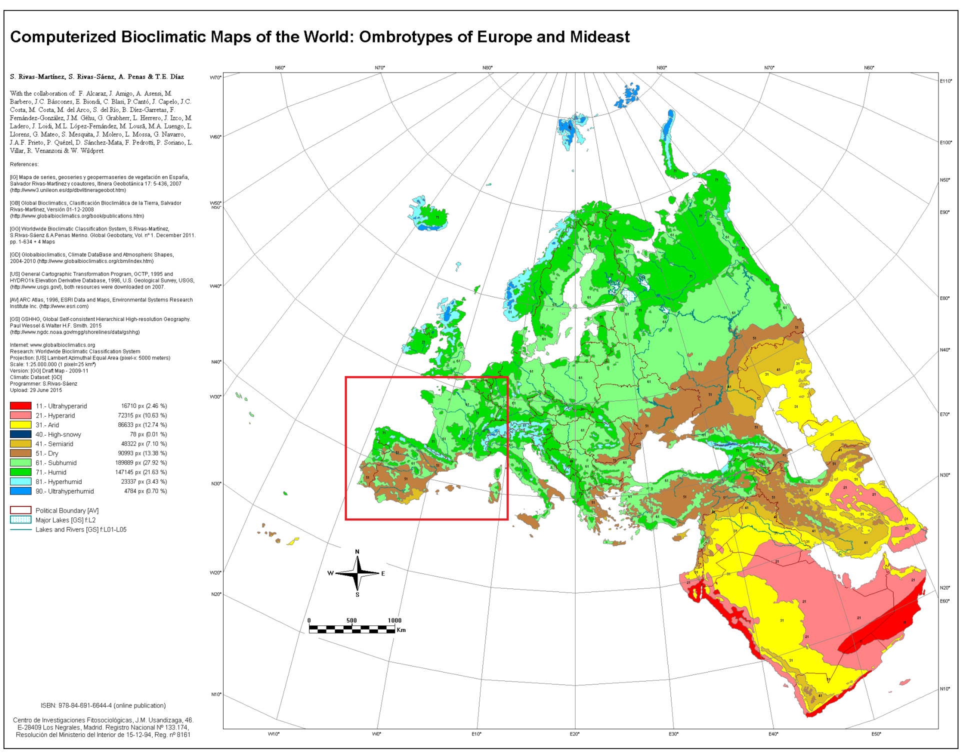


Figure 2.1: European Ombrotpyes - Regions of mean precipitation rates of biological relevance across Europe according to Rivas-Martínez et al.[104] and retrieved via the Worldwide Bioclimatic Classification System[106]. The Iberian Region is marked with a red rectangle.

2.1.2 Caatinga, Brazil

The Caatinga in Brazil - a dryland region located within northeastern Brazil - has been selected as a study region due to a predominance in dry ombrotypes and gradients of ombrotypes at its edges (see figure 2.2) as well as previously identified strong vegetation memory effects [3]. Additionally, Seddon et al. identified strong vegetation sensitivity towards water availability across this region [7] with stronger water sensitivity across the dryland ombrotype areas and lower water sensitivity in the adjacent humid areas.

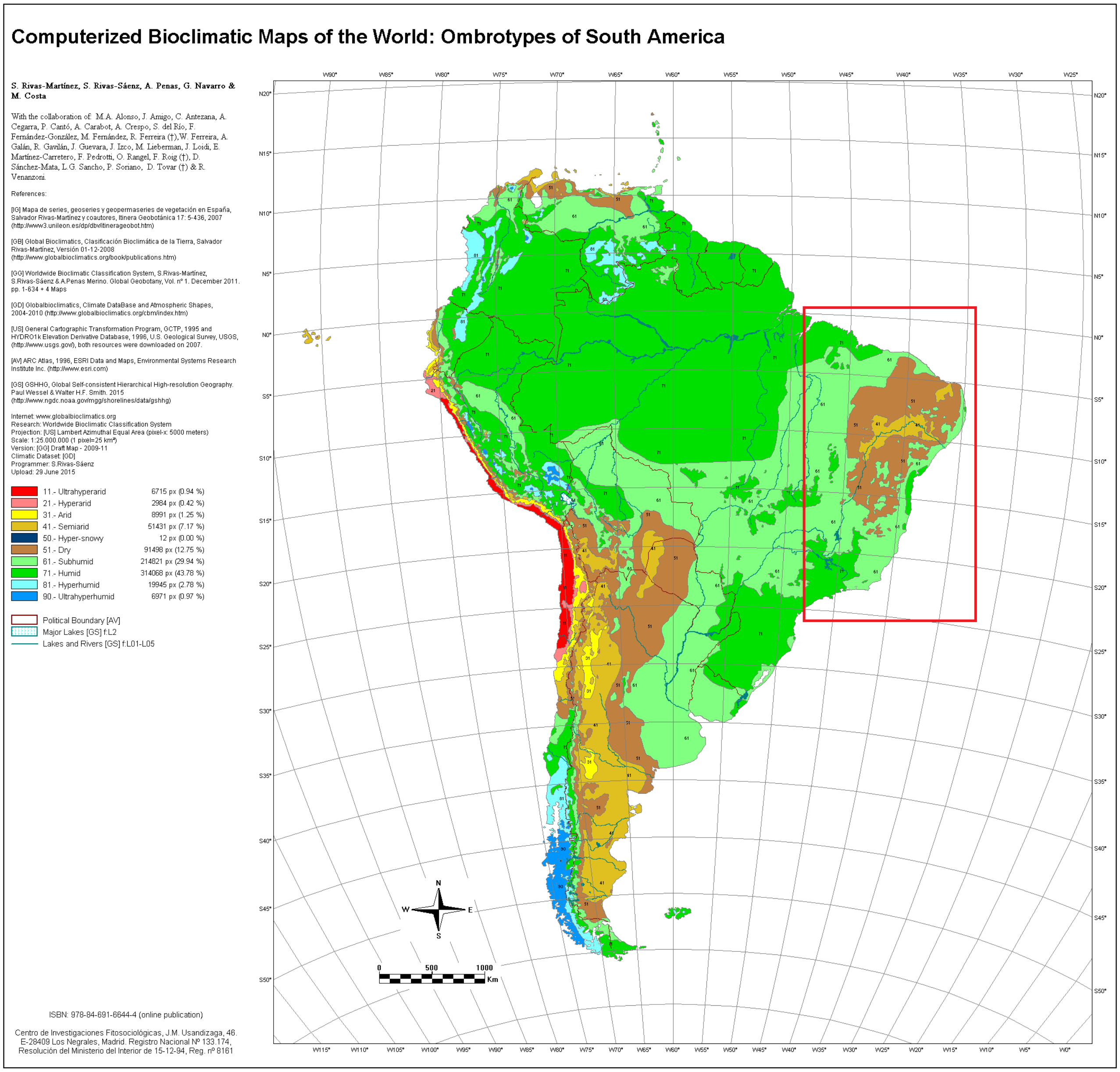


Figure 2.2: South-American Ombrotpyes - Regions of mean precipitation rates of biological relevance across South-America according to Rivas-Martínez et al. [104] and retrieved via the Worldwide Bioclimatic Classification System [107]. The Caatinga is marked with a red rectangle.

See figures A.6 and A.7 for an overview of GIMMS NDVI, ERA5 data, TRY PFT (both raw and extrapolated), and COMPADRE LHT data across the Caatinga.

2.1.3 Australia

Australia has been selected as a study region due to a predominance in dry and even arid ombrotypes as well as gradients of ombrotypes on its eastern coast (see figure 2.3). Previous studies have identified clear patterns of vegetation memory (overall strong memory with the strongest vegetation memory throughout the continental region of Australia [3,57]), vegetation sensitivity (highest autoregressive coefficients in western Australia, and strongest vegetation-water sensitivity in western Australia [7]), and water memory length and strength (strongest water memory in northern Australia, shortest memory in eastern Australia [6]) across Australia.

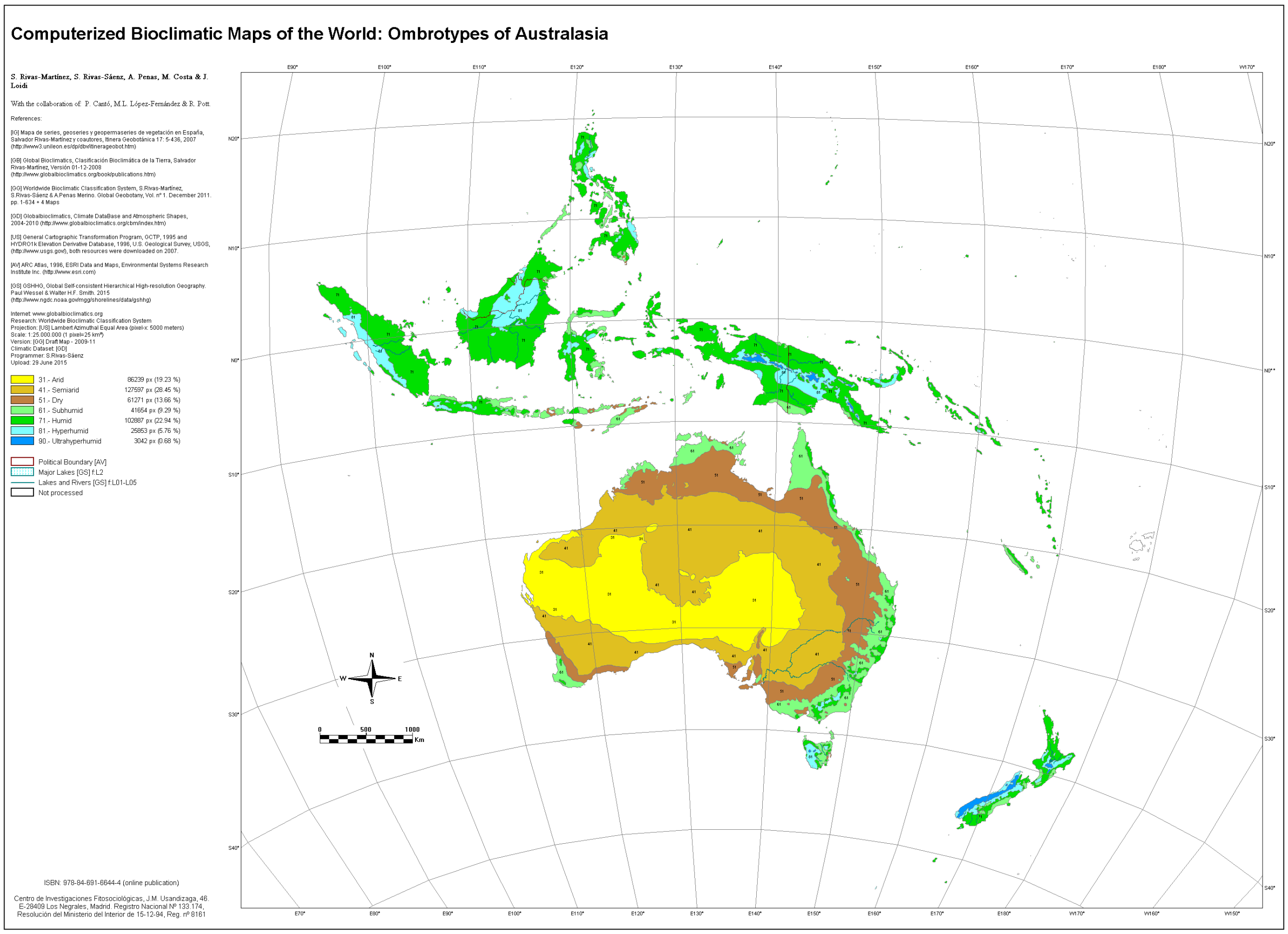


Figure 2.3: Australasian Ombrotpyes - Regions of mean precipitation rates of biological relevance across Australasia according to Rivas-Martínez et al. [104] and retrieved via the Worldwide Bioclimatic Classification System [108].

See figures A.8 and A.9 for an overview of GIMMS NDVI, ERA5 data, TRY PFT (both raw and extrapolated), and COMPADRE LHT data across Australia.

2.2 Identifying Vegetation Memory

2.2.1 Data

2.2.1.1 Vegetation Indices

Patterns of vegetation characteristics - both in time and space - can be readily assessed using remote sensing approaches [7,57,109–113]. A Vegetation Index (VI) is a remote-sensing proxy for different vegetation characteristics depending on the choice of VI. VIs have seen various applications in remote sensing including deriving vegetation dynamics on the scale of countries [114–116] or regions [112,117], and analysing resilience of vegetation assemblages to short-term climate anomalies [3]. Biome classifications have undoubtedly been one of the most prominent usages of the VIs [113,118–120] as have applications in agriculture [30,121].

Whilst a vast arsenal of VIs are available to macroecologists (see Cammarano et al. for an overview [112]), one of the main considerations in choosing an appropriate VI is that of **spatial resolution versus length of time series**. As satellite-born sensors receive hardware updates or other improvements - e.g. Advanced Very High Resolution Radiometer (AVHRR) versus Moderate-Resolution Imaging Spectroradiometer (MODIS) sensors - spatial resolution capabilities of VIs increase. However, we can't enhance the resolution of older records of VIs and are thus limited to lower resolution VIs if longer time series are desired. Since vegetation memory is a temporal phenomenon first and foremost, I have prioritised **time series length** over spatial resolution. Within the confines of my thesis, VIs represent the intrinsic characteristics of vegetation systems which will be used both as response and explanatory variables as outlined in figure 1.4.

Normalised Difference Vegetation Index (NDVI)

The NDVI has been selected as a proxy of vegetation characteristics in this study due to (1) its nature as an information criterion of biomass [122] and vegetation cover [112] (indicators of vegetation performance and composition), (2) the availability of a long time series data with global coverage [123], and (3) its demonstrated utility in various ecosystem studies [111] including assessments of vegetation sensitivity and memory [3,6,7]. See table 2.1 for an overview of the core characteristics of the NDVI data set.

Rouse et al. [124] first introduced the NDVI in 1974 and initially coined it 'Vegetation Index'. The NDVI is a composite VI that factors in measurements in the near infra-red wave band (NIR, 0.58-0.68 μm) and the red wave band (RED, 0.75-1.10 μm). These bands belong to the spectrum of light that is absorbed by chlorophyll and thus provide information on green vegetation [125]. The formula for NDVI is as follows [114,125]:

$$NDVI = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}} \tag{2.1}$$

with ρ_{NIR} being the reflection in the NIR band (reflectance from vegetation) and ρ_{RED} being the reflection in the RED band (background/dead biomass reflectance). The NDVI is bound between 0 and 1 with higher values representing greener vegetation. For a more detailed mathematical background on the calculation of NDVI scores see Matsushita et al [126].

Table 2.1: Core Information about NDVI data - Characteristics of the GIMMS NDVI3g data set (v.1).

Characteristic	Data
Resolution	$0.083° \times 0.083° \sim 9.27km \times 9.27km$
Time Series	Bi-weekly intervals from January 1982 to December 2015
Source	https://ecocast.arc.nasa.gov/data/pub/gimms/3g.v1/

NDVI data sets are readily available via the GIMMS. For this study the GIMMS3g data set was used (available at https://ecocast.arc.nasa.gov/data/pub/gimms/3g.v1/), which provides AVHRR NDVI data as 15-day maximum-value composites from 1982 to 2015 at $0.083° \times 0.083°$ resolution [127]. The data was processed in `R`, downloaded and compiled into monthly composites using the raster [128] and `gimms` package [129]. See figure 2.4 for a representation of one such global monthly composite. Chunk 8 contains the `R`-code used to process the data.

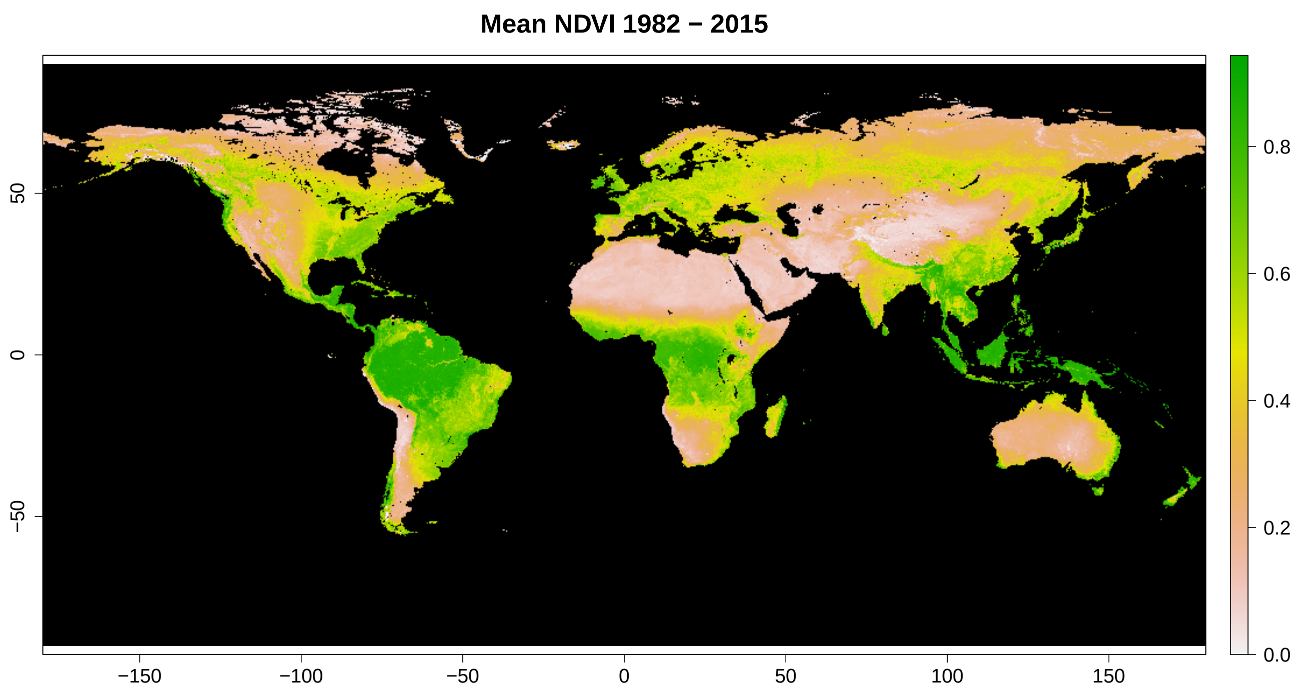


Figure 2.4: Global Representation of the Normalised Difference Vegetation Index (NDVI) - Mean of monthly maximum composite representations of NDVI from January 1982 to December 2015 at global coverage. Higher values of NDVI indicate higher vegetation coverage, plant biomass, and primary production. Figure established via Chunk 14.

2.2.1.2 European Centre for Medium-range Weather Forecasts ReAnalysis 5 (ERA5)

Extrinsic memory effects have largely been understood through processes of drought and water limitation [5,51,53,55,130]. Hence a large number of contemporary studies of vegetation memory have included some form of water-availability proxy, through precipitation records [7,54], or drought indices [3,5,57]. Additionally, temperature has been found to be a major driver of drought severity, plant morphology, and vegetation memory effects [7,54,56] particularly in colder regions where temperature acts as a limiting factor. ERA5 data is implemented into my analyses of vegetation memory using two environmental characteristics: (1) Soil Moisture (in four layers), and (2) Air Temperature (at a height of 2m above the ground). See table 2.2 for an overview of the core characteristics of the ERA5 data set. See Chunk 4, Chunk 5, and Chunk 6 for the codes used to download ERA5 data from the European Centre for Medium-range Weather Forecasts (ECMW) servers, aggregate data to full time series and fix gridding mismatches to be comparable to GIMMS data.

Table 2.2: Core Information about ERA5 data - Characteristics of the ERA5 data set.

Characteristic	Data
Resolution	$\sim 30km \times 30km$
Time Series	Hourly intervals from January 1950 to TODAY
Source	https://apps.ecmwf.int/data-catalogues/era5

Air Temperature recorded in K

Temperature indices have been linked successfully to vegetation sensitivity [7], tree-ring growth [56,131], global primary production [110], as well as severe drought events with possibly devastating consequences to local vegetation [12]. ERA5 includes several temperature variables (e.g. soil, snow and air temperature) [65]. Within this study, I use Air Temperature (at 2m above ground) (Tair) as contained within the ERA5 data set, due to the demonstrated impact of Tair on different aspects of plant physiology and plant morphology which may manifest in vegetation memory effects. See figure 2.5 for an overview of the ERA5 Tair data.

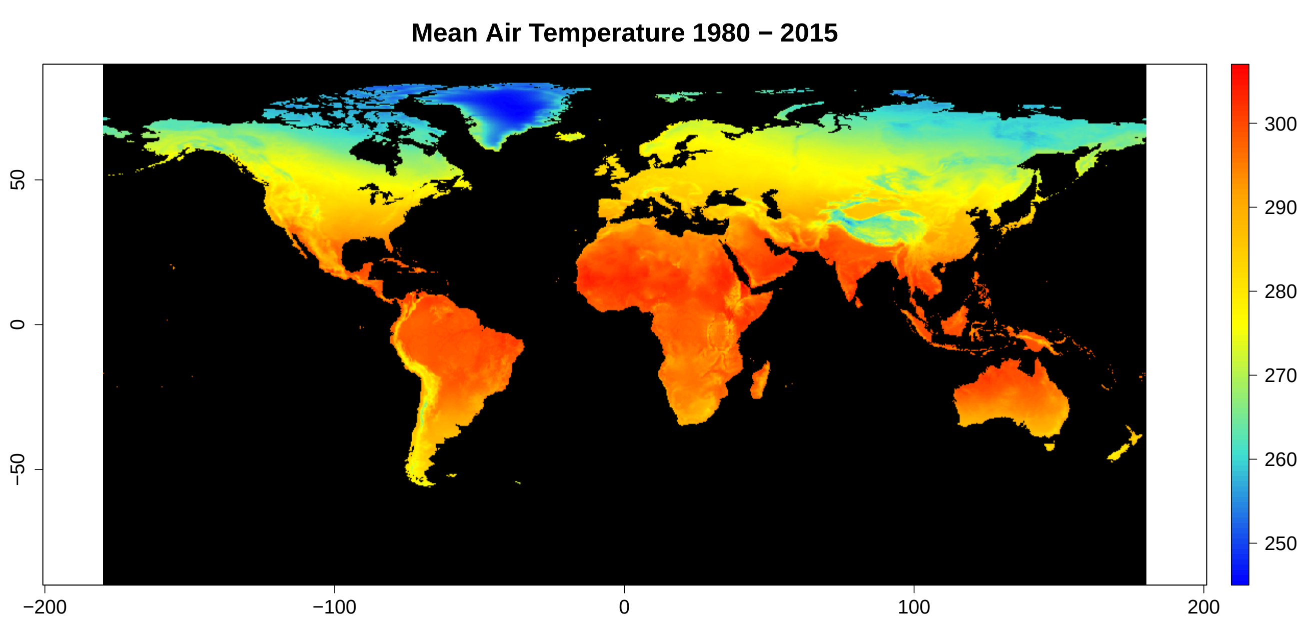


Figure 2.5: Overview of air temperature data - Air temperature data presented as obtainable via the ERA5 reanalysis product (Tair). Figure established via Chunk 16.

Soil Moisture (Qsoil) recorded in $\frac{m^3_{H_2O}}{m^3_{soil}}$

My study uses *Soil Moisture (Qsoil)* as a proxy of local water regimes. I suggest that water-dependent vegetation memory largely relies on water availability to root systems. Precipitation events are subject to further soil processes such as pore connectivity for precipitated water to be available to plant roots [132]. Soil moisture should thus serve as a much more direct proxy of local water regimes. Additionally, ERA5 includes four distinct layers in the soil for the calculation of Qsoil indices (see figure 2.6 for an overview): (1) Soil Moisture (0-7cm) (Qsoil1), (2) Soil Moisture (7-28cm) (Qsoil2), (3) Soil Moisture (28-100cm) (Qsoil3), and (4) Soil Moisture (100-255cm) (Qsoil4). Typical drought indices (i.e. SPEI) do not allow for this additional distinction. Within ERA5, unfrozen ground water (θ) across all four soil layers (k) is defined as:

$$\overline{\theta} = \sum_{k=1}^{4}(R_k * max[f_{liq;k}\theta_k, \theta_{pwp}]) \tag{2.2}$$

with R_k being the root fraction of soil layer k which is a fixed metric according to Land-Cover Classification Systems (LCCSs), and the statement $max[f_{liq;k}\theta_k, \theta_{pwp}]$ calculating the amount of unfrozen soil water in soil layer k. $f_{liq;k}$ is a parametrised function of soil temperature; θ_{pwp} denotes the permanent wilting point according to soil texture. For a more in-depth explanation of how Qsoil is calculated within ERA5, see the IFS Documentation CY45R1 Chapter 4 *Physical Processes* [65].

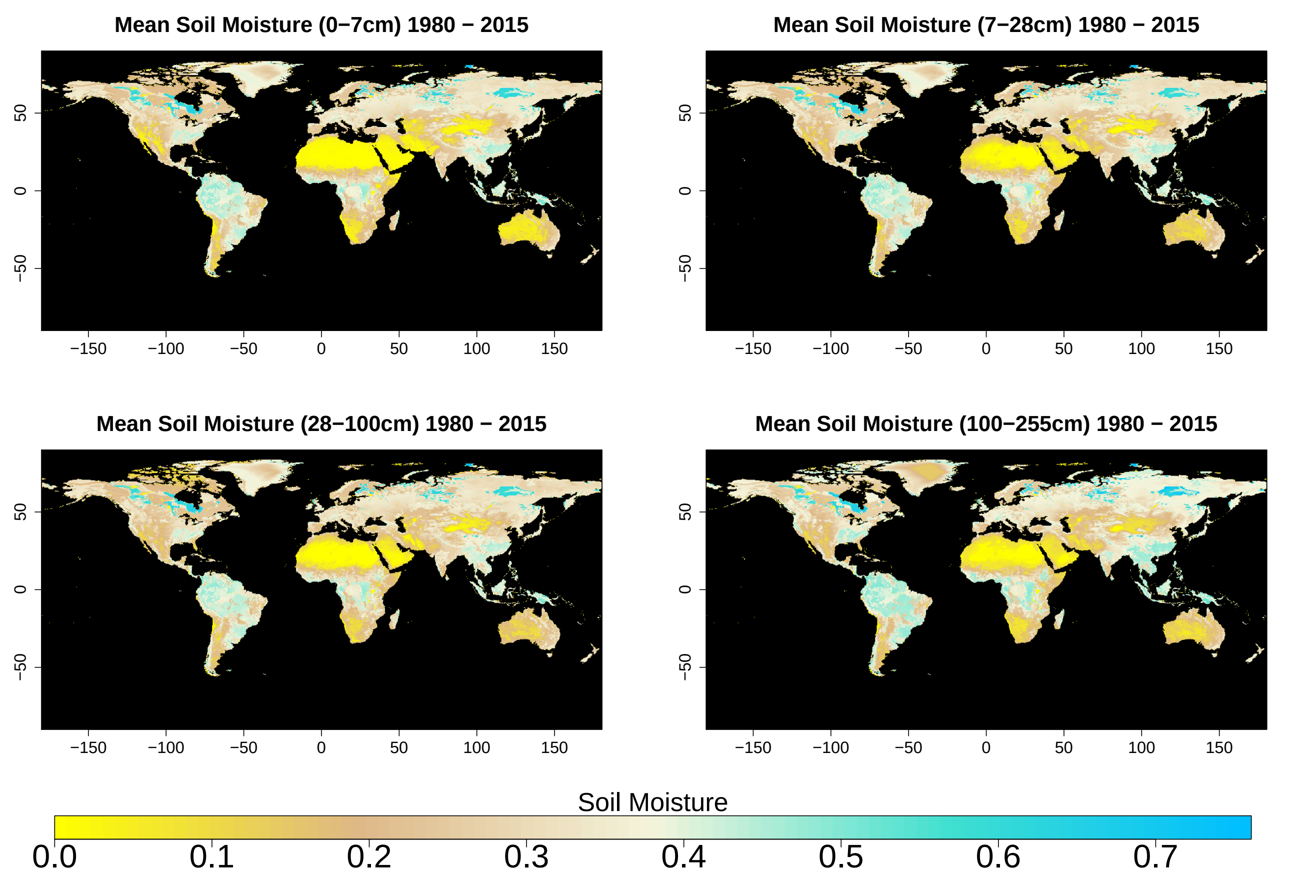


Figure 2.6: Overview of soil moisture data - Soil moisture data presented in layers as obtainable via the ERA5 reanalysis product (Tair, Qsoil1, Qsoil2, Qsoil3, Qsoil4). Figure established via Chunk 15.

Due to the focus on dryland regions, the **main focus of my work is on Qsoil effects**. Following the vegetation sensitivity index by Seddon et al. [7], these regions can easily be postulated to be heavily influenced by water availability regimes and less so by Tair regimes.

2.2.1.3 Digital Elevation Models

Digital Elevation Model (DEM) data preparation is outlined in section 2.2.2.1. To this end, I am using DEM data obtained via the Harmonized World Soil Database (HWSD) which contains different DEM outputs at 3 and 30 arc-second resolution [133,134]. See table 2.3 for an overview of HWSD DEM data characteristics. For an overview of all variables from the HWSD used within this study see table A.2. See figure 2.7 for a global representation of HWSD DEM data and figure A.2 and A.3 for an overview of HWSD slope aspect and incline data respectively as described in table A.2. These haven been rescaled to match GIMMS resolution (see table 2.1).

Table 2.3: Core Information about HWSD data - Characteristics of the HWSD data set.

Characteristic	Data
Resolution	$3arc \times 3arc \sim 31m \times 31m$
Source	http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/ harmonized-world-soil-database-v12/en/

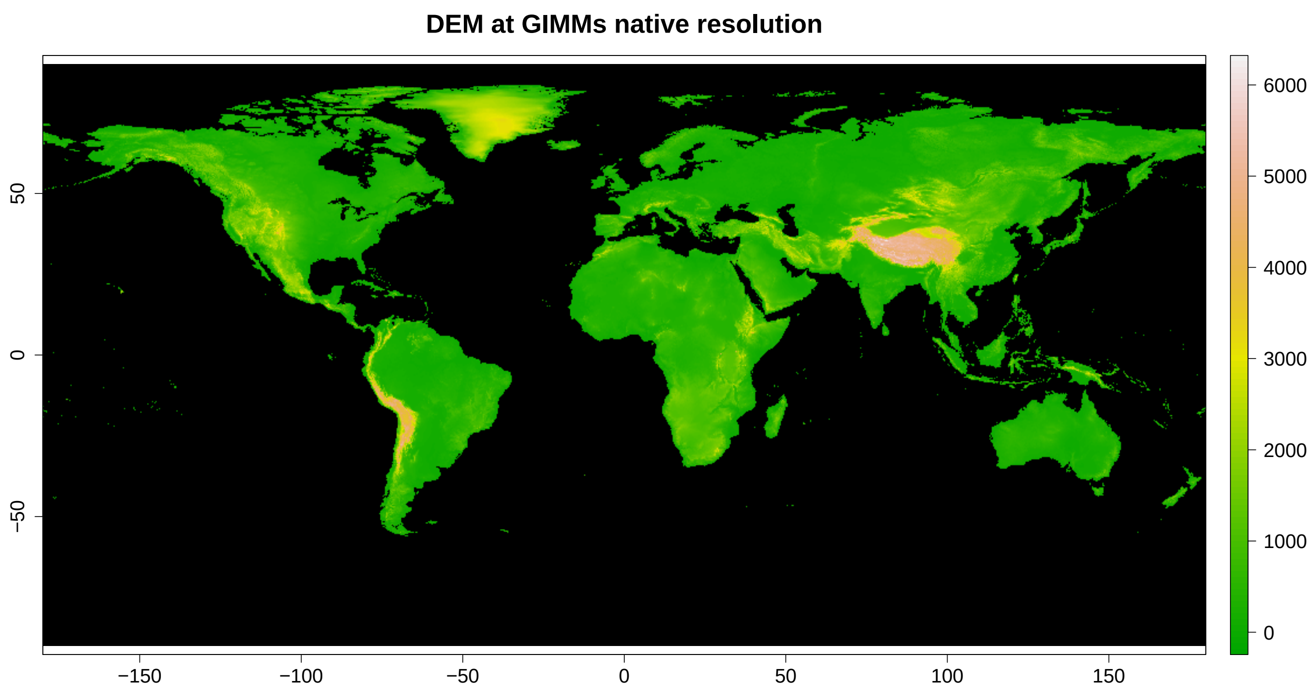


Figure 2.7: DEM elevation data - HWSD elevation data at GIMMS resolution. Figure established via Chunk 17.

2.2.2 Analyses

2.2.2.1 Data Preparation

Monthly Maximum Composites

After retrieving GIMMS NDVI 3g data using the `gimms` package in `R`, bi-weekly NDVI data is aggregated to monthly composites using the *maximum composite* method (see Chunk 8). Doing so is a method for removing data uncertainty due to atmospheric interference which has seen application in recent remote-sensing-based studies of vegetation memory and resilience[5,6,130].

Downscaling

In order to calculate vegetation memory effects (**first research goal**), vegetation data and abiotic data should be represented at the same spatial and temporal resolution. Whilst both NDVI and ERA5 data are available at monthly intervals (after pre-processing as already described), their spatial resolutions differ (e.g. $\sim 9.27km\times \sim 9.27km$, vs. $30km \times 30km$, respectively; see tables 2.1 and 2.2).

There are two ways of remedying this spatial mismatch: (1) aggregating GIMMS data to the coarser resolution of ERA5 data, or (2) downscaling of ERA5 data to GIMMS resolution. Whilst aggregating data to coarser resolutions is much easier and less computationally expensive, I have opted to apply **downscaling** to the ERA5 data used within my study so as to retain valuable information within the GIMMS NDVI 3g data set. This will allow more precise vegetation memory identification and result in more biologically relevant patterns of vegetation memory characteristics.

A large host of downscaling methods are available and before choosing any method one has to settle on whether to perform *statistical* or *dynamic* downscaling[60,135]:

1. *Statistical Downscaling* makes use of statistical links between global and local climate patterns[136], comparatively computationally cheap and easily transferable to different study regions, but suffers from a strong dependency on the choice of predictors/co-variates[137].

2. *Dynamic Downscaling* is built around regional climate models being implemented into global climate models[136] (in this case, dynamical models with boundary conditions obtained via ERA5), produces results based on physically consistent processes, but it is computationally expensive[137].

This study is focussed on three study regions (i.e. the Iberian Region, the Caatinga, and Australia) and five ERA5 variables (Tair, Qsoil1, Qsoil2, Qsoil3, and Qsoil4) in monthly intervals from 1981 to 2015. ERA5 data was downscaled to match GIMMS resolution to calculate cumulative lags of memory effect drivers (see section 2.2.2.2). With this, downscaling needs to be performed three times (study regions), for five variables, each with 420 individual time steps (twelve months per year for 35 years). In total, the downscaling process is thus required to be run 6,300 times. Using a computationally expensive method such as dynamical downscaling is thus undesirable and I have elected to employ a **statistical downscaling** method to match ERA5 data with GIMMS resolution.

More specifically, I use **Kriging** - a method that is well-understood and has long been used in non-biological sciences for geostatistical interpolation purposes[138]. Kriging is a two-step process. First, one establishes statistical relationships between data which is to be kriged at its native resolution with covariate data at the same resolution. The second step sees the extrapolation of these relationships using covariate data at the target resolution. The way in which Kriging marks an improvement over other statistical downscaling methods lies in the fact that the Kriging methodology not only *extrapolates* relationships but *residuals* as well (see figure 2.8 for a visual representation).

Therefore, Kriging within my analyses requires three inputs: (1) ERA5 data at native resolution, (2) HWSD covariates (see table A.2) at ERA5 resolution, and (3) HWSD covariates (see table A.2) at GIMMS resolution to produce sets of ERA5 data at GIMMS resolution. Kriging operations are built around formulae which establish response and predictor relationships. My kriging approaches are specified as follows:

$$Var_{ERA5} = \alpha + \sum_{i=1}^{14} (\beta_i * Cov_{HWSD;i}) \tag{2.3}$$

with Var_{ERA5} identifying any of the five ERA5 variables Tair, Qsoil1, Qsoil2, Qsoil3, or Qsoil4 at any given monthly interval between January 1981 and December 2015. $Cov_{HWSD;i}$ indexes the i^{th} HWSD covariate ranging from elevation and slope aspects to slope incline levels. Due to computational expense the kriging procedure only considers interaction effects between slope incline levels and slope aspects (not shown in equation 2.3, see Chunk 9).

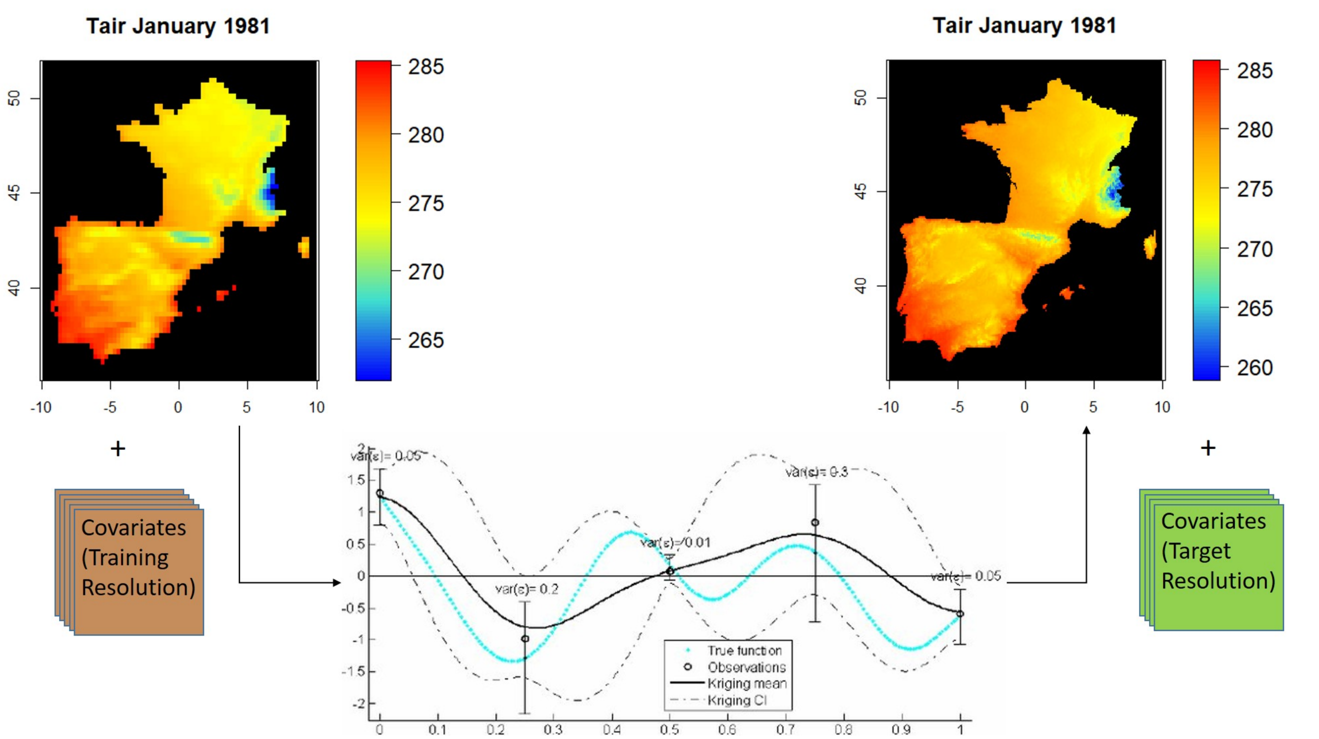


Figure 2.8: Kriging Concept - Statistical downscaling effects of Tair for the time step of January 1981 using HWSD covariate data across the Iberian Region. Diagram source: Le Riche et al., 2012 [139]. Some parts of this figure have been generated via Chunk 35.

For an in-depth mathematical explanation of the Kriging methodology, see Hengl, 2011 [138]. A practical example of its use can be retrieved in Lichtenstern, 2013 [140].

2.2.2.2 Memory Models

1. *Vegetation Response Coefficients* are used to identify intrinsic/extrinsic forcing factors answering *research question I.1.*
2. *Model Comparisons* and *Variance Partitioning* are used to assess the relative importance of different factors answering *research question I.1* and *I.2*.

Vegetation-Response Coefficients

The vegetation-memory analyses in this study have been informed heavily by DeKeersmaecker et al [3]. Vegetation memory models of this study are built upon the following basic specification:

$$NDVI_t = \beta_{t-1} * NDVI_{[t-1]} + \beta_{Qsoil} * Qsoil_{k;m} + \beta_{Tair} * Tair_t \tag{2.4}$$

with $NDVI_t$ and the Autoregressive NDVI Coefficient ($NDVI_{[t-1]}$) being standardised NDVI anomalies at time step t and $t-1$, respectively; $Qsoil_{k;m}$ denoting Qsoil data at depth level k (translating to Qsoil1- Qsoil4) and cumulative lag of standardised anomalies of lag m, and $Tair_t$ denoting Tair data at time t. See table 2.4 for an interpretation of the model coefficients. The duality of memory effects as *intrinsic* and *extrinsic* as proposed by Ogle et al. [8] is embraced as follows:

1. **Intrinsic Memory** is identified as coefficients of $NDVI_{[t-1]}$ [3,5] (see model formula 2.4).
2. **Extrinsic Memory** are implemented via ERA5 data. See section 2.2.1.2 for the rationale behind the inclusion of the different variables and model formula 2.4 for how they are included in vegetation memory models.
   (a) *Tair* is implemented as an *instantaneous* effect on plant performance as I am prioritising Qsoil-driven vegetation-memory effects.
   (b) *Qsoil1- Qsoil4* effects are implemented as *cumulative lag* effects ranging from instantaneous impacts to lags on annual time windows [6].

Table 2.4: Interpretation of Memory Model Coefficients - Biological Interpretation of Vegetation Response Coefficients as portrayed in model formula 2.4.

Coefficient	Magnitude	Sign
$\boldsymbol{\beta_{t-1}}$	Absolute values depict the speed at which systems return to equilibrium/pre-disturbance state. Large absolute values are assumed to represent low resilience (i.e. slow return) [3].	*Positive* - NDVI anomalies resemble previous ones. NDVI anomalies gradually diminish over time. *Negative* - NDVI anomalies resemble previous ones, but with the opposite sign. The return to pre-disturbance is characterised through oscillations.
$\boldsymbol{\beta_{Qsoil}}$	Absolute values depict the resistance to anomalies in Qsoil. Large absolute values indicate low resistance (i.e. strong vegetation responses) to Qsoil anomalies.	*Positive* - Wetter soil conditions than average induce positive NDVI anomalies; drier soil conditions than average induce negative NDVI anomalies. *Negative* - Drier soil conditions than average induce positive NDVI anomalies; wetter soil conditions than average induce negative NDVI anomalies.
$\boldsymbol{\beta_{Tair}}$	Absolute values depict the resistance to anomalies in air temperature. Large absolute values indicate low resistance (i.e. strong vegetation responses) to air temperature anomalies.	*Positive* - Warmer air temperature than average induces positive NDVI anomalies; colder air temperature than average induces negative NDVI anomalies. *Negative* - Colder air temperature than average induces positive NDVI anomalies; warmer air temperature than average induces negative NDVI anomalies.

Additionally, $\beta_{\mathbf{t-1}}$ is indicative of *intrinsic vegetation memory*[3].

The coefficients of the above drivers of vegetation memory (see table 2.4 and formula 2.4) are identified for each data pixel in the data rasters of the three study regions in four separate models (one for each Qsoil layer). A visual representation of the automated modelling approach used within this study can be seen in figure 2.9. This approach is carried out for each pixel as follows using the code contained within Chunk 10:

1. Data for each variable (NDVI, Tair, Qsoil) is **extracted and detrended** using linear detrending with the pracma package in 'R' to avoid effects of changing abiotic conditions over long time-series[3].

2. Detrended data is **standardised to Z-Scores** to obtain deviations of monthly means/monthly anomalies for each variable[3]:
$$Anomaly_i = \frac{Detrended_i - \overline{Detrended_{month}}}{SD_{Detrended,month}} \tag{2.5}$$
with i indexing individual, detrended data records. In the case of NDVI, one additionally calculates monthly means of untreated NDVI data across the entire data range. See figure 2.10 for an exemplary overview of NDVI data treatment.

3. Calculation of **lagged effects**:
   (a) $NDVI_{[t-1]}$ is calculated from Z-Score NDVI data.
   (b) Cumulative lags of Qsoil data are established for lags ranging from 0 (instantaneous effects) to annual effects (aggregated over twelve months of detrended Qsoil records) in steps on one month at a time. These lagged effects are subsequently standardised to Z-Scores. See figure 2.11 for an exemplary overview of Qsoil data treatment.
   (c) Tair data is implemented as instantaneous effects and so no lagged effects have to be calculated. Technically, this means that Tair-driven vegetation responses aren't memory effects in the sense of antecedent conditions influencing present ones. I will refer to these effects as memory effects nonetheless for the sake of brevity.

4. Variables in nature are often collinear[86,141]. Failing to address this issue results in masking of information which might significantly influence our understanding of processes in nature[142]. One method of circumventing this issue lies with *PCA*. More specifically, if regression modelling is the target, one may wish to employ **PCA regression** as a three-step process[142] for each of the cumulative Qsoil lags across all four Qsoil layers, effectively adding the necessity for a model selection step:
   (a) Z-Score data for NDVI, Tair, and Qsoil are fed to PCA via the vegan package in R. See figure 2.12 for an exemplary overview of data and a PCA result.
   (b) Regression is performed as follows:
$$NDVI_t = \beta_1 * PC_1 + \beta_2 * PC2 + \beta_3 * PC3 \tag{2.6}$$
   with $NDVI_t$ representing NDVI anomalies which have been set to NA (skipped in models) in months for which $Thresholds_i < 0.1$ with $Thresholds_i = \overline{Raw_{NDVI,month}}$[5]. PC_1 through PC_3 and β_1 through β_3 indicate principal components 1 through 3 and coefficients of their effects in the model respectively.
   (c) Model selection is performed to identify the cumulative soil lag which presents the most explanatory power through comparison of model Akaike Information Criterion (AIC) values. The model with the lowest AIC is determined to be performing the best[143]. This results in a proxy of Qsoil memory length of local vegetation.
   (d) The regression coefficients β_1 through β_3 can be back-transformed to represent PCA input variable effects (see formula 2.4) using PCA loadings and PCA model coefficients (see figure 2.13).

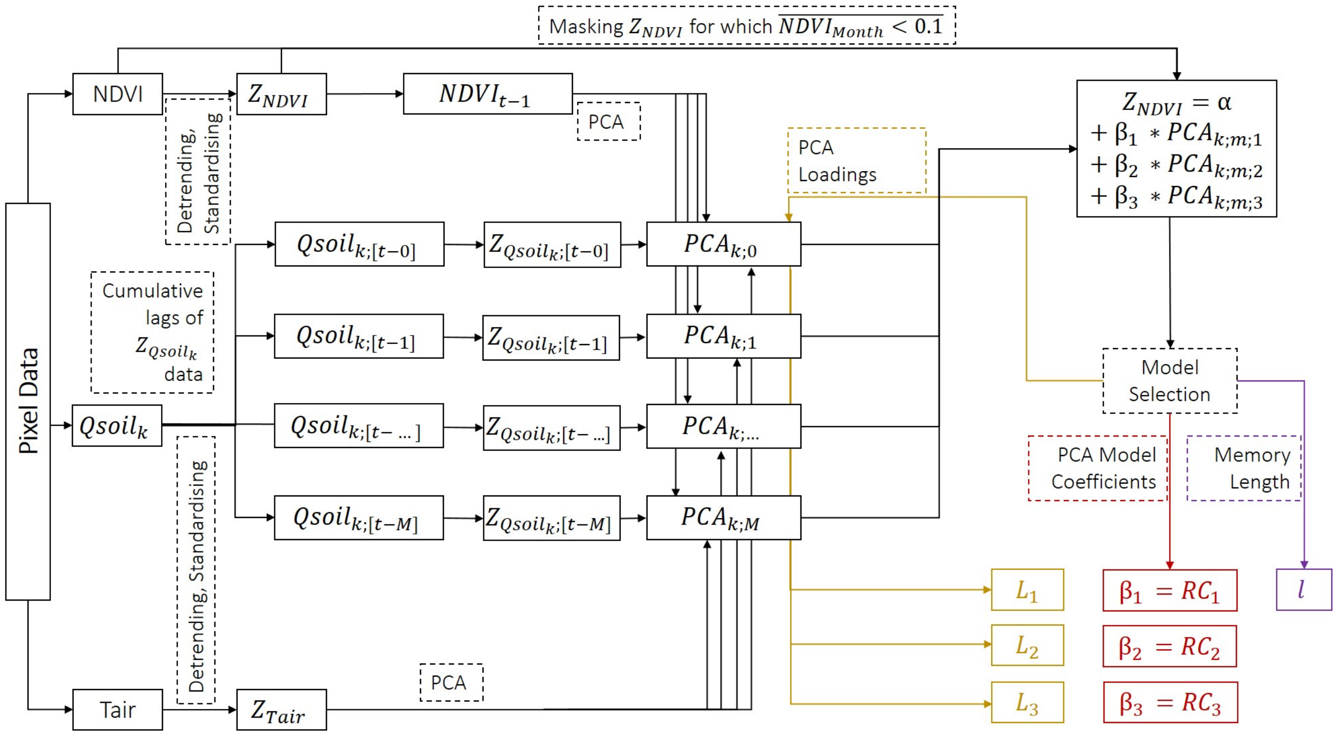
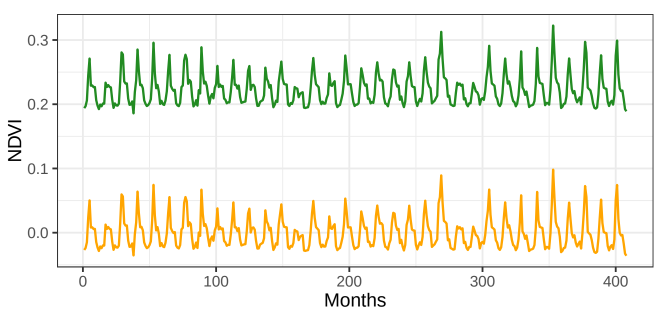
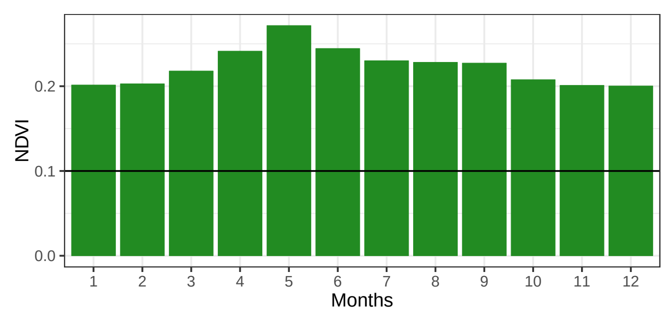


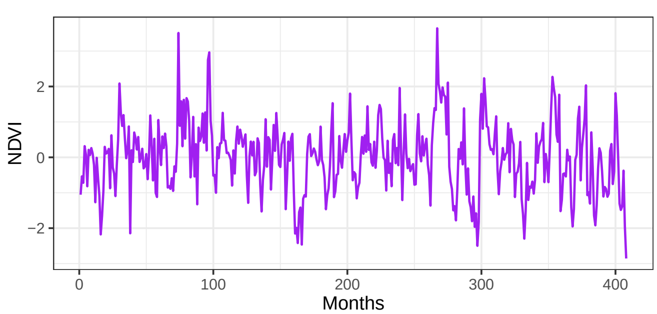
Figure 2.9: Vegetation Memory Model Flowchart - Visual workflow of pixel-wise iterated vegetation memory model. m denotes the currently considered cumulative lag of Qsoil data with M being the maximum cumulative lag. Qsoil layers Qsoil1- Qsoil4 are identified via k. L_1 through L_3 are the loadings of each detrended and standardised variable (NDVI, Tair, and Qsoil) onto the principal components 1 through 3, respectively. PCA model coefficients are identified as β_1 through β_3. l denotes the cumulative Qsoil lag offering the most explanatory power in terms of AIC values of PCA regression models and is thus a proxy for vegetation memory in terms of Qsoil.



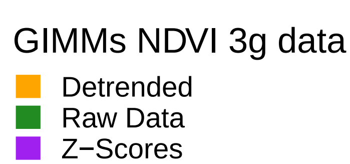
(a) NDVI Time-Series (Untreated and Detrended)



(b) Monthly Means of Untreated NDVI Data with 0.1 Threshold

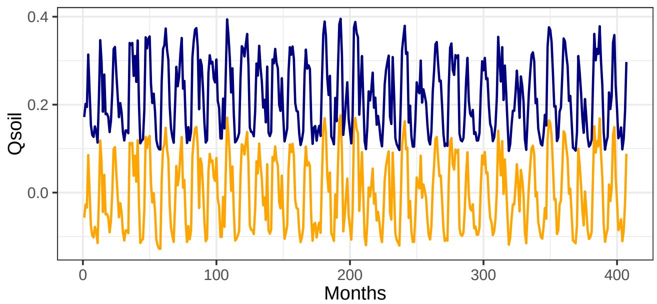


(c) Standardised Detrended NDVI Data (Z-Scores)

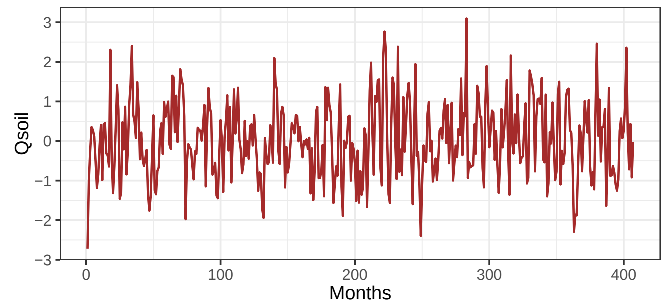


(d) Plotting Colour Reference

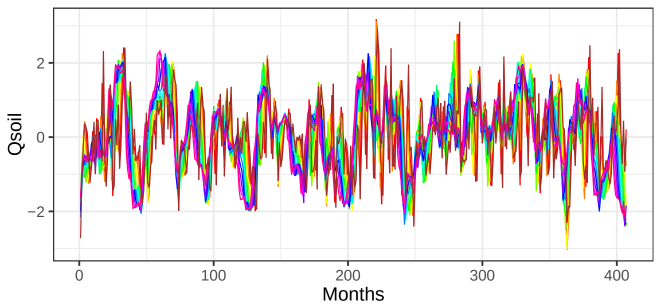
Figure 2.10: Memory Model NDVI Data Treatment - Overview of NDVI data treatment as outlined in figure 2.9 and explained above. The data presented here represents a single pixel in the Iberian dryland region (see figure 2.12). Figure established via Chunk 22.



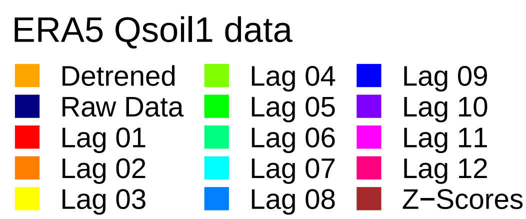
(a) Qsoil1 Time-Series (Untreated and Detrended)



(b) Standardised Detrended Qsoil1 Data (Z-Scores)



(c) Cummulative Qsoil1 Lags



(d) Plotting Colour Reference

Figure 2.11: Memory Model Qsoil Data Treatment - Overview of Qsoil data treatment as outlined in figure 2.9 and explained above. Only Qsoil1 data is represented. The data presented here represents a single pixel in the Iberian dryland region (see figure 2.12). Figure established via Chunk 23.

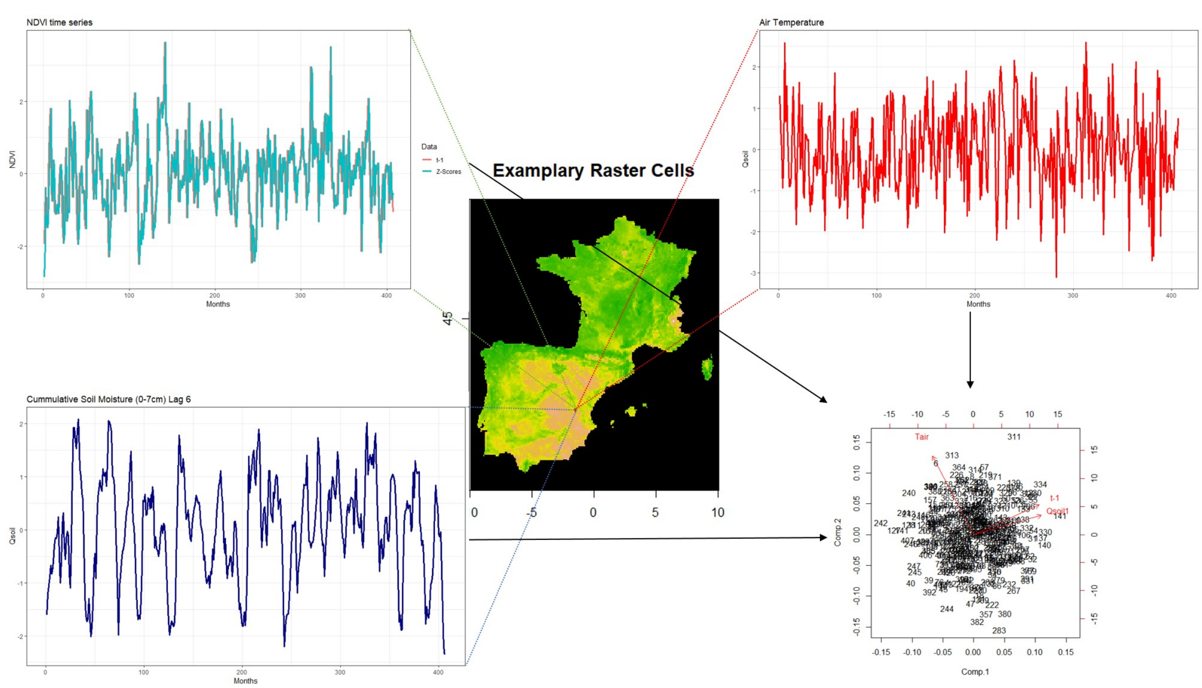


Figure 2.12: exemplary Vegetation Memory PCA - Overview of Z-Score data for $NDVI_{[t-1]}$, $Qsoil_{1;6}$, and $Tair$ for a single pixel in the dryland region of the Iberian study region (red colouring on NDVI background in central plot) as well as their representation in a PCA. Also pictured: $NDVI_t$ Z-Scores required for PCA regression models (see figure 2.9 and 2.13). Some parts of this figure have been generated via Chunk 35.

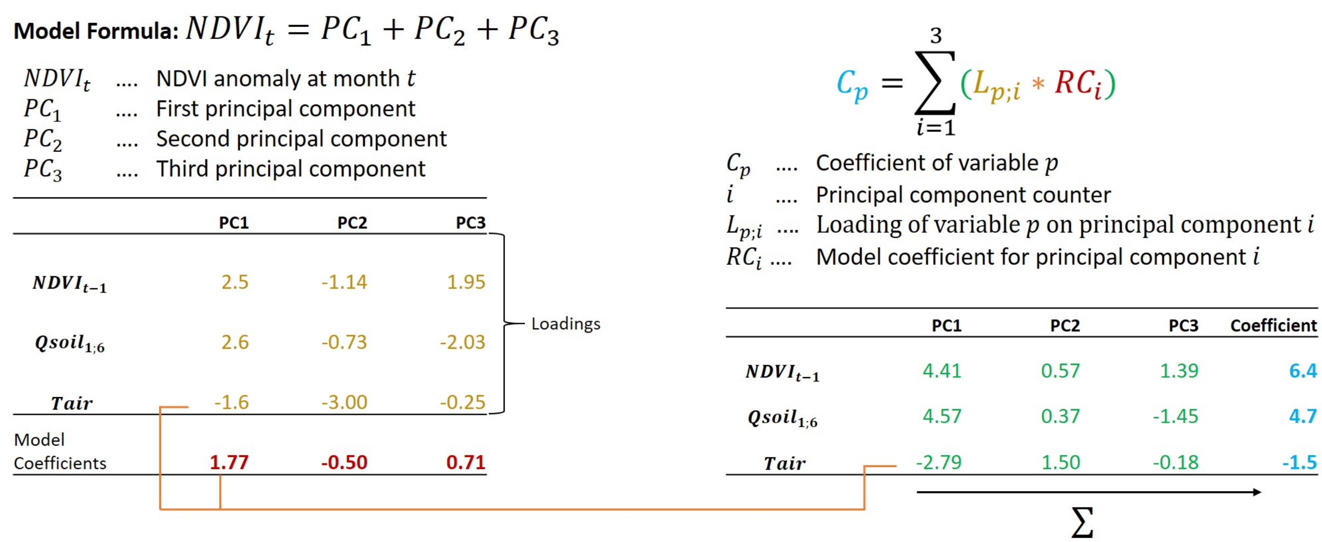


Figure 2.13: PCA Regression Coefficients - Theoretical back-calculation of regression coefficients from PCA regression coefficients as lined out by Zuur et al [142]. The data presented here (PCA loadings, PCA model coefficients, and final variable coefficients) represents a single pixel in the Iberian dryland region (see figure 2.12).

By following the vegetation memory modelling procedure outlined in figures 2.9 and 2.13 and explained above for each study region one obtains rasters containing the following information for each pixel:

1. **Vegetation-Response Coefficients** as established in formula 2.4 used to identify meaningful drivers of vegetation memory (*research question I.1*).
2. **Vegetation-Memory Length** in regards to Qsoil1, Qsoil2, Qsoil3, and Qsoil4 again being used to answer *research question I.1*.

Model Comparisons

Assessments of differences in relative importance of vegetation-response coefficients in individual models and between models serve to answer the following questions:

1. **Which model variable exerts the greatest influence on vegetation anomalies?** - To answer this question, I am comparing absolute values of variable coefficients across all pixels *within* each *model* individually, separately for each study region. Doing so serves to answer *research question I.1* and *I.2*.
2. **Which Qsoil layer is the most biologically influential?** - The answer to this question can be retrieved by comparing absolute values of Qsoil coefficients for all pixel per model *between* all four *models* for each study region which answers *research question I.1.* by identifying a robust extrinsic memory metric.

Differences in absolute variable coefficient values within and between models has been assessed using Mann-Whitney-U Test (`wilcox.test(..., paired = FALSE)` in `R`). I have chosen the Mann-Whitney U-Test, since vegetation memory coefficient data cannot be expected to be normal distributed, nor to fall onto symmetrical distributions (hence, one should contrast median values rather than mean values). The code for these assessments can be retrieved in Chunk 10 (`CoeffScaling`).

Although allowing for region-wide generalisations, these assessments of statistical significance can not be used to identify or display spatial patterns of relative memory coefficient importance.

Variance Partitioning

Variance partitioning is a model-driven method of assessing relative importance of vegetation memory model variables. As opposed to Mann-Whitney U model comparisons, variance partitioning can be carried out for each pixel in my study region data rasters individually. This, in turn, results in the identification of patterns of relative model predictor importance.

It was my goal to assess the relative importance of intrinsic and extrinsic vegetation memory components (*research question I.2*). I have identified Qsoil layers to be of special interest in representing extrinsic vegetation memory (see section 3). Therefore, I am assessing the relative information contained in the vegetation memory model predictors $NDVI_{[t-1]}$ (intrinsic memory) and $Qsoil_{k;m}$ (extrinsic soil moisture memory of layer k and cumulative lag m).

Zuur et al. [142] present a method developed by Legendre & Legendre [144] for the purpose of variance partitioning between two explanatory variables. This approach (represented in figure 2.14) is carried out as follows:

1. Apply the full regression model $Z_{NDVI} = NDVI_{t-1} + Qsoil_{k;m}$ and obtain R^2 (the coefficient of determination). This is R^2_{Full} and equal to all explained variance. Unexplained variance can then be calculated as $1 - R^2_{Full}$
2. Obtain R^2 of $Z_{NDVI} = NDVI_{t-1}$. This is $R^2_{NDVI_{t-1}}$ and equal to all variance explained by $NDVI_{[t-1]}$.
3. Obtain R^2 of $Z_{NDVI} = Qsoil_{k;m}$. This is $R^2_{Qsoil_{k;m}}$ and equal to all variance explained by $Qsoil_{k;m}$.

The variance shared between $NDVI_{[t-1]}$ and $Qsoil_{k;m}$ - R^2_{Shared} can then be calculated as:

$$R^2_{Shared} = R^2_{NDVI_{t-1}} + R^2_{Qsoil_{k;m}} - R^2_{Full} \tag{2.7}$$

The pure information contained within $NDVI_{[t-1]}$ (R^2_{t-1}) and $Qsoil_{k;m}$ ($R^2_{k;m}$) can then be calculated as follows:

$$R^2_{t-1} = R^2_{NDVI_{t-1}} - R^2_{Shared} \tag{2.8}$$

$$R^2_{k;m} = R^2_{Qsoil_{k;m}} - R^2_{Shared} \tag{2.9}$$

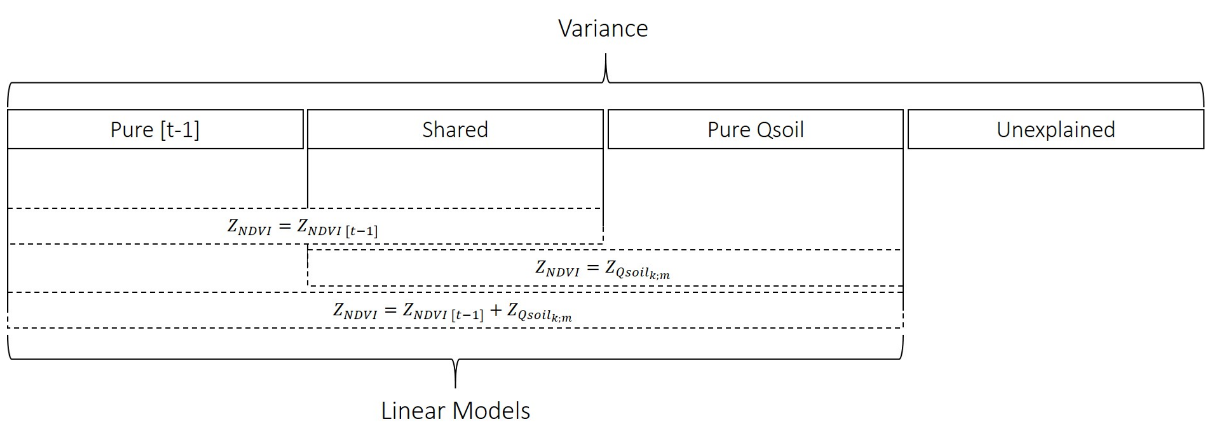


Figure 2.14: Variance Partitioning - Partitioning of variance in pure influence of $NDVI_{[t-1]}$ (Pure $[t-1]$), $Qsoil_{k;m}$ (Pure $Qsoil$, with k identifying the Qsoil layer, and m denoting the cumulative lag of Qsoil data), shared variance between the two (Shared), and residuals (Unexplained). The figure concept has been lifted from Zuur et al. [142] and adjusted to reflect the purpose of this study. Model specifications are contained within dashed boxes.

This form of variance partitioning is carried out for each pixel across all study regions and contained in the vegetation memory model output of Chunk 10 alongside PCA regression models.

2.2.2.3 Vegetation Memory Sensitivity

Following the notion of climate adaptation leading to altered disturbance-responses of local vegetation [51] I am assessing the relationship of vegetation memory characteristics and the mean value of their drivers across the entire time period from 1981 to 2015. This is done using linear regression.

2.3 Functional Aspects of Vegetation Memory

2.3.1 Data

2.3.1.1 COMPADRE

LHT information for plants and animals can be retrieved via COMPADRE; an extensive data base of different species and plot-scale ecosystems at different locations around the Earth [72]. COMPADRE is built around observational data and matrix models [145] which can be used to extract valuable information about temporal processes in biological communities [146,147]. The outputs of COMPADRE matrix models are manifold [73] offering access to a host of biologically relevant proxies of population processes.

For this study, I have selected the following COMPADRE outputs as the most likely to be related to vegetation memory characteristics alongside hypotheses of proposed relationships to vegetation memory:

1. The **Fast-Slow Continuum (FSC)**.
   (a) *FSC-1* - Species of fast life histories exert shorter/weaker vegetation memory.
   (b) *FSC-2* - Species of low reproductive output show longer/stronger vegetation memory.
2. **Reactivity** - Species of higher reactivity exert shorter/weaker vegetation memory.
3. ρ - Species of higher ρ exert shorter/weaker vegetation memory.
4. π - Species of higher π show shorter/weaker vegetation memory.

Data for almost all variables contained within COMPADRE can be retrieved for all locations contained within the current release of the data base. Sampling effort for LHTs within the COMPADRE scheme can only cover a limited range of geological locations (hence I am not presenting a global data overview). The best sampled for regions on Earth - in terms of COMPADRE sites - are the Iberian Peninsula and the contiguous United States of America. To date, no mapping approach beyond the rasterising of plot-level data of COMPADRE LHTs has been proposed.

2.3.1.2 TRY

Here, I test the relationships between vegetation memory effects and two different PFTs which have been selected because they represent two important plant characteristics/trade-off axes in plant function (1) Vegetative Height, and (2) Leaf nitrogen content per leaf dry mass [83].

Additionally, these two PFTs correspond to two of the three PFT domains (i.e. stems and leaves) layed out in Westoby & Wright's PFT framework [84]. Examining PFTs of multiple aspects of plant function in the face of adverse events is especially important as plant performance levels during and after disturbances often incur responses of all aspects of individual plant functional domains [81].

TRY (and other PFT databases) suffer from the same limitations as COMPADRE- sampling bias and effort [99]. Certain regions of the earth are well-sampled for PFTs (e.g. the Pyrenees) although these sampling schemes are often limited to plot-level sampling campaigns. This data limitation highlights the potential importance of a mapping approach as outlined in figure 1.5. The TRY data used for my thesis has been obtained via TRY on 07/08/2018.

2.3.1.3 Floral Data

Extrapolating PFT records from species-referenced records to spatial products using species-specific PFT means and species occurrence records as layed out by Ordonoez & Svenning [82] (see figure 1.5) may prove to overcome the data limitations of TRY PFT data.

Plant species occurrence records can be obtained via floral data repositories such as GBIF (see figure 2.15 for a representation of GBIF plant occurrence data). Data in GBIF are stored as geo-referenced records of species presence. Generating maps from geo-referenced data points can be achieved by aggregating data points to rasters of a desired extent and resolution (i.e. GIMMS resolution).

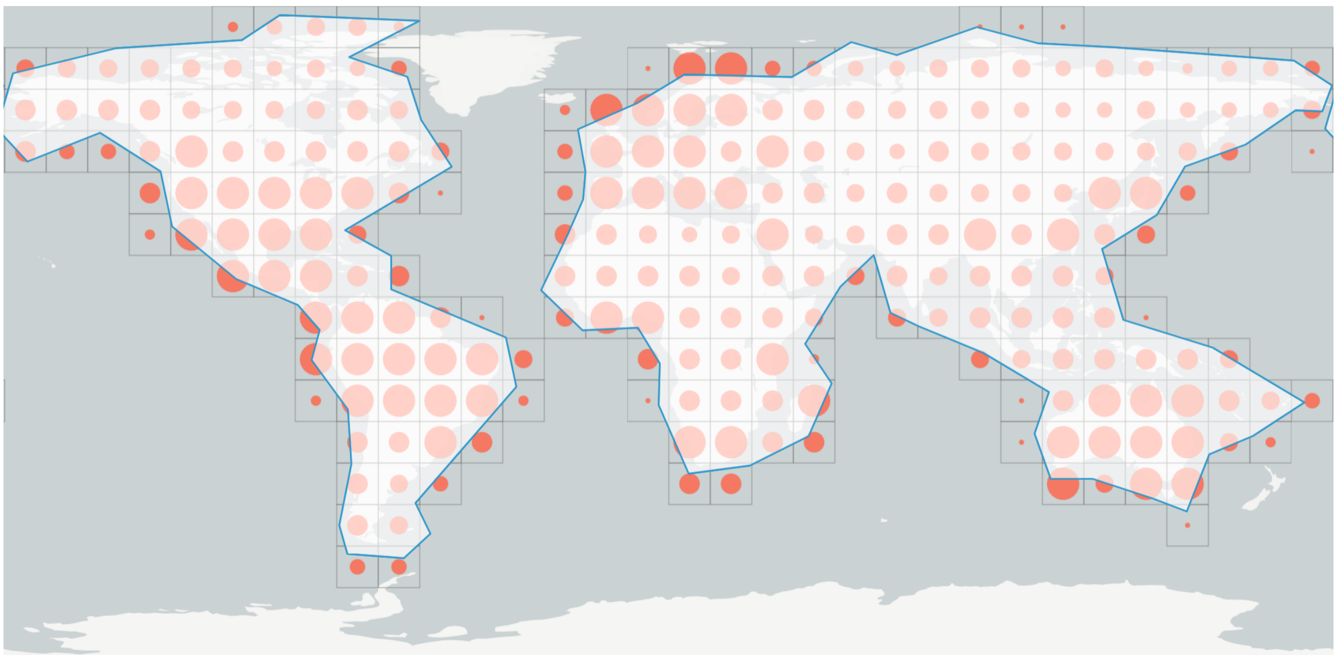


Figure 2.15: GBIF *Plantae* Occurrence Overview - Global representation of occurrence records for species within the *plantae* kingdom available via GBIF [98] as of 21/04/19. Figure generated using the GBIF occurrence data exploration tool [148] obtaining records from 1982 to 2015 (same as GIMMS data availability). Polygons have been drawn to omit occurrence records of marine plantae species. Larger red plotting symbols indicate greater amounts of geo-referenced occurrence records.

The floral data used within this study have been obtained for the time period of 1982 to 2015 (the same time span as the data availability for GIMMS NDVI data) via GBIF on 19/04/19 through to 21/04/19.

2.4 Analyses

2.4.1 Life History Traits

LHT data has been obtained via COMPADRE on 28/01/19 in geo-referenced table format and is used to achieve *research goal 2*.

Data Extraction

Target COMPADRE LHTs are extracted using Chunk 12 by aggregating geo-referenced LHT records to rasters of GIMMS resolution using a mean function for each study region.

Models

I use regression models to assess causal links between LHTs expressions.

2.4.2 Plant Functional Traits

PFT data has been obtained via TRY on 07/08/18 in geo-referenced data table format and is used to achieve *research goal 2*.

Data Extraction

Vegetation-memory characteristics are stored as raster data sets to retain spatial patterns of response coefficients. Therefore, PFT data needs rasterising for comparability. PFT data is extracted and rasterised in two ways:

1. **Raw geo-referenced PFT records**. Geo-referenced TRY data points are aggregated to rasters of GIMMS resolution using a mean function (if multiple PFT records fall onto the same raster cell, the mean value of these is assigned to the cell). This results in rasters of low data-coverage but peer-reviewed data records.

2. **Species-specific mean PFT records**. These are extracted for each species within the TRY data set to enable PFT mapping as depicted in figure 1.5.

Data extraction of PFT records is handled via Chunk 11.

Data Mapping

Data mapping is carried out via Chunk 11 and extrapolates species-specific PFT mean values according to GBIF occurrence data (see figure 1.5). Using the approach presented by Ordonez & Svenning[82], I am establishing rasters of mean PFT expressions by:

1. Assigning species-specific mean PFT values to all cells of individual species occurrence according to GBIF.

2. Computing the mean PFT value for each cell in the rasters.

3. Removing all cells whose values exceed the upper 95% quantile of mean PFT records in these final mean rasters to remove outliers which may be due to sampling bias. The lower 5% quantile remains unaltered to retain grassland regions which are typical of dryland regions and characterised by low Vegetative Height (H) records.

This results in rasters of high data coverage but at a loss of data reliability.

Models

Like with LHTs, I use regression models to assess causal links between PFT expressions (both geo-referenced records directly obtained via TRY and mean-extrapolated as presented in figure 1.5).

3. Results

3.1 Identifying Vegetation Memory

3.1.1 Vegetation-Memory Models

Across all three study regions, I have identified Qsoil1 to be the most informative Qsoil layer (see figures A.13, A.17, and A.21 and tables A.6, A.10, and A.14). Therefore, I am only presenting Qsoil1 results here. Qsoil2, Qsoil3, and Qsoil4 results can be found in section A.4.1. Additionally, patterns of $NDVI[t-1]$ and Tair coefficients do not change (quantitatively or qualitatively) between different soil layer models and are thus reported only in this results section.

3.1.1.1 Iberian Region

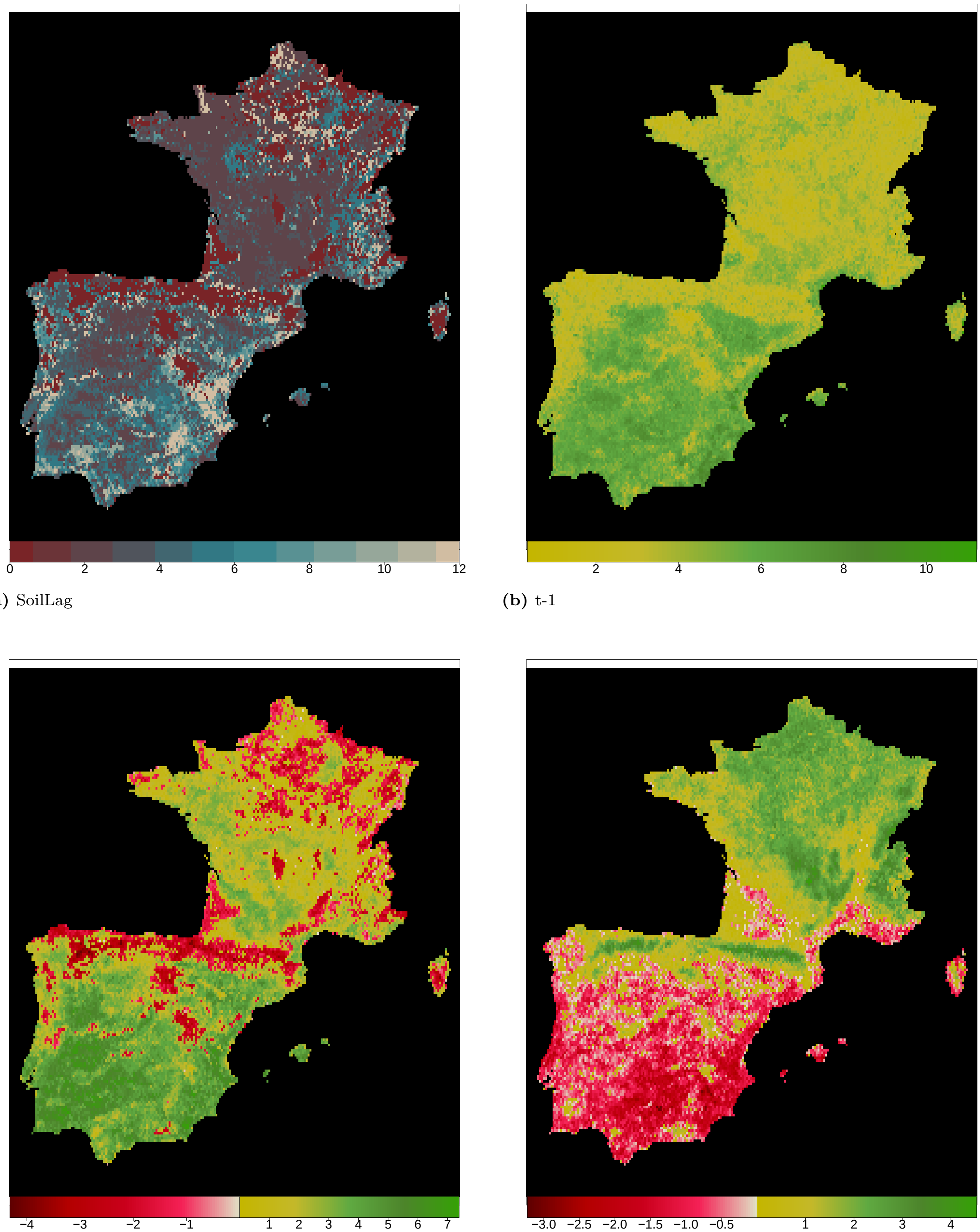
Across the Iberian Region, my analyses have identified the following vegetation-memory patterns:

1. Qsoil1 memory length (identified as cumulative lags according to AIC values, figure 3.1a) exhibits low values (short memory) in the region of the Pyrenees with a noticeable shift towards slightly longer memory lengths when moving to dryland regions. Particularly across Spain, this manifests in a north-west to south-east gradient of increasing Qsoil1 memory length.
2. $NDVI_{[t-1]}$ memory follows the same pattern as presented by De Keersmaecker et al. [3] with strong, positive memory effects in the south of the Iberian region which diminish but remain positive towards the north (figure 3.1b). Across Spain, the $NDVI_{[t-1]}$ pattern is a close match of the Qsoil1 memory-length pattern with longest Qsoil1 memory being associated with stronger $NDVI_{[t-1]}$ memory effects and vice versa.
3. Qsoil1 memory (figure 3.1c) is predominantly positive in sign meaning that vegetation will react positively (increased NDVI values) to positive anomalies in Qsoil1 conditions (wetter conditions). A notable exception of this are the Pyrenees. Especially strong Qsoil1 memory can be observed in the dryland regions of Spain and Portugal. Qsoil1 memory patterns closely resemble $NDVI_{[t-1]}$ memory patterns.
4. Tair vegetation responses falls onto a clear latitudinal gradient with positive memory effects in the north to negative memory effects of almost equal absolute values in the south (figure 3.1d). Additionally, there seems to be an altitudinal gradient with negative memory effects in the Pyrenees.

$NDVI_{[t-1]}$ memory is stronger than Qsoil1 memory which, in turn, is stronger than Tair vegetation responses as indicated by the coefficient sizes (see table 3.1).

Table 3.1: Mann-Whitney U-Test (Iberian Region, Qsoil1 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil1	Tair
NDVI t-1	3.854	NA	2.58e+08	316643639
Qsoil	2.228	0	NA	248234384
Tair	1.192	0	0.00e+00	NA



(a) SoilLag

(b) t-1

(c) Qsoil

(d) Tair

Figure 3.1: Vegetation Response Coefficients (Iberian Region; Qsoil1) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

3.1.1.2 Caatinga

The vegetation-memory patterns I have identified throughout the greater Caatinga region are as follows:

1. Qsoil1 memory length (identified as cumulative lags according to AIC values, figure 3.2a) exhibits lower values (shorter memory) in the north-eastern dryland regions (this is the Caatinga region itself) as well as the south-western sub-humid and humid regions (although not part of the Caatinga itself, these have been included to show possible differences in vegetation behaviour of dryland and non-dryland regions).

2. $NDVI_{[t-1]}$ memory depicts a different pattern as the vegetation sensitivity index developed by Seddon et al[7] with the north-eastern dryland region being characterised by strong intrinsic-memory effects with a marked drop-off towards the south-west (figure 3.2b). The $NDVI_{[t-1]}$ analysis by Seddon et al. identified strong intrinsic vegetation memory to the west of the pattern identified here. Additionally, $NDVI_{[t-1]}$ memory patterns closely resemble an inverse of Qsoil1 memory-length patterns with increased $NDVI_{[t-1]}$ effects in areas of short Qsoil1 memory length.

3. Qsoil1 memory is mostly positive across the entire Caatinga range with the highest absolute values being found in the north-eastern dryland region (figure 3.2c). Take note that the maximum Qsoil1 memory coefficient is larger than the absolute of the corresponding minimum value. This points towards an overall positive effect of wetter conditions on vegetation performance across the Caatinga. Additionally, there are some changes in Qsoil memory effect sign between Qsoil1 and Qsoil2, and Qsoil3 and Qsoil4 with negative Qsoil effects being identified around the region of Brasilia in the latter two layers (see section A.4.1). Overall, the Qsoil1 memory pattern closely resembles that of $NDVI_{[t-1]}$ effects.

4. Tair vegetation-response patterns (figure 3.2d) follow a pattern that is close to the inverse of the Qsoil1 pattern with large negative values in the north-eastern dryland region. Take note that the absolute of the minimum Tair vegetation-response coefficient is larger than the corresponding maximum value identifying a predominantly negative effect of increased Tair on vegetation in this region. Tair vegetation-response patterns are a close match to an inverse Qsoil1 memory pattern.

Again, region-wide assessments of mean vegetation response coefficients has revealed that $NDVI_{[t-1]}$ memory is stronger than Qsoil1 memory which, in turn, is stronger than Tair vegetation responses (see table 3.2).

Table 3.2: Mann-Whitney U-Test (Caatinga, Qsoil1 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil1	Tair
NDVI t-1	4.206	NA	882317448	1.119e+09
Qsoil	3.089	0	NA	9.731e+08
Tair	1.264	0	0	NA

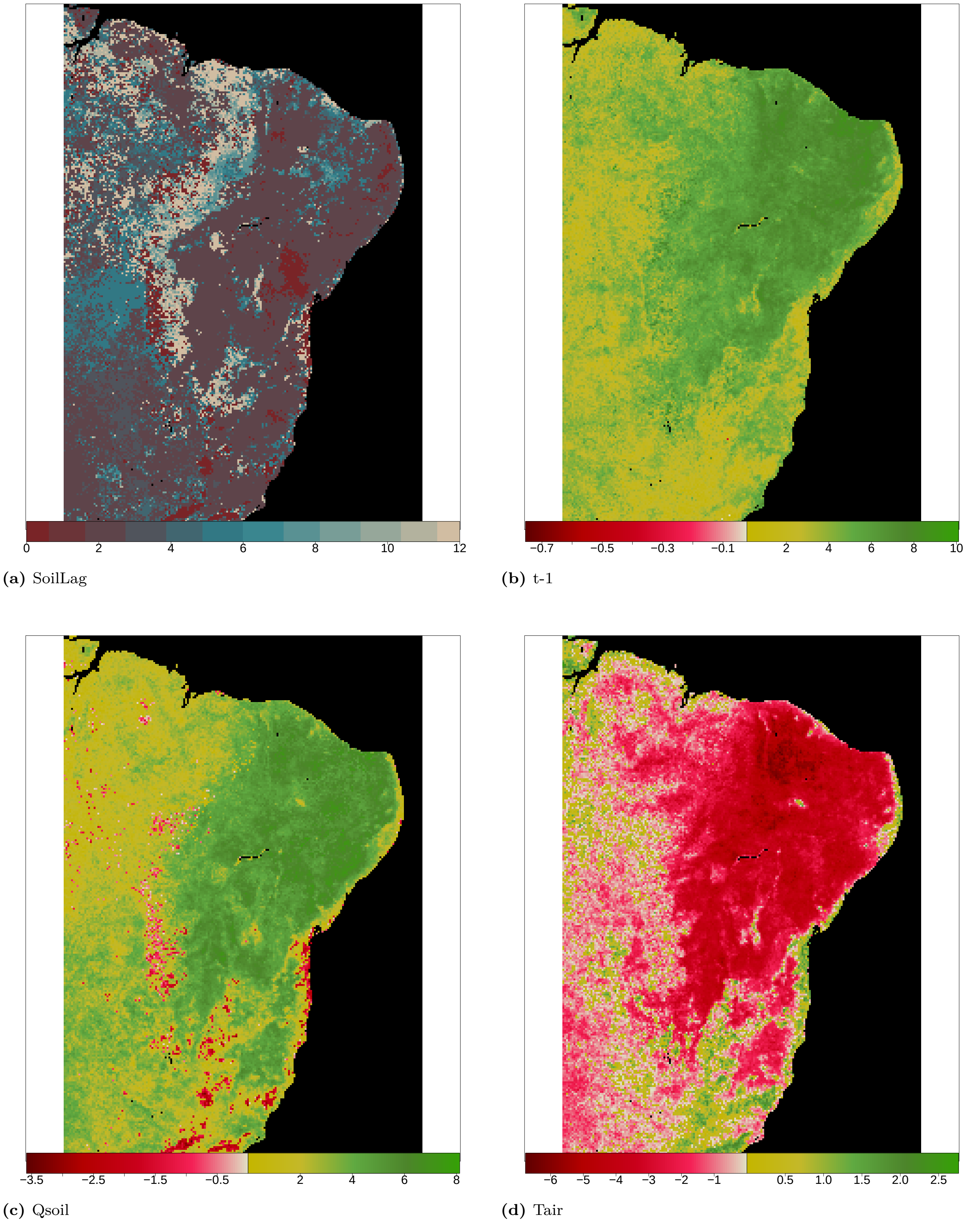


Figure 3.2: Vegetation Response Coefficients (Caatinga; Qsoil1) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

3.1.1.3 Australia

Vegetation-memory patterns across Australia have been identified as follows:

1. Qsoil1 memory length (identified as cumulative lags according to AIC values, figure 3.3a) exhibits larger values (longer memory) in the western regions whilst lower memory length indices in the eastern regions of Australia identify shorter memory. Take note that these patterns do not overlap well with the previously identified memory lags of Vicente-Serrano et al.[6], or Liu et al.[5]. However, this pattern closely mimics that of the autoregressive coefficient ($NDVI_{[t-1]}$) employed by Seddon et al[7].

2. $NDVI_{[t-1]}$ memory follows the same pattern as presented by De Keersmaecker et al.[3] with strong, positive memory effects all across Australia with stronger $NDVI_{[t-1]}$ memory effects in the continental region of Australia and smaller effects across the coastlines (figure 3.3b). Unlike before, there is no obvious match between $NDVI_{[t-1]}$ memory patterns and Qsoil1 memory length.

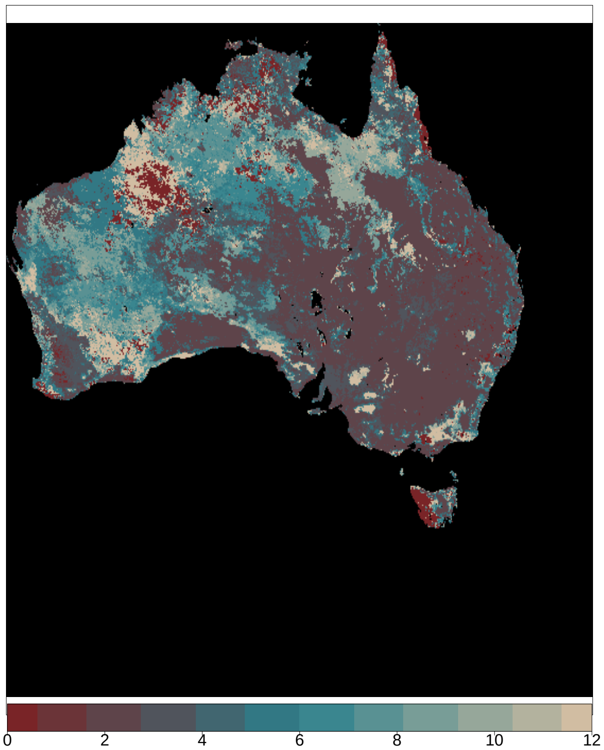
3. Qsoil1 memory effects are mostly positive across the entirety of Australia with the highest values being located in the continental region of Australia and negative Qsoil1 memory effects being found on Tasmania, along coastlines and south of the St. George Ranges (figure 3.3c). Take note that the maximum Qsoil1 memory coefficient is larger than the absolute of the corresponding minimum value. Therefore, positive Qsoil1 anomalies lead to stronger positive NDVI anomalies than negative Qsoil1 anomalies of the same strength. Additionally, there are some changes in Qsoil memory-effect sign between Qsoil1 and Qsoil2, and Qsoil3 and Qsoil4 with negative Qsoil effects being identified throughout the entire continental region of Australia particularly within the Qsoil4 layer (see section A.4.1). Again, the Qsoil1 memory pattern closely resembles that of $NDVI_{[t-1]}$ memory effects.

4. Tair vegetation-response patterns (figure 3.3d) follow a pattern that is close to the inverse of the Qsoil1 pattern with large negative values across all of Australia. Take note that the absolute of the minimum Tair vegetation-response coefficient is larger than the corresponding maximum value.

The previously identified hierarchy of vegetation-response coefficient importance stays true for the Australia: $NDVI_{[t-1]}$ memory is stronger than Qsoil1 memory which, in turn, is stronger than Tair vegetation responses (see table 3.3).

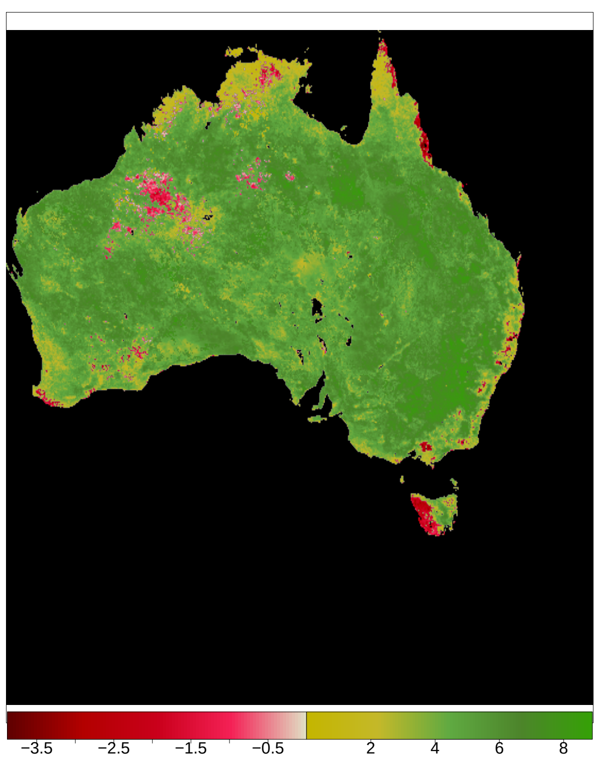
Table 3.3: Mann-Whitney U-Test (Australia, Qsoil1 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil1	Tair
NDVI t-1	7.453	NA	8.424e+09	9.680e+09
Qsoil	5.326	0	NA	8.483e+09
Tair	2.852	0	0.000e+00	NA



(a) SoilLag

(b) t-1



(c) Qsoil

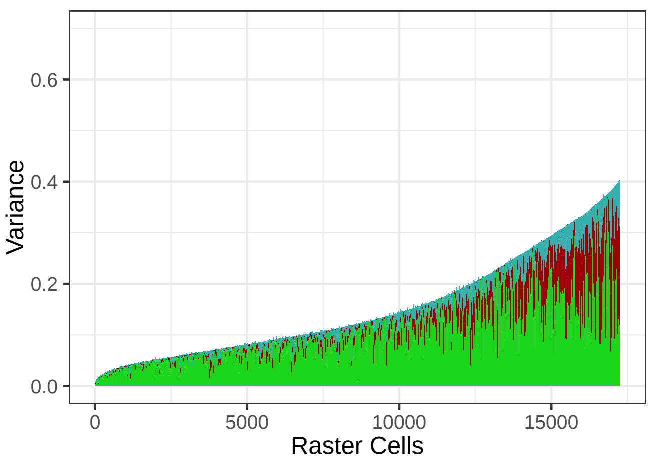
(d) Tair

Figure 3.3: Vegetation Response Coefficients (Australia; Qsoil1) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

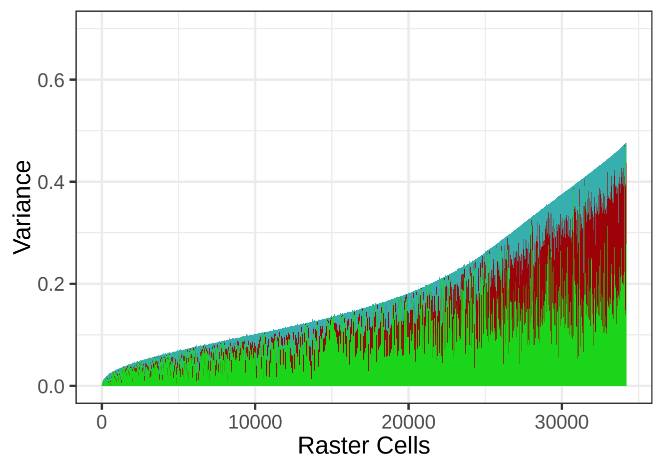
3.1.2 Variance Partitioning

Whilst the above region-wide Mann-Whitney U analyses allow for region-encompassing comparisons of relative importance of model components, variance partitioning allows for far more advanced analyses of relative model-coefficient importance. As mentioned previously, I have elected to analyse only $NDVI_{[t-1]}$ and Qsoil1 variance as these two are representative of the target processes I am assessing across dryland regions and have been identified as the most important vegetation memory metrics across all three study regions (see tables 3.1, 3.2, and 3.3).

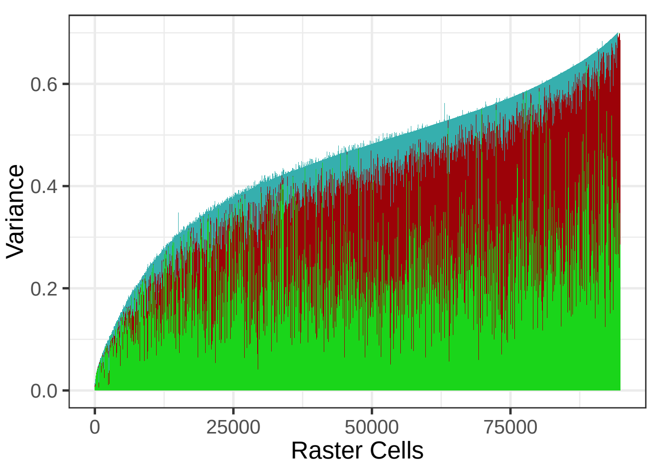
Comparing variance-partitioning results across study regions (see figure 3.4) patterns of different ecological system processes emerge. Whilst Qsoil and $NDVI_{[t-1]}$ share little to no variance across the Iberian Region, a moderate amount of variance is shared by these two variables when analysing the Caatinga range in the areas where the variance explained is highest. Across Australia, however, large amounts of variance are shared by $NDVI_{[t-1]}$ and Qsoil1. Anomalies of Qsoil1 and $NDVI_{[t-1]}$ are the most meaningful predictors of NDVI anomalies across Australia, followed by the Caatinga suggesting different ecological processes to be in effect in these regions. Overall, $NDVI_{[t-1]}$ explains more variation in records of NDVI than Qsoil1. Additionally, $NDVI_{[t-1]}$ and Qsoil1 explain the most variation in NDVI anomalies in dryland regions across all three study regions (see figures A.23, A.25, and A.27).



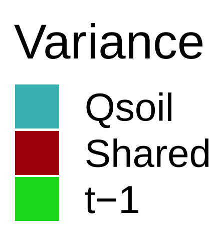
(a) Iberian Region



(b) Caatinga



(c) Australia



(d) Plotting Colour Reference

Figure 3.4: Variance Partitioning Across Study Regions - Variance in dvi explained by $NDVI_{[t-1]}$ and Qsoil1 records across all three study regions. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.

Finally, variance partitioning also highlights the overall importance of Qsoil1 when compared to the variance explained by Qsoil2, Qsoil3, and Qsoil4 (see figures A.22, A.24, and A.26).

3.1.3 Vegetation Memory Sensitivity

3.1.3.1 $NDVI_{[t-1]}$

In general, $NDVI_{[t-1]}$ memory effects decrease as mean NDVI records increase. These relationships (shown in figure 3.5) stay consistent (qualitatively and quantitatively with slight deviations) across all three study regions. The intercepts (shown alongside regression slopes in table 3.4) identify near identical baseline vegetation response coefficients for unvegetated regions (regions where NDVI = 0). The regression intercept across the Iberian Region exhibits a smaller value. Whilst the intercepts themselves can't accurately represent nature since no vegetation memory can be observed when no vegetation is present. However, they can be used to identify different vegetation memory sensitivity across all three study regions with $NDVI_{[t-1]}$ memory being more sensitive towards lower mean NDVI throughout Australia, and the Caatinga than across the Iberian Region.

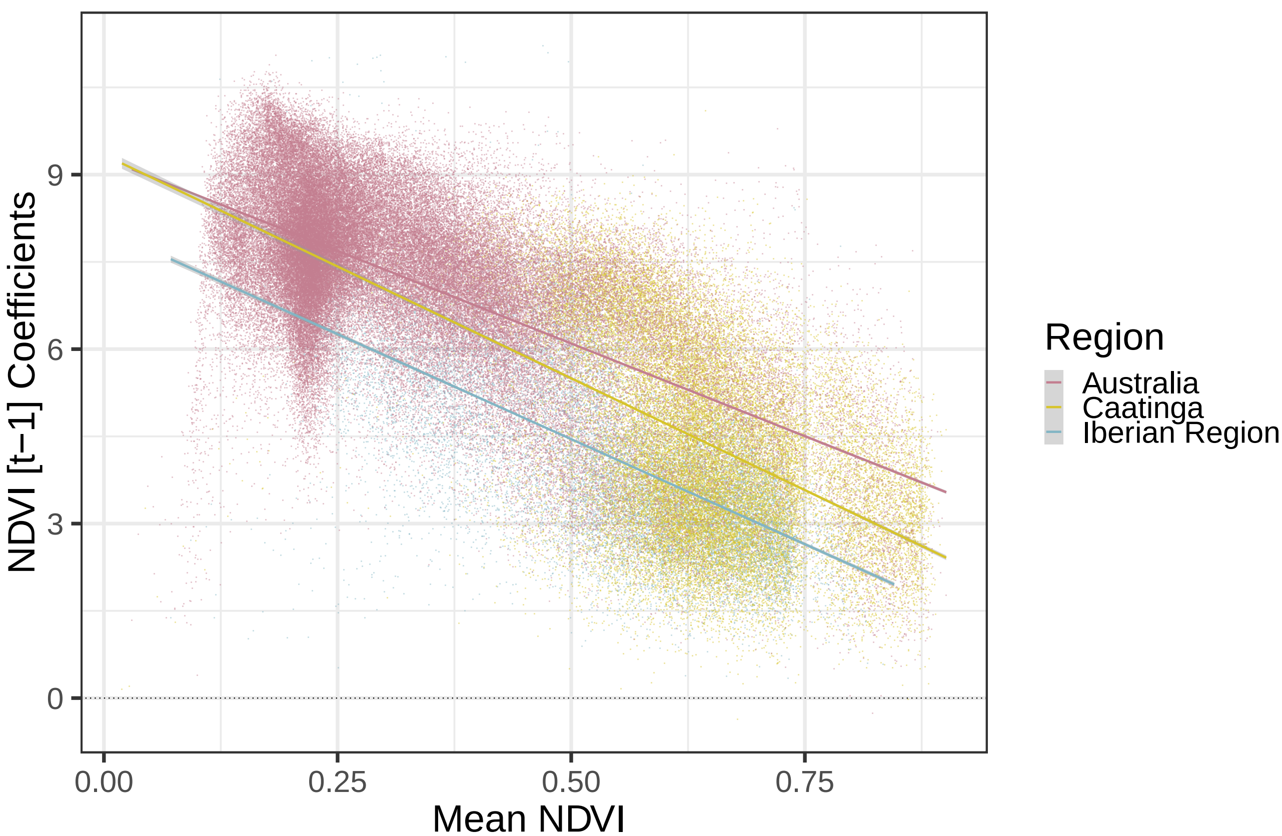


Figure 3.5: Intrinsic Vegetation Memory Sensitivity - $NDVI_{[t-1]}$ vegetation response coefficients of figure 3.1, 3.2, and 3.3 plotted against mean intrinsic drivers (see figure A.4, A.6, and A.8). Figure established via Chunk 29.

Take note that NDVI records are bound between zero and one. Given the intercepts and slopes of the linear regressions, $NDVI_{[t-1]}$ memory effects will not dip below zero in areas of high mean NDVI.

Table 3.4: Vegetation Memory Sensitivity (Mean NDVI and $NDVI_{[t-1]}$ Coefficients) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		Caatinga		Australia	
	I	S	I	S	I	S
Value	8.063	-7.223	9.342	-7.688	9.282	-6.372
p_{Value}	0.000	0.000	0.000	0.000	0.000	0.000

3.1.3.2 Air Temperature

Overall, values of Tair vegetation response effects decrease as mean Tair records increase (see figure 3.6). This relationship is most pronounced across the Iberian Region and least intense across Australia (see table 3.5 for regression coefficients). This may be due to the presence of altitudinal gradients across the Iberian Region which are largely absent from Australia thus limiting the range of mean Tair values. On the other hand, this may be indicative of different patterns of vegetation adaptation to Tair conditions. Take note that most Tair vegetation response effects are negative across the selected dryland study regions.

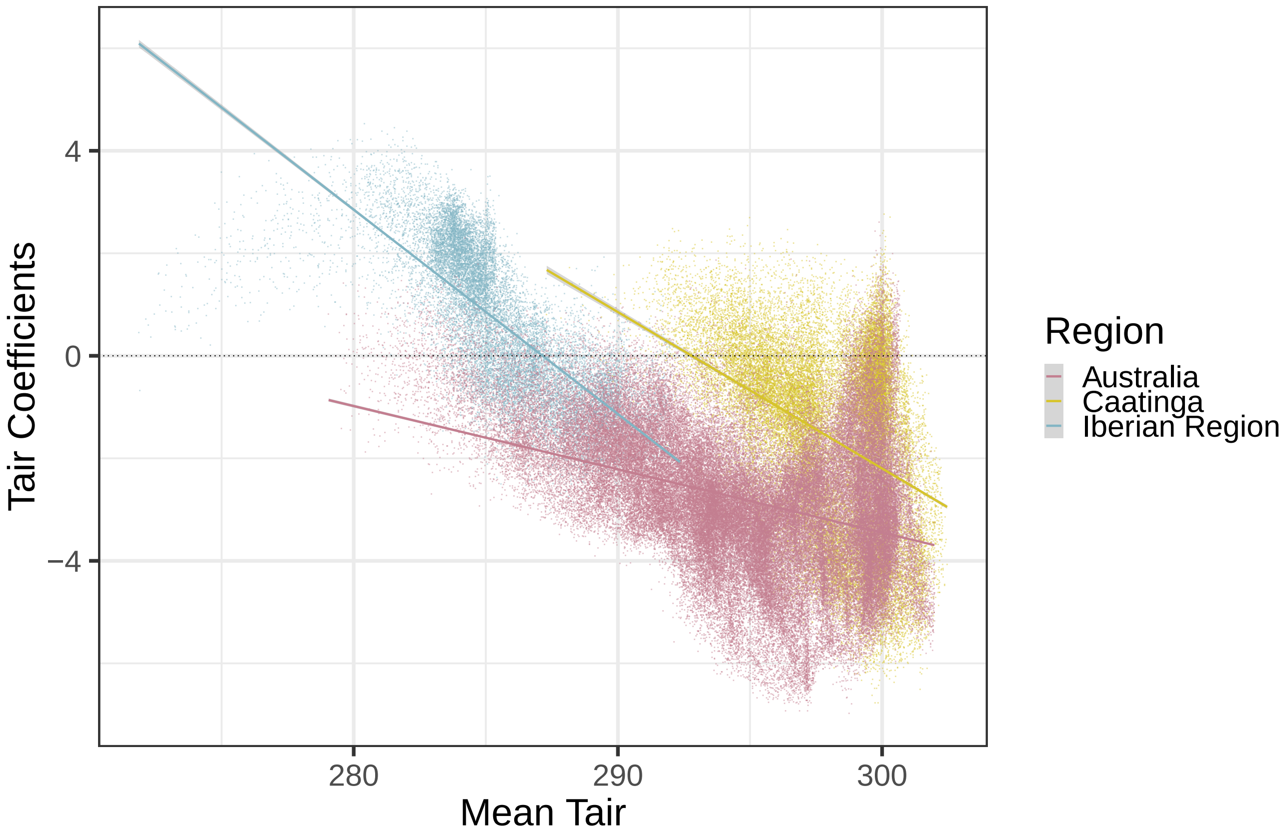


Figure 3.6: Tair Vegetation Memory Sensitivity - Tair vegetation response coefficients of figure 3.1, 3.2, and 3.3 plotted against mean Tair drivers (see figure A.4, A.6, and A.8). Figure established via Chunk 29.

When interpreting the intercepts of these regressions, it is important to take note of Tair records being identified in K instead of $°C$.

Table 3.5: Vegetation Memory Sensitivity (Mean Tair and Tair Coefficients) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		**Caatinga**		**Australia**	
	I	S	I	S	I	S
Value	114.5	-0.3987	89.34	-0.3051	33.59	-0.1234
p_{Value}	0.0	0.0000	0.00	0.0000	0.00	0.0000

3.1.3.3 Soil Moisture (0-7cm)

Whilst values of Qsoil1 memory effect decrease as mean Qsoil1 increases, the rate at which they do differs hugely between study regions (see figure 3.7). Qsoil1 memory effects across Australia seem largely de-coupled from from Qsoil1 conditions. Across the Iberian Region and Caatinga, on the other hand, the negative relationship between Qsoil1 memory effects and mean Qsoil1 records are pronounced with Qsoil1 memory effects dipping into the negative range across the Iberian Region.

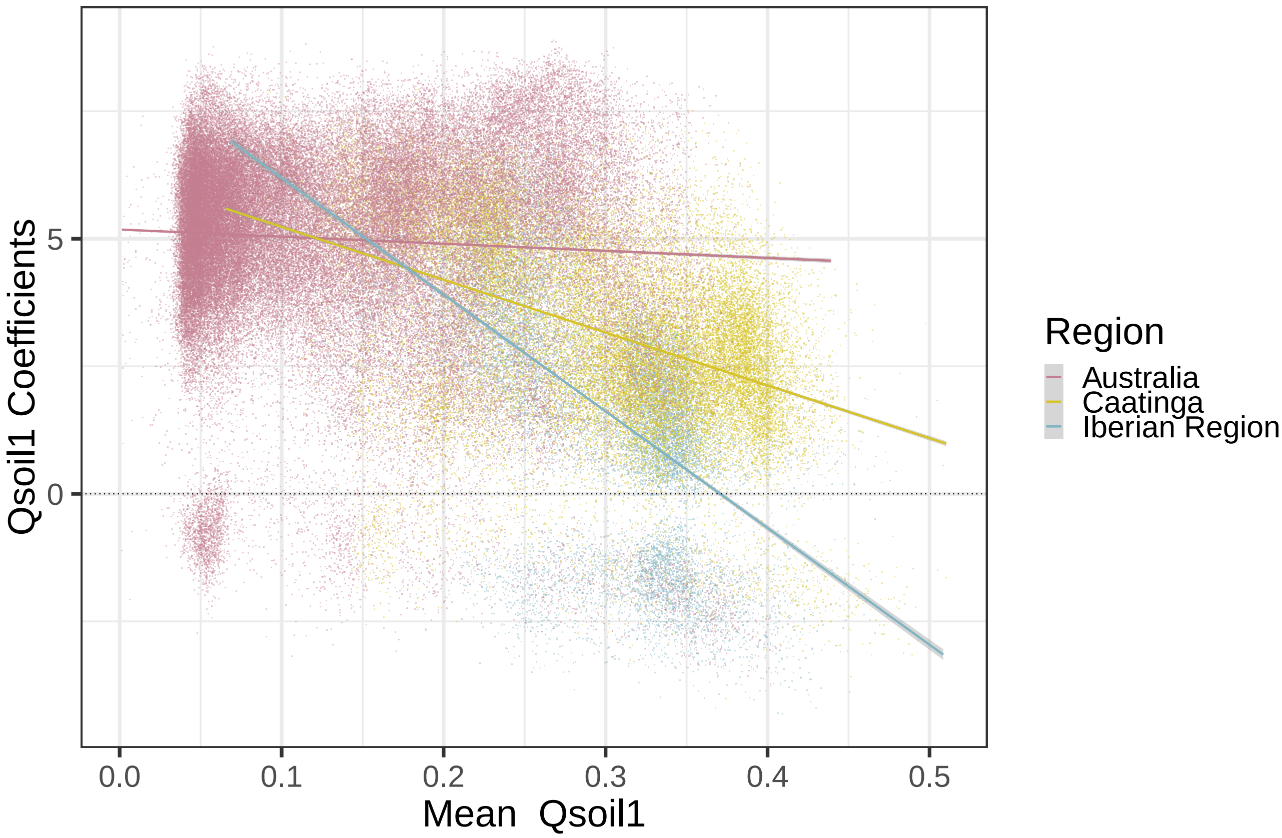


Figure 3.7: Qsoil1 Vegetation Memory Sensitivity - Qsoil1 vegetation response coefficients of figure 3.1, 3.2, and 3.3 plotted against mean Qsoil1 drivers (see figure A.4, A.6, and A.8). Figure established via Chunk 29.

Regression slopes of Qsoil1 memory coefficients and mean Qsoil1 reported in table 3.6 vary widely and support the patterns reported in figure 3.7. Despite the disparity of regressions slopes, the regression intercepts are somewhat similar suggesting a similar baseline Qsoil1 vegetation response in the driest regions.

Table 3.6: Vegetation Memory Sensitivity (Mean Qsoil1 and Qsoil1 Coefficients) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		**Caatinga**		**Australia**	
	I	S	I	S	I	S
Value	8.482	-22.88	6.268	-10.35	5.181	-1.393
p_{Value}	0.000	0.00	0.000	0.00	0.000	0.000

3.1.3.4 Memory Length

Across all three study regions, Qsoil1 vegetation-memory length (assessed via Qsoil1 lags) decreases as mean Qsoil1 increases (see figure 3.8). This effect is most pronounced across the Iberian Region, followed by Australia, and the Caatinga (see regression slopes in table 3.7).

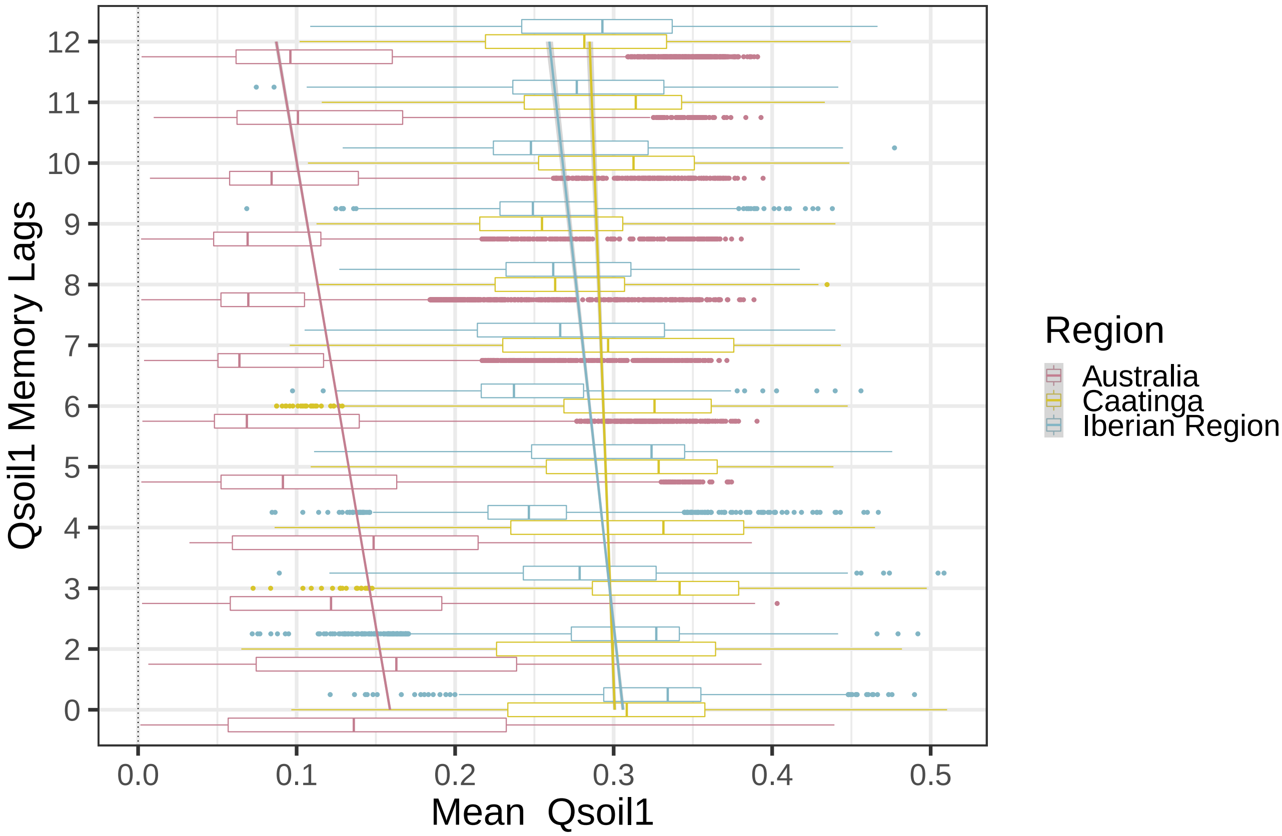


Figure 3.8: Qsoil1 Vegetation Memory Length Sensitivity - Qsoil1 vegetation memory length coefficients of figure 3.1, 3.2, and 3.3 plotted against mean Qsoil1 drivers (see figure A.4, A.6, and A.8). Take note that the above trend-lines are smoothers and not the linear regresseions outlined in table 3.7. Figure established via Chunk 29.

The intercepts of these linear regressions follow the same hierarchy, effectively suggesting a stronger influence of Qsoil1 conditions on memory length across the Iberian Region and Australia when compared to the Caatinga (see table 3.7).

Table 3.7: Vegetation Memory Sensitivity (Mean Qsoil1 and Qsoil1 Memory Lags) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		**Caatinga**		**Australia**	
	I	S	I	S	I	S
Value	7.891	-13.99	4.635	-2.52	6.197	-10.44
p_{Value}	0.000	0.00	0.000	0.00	0.000	0.00

3.1.3.5 $NDVI_{[t-1]}$ and Memory Length

Patterns of $NDVI_{[t-1]}$ and Qsoil1 memory length have been identified to overlap in region-specific manners. Relationships of the two are depicted alongside raw data in figure 3.9. As Qsoil1 memory length increases across the Iberian Region, $NDVI_{[t-1]}$ memory effects increase as well. Across Australia and the Caatinga, the reverse is true and $NDVI_{[t-1]}$ memory effects decrease as Qsoil1 memory length increases (see table 3.8). These relationships are strongest across the Iberian Region and the Caatinga. Although these relationships are significant, the boxplots in figure 3.9 indicate that linear fits might not be the most appropriate.

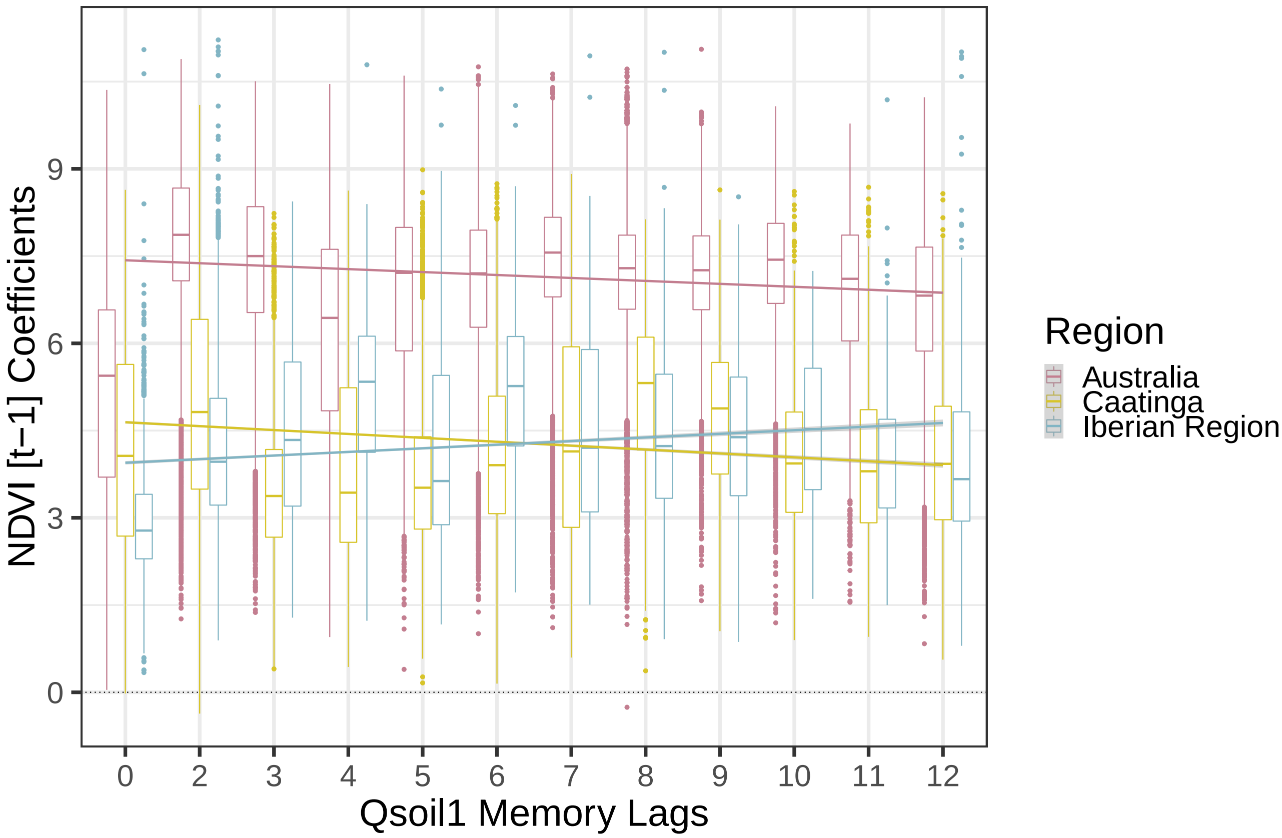


Figure 3.9: Intrinsic Vegetation Memory and Qsoil1 Memory Length - $NDVI_{[t-1]}$ vegetation response coefficients of figure 3.1, 3.2, and 3.3 plotted against Qsoil1 vegetation memory length of the same figures. Figure established via Chunk 29.

The intercepts of the regression analyses identify quickly reacting vegetation (short Qsoil1 memory length) to exhibit strong intrinsic vegetation memory ($NDVI_{[t-1]}$) in Australia with markedly $NDVI_{[t-1]}$ coefficients across the Caatinga, and the Iberian Region.

Table 3.8: Vegetation Memory Sensitivity (Qsoil1 Memory Lags and $NDVI_{[t-1]}$ Coefficients) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		**Caatinga**		**Australia**	
	I	S	I	S	I	S
Value	3.847	0.0748	4.691	-0.0631	7.434	-0.0419
p_{Value}	0.000	0.0000	0.000	0.0000	0.000	0.0000

3.1.3.6 Soil Moisture and Memory Length

When assessing the relationships of Qsoil1 memory coefficients and Qsoil1 memory length qualitatively (see figure 3.10) and quantitatively (see table 3.9), the same patterns as previously described emerge:

- As Qsoil1 memory length increases (i.e. vegetation reacts slower), Qsoil1 model coefficients intensify across the Iberian Region and Australia. This effect is much stronger across the Iberian Region, whereas the effect size across Australia is much smaller and very close to zero.
- Throughout the Caatinga region, the faster vegetation reacts to Qsoil1 anomalies, the more intense the Qsoil1 vegetation memory response.

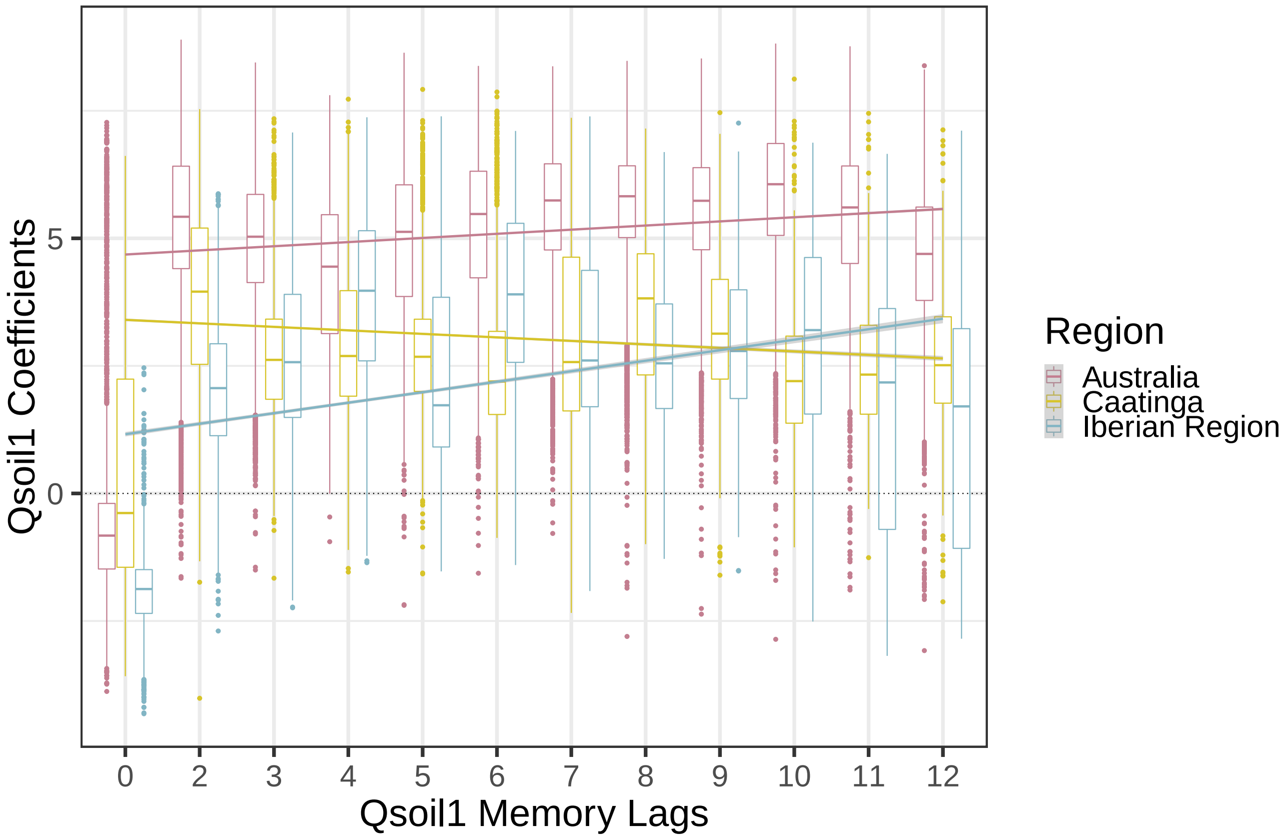


Figure 3.10: Soil Moisture Vegetation Memory and Qsoil1 Memory Length - Qsoil1 vegetation response coefficients of figure 3.1, 3.2, and 3.3 plotted against Qsoil1 vegetation memory length of the same figures. Figure established via Chunk 29.

Vegetation characterised by instantaneous responses to Qsoil1 anomalies (i.e. Qsoil1 memory length = 0) exerts NDVI anomalies close to 0 in response (i.e. Qsoil1 memory coefficients ~ 0) across the Iberian Region. Throughout the remaining two study regions, instantaneous Qsoil1 responses are much stronger, signalling low resistance towards Qsoil1 anomalies.

Table 3.9: Vegetation Memory Sensitivity (Qsoil1 Memory Lags and Qsoil1 Coefficients) - Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as I, the slope of the regression as S. Established via Chunk 30.

	Iberian Region		Caatinga		Australia	
	I	S	I	S	I	S
Value	0.8578	0.2407	3.402	-0.0522	4.525	0.0978
p_{Value}	0.0000	0.0000	0.000	0.0000	0.000	0.0000

3.1.4 Summary

3.1.4.1 Vegetation Memory Models

Across all three study regions, the **Qsoil1** layer has been **identified as the most informative** of all the Qsoil layers.

Some persistent vegetation memory patterns have emerged across all three study regions:

1. $NDVI_{[t-1]}$ **memory** follows similar patterns as identified in previous studies [3,7] with **strong and positive** memory effects **across dryland regions**.
2. **Qsoil1 memory** is **strong and positive across dryland regions**, largely coinciding with strong and positive $NDVI_{[t-1]}$ effects.
3. **Tair vegetation response** patterns are usually **strong and negative in dryland regions**.

Model coefficient comparison via Mann-Whitney U-Tests revealed that $NDVI_{[t-1]}$ is a larger memory component than Qsoil coefficients which, in turn, are larger components than Tair. Overall, $NDVI_{[t-1]}$ **explains more variance in NDVI z-scores than Qsoil data** or variance shared by $NDVI_{[t-1]}$ or Qsoil data at different rates across the three study regions with the most shared variance across Australia, followed by the Caatinga, and the Iberian Region.

3.1.4.2 Vegetation Memory Sensitivity

In terms of vegetation memory sensitivity:

1. $NDVI_{[t-1]}$ response coefficients are **negatively correlated** to mean NDVI records.
2. **Qsoil** response coefficients are **negatively correlated** to mean Qsoil records at differing intensities across the three study regions.
3. **Tair** response coefficients are **negatively correlated** to mean Tair records.

3.1.4.3 Regional Differences

Overall, whilst **general patterns of vegetation memory sensitivity** can be identified, the intensity of these **can vary strongly from region to region**. When comparing the overlap of vegetation memory coefficient patterns against one another (e.g. $NDVI_{[t-1]}$ response coefficients and Qsoil1 memory length) across all three study regions, noticeable differences arise:

1. $NDVI_{[t-1]}$ **memory** and **Qsoil1 memory length** are correlated negatively across Australia and the Caatinga, but positively throughout the Iberian Region.
2. **Qsoil1** response coefficients and **Qsoil1 memory length** are correlated negatively across Australia and the Iberian Region, but positively across the Caatinga.

These may identify the relative importance of different biological processes across all study regions.

3.2 Functional Aspects of Vegetation Memory

LHT and PFT data limitations across two of my three study regions - the Caatinga and Australia (see figures A.7 and A.9) are prevalent. The TRY data obtained for this project contains no geo-referenced records of H across these regions and coverage of N_{mass} records is minute at best. Additionally, COMPADRE stations are spread unevenly across the three study regions with 69 stations across the Iberian region (most of which are located across Spain), 10 stations throughout the Caatinga region and 16 COMPADRE sites representing all of Australia.

To avoid drawing conclusions from insufficient data, I am focussing my analysis of how functional aspects of vegetation and vegetation memory are linked only on the Iberian Region.

3.2.1 Life History Traits

Fast-Slow Continuum

Linking FSC-1 (life history speed) records to vegetation-response coefficients revealed limited strong correlations. See figure 3.11a for a visual representation. Whilst most intercepts of these regressions are statistically significant, only the regression slope of Tair vegetation response coefficients and FSC-1 is statistically significant but at a small effect size (see table 3.10). This suggest a weak link between Tair response coefficients and FSC-1 records with faster life histories leading to larger vegetation responses to Tair anomalies.

Table 3.10: COMPADRE FSC-1 and Vegetation Memory (Iberian Region) - Linear regression coefficients of COMPADRE FSC-1 data and vegetation response coefficients. Established via Chunk 32.

	NDVI [t-1]		**Tair**		**Qsoil1**		**Lag1**	
	I	S	I	S	I	S	I	S
Value	5.443	-0.0344	-0.3326	0.0262	1.8599	-0.0038	4.0145	0.0311
p_{Value}	0.000	0.1180	0.2618	0.0438	0.0024	0.8810	0.0001	0.4680

Linear relationships of FSC-2 (reproductive strategy) and vegetation-response coefficients across the Iberian region (see figure 3.11b) identified all intercepts (see table 3.11) to be statistically significant, only the regression slopes of $NDVI_{[t-1]}$ and Tair vegetation response coefficients and FSC-2. $NDVI_{[t-1]}$ memory decreases with increasing FSC-2 records. Tair vegetation responses increase with increasing FSC-2 records.

Table 3.11: COMPADRE FSC-2 and Vegetation Memory (Iberian Region) - Linear regression coefficients of COMPADRE FSC-2 data and vegetation response coefficients. Established via Chunk 32.

	NDVI [t-1]		**Tair**		**Qsoil1**		**Lag1**	
	I	S	I	S	I	S	I	S
Value	5.921	-0.1463	-0.6467	0.1053	1.9640	-0.0227	3.8272	0.1019
p_{Value}	0.000	0.0179	0.0425	0.0038	0.0035	0.7559	0.0007	0.4035

Period of Oscillation (π)

My analyses failed to identify links between π and vegetation response coefficients (see figure 3.11c). Only intercepts of Qsoil1 memory length and $NDVI_{[t-1]}$ regressions are statistically significant (see table 3.12). π may not be an informative criterion when concerned with vegetation memory.

Table 3.12: COMPADRE Pi and Vegetation Memory (Iberian Region) - Linear regression coefficients of COMPADRE Pi data and vegetation response coefficients. Established via Chunk 32.

	NDVI [t-1]		**Tair**		**Qsoil1**		**Lag1**	
	I	*S*	*I*	*S*	*I*	*S*	*I*	*S*
Value	5.106	-0.0435	0.6520	-0.0134	-0.1074	0.0796	3.4905	0.061
p_{Value}	0.000	0.3069	0.0875	0.5598	0.8773	0.0669	0.0113	0.454

Damping Ratio (ρ)

Much like with π, only the intercepts of some of the linear regressions of ρ and vegetation response coefficients (figure 3.11d and table 3.14) are statistically significant.Beyond this, I have not identified any meaningful links of vegetation memory and ρ.

Table 3.13: COMPADRE Rho and Vegetation Memory (Iberian Region) - Linear regression coefficients of COMPADRE Rho data and vegetation response coefficients. Established via Chunk 32.

	NDVI [t-1]		**Tair**		**Qsoil1**		**Lag1**	
	I	*S*	*I*	*S*	*I*	*S*	*I*	*S*
Value	3.6586	0.5212	0.7530	-0.2896	2.1915	-0.2479	6.6103	-0.9922
p_{Value}	0.0001	0.2784	0.2105	0.3580	0.0707	0.6937	0.0014	0.3456

Reactivity

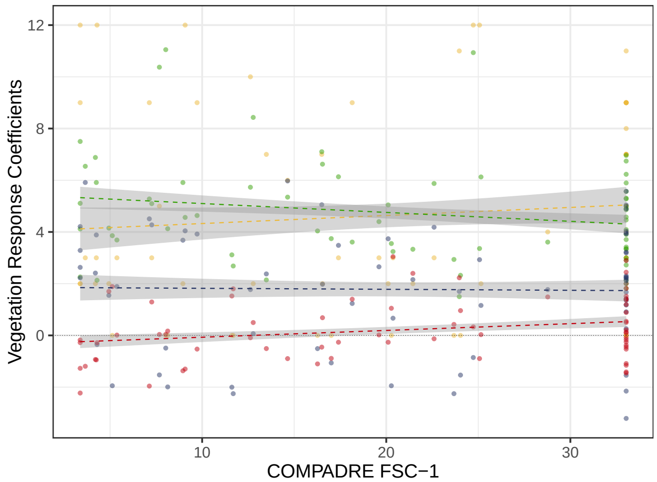
Whilst COMPADRE reactivity is not linked statistically significantly to Qsoil memory length, relationships of Tair vegetation-response coefficients and COMPADRE reactivity are (see table 3.14). Tair vegetation-response effects increase with increasing COMPADRE reactivity.

Table 3.14: COMPADRE Reactivity and Vegetation Memory (Iberian Region) - Linear regression coefficients of COMPADRE Reactivity data and vegetation response coefficients. Established via Chunk 32.

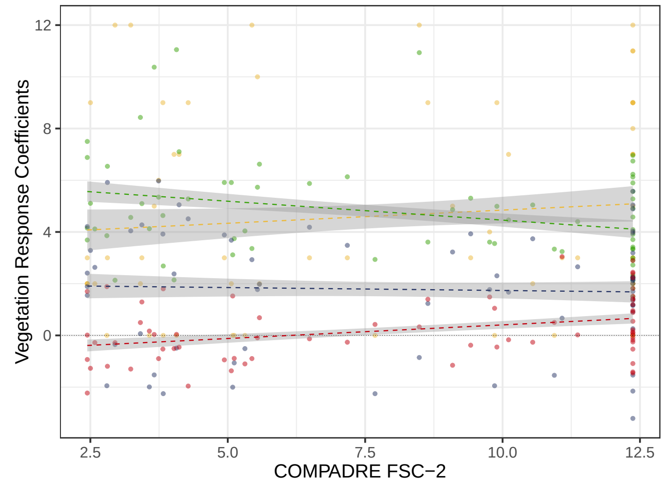
	NDVI [t-1]		**Tair**		**Qsoil1**		**Lag1**	
	I	*S*	*I*	*S*	*I*	*S*	*I*	*S*
Value	5.81	-0.6010	-0.9483	0.6209	3.1582	-0.7768	5.5325	-0.5127
p_{Value}	0.00	0.0845	0.0186	0.0029	0.0002	0.0639	0.0001	0.4679

Whilst relationships of $NDVI_{[t-1]}$, and Qsoil1 memory effects and COMPADRE reactivity aren't statistically significant, their patterns link up with what I hypothesised. $NDVI_{[t-1]}$ and Qsoil memory effects decrease with increasing COMPADRE reactivity. See figure 3.11e for a visual representation.

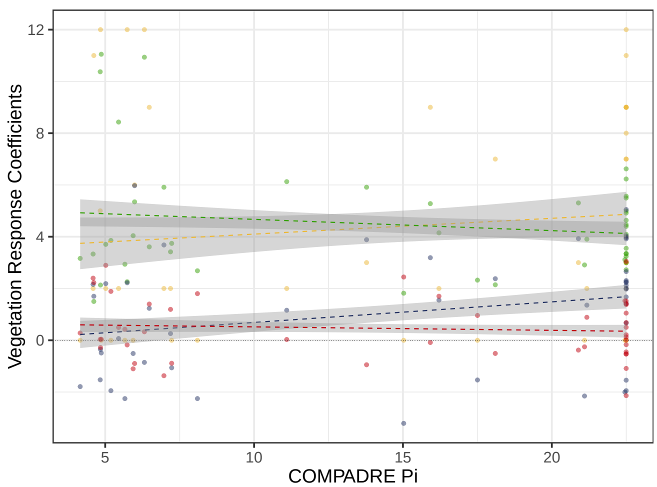
Overview



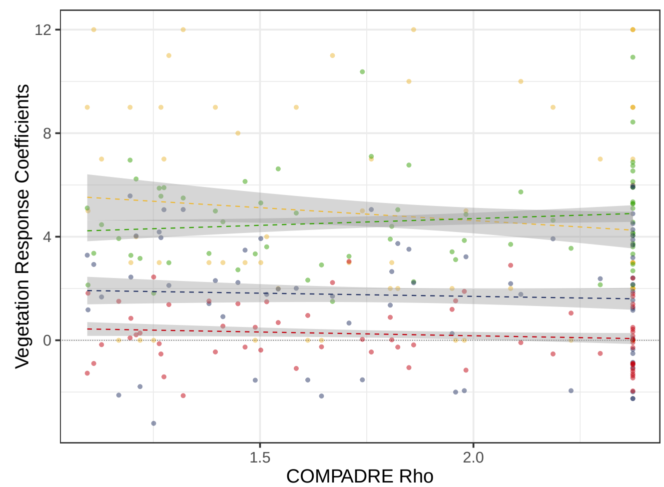
(a) FSC-1



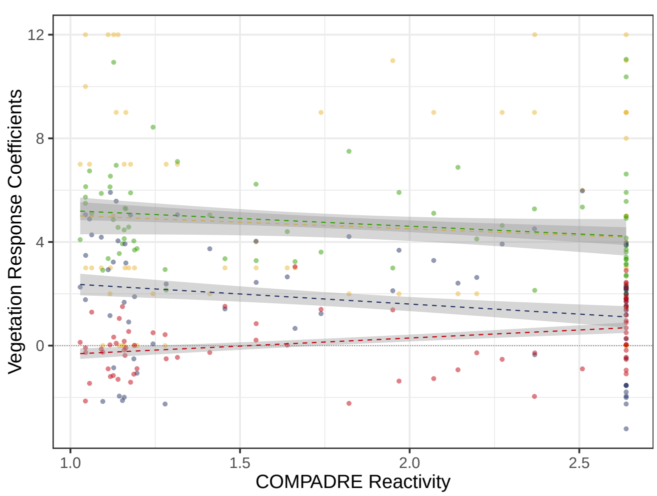
(b) FSC-2



(c) Period of Oscillation



(d) Damping Ratio



(e) Reactivity

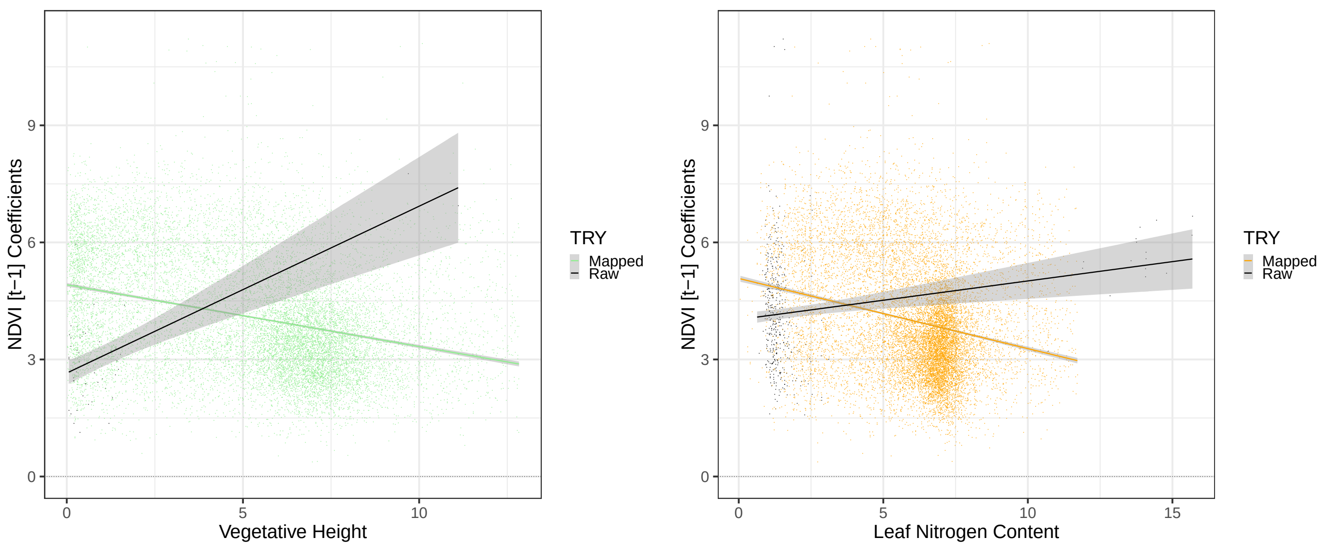
(f) Potting Colour Reference

Figure 3.11: Short - Linear regression coefficients of COMPADRE variables and vegetation response coefficients. Vegetation memory characteristics for each geo-referenced COMPADRE record across the Iberian Region are shown. Statistically non-significant regression slopes are presented as dashed lines. Figure established via Chunk 31.

3.2.2 Plant Functional Traits

Intrinsic Memory

Untreated geo-referenced records of H and N_{mass} are correlated positively with $NDVI_{[t-1]}$ memory effects (figure 3.12). This is not in accordance with my earlier hypotheses of H and N_{mass} enabling higher engineering resilience through material legacies but based on the most reliable (i.e. peer-reviewed, field-sampled, and not generated via statistical procedures) data and statistically significant (table 3.15).



(a) Height **(b)** Nitrogen

Figure 3.12: PFTs and Intrinsic Memory (Iberian Region) - $NDVI_{[t-1]}$ (figure 3.1) plotted against expressions of (a) H (vegetative height), and (b) N_{mass} (leaf nitrogen content) both via geo-referenced TRY records and mapped species-specific trait means (figure A.5). Figure established via Chunk 33.

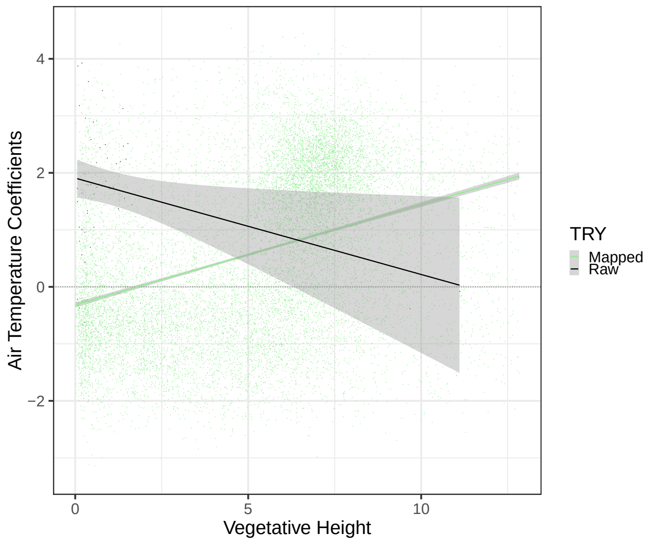
Using the PFT maps obtained via the PFT mapping approach outlined in figure 1.5, one obtains negative correlations of H, and N_{mass} with $NDVI_{[t-1]}$ memory effects. These relationships can arguably be seen as synonymous with the $NDVI_{[t-1]}$ vegetation memory sensitivity relationship: Areas of lower biomass (low mean NDVI, low H, and low N_{mass}) exhibit stronger intrinsic vegetation memory.

Table 3.15: PFTs and $NDVI_{[t-1]}$ Memory (Iberian Region) - Coefficients of linear regressions of $NDVI_{[t-1]}$ response coefficients against PFT data (vegetative height H, and leaf nitrogen mass Nmass) both as raw geo-referenced data (R), and extended maps of species-specific trait means (M). Established via Chunk 34.

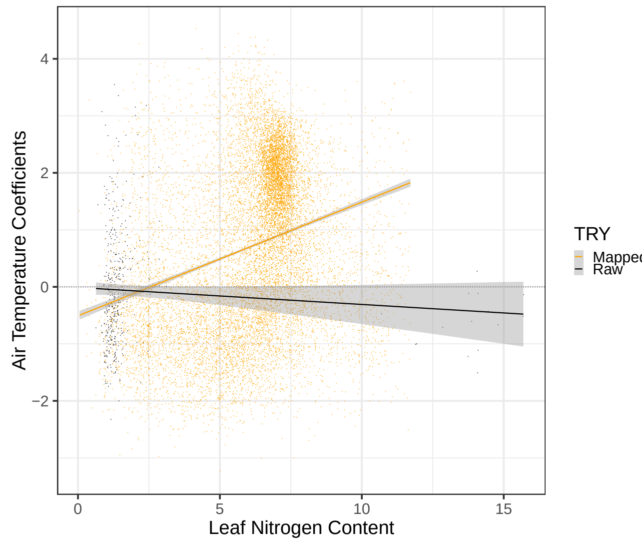
	Vegetative Height				**Nitrogen Content**			
	R		M		R		M	
	I	S	I	S	I	S	I	S
Value	2.651	0.4276	4.914	-0.1582	4.023	0.0989	5.077	-0.1799
p_{Value}	0.000	0.0000	0.000	0.0000	0.000	0.0004	0.000	0.0000

Air Temperature

Untreated, geo-referenced records of H, and N_{mass} are correlated negatively with Tair vegetation memory effects (see figure 3.13) - not statistically significant in the case of the latter (table 3.16). This in accordance with what I postulated earlier of material legacies (i.e. higher H and N_{mass} records) leading to enhanced vegetation resistance.



(a) Height



(b) Nitrogen

Figure 3.13: PFTs and Tair Memory (Iberian Region) - Tair Memory (figure 3.1) plotted against expressions of (a) H (vegetative height), and (b) N_{mass} (leaf nitrogen content) both via geo-referenced TRY records and mapped species-specific trait means (figure A.5). Figure established via Chunk 33.

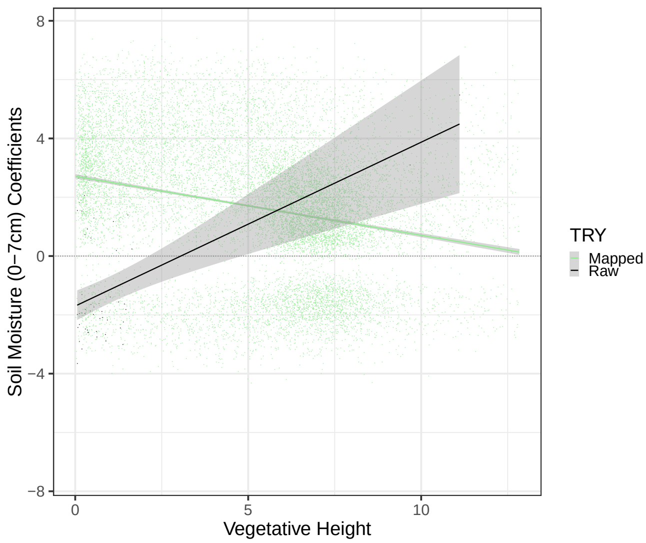
When using mapped species-specific mean PFT records, these relationships are turned around with H and N_{mass} now correlated positively with Tair vegetation response coefficients.

Table 3.16: PFTs and Tair Memory (Iberian Region) - Coefficients of linear regressions of Tair response coefficients against PFT data (vegetative height H, and leaf nitrogen mass Nmass) both as raw geo-referenced data (R), and extended maps of species-specific trait means (M). Established via Chunk 34.

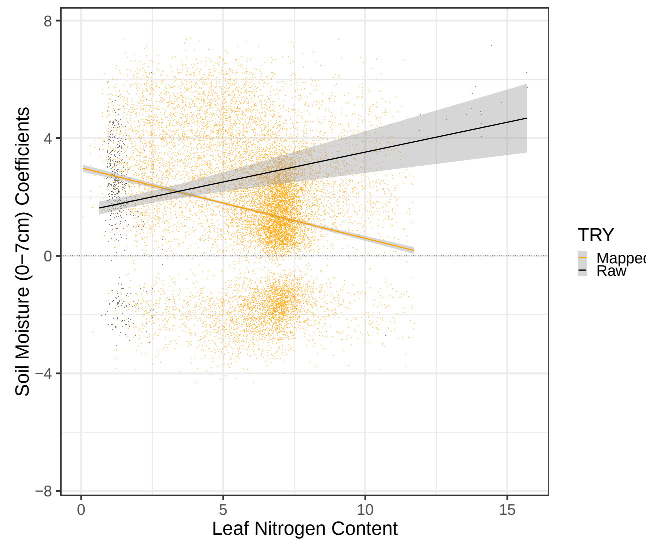
	Vegetative Height				**Nitrogen Content**			
	R		M		R		M	
	I	S	I	S	I	S	I	S
Value	1.908	-0.1691	-0.317	0.1761	-0.0113	-0.0298	-0.5085	0.1994
p_{Value}	0.000	0.0276	0.000	0.0000	0.8550	0.1482	0.0000	0.0000

Soil Moisture (0-7cm)

My analyses identified statistically significant relationships for all Qsoil1 related vegetation response coefficients with the exception of untreated H records and Qsoil1 memory length (see figure 3.14).



(a) Height



(b) Nitrogen

Figure 3.14: PFTs and Qsoil1 Memory (Iberian Region) - Qsoil1 Memory (figure 3.1) plotted against expressions of (a) H (vegetative height), and (b) N_{mass} (leaf nitrogen content) both via geo-referenced TRY records and mapped species-specific trait means (figure A.5). Figure established via Chunk 33.

Whilst geo-referenced data records of H and N_{mass} establish themselves in positive correlations with Qsoil1 memory effects and Qsoil1 memory length in the case of H, mapped mean trait values of H and N_{mass} are correlated negatively to Qsoil1 related vegetation response coefficients (see table 3.17).

Table 3.17: PFTs and Qsoil1 Memory (Iberian Region) - Coefficients of linear regressions of Qsoil1 response coefficients against PFT data (vegetative height H, and leaf nitrogen mass Nmass) both as raw geo-referenced data (R), and extended maps of species-specific trait means (M). Established via Chunk 34.

	Qsoil				**Lag**			
	H		Nmass		H		Nmass	
	R	M	R	M	R	M	R	M
Intercept	-1.7021	2.7139	1.4959	2.9909	2.3419	4.2748	5.3394	4.8624
$p_{Intercept}$	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0000
Slope	0.5574	-0.2007	0.2031	-0.2398	0.1256	-0.0851	-0.2027	-0.1781
p_{Slope}	0.0000	0.0000	0.0000	0.0000	0.6681	0.0000	0.0127	0.0000

3.2.3 Summary

3.2.3.1 Life History Traits

Linear regressions of COMPADRE data against vegetation response coefficients have proven **largely inconclusive**.

Making use of COMPADRE data across the **Iberian region**, COMPADRE **FSC-2** (reproductive strategy) and **reactivity** emerged as **important factors in explaining vegetation memory patterns**:

- *FSC-2* - Despite small effect sizes, the following correlations of FSC-2 and vegetation-memory characteristics have proven statistically significant:

  1. $NDVI_{[t-1]}$ memory strength decreases as species rank higher on the FSC-2 spectrum.
  2. Tair vegetation response strength increases as species rank higher on the FSC-2 spectrum.

- *Reactivity* - My vegetation memory hypotheses concerning the FSC spectrum are built on the assumption of perturbations leading to a die-off of local vegetation which will be counteracted by (1) fast growth (FSC-1), or (2) fast reproduction of surviving plants (FSC-2). On the other hand, I postulated reactivity, to characterise the responses of communities exerted by surviving communities in response to perturbations which do not necessarily pose a selection pressure. Despite my analyses of reactivity and vegetation-memory characteristics largely being statistically non-significant, some hints at biological reactions to perturbations have emerged which I am including at this point to highlight the possible explanatory power of the COMPADRE data set:

  1. $NDVI_{[t-1]}$ memory strength decreases as reactivity increases.
  2. Tair vegetation response strength increases as reactivity increases.
  3. Qsoil memory strength decreases as reactivity increases.

Overall, **vegetation memory strength** seems to **decrease** as **reactivity increases**.

3.2.3.2 Plant Functional Traits

Analyses of the relationships of PFT expressions and vegetation memory characteristics revealed that:

1. **Geo-referenced PFT records can be linked to vegetation memory characteristics** in a statistically significant way:
   - $NDVI_{[t-1]}$ memory effects increase as H and N_{mass} increase.
   - Tair vegetation response effects decrease as H and N_{mass} increase.
   - Qsoil1 memory effects increase as H and N_{mass} increase.
2. **Mapped species-specific PFT means can be linked to patterns of vegetation memory** characteristics:
   - $NDVI_{[t-1]}$ memory effects decrease as H and N_{mass} increase.
   - Tair vegetation response effects increase as H and N_{mass} increase.
   - Qsoil1 memory effects decrease as H and N_{mass} increase.
3. There is a **mismatch between geo-referenced PFT records and mapped species-specific trait means** in that relationships of the two towards the same vegetation memory characteristics often follow correlations of opposite sign.

4. Discussion

The first research goal of my thesis was to identify and map vegetation memory characteristics across dryland regions. Additionally, I was aiming to analyse the key differences in the contributions of intrinsic and extrinsic memory across the different dryland regions (i.e. first research goal).

4.1 Vegetation Memory

4.1.1 Spatial Patterns of Vegetation Memory

4.1.1.1 Intrinsic Vegetation Memory

The strongest intrinsic vegetation-memory effects can be found throughout the dryland areas of the three study regions. According to the proposed inverse link between intrinsic vegetation memory and engineering resilience, my study would have identified **dryland regions** to **exert lower engineering resilience** (i.e. slower recovery rates) than surrounding non-dryland regions. These findings link up well with De Keersmaecker et al. across the Iberian Region and Australia [3].

However, my analyses also demonstrated that some biological processes previously believed to be depicting intrinsic memory might be best explained through extrinsic vegetation memory. Therefore, I believe that **the assumption that intrinsic vegetation memory is a direct proxy of engineering resilience may be an oversimplification**.

4.1.1.2 Extrinsic Vegetation Memory

Soil-Moisture Effects

I have identified positive effects of extrinsic Qsoil vegetation memory across the dryland portions of all three study regions. These patterns are almost identical to the distributions of arid and semi-arid ombrotypes (see figures 2.1, 2.2, and 2.3). Thus, dryland vegetation reacts positively to positive anomalies in soil moisture depicting the water limitation of drylands. These findings partially mimic the findings by Seddon et al. [7] in showing similar patterns to the precipitation climate weight of their vegetation sensitivity index which highlights the importance of **water limitations in dryland regions**. Whilst some positive Qsoil vegetation memory effects have been identified for the sub-humid typified areas across the study regions (e.g. south-western France) these are much weaker than the Qsoil effects within the adjacent dryland regions. Conclusively, dryland vegetation reacts stronger and more positively to positive Qsoil anomalies than plants located in sub-humid regions.

Air-Temperature Driven Responses

Instantaneous vegetation response to positive Tair anomalies is negative across dryland areas of all three study regions. This depicts **negative effects of increased evaporation and levels of aridity** (that are Tair-driven) **on dryland vegetation**. In northern regions and high altitude areas of the Iberian Region, these effects are positive signalling that vegetation in cold/wet regions reacts positively to positive anomalies in Tair. This, in turn, depicts positive effects of increased solar radiation (coupled with higher Tair records) and increased speeds of biological processes in higher temperatures.

4.1.1.3 Regional Differences

Whilst the above patterns hold true for all three study regions, all three exhibit largely different spatial patterns of Qsoil memory length. This manifests in two important relationships of vegetation memory characteristics:

1. $NDVI_{[t-1]}$ memory and Qsoil memory length are correlated negatively across Australia and the Caatinga. This suggests that, across these regions, intrinsic vegetation memory ($NDVI_{[t-1]}$) is lower, the longer vegetation responses to Qsoil anomalies. Throughout the Iberian Region, this pattern is reversed with lower intrinsic vegetation memory in regions of shorter Qsoil vegetation memory.
2. Qsoil memory and Qsoil memory length are linked differently across the three study regions. Whilst short Qsoil vegetation memory overlaps with strong Qsoil response coefficients across the Caatinga, the reverse is true considering the Iberian Region and Australia. I conclude that vegetation in the Caatinga region reacts almost instantaneously to Qsoil anomalies whereas the vegetation of the drier areas of the Iberian Region exerts strongest Qsoil memory at long time scales. Australian vegetation exhibits a clear west-to-east pattern of decreasing Qsoil vegetation memory length.

The effect sizes of any of the relationships mentioned above are small which is most likely due to regressions being built across biomes. Further sub-setting of dryland regions may alleviate this. Nevertheless, this highlights that even though all study regions share similar base characteristics of **vegetation memory**, the **processes** of how these establish themselves **differ** greatly **between the regions**. This may be because of different physical systems, different periodicity of the environments of each region, contrasting perturbation histories, or diverging functional aspects of local vegetation between these regions. Further high-resolution local studies may be needed to uncover the causes for the observed differences in vegetation memory characteristics.

4.1.2 Intrinsic and Extrinsic Vegetation Memory

Regarding the $NDVI_{[t-1]}$ analysis carried out by Seddon et al. [7], across Australia, my findings of $NDVI_{[t-1]}$ memory patterns deviate significantly from the west-to-east drop-off in $NDVI_{[t-1]}$ intensity identified by Seddon et al. Additionally, the the peak of $NDVI_{[t-1]}$ memory strength of my analyses in the greater Caatinga region is located closer to the eastern Brazilian coast than identified by Seddon et al. This pattern coincides with the short Qsoil1 memory lengths observed in this region. Thus, I postulate that **intrinsic vegetation memory might be masked extrinsic vegetation memory**. Within this context, my analyses have identified vegetation memory information to be assigned to Qsoil-driven effects which was previously believed to be contained within $NDVI_{[t-1]}$ effects. Additional support for this hypothesis stems from the variance partitioning results with $NDVI_{[t-1]}$ and Qsoil sharing the most variance across Australia and the second-most throughout the Caatinga region.

That being said, **distinguishing intrinsic and extrinsic vegetation memory remains challenging** since I have shown that some intrinsic memory patterns can in fact be best explained through extrinsic memory effects so long as the right extrinsic variables are taken into account. This *may indicate that intrinsic vegetation memory is not a measure of engineering resilience but a combination of vegetation responses to different abiotic drivers.*

More climate variables are needed to identify additional extrinsic vegetation memory effects which can explain NDVI anomaly patterns by capturing variation currently contained in $NDVI_{[t-1]}$ effects. Since my analyses have shown that even dryland regions adhere to different spatio-temporal processes, such supplementary variables should ideally contain information about a host of ecosystem-processes. The importance of this consideration is boosted when one is concerned with non-dryland region which are governed by different processes than dryland vegetation.

Additionally, my analysis has been limited to monthly intervals for 33 years. Sub-monthly time-step analyses will allow for a more refined identification of bio-climatic processes thus increasing the chances of identifying additional extrinsic memory effects particularly in fast-responding ecosystems such as the Caatinga. Longer time series analyses would allow for more confident distinction of intrinsic and extrinsic memory given a higher rate of bioclimatic anomalies than in shorter time series.

4.1.3 Climate Reanalysis Data in Vegetation Memory Analyses

Starting this analyses, I argued the case that implementing climate data reanalysis data sets into vegetation memory analyses will enhance the same when compared to using observational data sets. Having finished the vegetation memory analyses and identified hitherto unrecognised patterns of vegetation memory, I conclude that **climate reanalysis data sets enhance vegetation memory analyses** when compared to observational data sets. Key advantages of the ERA5 dataset are as follows:

1. *Variables* - ERA5 contains more than a hundred climatic variables including all bioclimatic variables recognised by the WorldClim data set. In addition, due to ERA5 being the result of a model procedure, one can obtain a host of information criteria pertaining to complete hydrological cycles, radiation regimes, and surface conditions.
2. *Temporal Resolution* - ERA5 supplies information of each variable at hourly intervals thus much improving over the monthly time-steps used here.
3. *Time Series Length* - ERA5 data is available for the years of 1950 to today, effectively offering 32 years worth of data I couldn't make use of due to the limitation of NDVI time series availability.
4. *Uncertainty Records* - Uncertainty records are proxies of how sure one can be of any given data record. These data are provided in ERA5 and will make for more refined analyses especially when considering Bayesian frameworks which handle data uncertainty in a very favourable way.

Therefore, ERA5 makes for a great resource of important bioclimatic variables as well as their respective uncertainty records at favourable temporal resolutions and time scales. Kriging methods can be employed to downscale ERA5 outputs to almost any spatial target resolution whilst retaining uncertainty records thus making the two a strong duo in bioclimatic analyses of the future.

4.1.4 Vegetation Memory Sensitivity

4.1.4.1 Intrinsic Vegetation Memory

The negative correlation of $NDVI_{[t-1]}$ memory effects to mean NDVI records suggests that **high intrinsic vegetation memory** is **linked to regions of low biomass** (low mean NDVI records). This effect holds true across all three study regions thus identifying a global link between vegetation memory and plant biomass reminiscent of the concept of **material legacies** proposed by Johnstone et al [9].

4.1.4.2 Extrinsic Vegetation Memory

Soil Moisture Effects

The negative correlation of Qsoil memory effects to mean Qsoil highlights an increased resistance potential towards Qsoil anomalies exerted by vegetation in sub-humid regions when compared to semi-arid and arid regions. The humid areas and high-altitude areas of the Iberian Region as well as the western Tasmanian coastline are characterised by weakened vegetation resistance against Qsoil anomalies. This highlights a climate sensitivity of ecosystems with **low resistance toward soil moisture anomalies at upper and lower ends of the moisture spectrum, and high resistance to Qsoil anomalies at intermediate moisture levels**. Throughout Australia, this relationship is much weaker ($\sim \frac{1}{10}$ of the forcing across the Caatinga and $\sim \frac{1}{20}$ of the forcing across the Iberian Region). This may indicate that whilst one can describe the vegetation-memory sensitivity of vegetation across the Iberian Region and the Caatinga with intuitive bell-shaped functions, these relationships break down across Australia pointing towards additional factors I have not assessed here.

Air Temperature Driven Responses

Much like with the Qsoil spectrum, the negative correlation of instantaneous Tair effects to mean Tair records points to ecosystem sensitivity patterns with **low resistance at upper and lower Tair spectrum, and high resistance to Tair anomalies at intermediate Tair levels**.

4.2 Functional Aspects of Vegetation Memory

4.2.1 Life History Traits

COMPADRE data has proven largely difficult to link to vegetation memory characteristics. Despite having found the relationships described below (which are highly biologically feasible), it remains interesting to address why the other COMPADRE variables did not link up as well with vegetation memory characteristics. This is probably due to a scale mismatch between COMPADRE plot sites and the NDVI spatial grid used here. It is uncertain how well each COMPADRE plot represent the NDVI grid it falls into and from which vegetation memory characteristics have been identified. In addition, data coverage of COMPADRE sites is relatively low and not spaced along gradients of vegetation-memory patterns. Despite these issues, I have identified two COMPADRE variables of special interest in explaining vegetation memory:

1. **FSC-2** - Reproductive strategy within the Fast-Slow continuum.
   - The faster/more species reproduce, the lower the intrinsic vegetation memory of their ecosystems. Thus, **ecosystems populated by species of fast reproductive strategies exhibit lower intrinsic vegetation memory**.
   - Species of faster/increased reproduction exhibit higher Tair vegetation response coefficients, signalling decreased resistance towards Tair anomalies. Therefore, I conclude that **ecosystems populated by species of fast reproductive strategies exhibit lower resistance towards Tair anomalies**.
2. **Reactivity** - First time-step biological response to perturbations which I am focussing on particularly due to its strong theoretical link to core processes of vegetation memory especially considering short-time scale responses across dryland regions.
   - **As reactivity increases, intrinsic vegetation memory ($NDVI_{[t-1]}$) and Qsoil vegetation memory decrease** thus enhancing ecosystem recovery as well as ecosystem resistance. Therefore, first-time step responses of greater magnitude allow for faster recovery rates by means of stronger autocorrelation of vegetation features. Additionally, stronger first-time step responses can serve to diminish the effects Qsoil anomalies.
   - **Higher reactivity** prompts **increased Tair vegetation-response intensity** signalling decreasing instantaneous resistance to Tair anomalies in system of higher reactivity. This is in contrast to the Qsoil resistance and I argue that there may be two reasons for this: (1) increased reactivity comes at the price of decreased Tair resistance, or (2) instantaneous response can't be captured by first-time-step responses (i.e. reactivity).

Overall, I have highlighted the biological feasibility of linking COMPADRE data and vegetation response coefficients. I hypothesize that additional COMPADRE data or vegetation memory analyses at finer spatial resolutions will find these patterns to be statistically significant.

4.2.2 Plant Functional Traits

Expressions of PFTs can hardly be linked to vegetation memory in this study which mirrors the findings of Lauglin et al.[149] who identified PFTs to be unreliable estimates of community adaptation to climate characteristics. My inability to find meaningful links between PFTs and vegetation response coefficients might additionally be caused by low data availability and coverage or due to changes in PFT compositions over time which have not been taken into account. Furthermore, I observed a mismatch in identified (non-significant) links of PFTs and vegetation memory between untreated, geo-referenced PFT records and mapped species-specific PFT means. This suggests that the PFT mapping approach may require further developing or that data quality (either of untreated, geo-referenced records, or floral occurrence records) is sub par. It is likely that a combination of higher quality data and more refined mapping methods will solve this issue. Additionally, this may be a sign of material legacies being overpowered by information legacies which the selected PFTs do not capture. Further research will aid identifying which is the case.

5. Conclusion

Understanding spatio-temporal patterns of vegetation performance in the face of environmental anomalies which are expected to be exacerbated by climate change is paramount to successful ecosystem management decisions.

Within this study, I have synthesized contemporary approaches of identifying vegetation memory (proposed as a proxy of engineering resilience [3]), state-of-the-art climate reanalysis data sets, and functional ecology approaches. The aims was to generate an enhanced understanding of how antecedent environmental conditions influence vegetation performance through time across three study regions spanning three continents and consisting of predominantly dryland areas.

5.1 Vegetation Memory

According to U-Tests and variance partitioning results, region-wide vegetation memory effects sizes are ranked as follows: $NDVI_{[t-1]}$ > Qsoil (Qsoil1 > Qsoil2 > Qsoil3 > Qsoil4) > Tair. This highlights that, across all three study regions, **intrinsic vegetation memory drives systems more than extrinsic vegetation memory**. This can be used to answer **research question I.1** (*To what extent can extrinsic and intrinsic vegetation memory be identified in dryland regions?*):

$NDVI_{[t-1]}$ and Qsoil1 are the most robust intrinsic and extrinsic vegetation memory metrics.

However, the relative strength of these effects, and their relationships, depend on the study region. I have merely assessed dryland regions and hypothesise that Arctic or alpine region will be driven more by temperature than by water availability, for example. The clear overlap in patterns of $NDVI_{[t-1]}$ and Qsoil1 memory effects suggest that **intrinsic memory might be a masked extrinsic memory effect** in some areas. Additional support for this arguments stems from the fact that Seddon et al.'s [7] $NDVI_{[t-1]}$ pattern across Australia is reminiscent of the Qsoil1 memory length pattern identified here with high autoregressive coefficients in Seddon et al. coinciding with long Qsoil1 memory lags within my study. Variance partitioning further underlines this by depicting great amounts of variance in my vegetation memory models being shared between $NDVI_{[t-1]}$ and Qsoil1 across Australia and less so across the other two study regions. This information can be used in answering **research question I.2** (*How well can we distinguish between intrinsic and extrinsic memory in dryland regions?*):

It is difficult to distinguish intrinsic and extrinsic memory in certain regions.

This is due to the high portion of shared variance of $NDVI_{[t-1]}$ and extrinsic memory components. Further research into robust extrinsic vegetation memory metrics will undoubtedly serve to diminish the importance of $NDVI_{[t-1]}$ in these model approaches and aid in disentangling intrinsic and extrinsic vegetation memory. For example, including model selection for a range of lagged responses within this study has led to more uniform/less intense patterns of $NDVI_{[t-1]}$ memory (especially across Australia when compared to Seddon et al. [7]). Future models may approach this in a Generalised Linear Mixed Effect Model setting with Qsoil lags being implemented as random effects rather than separate models. In general, however, this suggest that the **assumption of intrinsic vegetation memory as a proxy for engineering resilience may be an oversimplification** and the $NDVI_{[t-1]}$ coefficient may be a catch-all variable for a host of extrinsic memory coefficients.

The use of ERA5 data has proven highly informative in identifying vegetation memory components and their respective characteristics. Future vegetation memory modelling procedures may investigate the usefulness of soil moisture-driven drought indices [150] calculated directly from ERA5 data. Additionally, leveraging the uncertainty records in the ERA5 data base may allow for more refined predictions of vegetation performance and responses to perturbations by more accurately representing such forcing. Arguably, the usefulness of ERA5-driven approaches in biological sciences spans far beyond the reach of vegetation memory analyses.

5.2 Functional Aspects of Vegetation Memory

Furthermore, I aimed to **identify a set of LHTs which are linked to vegetation memory characteristics**. Placement along the FSC-2 spectrum (ranging from low reproductive output to fast reproducing species) and COMPADRE reactivity (biological first time-step response to a biologically relevant event) are linked to vegetation memory characteristics. Thus I answer **research question II.1** (*Which traits of biological function (PFT and LHT) are related to vegetation memory characteristics?*) as follows:

Measures of plant reproductive strategy and first time-step response are related to vegetation memory characteristics.

Whilst this is a small subset of many LHTs and PFTs which can be argued to be linked to vegetation memory in biologically feasible way, **neither of them serve as direct proxies of either intrinsic or extrinsic vegetation memory components**. Overall, populations ranking higher on the FSC-2 spectrum exert decreased intrinsic vegetation memory and decreased resistance against Tair anomalies suggesting a trade-off between the two. Further proof of this trade-off is supplied by the fact that higher COMPADRE reactivity prompts decreased intrinsic vegetation memory and decreased resistance to Tair anomalies but increased resistance to Qsoil anomalies. This, in turn, identifies a trade-off between maximising vegetation memory (both as recovery and resistance) to certain environmental drivers and other environmental drivers instead of a general trade-off of intrinsic vegetation memory and extrinsic vegetation memory. Thus, I answer my **research question II.2** (*What biological traits cause areas to exert intrinsic and extrinsic memory respectively?*)

Assigning intrinsic/extrinsic memory effects to separate aspects of plant function/life histories is difficult and needs further research

My analyses of PFTs and possible links to vegetation memory characteristics have proven inconclusive. I argue that further research into this field should leverage the BIEN data base to increase the geo-spatial range of PFT records. Additionally, the basic mapping approach of PFTs from PFT data bases and floral data sources employed here has yielded contrasting links to vegetation memory when compared to geo-referenced PFT records. Therefore, it is paramount that future studies refine this approach if applied.

Bibliography

[1] Convention on Biological Diversity. Aichi Biodiversity Targets (2013).

[2] Newton, A. C. Biodiversity Risks of Adopting Resilience as a Policy Goal. *Conservation Letters* **9**, 369–376 (2016).

[3] De Keersmaecker, W. *et al.* A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Global Ecology and Biogeography* **24**, 539–548 (2015).

[4] Hodgson, D., McDonald, J. L. & Hosken, D. J. What do you mean, 'resilient'? *Trends in Ecology & Evolution* **30**, 503–506 (2015).

[5] Liu, L., Zhang, Y., Wu, S., Li, S. & Qin, D. Water memory effects and their impacts on global vegetation productivity and resilience. *Scientific Reports* **8**, 1–9 (2018).

[6] Vicente-Serrano, S. M. *et al.* Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences* **110**, 52–57 (2013).

[7] Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D. & Willis, K. J. Sensitivity of global terrestrial ecosystems to climate variability. *Nature* **531**, 229–232 (2016).

[8] Ogle, K. *et al.* Quantifying ecological memory in plant and ecosystem processes. *Ecology Letters* **18**, 221–235 (2015).

[9] Johnstone, J. F. *et al.* Changing disturbance regimes, ecological memory, and forest resilience. *Frontiers in Ecology and the Environment* **14**, 369–378 (2016). arXiv:1011.1669v3.

[10] Core Writing Team, R.K. Pachauri & L.A. Meyer. Climate Change 2014: Synthesis Report. Contribution of Working Groups I; II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. Rep., IPCC, Geneva; Switzerland (2014).

[11] Santiago, L. S., Bonal, D., Guzman, M. E. D. & Eleinis, Á. Drought Survival Strategies of Tropical Trees. In Goldstein, G. & Santiago, L. S. (eds.) *Tropical Tree Physiology*, vol. 6 of *Tree Physiology*, 243–258 (Springer International Publishing, Cham, 2016).

[12] Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* **259**, 660–684 (2010). arXiv:1011.1669v3.

[13] Kéfi, S., Holmgren, M. & Scheffer, M. When can positive interactions cause alternative stable states in ecosystems? *Functional Ecology* **30**, 88–97 (2016).

[14] Seidl, R., Spies, T. A., Peterson, D. L., Stephens, S. L. & Hicke, J. A. REVIEW: Searching for resilience: addressing the impacts of changing disturbance regimes on forest ecosystem services. *Journal of Applied Ecology* **53**, 120–129 (2016).

[15] Senf, C. & Seidl, R. Natural disturbances are spatially diverse but temporally synchronized across temperate forest landscapes in Europe. *Global Change Biology* **24**, 1201–1211 (2018). 0608246v3.

[16] Scheffer, M. Foreseeing tipping points. *Nature* **467**, 411–412 (2010).

[17] Xu, C., Van Nes, E. H., Holmgren, M., Kéfi, S. & Scheffer, M. Local Facilitation May Cause Tipping Points on a Landscape Level Preceded by Early-Warning Indicators. *The American Naturalist* **186**, E81–E90 (2015).

[18] Vasilakopoulos, P. & Marshall, C. T. Resilience and tipping points of an exploited fish population over six decades. *Global Change Biology* **21**, 1834–1847 (2015).

[19] Côté, I. M. & Darling, E. S. Rethinking Ecosystem Resilience in the Face of Climate Change. *PLoS Biology* **8**, e1000438 (2010).

[20] Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).

[21] Genkai-Kato, M. Regime shifts: catastrophic responses of ecosystems to human impacts. *Ecological Research* **22**, 214–219 (2007).

[22] Australian Government Department of the Environment. National Climate Resilience and Adaptation Strategy 2015 Vision for a Climate Resilient Australia (2015). URL www.environment.gov.au.

[23] Holling, C. S. Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics* **4**, 1–23 (1973).

[24] Mumby, P. J. & Anthony, K. R. N. Resilience metrics to inform ecosystem management under global change with application to coral reefs. *Methods in Ecology and Evolution* 1–9 (2015).

[25] Scheffer, M., Carpenter, S. R., Dakos, V. & van Nes, E. H. Generic Indicators of Ecological Resilience: Inferring the Chance of a Critical Transition. *Annual Review of Ecology, Evolution, and Systematics* **46**, 145–167 (2015).

[26] Standish, R. J. *et al.* Resilience in ecology: Abstraction, distraction, or where the action is? *Biological Conservation* **177**, 43–51 (2014).

[27] Gunderson, L. H. Ecological Resilience—In Theory and Application. *Annual Review of Ecology and Systematics* **31**, 425–439 (2000).

[28] Oliveras, I. *et al.* Andean grasslands are as productive as tropical cloud forests. *Environmental Research Letters* **9**, 115011 (2014).

[29] Foley, J. A., Coe, M. T., Scheffer, M. & Wang, G. Regime Shifts in the Sahara and Sahel: Interactions between Ecological and Climatic Systems in Northern Africa. *Ecosystems* **6**, 524–532 (2003).

[30] Ares, J., Bertiller, M. & Del Valle, H. Functional and structural landscape indicators of intensification, resilience and resistance in agroecosystems in southern Argentina based on remotely sensed data. *Landscape Ecology* **16**, 221–234 (2001).

[31] Harris, A., Carr, A. & Dash, J. Remote sensing of vegetation cover dynamics and resilience across southern Africa. *International Journal of Applied Earth Observation and Geoinformation* **28**, 131–139 (2014).

[32] Redmond, M. D., Weisberg, P. J., Cobb, N. S. & Clifford, M. J. Woodland resilience to regional drought: Dominant controls on tree regeneration following overstorey mortality. *Journal of Ecology* **106**, 625–639 (2018).

[33] Papagiannopoulou, C. *et al.* A non-linear Granger-causality framework to investigate climate–vegetation dynamics. *Geoscientific Model Development* **10**, 1945–1960 (2017).

[34] Gazol, A., Camarero, J. J., Anderegg, W. R. L. & Vicente-Serrano, S. M. Impacts of droughts on the growth resilience of Northern Hemisphere forests. *Global Ecology and Biogeography* **26**, 166–176 (2017).

[35] Scheffer, M. & Carpenter, S. R. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution* **18**, 648–656 (2003).

[36] Rist, L. *et al.* Applying resilience thinking to production ecosystems. *Ecosphere* **5**, art73 (2014).

[37] Tabacchi, E., Steiger, J., Corenblit, D., Monaghan, M. T. & Planty-Tabacchi, A.-M. Implications of biological and physical diversity for resilience and resistance patterns within Highly Dynamic River Systems. *Aquatic Sciences* **71**, 279–289 (2009).

[38] Mumby, P. J., Chollett, I., Bozec, Y. M. & Wolff, N. H. Ecological resilience, robustness and vulnerability: How do these concepts benefit ecosystem management? (2014). URL http://dx.doi.org/10.1016/j.cosust.2013.11.021.

[39] Angeler, D. G. & Allen, C. R. Quantifying resilience. *Journal of Applied Ecology* **53**, 617–624 (2016).

[40] Mouillot, D., Graham, N. A. J., Villéger, S., Mason, N. W. H. & Bellwood, D. R. A functional approach reveals community responses to disturbances. *Trends in Ecology and Evolution* **28**, 167–177 (2013). 2072.

[41] Grimm, V. & Wissel, C. Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia* **109**, 323–334 (1997).

[42] Wardle, D. A. & Jonsson, M. Long-term resilience of above- and below-ground ecosystem components among contrasting ecosystems. *Ecology* **95**, 1836–1849 (2014).

[43] Seddon, A. W. R., Froyd, C. A., Witkowski, A. & Willis, K. J. A quantitative framework for analysis of regime shifts in a Galápagos coastal lagoon. *Ecology* **95**, 3046–3055 (2014).

[44] Mori, A. S., Furukawa, T. & Sasaki, T. Response diversity determines the resilience of ecosystems to environmental change. *Biological Reviews* **88**, 349–364 (2013).

[45] Selkoe, K. A. *et al.* Principles for managing marine ecosystems prone to tipping points. *Ecosystem Health and Sustainability* **1**, 1–18 (2015).

[46] Peterson, G., Allen, C. R. & Holling, C. S. Original Articles: Ecological Resilience, Biodiversity, and Scale. *Ecosystems* **1**, 6–18 (1998). 9610022.

[47] Brown, E. D. & Williams, B. K. Resilience and Resource Management. *Environmental Management* **56**, 1416–1427 (2015).

[48] Cole, L. E. S., Bhagwat, S. A. & Willis, K. J. Recovery and resilience of tropical forests after disturbance. *Nature Communications* **5**, 3906 (2014).

[49] Reyer, C. P. O. *et al.* Forest resilience and tipping points at different spatio-temporal scales: approaches and challenges. *Journal of Ecology* **103**, 5–15 (2015).

[50] van Belzen, J. *et al.* Disturbance-recovery experiments to assess resilience of ecosystems along a stress gradient. *Protocol Exchange* (2017).

[51] Walter, J., Jentsch, A., Beierkuhnlein, C. & Kreyling, J. Ecological stress memory and cross stress tolerance in plants in the face of climate extremes. *Environmental and Experimental Botany* **94**, 3–8 (2013).

[52] Nyström, M. & Folke, C. Spatial Resilience of Coral Reefs. *Ecosystems* **4**, 406–417 (2001).

[53] Munné-Bosch, S. & Alegre, L. Cross-stress tolerance and stress "memory" in plants: An integrated view. *Environmental and Experimental Botany* **94**, 1–2 (2013). arXiv:1011.1669v3.

[54] Papagiannopoulou, C. *et al.* Vegetation anomalies caused by antecedent precipitation in most of the world. *Environmental Research Letters* **12**, 074016 (2017).

[55] Rudgers, J. A. *et al.* Climate sensitivity functions and net primary production: A framework for incorporating climate mean and variability. *Ecology* **99**, 576–582 (2018). 0608246v3.

[56] Esper, J., Schneider, L., Smerdon, J. E., Schöne, B. R. & Büntgen, U. Signals and memory in tree-ring width and density data. *Dendrochronologia* **35**, 62–70 (2015).

[57] De Keersmaecker, W. *et al.* Assessment of Regional Vegetation Response to Climate Anomalies: A Case Study for Australia Using GIMMS NDVI Time Series between 1982 and 2006. *Remote Sensing* **9**, 34 (2017).

[58] Kéfi, S. *et al.* Early warning signals of ecological transitions: Methods for spatial patterns. *PLoS ONE* **9**, 10–13 (2014).

[59] Filotas, E. *et al.* Viewing forests through the lens of complex systems science. *Ecosphere* **5** (2014).

[60] Trzaska, S. & Schnarr, E. A review of downscaling methods for climate change projections. Tech. Rep.

September (2014).

[61] ECMWF. Era5: Fifth generation of ECMWF atmospheric reanalyses of the global climate 4. (2019). URL https://cds.climate.copernicus.eu/cdsapp. 2018.

[62] ECMWF. PART I: OBSERVATIONS. In *IFS Documentation CY45R1*, March, 1–76 (2018).

[63] ECMWF. Part II : DATA ASSMILIATION. In *IFS Documentation CY45R1*, vol. March, 1–103 (2018).

[64] ECMWF. PART III: DYNAMICS AND NUMERICAL PROCEDURES. In *IFS Documentation CY45R1*, vol. March, 1–31 (2018).

[65] ECMWF. PART IV: PHYSICAL PROCESSES. In *IFS Documentation CY45R1*, March, 1–189 (2018).

[66] Garris, H. W., Mitchell, R. J., Fraser, L. H. & Barrett, L. R. Forecasting climate change impacts on the distribution of wetland habitat in the Midwestern United states. *Global Change Biology* **21**, 766–776 (2015).

[67] Bruelheide, H. *et al.* Global trait–environment relationships of plant communities. *Nature Ecology & Evolution* **2**, 1906–1917 (2018).

[68] De Boeck, H. J. *et al.* Patterns and drivers of biodiversity-stability relationships under climate extremes. *Journal of Ecology* **106**, 890–902 (2018).

[69] Truchy, A., Angeler, D. G., Sponseller, R. A., Johnson, R. K. & McKie, B. G. Linking Biodiversity, Ecosystem Functioning and Services, and Ecological Resilience. In *Advances in Ecological Research*, vol. 53, 55–96 (Elsevier Ltd., 2015), 1 edn.

[70] Mace, G. M. *et al.* Approaches to defining a planetary boundary for biodiversity. *Global Environmental Change* **28**, 289–297 (2014).

[71] Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: revisting the Holy Grail. *Functional Ecology* **16**, 545–556 (2002).

[72] Salguero-Gómez, R. *et al.* The COMPADRE Plant Matrix Database: an open online repository for plant demography. *Journal of Ecology* **103**, 202–218 (2015).

[73] Salguero-Gómez, R. Applications of the fast-slow continuum and reproductive strategy framework of plant life histories. *New Phytologist* **213**, 1618–1624 (2017).

[74] Salguero-Gómez, R. *et al.* Fast–slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proceedings of the National Academy of Sciences* **113**, 230–235 (2016).

[75] Stott, I., Townley, S. & Hodgson, D. J. A framework for studying transient dynamics of population projection matrix models. *Ecology Letters* **14**, 959–970 (2011).

[76] Caswell, H. *Matrix Population Models: Construction, Analysis, and Interpretation* (Sinauer Associates, Elsevier Inc., 2001).

[77] Ackerly, D. D. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* **164**, S165–S184 (2003).

[78] Laughlin, D. C. The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology* **102**, 186–193 (2014).

[79] Wright, I. J. *et al.* Modulation of leaf economic traits and trait relationships by climate. *Global Ecology and Biogeography* **14**, 411–421 (2005).

[80] Nock, C. A., Vogt, R. J. & Beisner, B. E. Functional Traits. In *eLS*, 1–8 (John Wiley & Sons, Ltd, Chichester, UK, 2016).

[81] Bonser, S. P. Form defining function: interpreting leaf functional variability in integrated plant phenotypes. *Oikos* **114**, 187–190 (2006).

[82] Ordonez, A. & Svenning, J.-C. Geographic patterns in functional diversity deficits are linked to glacial-

interglacial climate stability and accessibility. *Global Ecology and Biogeography* **24**, 826–837 (2015).

[83] Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).

[84] Westoby, M. & Wright, I. J. Land-plant ecology on the basis of functional traits. *Trends in Ecology & Evolution* **21**, 261–268 (2006). arXiv:1011.1669v3.

[85] Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Sciences* **96**, 1463–1468 (1999).

[86] Butler, E. E. *et al.* Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences* **114**, E10937–E10946 (2017).

[87] Violle, C. *et al.* The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution* **27**, 244–252 (2012).

[88] Dantas, V. d. L., Pausas, J. G., Batalha, M. A., de Paula Loiola, P. & Cianciaruso, M. V. The role of fire in structuring trait variability in Neotropical savannas. *Oecologia* **171**, 487–494 (2013).

[89] Kattge, J. *et al.* TRY - a global database of plant traits. *Global Change Biology* **17**, 2905–2935 (2011).

[90] Capdevila, P. *et al.* Life history strategies across realms : marine animals are slower but more reproductive than terrestrial ones (2018).

[91] Healy, K., Ezard, T. H. G., Jones, R., Salguero-Gómez, R. & Buckley, Y. M. Beyond the fast-slow continuum in animal life-histories (2018).

[92] Paniw, M., Ozgul, A. & Salguero-Gómez, R. Interactive life-history traits predict sensitivity of plants and animals to temporal autocorrelation. *Ecology Letters* **21**, 275–286 (2018).

[93] Pérez-Harguindeguy, N. *et al.* New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* **61**, 167 (2013).

[94] Gerhard, B. & Kattge, J. Try plant trait data base (2019). URL https://www.try-db.org/TryWeb/About.php.

[95] Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M. & Thiers, B. The Botanical Information and Ecology Network (BIEN): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. Tech. Rep. (2009).

[96] Botanical information and ecology network (2019). URL http://bien.nceas.ucsb.edu/bien/biendata/.

[97] COMPADRE. Compadre plant matrix data base (2019). URL https://www.compadre-db.org/Data/Compadre.

[98] GBIF. Gbif - global biodiversity information facility (2019). URL https://www.gbif.org/.

[99] Sandel, B. *et al.* Estimating themissing species bias in plant trait measurements. *Journal of Vegetation Science* **26**, 828–838 (2015).

[100] Newbold, T., Butchart, S. H. M., Şekercioğlu, Ç. H., Purves, D. W. & Scharlemann, J. P. W. Mapping Functional Traits: Comparing Abundance and Presence-Absence Estimates at Large Spatial Scales. *PLoS ONE* **7**, e44019 (2012).

[101] Prestele, R. *et al.* Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison. *Global Change Biology* **22**, 3967–3983 (2016).

[102] Peterson, G. D. Contagious Disturbance, Ecological Memory, and the Emergence of Landscape Pattern. *Ecosystems* **5**, 329–338 (2002).

[103] Bell, D. M., Cohen, W. B., Reilly, M. & Yang, Z. Visual interpretation and time series modeling of Landsat imagery highlight drought's role in forest canopy declines. *Ecosphere* **9**, e02195 (2018).

[104] Rivas-Martínez, S., Rivas-Sáenz, S. & Penas, Á. Worldwide bioclimatic classification system. *Global Geobotany* **1**, 1–638 (2011).

[105] Bivand, R., Keitt, T. & Rowlingson, B. *rgdal: Bindings for the Geospatial Data Abstraction Library* (2016).

URL https://CRAN.R-project.org/package=rgdal. R package version 1.4-2.

[106] Rivas-Martínez, S., Rivas-Sáenz, S. & Penas, Á. Computerized bioclimatic maps of the world - ombrotypes of europe and the mideast (2019). URL http://www.globalbioclimatics.org/cbm/static/conti/Europe_Ombrotypes_gb.png.

[107] Rivas-Martínez, S., Rivas-Sáenz, S. & Penas, Á. Computerized bioclimatic maps of the world - ombrotypes of south america (2019). URL http://www.globalbioclimatics.org/cbm/static/conti/Europe_Ombrotypes_gb.png.

[108] Rivas-Martínez, S., Rivas-Sáenz, S. & Penas, Á. Computerized bioclimatic maps of the world - ombrotypes of australasia (2019). URL http://www.globalbioclimatics.org/cbm/static/conti/Europe_Ombrotypes_gb.png.

[109] Nioti, F., Xystrakis, F., Koutsias, N. & Dimopoulos, P. A Remote Sensing and GIS Approach to Study the Long-Term Vegetation Recovery of a Fire-Affected Pine Forest in Southern Greece. *Remote Sensing* **7**, 7712–7731 (2015).

[110] Prince, S. D. & Goward, S. N. Global Primary Production: A Remote Sensing Approach. *Journal of Biogeography* **22**, 815 (1995).

[111] Garroutte, E., Hansen, A. & Lawrence, R. Using NDVI and EVI to Map Spatiotemporal Variation in the Biomass and Quality of Forage for Migratory Elk in the Greater Yellowstone Ecosystem. *Remote Sensing* **8**, 404 (2016).

[112] Cammarano, D., Fitzgerald, G., Casa, R. & Basso, B. Assessing the Robustness of Vegetation Indices to Estimate Wheat N in Mediterranean Environments. *Remote Sensing* **6**, 2827–2844 (2014).

[113] Friedl, M. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* **83**, 287–302 (2002).

[114] Evrendilek, F. & Gulbeyaz, O. Deriving Vegetation Dynamics of Natural Terrestrial Ecosystems from MODIS NDVI/EVI Data over Turkey. *Sensors* **8**, 5270–5302 (2008).

[115] Colditz, R. R. *et al.* Generation and analysis of the 2005 land cover map for Mexico using 250m MODIS data. *Remote Sensing of Environment* **123**, 541–552 (2012).

[116] Krishna Prasad, V., Badarinath, K. V. S. & Eaturu, A. Spatial patterns of vegetation phenology metrics and related climatic controls of eight contrasting forest types in India – analysis from remote sensing datasets. *Theoretical and Applied Climatology* **89**, 95–107 (2007).

[117] Stow, D. A. *et al.* Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. *Remote Sensing of Environment* **89**, 281–308 (2004).

[118] Chen, B., Chen, J. M. & Ju, W. Remote sensing-based ecosystem–atmosphere simulation scheme (EASS)—Model formulation and test with multiple-year data. *Ecological Modelling* **209**, 277–300 (2007).

[119] *Classification of Land Cover Using Decision Trees and Multiple Reference Data Sources*, vol. 20-24.

[120] Friedl, M. A. *et al.* MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* **114**, 168–182 (2010).

[121] Devadas, R., Lamb, D. W., Simpfendorfer, S. & Backhouse, D. Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. *Precision Agriculture* **10**, 459–470 (2009).

[122] Fuller, D. O. & Prince, S. D. Rainfall and foliar dynamics in tropical Southern Africa: Potential impacts of global climatic change on savanna vegetation. *Climatic Change* **33**, 69–96 (1996).

[123] Pinzon, J. & Tucker, C. A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series. *Remote Sensing* **6**, 6929–6960 (2014).

[124] Rouse, J. W., Haa, R. H., Schell, J. A. & Deering, D. W. Monitoring vegetation systems in the great plains

with erts. In *Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations.* (1974).

[125] Hao, C., Wu, S. & Xu, C. Comparison of some vegetation indices in seasonal information. *Chinese Geographical Science* **18**, 242–248 (2008).

[126] Matsushita, B., Yang, W., Chen, J., Onda, Y. & Qiu, G. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors* **7**, 2636–2651 (2007).

[127] Higgins, S. I., Buitenwerf, R. & Moncrieff, G. R. Defining functional biomes and monitoring their change globally. *Global Change Biology* **22**, 3583–3593 (2016).

[128] Hijmans, R. J. *raster: Geographic Data Analysis and Modeling* (2016). URL https://CRAN.R-project.org/package=raster. R package version 2.8-19.

[129] Detsch, F. *gimms: Download and Process GIMMS NDVI3g Data* (2016). URL https://CRAN.R-project.org/package=gimms. R package version 1.1.1.

[130] Gazol, A., Camarero, J. J., Sangüesa-Barreda, G. & Vicente-Serrano, S. M. Post-drought Resilience After Forest Die-Off: Shifts in Regeneration, Composition, Growth and Productivity. *Frontiers in Plant Science* **9**, 1–12 (2018).

[131] Pagotto, M. A., Roig, F. A., de Souza Ribeiro, A. & Lisi, C. S. Influence of regional rainfall and Atlantic sea surface temperature on tree-ring growth of *Poincianella pyramidalis*, semiarid forest from Brazil. *Dendrochronologia* **35**, 14–23 (2015).

[132] Smith, A. P. *et al.* Shifts in pore connectivity from precipitation versus groundwater rewetting increases soil carbon loss after drought. *Nature Communications* **8**, 1335 (2017).

[133] CGIAR-CSI. Nasa shuttle radar topographic mission (srtm) (2018). URL http://srtm.csi.cgiar.org/. 2006.

[134] USGS. Gtopo30 – global 30 arc second elevation data (2018). URL https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30. 2002.

[135] Onyutha, C., Tabari, H., Rutkowska, A., Nyeko-Ogiramoi, P. & Willems, P. Comparison of different statistical downscaling methods for climate change rainfall projections over the Lake Victoria basin considering CMIP3 and CMIP5. *Journal of Hydro-environment Research* **12**, 31–45 (2016).

[136] Maraun, D. *et al.* Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* **48**, 1–38 (2010).

[137] Fowler, H. J., Blenkinsop, S. & Tebaldi, C. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* **27**, 1547–1578 (2007). joc.1556.

[138] Hengl, T. *A Practical Guide to Geostatistical Mapping*, vol. 44 (2011). arXiv:1011.1669v3.

[139] Le Riche, R., Picheny, V., Meyer, A., Kim, N.-H. & Ginsbourger, D. Gears Design with Shape Uncertainties Using Controlled Monte Carlo Simulations and Kriging (2012).

[140] Lichtenstern, A. Kriging methods in spatial statistics. *TUM Media Online* 97 (2013).

[141] Goldsmith, G. R. *et al.* Spatial variation in throughfall, soil, and plant water isotopes in a temperate forest. *Ecohydrology* e2059 (2018).

[142] Zuur, A. F., Ieno, E. N. & Smith, G. M. *Analysing Ecological Data*, vol. 75 of *Statistics for Biology and Health* (Springer New York, New York, NY, 2007). arXiv:1305.6995v1.

[143] Cailleret, M. *et al.* A synthesis of radial growth patterns preceding tree mortality. *Global Change Biology* **23**, 1675–1690 (2017).

[144] Legendre, P. & Legendre, L. *Numerical Ecology*, vol. 24 (Elsevier Science, 1998), 2 edn.

[145] *User guide to the COMPADRE Plant Matrix Database* (2017).

[146] Jongejans, E. & de Kroon, H. Matrix Models. In *Encyclopedia of Theoretical Ecology*, 1–14.

[147] de Kroon, H. *et al.* Elasticities: A Review of Methods and Model Limitations. *Ecology* **81**, 607–618 (2000).

[148] GBIF. Gbif occurences (2019). URL https://www.gbif.org/occurrence/map.

[149] Laughlin, D. C., Strahan, R. T., Adler, P. B. & Moore, M. M. Survival rates indicate that correlations between community-weighted mean traits and environments can be unreliable estimates of the adaptive value of traits. *Ecology Letters* **21**, 411–421 (2018).

[150] Sohrabi, M. M., Ryu, J. H., Abatzoglou, J. & Tracy, J. Development of Soil Moisture Drought Index to Characterize Droughts. *Journal of Hydrologic Engineering* **20**, 04015025 (2015).

[151] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2016). URL https://www.R-project.org/.

[152] Calaway, R., Weston, S. & Tenenbaum, D. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package* (2015). URL https://CRAN.R-project.org/package=doParallel. R package version 1.0.10.

[153] Hiemstra, P. *automap: Automatic interpolation package*. URL https://cran.r-project.org/web/packages/automap/index.html. R package version 1.0.-14.

[154] Calaway, R., Microsoft & Weston, S. *foreach: Provides Foreach Looping Construct for R*. URL https://cran.r-project.org/web/packages/foreach/index.html. R package version 1.4.4.

[155] Rico, A. J. *gameofthrones: "Palettes Inspired in the TV Show "Game of Thrones""* (2018). URL https://github.com/aljrico/gameofthrones. R package version 1.0.0.

[156] Wickham, H. *et al.* *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (2016). URL https://cran.r-project.org/web/packages/ggplot2/index.html. R package version 3.3.0.

[157] Pierce, D. *ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files* (2016). URL https://cran.r-project.org/web/packages/ncdf4/index.html. R package version 1.16.1.

[158] Borchers, H. W. *pracma: Practical Numerical Math Functions* (2016). URL https://cran.r-project.org/web/packages/pracma/index.html. R package version 2.2.2.

[159] Chamberlain, S. *et al.* *rgbif: Interface to the Global 'Biodiversity' Information Facility API* (2016). URL https://cran.r-project.org/web/packages/rgbif/index.html. R package version 1.2.0.

[160] Pebesma, E. *et al.* *sp: Classes and methods for spatial data in R* (2016). URL https://CRAN.R-project.org/package=sp. R package version 1.3-1.

[161] Oksanen, J. *et al.* *vegan: Community Ecology Package* (2016). URL https://cran.r-project.org/web/packages/vegan/index.html. R package version 2.5-4.

[162] Dragulescu, A. A. & Arendt, C. *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files* (2016). URL https://cran.r-project.org/web/packages/xlsx/index.html. R package version 0.6.1.

Appendix

A.1 Project Requirements

Following these steps ensures full reproducibility of the entire analysis. Alternatively, codes can be retrieved via https://github.com/ErikKusch/M.Sc.-Thesis.

With the exception of python, bash, and matlab scripts to obtain ERA5 data, have been carried out in `R`[151] using the functionality contained within the base installation as well as the packages outlined in table A.1. Chunk 1 contains the `R` commands used to install all packages needed to reproduce the analyses of this study from a base installation of `R`. Data sources and methods of data retrieval are stated for each data set individually. All codes needed to reproduce the analysis at the core of this thesis have been included to this document in section A.3. Note that the computation can be sped up by altering the `Cores` argument to enable parallel processing via the `foreach` and `doparallel`[152] packages.

A.1.1 `R` Requirements

Packages

Table A.1: R Packages - Packages which need to be loaded into R to fully reproduce the analyses within this study.

Package	Version	Use
automap[153]	1.0-14	Statistical Downscaling of ERA5 5 data
doParallel[152]	1.0.14	Paralell processing
foreach[154]	1.4.4	Paralell processing
gameofthrones[155]	1.0.0	Data visualisation
ggplot2[156]	3.1.0	Data visualisation
gimms[129]	1.1.1	Downloading GIMMs NDVI3g data
ncdf4[157]	1.16.1	Namespace for NetCDF files
pracma[158]	2.2.2	Detrending time series
raster[128]	2.8-19	Rasterising NetCDF data
rgbif[159]	1.2.0	Downloading floral occurence data
rgdal[105]	1.4-2	Loading and using shapefiles
sp[160]	1.3-1	Converting point data to rasterised data
vegan[161]	2.5-4	PCA approach for model building
xlsx[162]	0.6.1	Export of numeric results

Chunk 1: Installing and loading of R packages needed to reproduce the analyses of this study.

```
install.load.package <- function(x) {
    if (!require(x, character.only = TRUE))
        install.packages(x, repos = "http://cran.us.r-project.org")
    require(x, character.only = TRUE)
}
package_vec <- c("automap", "doParallel", "foreach", "gameofthrones", "ggplot2",
    "gimms", "gridExtra", "ndcf4", "pracma", "raster", "rgbif", "rgdal", "sp", "vegan",
    "xlsx")
sapply(package_vec, install.load.package)
```

Project Directories

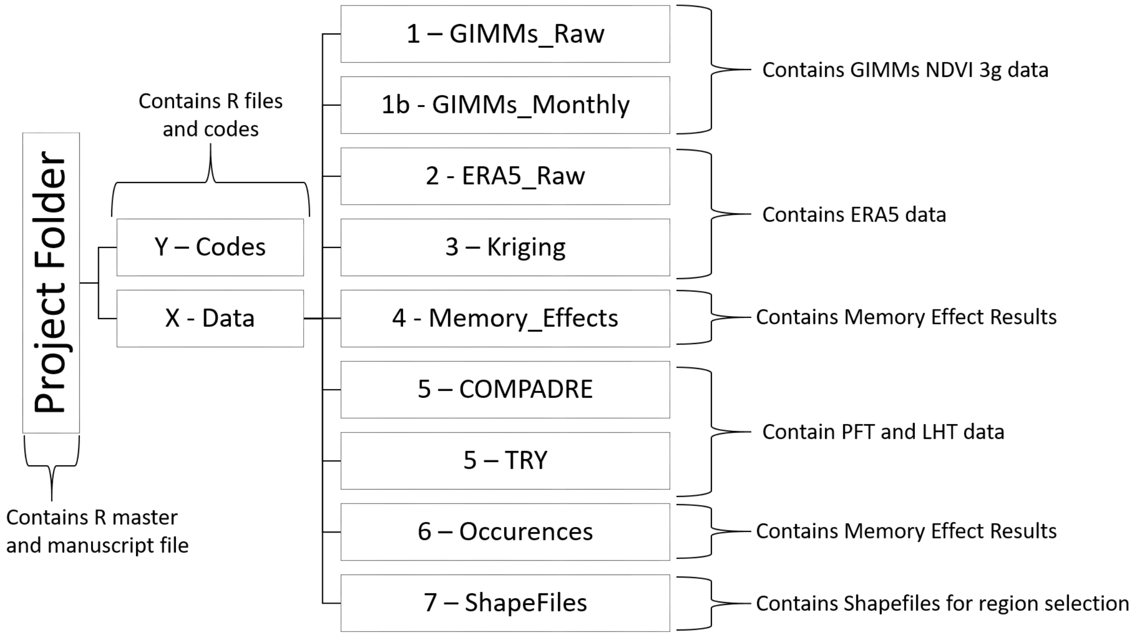


Figure A.1: Working Directories - An overview of working directories established/required by Chunk 2.

Chunk 2: Identifing and generating the folder structure used within the computational steps. A visual overview is presented in A.1.

```r
mainDir <- getwd() # extract the project folder location
# WORKING DIRECTORY FOR CODES
Dir.Codes <- paste(mainDir, "/Y - Codes", sep="")
# WORKING DIRECTORY FOR DATA
Dir.Data <- paste(mainDir, "/X - Data", sep="")
# WORKING DIRECTORY FOR RAW GIMMS DATA
Dir.Gimms <- paste(Dir.Data, "/1 - GIMMs_Raw", sep="")
if(!dir.exists(Dir.Gimms)){dir.create(Dir.Gimms)}
# WORKING DIRECTORY FOR PROCESSED GIMMS DATA
Dir.Gimms.Monthly <- paste(Dir.Data, "/1b - GIMMs_Monthly", sep="")
if(!dir.exists(Dir.Gimms.Monthly)){dir.create(Dir.Gimms.Monthly)}
# WORKING DIRECTORY FOR RAW ERA5 DATA
Dir.ERA <- paste(Dir.Data, "/2 - ERA5_Raw", sep="")
# WORKING DIRECTORY FOR PROCESSED ERA5 DATA
Dir.ERA.Monthly <- paste(Dir.Data, "/2b - ERA5_Monthly", sep="")
if(!dir.exists(Dir.ERA.Monthly)){dir.create(Dir.ERA.Monthly)}
# WORKING DIRECTORY FOR KRIGING COVARIATES
Dir.KrigCov <- paste(Dir.Data, "/3 - Kriging", sep="")
# WORKING DIRECTORY FOR MEMORY EFFECT DATA
Dir.Memory <- paste(Dir.Data, "/4 - Memory_Effects", sep="")
# WORKING DIRECTORY FOR COMPADRE DATA
Dir.Compadre <- paste(Dir.Data, "/5 - COMPADRE", sep="")
# WORKING DIRECTORY FOR TRY PFT DATA
if(!dir.exists(Dir.Memory)){dir.create(Dir.Memory)}
Dir.TRY <- paste(Dir.Data, "/5 - TRY", sep="")
# WORKING DIRECTORY FOR OCCURENCE DATA
Dir.OCCs <- paste(Dir.Data, "/6 - Occurences", sep="")
if(!dir.exists(Dir.OCCs)){dir.create(Dir.OCCs)}
# WORKING DIRECTORY FOR SHAPEFILES (contains masking file for water bodies)
Dir.Mask <- paste(Dir.Data, "/7 - ShapeFiles", sep="")
```

Functions for Region Selection and Raster Names

Chunk 3: User-defined functions used to optimise further functions of this analysis when (1) Limitting to study regions (RegionSelection), and (2) assigning names to model rasters (Fun__NamesRas).

```r
### RegionSelection [Region, RegionFile, Extent] (selecting region and extent from
### shapefiles)
RegionSelection <- function(Region, RegionFile, Extent) {
  ## loading shapefiles
  Shapes <- readOGR(Dir.Mask, "ne_50m_admin_0_countries", verbose = FALSE)
  ## selecting region from shapefile run global analysis read user-defined extent
  ## (if applicable)
  if (Region == "Global") {
    if (is.null(Extent)) {
      area <- extent(-180, 180, -90, 90)
    } else {
      area <- Extent
    }
    location <- 1:length(Shapes) # selecting all countries contained within the shapefile
  } else {
    Where <- Region # countries to consider
    location <- NA # position vector in shapefile list
    for (i in 1:length(Where)) {
      # select region from Shapefiles
      location[i] <- which(as.vector(Shapes$NAME) == Where[i])
    }
    if (is.null(Extent)) {
      # read user-defined extent (if applicable)
      area <- extent(Shapes[location, ])
    } else {
      area <- Extent
    }
  }
  if (is.null(RegionFile)) {
    # if no file name has been specified
    RegionFile <- toString(Region) # take name of region
  }
  # returning parameters
  return(list(area, location, RegionFile))
}

### Fun_NamesRas [raster, ClimVar, ClimVar2]
# (assigning layer names to model rasters) ----
Fun_NamesRas <- function(raster, ClimVar, ClimVar2, rasiter = 1){
  names(raster) <- c(paste("Most informative", ClimVar[[rasiter]],"lag", sep=" "),
                     "Model AICs", "Model p-value", "Antecedent NDVI (c_NDVI)",
                     paste("Antecedent" , ClimVar[[rasiter]], "(c_clim)", sep=" "),
                     paste("Antecedent" , ClimVar2[[rasiter]], "(c_clim2)", sep=" "),
                     "Explained Variance", "Variance (NDVI)", "Variance (Shared)",
                     paste("Variance (" , ClimVar[[rasiter]], ")", sep=""))
  return(raster)} # Fun_NamesRas end
```

A.1.2 ERA5 Data

Obtaining ERA5 Data

Chunk 4: Obtaining ERA5 data from the ECMW servers. Substitute the date statement to download the full data set. Python script.

```
import cdsapi

c = cdsapi.Client()
c.retrieve('reanalysis-era5-complete', {    # do not change this!
    'class'   : 'ea',
    'expver'  : '1',
    'stream'  : 'moda',
    'type'    : 'an',
    'param'   : '39.128/40.128/41.128/42.128/167.128',
    'levtype' : 'sfc',
    'date'    : '19800101/19800201/19800301/19800401/19800501/19800601/19800701/19800801/19800901/19801001/19801101/
19801201/19810101/19810201/19810301/19810401/19810501/19810601/19810701/19810801/19810901/19811001/19811101/
19811201/19820101/19820201/19820301/19820401/19820501/19820601/19820701/19820801/19820901/19821001/19821101/
19821201/19830101/19830201/19830301/19830401/19830501/19830601/19830701/19830801/19830901/19831001/19831101/
19831201/19840101/19840201/19840301/19840401/19840501/19840601/19840701/19840801/19840901/19841001/19841101/
19841201/19850101/19850201/19850301/19850401/19850501/19850601/19850701/19850801/19850901/19851001/19851101/
19851201/19860101/19860201/19860301/19860401/19860501/19860601/19860701/19860801/19860901/19861001/19861101/
19861201/19870101/19870201/19870301/19870401/19870501/19870601/19870701/19870801/19870901/19871001/19871101/
19871201/19880101/19880201/19880301/19880401/19880501/19880601/19880701/19880801/19880901/19881001/19881101/
19881201/19890101/19890201/19890301/19890401/19890501/19890601/19890701/19890801/19890901/19891001/19891101/19891201',
    'decade'  : '1980',
}, 'All_1980s.grib')
```

Pre-Processing ERA5 Data

Chunk 5: Processing ERA5 data from local downloads to merge by time step. Bash script.

```
#!/bin/bash

. /opt/.profiles/cdo.bash

# Convert files to netcdf format
cdo -f nc setgridtype,regular 'All_1980s.grib' 'All_1980s.nc'
cdo -f nc setgridtype,regular 'All_1990s.grib' 'All_1990s.nc'
cdo -f nc setgridtype,regular 'All_2000s.grib' 'All_2000s.nc'
cdo -f nc setgridtype,regular 'All_2010s.grib' 'All_2010s.nc'

# Change variable names to identifiable ones
cdo chname,var167,Tair,var39,Qsoil1,var40,Qsoil2,var41,Qsoil3,var42,Qsoil4 'All_1980s.nc' 'All_1980s_renamed.nc'
cdo chname,var167,Tair,var39,Qsoil1,var40,Qsoil2,var41,Qsoil3,var42,Qsoil4 'All_1990s.nc' 'All_1990s_renamed.nc'
cdo chname,var167,Tair,var39,Qsoil1,var40,Qsoil2,var41,Qsoil3,var42,Qsoil4 'All_2000s.nc' 'All_2000s_renamed.nc'
cdo chname,var167,Tair,var39,Qsoil1,var40,Qsoil2,var41,Qsoil3,var42,Qsoil4 'All_2010s.nc' 'All_2010s_renamed.nc'

# Merge the files over time
cdo -r mergetime `ls *'_renamed.nc'` 'All_variables_1980_2016.nc'

# Cleanup
rm *renamed.nc
```

Chunk 6: Regridding ERA5 data to fit GIMMS and HWSD extents. Matlab script.

```
clear
clc
fn              = 'All_variables_1980_2016.nc';
ncin            = netcdf.open(fn,'NC_NOWRITE');
```

```
lon_orig        = double(netcdf.getVar(ncin,0));
lat             = double(netcdf.getVar(ncin,1));
time            = double(netcdf.getVar(ncin,4));
Tair_orig(:,:,:)   = double(netcdf.getVar(ncin,8));
Qsoil1_orig(:,:,:) = double(squeeze(netcdf.getVar(ncin,5)));
Qsoil1_orig(:,:,:) = double(squeeze(netcdf.getVar(ncin,5)));
Qsoil1_orig(:,:,:) = double(squeeze(netcdf.getVar(ncin,5)));
Qsoil1_orig(:,:,:) = double(squeeze(netcdf.getVar(ncin,5)));
netcdf.close(ncin)

lon_orig(lon_orig>180)=lon_orig(lon_orig>180)-360; % Convert longitude to the -180 to 180 reference

% Re-organise by longitude
[lon,I]   = sort(lon_orig);
Qsoil1 = Qsoil1_orig(I,:,:);
Qsoil2 = Qsoil2_orig(I,:,:);
Qsoil3 = Qsoil3_orig(I,:,:);
Qsoil4 = Qsoil4_orig(I,:,:);
Tair   = Tair_orig(I,:,:);

clear Qsoil1_orig Qsoil2_orig Qsoil3_orig Qsoil4_orig Tair_orig

%%
for ivar = 1:5
    if ivar==1; var = Qsoil1; VarName='Qsoil1'; end
    if ivar==2; var = Qsoil2; VarName='Qsoil2'; end
    if ivar==3; var = Qsoil3; VarName='Qsoil3'; end
    if ivar==4; var = Qsoil4; VarName='Qsoil4'; end
    if ivar==5; var = Tair;   VarName='Tair';   end

    fn = [Variables{ivar},'_TrainingResolution.nc'];
    [lon_dim,lat_dim,tim_dim] = size(Tair);
    ncout = netcdf.create(fn,'CLOBBER');
    lonID = netcdf.defDim(ncout,'lon',lon_dim);
    latID = netcdf.defDim(ncout,'lat',lat_dim);
    timID = netcdf.defDim(ncout,'time',tim_dim);
    varid1  = netcdf.defVar(ncout,'lon',      'nc_float',[lonID]); %#ok<*NBRAK>
    varid2  = netcdf.defVar(ncout,'lat',      'nc_float',[latID]);
    varid3  = netcdf.defVar(ncout,'time',     'nc_float',[timID]);
    varid5  = netcdf.defVar(ncout,VarName,    'nc_float',[lonID,latID,timID]);
    % Put the attributes for the lon dimension
    netcdf.putAtt(ncout,varid1,'units','degrees_east');
    netcdf.putAtt(ncout,varid1,'standard_name','longitude');
    % Put the attributes for the lat dimension
    netcdf.putAtt(ncout,varid2,'units','degrees_north');
    netcdf.putAtt(ncout,varid2,'standard_name','latitude');
    % Put the attributes for the Time
    netcdf.putAtt(ncout,varid3,'units','hours since 2008-01-01 06:00:00');
    netcdf.putAtt(ncout,varid3,'calendar','proleptic_gregorian');
    netcdf.putAtt(ncout,varid3,'standard_name','time');
    netcdf.endDef(ncout)
    netcdf.putVar(ncout,varid1, lon)
    netcdf.putVar(ncout,varid2, lat)
    netcdf.putVar(ncout,varid3, time)
    netcdf.putVar(ncout,varid5, var)
    netcdf.close(ncout)
end
```

A.2 Data

A.2.1 HWSD Data

Table A.2: HWSD Variables and Explanations - HWSD variables used within this study and their explanations. Slope aspects and inclines are recorded as number of 3 arc-second cells falling into 5 minute cells

HWSD Variable	**Explanation**
Elevation	Altitude in metres as measured from sea-level
Slope_aspect_N	$(0°; 45°]$; $(315°; 360°]$
Slope_aspect_E	$(45°; 135°]$
Slope_aspect_S	$(135°; 225°]$
Slope_aspect_W	$(225°; 315°]$
Slope_aspect_U	Slope undefined or slope incline is less than 2%
Slopes1	$0\% \leq$ incline $\leq 0.5\%$
Slopes2	$0.5\% \leq$ incline $\leq 2\%$
Slopes3	$2\% \leq$ incline $\leq 5\%$
Slopes4	$5\% \leq$ incline $\leq 10\%$
Slopes5	$10\% \leq$ incline $\leq 15\%$
Slopes6	$15\% \leq$ incline $\leq 30\%$
Slopes7	$30\% \leq$ incline $\leq 45\%$
Slopes8	Incline $> 45\%$

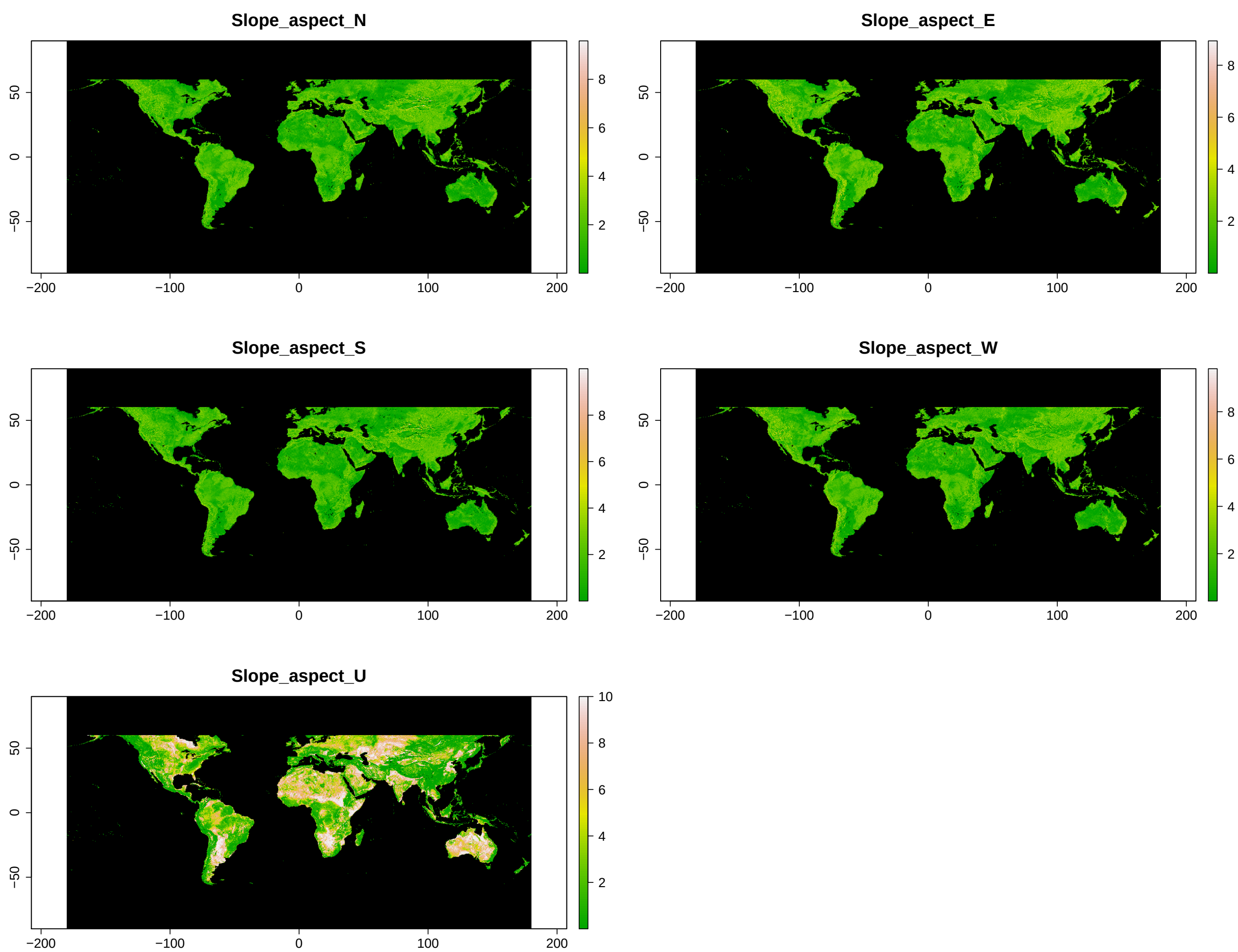


Figure A.2: HWSD slope aspect data - HWSD data at GIMMS resolution divided by 1000 for displaying. Figure established via Chunk 18.

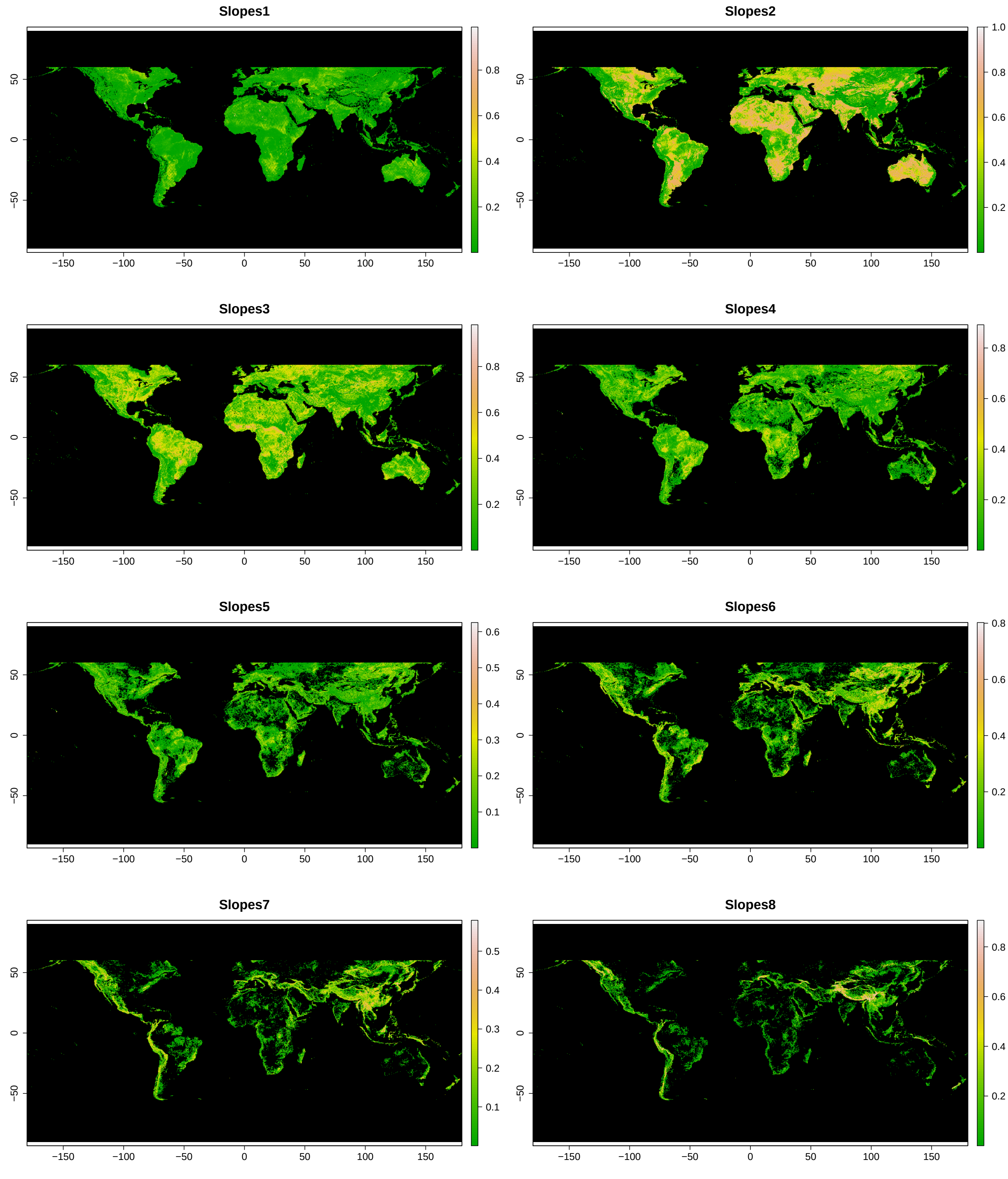


Figure A.3: HWSD slope incline data - HWSD data at GIMMS resolution. Data has been divided by 1000 for displaying. Figure established via Chunk 19.

A.2.2 Study Regions

A.2.2.1 Iberian Region

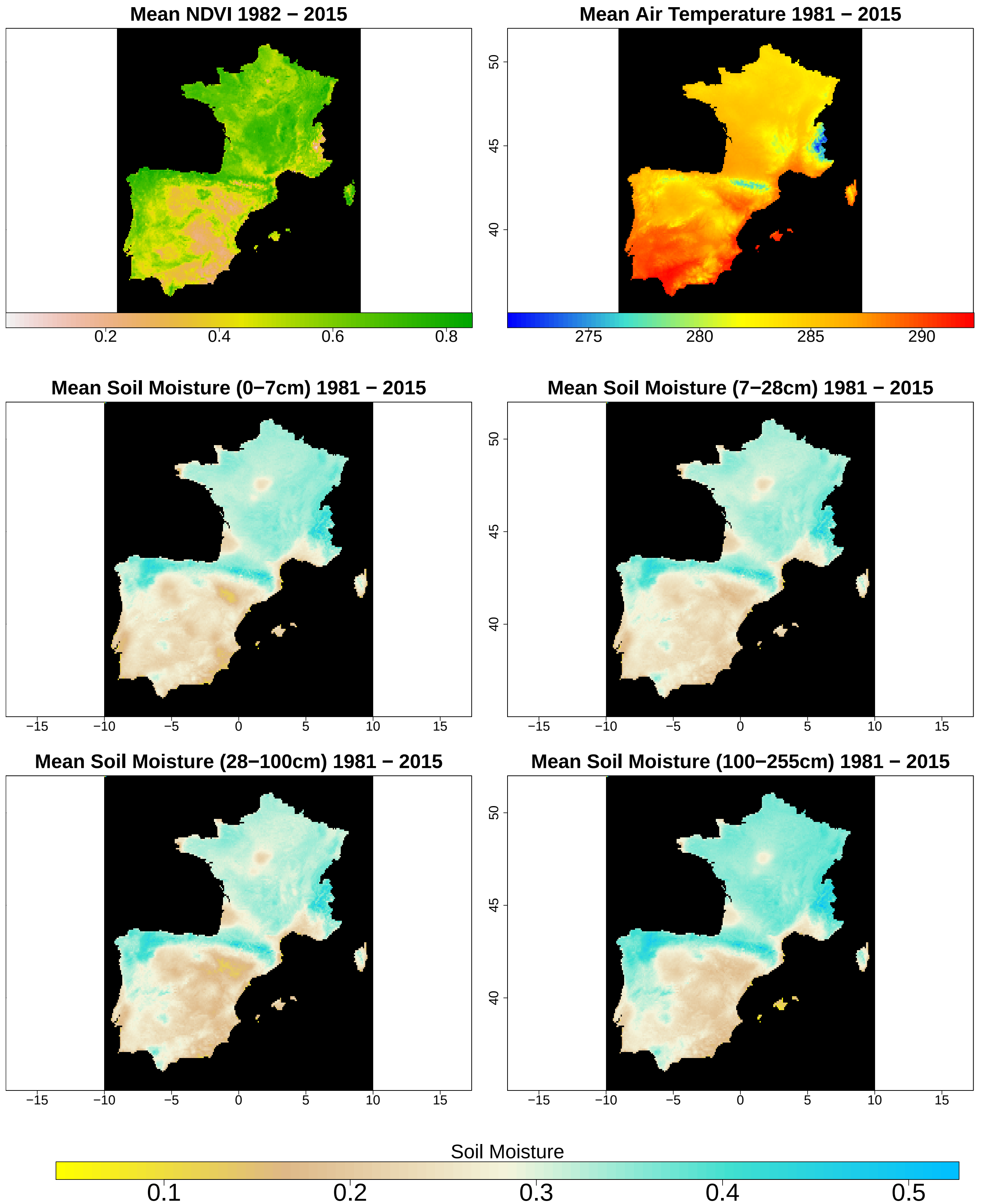


Figure A.4: Data Overview (Iberian Region) - Data required for identificiation of vegetation memory: (1) GIMMS NDVI 3g; and (2) ERA5: Tair, Qsoil1, Qsoil2, Qsoil3, Qsoil4. Figure established via Chunk 20.

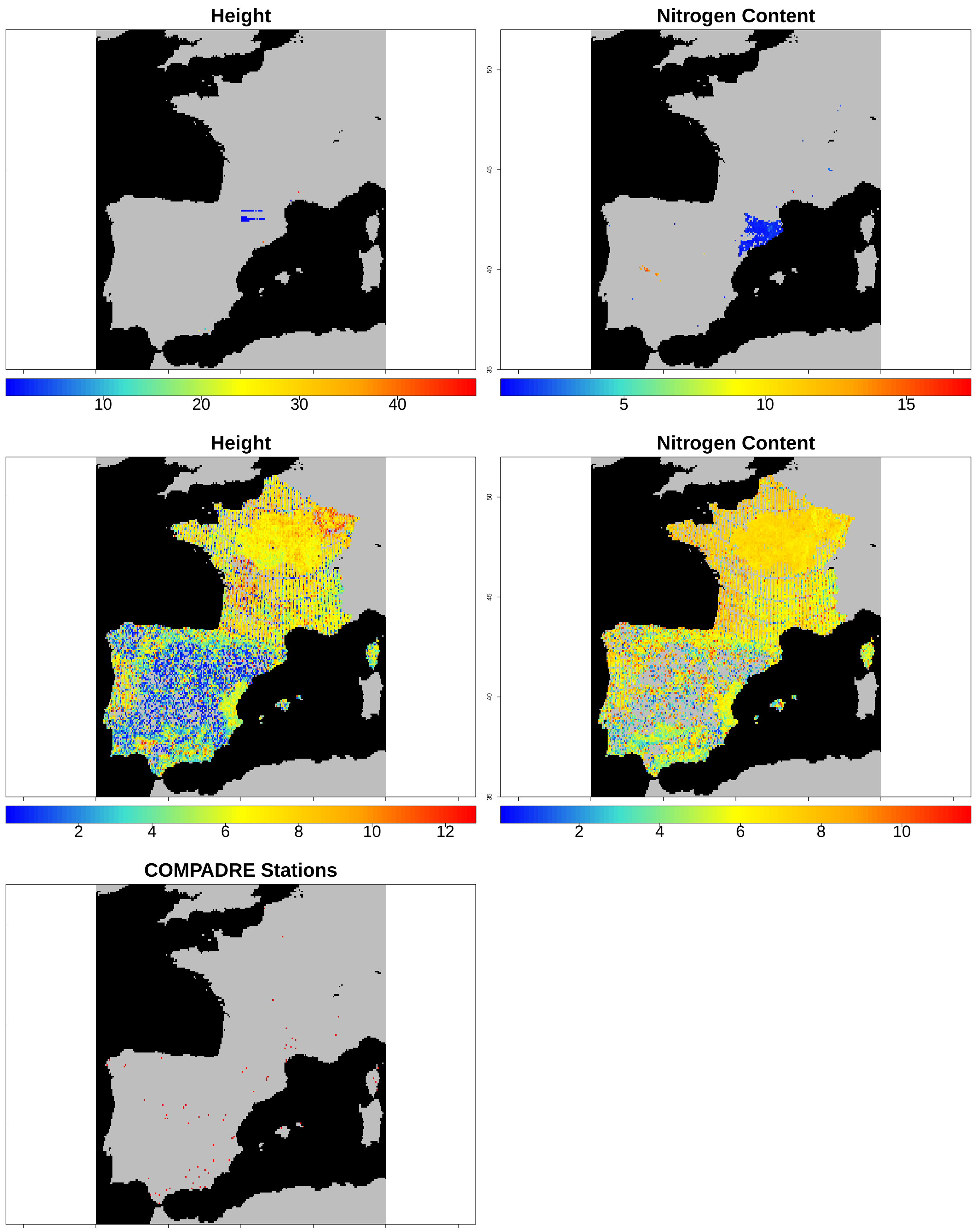


Figure A.5: TRY/COMPADRE Data (Iberian Region) - TRY PFT data across the Iberian Region: (1) Geo-referenced records (upper row), and (2) Extrapolated species-specific mean PFT records (middle row) as well as COMPADRE stations (lower plot). Figure established via Chunk 21.

A.2.2.2 Caatinga, Brazil

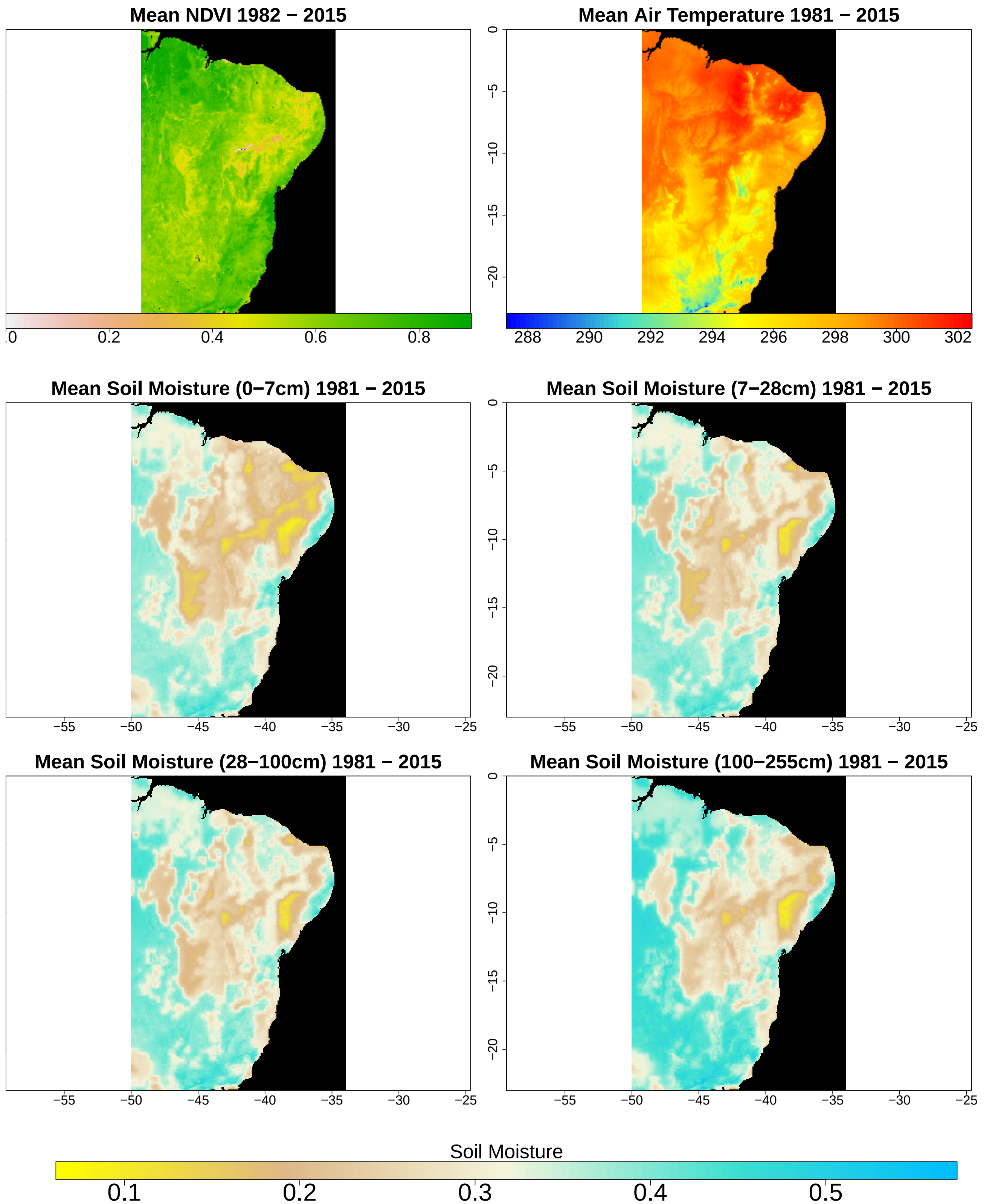


Figure A.6: Data Overview (Caatinga, Brazil) - Data required for identificiation of vegetation memory: (1) GIMMS NDVI 3g; and (2) ERA5: Tair, Qsoil1, Qsoil2, Qsoil3, Qsoil4. Figure established via Chunk 20.

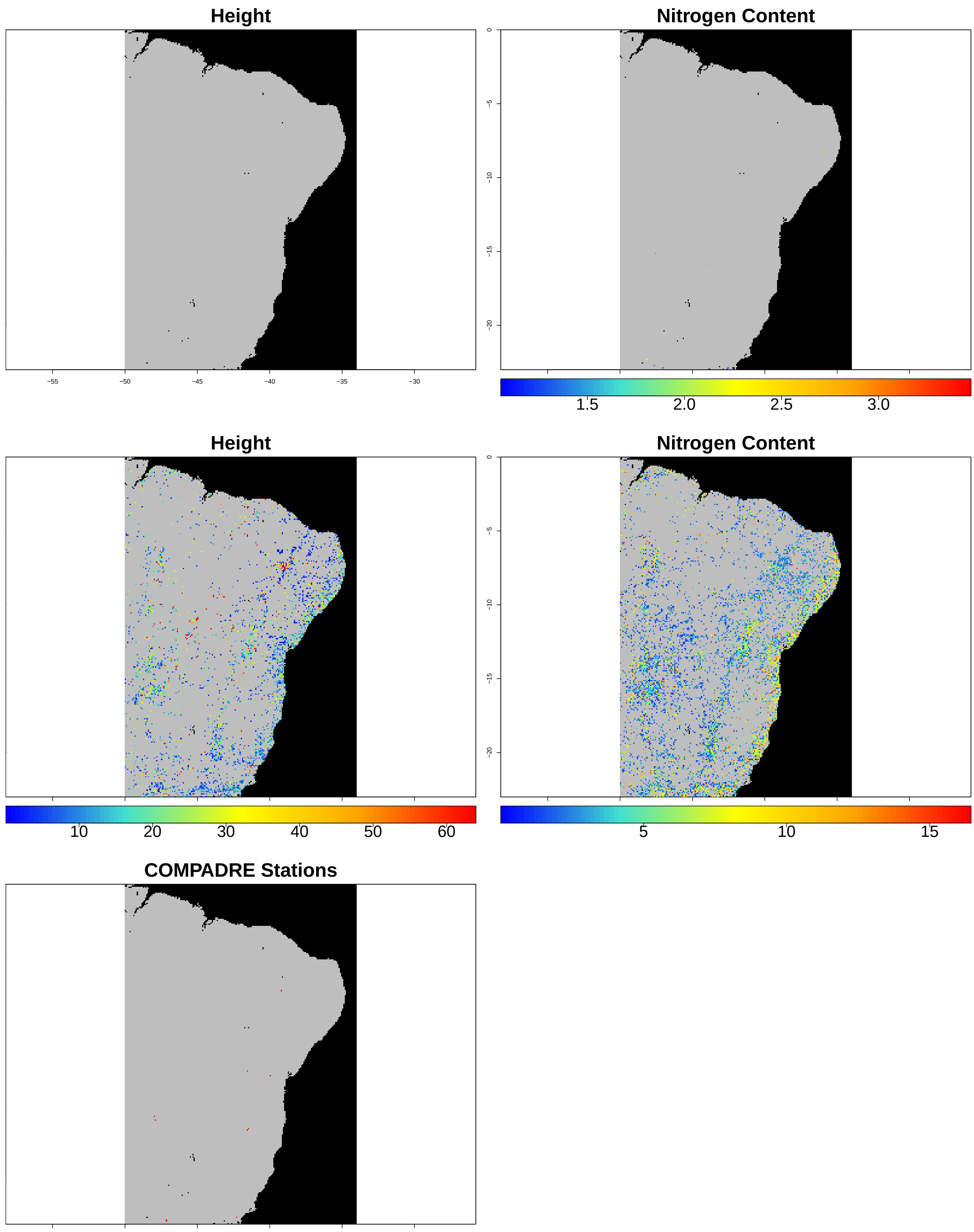


Figure A.7: TRY/COMPADRE Data (Caatinga, Brazil) - TRY PFT data across Caatinga, Brazil: (1) Geo-referenced records (upper row), and (2) Extrapolated species-specific mean PFT records (middle row) as well as COMPADRE stations (lower plot). Figure established via Chunk 21.

A.2.2.3 Australia

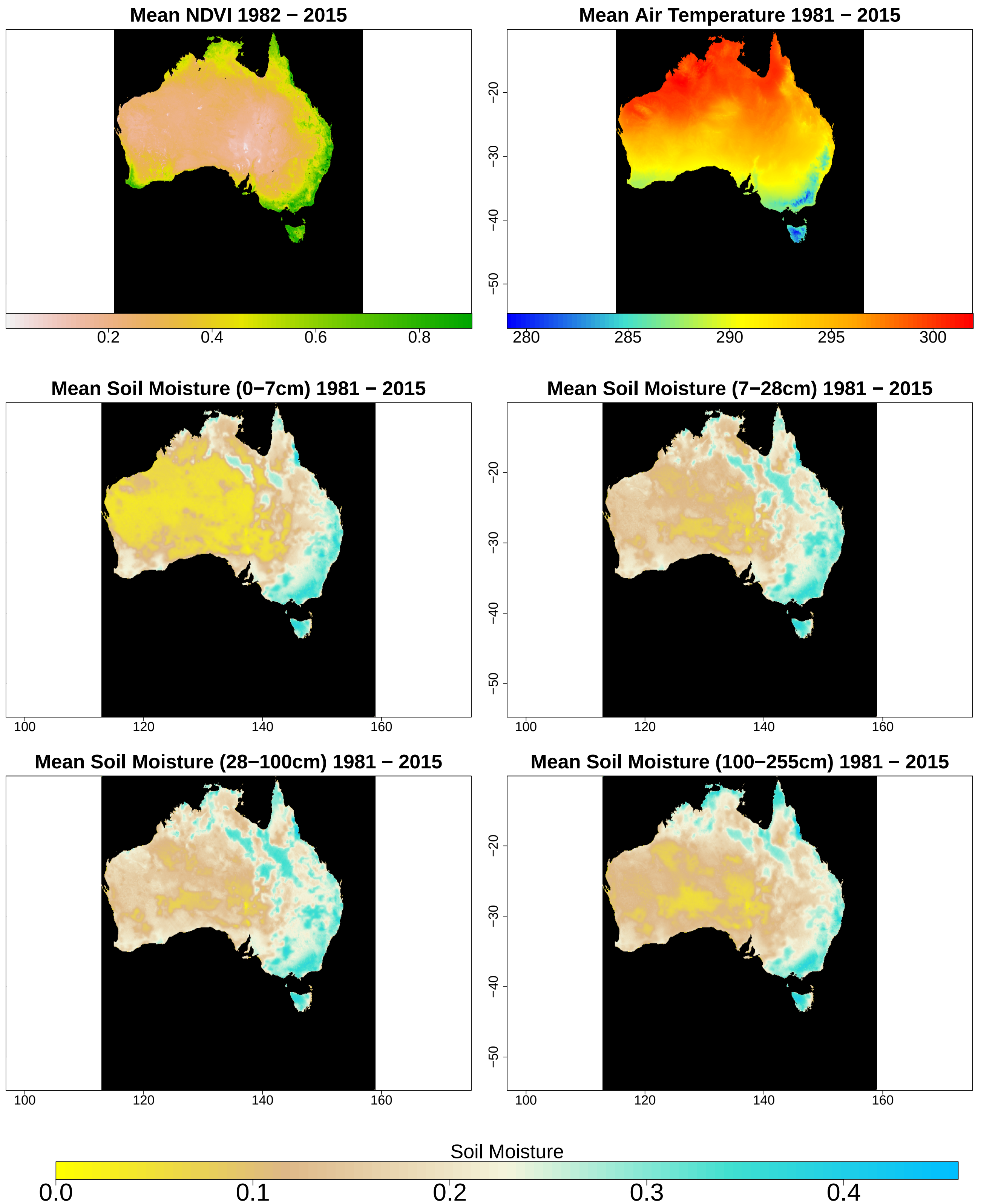


Figure A.8: Data Overview (Australia) - Data required for identificiation of vegetation memory: (1) GIMMS NDVI 3g; and (2) ERA5: Tair, Qsoil1, Qsoil2, Qsoil3, Qsoil4. Figure established via Chunk 20.

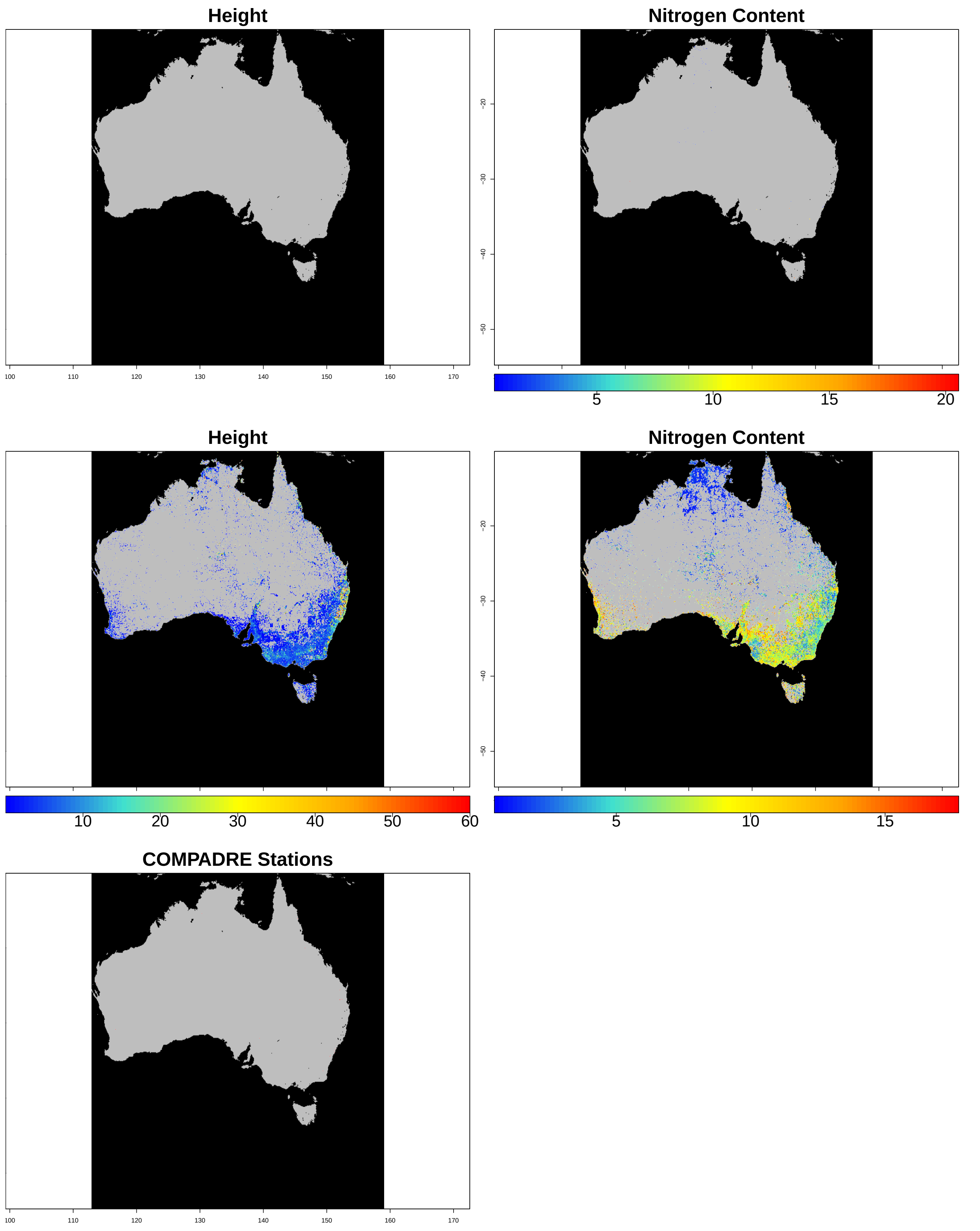


Figure A.9: TRY/COMPADRE Data (Australia) - TRY PFT data across Australia: (1) Geo-referenced records (upper row), and (2) Extrapolated species-specific mean PFT records (middle row) as well as COMPADRE stations (lower plot). Figure established via Chunk 21.

A.3 R Analyses Codes

A.3.1 Master File

Chunk 7: Running the entire analysis once data has been allocated to directories. S0a_Packages.R can be found in Chunk 1, S0b_Directories.R is located in Chunk 2 whilst S0c_Functions.R is contained in Chunk 3. S1_GIMMs.R, S2_ERA5.R, S3_VegetationMemory.R, S4_PFTs.R, and S4_COMPADRE.R are included in Chunk 8, Chunk 9, Chunk 10, Chunk 11, and Chunk 12, respectively.

```
rm(list = ls()) # clearing environment
####--------------- PACKAGES ----
source("Y - Codes/S0a_Packages.R") # loading packages
####--------------- DIRECTORIES ----
source("Y - Codes/S0b_Directories.R") # setting directories
####--------------- FUNCTIONS ----
source("Y - Codes/S0c_Functions.R") # Loading miscellaneous functions
####--------------- VARIABLE VECTORS ----
ModVars <- c("Tair", "Qsoil1", "Qsoil2", "Qsoil3", "Qsoil4")
ClimVars = list("Qsoil1_mean", "Qsoil2_mean", "Qsoil3_mean", "Qsoil4_mean")
ClimVars2 = list("Tair_mean", "Tair_mean", "Tair_mean", "Tair_mean")
###---------------- FUNCTIONS ---------------------------------------------------------
####--------------- Fun_Vegetation [Regions, RegionFiles, Extents, From, To, Lags, Cores]
# (selecting and preparing data, and calculating vegetation memory) ----
Fun_Vegetation <- function(Regions, RegionFiles, Extents, From, To, Lags, Cores) {
  FromY <- (From - ceiling(1/12 * max(Lags))) # Figuring out real start year after factoring in lags
  ##### GIMMs NDVI -----
  print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  print("HANDLING GIMMs NDVI DATA")
  print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  setwd(mainDir)
  source(paste(Dir.Codes, "S1_GIMMs.R", sep="/"))
  # Download NDVI data for full years (from, to), turn it into monthly composite
  # rasters checking if all files are already there and not running this function
  # if they are
  froms <- c(1982, 1986, 1991, 1996, 2001, 2006, 2011)
  tos <- c(1985, 1990, 1995, 2000, 2005, 2010, 2015)
  for(RasGimms in 1:length(tos)){
    if(file.exists(paste(Dir.Gimms.Monthly, "/GlobalNDVI_", froms[RasGimms], tos[RasGimms], ".nc", sep=""))){
      print(paste("Global NDVI raster from", froms[RasGimms], "to", tos[RasGimms], "has already been established."))
      next()
    }else{
      RasterGIMMs(from = froms[RasGimms], to = tos[RasGimms])
    }
  }
  # Load composite NDVI data and limit to extent of a study region saving the
  # resulting data
  for (CombineRun in 1:length(Regions)) {
    # Checking if this particular data has been produced already
    if (file.exists(paste(Dir.Gimms.Monthly, "/NDVI_", RegionFiles[[CombineRun]],
                          ".nc", sep = ""))) {
      print(paste("NDVI raster stack already cropped for: ", RegionFiles[[CombineRun]],
                  sep = ""))
      (next)()
    }
    CombineCDFs(Region = Regions[[CombineRun]], RegionFile = RegionFiles[[CombineRun]],
                Extent = Extents[[CombineRun]])
  } # CombineCDFs
  ##### ERA5 -----
  print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
```

```r
print("HANDLING ERA5 DATA")
print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
setwd(mainDir)
source(paste(Dir.Codes, "S2_ERA5.R", sep="/"))
# kriging ERA5 variable data across study regions for selected time period
# parallel
if (Cores > 1) {
  cl <- makeCluster(Cores) # Assuming X node cluster
  registerDoParallel(cl) # Register cores
  for (KrigRegion in 1:length(Regions)) {
    # looping over regions
    print("##################################################")
    print(paste("Kriging ERA5 ", toString(ModVars), " data from across ",
                RegionFiles[[KrigRegion]], sep = ""))
    foreach(Krigrun = 1:length(ModVars)) %dopar% {
      # looping over variables
      source(paste("Y - Codes", "S0a_Packages.R", sep="/")) # load packages to each core
      source(paste("Y - Codes", "S0b_Directories.R", sep="/")) # set packages for each core
      source(paste(Dir.Codes, "S0c_Functions.R", sep="/")) # Loading miscellaneous functions
      source(paste(Dir.Codes, "S2_ERA5.R", sep="/")) # source function for each core
      ModVars <- c("Tair", "Qsoil1", "Qsoil2", "Qsoil3", "Qsoil4")
      # Checking if this particular data has been kriged already
      if (!file.exists(paste(Dir.ERA.Monthly, "/", ModVars[[Krigrun]],
                             "_mean_", RegionFiles[[KrigRegion]], "_", 1, FromY, "_", 12, To,
                             ".nc", sep = "")))
      {
        sapply(package_vec, install.load.package)
        RasterEra5(Variable = ModVars[[Krigrun]], Region = Regions[[KrigRegion]],
                   RegionFile = RegionFiles[[KrigRegion]], Extent = Extents[[KrigRegion]],
                   FromY = FromY, FromM = 1, ToY = To, ToM = 12, Temporary = "Keep")
      } # check if already kriged
    } # parallel run
  } # Region-loop
  stopCluster(cl) # stop cluster
} else {
  # non-parallel looping over regions looping over variables Checking if this
  # particular data has been kriged already
  for (KrigRegion in 1:length(Regions)) {
    for (Krigrun in 1:length(ModVars)) {
      if (file.exists(paste(Dir.ERA.Monthly, "/", ModVars[[Krigrun]], "_mean_",
                            RegionFiles[[KrigRegion]], "_", 1, FromY, "_", 12, To, ".nc", sep = ""))) {
        print(paste(ModVars[[Krigrun]], " data already kriged for: ", RegionFiles[[KrigRegion]],
                    sep = ""))
        (next)()
      }
      RasterEra5(Variable = ModVars[[Krigrun]], Region = Regions[[KrigRegion]],
                 RegionFile = RegionFiles[[KrigRegion]], Extent = Extents[[KrigRegion]],
                 FromY = FromY, FromM = 1, ToY = To, ToM = 12, Temporary = "Keep")
    } # Variable-loop
  } # region-loop
} # RasterEra5 function
##### VEGETATION MEMORY -----
print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
print("IDENTIFYING VEGETATION MEMORY")
print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
setwd(mainDir)
source(paste(Dir.Codes, "S3_VegetationMemory.R", sep="/"))
# Calculating vegetation memory
```

```r
  `%nin%` = Negate(`%in%`) # create a 'not in' statement
  if (Cores > 1) {
    # parallel
    cl <- makeCluster(Cores) # Assuming X node cluster
    registerDoParallel(cl) # registering cores
    for (MemReg in 1:length(Regions)) {
      # looping over regions
      print("##################################################")
      print(paste("Calculating vegetation memory according to ", toString(ModVars),
                  " across ", RegionFiles[[MemReg]], sep = ""))
      foreach(Memrun = 2:length(ModVars)) %dopar% {
        # looping over variables
        source(paste("Y - Codes", "S0a_Packages.R", sep="/")) # load packages to each core
        source(paste("Y - Codes", "S0b_Directories.R", sep="/")) # register directories with each core
        source(paste(Dir.Codes, "S0c_Functions.R", sep="/")) # Loading miscellaneous functions
        source(paste(Dir.Codes, "S3_VegetationMemory.R", sep="/")) # source functions for each core
        ModVars <- c("Tair", "Qsoil1", "Qsoil2", "Qsoil3", "Qsoil4")
        # check if already computed
        if (paste(RegionFiles[[MemReg]], "_Tair_mean-", ModVars[Memrun],
                  "_mean", paste(Lags, collapse = "_"), "_", FromY, "-", To, ".nc",
                  sep = "") %nin% list.files(Dir.Memory)) {
          VegMem(ClimVar = paste(ModVars[Memrun], "_mean", sep = ""), ClimVar2 = "Tair_mean",
                 Region = RegionFiles[[MemReg]], Cumlags = Lags, FromY = FromY,
                 ToY = To)
        }
      } # parallel loop
    } # region-loop
    stopCluster(cl) # stop cluster
  } else {
    # non-parallel looping over regions looping over coefficients of soil layers
    # check if already computed
    for (MemReg in 1:length(Regions)) {
      for (Memrun in 2:length(ModVars)) {
        if (paste(RegionFiles[[MemReg]], "_Tair_mean-", ModVars[Memrun],
                  "_mean", paste(Lags, collapse = "_"), "_", FromY, "-", To, ".nc",
                  sep = "") %nin% list.files(Dir.Memory)) {
          VegMem(ClimVar = paste(ModVars[Memrun], "_mean", sep = ""), ClimVar2 = "Tair_mean",
                 Region = RegionFiles[[MemReg]], Cumlags = Lags, FromY = FromY,
                 ToY = To)
        } else {
          print(paste("Vegetation memory already computed for", ModVars[Memrun],
                      " across:", RegionFiles[[MemReg]], sep = " "))
        }
      } # memory-loop
    } # region-loop
  } # VegMem function
  # scaling coefficients per region to be represented on fixed scales looping over
  # regions
  for (MemReg in 1:length(Regions)) {
    CoeffScaling(ClimVar = ClimVars, ClimVar2 = ClimVars2, Region = list(RegionFiles[[MemReg]],
                                                                          RegionFiles[[MemReg]],
                                                                          RegionFiles[[MemReg]],
                                                                          RegionFiles[[MemReg]]),
                 Cumlags = list(Lags, Lags, Lags, Lags), FromY = FromY, ToY = To, UAbs = TRUE)
  } # CoeffScaling function
} # Fun_Vegetation
####--------------- Fun_PFTs [Regions, RegionFiles, Extents, From, To, Occ]
# (aggregating PFT data, downloading species occurences, building PFT rasters)
```

```r
# ----
Fun_PFTs <- function(Regions, RegionFiles, Extents, From, To, Occ) {
  source(paste(Dir.Codes, "S4_PFTs.R", sep="/"))
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  print("CALCULATING SPECIES SPECIFIC-TRAIT MEANS")
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  # calculating species-specific trait means
  if (!file.exists(paste(Dir.TRY, "/SpeciesTraits.RData", sep = ""))) {
    PFTs() # species-specific trait means
  } else {
    print("Species-specific trait means already calculated")
  }
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  print("OBTAINING SPECIES OCCURENCE RECORDS FROM GBIF")
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  # figuring out ISO3166 codes for selected regions
  ISO3166_df <- read.csv(paste(Dir.Mask, "/ISO3166.csv", sep = "")) # read ISO3166 country code list
  CountCodes <- ISO3166_df$ISO.3166.ALPHA.2[which(ISO3166_df$Country %in% unlist(Regions))]
  CountCodes <- toString(CountCodes)
  CountCodes <- gsub(pattern = ", ", replacement = ";", x = CountCodes)
  # Download occurence files from GBIF
  if (Occ == "Download") {
    GbifStat <- NULL
    # species occurences
    Attempt <- 0
    while (is.null(GbifStat) || GbifStat != "Done") {
      if (Attempt > 0) {
        print("Encountered an error and starting the downloading of GBIF occurence data again.
              This is usually due to issues with the GBIF server connection and you don't have
              to worry as long as your internet connection is stable.")
      }
      Attempt <- Attempt + 1
      try(GbifStat <- DistMaps(Species = "All", Years = From:To, CountCodes = CountCodes))
      }
    } else {
      print("Occurence data will not be downloaded according to function call.")
    }
  # generate mean rasters of PFTs across specifiec regions looping over regions
  for (CompReg in 1:length(Regions)) {
    if (!file.exists(paste(Dir.TRY, "/TRY-", RegionFiles[[CompReg]], ".nc", sep = ""))) {
      PFTRasters(Region = Regions[[CompReg]], RegionFile = RegionFiles[[CompReg]],
                 Extent = Extents[[CompReg]], CountCodes = CountCodes)
    } else {
      print(paste("TRY data already aggregated to mean raster for region ",
                  RegionFiles[[CompReg]], sep = ""))
    }
  } # region-loop
} # Fun_PFTs
####--------------- Fun_Compadre [Variables, Regions, RegionFiles, Extents]
# (selecting and preparing data, and making COMPADRE data into rasters) ----
Fun_COMPADRE <- function(Variables, Regions, RegionFiles, Extents) {
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  print("COMPADRE ANALYSES")
  print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
  source(paste(Dir.Codes, "S4_COMPADRE.R", sep="/"))
  # Build rasters of compadre variables across study regions looping over Compadre
  # variables looping over regions
  for (CompVar in 1:length(Variables)) {
```

```
    for (CompReg in 1:length(Regions)) {
      if (!file.exists(paste(Dir.Compadre, "/", Variables[[CompVar]], "/",
                             Variables[[CompVar]], "_", RegionFile = RegionFiles[[CompReg]], ".nc",
                             sep = ""))) {
        RasterCOMPADRE(Variable = Variables[[CompVar]], Region = Regions[[CompReg]],
                       RegionFile = RegionFiles[[CompReg]], Extent = Extents[[CompReg]])
      } else {
        print(paste(Variables[[CompVar]], " already rasterised across ",
                    RegionFiles[[CompReg]], sep = ""))
      }
    } # region-loop
  } # CompVar-loop
} # Fun_Compadre
####--------------- FUNCTION CALLS ----
Fun_Vegetation(Regions = list(c("Portugal", "Spain", "France", "Andorra"), "Brazil", "Australia"),
               RegionFiles = list("Iberian Region", "Caatinga", "Australia"),
               Extents = list(extent(-10,10,35,52), extent(-50,-34,-23,0), NULL),
               From = 1982, To = 2015, Lags = 0:12, Cores = 5)

Fun_PFTs(Regions = list(c("Portugal", "Spain", "France", "Andorra"), "Brazil", "Australia"),
         RegionFiles = list("Iberian Region", "Caatinga", "Australia"),
         Extents = list(extent(-10,10,35,52), extent(-50,-34,-23,0), NULL),
         From = 1982, To = 2015, Occ = "Download")

Fun_COMPADRE(Variables = list("Reactivity", "Rho", "Pi", "FastSlow"),
             Regions = list(c("Portugal", "Spain", "France", "Andorra"), "Brazil", "Australia"),
             RegionFiles = list("Iberian Region", "Caatinga", "Australia"),
             Extents = list(extent(-10,10,35,52), extent(-50,-34,-23,0), NULL))
```

A.3.2 Vegetation Memory

A.3.2.1 GIMMs Data

Chunk 8: Downloading GIMMs NDVI3g data, establishing rasters of monthly composites (RasterGIMMs), and creating NDVI raster stacks for each study region (CombineCDFs).

```r
setwd(Dir.Gimms)
gimms_files <- updateInventory
####--------------- RasterGIMMs [from, to] ----
RasterGIMMs <- function(from, to){
  print("#################################################")
  print(paste("Rasterising GIMMs NDVI data from ", from, " to ", to, sep=""))
  setwd(Dir.Gimms) # set working directory to base GIMMs folder
  # PREPARING DATA----
  invisible(capture.output(gimms_files <- downloadGimms(x = as.Date(paste(from,"-01-01",sep="")), # start date
                                                        y = as.Date(paste(to,"-12-31",sep="")), # end date
                                                        dsn = Dir.Gimms))) # where to store the files
  gimms_raster <- rasterizeGimms(x = gimms_files, remove_header = TRUE) # rasterising
  indices <- monthlyIndices(gimms_files) # extract month indices from file list (should be two of each)
  gimms_raster_mvc <- monthlyComposite(gimms_raster, indices = indices) # create composites according to indices
  # Fix NDVI misbehaviours
  gimms_raster_mvc[gimms_raster_mvc<0] <- 0 # set threshold for barren land (NDVI<0)
  gimms_raster_mvc[gimms_raster_mvc>1] <- 1 # set threshold for saturated NDVI (NDVI > 1)
  # Indices
  Years <- rep(seq(from, to, 1), each = 12) # The year corresponding to each month in the stack
  names(gimms_raster_mvc) <- paste(month.abb, Years, sep = "") # create names for rasters
  # SAVING DATA----
  writeRaster(gimms_raster_mvc, paste(Dir.Gimms.Monthly, "/GlobalNDVI_", from, to, sep=""),
              overwrite=TRUE, format="CDF", varname="GIMMsNDVI")
  setwd(mainDir)}# end of RasterGIMMS-function
####--------------- CombineCDFS [Region, RegionFile, Extent] ----
CombineCDFs <- function(Region, RegionFile, Extent){
  print("#################################################")
  print(paste("Producing cropped GIMMs NDVI raster stacks for ", RegionFile, sep=""))
  # SELECTING FILES----
  files <- list.files(Dir.Gimms.Monthly)
  files.pos <- grep("GlobalNDVI", files)
  # REGION SELECTION----
  Shapes <- readOGR(Dir.Mask,'ne_50m_admin_0_countries', verbose = FALSE)
  RegObj <- RegionSelection(Region = Region, RegionFile = RegionFile, Extent = Extent)
  area <- RegObj[[1]]
  location <- RegObj[[2]]
  RegionFile <- RegObj[[3]]
  # LOADING, CROPPING AND MASKING----
  setwd(Dir.Gimms.Monthly)
  ras <- list()
  for(i in 1:length(files.pos)){
    rasinter <- brick(files[files.pos[i]]) # load i-th ndvi file
    rasinter <- crop(rasinter, area) # cropping to extent
    rasinter <- mask(rasinter, Shapes[location,]) # masking via Shapefile
    ras[[i]] <- rasinter # save masked ndvi to list
  }
  ras <- brick(ras) # create one big regional ndvi raster
  # SAVING DATA----
  writeRaster(ras, paste(Dir.Gimms.Monthly, "/NDVI_", RegionFile, sep=""),
              overwrite=TRUE, format="CDF", varname="NDVI",
              longname= paste("Monthly NDVI means across ", Region, sep=""))
  setwd(mainDir)} # CombineCDFs
```

A.3.2.2 ERA5

Chunk 9: Kriging ERA5 data from ERA5 to GIMMS resolution using HWSD covariates (table A.2) (RasterEra5).

```r
###---------------- PREAMBLE --------------------------------------------------------
Variables_vec <- c("Tair_mean", "Qsoil1_mean", "Qsoil2_mean", "Qsoil3_mean", "Qsoil4_mean")
VariablesNames_vec <- c("Air Temperature", "Soil Moisture (0-7cm)", "Soil Moisture (7-28cm)",
                        "Soil Moisture (28-100cm)", "Soil Moisture (100-255cm)")
Covariates_vec <- c("Slopes1", "Slopes2", "Slopes3", "Slopes4", "Slopes5", "Slopes6", "Slopes7", "Slopes8",
                    "Slope_aspect_N", "Slope_aspect_E", "Slope_aspect_S", "Slope_aspect_W", "Slope_aspect_U",
                    "Elevation")
Variables_vec <- c("Tair_mean", "Qsoil1_mean", "Qsoil2_mean", "Qsoil3_mean", "Qsoil4_mean")
VariablesNames_vec <- c("Air Temperature", "Soil Moisture (0-7cm)", "Soil Moisture (7-28cm)",
                        "Soil Moisture (28-100cm)", "Soil Moisture (100-200cm)")
Covariates_vec <- c("Slopes1", "Slopes2", "Slopes3", "Slopes4", "Slopes5", "Slopes6", "Slopes7", "Slopes8",
                    "Slope_aspect_N", "Slope_aspect_E", "Slope_aspect_S", "Slope_aspect_W", "Slope_aspect_U",
                    "Elevation")
####--------------- RasterEra5 [Variable, Region, FromY, FromM, ToY, ToM, Temporary]
# (selecting data, downscaling, exporting rasters) ----
RasterEra5 <- function(Variable, Region, RegionFile, Extent, FromY, FromM, ToY, ToM, Temporary){
  VarPos <- which(Variables_vec == paste(Variable, "mean", sep="_")) # position for indexing of variable
  YearVec <- rep(1980:2015, each = 12) # Year vector to indicate months for time frame selection
  # CONSOLE MESSAGE----
  print("#################################################")
  print(paste("Kriging ERA5 ", VariablesNames_vec[VarPos], " data from ", FromM, "/", FromY, " to ", ToM, "/", ToY,
              " across ", RegionFile, sep=""))
  # FORMULAE VECTORS----
  Krig_formula <- "ERA5 ~ Slopes1+Slopes2+Slopes3+Slopes4+Slopes5+Slopes6+Slopes7+Slopes8+Slope_aspect_N+
  Slope_aspect_E+Slope_aspect_S+Slope_aspect_W+Slope_aspect_U+Elevation+
  Slopes1:Slope_aspect_N+Slopes2:Slope_aspect_N+Slopes3:Slope_aspect_N+
  Slopes4:Slope_aspect_N+Slopes5:Slope_aspect_N+Slopes6:Slope_aspect_N+
  Slopes7:Slope_aspect_N+Slopes8:Slope_aspect_N+Slopes1:Slope_aspect_S+
  Slopes2:Slope_aspect_S+Slopes3:Slope_aspect_S+Slopes4:Slope_aspect_S+
  Slopes5:Slope_aspect_S+Slopes6:Slope_aspect_S+Slopes7:Slope_aspect_S+
  Slopes8:Slope_aspect_S"
  # LOAD DATA----
  ## Era5 data
  FirstMonth <- which(YearVec == FromY)[FromM] # first month to consider
  LastMonth <- which(YearVec == ToY)[ToM] # last month to consider
  ras <- list() # create empty list for era5 raster data
  Montquence <- FirstMonth:LastMonth
  ras <- brick(paste(Dir.ERA, "/", Variable ,"_TrainingResolution.nc", sep="")) # loading data
  ras <- ras[[Montquence]] # limitting to sought-after months
  extent(ras) <- c(-180,180,-90,90) # fix extent
  ## Covariates for Kriging
  Cov_coarse <- list() # create empty list
  for(c in 1:length(Covariates_vec)){ # cycle through all covariates and load the data
    Cov_coarse[[c]] <- raster(paste(Dir.KrigCov, "/Co-variates_TrainingResolution.nc", sep=""),
                              varname = Covariates_vec[c])}
  Cov_coarse <- brick(Cov_coarse) # make coarse covariate data into one big brick
  extent(Cov_coarse) <- c(-180,180,-90,90) # fix extent
  Cov_fine <- list() # create empty list
  for(c in 1:length(Covariates_vec)){ # cycle through all covariates and load the data
    Cov_fine[[c]] <- raster(paste(Dir.KrigCov, "/Co-variates_NativeResolution.nc", sep=""),
                            varname = Covariates_vec[c])}
  Cov_fine <- brick(Cov_fine) # make fine covariate data into one big brick
  # REGION SELECTION----
  Shapes <- readOGR(Dir.Mask,'ne_50m_admin_0_countries', verbose = FALSE)
  RegObj <- RegionSelection(Region = Region, RegionFile = RegionFile, Extent = Extent)
```

```
area <- RegObj[[1]]
location <- RegObj[[2]]
RegionFile <- RegObj[[3]]
# CROPPING AND MASKING----
## Era5 cropping and masking
ras <- crop(ras, area) # cropping to extent
ras <- mask(ras, Shapes[location,]) # masking via Shapefile
## Coarse covariate cropping and masking
Cov_coarse <- crop(Cov_coarse, area) # cropping to extent
Cov_coarse <- mask(Cov_coarse, Shapes[location,]) # masking via Shapefile
## Fine covariate cropping and masking
Cov_fine <- crop(Cov_fine, area) # cropping to extent
Cov_fine <- mask(Cov_fine, Shapes[location,]) # masking via Shapefile
# KRIGING----
Months1 <- (ToY-FromY-1)*12 # how many months to cover just by years
Months2 <- abs(ToM-FromM+1) # how many months to cover only taking months of time frame into account
FullMonths <- Months1+Months2 # total count of months that are covered
if(FromY == ToY){# if range doesn't exceed a calendar year
  Years <- rep(YearVec[FirstMonth], Months2) }else{
  FromLeft <- 12-FromM+1 # months left in starting year
  ToCovered <- ToM # months to be covered in final year
  Years1 <- rep(FromY, FromLeft)
  Years2 <- rep(ToY, ToCovered)
  if(ToY-FromY > 1){
    Years3 <- rep((FromY+1):(ToY-1), each = 12) # months in full years
    Years <- c(Years1, Years3, Years2) }else{
    Years <- c(Years1, Years2)}}
Months <- rep(c(1:12), length = length(Years))
Names <- paste(month.abb, Years, sep="") # combination of month names and years
## Preparing Kriging
Dir.Temp <- paste(Dir.ERA.Monthly,"/Temp_",Variable,"_",RegionFile, sep="")
dir.create(Dir.Temp)
TempNames <- paste(rep(YearVec),
                   rep(c("01","02","03","04","05","06","07","08","09","10","11","12")),sep="_")
# figuring out where to begin with the names
if(FromM < 10){
  TempStart <- which(TempNames == paste(FromY,"_0",FromM, sep="")) }else{
  TempStart <- which(TempNames == paste(FromY,"_",FromM, sep=""))}
# figuring out where to stop with the names
if(ToM < 10){
  TempStop <- which(TempNames == paste(ToY,"_0",ToM, sep="")) }else{
  TempStop <- which(TempNames == paste(ToY,"_",ToM, sep=""))}
TempNames <- TempNames[TempStart:TempStop]
Ras_Krig <- list()
### Actual Kriging
counter <- 0
for(i in 1:length(names(ras))){
  if(paste(TempNames[i], ".nc", sep="") %in% list.files(Dir.Temp)){ # check if this file has already been produced
    print(paste(TempNames[i], "already kriged", sep=" "))
    Ras_Krig[[i]] <- raster(paste(Dir.Temp, "/", TempNames[i], ".nc", sep=""))
    next()}
  counter <- counter + 1
  T_Begin <- Sys.time()
  RasterX <- ras[[i]]# extracting raster from Era5 stack
  # Base and Covariate Coarse Data
  Origin <- as.data.frame(RasterX, xy = TRUE)
  Origin <- na.omit(Origin)
  for(c in 1:length(Covariates_vec)){
```

```
    Cov_coarse[[c]][!is.na(RasterX) & is.na(Cov_coarse[[c]])] <- 0 # 0 cells where no info
    Cov_coarse[[c]][is.na(RasterX)] <- NA # ensure same NAs
    Covariate <- as.data.frame(Cov_coarse[[c]], xy = TRUE)
    Covariate <- na.omit(Covariate)
    Origin <- cbind(Origin, Covariate[,3])}
  colnames(Origin) <- c("x","y", "ERA5", Covariates_vec)
  # checking data availability
  for(it_check in 1:length(colnames(Origin))){
    if(length(which(Origin[,it_check] != 0)) < 2){
      stop(paste("The native resolution data does not support kriging using the formula you have specified
                 because ", colnames(Origin)[it_check], " does not contain enough data records for kriging
                 to be performed across the region you have specified (", Region, ").", " You can resolve
                 this issue by either removing the interaction effects containing this variable from the
                 formula or choosing a bigger study region.", sep=""))}}
  OriginK <- Origin
  gridded(OriginK) <- ~x+y
  # CROPPING TARGET
  Cov_fine[[1]][which(is.na(as.vector(Cov_fine[[1]])))] <- 0
  Cov_fine[[1]] <- mask(Cov_fine[[1]], Shapes[location,])
  Target <- as.data.frame(Cov_fine[[1]], xy = TRUE)
  for(c in 2:length(Covariates_vec)){
    Cov_fine[[c]][which(is.na(as.vector(Cov_fine[[c]])))] <- 0
    Cov_fine[[c]] <- mask(Cov_fine[[c]], Shapes[location,])
    Covariate <- as.data.frame(Cov_fine[[c]], xy = TRUE)
    Target <- cbind(Target, Covariate[,3])}
  colnames(Target) <- c("x","y", Covariates_vec)
  TargetK <- Target
  gridded(TargetK) <- ~x+y
  # KRIGING
  invisible(capture.output(
    kriging_result <- autoKrige(
      as.formula(Krig_formula),OriginK, TargetK, verbose = FALSE)))
  Krig_ras <- raster(kriging_result$krige_output)
  Ras_Krig[[i]] <- Krig_ras
  # writing the raster
  writeRaster(Krig_ras, filename = paste(Dir.Temp,"/",TempNames[i],sep=""), overwrite=TRUE, format="CDF")
  if(counter == 1){
    T_End <- Sys.time()
    Duration <- as.numeric(T_End)-as.numeric(T_Begin)
    print(paste("Calculating monthly ERA5 ", VariablesNames_vec[VarPos], " rasters from ", FromM, "/", FromY,
                " to ", ToM, "/", ToY, " across ", RegionFile, " should finish around: ",
                as.POSIXlt(T_Begin + Duration*(length(names(ras))-i), tz = Sys.timezone(location=TRUE)), sep=""))
    pb <- txtProgressBar(min = 0, max = length(names(ras)), style = 3)}
  setTxtProgressBar(pb, i)} # kriging loop
# COMBINING KRIGED ENSEMBLES FROM MEMORY----
Ras_Krig <- brick(Ras_Krig)
ras <- Ras_Krig
# ELIMINATE KRIGING ARTIFACTS OF SOIL MOISTURE BY BOUNDING
if(Variable == "Qsoil1" | Variable == "Qsoil2" | Variable == "Qsoil3" | Variable == "Qsoil4"){
  values(ras)[which(values(ras) < 0)] <- 0}
# SAVING DATA----
setwd(Dir.ERA.Monthly)
writeRaster(ras, paste(Variable, "_mean_", RegionFile, "_", FromM, FromY, "_", ToM, ToY, sep=""),
            overwrite=TRUE, format="CDF", varname=Variable,
            longname= paste(Variables_vec[VarPos], " mean for years ", FromM, "/",FromY, " to ", ToM,"/",ToY,
                            " across ", Region, sep=""))
if(Temporary == "Delete"){unlink(Dir.Temp, recursive = TRUE)}
setwd(mainDir)}# end of RasterEra5 function
```

A.3.2.3 Calculation of Vegetation Memory

Chunk 10: Computing vegetation memory for study regions as specified by lags, time frame, and variables to be considered (VegMem), scaling rasters of vegetation memory to be on the same scale for each study region (CoeffScaling).

```
Variables_vec <- c("Tair_mean", "Qsoil1_mean", "Qsoil2_mean", "Qsoil3_mean", "Qsoil4_mean")
VariablesNames_vec <- c("Air Temperature", "Soil Moisture (0-7cm)", "Soil Moisture (7-28cm)",
                        "Soil Moisture (28-100cm)", "Soil Moisture (100-200cm)")
####--------------- VegMem [ClimVar, Region, Cumlags, FromY, ToY]
# (selecting data, calculating vegetation memory according to specified lags, exporting rasters) ----
VegMem <- function(ClimVar, ClimVar2, Region, Cumlags, FromY, ToY){
  print("##################################################")
  print(paste("Identifying vegetation memory effects of NDVI based on antecedent NDVI and ",
              VariablesNames_vec[which(Variables_vec == ClimVar2)], " (immediate effects) and ",
              VariablesNames_vec[which(Variables_vec == ClimVar)], " at lags: ", toString(Cumlags)," across ",
              Region, sep=""))
  # LOAD DATA----
  ## NDVI/GIMMs
  NDVI_ras <- brick(paste(Dir.Gimms.Monthly,"/NDVI_", Region,".nc", sep=""))
  ## ERA5
  Clim <- list.files(Dir.ERA.Monthly)[grep(pattern = ClimVar, list.files(Dir.ERA.Monthly))]
  Clim <- Clim[grep(pattern = Region, Clim)] # files in target region
  Clim <- Clim[grep(pattern = FromY, Clim)] # files with correct start date
  Clim <- Clim[grep(pattern = ToY, Clim)] # files with correct end date
  Clim_mean_ras <- brick(paste(Dir.ERA.Monthly, "/", Clim, sep="")) # rasterise
  Clim2 <- list.files(Dir.ERA.Monthly)[grep(pattern = ClimVar2, list.files(Dir.ERA.Monthly))]
  Clim2 <- Clim2[grep(pattern = Region, Clim2)]
  Clim2 <- Clim2[grep(pattern = FromY, Clim2)]
  Clim2 <- Clim2[grep(pattern = ToY, Clim2)]
  Clim2_mean_ras <- brick(paste(Dir.ERA.Monthly, "/", Clim2, sep=""))
  # PREPARE DATA----
  ## Limit NDVI data to ERA5 time frame
  NDVIYears <- rep(1982:2015, each = 12) # Year vector to indicate months for time frame selection
  NDVITo <- max(which(NDVIYears %in% ToY))
  if(min(which(NDVIYears %in% FromY)) == Inf){
    NDVIFrom <- 1 }else{ NDVIFrom <- min(which(NDVIYears %in% FromY))}
  NDVI_ras <- NDVI_ras[[NDVIFrom:NDVITo]] # NDVI data is limitted
  ## Identify data positions
  # establish a mean raster (this sets every cell to NA where any NA is within the time series)
  NATest_ras <- mean(NDVI_ras)
  NATest_vec <- values(NATest_ras) # set values as vector
  Data_Pos <- which(!is.na(NATest_vec)) # select non-NA positions (these are the ones we should build models on)
  # PREPARE RASTERS----
  ModelEval_ras <- NDVI_ras[[1:10]] # select six raster layers
  # put names on the layers to tell us what they contain later
  ModelEval_ras <- Fun_NamesRas(raster = ModelEval_ras, ClimVar = ClimVar, ClimVar2 = ClimVar2)
  # MODELS----
  for(pixel in Data_Pos){ # loop non-NA pixels
    T_Begin <- Sys.time() # note time when calculation is started (needed for estimation of remaining time)
    ## DATA ----
    ### NDVI stuff -----
    NDVI_vecraw <- as.vector(NDVI_ras[pixel]) # eytract data
    NDVI_vecdet <- detrend(NDVI_vecraw, tt = 'linear') # linear detrending
    # create NDVI data frame
    NDVI_df <- data.frame(Month = rep(1:12, length(NDVI_vecraw)/12), NDVI_raw = NDVI_vecraw, NDVI_de = NDVI_vecdet)
    ## calculate anomalies (Z-scores) and monthly means
    NDVI_df <- transform(NDVI_df, NDVI_Anomalies = ave(NDVI_de, Month, FUN=scale),
                         NDVI_Threshold = ave(NDVI_raw, Month, FUN=function(t) mean(t, na.rm=TRUE)))
    NDVI_df <- NDVI_df[nrow(NDVI_df):1,] # reverse order to read "present to past"
```

```r
NDVI_anom <- c(NDVI_df$NDVI_Anomalies, rep(NA, max(Cumlags))) # extract anomalies, adding cumlag NAs
ThreshPos <- which(NDVI_df$NDVI_Threshold < 0.1) # positions which should be excluded
if(length(ThreshPos) == length(NDVI_vecraw)){ # if all months should be masked due to NDVImean < 0.1
  # set all in model raster layers to NA for this pixel
  ModelEval_ras[pixel] <- as.numeric(rep(NA, dim(ModelEval_ras)[3]))
  next()}
# calculate lag 1
NDVI_Lag1 <- c(NDVI_anom[-1], NA) # adding one NA for month preceeding data range of NDVI itself
### Climate stuff -----
#### ClimVar ------
Clim_vec <- as.vector(Clim_mean_ras[pixel]) # extract raw data for pixel (instantenous predictor)
Clim_vec <- detrend(Clim_vec, tt = 'linear') # linear detrending
Clim_vec <- Clim_vec[nrow(Clim_vec):1,] # reverse order to read "present to past"
# calculate cumulative climate indices (antecedent predictor)
Clim_cum <- rep(NA, length(Cumlags))
Clim_cum <- as.list(Cumlags)
position <- 1
for(lag in Cumlags){
  for(i in 1:(length(Clim_vec)-lag)){
    Clim_cum[[position]] <- c(Clim_cum[[position]], sum(Clim_vec[i:(i+lag)]))}
  Clim_cum[[position]] <- Clim_cum[[position]][-1] # removing initial NA
  # adding enough NAs to bring it up to full length
  Clim_cum[[position]] <- c(Clim_cum[[position]], rep(NA , length(Clim_vec)-length(Clim_cum[[position]])))
  position <- position+1}
# make data frame of climate stuff
Clim_df <- as.data.frame(Clim_cum) # make list into data frame
Clim_df <- cbind(rep(12:1, length(Clim_vec)/12), Clim_vec, Clim_df) # append month index and raw data
colnames(Clim_df) <- c("Month", "Clim_raw", paste(rep("ClimCum_",length(Cumlags)),Cumlags, sep="")) # column names
# calculate anomalies
for(anomaly in 2:length(Clim_df)){# cycle through all the columns of the climate data frame except the month column
  Clim_iter <- with(Clim_df, cbind(Month, Clim_df[,anomaly])) # extract necessary data
  colnames(Clim_iter) <- c("Month", "AnomalyCalc") # set column names
  Clim_iter <- transform(Clim_iter, # calculate anomaly for each month
                       AnomalyCalc = ave(AnomalyCalc, Month, FUN=scale))
  # save to original data frame
  Clim_df[,anomaly] <- Clim_iter$AnomalyCalc}
#### ClimVar2 ------
Clim2_vec <- as.vector(Clim2_mean_ras[pixel]) # extract raw data for pixel (instantenous predictor)
Clim2_vec <- detrend(Clim2_vec, tt = 'linear') # linear detrending
Clim2_df <- data.frame(Month = rep(1:12, length(Clim2_vec)/12), Clim2_raw = Clim2_vec)
# calculate anomalies
Clim2_df <- transform(Clim2_df,
                    Clim2_Anomalies = ave(Clim2_raw, Month, FUN=scale))
Clim2_df <- Clim2_df[nrow(Clim2_df):1,] # reverse order to read "present to past"
Clim2_anom <- Clim2_df$Clim2_Anomalies
### Combining all the data -----
ModData_df <- cbind(NDVI_anom[1:length(Clim2_anom)], NDVI_Lag1[1:length(Clim2_anom)], Clim_df, Clim2_anom)
if(length(ThreshPos) > 0){ # set threshold months to NA if necessary
  ModData_df$NDVI_anom[ThreshPos] <- NA}
ModData_df <- na.omit(ModData_df) # get rid of NA rows
## MODELS ----
### Establishing models----
# list to save Model objects
Mods_ls <- as.list(rep(NA, length(Cumlags))) # List of models
ps <- rep(NA, length(Cumlags)) # p-values
coeffst1 <- rep(NA, length(Cumlags)) # coefficients of NDVI-1
coeffsC <- rep(NA, length(Cumlags)) # coefficients of ClimVar
coeffsC2 <- rep(NA, length(Cumlags)) # coefficients of ClimVar2
```

```r
# itterate over all climate lags
counter <- 0 # create a counter variable
for(ModelIter in Cumlags){ # go through all possible cumulative lags
  ## PCA of our variables
  pca_mat <- matrix(cbind(ModData_df$NDVI_Lag1,ModData_df[, counter+4], ModData_df$Clim2_anom),
                    ncol = 3, byrow = FALSE, dimnames = list(1:length(ModData_df$NDVI_Lag1),
                                                            c("t-1", ClimVar, ClimVar2))) # pca matrix
  pca <- rda(pca_mat) # running pca
  ## Extracting PC axes
  pc1 <- summary(pca)[["sites"]][,1]
  pc2 <- summary(pca)[["sites"]][,2]
  pc3 <- summary(pca)[["sites"]][,3]
  ## Building models
  Mod0 <- lm(ModData_df$NDVI_anom ~ 1) # null model
  Mod <- lm(ModData_df$NDVI_anom ~ pc1 + pc2 + pc3) # full model
  loadings <- summary(pca)[["species"]] # extract loadings
  coefficients <- Mod$coefficients[2:(dim(summary(pca)[["sites"]])[2]+1)] # extract coefficients
  ## Make coefficients representative by multiplying them with the loadings
  t1newCof <- loadings[1,] * coefficients
  CnewCof <- loadings[2,] * coefficients
  C2newCof <- loadings[3,] * coefficients
  ## Saving information to vectors
  coeffst1[counter+1] <- sum(t1newCof) # NDVI-1
  coeffsC[counter + 1] <- sum(CnewCof) # ClimVar
  coeffsC2[counter + 1] <- sum(C2newCof) # ClimVar2
  if(anova(Mod0, Mod)$RSS[1] > anova(Mod0, Mod)$RSS[2]){ # only save p value if model is an imporvement
    ps[counter+1] <- anova(Mod0, Mod)$'Pr(>F)'[2]
  }else{ # if model is not an improvement over null, set p to 1
    ps[counter+1] <- 1}
  Mods_ls[[counter+1]] <- Mod # save model to list of models
  counter <- counter + 1 }
### Selecting best model -----
AICs <- sapply(X = Mods_ls, FUN = AIC) # calculate AICs for each model
Best <- which(abs(AICs) == min(abs(AICs), na.rm = TRUE))[1] # best model, if same value present use first
c_NDVI <- coeffst1[Best] # ndvi coefficient
c_Clim <- coeffsC[Best] # climate coefficient
c_Clim2 <- coeffsC2[Best] # p-value of climate coefficient
p_Mod <- ps[Best] # p-value is set to p-value of best model
AICMod <- AICs[Best] # AIC is set to best AIC
Bestlag <- Cumlags[Best] # this is the lag at which best model was observed
## EXPLAINED VARIANCE----
colnames(ModData_df)[c(1:2)] <- c("NDVI_anom", "NDVI_Lag1")
# Legendre & Legendre
Explainedvar <- lm(data = ModData_df,
                   NDVI_anom ~ NDVI_Lag1 + ModData_df[,Best+5])
Explainedvar <- summary(Explainedvar)[["r.squared"]]
VarTotalNDVI <- lm(data = ModData_df,
                   NDVI_anom ~ NDVI_Lag1)
VarTotalNDVI <- summary(VarTotalNDVI)[["r.squared"]]
VarTotalQsoil <- lm(data = ModData_df,
                    NDVI_anom ~ ModData_df[,Best+5])
VarTotalQsoil <- summary(VarTotalQsoil)[["r.squared"]]
VarShared <- VarTotalQsoil + VarTotalNDVI - Explainedvar
VarNDVI <- VarTotalNDVI - VarShared
VarQsoil <- VarTotalQsoil - VarShared
## WRITING INFORMATION TO RASTERS----
ModelEval_ras[pixel] <- as.numeric(c((Bestlag), AICMod, p_Mod, c_NDVI, c_Clim, c_Clim2, Explainedvar,
                                     VarNDVI, VarShared, VarQsoil)) # saving model information to raster
```

```r
    ## Updating progress bar----
    if(pixel == Data_Pos[1]){ # if we are currently on the first pixel
      T_End <- Sys.time() # note end time
      Duration <- as.numeric(T_End)-as.numeric(T_Begin) # calculate the time it took to establish and select models
      ## Put an estimator up on the console that tells the user when to expect the program to finish its current run
      print(paste("Calculating Vegetation Memory effects across ", Region, " should finish around: ",
                  as.POSIXlt(T_Begin + Duration*length(Data_Pos), tz = Sys.timezone(location=FALSE)), sep=""))
      ## Update progress bar
      pb <- txtProgressBar(min = 0, max = length(Data_Pos), style = 3)
      pbi <- 0}
    pbi <- pbi + 1 ## Update progress bar
    setTxtProgressBar(pb, pbi)} # end of pixel loop
  ### Save raster -----
  writeRaster(ModelEval_ras, filename = paste(Dir.Memory,"/", Region, "_", ClimVar2, "-", ClimVar,
                                              paste(Cumlags, collapse="_"),"_",FromY,"-", ToY, ".nc",sep=""),
              overwrite=TRUE, format="CDF")
  setwd(mainDir)}# end of VegMem function

####--------------- CoeffScaling [ClimVar, ClimVar2, Region, Cumlags, FromY, ToY, UAbs]
# (loading previously saved rasters and making visualisation of model coefficients better) ----
CoeffScaling <- function(ClimVar, ClimVar2, Region, Cumlags, FromY, ToY, UAbs){
  print("##################################################")
  print(paste("Producing composites of vegetation memory effects across ", unique(Region), sep=""))
  # PREPARATIONS ----
  Rasters <- ClimVar
  minmaxNDVIn <- rep(NA, length(ClimVar)*2)
  minmaxNDVIs <- rep(NA, length(ClimVar)*2)
  minmaxCVn <- rep(NA, length(ClimVar)*2)
  minmaxCVs <- rep(NA, length(ClimVar)*2)
  minmaxCV2n <- rep(NA, length(ClimVar)*2)
  minmaxCV2s <- rep(NA, length(ClimVar)*2)
  minmaxPos <- 1
  # LOADING DATA ----
  for(rasiter in 1:length(ClimVar)){ # cycle through specified vegetation memory rasters
    # load raster
    Alter_ras <- brick(paste(Dir.Memory,"/", Region[[rasiter]], "_", ClimVar2[[rasiter]], "-", ClimVar[[rasiter]],
                             paste(Cumlags[[rasiter]], collapse="_"),"_",FromY,"-", ToY, ".nc",sep=""))
    Alter_ras <- Fun_NamesRas(raster = Alter_ras, ClimVar = ClimVar, ClimVar2 = ClimVar2, rasiter = rasiter)
    # PREPARING DATA ----
    P_ras <- Alter_ras$Model.p.value # extract p-value layer
    C_clim <- Alter_ras[[5]] # extract ClimVar coefficients
    C_climNon <- C_clim
    C_climNon[which(values(P_ras) < 0.05)] <- NA # set everything that significant to NA
    C_climSig <- C_clim
    C_climSig[which(values(P_ras) >= 0.05)] <- NA # set everything that's not significant to NA
    C_clim2 <- Alter_ras[[6]] # extract ClimVar2 coefficients
    C_climNon2 <- C_clim2
    C_climNon2[which(values(P_ras) < 0.05)] <- NA # set everything that significant to NA
    C_climSig2 <- C_clim2
    C_climSig2[which(values(P_ras) >= 0.05)] <- NA # set everything that's not significant to NA
    C_NDVI <- Alter_ras$Antecedent.NDVI..c_NDVI. # extract NDVI-1 coefficients
    C_NDVINon <- C_NDVI
    C_NDVINon[which(values(P_ras) < 0.05)] <- NA # set everything that significant to NA
    C_NDVISig <- C_NDVI
    C_NDVISig[which(values(P_ras) >= 0.05)] <- NA # set everything that's not significant to NA
    C_Lags <- Alter_ras[[1]] # extract Lags coefficients
    C_LagsNon <- C_Lags
    C_LagsNon[which(values(P_ras) < 0.05)] <- NA # set everything that significant to NA
```

```
  C_LagsSig <- C_Lags
  C_LagsSig[which(values(P_ras) >= 0.05)] <- NA # set everything that's not significant to NA
  # SAVING PARAMETERS ----
  Rasters[[rasiter]] <- list(C_NDVINon, C_NDVISig, C_climNon, C_climSig,
                             C_climNon2, C_climSig2, C_LagsNon, C_LagsSig)
  ## Identify maximum values of each coefficient raster
  minmaxNDVIn[minmaxPos] <- max(values(C_NDVINon), na.rm = TRUE)
  minmaxNDVIs[minmaxPos] <-max(values(C_NDVISig), na.rm = TRUE)
  minmaxCVn[minmaxPos] <- max(values(C_climNon), na.rm = TRUE)
  minmaxCVs[minmaxPos] <- max(values(C_climSig), na.rm = TRUE)
  minmaxCV2n[minmaxPos] <- max(values(C_climNon2), na.rm = TRUE)
  minmaxCV2s[minmaxPos] <- max(values(C_climSig2), na.rm = TRUE)
  minmaxPos <- minmaxPos + 1 # +1 to counter in vector
  ## Identify minimum values of each coefficient raster
  minmaxNDVIn[minmaxPos] <- min(values(C_NDVINon), na.rm = TRUE)
  minmaxNDVIs[minmaxPos] <-min(values(C_NDVISig), na.rm = TRUE)
  minmaxCVn[minmaxPos] <- min(values(C_climNon), na.rm = TRUE)
  minmaxCVs[minmaxPos] <- min(values(C_climSig), na.rm = TRUE)
  minmaxCV2n[minmaxPos] <- min(values(C_climNon2), na.rm = TRUE)
  minmaxCV2s[minmaxPos] <- min(values(C_climSig2), na.rm = TRUE)
  minmaxPos <- minmaxPos + 1 # +1 to counter in vector
} # end of loop cycling through specified vegetation memory rasters
# MANN-WHITNEY U ----
## setting up directory
Dir.Memory.Reg <- paste(Dir.Memory,"/",unique(Region),"-",FromY,"_",ToY, sep="")
dir.create(Dir.Memory.Reg)
## cleaning directory o potential earlier runs
if(paste("U-Variables_Abs",UAbs,".xlsx",sep="") %in% list.files(Dir.Memory.Reg)){
  file.remove(paste(Dir.Memory.Reg,"/U-Variables_Abs",UAbs,".xlsx",sep=""))}
# Establish matrices and vectors for naming
UModMat <- matrix(rep(NA, length(Rasters)^2), nrow=length(Rasters)) # for saving U outputs
UVariables <- c("NDVI t-1", "Qsoil", "Tair", "Lags") # for naming purposes
UMedians <- matrix(rep(NA, length(Rasters)^2), nrow=length(Rasters)) # for saving variable value medians
dimnames(UMedians) <- list(c(1:4), UVariables) # set names
# variable-wise comparison
for(UVar in 1:length(Rasters)){ # loop over all variables
  for(UTest in 1:(length(Rasters)-1)){ # loop over the model layers
    UTest2 <- UTest + 1 # create seperate counter for variable with which to compare
    while(UTest2 <= length(Rasters)){ # cycle so long as second counter does not exceed range of specified models
      if(UAbs == TRUE){ # if absolute values should be used
        Test1 <- abs(values(Rasters[[UTest]][[UVar * 2]])) # data extraction
        Test2 <- abs(values(Rasters[[UTest2]][[UVar * 2]])) # data extraction
      }else{ # if absolute values are not desired
        Test1 <- values(Rasters[[UTest]][[UVar * 2]])
        Test2 <- values(Rasters[[UTest2]][[UVar * 2]])}
      test <- wilcox.test(Test1, Test2, paired = FALSE) # WHitney-U Test
      UModMat[UTest, UTest2] <- test$statistic # Extract test statistic
      UModMat[UTest2, UTest] <- test$p.value # Extract p-value
      Med1 <- median(Test1, na.rm=TRUE) # extract median
      Med2 <- median(Test2, na.rm=TRUE) # extract median
      UMedians[UTest,UVar] <- Med1 # write median of first object
      if(UTest2 == length(Rasters)){ # only write median of last variable
        UMedians[UTest2,UVar] <- Med2}# if statement
      UTest2 <- UTest2 + 1 } # while statement
  } # for UTest statement
  # saving output
  write.xlsx(UModMat, sheetName = UVariables[UVar],
             file = paste(Dir.Memory.Reg,"/U-Variables_Abs",UAbs,".xlsx",sep=""), append = TRUE)
```

```
} # for UVar statement
write.xlsx(UMedians, sheetName = "Variable medians", # saving output
           file = paste(Dir.Memory.Reg,"/U-Variables_Abs",UAbs,".xlsx",sep=""), append = TRUE)
UModelMat <- matrix(rep(NA,3^2),nrow=3) # matrix for model-internal comparisons
dimnames(UModelMat) <- list(UVariables[1:3], UVariables[1:3]) # set names
# model-wise comparison
for(UTest in 1:(length(Rasters))){
  if(UAbs == TRUE){ # if absolute values should be used
    ND <- abs(values(Rasters[[UTest]][[2]]))
    QS <- abs(values(Rasters[[UTest]][[4]]))
    TA <- abs(values(Rasters[[UTest]][[6]]))
  }else{ # if absolute values are not desired
    ND <- values(Rasters[[UTest]][[2]])
    QS <- values(Rasters[[UTest]][[4]])
    TA <- values(Rasters[[UTest]][[6]])}
  UModelMat[2,1] <- wilcox.test(ND, QS, paired = FALSE)$p.value # WHitney-U Test
  UModelMat[3,2] <- wilcox.test(QS, TA, paired = FALSE)$p.value # WHitney-U Test
  UModelMat[3,1] <- wilcox.test(ND, TA, paired = FALSE)$p.value # WHitney-U Test
  UModelMat[1,2] <- wilcox.test(ND, QS, paired = FALSE)$statistic # WHitney-U Test
  UModelMat[2,3] <- wilcox.test(QS, TA, paired = FALSE)$statistic # WHitney-U Test
  UModelMat[1,3] <- wilcox.test(ND, TA, paired = FALSE)$statistic # WHitney-U Test
  write.xlsx(UModelMat, sheetName = paste("Model", UTest, sep=" "), # saving output
             file = paste(Dir.Memory.Reg,"/U-Variables_Abs",UAbs,".xlsx",sep=""), append = TRUE)
} # for UTest statement
# SAVING DATA FOR LATER PLOTTING ----
for(iterplot in 1:length(ClimVar)){ # cycle through all specified vegetation memory raster for plotting
  ## lags -----
  Lag_ras <- brick(paste(Dir.Memory,"/", Region[[iterplot]], "_", ClimVar2[[iterplot]], "-", ClimVar[[iterplot]],
                       paste(Cumlags[[iterplot]], collapse="_"),"_",FromY,"-", ToY, ".nc",sep=""))
  ## NDVI[t-1] -----
  Rasters[[iterplot]][[1]][1] <- max(minmaxNDVIn)
  Rasters[[iterplot]][[1]][2] <- min(minmaxNDVIn)
  Rasters[[iterplot]][[2]][3] <- max(minmaxNDVIs)
  Rasters[[iterplot]][[2]][4] <- min(minmaxNDVIs)
  ### ClimVar -----
  Rasters[[iterplot]][[3]][1] <- max(minmaxCVn)
  Rasters[[iterplot]][[3]][2] <- min(minmaxCVn)
  Rasters[[iterplot]][[4]][3] <- max(minmaxCVs)
  Rasters[[iterplot]][[4]][4] <- min(minmaxCVs)
  ### ClimVar2 -----
  Rasters[[iterplot]][[5]][1] <- max(minmaxCV2n)
  Rasters[[iterplot]][[5]][2] <- min(minmaxCV2n)
  Rasters[[iterplot]][[6]][3] <- max(minmaxCV2s)
  Rasters[[iterplot]][[6]][4] <- min(minmaxCV2s)
  ### sighnifciant coefficient rasters (already with max/min dots)
  Save_ras <- brick(Lag_ras[[1]], # Lags
                  Rasters[[iterplot]][[2]], # NDVI
                  Rasters[[iterplot]][[4]], # climvar
                  Rasters[[iterplot]][[6]]) # climvar2
  ### Saving significant effects -----
  writeRaster(Save_ras, filename = paste(Dir.Memory.Reg ,"/", ClimVar[[iterplot]], "_", ClimVar2[[iterplot]],
                                       paste(Cumlags[[iterplot]], collapse="_"), "Plots.nc",sep=""),
              overwrite=TRUE, format="CDF")} # plotting loop
setwd(mainDir)} # CoeffScaling end
```

A.3.3 Plant Functional Traits

Chunk 11: Calculating species-specific PFT means (PFTs), creating distribution maps from floral occurence data obtained from GBIF (DistMaps), and combine occurence maps with species trait means (PFTRasters).

```
####--------------- PFTs []
# (loading data, building species-specific trait means and saving the result) ----
PFTs <- function(){
  # LOADING DATA ----
  ## NDVI (reference raster) ---
  NDVI_ras <- brick(paste(Dir.Gimms.Monthly,"/GlobalNDVI_20112015.nc", sep=""))
  ref_ras <- NDVI_ras[[6]]
  ref_ras[which(values(ref_ras) > -1)] <- 8888 # identify land pixels
  ## Master PFT data from TRY
  PFT_Master <- read.table(file = paste(Dir.TRY, "/4704.txt", sep=""), stringsAsFactors = FALSE, fill = TRUE,
                           sep="\t", header = TRUE)
  ## Extracting necessary data to handle smaller data frame
  PFTs_df <- data.frame(Species = PFT_Master$SpeciesName, ObsID = PFT_Master$ObservationID,
                        Variable = PFT_Master$DataName, Value = PFT_Master$OrigValueStr, Unit = PFT_Master$UnitName)
  ## fixing factor to numeric
  as.numeric.factor <- function(x) {as.numeric(levels(x))[x]}
  PFTs_df$Value <- as.numeric.factor(PFTs_df$Value)
  # CALCULATING SPECIES-SPECIFIC MEAN TRAIT VALUES ----
  ## Preparations for calculations
  Species <- unique(PFTs_df$Species) # all species to consider
  FullSpec_df <- data.frame(Species = NA, Nmass = NA, Height = NA)
  ## Calculations
  for(Iter in 1:length(Species)){ # loop over all species that need consideration
    T_Begin <- Sys.time() # read start time (needed for expected finishing time)
    ### Nitrogen mean
    Nitro <- which(PFTs_df$Species == Species[Iter] & # positions of nitrogen rows for species
                   PFTs_df$Variable == "Leaf nitrogen content per dry mass (Nmass)")
    Mean_Nitro <- mean(PFTs_df$Value[Nitro], na.rm = TRUE) # mean
    ### Height mean
    Height <- which(PFTs_df$Species == Species[Iter] & # positions of height rows for species
                    PFTs_df$Variable == "Plant height vegetative")
    Mean_Height <- mean(PFTs_df$Value[Height], na.rm = TRUE) # mean
    ### Combining data into a data frame
    Spec_df <- data.frame(Species = Species[Iter], Nmass = Mean_Nitro, Height = Mean_Height)
    FullSpec_df <- rbind(FullSpec_df, Spec_df)
    ### Updating progress bar
    if(Iter == 1){ # estimate on finishing time on first loop
      T_End <- Sys.time() # read end time
      Duration <- as.numeric(T_End)-as.numeric(T_Begin) # duration between date points
      ### Output to console
      print(paste("Calculating species-specific trait mean values should finish around ",
                  as.POSIXlt(T_Begin + Duration*length(Species), tz = Sys.timezone(location=TRUE)), sep=""))
      ### Removing empty initial row
      FullSpec_df <- FullSpec_df[-1,]}# end of estimator if-statement
    pb <- txtProgressBar(min = 0, max = length(Species), style = 3) # Setting up a progress bar
    setTxtProgressBar(pb, Iter)}# end of species-specific mean trait value calculation for-loop
  FullNA <- which(is.na(FullSpec_df$Nmass) & is.na(FullSpec_df$Height)) # identifying species where both means are NA
  CorrectedSpec_df <- FullSpec_df[-FullNA,] # removing full NA species records
  # Save the species-specific and NA-freed data
  save(CorrectedSpec_df, file = paste(Dir.TRY, "/SpeciesTraits.RData", sep=""))
  setwd(mainDir)

  # BUILDING RASTERS FROM RAW TRAIT MEASURES ----
  ## create empty data frame and filling it
```

```
  Locs_df <- data.frame(H = NA, Nmass = NA, Lat = NA, Lon = NA)
  `%nin%` = Negate(`%in%`) # create a 'not in' statement
  ## eliminate observations with less than three records (these can't have enough data)
  IDs <- PFTs_df$ObsID
  exclude <- names(table(IDs))[which(table(IDs) < 3)]
  IDs <- unique(IDs)[which(as.character(unique(IDs)) %nin% exclude)]
  ## loop over all observations with enough data
  for(i in 1:length(IDs)){
    T_Begin <- Sys.time() # read start time (needed for expected finishing time)
    Iter_df <- PFTs_df[which(PFTs_df$ObsID == IDs[i]),] # extract all data for current observation
    ## data extraction
    Nmass <- Iter_df$Value[which(Iter_df$Variable == "Leaf nitrogen content per dry mass (Nmass)")]
    if(length(Nmass) == 0){Nmass <- NA}
    H <- Iter_df$Value[which(Iter_df$Variable == "Plant height vegetative")]
    if(length(H) == 0){H <- NA}
    Lat <- as.numeric(Iter_df$Value[which(Iter_df$Variable == "Latitude")])
    Lon <- as.numeric(Iter_df$Value[which(Iter_df$Variable == "Longitude")])
    Locs_df <- rbind(Locs_df, c(H, Nmass, Lat, Lon)) # bin data
    ## Updating progress bar
    if(i == 1){ # estimate on finishing time on first loop
      T_End <- Sys.time() # read end time
      Duration <- as.numeric(T_End)-as.numeric(T_Begin) # duration between date points
      ### Output to console
      print(paste("Extracting raw geo-referenced data should finish around ",
                  as.POSIXlt(T_Begin + Duration*length(IDs), tz = Sys.timezone(location=TRUE)), sep=""))
      ### Removing empty initial row
      Locs_df <- Locs_df[-1,]
      pb <- txtProgressBar(min = 0, max = length(IDs), style = 3)}
    setTxtProgressBar(pb, i)
  } # obvservation loop
  ## converting to spatial points objects
  H_df <- na.omit(Locs_df[which(!is.na(Locs_df$H)),-2])
  H_pts <- data.frame(y = H_df$Lat, x = H_df$Lon, z = H_df$H)
  H_pts <- na.omit(H_pts)
  coordinates(H_pts) = ~x+y # convert x and y to coordinates
  Nmass_df <- na.omit(Locs_df[which(!is.na(Locs_df$Nmass)),-1])
  Nmass_pts <- data.frame(y = Nmass_df$Lat, x = Nmass_df$Lon, z = Nmass_df$Nmass)
  Nmass_pts <- na.omit(Nmass_pts)
  coordinates(Nmass_pts) = ~x+y # convert x and y to coordinates
  ## rasterising
  rast <- raster(ext=extent(ref_ras), resolution=res(ref_ras)) # create raster to be filled
  H_rasOut <- rasterize(x = H_pts, y = rast, field = H_pts$z, fun = mean) # rasterize irregular points
  Nmass_rasOut <- rasterize(x = Nmass_pts, y = rast, field = Nmass_pts$z, fun = mean) # rasterize irregular points
  Means_ras <- brick(H_rasOut, Nmass_rasOut)
  ## saving data
  writeRaster(x=Means_ras, filename = paste(Dir.TRY,"/RawTRY-Global", sep=""),overwrite=TRUE, format="CDF")
}# end of PFTs-function

####--------------- DistMaps [Species, Extent, Years, CountCodes]
# (Obtaining occurence data via GBIF, rasterising, saving the raster, limitting to a region) ----
DistMaps <- function(Species, Years, CountCodes){
  print("##################################################")
  print(paste("Downloading occurence data of species: ", Species, " across: ", CountCodes, sep=""))
  # LOADING SPECIES DATA FRAME ----
  load(paste(Dir.TRY, "/SpeciesTraits.RData", sep="")) # load data frame 'CorrectedSpec_df'
  if(Species == "All"){ # selecting all species contained in 'CorrectedSpec_df'
    Species <- sort(CorrectedSpec_df$Species)
    Species <- Species[-which(Species == "-")] # remove this error of a species name
```

```
  SP <- "All"
}else{
  SP <- "Dummy" # only used if single species is targeted
}
# GLOBAL REFERENCE DATA (needed for rasterising and masking) ----
## NDVI (reference raster) ---
NDVI_ras <- brick(paste(Dir.Gimms.Monthly,"/GlobalNDVI_20112015.nc", sep=""))
ref_ras <- NDVI_ras[[6]]
ref_ras[which(values(ref_ras) > -1)] <- 8888 # identify land pixels
# IDENTIFYING GBIF KEY(S) ----
# if data is already present and all species are sought-after
if("SpeciesGBIFKeys.rda" %in% list.files(path=Dir.TRY) & SP == "All"){
  print("Loading species-specific GBIF keys from local storage") # output to console
  load(paste(Dir.TRY, "/SpeciesGBIFKeys.rda", sep=""))
}else{
  ## Preparations ---
  Key_vec <- NA # create empty vector for gbif key(s)
  Species_Pres <- NA # create empty vector for all species which we have occurence records for
  print("Identifying species-specific GBIF keys from GBIF repository") # output to console
  pb <- txtProgressBar(min = 0, max = length(Species), style = 3) # Setting up a progress bar
  for(Iter in 1:length(Species)){ # cycle through all species specified
    key <- name_suggest(q=Species[Iter], rank='species')$key[1] # pull gbif key
    Key_vec <- c(Key_vec, key) # append key to key vector
    if(!is.null(key)){ # if we have occurence data in the gbif records
      Species_Pres <- c(Species_Pres, Species[Iter]) # append species to species vector
    }
    setTxtProgressBar(pb, Iter)}
  ## Fixing vectors
  Key_vec <- Key_vec[-1] # Removing empty initial element
  Species_Pres <- Species_Pres[-1] # removing empty initial element
  # Save the species-specific and NA-freed data
  if(SP == "All"){ # if we want all species, we might as well save the names and key objects for later saving of time
    save(list = c("Species_Pres", "Key_vec"), file = paste(Dir.TRY, "/SpeciesGBIFKeys.rda", sep=""))}} # GBIF keys
# OCCURENCE DATA ----
## Preparation
print("Downloading species-specific occurence records from GBIF") # output to console
pb <- txtProgressBar(min = 0, max = length(Species_Pres), style = 3) # Setting up a progress bar
## If an error occured previously
if("Breakage.txt" %in% list.files(path = mainDir)){ # this file is only present if the run finished prematurely
  OccIter <- read.table(paste(mainDir, "/Breakage.txt", sep=""))[1,1] # position at which it failed previously - 1
}else{
  OccIter <- 0 # set to 0 if it didn't fail previously
}
if(OccIter > 1){ # if previous run (OccIter) failed at the second step or later
  Start <- OccIter + 1 # start from where it failed, OccIter is the last one that got done
}else{ # if it failed at the first one
  Start <- 1 # start at the first species
}
for(OccIter in Start:length(Key_vec)){ # cycling through all species to obtain occurence data
  # if species name cannot be put into a file name due to special characters, this excludes 61 species records
  if(grepl('[^[:alnum:]. .).(.-]', Species_Pres[OccIter]) == TRUE){
    next()}
  if(paste(Species_Pres[OccIter], "_", CountCodes, ".rda", sep="") %in% list.files(path=Dir.OCCs)){
  }else{ # data not present locally yet
    ## Downloading Data
    key <- Key_vec[OccIter] # select GBIF key
    Gbif <- occ_data(key, limit=200000, hasCoordinate = TRUE, year = Years,
                     hasGeospatialIssue = FALSE, country = CountCodes) # download data
```

```
    ## Dealing with separate data frames of years
    BaseOcc <- rep(NA, 3) # create empty vector
    BaseOcc_df <- t(as.data.frame(BaseOcc)) # make empty vector into empty data frame
    colnames(BaseOcc_df) <- c("decimalLatitude", "decimalLongitude", "year") # set column names
    for(i in 1:length(Years)){
      # create a data frame of latitude and longitude records of currently iterated year
      GbifFrame <- data.frame(decimalLatitude = Gbif[[i]]$data$decimalLatitude,
                              decimalLongitude = Gbif[[i]]$data$decimalLongitude,
                              year = rep(Years[i], length(Gbif[[i]]$data$decimalLatitude)))
      BaseOcc_df <- rbind(BaseOcc_df, GbifFrame)}
    ## Sanity check
    if(dim(BaseOcc_df)[1] == 1){ # if there is no occurence data
      next()}
    BaseOcc_df <- BaseOcc_df[-1,] # remove initial NA row
    ## Saving data frame
    save(BaseOcc_df, file = paste(Dir.OCCs,"/", Species_Pres[[OccIter]], "_", CountCodes,".rda",sep=""))}
   setTxtProgressBar(pb, OccIter) # pdate progress bar
   # save current iteration number to disk (used for jumping right back in if errors occur)
   write.table(OccIter, file = paste(mainDir, "/Breakage.txt", sep=""))
  }# occurence data loop
  file.remove(paste(mainDir, "/Breakage.txt", sep=""))
  setwd(mainDir)
  GbifStat <- "Done"
  return(GbifStat)
}# end of Mapping function

####--------------- PFTRasters [Region, Extent, RegionFile, CountCodes]
# (loading data, building species-specific trait mean rasters for study regions) ----
PFTRasters <- function(Region, Extent, RegionFile, CountCodes){
  print("##################################################")
  print(paste("Building mean trait rasters across ", RegionFile,sep=""))
  load(paste(Dir.TRY, "/SpeciesTraits.RData", sep="")) # load data
  RawTry_ras <- brick(paste(Dir.TRY,"/RawTRY-Global.nc", sep=""))
  # GLOBAL REFERENCE DATA (needed for rasterising and masking) ----
  ## NDVI (reference raster) ---
  NDVI_ras <- brick(paste(Dir.Gimms.Monthly,"/GlobalNDVI_20112015.nc", sep=""))
  ref_ras <- NDVI_ras[[6]]
  ref_ras[which(values(ref_ras) > -1)] <- 8888 # identify land pixels
  # REGION SELECTION----
  Shapes <- readOGR(Dir.Mask,'ne_50m_admin_0_countries', verbose = FALSE)
  RegObj <- RegionSelection(Region = Region, RegionFile = RegionFile, Extent = Extent)
  area <- RegObj[[1]]
  location <- RegObj[[2]]
  RegionFile <- RegObj[[3]]
  # CROPPING AND MASKING ----
  ## Reference cropping and masking
  ref_rasC <- crop(ref_ras, area) # cropping to extent
  ref_rasF <- mask(ref_rasC, Shapes[location,]) # masking via Shapefile
  # RAW TRY DATA ----
  RawTry_rasC <- crop(RawTry_ras, area) # cropping to extent
  RawTry_rasF <- mask(RawTry_rasC, Shapes[location,]) # masking via Shapefile
  writeRaster(x=RawTry_rasF, filename = paste(Dir.TRY,"/RawTRY-",RegionFile, sep=""),overwrite=TRUE, format="CDF")
  # CALCULATING MEAN RASTERS WITH DISTRIBUTION MAPS ----
  # create empty mean raster
  BaseMeans <- ref_rasF
  values(BaseMeans)[!is.na(values(BaseMeans))] <- 0
  # build brick for mean calculations
  BaseMeans <- brick(BaseMeans, BaseMeans, BaseMeans, BaseMeans)
```

```
  names(BaseMeans) <- c("Height", "NMass", "HCount", "NCount")
  # progress bar
  pb <- txtProgressBar(min = 0, max = length(list.files(Dir.OCCs)), style = 3) #
  # looping over all .rda occurence files previously downloaded
  for(OccRast in 1:length(list.files(Dir.OCCs))){
    # OCCURENCE ----
    load(paste(Dir.OCCs, "/", list.files(Dir.OCCs)[OccRast], sep=""))
    ## Converting to SpatialPoints
    pts <- data.frame(y = BaseOcc_df$decimalLatitude, x = BaseOcc_df$decimalLongitude,
                      z = rep(1, length(BaseOcc_df$decimalLongitude)))
    pts <- na.omit(pts) # remove NA rows
    coordinates(pts) = ~x+y # convert x and y to coordinates
    # RASTERISING ----
    # create raster to be filled
    rast <- raster(ext=extent(ref_ras), resolution=res(ref_ras))
    # rasterize irregular points
    # we use a mean function here to regularly grid the irregular input points
    rasOut<-rasterize(x = pts, y = rast, field = pts$z, fun = max)
    ## Occurence cropping and masking
    rasC <- crop(rasOut, area) # cropping to extent
    rasF <- mask(rasC, Shapes[location,]) # masking via Shapefile
    # TRAIT MEANS ----
    # loading data of currently iterated on species
    Grep <- list.files(Dir.OCCs)[OccRast]
    Grep <- gsub(x = Grep, pattern = CountCodes, replacement = "")
    Grep <- gsub(x = Grep, pattern = "_.rda", replacement = "")

    NMass <- CorrectedSpec_df$Nmass[which(CorrectedSpec_df$Species == Grep)]
    Height <- CorrectedSpec_df$Height[which(CorrectedSpec_df$Species == Grep)]
    # BUILDING MAP
    Identifier <- which(!is.na(values(rasF)))
    if(length(NMass) != 0){
      if(!is.nan(NMass)){ # add current species-NMass to raster layer and bump up count by 1
        values(BaseMeans$NMass)[Identifier] <- values(BaseMeans$NMass)[Identifier] + NMass
        values(BaseMeans$NCount)[Identifier] <- values(BaseMeans$NCount)[Identifier] + 1}}
    if(length(Height) != 0){
      if(!is.nan(Height)){ # add current species-Height to raster layer and bump up count by 1
        values(BaseMeans$Height)[Identifier] <- values(BaseMeans$Height)[Identifier] + Height
        values(BaseMeans$HCount)[Identifier] <- values(BaseMeans$HCount)[Identifier] + 1}}
    setTxtProgressBar(pb, OccRast) # update progress bar
  } # OccRast-loop
  # CALCULATE MEANS ----
  TestHeight <- BaseMeans$Height/BaseMeans$HCount
  values(TestHeight)[which(values(TestHeight) > quantile(values(TestHeight), .95, na.rm = TRUE))] <- NA
  TestNMass <- BaseMeans$NMass/BaseMeans$NCount
  values(TestNMass)[which(values(TestNMass) > quantile(values(TestNMass), .95, na.rm = TRUE))] <- NA
  Means_ras <- brick(TestHeight, TestNMass)
  # SAVING DATA ----
  writeRaster(x=Means_ras, filename = paste(Dir.TRY,"/TRY-",RegionFile, sep=""),overwrite=TRUE, format="CDF")
}# PFTRasters
```

A.3.4 COMPADRE

Chunk 12: Extracting and rasterising COMPADRE data from COMPADRE data base for each study region (RasterCOMPADRE).

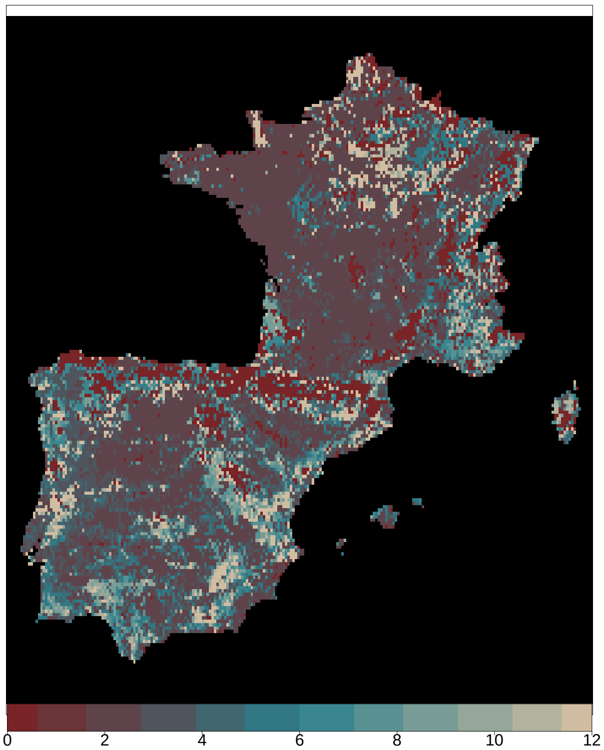
```
####--------------- RasterCOMPADRE [Variable, Region, RegionFile, Extent]
# (Selecting specified COMPADRE data, rasterising, saving the raster, limitting to a region) ----
RasterCOMPADRE <- function(Variable, Region, RegionFile, Extent){
  print("#################################################")
  print(paste("Rasterising COMPADRE ", Variable, " across ", RegionFile, sep=""))
  # LOADING DATA ----
  Compadre_df <- read.csv(paste(Dir.Compadre, "/allCOMPADREOutput.csv", sep="")) # load data frame
  NDVI_ras <- brick(paste(Dir.Gimms.Monthly,"/GlobalNDVI_20112015.nc", sep="")) # reference raster
  ref_ras <- NDVI_ras[[6]] # select only one years data
  ref_ras[which(values(ref_ras) > -1)] <- 8888 # select only land pixels and set them to -8888
  # REGION SELECTION----
  Shapes <- readOGR(Dir.Mask,'ne_50m_admin_0_countries', verbose = FALSE)
  RegObj <- RegionSelection(Region = Region, RegionFile = RegionFile, Extent = Extent)
  area <- RegObj[[1]]
  location <- RegObj[[2]]
  RegionFile <- RegObj[[3]]
  # DATA MANIPULATION ----
  if(Variable == "FastSlow"){ # analysis of PCA axes
    FSVars <- c("GenT", "H", "La", "GrowSSD", "ShriSSD", "RepSSD", "S", "R0", "Lmean")
    VariableCol <- match(FSVars, colnames(Compadre_df))
    FSLoads <- list(c(.87,.53,.7,-.8,.04,-.81,-.25,-.03,.12), # PCA 1 according to Salguero-Gomez, 2017
                    c(.15,.27,.28,-.05,-.79,.32,.65,.7,.26)) # PCA 2 according to Salguero-Gomez, 2017
    pts <- na.omit(data.frame(y = Compadre_df$Lat, x = Compadre_df$Lon, z = Compadre_df[,VariableCol]))
    PCA1_df <- t(t(pts[,-1:-2]) * FSLoads[[1]]) # multiplying by first axis loadings
    PCA1_df <- rowSums(PCA1_df) # build sums for single index along PCA 1
    PCA1_df <- cbind(pts[,1:2], PCA1_df) # binding with coordinates
    coordinates(PCA1_df) = ~x+y # convert x and y to coordinates
    PCA2_df <- t(t(pts[,-1:-2]) * FSLoads[[2]]) # multiplying by second axis loadings
    PCA2_df <- rowSums(PCA2_df) # build sums for single index along PCA 2
    PCA2_df <- cbind(pts[,1:2], PCA2_df) # binding with coordinates
    coordinates(PCA2_df) = ~x+y # convert x and y to coordinates
    ## Rasterizing ----
    rast <- raster(ext=extent(ref_ras), resolution=res(ref_ras)) # create raster to be filled
    rasOut1 <-rasterize(x = PCA1_df, y = rast, field = PCA1_df$PCA1_df, fun = mean) # rasterize irregular points
    rasOut2 <-rasterize(x = PCA2_df, y = rast, field = PCA2_df$PCA2_df, fun = mean) # rasterize irregular points
    rasOut <- brick(rasOut1, rasOut2)
    names(rasOut) <- c("FS PCA1", "FS PCA2")
  }else{ # single variable desired
    VariableCol <- which(colnames(Compadre_df) == Variable) # figure out the position of the desired Variable
    pts <- na.omit(data.frame(y = Compadre_df$Lat, x = Compadre_df$Lon, z = Compadre_df[,VariableCol]))
    coordinates(pts) = ~x+y # convert x and y to coordinates
    rast <- raster(ext=extent(ref_ras), resolution=res(ref_ras)) # create raster to be filled
    rasOut<-rasterize(x = pts, y = rast, field = pts$z, fun = mean) # rasterize irregular points
  }
    # CROPPING AND MASKING ----
    rasC <- crop(rasOut, area) ## Occurence cropping and masking
    rasF <- mask(rasC, Shapes[location,]) # masking via Shapefile
    # DATA EXPORT ----
    Dir.Temp.Compadre <- paste(Dir.Compadre, Variable, sep="/")
    dir.create(Dir.Temp.Compadre)
    values(rasF)[which(values(rasF) == Inf)] <- NA # get rid off Inf values (when dealing with Rho)
    invisible(writeRaster(rasF, filename = paste(Dir.Temp.Compadre, "/",Variable,"_",RegionFile,sep=""),
                          overwrite=TRUE, format="CDF"))
}# end of RasterCOMPADRE
```

A.4 Results

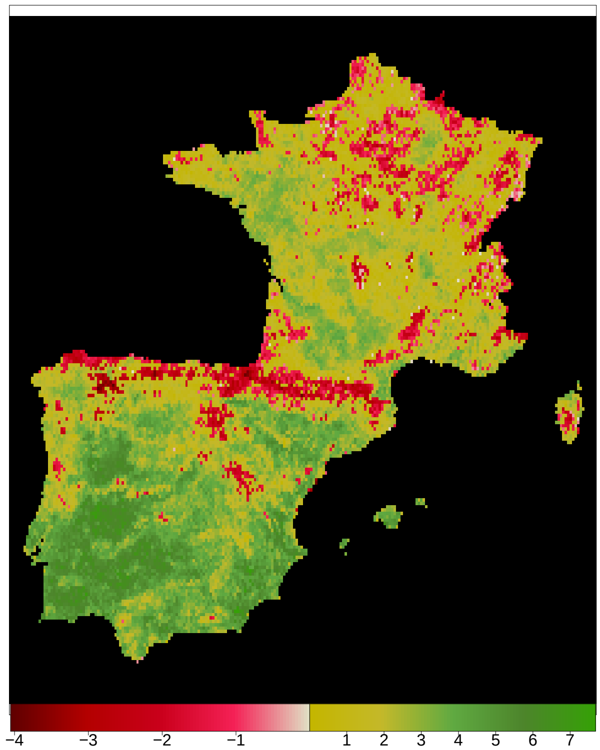
A.4.1 Vegetation Memory Models

A.4.1.1 Iberian Region

Model Coefficients (Soil Layer 2)



(a) SoilLag



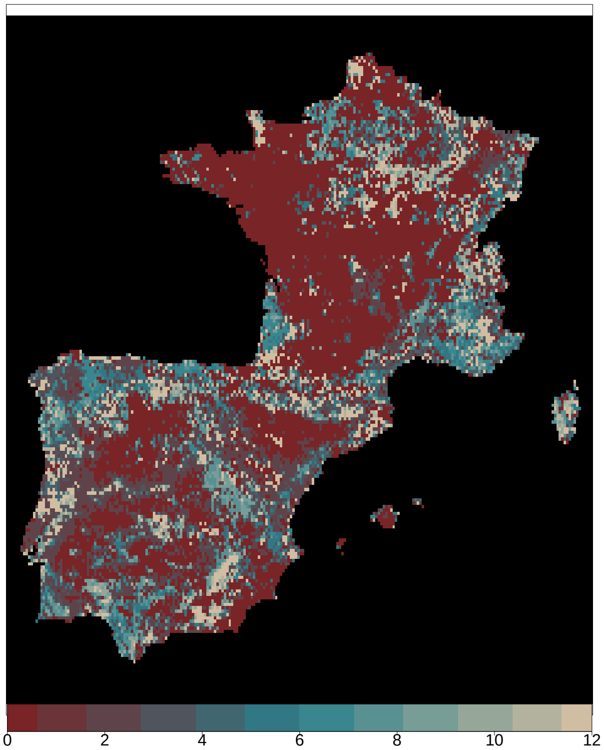
(b) Qsoil

Figure A.10: Vegetation Response Coefficients (Iberian Region; Qsoil2) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). Figure established via Chunk 24.

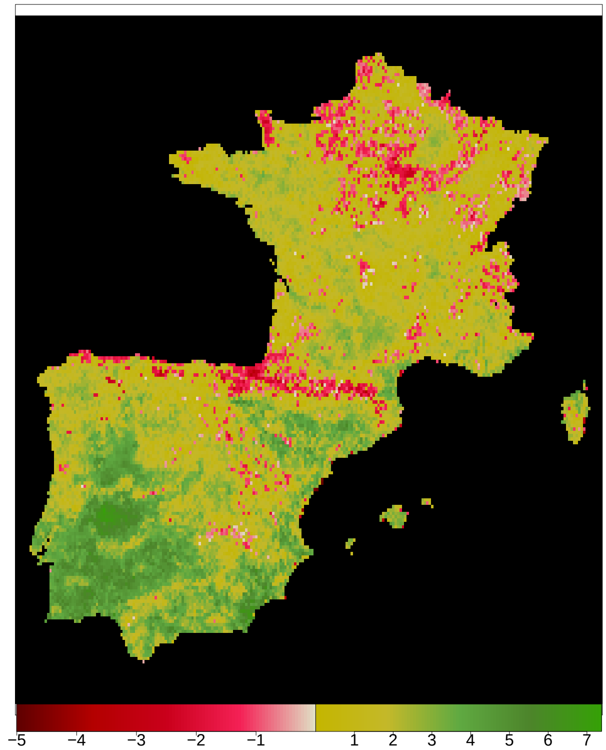
Table A.3: Mann-Whitney U-Test (Iberian Region, Qsoil2 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil2	Tair
NDVI t-1	3.855	NA	259724355	316424407
Qsoil	2.166	0	NA	241657391
Tair	1.195	0	0	NA

Model Coefficients (Soil Layer 3)



(a) SoilLag



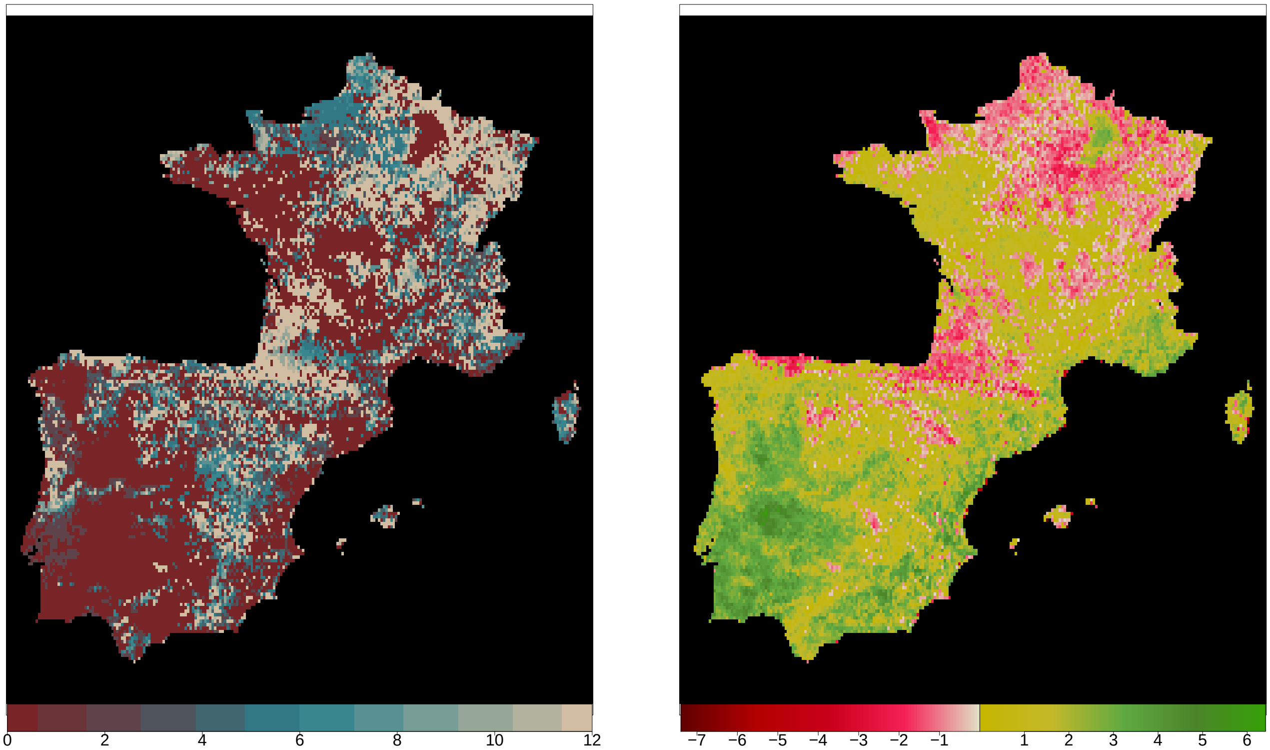
(b) Qsoil

Figure A.11: Vegetation Response Coefficients (Iberian Region; Qsoil3) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). Figure established via Chunk 24.

Table A.4: Mann-Whitney U-Test (Iberian Region, Qsoil3 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil3	Tair
NDVI t-1	3.863	NA	285161401	316091617
Qsoil	1.680	0	NA	213476555
Tair	1.199	0	0	NA

Model Coefficients (Soil Layer 4)



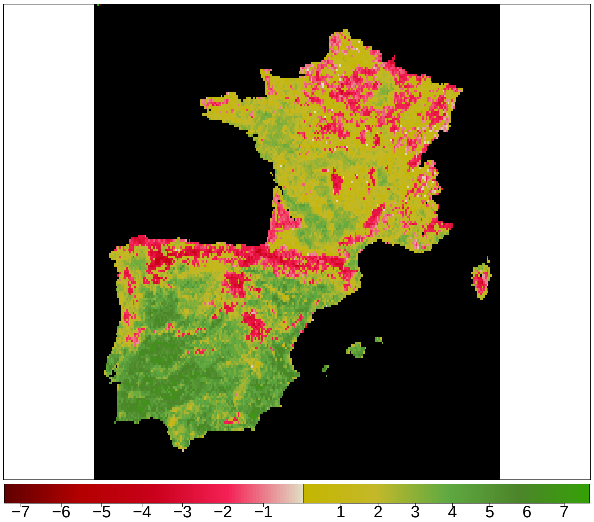
(a) SoilLag **(b)** Qsoil

Figure A.12: Vegetation Response Coefficients (Iberian Region; Qsoil4) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). Figure established via Chunk 24.

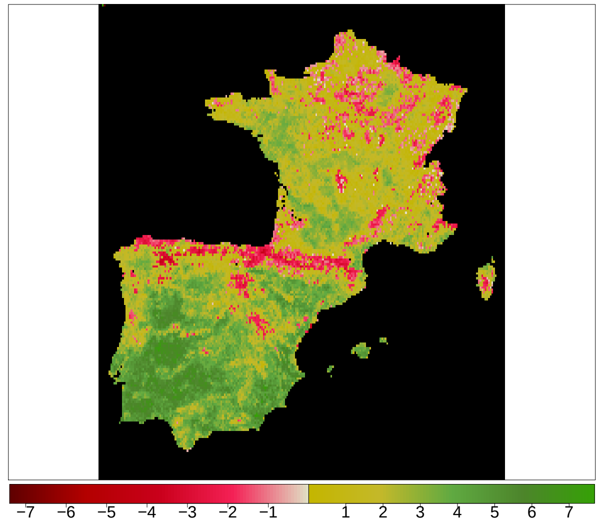
Table A.5: Mann-Whitney U-Test (Iberian Region, Qsoil4 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil4	Tair
NDVI t-1	3.852	NA	306106900	3.16e+08
Qsoil	1.229	0	NA	1.80e+08
Tair	1.191	0	0	NA

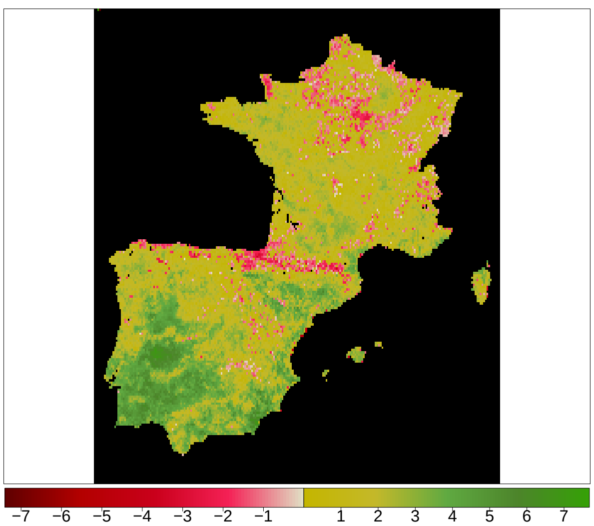
Soil Layer Comparison



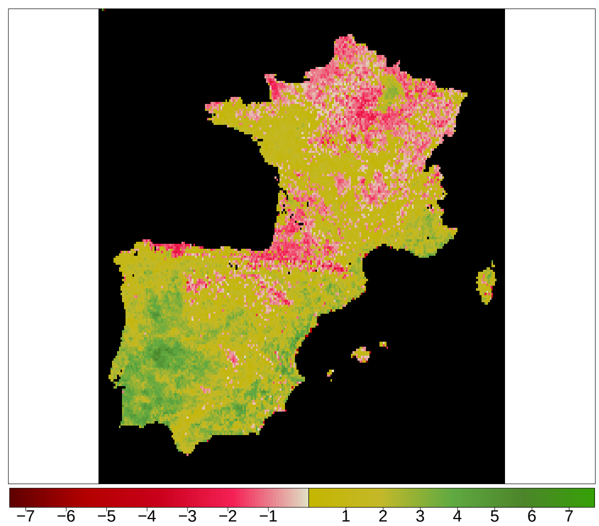
(a) Qsoil1



(b) Qsoil2



(c) Qsoil3



(d) Qsoil4

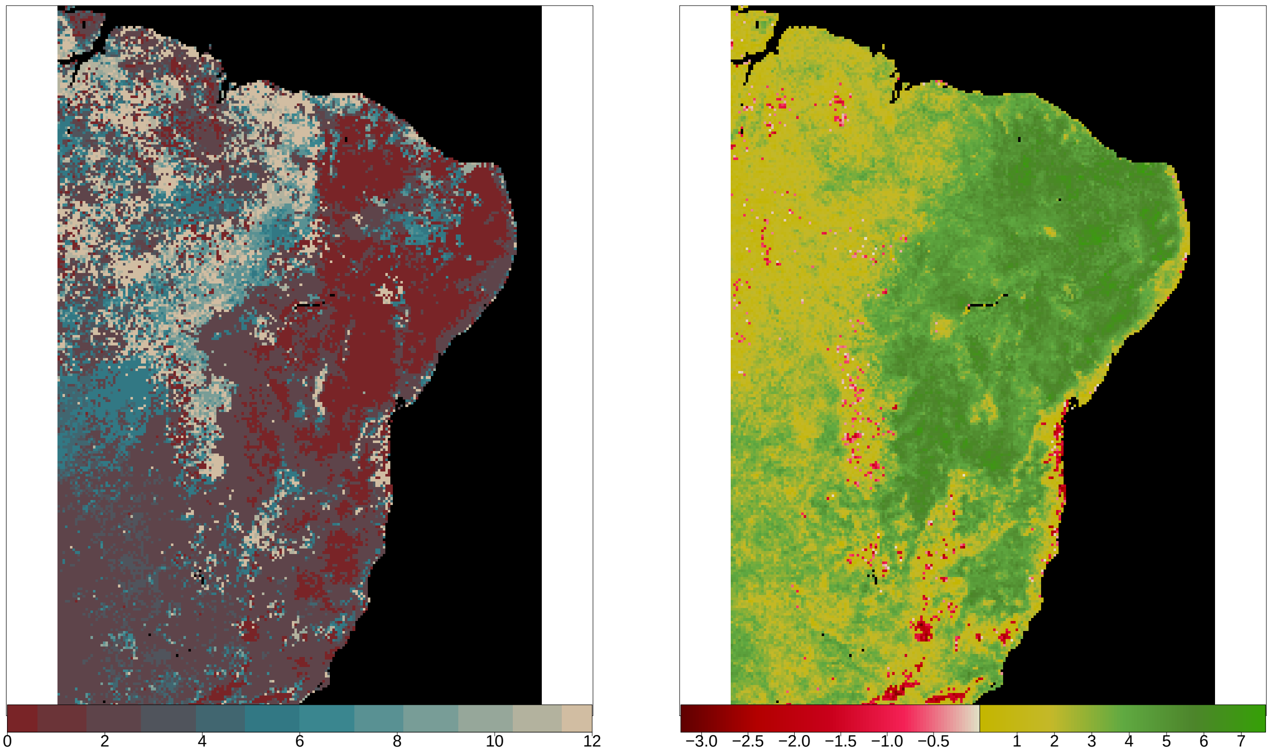
Figure A.13: Vegetation Response Coefficients (Iberian Region, QSoil Layers) - Vegetation response coefficients of different Qsoil layers. These are also contained in figures 3.1, A.10, A.11, and A.12 and have been scaled to be represented on the same colour axis for comparability. Figure established via Chunk 27.

Table A.6: Mann-Whitney U-Test (Iberian Region, Qsoil Layers) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 28.

		U-Test			
	Medians	Qsoil1	Qsoil2	Qsoil3	Qsoil4
Qsoil1	2.228	NA	169919584	201893947	235372117
Qsoil2	2.166	0	NA	195575594	228882554
Qsoil3	1.680	0	0	NA	200374929
Qsoil4	1.229	0	0	0	NA

A.4.1.2 Caatinga

Model Coefficients (Soil Layer 2)



(a) SoilLag

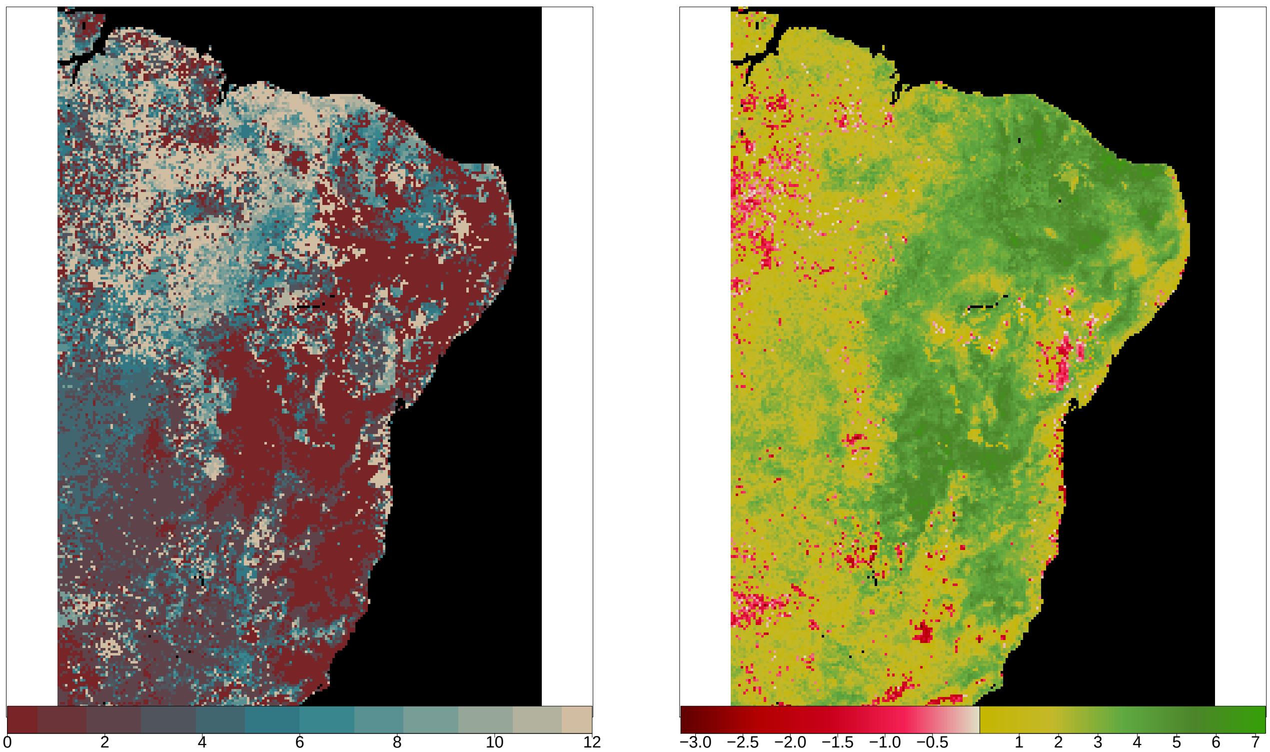
(b) Qsoil

Figure A.14: Vegetation Response Coefficients (Caatinga; Qsoil2) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

Table A.7: Mann-Whitney U-Test (Caatinga, Qsoil2 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil2	Tair
NDVI t-1	4.205	NA	899172763	1.12e+09
Qsoil	3.053	0	NA	9.65e+08
Tair	1.264	0	0	NA

Model Coefficients (Soil Layer 3)



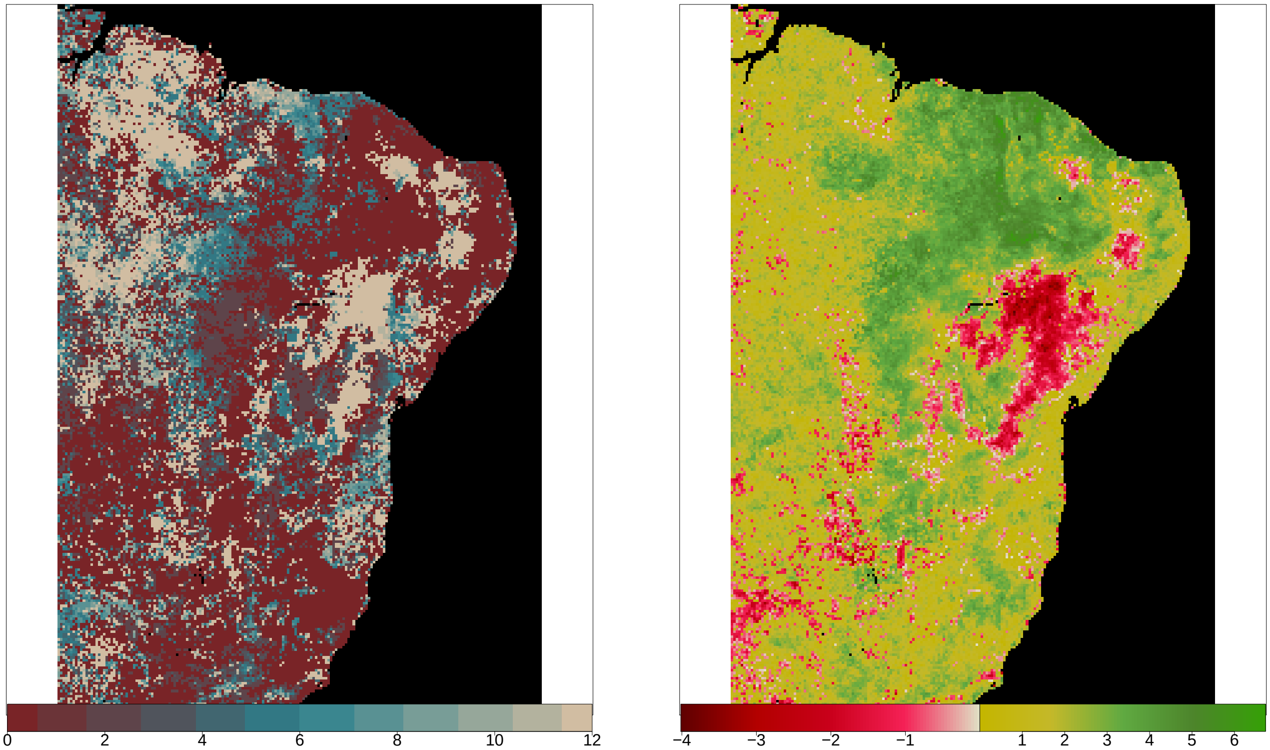
(a) SoilLag **(b)** Qsoil

Figure A.15: Vegetation Response Coefficients (Caatinga; Qsoil3) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

Table A.8: Mann-Whitney U-Test (Caatinga, Qsoil3 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil3	Tair
NDVI t-1	4.207	NA	1.028e+09	1.116e+09
Qsoil	2.352	0	NA	8.525e+08
Tair	1.268	0	0.000e+00	NA

Model Coefficients (Soil Layer 4)



(a) SoilLag

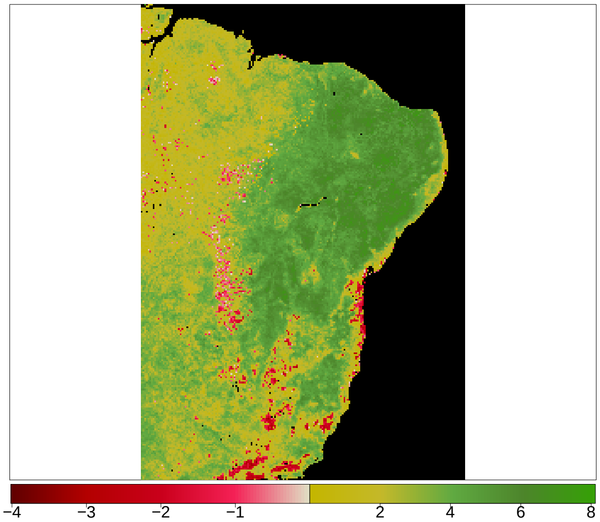
(b) Qsoil

Figure A.16: Vegetation Response Coefficients (Caatinga; Qsoil4) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

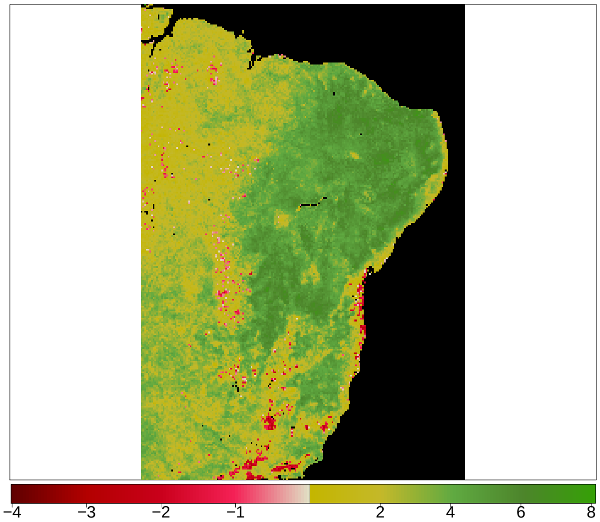
Table A.9: Mann-Whitney U-Test (Caatinga, Qsoil4 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil4	Tair
NDVI t-1	4.208	NA	1.152e+09	1.116e+09
Qsoil	1.730	0	NA	7.298e+08
Tair	1.268	0	0.000e+00	NA

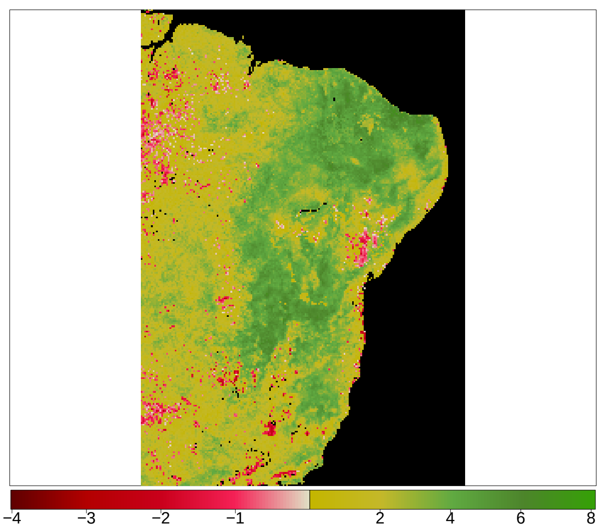
Soil Layer Comparison



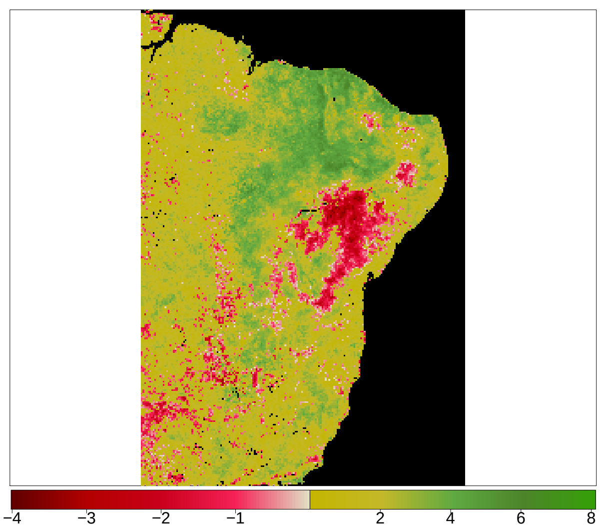
(a) Qsoil1



(b) Qsoil2



(c) Qsoil3



(d) Qsoil4

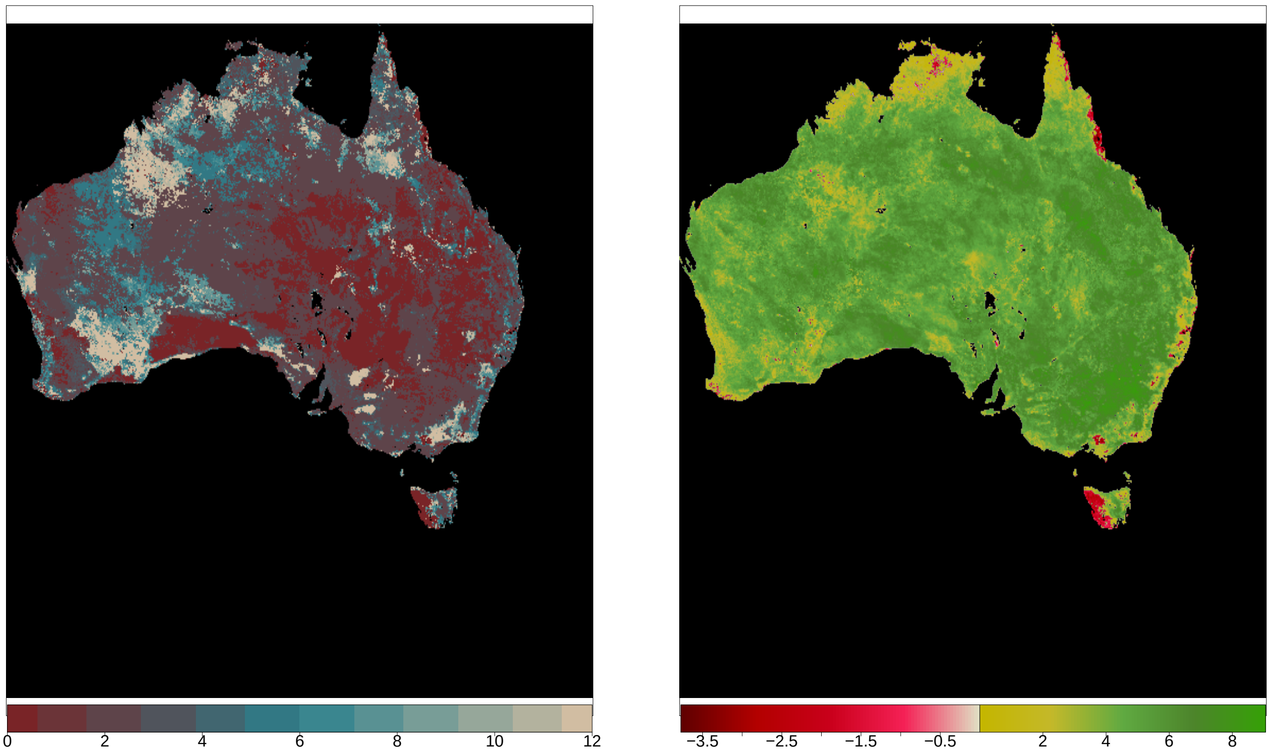
Figure A.17: Vegetation Response Coefficients (Caatinga, QSoil Layers) - Vegetation response coefficients of different Qsoil layers. These are also contained in figures 3.2, A.14, A.15, and A.16 and have been scaled to be represented on the same colour axis for comparability. Figure established via Chunk 27.

Table A.10: Mann-Whitney U-Test (Caatinga, Qsoil Layers) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. *p*-values belonging to these *U*-values are represented in the lower-lefthand block of cells. Established via Chunk 28.

		U-Test			
	Medians	Qsoil1	Qsoil2	Qsoil3	Qsoil4
Qsoil1	3.089	NA	662588655	810071500	963908394
Qsoil2	3.053	0	NA	797671685	954810187
Qsoil3	2.352	0	0	NA	812573884
Qsoil4	1.730	0	0	0	NA

A.4.1.3 Australia

Model Coefficients (Soil Layer 2)



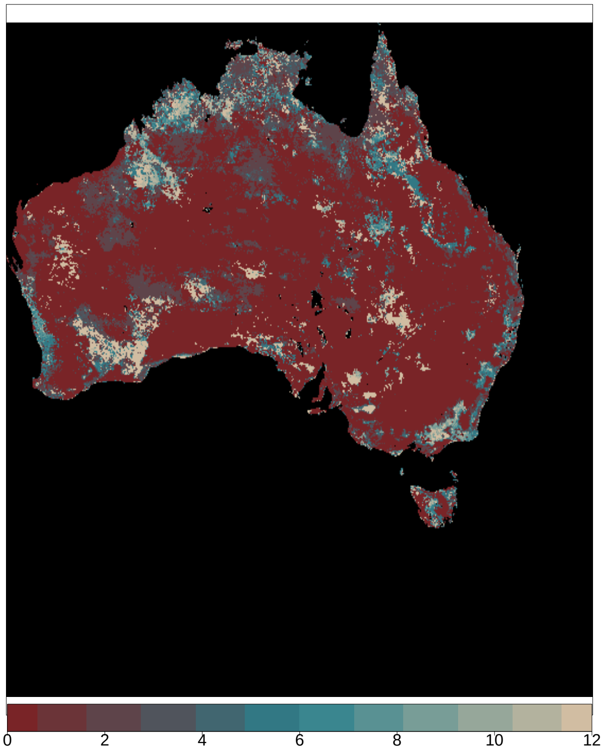
(a) SoilLag **(b)** Qsoil

Figure A.18: Vegetation Response Coefficients (Australia; Qsoil2) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

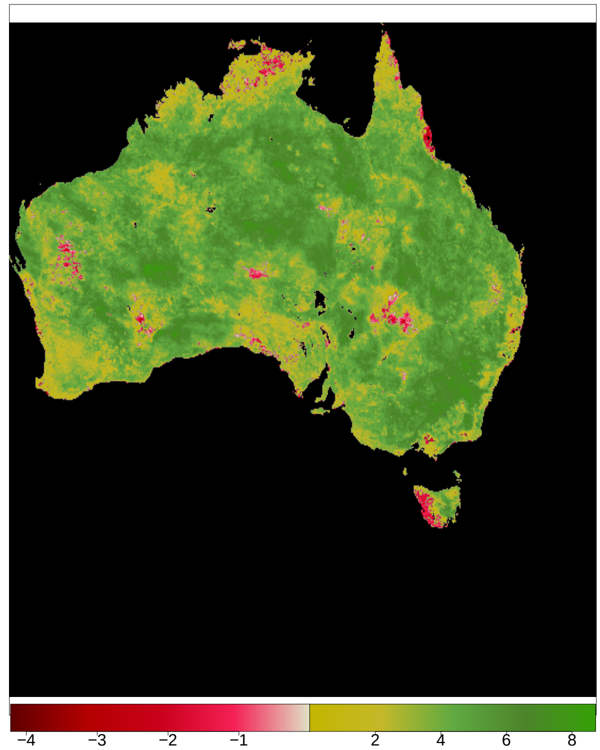
Table A.11: Mann-Whitney U-Test (Australia, Qsoil2 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil2	Tair
NDVI t-1	7.452	NA	8.541e+09	9.680e+09
Qsoil	5.224	0	NA	8.556e+09
Tair	2.852	0	0.000e+00	NA

Model Coefficients (Soil Layer 3)



(a) SoilLag



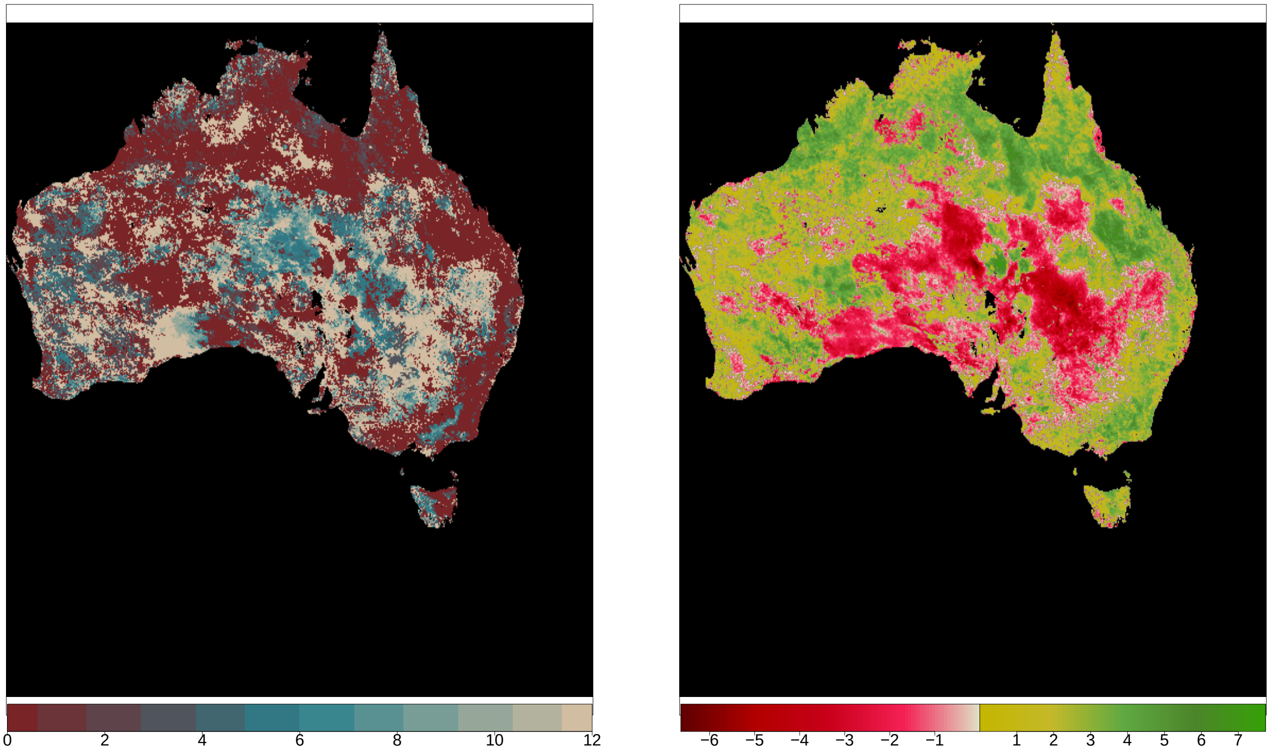
(b) Qsoil

Figure A.19: Vegetation Response Coefficients (Australia; Qsoil3) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

Table A.12: Mann-Whitney U-Test (Australia, Qsoil3 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil3	Tair
NDVI t-1	7.454	NA	9.064e+09	9.678e+09
Qsoil	4.415	0	NA	7.490e+09
Tair	2.853	0	0.000e+00	NA

Model Coefficients (Soil Layer 4)



(a) SoilLag

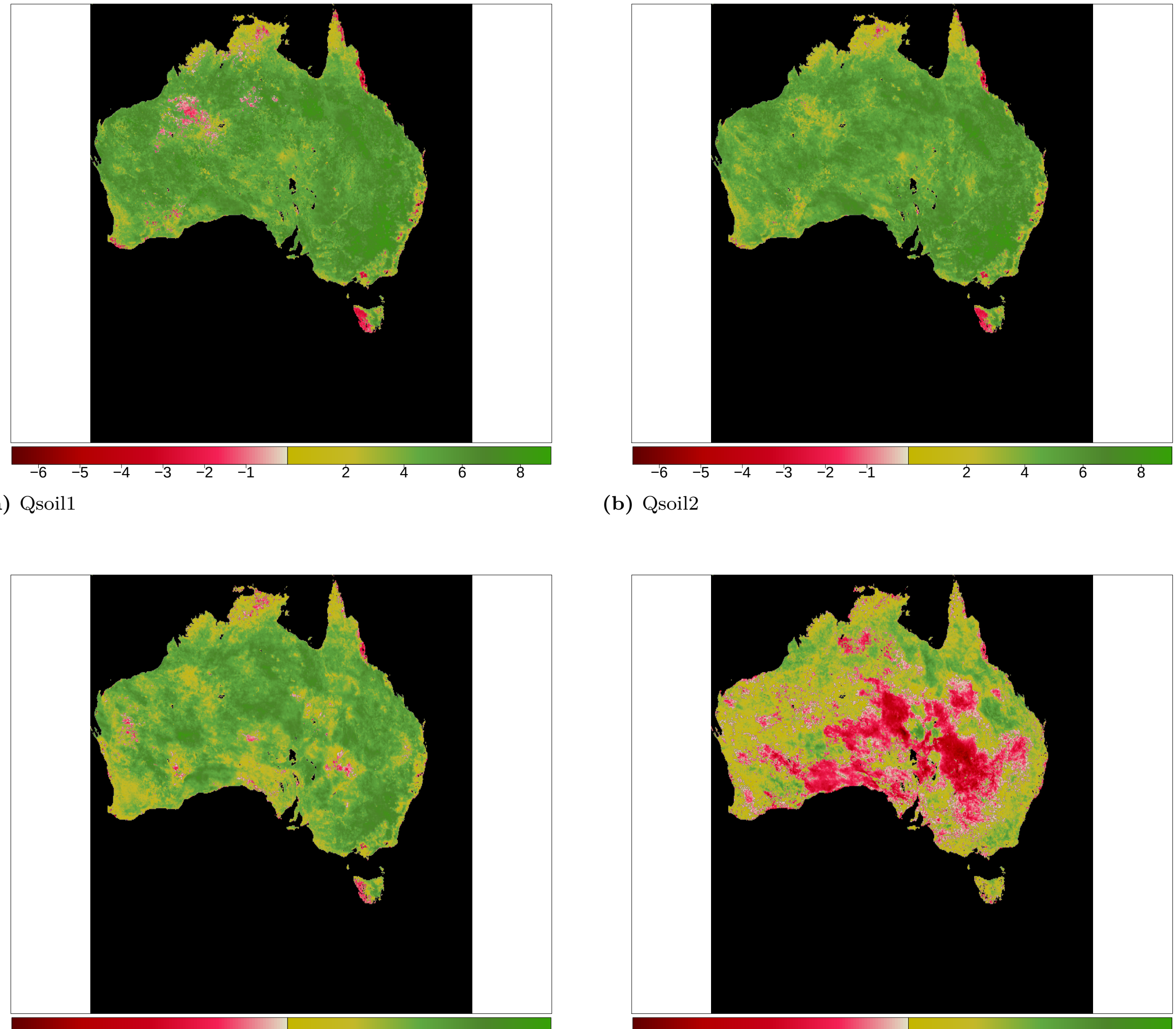
(b) Qsoil

Figure A.20: Vegetation Response Coefficients (Australia; Qsoil4) - Coefficients of vegetation memory obtained via model selection (figure 2.9) and PCA regression (figure 2.13). For an interpretation of these coefficients, see table 2.4. Figure established via Chunk 24.

Table A.13: Mann-Whitney U-Test (Australia, Qsoil4 Model) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the column name variable. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 25.

		U-Test		
	Medians	NDVI [t-1]	Qsoil4	Tair
NDVI t-1	7.469	NA	9.797e+09	9.674e+09
Qsoil	1.877	0	NA	3.235e+09
Tair	2.860	0	0.000e+00	NA

Soil Layer Comparison



(a) Qsoil1 **(b)** Qsoil2

(c) Qsoil3 **(d)** Qsoil4

Figure A.21: Vegetation Response Coefficients (Australia, QSoil Layers) - Vegetation response coefficients of different Qsoil layers. These are also contained in figures 3.3, A.18, A.19, and A.20 and have been scaled to be represented on the same colour axis for comparability. Figure established via Chunk 27.

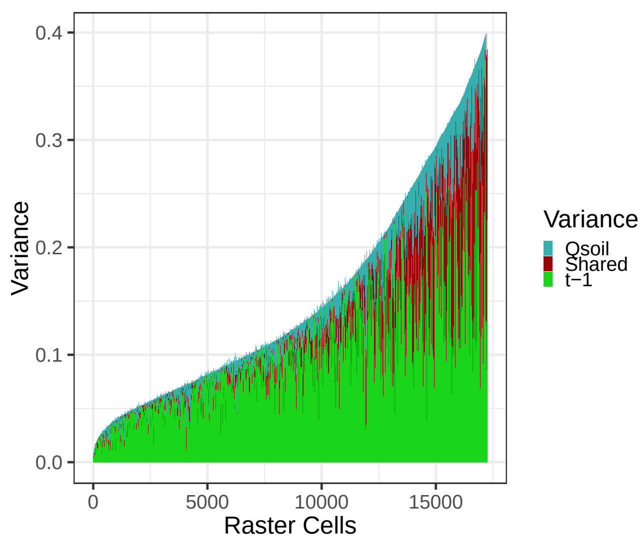
Table A.14: Mann-Whitney U-Test (Australia, Qsoil Layers) - U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand cells. p-values belonging to these U-values are represented in the lower-lefthand block of cells. Established via Chunk 28.

		U-Test			
	Medians	Qsoil1	Qsoil2	Qsoil3	Qsoil4
Qsoil1	5.326	NA	5.141e+09	6.445e+09	9.032e+09
Qsoil2	5.224	0	NA	6.353e+09	9.135e+09
Qsoil3	4.415	0	0.000e+00	NA	8.428e+09
Qsoil4	1.877	0	0.000e+00	0.000e+00	NA

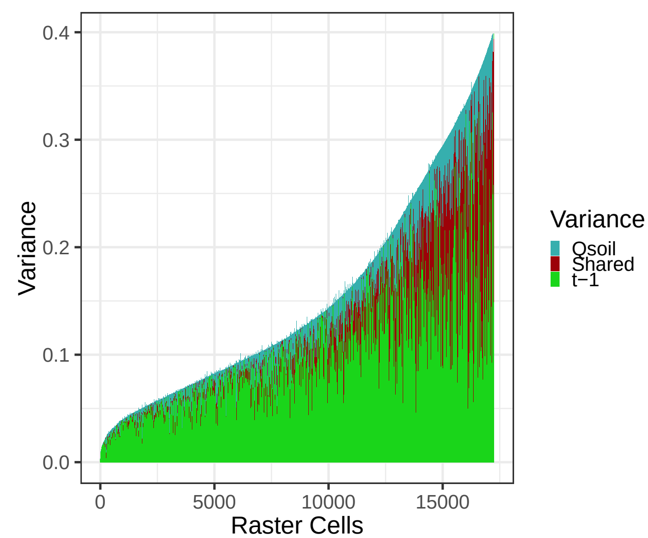
A.4.2 Variance Partitioning

A.4.2.1 Iberian Region

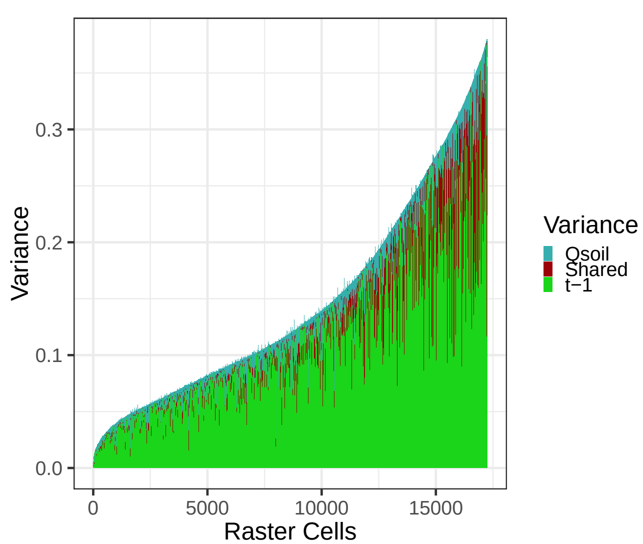
Variance partitioning of $NDVI_{[t-1]}$ and Qsoil has been assessed for all Qsoil layers. These results are presented in figure A.22 and show a clear pattern of $NDVI_{[t-1]}$ explaining an overwhelming majority of NDVI z-scores. Variance explained by Qsoil and variance shared by Qsoil and $NDVI_{[t-1]}$ decrease through the soil layers.
Spatial patterns of explained variance (see figure A.23) reveal that $NDVI_{[t-1]}$ and Qsoil are the most valuable as predictors NDVI across the southern dryland regions of Spain and Portugal.



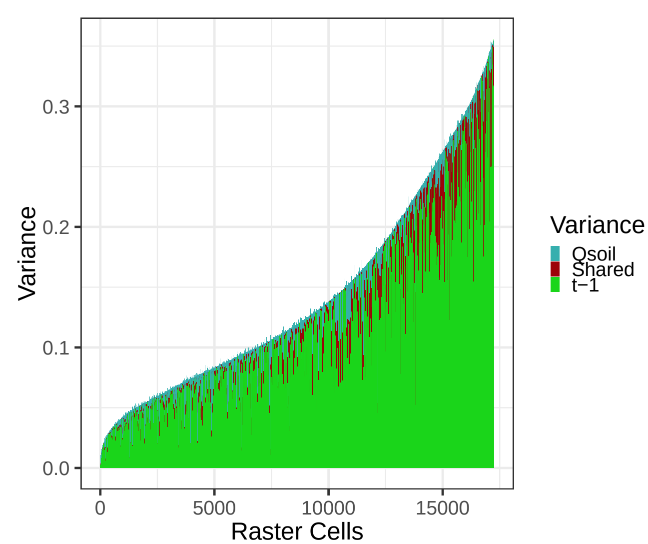
(a) Soil Layer 1



(b) Soil Layer 2

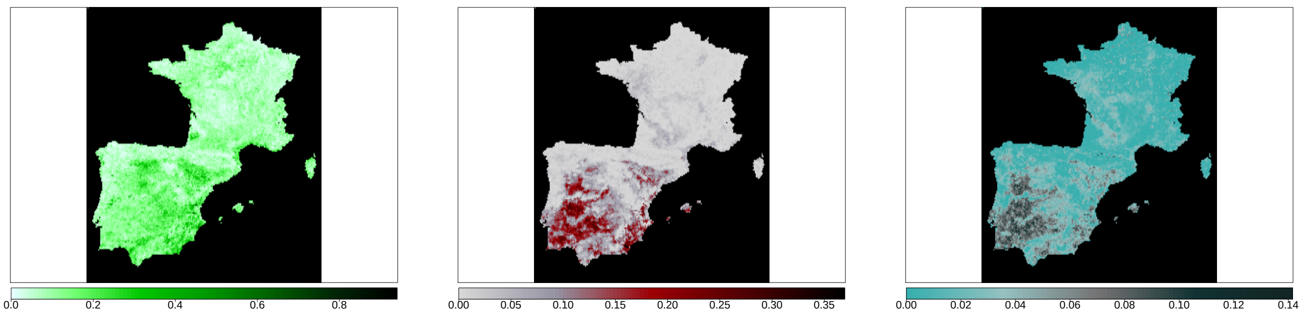


(c) Soil Layer 3

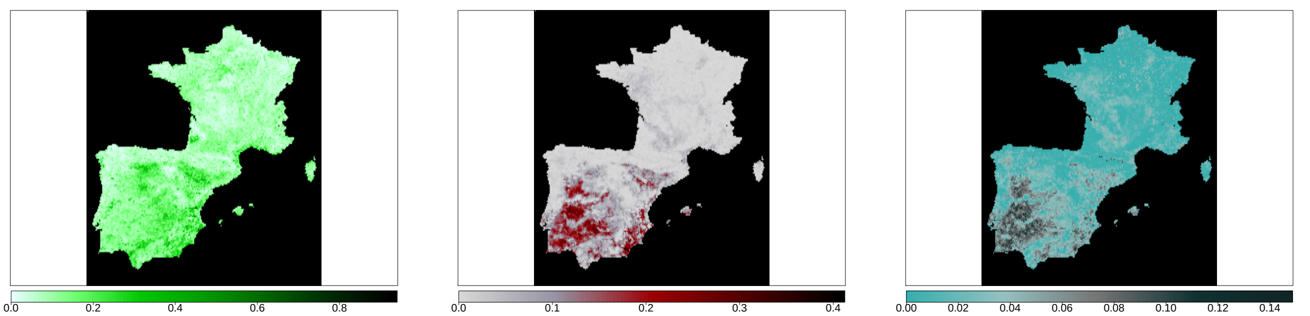


(d) Soil Layer 4

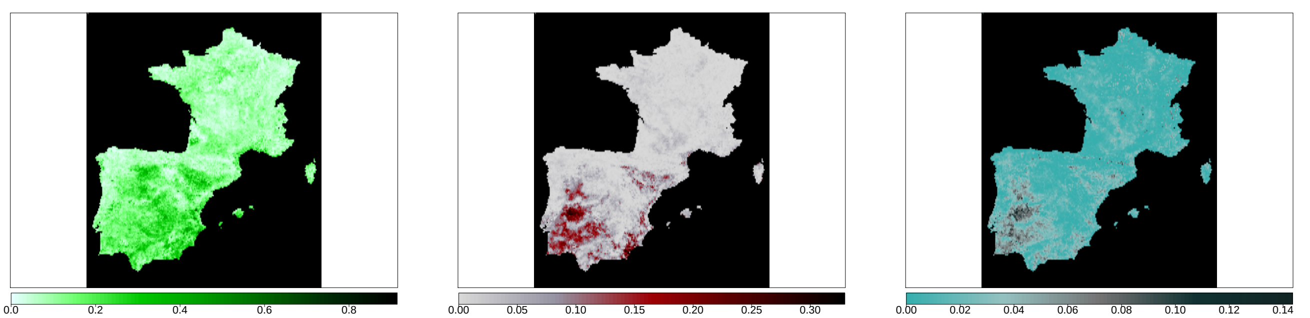
Figure A.22: Variance Partitioning (Iberian Region; Qsoil1) - Variance of NDVI anomalies explained by (a) full models of intrinsic and extrinsic memory, (b) intrinsic memory, (c) shared variance, and (d) extrinsic memory. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.



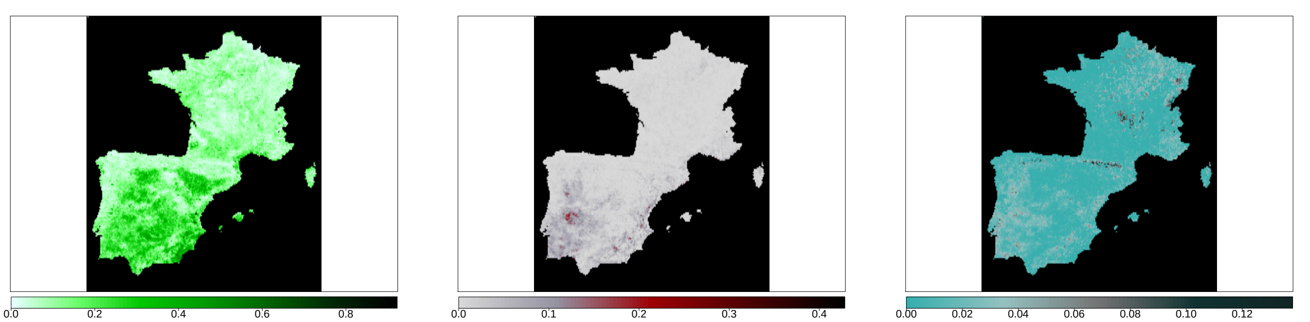
(a) Total **(b)** t-1 **(c)** Shared



(d) Qsoil **(e)** Total **(f)** t-1



(g) Shared **(h)** Qsoil **(i)** Total

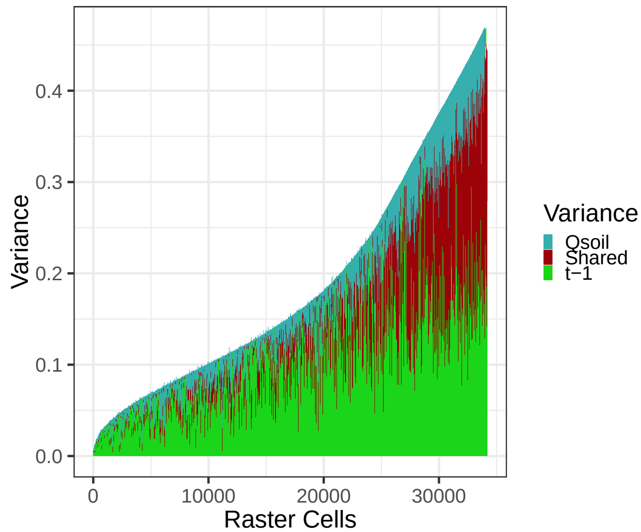


(j) t-1 **(k)** Shared **(l)** Qsoil

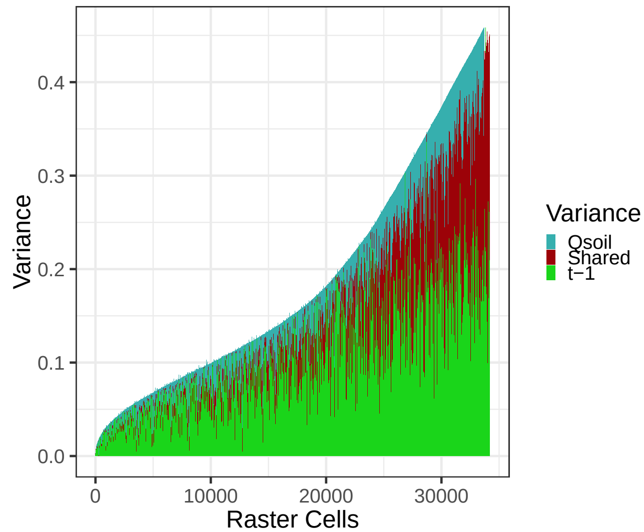
Figure A.23: Variance Partitioning (Iberian Region) - Variance of NDVI anomalies explained by (a,d,g,j) full models of intrinsic and extrinsic memory, (b,e,h,k) intrinsic memory, (c,f,i,l) shared variance, and (d) extrinsic memory across soil layers Qsoil1 to Qsoil4. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.

A.4.2.2 Caatinga

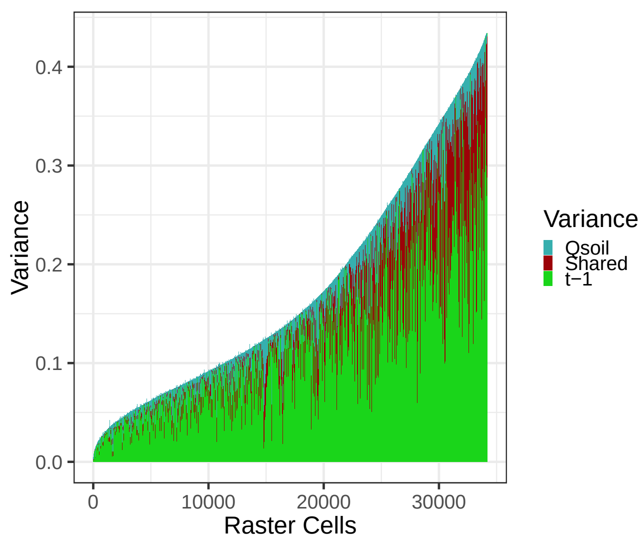
Variance partitioning of $NDVI_{[t-1]}$ and Qsoil has been assessed for all Qsoil layers. These results are presented in figure A.24 and show a clear pattern of $NDVI_{[t-1]}$ explaining an overwhelming majority of NDVI z-scores. Variance explained by Qsoil and variance shared by Qsoil and $NDVI_{[t-1]}$ decrease through the soil layers.
Spatial patterns of explained variance (see figure A.25) reveal that $NDVI_{[t-1]}$ and Qsoil are the most valuable as predictors NDVI across the north-eastern dryland region.



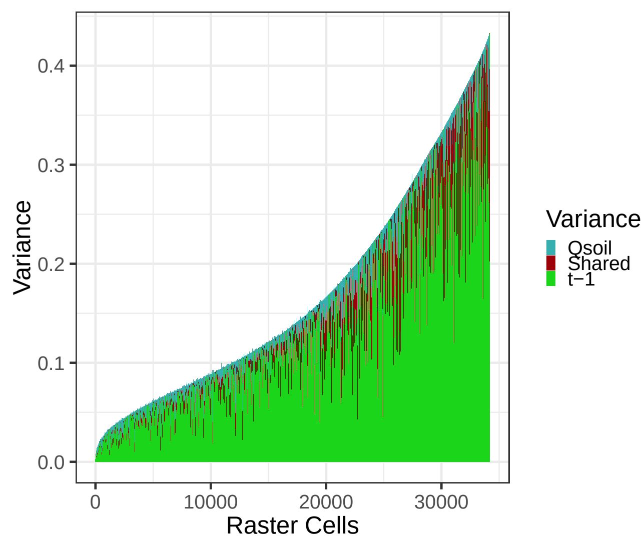
(a) Soil Layer 1



(b) Soil Layer 2



(c) Soil Layer 3



(d) Soil Layer 4

Figure A.24: Variance Partitioning (Caatinga; Qsoil1) - Variance of NDVI anomalies explained by (a) full models of intrinsic and extrinsic memory, (b) intrinsic memory, (c) shared variance, and (d) extrinsic memory. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.

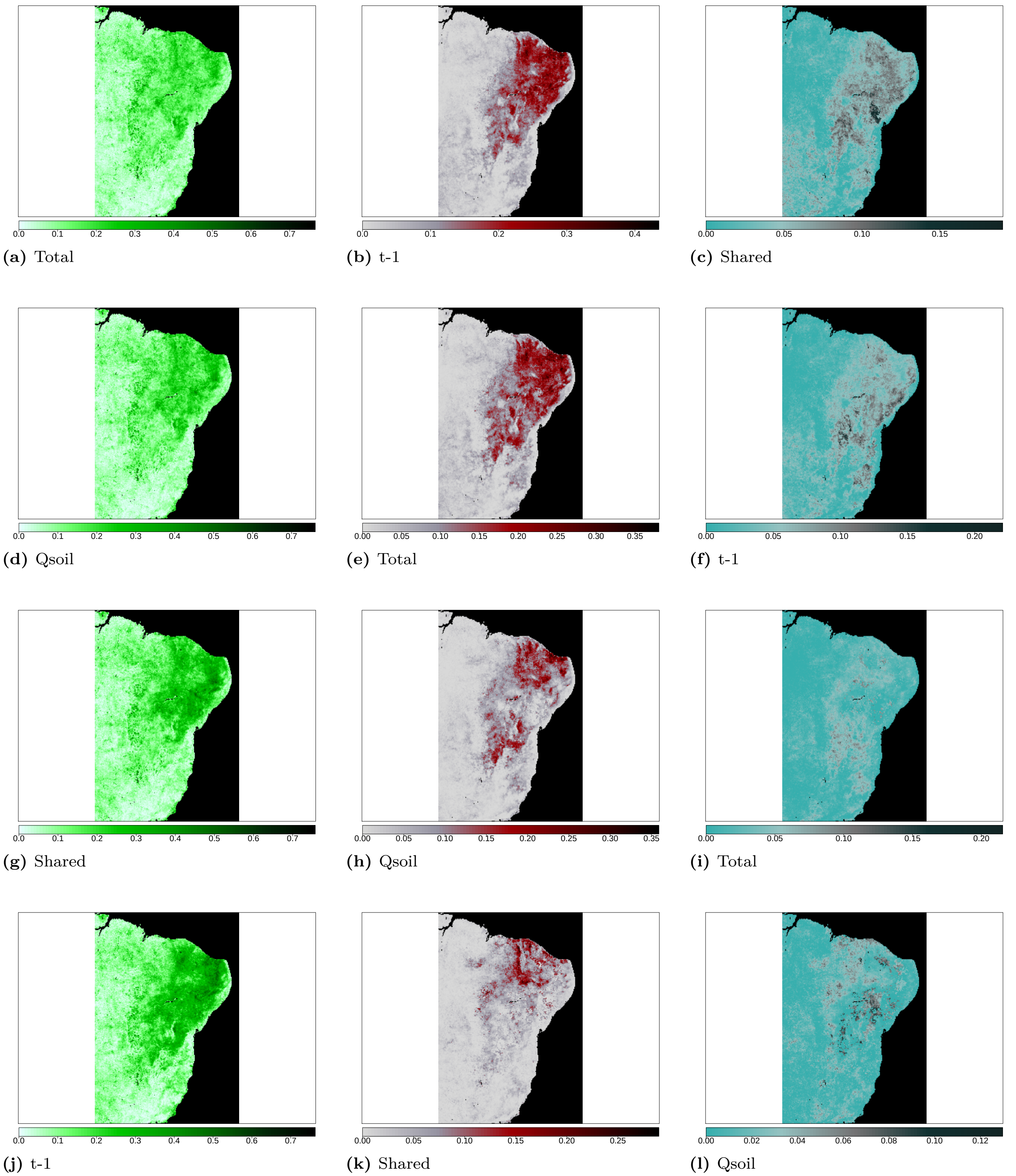
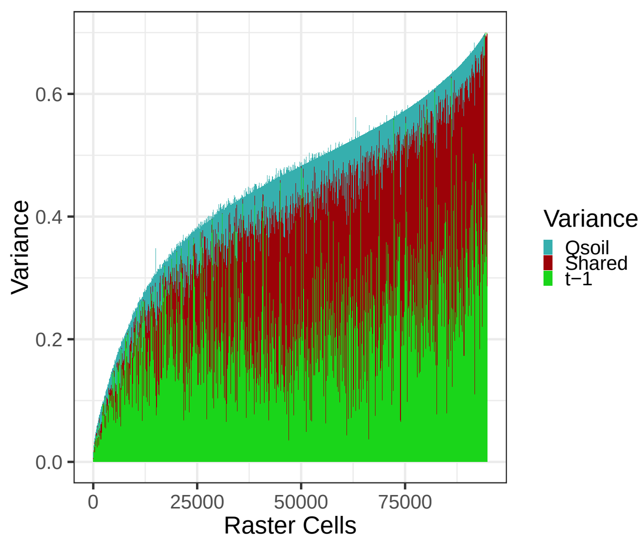


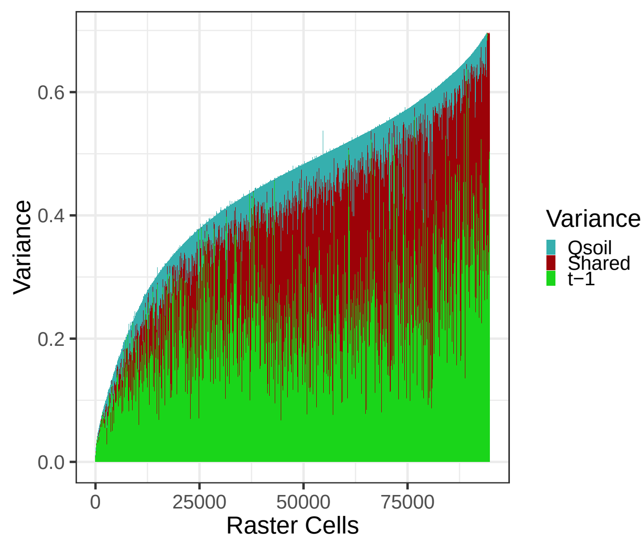
Figure A.25: Variance Partitioning (Caatinga) - Variance of NDVI anomalies explained by (a,d,g,j) full models of intrinsic and extrinsic memory, (b,e,h,k) intrinsic memory, (c,f,i,l) shared variance, and (d) extrinsic memory across soil layers Qsoil1 to Qsoil4. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.

A.4.2.3 Australia

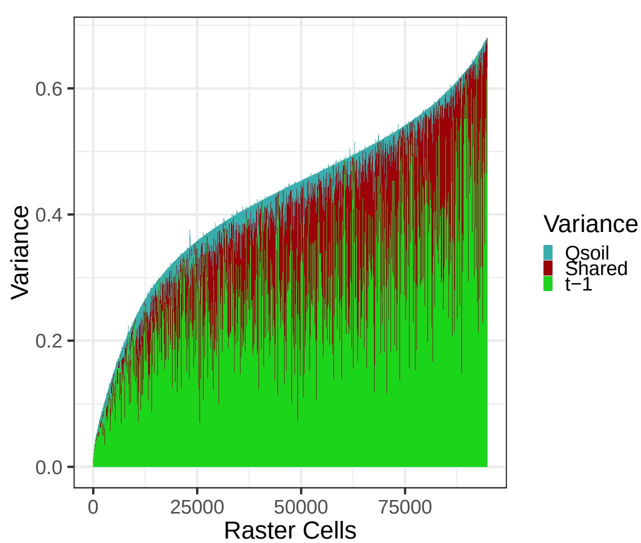
Variance partitioning of $NDVI_{[t-1]}$ and Qsoil has been assessed for all Qsoil layers. These results are presented in figure A.26 and show a clear pattern of $NDVI_{[t-1]}$ explaining an overwhelming majority of NDVI z-scores. Variance explained by Qsoil and variance shared by Qsoil and $NDVI_{[t-1]}$ decrease through the soil layers.
Spatial patterns of explained variance (see figure A.27) reveal that $NDVI_{[t-1]}$ and Qsoil are the most valuable as predictors NDVI across the continental region of Australia with patchy distributions.



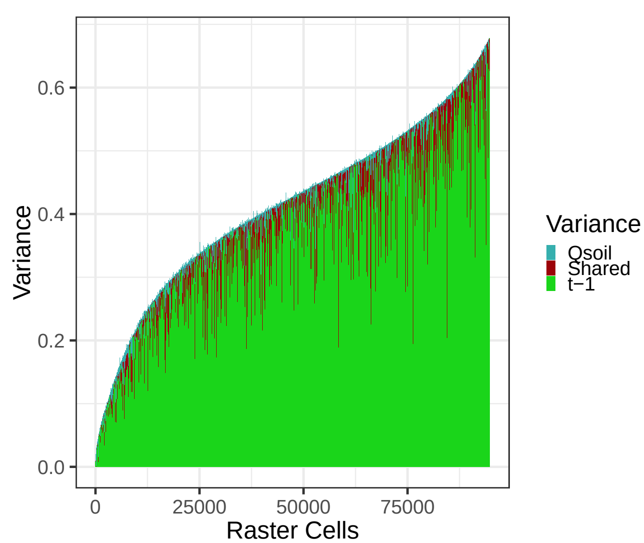
(a) Soil Layer 1



(b) Soil Layer 2

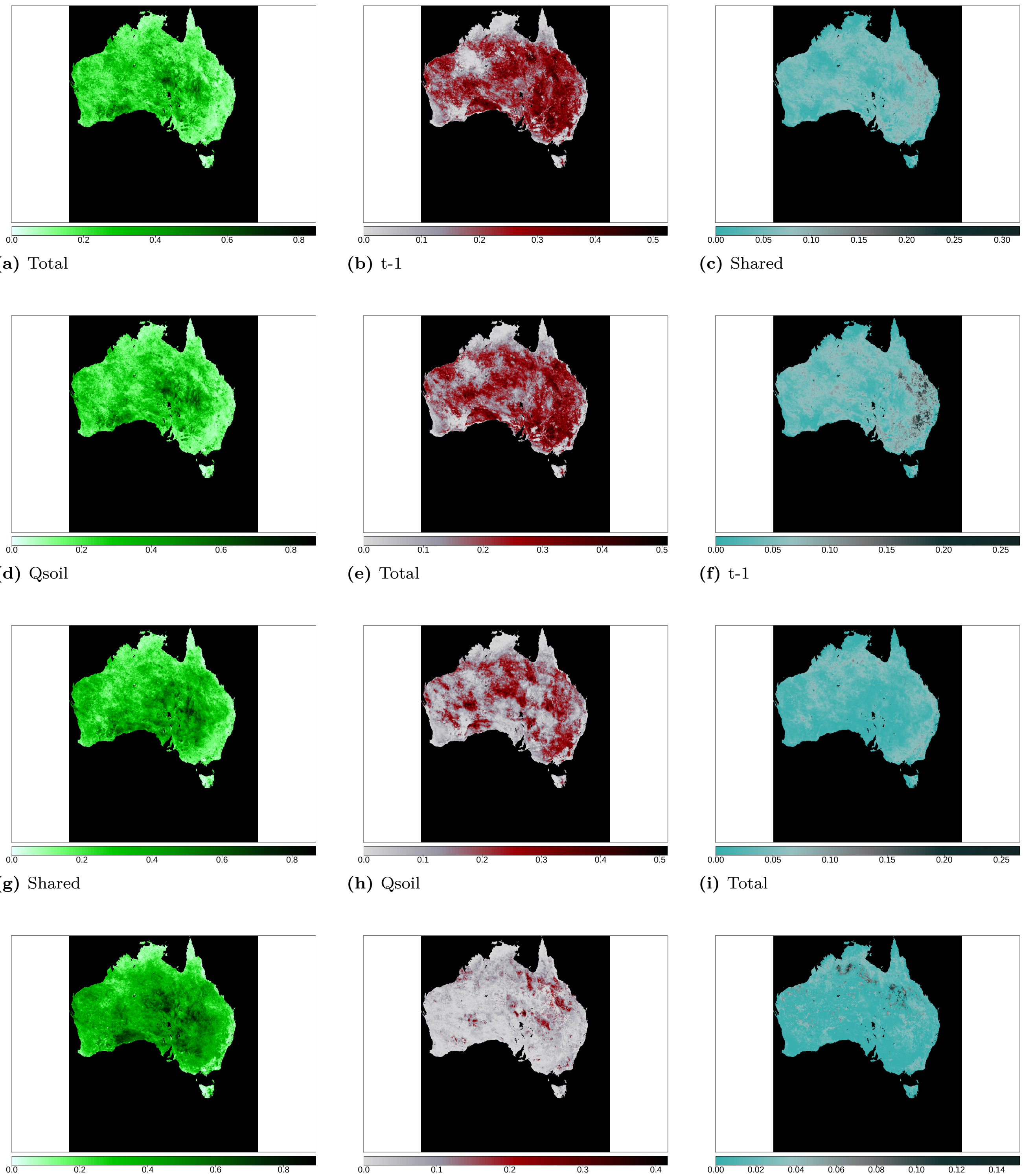


(c) Soil Layer 3



(d) Soil Layer 4

Figure A.26: Variance Partitioning (Australia; Qsoil1) - Variance of NDVI anomalies explained by (a) full models of intrinsic and extrinsic memory, (b) intrinsic memory, (c) shared variance, and (d) extrinsic memory. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.



(a) Total **(b)** t-1 **(c)** Shared

(d) Qsoil **(e)** Total **(f)** t-1

(g) Shared **(h)** Qsoil **(i)** Total

(j) t-1 **(k)** Shared **(l)** Qsoil

Figure A.27: Variance Partitioning (Australia) - Variance of NDVI anomalies explained by (a,d,g,j) full models of intrinsic and extrinsic memory, (b,e,h,k) intrinsic memory, (c,f,i,l) shared variance, and (d) extrinsic memory across soil layers Qsoil1 to Qsoil4. A representation of how these were calculated can be retrieved in figure 2.14. Figure established via Chunk 26.

A.5 R Data Visualisation Codes

A.5.1 Preamble

Chunk 13: Preamble needed for processing the following chunks within this manuscript.

```
source("S0a_Packages.R")  # loading packages
source("S0b_Directories.R")  # setting directories
source("S0c_Functions.R")  # Loading miscellaneous functions
col.NDVI <- rev(terrain.colors(100))
col.qsoil <- colorRampPalette(c("yellow", "burlywood", "beige", "turquoise", "deepskyblue"))(100)
col.tair <- colorRampPalette(c("blue", "turquoise", "yellow", "orange", "red"))(100)
SR_cols <- list(col.NDVI, col.tair, col.qsoil, col.qsoil, col.qsoil, col.qsoil)
SR_Titles <- list("NDVI", "Air Temperature", "Soil Moisture (0-7cm)", "Soil Moisture (7-28cm)",
    "Soil Moisture (28-100cm)", "Soil Moisture (100-255cm)")
setwd(Dir.KrigCov)
Elevation <- raster(list.files()[2], varname = "Elevation")
ElevationF <- raster(list.files()[1], varname = "Elevation")
```

A.5.2 Data Overview

A.5.2.1 NDVI

Chunk 14: Plotting global NDVI mean for the time frame of 1982 - 2015.

```
setwd(Dir.Gimms.Monthly)
NDVIMean_ras <- mean(brick(list.files()[1:7]), na.rm = TRUE)
plot(NDVIMean_ras, colNA = "black", main = "Mean NDVI 1982 - 2015", cex.main = 2,
    legend.width = 1.5, legend.shrink = 1, axis.args = list(cex.axis = 1.5), cex.axis = 1.5)
```

A.5.2.2 ERA5

Chunk 15: Two by two plotting of global soil moisture indices across four different soil layers all on the same scale.

```
Variable <- c("Qsoil1", "Qsoil2", "Qsoil3", "Qsoil4")
colour = col.qsoil
par(mfrow = c(2,2))
setwd(Dir.ERA)
Era_list <- list(NA, NA, NA, NA)
for(i in 1:4){
  Era5ras <- mean(brick(list.files()[grep(pattern = Variable[i], list.files())]), na.rm = TRUE)
  values(Era5ras)[which(is.na(values(Elevation)))] <- NA
  Era_list[[i]] <- Era5ras
}
Era5ras <- brick(Era_list)
Era5ras[1] <- max(Era5ras)@data@max
for(i in 1:4){
  plot(Era5ras[[i]], colNA = "black", main = paste("Mean", SR_Titles[i+2], "1980 - 2015"),
       cex.main = 2, legend.width = 1.5, legend.shrink=1,
       axis.args=list(cex.axis=1.5), cex.axis = 1.5, col = colour, legend = FALSE)
}
```

Chunk 16: Plotting mean air temperature from 1982 - 2015 globally.

```
Variable <- c("Tair")
colour = col.tair
setwd(Dir.ERA)
Era5ras <- mean(brick(list.files()[grep(pattern = Variable, list.files())]), na.rm = TRUE)
values(Era5ras)[which(is.na(values(Elevation)))] <- NA
plot(Era5ras, colNA = "black", main = paste("Mean", SR_Titles[2], "1980 - 2015"),
     cex.main = 2, legend.width = 1.5, legend.shrink=1,
     axis.args=list(cex.axis=1.5), cex.axis = 1.5, col = colour, legend = TRUE)
```

A.5.2.3 HWSD

Chunk 17: Plotting HWSD elevation data.

```
plot(ElevationF, colNA = "black", cex.main = 2, legend.width = 1.5, legend.shrink = 1,
    axis.args = list(cex.axis = 1.5), cex.axis = 1.5, legend = TRUE, col = terrain.colors(100),
    main = "DEM at GIMMs native resolution")
```

Chunk 18: Loading HWSD data and plotting HWSD slope incline data.

```
Covariates_vec <- c("Slopes1", "Slopes2", "Slopes3", "Slopes4", "Slopes5", "Slopes6", "Slopes7", "Slopes8",
                    "Slope_aspect_N", "Slope_aspect_E", "Slope_aspect_S", "Slope_aspect_W", "Slope_aspect_U",
                    "Elevation")
Cov_fine <- list() # create empty list
for(c in 1:length(Covariates_vec)){ # cycle through all covariates and load the data
      Cov_fine[[c]] <- raster(paste(Dir.KrigCov, "/Co-variates_NativeResolution.nc", sep=""),
                              varname = Covariates_vec[c])
      }
Cov_fine <- brick(Cov_fine) # make fine covariate data into one big brick

for(i in 1:13){
values(Cov_fine[[i]]) <- values(Cov_fine[[i]])/1000
}
par(mfrow = c(3,2))
for(i in 9:13){
  plot(Cov_fine[[i]], colNA = "black", cex.main = 2, legend.width = 1.5, legend.shrink=1,
       axis.args=list(cex.axis=1.5), cex.axis = 1.5, legend = TRUE, col = terrain.colors(100),
       main = Covariates_vec[[i]])
}
```

Chunk 19: Plotting HWSD slope aspect data

```
par(mfrow = c(4, 2))
for (i in 1:8) {
    plot(Cov_fine[[i]]/10, colNA = "black", cex.main = 2, legend.width = 1.5, legend.shrink = 1,
        axis.args = list(cex.axis = 1.5), cex.axis = 1.5, legend = TRUE, col = terrain.colors(100),
        main = Covariates_vec[[i]])
}
```

A.5.2.4 Study Regions

Chunk 20: Function for plotting data overviews in three-by-two plots for study regions.

```
col.NDVI <- rev(terrain.colors(100))
col.qsoil <- colorRampPalette(c("yellow", "burlywood", "beige", "turquoise", "deepskyblue"))(100)
col.tair <- colorRampPalette(c("blue", "turquoise", "yellow", "orange", "red"))(100)
SR_cols <- list(col.NDVI, col.tair, col.qsoil, col.qsoil, col.qsoil, col.qsoil)
SRData <- function(Region) {
    ### LOADING DATA ----
    setwd(Dir.Gimms.Monthly)
    SR_NDVI <- mean(brick(list.files()[grep(pattern = Region, list.files())]), na.rm = TRUE)
    setwd(Dir.ERA.Monthly)
    SR_Qsoil1 <- mean(brick(list.files()[grep(pattern = paste("Qsoil1_mean_", Region,
        sep = ""), list.files())]), na.rm = TRUE)
    SR_Qsoil2 <- mean(brick(list.files()[grep(pattern = paste("Qsoil2_mean_", Region,
        sep = ""), list.files())]), na.rm = TRUE)
    SR_Qsoil3 <- mean(brick(list.files()[grep(pattern = paste("Qsoil3_mean_", Region,
        sep = ""), list.files())]), na.rm = TRUE)
    SR_Qsoil4 <- mean(brick(list.files()[grep(pattern = paste("Qsoil4_mean_", Region,
        sep = ""), list.files())]), na.rm = TRUE)
    SR_Tair <- mean(brick(list.files()[grep(pattern = paste("Tair_mean_", Region,
        sep = ""), list.files())]), na.rm = TRUE)
    ### FIXING VALUES ----
    QsoilStack <- stack(SR_Qsoil1, SR_Qsoil2, SR_Qsoil3, SR_Qsoil4)
    SR_Qsoil1[1:2] <- c(min(QsoilStack, na.rm = TRUE)@data@min, max(QsoilStack, na.rm = TRUE)@data@max)
    SR_Qsoil2[1:2] <- c(min(QsoilStack, na.rm = TRUE)@data@min, max(QsoilStack, na.rm = TRUE)@data@max)
    SR_Qsoil3[1:2] <- c(min(QsoilStack, na.rm = TRUE)@data@min, max(QsoilStack, na.rm = TRUE)@data@max)
    SR_Qsoil4[1:2] <- c(min(QsoilStack, na.rm = TRUE)@data@min, max(QsoilStack, na.rm = TRUE)@data@max)
    ### PLOTTING ----
    Plot_Stack <- stack(SR_NDVI, SR_Tair, SR_Qsoil1, SR_Qsoil2, SR_Qsoil3, SR_Qsoil4)
    par(mfrow = c(3, 2), mai = c(1, 0, 0.5, 0))
    for (i in 1:length(SR_cols)) {
        if (i == 1) {
            plot(Plot_Stack[[i]], colNA = "black", main = paste("Mean", SR_Titles[[i]],
                "1982 - 2015"), cex.main = 3, legend.width = 3, legend.shrink = 1,
                axis.args = list(cex.axis = 2.5), cex.axis = 2, col = SR_cols[[i]],
                axes = TRUE, legend = FALSE)
        } else {
            plot(Plot_Stack[[i]], colNA = "black", main = paste("Mean", SR_Titles[[i]],
                "1981 - 2015"), cex.main = 3, legend.width = 3, legend.shrink = 1,
                axis.args = list(cex.axis = 2.5), cex.axis = 2, col = SR_cols[[i]],
                axes = TRUE, legend = FALSE)
        }
        if (i <= 2) {
            plot(Plot_Stack[[i]], legend.only = TRUE, smallplot = c(0, 0.93, 0.12,
                0.16), horizontal = TRUE, axis.args = list(cex.axis = 2.5), col = SR_cols[[i]])
        }
        if (i == 2) {
            par(mai = c(0.5, 0, 0.5, 0))
        }
    }
    return(SR_Qsoil4)
}
```

Chunk 21: Function for plotting TRY data overviews in two-by-two plots for study regions.

```
col.tair <- colorRampPalette(c("blue", "turquoise", "yellow", "orange", "red"))(100)
TRYRegions <- function(Region) {
    ### LOADING DATA ----
    setwd(Dir.Gimms.Monthly)
    Back_ras <- mean(brick(list.files()[1:7]), na.rm = TRUE)  # background raster for later plots
    values(Back_ras)[which(!is.na(values(Back_ras)))] <- 8888
    Titles <- c("Height", "Nitrogen Content")
    ## compadre
    Variable = "FastSlow"
    Dir.Comp <- paste(Dir.Compadre, Variable, sep = "/")
    Compad_ras <- list.files(Dir.Comp)[grep(list.files(Dir.Comp), pattern = Region)]
    Compad_ras <- raster(paste(Dir.Comp, Compad_ras, sep = "/"))[[1]]
    ### PLOTTING ----
    par(mfrow = c(3, 2), mai = c(1, 0, 0.5, 0))
    setwd(Dir.TRY)
    Plot_ras <- brick(list.files(Dir.TRY)[grep(pattern = Region, list.files(Dir.TRY))][1])  # Raw file
    Plot1_ras <- brick(list.files(Dir.TRY)[grep(pattern = Region, list.files(Dir.TRY))][2])  # Distrib file
    Back_ras1 <- crop(Back_ras, extent(Plot_ras))
    Back_ras1 <- stack(Back_ras1, Back_ras1)
    names(Back_ras1) <- c("Height", "Nitrogen")
    for (i in 1:2) {
        plot(Back_ras1[[i]], colNA = "black", main = Titles[i], cex.main = 3, legend = FALSE,
            col = "grey")
        plot(Plot_ras[[i]], add = TRUE, legend = FALSE, col = col.tair, axes = TRUE)
        plot(Plot_ras[[i]], legend.only = TRUE, smallplot = c(0, 0.95, 0.075, 0.115),
            horizontal = TRUE, axis.args = list(cex.axis = 2.5), col = col.tair)
    }
    for (i in 1:2) {
        plot(Back_ras1[[i]], colNA = "black", main = Titles[i], cex.main = 3, legend = FALSE,
            col = "grey")
        plot(Plot1_ras[[i]], add = TRUE, legend = FALSE, col = col.tair, axes = TRUE)
        plot(Plot1_ras[[i]], legend.only = TRUE, smallplot = c(0, 0.95, 0.075, 0.115),
            horizontal = TRUE, axis.args = list(cex.axis = 2.5), col = col.tair)
    }
    plot(Back_ras1[[1]], colNA = "black", main = "COMPADRE Stations", cex.main = 3,
        legend = FALSE, col = "grey")
    plot(Compad_ras, add = TRUE, legend = FALSE, col = "red", axes = TRUE)
}
```

A.5.3 Vegetation Memory Models

Chunk 22: Plotting NDVI data treatment in two-by-two plots.

```
## load data
IbRas <- grep(list.files(Dir.Gimms.Monthly), pattern = "Iberian Region")
IBNDVI <- brick(paste(Dir.Gimms.Monthly, list.files(Dir.Gimms.Monthly)[IbRas],
    sep = "/"))
## extract data
NDVI_vecraw <- as.vector(IBNDVI[35363])
NDVI_vecdet <- detrend(NDVI_vecraw, tt = "linear")  # linear detrending
NDVI_df <- data.frame(Month = rep(1:12, length(NDVI_vecraw)/12),
    NDVI_raw = NDVI_vecraw, NDVI_de = NDVI_vecdet)  # create NDVI data frame
## calculate anomalies (Z-scores) and monthly means
NDVI_df <- transform(NDVI_df, NDVI_Anomalies = ave(NDVI_de,
    Month, FUN = scale), NDVI_Threshold = ave(NDVI_raw,
    Month, FUN = function(t) mean(t, na.rm = TRUE)))
## plotting data frame
NDVIplot_df <- data.frame(NDVI = c(NDVI_vecraw, NDVI_vecdet,
    NDVI_df$NDVI_Anomalies, NDVI_df$NDVI_Threshold[1:12]),
    Ident = c(rep("Raw Data", length(NDVI_vecraw)),
        rep("Detrended", length(NDVI_vecraw)), rep("Z-Scores",
            length(NDVI_vecraw)), rep("Monthly Means",
            12)), Months = c(rep(1:length(NDVI_vecraw),
        3), 1:12))
## raw and detrended
ggplot(NDVIplot_df[1:(length(NDVI_vecraw) * 2), ],
    aes(x = Months, y = NDVI, col = Ident)) + geom_line(size = 2) +
    theme_bw(base_size = 35) + theme(legend.position = "none") +
    scale_color_manual(values = c("orange", "forestgreen"))
## monthly means
ggplot(NDVIplot_df[(length(NDVI_vecraw) * 3 + 1):(dim(NDVIplot_df)[1]),
    ], aes(x = as.factor(Months), y = NDVI, col = Ident)) +
    geom_bar(stat = "identity", fill = "forestgreen") +
    theme_bw(base_size = 35) + theme(legend.position = "none") +
    scale_color_manual(values = c("forestgreen")) +
    xlab("Months") + geom_hline(aes(yintercept = 0.1),
    size = 1.5)
## z-scores
ggplot(NDVIplot_df[(length(NDVI_vecraw) * 2 + 1):(length(NDVI_vecraw) *
    3), ], aes(x = Months, y = NDVI, col = Ident)) +
    geom_line(size = 2) + theme_bw(base_size = 35) +
    theme(legend.position = "none") + scale_color_manual(values = c("purple"))
## legend
leg <- ggplot(NDVIplot_df[1:(length(NDVI_vecraw) *
    3), ], aes(x = Months, y = NDVI, col = Ident)) +
    geom_line(size = 2) + scale_color_manual(values = c("orange",
    "forestgreen", "purple")) + theme_bw(base_size = 65) +
    guides(colour = guide_legend(override.aes = list(size = 20))) +
    labs(col = "GIMMs NDVI 3g data")
legend <- cowplot::get_legend(leg)
grid.newpage()
grid.draw(legend)
```

Chunk 23: Plotting Qsoil 1 data treatment in two-by-two plots.

```
setwd(Dir.ERA.Monthly)
Qsoil <- brick(list.files()[grep(list.files(), pattern = "Iberian Region_11981_122015.nc")[1]])
load(paste(Dir.Data, "ModData_df.RData", sep = "/"))  # model data for pixel 35363
Clim <- as.vector(Qsoil[35363])[1:dim(ModData_df)[1]]
Clim_vec <- detrend(Clim, tt = "linear")  # linear detrending
## raw and detrended
plot1_df <- with(ModData_df, data.frame(Months = rep(1:dim(ModData_df)[1],
    2), Qsoil = c(Clim, Clim_vec), Ident = rep(c("Raw Data",
    "Detrened"), each = dim(ModData_df)[1])))
ggplot(plot1_df, aes(x = Months, y = Qsoil, col = Ident)) +
    geom_line(size = 2) + theme_bw(base_size = 35) +
    theme(legend.position = "none") + scale_color_manual(values = c("orange",
    "navyblue"))
## raw and detrended
plot2_df <- with(ModData_df, data.frame(Months = rep(1:dim(ModData_df)[1],
    1), Qsoil = c(Clim_raw), Ident = rep(c("Z-Scores"),
    each = dim(ModData_df)[1])))
ggplot(plot2_df, aes(x = Months, y = Qsoil, col = Ident)) +
    geom_line(size = 2) + theme_bw(base_size = 35) +
    theme(legend.position = "none") + scale_color_manual(values = c("brown"))
## Cummulative soil moisture lags
plot3_df <- with(ModData_df, data.frame(Months = rep(1:dim(ModData_df)[1],
    13), Qsoil = c(ClimCum_0, ClimCum_1, ClimCum_2,
    ClimCum_3, ClimCum_4, ClimCum_5, ClimCum_6, ClimCum_7,
    ClimCum_8, ClimCum_9, ClimCum_10, ClimCum_11, ClimCum_12),
    Ident = as.factor(rep(c("Z-Scores", "Lag 01", "Lag 02",
        "Lag 03", "Lag 04", "Lag 05", "Lag 06", "Lag 07",
        "Lag 08", "Lag 09", "Lag 10", "Lag 11", "Lag 12"),
        each = dim(ModData_df)[1]))))
ggplot(plot3_df, aes(x = Months, y = Qsoil, col = Ident)) +
    geom_line(size = 1) + theme_bw(base_size = 35) +
    theme(legend.position = "none") + scale_color_manual(values = c(rainbow(12),
    "brown"))
## legend
Qsoilplot_df <- rbind(plot1_df, plot3_df)
leg <- ggplot(Qsoilplot_df, aes(x = Months, y = Qsoil,
    col = Ident)) + geom_line(size = 2) + scale_color_manual(values = c("orange",
    "navyblue", rainbow(12), "brown")) + theme_bw(base_size = 65) +
    guides(colour = guide_legend(ncol = 3, override.aes = list(size = 20))) +
    labs(col = "ERA5 Qsoil1 data")
legend <- cowplot::get_legend(leg)
grid.newpage()
grid.draw(legend)
```

A.5.4 Results

A.5.4.1 Memory Models

Chunk 24: Plotting vegetation memory model outputs.

```
ModelRes <- function(Region, SoilLayer) {
    col.signeg <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "targaryen2")
    col.sigpos <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "tyrell")
    col.nonsig <- colorRampPalette(c("grey"))(1)
    col.lags <- got(n = 12, alpha = 1, begin = 0, end = 1, direction = 1, option = "daenerys")
    smallplotxpos <- c(0.5, 0.93, 0.03, 0.065)  # where to put colour scales
    smallplotxneg <- c(0.05, 0.5, 0.03, 0.065)  # where to put colour scales
    Files <- list.files(Dir.Memory)[grep(list.files(Dir.Memory), pattern = Region)]
    Files <- Files[grep(Files, pattern = ".nc")]
    plot_ras <- brick(paste(Dir.Memory, Files[SoilLayer], sep = "/"))
    plot(plot_ras[[1]], col = col.lags, colNA = "black", legend = FALSE, axes = FALSE)
    plot(plot_ras[[1]], legend.only = TRUE, col = col.lags, smallplot = c(0.05, 0.93,
        0.03, 0.065), horizontal = TRUE, axis.args = list(cex.axis = 2.5))
    if (SoilLayer == 1) {
        RasLay <- 4:6
    } else {
        RasLay <- c(5)
    }
    for (Plot in RasLay) {
        Neg_ras <- plot_ras[[Plot]]
        Neg_ras[which(values(Neg_ras) >= 0)] <- NA
        Pos_ras <- plot_ras[[Plot]]
        Pos_ras[which(values(Pos_ras) < 0)] <- NA
        plot(Neg_ras, col = col.signeg, colNA = "black", legend = FALSE, axes = FALSE)
        if (Plot == 4 & Region == "Iberian Region") {
            plot(Pos_ras, col = col.sigpos, colNA = "black", legend = FALSE, axes = FALSE,
                add = TRUE)
        } else {
            plot(Pos_ras, col = col.sigpos, legend = FALSE, axes = FALSE, add = TRUE)
        }
        plot(Neg_ras, legend.only = TRUE, col = col.signeg, colNA = "black", smallplot = smallplotxneg,
            horizontal = TRUE, axis.args = list(cex.axis = 2.5))
        if (Plot == 4 & Region == "Iberian Region") {
            plot(Pos_ras, legend.only = TRUE, col = col.sigpos, smallplot = c(0.05,
                0.93, 0.03, 0.065), horizontal = TRUE, axis.args = list(cex.axis = 2.5))
        } else {
            plot(Pos_ras, legend.only = TRUE, col = col.sigpos, smallplot = smallplotxpos,
                horizontal = TRUE, axis.args = list(cex.axis = 2.5))
        }
    }
}
```

Chunk 25: Loading U-Test results of model coefficients and displaying as a table.

```
Ures <- function(Region, SoilLayer) {
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".xlsx")]
    Utab <- read.xlsx(file = paste(Dir.Reg, Files, sep = "/"), sheetName = paste("Model",
        SoilLayer))
    Medians <- read.xlsx(file = paste(Dir.Reg, Files, sep = "/"), sheetName = "Variable medians")[SoilLayer,
        ]
    Medians <- as.numeric(Medians[1, c(2:4)])
    rownames(Utab) <- Utab[, 1]
    Utab[, 1] <- round(Medians, 4)
    colnames(Utab) <- c("Medians", "NDVI [t-1]", paste("Qsoil", SoilLayer, sep = ""),
        "Tair")
    Utab <- round(Utab, 4)
    Short <- paste("Mann-Whitney U-Test (", Region, ", Qsoil", SoilLayer, " Model)",
        sep = "")
    Long <- paste("U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand
              cells. This statistic is the sum of of all pixels in which the rowname variable is bigger than the
              column name variable. $p$-values belonging to these $U$-values are represented in the lower-lefthand
              block of cells. Established via \\ref{rUresChunk}.")
    kab <- kable(Utab, booktabs = TRUE, caption = paste("\\textbf{", Short, " -} ",
        Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
    kab <- add_header_above(kable_input = kab, header = c(` ` = 2, `U-Test` = 3),
        bold = TRUE, align = "c")
    kab <- kable_styling(kab, latex_options = c("hold_position"))
    print(kab)
}
```

A.5.4.2 Variance Partitioning

Chunk 26: Plotting variance partitioning results.

```
VarParRes <- function(Region, SoilLayer, Plot = 1, Legend = TRUE, lim = FALSE) {
    col.varpar1 <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "wildfire")
    col.varpar2 <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "targaryen")
    col.varpar3 <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "jon_snow")
    col.list <- list(col.varpar1, col.varpar2, col.varpar3)
    Files <- list.files(Dir.Memory)[grep(list.files(Dir.Memory), pattern = Region)]
    Files <- Files[grep(Files, pattern = ".nc")]
    Alter_ras <- brick(paste(Dir.Memory, Files[SoilLayer], sep = "/"))[[7:10]]
    Alter_ras[2] <- 0
    values(Alter_ras)[which(values(Alter_ras) < 0)] <- 0
    cells <- order(values(Alter_ras[[1]]))
    `%nin%` = Negate(`%in%`)  # create a 'not in' statement
    cells <- cells[which(cells %nin% which(values(Alter_ras[[1]]) > quantile(values(Alter_ras[[1]]),
        0.95, na.rm = TRUE)))]
    plot_df <- data.frame(Data = NA, Cell = NA, Variance = NA)
    Idents <- c("Total", "t-1", "Shared", "Qsoil")
    for (i in 1:4) {
        if (i > 1) {
            plot_df1 <- data.frame(Data = values(Alter_ras[[i]])[cells], Cell = 1:length(cells),
                Variance = rep(Idents[i], length(cells)))
            plot_df <- rbind(plot_df, plot_df1)
        }
    }
    if (lim == TRUE) {
        Lims <- c(0, 0.7)
    } else {
        Lims <- c(0, max(plot_df$Data, na.rm = TRUE))
    }
    plot_df <- na.omit(plot_df)
    if (Legend == TRUE) {
        p <- ggplot(data = plot_df, aes(y = Data, x = Cell, fill = Variance)) + geom_bar(stat = "identity") +
            theme_bw(base_size = 45) + xlab("Raster Cells") + ylab("Variance") +
            scale_fill_manual(values = c(col.list[[3]][1], col.list[[2]][50], col.list[[1]][30])) +
            ylim(Lims)
    } else {
        p <- ggplot(data = plot_df, aes(y = Data, x = Cell, fill = Variance)) + geom_bar(stat = "identity") +
            theme_bw(base_size = 45) + xlab("Raster Cells") + ylab("Variance") +
            scale_fill_manual(values = c(col.list[[3]][1], col.list[[2]][50], col.list[[1]][30])) +
            theme(legend.position = "none") + ylim(Lims)
    }
    if (Plot == 1) {
        return(p)
    } else {
        for (i in 2:4) {
            col.varpar <- col.list[[i - 1]]
            plot(Alter_ras[[i]], col = col.varpar, colNA = "black", legend = FALSE,
                axes = FALSE)
            plot(Alter_ras[[i]], legend.only = TRUE, col = col.varpar, smallplot = c(0.05,
                0.93, 0.03, 0.065), horizontal = TRUE, axis.args = list(cex.axis = 2.5))
        }
    }
}
```

A.5.4.3 Soil Layer Comparison

Chunk 27: Comparing soil layer response coefficients.

```
SoilLayersRes <- function(Region) {
    col.signeg <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "targaryen2")
    col.sigpos <- got(n = 100, alpha = 1, begin = 0, end = 1, direction = -1, option = "tyrell")
    col.nonsig <- colorRampPalette(c("grey"))(1)
    smallplotxpos <- c(0.5, 0.93, 0.03, 0.065)  # where to put colour scales
    smallplotxneg <- c(0.05, 0.5, 0.03, 0.065)  # where to put colour scales
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    SoilLayer <- 1
    Alter_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[3]]
    for (SoilLayer in 2:4) {
        Add_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[3]]
        Alter_ras <- stack(Alter_ras, Add_ras)
    }
    for (Plot in 1:4) {
        Neg_ras <- Alter_ras[[Plot]]
        Neg_ras[which(values(Neg_ras) >= 0)] <- NA
        Pos_ras <- Alter_ras[[Plot]]
        Pos_ras[which(values(Pos_ras) < 0)] <- NA
        plot(Neg_ras, col = col.signeg, colNA = "black", legend = FALSE, axes = FALSE)
        plot(Pos_ras, col = col.sigpos, legend = FALSE, axes = FALSE, add = TRUE)
        plot(Neg_ras, legend.only = TRUE, col = col.signeg, colNA = "black", smallplot = smallplotxneg,
            horizontal = TRUE, axis.args = list(cex.axis = 2.5))
        plot(Pos_ras, legend.only = TRUE, col = col.sigpos, smallplot = smallplotxpos,
            horizontal = TRUE, axis.args = list(cex.axis = 2.5))
    }
}
```

Chunk 28: Loading U-Test results of soil layer comparison and displaying as a table.

```
UQsoilres <- function(Region) {
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".xlsx")]
    Utab <- read.xlsx(file = paste(Dir.Reg, Files, sep = "/"), sheetName = "Qsoil")
    Medians <- read.xlsx(file = paste(Dir.Reg, Files, sep = "/"), sheetName = "Variable medians")[,
        3]
    rownames(Utab) <- paste("Qsoil", c(1:4), sep = "")
    Utab[, 1] <- c(round(Medians, 4))
    colnames(Utab) <- c("Medians", paste("Qsoil", c(1:4), sep = ""))
    Utab <- round(Utab, 4)
    Short <- paste("Mann-Whitney U-Test (", Region, ", Qsoil Layers)", sep = "")
    Long <- paste("U-Test statistics of vegetation memory coefficient values are contained within the upper-righthand
              cells. $p$-values belonging to these $U$-values are represented in the lower-lefthand block of cells.
              Established via \\ref{rUQsoilRes}.")
    kab <- kable(Utab, booktabs = TRUE, caption = paste("\\textbf{", Short, " -} ",
        Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
    kab <- kable_styling(kab, latex_options = c("hold_position"))
    kab <- add_header_above(kable_input = kab, header = c(` ` = 2, `U-Test` = 4),
        bold = TRUE, align = "c")
    print(kab)
}
```

A.5.4.4 Vegetation Memory Sensitivity

Chunk 29: Vegetation sensitivity plots.

```
PatcausRes <- function(SoilLayer = 1, X, Y) {
    Regions <- c("Iberian Region", "Caatinga", "Australia")
    Idents <- c("NDVI", "ENDVI", "Qsoil", "EQsoil", "Tair", "ETair", "Memory")
    Labs <- c("NDVI [t-1] Coefficients", "Mean NDVI ", paste("Qsoil", SoilLayer,
        " Coefficients", sep = ""), paste("Mean  Qsoil", SoilLayer, sep = ""),
        "Tair Coefficients", "Mean Tair", paste("Qsoil", SoilLayer, " Memory Lags",
            sep = ""))
    # data
    plot_df <- data.frame(X = NA, Y = NA, Region = NA)
    for (i in 1:length(Regions)) {
        Region <- Regions[i]
        Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
        Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
        Alter_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))
        Files <- list.files(Dir.ERA.Monthly)[grep(list.files(Dir.ERA.Monthly),
            pattern = Region)]
        Files <- Files[grep(Files, pattern = ".nc")]
        EQsoil <- Files[grep(Files, pattern = "Qsoil")]
        EQsoil <- brick(paste(Dir.ERA.Monthly, EQsoil[SoilLayer], sep = "/"))
        EQsoil <- values(mean(EQsoil))
        ETair <- Files[grep(Files, pattern = "Tair")]
        ETair <- brick(paste(Dir.ERA.Monthly, ETair, sep = "/"))
        ETair <- values(mean(ETair))
        ENDVI <- list.files(Dir.Gimms.Monthly)[grep(list.files(Dir.Gimms.Monthly),
            pattern = Region)]
        ENDVI <- brick(paste(Dir.Gimms.Monthly, ENDVI, sep = "/"))
        ENDVI <- values(mean(ENDVI))
        values <- list(values(Alter_ras[[2]]), ENDVI, values(Alter_ras[[3]]),
            EQsoil, values(Alter_ras[[4]]), ETair, values(Alter_ras[[1]]))
        plot_df1 <- data.frame(X = values[[which(Idents == X)]], Y = values[[which(Idents ==
            Y)]], Region = rep(Region, length(values[[which(Idents == X)]])))
        plot_df1 <- na.omit(plot_df1)
        plot_df <- rbind(plot_df, plot_df1)
        QsoilTitle <- c("(0-7cm)", "(7-28cm)", "(28-100cm)", "(100-255cm)")
    }
    plot_df <- na.omit(plot_df)
    # plotting
    p <- ggplot(plot_df, aes(x = X, y = Y, col = Region)) + geom_point(shape = ".",
        alpha = 0.4, size = 0.5) + theme_bw(base_size = 35) + ylab(Labs[which(Idents ==
        Y)]) + xlab(Labs[which(Idents == X)]) + stat_smooth(method = "lm",
        level = 0.95) + geom_hline(yintercept = 0, linetype = "dotted") +
        scale_color_manual(values = got(n = length(Regions), alpha = 1,
            direction = -1, option = "margaery"))
    if (X == "Memory") {
        p <- ggplot(plot_df, aes(x = as.factor(X), y = Y, col = Region)) +
            geom_boxplot() + theme_bw(base_size = 35) + ylab(Labs[which(Idents ==
            Y)]) + xlab(Labs[which(Idents == X)]) + geom_hline(yintercept = 0,
            linetype = "dotted") + scale_color_manual(values = got(n = length(Regions),
            alpha = 1, direction = -1, option = "margaery")) + stat_smooth(method = "lm",
            level = 0.95, aes(group = plot_df$Region))
    }
    if (Y == "Memory") {
        p <- ggplot(plot_df, aes(x = as.factor(Y), y = X, col = Region)) +
            geom_boxplot() + theme_bw(base_size = 35) + xlab(Labs[which(Idents ==
            Y)]) + ylab(Labs[which(Idents == X)]) + coord_flip() + geom_hline(yintercept = 0,
```

```
            linetype = "dotted") + scale_color_manual(values = got(n = length(Regions),
            alpha = 1, direction = -1, option = "margaery")) + stat_smooth(method = "lm",
            level = 0.95, aes(group = plot_df$Region))
    }
    print(p)
    # Regressions
    Output <- as.list(rep(NA, 4 * length(Regions)))
    pos <- 1
    for (i in 1:length(Regions)) {
        Output[[pos]] <- summary(lm(Y ~ X, data = plot_df[which(plot_df$Region ==
            Regions[i]), ]))[["coefficients"]][1, 1]
        pos <- pos + 1
        Output[[pos]] <- summary(lm(Y ~ X, data = plot_df[which(plot_df$Region ==
            Regions[i]), ]))[["coefficients"]][1, 4]
        pos <- pos + 1
        Output[[pos]] <- summary(lm(Y ~ X, data = plot_df[which(plot_df$Region ==
            Regions[i]), ]))[["coefficients"]][2, 1]
        pos <- pos + 1
        Output[[pos]] <- summary(lm(Y ~ X, data = plot_df[which(plot_df$Region ==
            Regions[i]), ]))[["coefficients"]][2, 4]
        pos <- pos + 1
    }
    return(Output)
}
```

Chunk 30: Linear regression coefficients of vegetation memory sensitivity.

```
PatCauseTab <- function(ModelCoeffs, SoilLayer = 1, X, Y) {
    # using Outputs of PatcauseRes for ModelCoeffs argument
    tabout <- data.frame(Column = ModelCoeffs[1:2])
    starts <- seq(1, length(ModelCoeffs), by = 2)
    for (i in 2:(length(ModelCoeffs)/2)) {
        tabout <- cbind(tabout, ModelCoeffs[starts[i]:(starts[i] + 1)])
    }
    Idents <- c("NDVI", "ENDVI", "Qsoil", "EQsoil", "Tair", "ETair", "Memory")
    Labs <- c("$NDVI_{[t-1]}$ Coefficients", "Mean NDVI ", paste("Qsoil", SoilLayer,
        " Coefficients", sep = ""), paste("Mean  Qsoil", SoilLayer, sep = ""), "Tair Coefficients",
        "Mean Tair", paste("Qsoil", SoilLayer, " Memory Lags", sep = ""))
    X <- Labs[which(Idents == X)]
    Y <- Labs[which(Idents == Y)]
    colnames(tabout) <- rep(c("$I$", "$S$"), 3)
    rownames(tabout) <- c("Value", "$p_{Value}$")
    Short <- paste("Vegetation Memory Sensitivity (", X, " and ", Y, ")", sep = "")
    Long <- paste("Coefficients of linear regressions of vegetation sensitivity. The intercept is marked as $I$,
              the slope of the regression as $S$. Established via \\ref{rPatcaustab}.",
        sep = "")
    kab <- kable(tabout, booktabs = TRUE, caption = paste("\\textbf{", Short, " -} ",
        Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
    kab <- add_header_above(kable_input = kab, header = c(` ` = 1, `Iberian Region` = 2,
        Caatinga = 2, Australia = 2), bold = TRUE, align = "c")
    kab <- kable_styling(kab, latex_options = c("hold_position"))
    print(kab)
}
```

A.5.5 Plant Function

A.5.5.1 Life History Traits

Chunk 31: Plotting COMPADRE data against vegetation response coefficients.

```
Comres <- function(Variable, Region, Legend = FALSE) {
    col.tair <- got(n = 1, alpha = 1, begin = 0, end = 1, direction = -1, option = "tully")
    col.ndvi <- got(n = 1, alpha = 1, begin = 0, end = 1, direction = 1, option = "tyrell")
    col.qsoil <- got(n = 1, alpha = 1, begin = 0, end = 1, direction = 1, option = "white_walkers")
    col.mem <- got(n = 1, alpha = 1, begin = 0, end = 1, direction = -1, option = "greyjoy")
    VarHold <- Variable
    if (Variable == "FSC-1" | Variable == "FSC-2") {
        Variable <- "FastSlow"
    }
    ## Memory Data Memory
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    SoilLayer <- 1
    Memory_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[1]]
    # t-1
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    SoilLayer <- 1
    NDVI_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[2]]
    # Qsoil
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    SoilLayer <- 1
    Qsoil_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[3]]
    # Tair
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    SoilLayer <- 1
    Tair_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[4]]
    ## COMPADRE data
    Dir.Comp <- paste(Dir.Compadre, Variable, sep = "/")
    Compad_ras <- list.files(Dir.Comp)[grep(list.files(Dir.Comp), pattern = Region)]
    if (VarHold != "FSC-2") {
        Compad_ras <- raster(paste(Dir.Comp, Compad_ras, sep = "/"))[[1]]
    } else {
        Compad_ras <- brick(paste(Dir.Comp, Compad_ras, sep = "/"))[[2]]
    }
    values(Compad_ras)[which(values(Compad_ras) > quantile(values(Compad_ras), 0.66,
        na.rm = TRUE))] <- quantile(values(Compad_ras), 0.66, na.rm = TRUE)
    if (VarHold == "FSC-1" | VarHold == "FSC-2") {
        values(Compad_ras)[which(values(Compad_ras) < quantile(values(Compad_ras),
            0.05, na.rm = TRUE))] <- quantile(values(Compad_ras), 0.05, na.rm = TRUE)
    }
    plot_df <- data.frame(Data = c(values(Memory_ras), values(NDVI_ras), values(Qsoil_ras),
        values(Tair_ras)), Identifiers = rep(c("Lag", "t-1", "Qsoil1", "Tair"), each = length(values(Memory_ras))),
        Compadre = rep(values(Compad_ras), 4))
    Variable <- VarHold
    plot_df <- na.omit(plot_df)
    Output <- as.list(rep(NA, 16))
    Idents <- c("t-1", "Tair", "Qsoil1", "Lag")
    for (i in 0:(length(Idents) - 1)) {
        Output[[(4 * i + 1)]] <- summary(lm(Data ~ Compadre, data = plot_df[which(plot_df$Identifiers ==
            Idents[(i + 1)]), ]))[["coefficients"]][1, 1]
```

```
        Output[[(4 * i + 2)]] <- summary(lm(Data ~ Compadre, data = plot_df[which(plot_df$Identifiers ==
            Idents[(i + 1)]), ]))[["coefficients"]][1, 4]
        Output[[(4 * i + 3)]] <- summary(lm(Data ~ Compadre, data = plot_df[which(plot_df$Identifiers ==
            Idents[(i + 1)]), ]))[["coefficients"]][2, 1]
        Output[[(4 * i + 4)]] <- summary(lm(Data ~ Compadre, data = plot_df[which(plot_df$Identifiers ==
            Idents[(i + 1)]), ]))[["coefficients"]][2, 4]
    }
    Linesa <- rep(1, 4)
    Linesa[which(unlist(Output)[c(4, 8, 12, 16)] > 0.05)] <- 2
    Lines <- c(Linesa[4], Linesa[1], Linesa[3], Linesa[2])
    Lines <- rep(Lines, each = length(which(plot_df$Identifiers == "Lag")))
    Lines <- as.factor(Lines)
    if (unique(Lines) == 2) {
        if (Legend == TRUE) {
            plot <- ggplot(data = plot_df, aes(x = Compadre, y = Data, col = Identifiers,
                linetype = Lines)) + geom_point(alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) +
                xlab(paste("COMPADRE", Variable)) + ylab("Vegetation Response Coefficients") +
                geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm",
                level = 0.66, linetype = 2) + scale_color_manual(values = c(col.mem,
                col.qsoil, col.ndvi, col.tair)) + labs(linetype = "p < .05", colour = "Identifiers") +
                scale_linetype_manual(values = c(1, 2), labels = c("Yes", "No")) +
                guides(colour = guide_legend(ncol = 2, override.aes = list(size = 5)),
                  linetype = guide_legend(override.aes = list(size = 3)))
        } else {
            plot <- ggplot(data = plot_df, aes(x = Compadre, y = Data, col = Identifiers,
                linetype = Lines)) + geom_point(alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) +
                xlab(paste("COMPADRE", Variable)) + ylab("Vegetation Response Coefficients") +
                geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm",
                level = 0.66, linetype = 2) + scale_color_manual(values = c(col.mem,
                col.qsoil, col.ndvi, col.tair)) + theme(legend.position = "none")
        }
    } else {
        if (Legend == TRUE) {
            plot <- ggplot(data = plot_df, aes(x = Compadre, y = Data, col = Identifiers,
                linetype = Lines)) + geom_point(alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) +
                xlab(paste("COMPADRE", Variable)) + ylab("Vegetation Response Coefficients") +
                geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm",
                level = 0.66) + scale_color_manual(values = c(col.mem, col.qsoil,
                col.ndvi, col.tair)) + labs(linetype = "p < .05", colour = "Identifiers") +
                scale_linetype_manual(values = c(1, 2), labels = c("Yes", "No")) +
                guides(colour = guide_legend(ncol = 2, override.aes = list(size = 5)),
                  linetype = guide_legend(override.aes = list(size = 3)))
        } else {
            plot <- ggplot(data = plot_df, aes(x = Compadre, y = Data, col = Identifiers,
                linetype = Lines)) + geom_point(alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) +
                xlab(paste("COMPADRE", Variable)) + ylab("Vegetation Response Coefficients") +
                geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm",
                level = 0.66) + scale_color_manual(values = c(col.mem, col.qsoil,
                col.ndvi, col.tair)) + theme(legend.position = "none")
        }
    }
    return(list(plot, Output, dim(plot_df)[1]/4))
}
```

Chunk 32: Linear regression coefficients of COMPADRE data and vegetation memory coeffcients.

```
CompadTab <- function(Region, ModelCoeffs, Variable) {
    tabout <- data.frame(Column = ModelCoeffs[1:2])
    starts <- seq(1, length(ModelCoeffs), by = 2)
    for (i in 2:(length(ModelCoeffs)/2)) {
        tabout <- cbind(tabout, ModelCoeffs[starts[i]:(starts[i] + 1)])
    }
    colnames(tabout) <- rep(c("$I$", "$S$"), 4)
    rownames(tabout) <- c("Value", "$p_{Value}$")
    tabout <- round(tabout, 5)
    Short <- paste("COMPADRE ", Variable, " and Vegetation Memory (", Region, ")",
        sep = "")
    Long <- paste("Linear regression coefficients of COMPADRE ", Variable, " data and vegetation
              response coefficients. Established via \\ref{rCOMPADTab}.",
        sep = "")
    kab <- kable(tabout, booktabs = TRUE, caption = paste("\\textbf{", Short, " -} ",
        Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
    kab <- add_header_above(kable_input = kab, header = c(` ` = 1, `NDVI [t-1]` = 2,
        Tair = 2, Qsoil1 = 2, Lag1 = 2), bold = TRUE, align = "c")
    kab <- kable_styling(kab, latex_options = c("hold_position"))
    print(kab)
}
```

A.5.5.2 Plant Functional Traits

Chunk 33: Plotting TRY data against vegetation response coefficients.

```
Tryres <- function(Region, SoilLayer, Variable) {
    Dir.Reg <- paste(Dir.Memory, "/", Region, "-1981_2015", sep = "")
    Variable_vec <- c("t-1", "Qsoil", "Tair")
    LocVar <- which(Variable_vec == Variable) + 1
    TryFiles <- list.files(Dir.TRY)[grep(list.files(Dir.TRY), pattern = Region)]
    TryRaw <- brick(paste(Dir.TRY, TryFiles[1], sep = "/"))
    values(TryRaw)[which(values(TryRaw) > quantile(values(TryRaw), 0.99, na.rm = TRUE))] <- NA
    TryDis <- brick(paste(Dir.TRY, TryFiles[2], sep = "/"))
    Files <- list.files(Dir.Reg)[grep(list.files(Dir.Reg), pattern = ".nc")]
    Alter_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[LocVar]]
    Memory_ras <- brick(paste(Dir.Reg, Files[SoilLayer], sep = "/"))[[1]]
    data_df <- data.frame(Height = c(values(TryRaw[[1]]), values(TryDis[[1]])), Nitrogen = c(values(TryRaw[[2]]),
        values(TryDis[[2]])), Qsoil = c(values(Alter_ras), values(Alter_ras)), Memory = c(values(Memory_ras),
        values(Memory_ras)), TRY = c(rep("Raw", length(values(Alter_ras))), rep("Mapped",
        length(values(Alter_ras)))))

    VarTitle <- c("NDVI [t-1] Coefficients", "Qsoil", "Air Temperature Coefficients")[(LocVar -
        1)]
    if (LocVar == 3) {
        VarTitle <- c("Soil Moisture (0-7cm) Coefficients", "Soil Moisture (7-28cm) Coefficients",
            "Soil Moisture (28-100cm) Coefficients", "Soil Moisture (100-255cm) Coefficients")[SoilLayer]
    }
    if (Variable != "Qsoil") {
        Output <- as.list(rep(NA, 16))
    } else {
        Output <- as.list(rep(NA, 32))
    }
    p1 <- ggplot(data_df, aes(x = Height, y = Qsoil, col = TRY)) + geom_point(shape = ".",
        alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) + xlab("Vegetative Height") +
        ylab(VarTitle) + geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm") +
        scale_color_manual(values = c("lightgreen", "black"))
    print(p1)
    if (length(which(!is.na(data_df$Height[which(data_df$TRY == "Raw")]))) < 2) {
        Output[[1]] <- NA
        Output[[2]] <- NA
        Output[[3]] <- NA
        Output[[4]] <- NA
    } else {
        Output[[1]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][1, 1]
        Output[[2]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][1, 4]
        Output[[3]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][2, 1]
        Output[[4]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][2, 4]
    }
    Output[[5]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 1]
    Output[[6]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 4]
    Output[[7]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][2, 1]
    Output[[8]] <- summary(lm(Qsoil ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][2, 4]
```

```r
p2 <- ggplot(data_df, aes(x = Nitrogen, y = Qsoil, col = TRY)) + geom_point(shape = ".",
    alpha = 0.5, size = 3.5) + theme_bw(base_size = 35) + xlab("Leaf Nitrogen Content") +
    ylab(VarTitle) + geom_hline(yintercept = 0, linetype = "dotted") + stat_smooth(method = "lm") +
    scale_color_manual(values = c("orange", "black"))
print(p2)
Output[[9]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Raw"), ]))[["coefficients"]][1, 1]
Output[[10]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Raw"), ]))[["coefficients"]][1, 4]
Output[[11]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Raw"), ]))[["coefficients"]][2, 1]
Output[[12]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Raw"), ]))[["coefficients"]][2, 4]
Output[[13]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Mapped"), ]))[["coefficients"]][1, 1]
Output[[14]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Mapped"), ]))[["coefficients"]][1, 4]
Output[[15]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Mapped"), ]))[["coefficients"]][2, 1]
Output[[16]] <- summary(lm(Qsoil ~ Nitrogen, data = data_df[which(data_df$TRY ==
    "Mapped"), ]))[["coefficients"]][2, 4]
if (Variable == "Qsoil") {
    if (length(which(!is.na(data_df$Height[which(data_df$TRY == "Raw")]))) <
        2) {
        Output[[17]] <- NA
        Output[[18]] <- NA
        Output[[19]] <- NA
        Output[[20]] <- NA
    } else {
        Output[[17]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][1, 1]
        Output[[18]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][1, 4]
        Output[[19]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][2, 1]
        Output[[20]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
            "Raw"), ]))[["coefficients"]][2, 4]
    }
    Output[[21]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 1]
    Output[[22]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 4]
    Output[[23]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][2, 1]
    Output[[24]] <- summary(lm(Memory ~ Height, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][2, 4]
    Output[[25]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Raw"), ]))[["coefficients"]][1, 1]
    Output[[26]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Raw"), ]))[["coefficients"]][1, 4]
    Output[[27]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Raw"), ]))[["coefficients"]][2, 1]
    Output[[28]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Raw"), ]))[["coefficients"]][2, 4]
    Output[[29]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 1]
    Output[[30]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
        "Mapped"), ]))[["coefficients"]][1, 4]
```

```
        Output[[31]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
            "Mapped"), ]))[["coefficients"]][2, 1]
        Output[[32]] <- summary(lm(Memory ~ Nitrogen, data = data_df[which(data_df$TRY ==
            "Mapped"), ]))[["coefficients"]][2, 4]
    }
    return(Output)
}
```

Chunk 34: Linear regression coefficients of TRY data and vegetation memory coeffcients.

```
TryTab <- function(ModelCoeffs, SoilLayer, Region, Variable) {
    ModelCoeffs <- unlist(ModelCoeffs)
    if (Variable != "Qsoil") {
        tabout <- data.frame(Column = ModelCoeffs[1:2])
        starts <- seq(1, length(ModelCoeffs), by = 2)
        for (i in 2:(length(ModelCoeffs)/2)) {
            tabout <- cbind(tabout, ModelCoeffs[starts[i]:(starts[i] + 1)])
        }
        colnames(tabout) <- rep(c("$I$", "$S$"), 4)
        rownames(tabout) <- c("Value", "$p_{Value}$")
        tabout <- round(tabout, 5)
        Short <- paste("PFTs and ", Variable, " Memory (", Region, ")", sep = "")
        Long <- paste("Coefficients of linear regressions of ", Variable, " response coefficients against PFT
             data (vegetative height H, and leaf nitrogen mass Nmass) both as raw geo-referenced data (R),
             and extended maps of species-specific trait means (M).
             Established via \\ref{rTRYresTab}.",
            sep = "")
        kab <- kable(tabout, booktabs = TRUE, caption = paste("\\textbf{", Short,
            " -} ", Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
        kab <- add_header_above(kable_input = kab, header = c(` ` = 1, R = 2, M = 2,
            R = 2, M = 2), align = "c")
        kab <- add_header_above(kable_input = kab, header = c(` ` = 1, `Vegetative Height` = 4,
            `Nitrogen Content` = 4), bold = TRUE, align = "c")
    } else {
        tabout <- data.frame(Column = ModelCoeffs[1:4])
        starts <- seq(1, length(ModelCoeffs), by = 4)
        for (i in 2:(length(ModelCoeffs)/4)) {
            tabout <- cbind(tabout, ModelCoeffs[starts[i]:(starts[i] + 3)])
        }
        colnames(tabout) <- rep(c("R", "M"), 4)
        rownames(tabout) <- c("Intercept", "$p_{Intercept}$", "Slope", "$p_{Slope}$")
        Short <- paste("PFTs and Qsoil", SoilLayer, " Memory (", Region, ")", sep = "")
        Long <- paste("Coefficients of linear regressions of Qsoil", SoilLayer, " response coefficients against PFT
             data (vegetative height H, and leaf nitrogen mass Nmass) both as raw geo-referenced data (R),
             and extended maps of species-specific trait means (M).
             Established via \\ref{rTRYresTab}.",
            sep = "")
        kab <- kable(tabout, booktabs = TRUE, caption = paste("\\textbf{", Short,
            " -} ", Long, sep = ""), caption.short = Short, escape = FALSE, format = "latex")
        kab <- add_header_above(kable_input = kab, header = c(` ` = 1, H = 2, Nmass = 2,
            H = 2, Nmass = 2), italic = FALSE, align = "c")
        kab <- add_header_above(kable_input = kab, header = c(` ` = 1, Qsoil = 4,
            Lag = 4), bold = TRUE, align = "c")
    }
    kab <- kable_styling(kab, latex_options = c("hold_position"))
    print(kab)
}
```

A.5.6 Miscellaneous Figures

Chunk 35: Producing miscellaneous figures used for flow charts and scheme overviews.

```
rm(list = ls())  # clearing environment
####--------------- PACKAGES ----
source("Y - Codes/S0a_Packages.R")  # loading packages
####--------------- DIRECTORIES ----
source("Y - Codes/S0b_Directories.R")  # setting directories
####--------------- FUNCTIONS ----
source("Y - Codes/S0c_Functions.R")  # Loading miscellaneous functions

## MEMORY COMPONENT SCHEMATIC -----
cell <- 34881
RegionFile <- "Iberian Region"
From <- 1982
To <- 2015
Lags <- 0:12
Yearvec <- rep(c(1982:2015), each = 12)
Time <- paste("1", (From - 1), "_12", To, sep = "")
DetVars <- c("NDVI", "Qsoil1", "Qsoil2", "Qsoil3", "Qsoil4", "Tair")
Dir.Memory.Reg <- paste(Dir.Memory, "/", RegionFile, "-", From - ceiling(1/12 * max(Lags)),
    "_", To, sep = "")
# Era5
Environment <- list.files(Dir.ERA.Monthly)
EnvRegion <- Environment[grep(RegionFile, Environment)]
EnvRegion <- EnvRegion[grep(Time, EnvRegion)]
setwd(Dir.ERA.Monthly)
Qsoil1 <- brick(EnvRegion[1])
Qsoil2 <- brick(EnvRegion[2])
Qsoil3 <- brick(EnvRegion[3])
Qsoil4 <- brick(EnvRegion[4])
Tair <- brick(EnvRegion[5])
# NDVI
Vegetation <- list.files(Dir.Gimms.Monthly)
VegRegion <- Vegetation[grep(RegionFile, Vegetation)]
setwd(Dir.Gimms.Monthly)
NDVI <- brick(VegRegion[1])
NDVI <- NDVI[[c(as.numeric(min(which(Yearvec == From)):max(which(Yearvec == To))))]]
setwd(mainDir)
# Memory
Memory <- list.files(Dir.Memory)
Memory <- Memory[grep(RegionFile, Memory)]
Memory <- Memory[grep(paste(From - ceiling(1/12 * max(Lags)), To, sep = "-"), Memory)]
## Plotting Frames
cells <- adjacent(NDVI, cells = cell, directions = 8, include = TRUE)
cells <- cells[10:18]
NDVI_vec <- as.numeric(apply(NDVI[cells], 2, mean))
NDVI_df <- data.frame(NDVI = c(as.numeric(NDVI_vec)))
NDVI_df <- transform(NDVI_df, DeNDVI = c(detrend(NDVI, tt = "linear")))
Data_df <- data.frame(Qsoil1 = as.numeric(apply(Qsoil1[cells], 2, mean)), Qsoil2 = as.numeric(apply(Qsoil2[cells],
    2, mean)), Qsoil3 = as.numeric(apply(Qsoil3[cells], 2, mean)), Qsoil4 = as.numeric(apply(Qsoil4[cells],
    2, mean)), Tair = as.numeric(apply(Tair[cells], 2, mean)))
# Data Manipulation
Data_df <- transform(Data_df, NDVI = c(NDVI_df$NDVI, rep(NA, 12)), DeNDVI = c(NDVI_df$DeNDVI,
    rep(NA, 12)), DeQsoil1 = detrend(Qsoil1, tt = "linear"), DeQsoil2 = detrend(Qsoil2,
    tt = "linear"), DeQsoil3 = detrend(Qsoil3, tt = "linear"), DeQsoil4 = detrend(Qsoil4,
    tt = "linear"), DeTair = detrend(Tair, tt = "linear"))  # linear detrending)
Data_df <- as.data.frame(cbind(Data_df[, 6], Data_df[, c(1:5, 7:12)]))
```

```r
colnames(Data_df)[1] <- "NDVI"
Data_df$Qsoil1 <- c(Data_df$Qsoil1[c(-1, -2)], NA, NA)
Data_df1 <- Data_df[1:24, ]
## Intrinsic
Intrinsic_df <- data.frame(Data = Data_df1$NDVI, Variable = rep("Response", dim(Data_df1)[1]),
    Time = rep(1:dim(Data_df1)[1], 1))
ggplot(data = Intrinsic_df, aes(x = Time, y = Data, color = Variable)) + geom_line() +
    theme_bw() + geom_point() + scale_colour_manual(values = c("darkgreen")) + ggtitle("Intrinsic Memory") +
    xlab("Time [months]")
## Extrinsic
Extrinsic_df <- data.frame(Data = c(Data_df1$NDVI, Data_df1$Qsoil1), Variable = c(rep("Response",
    dim(Data_df1)[1]), rep("Extrinsic", dim(Data_df1)[1])), Time = rep(1:dim(Data_df1)[1],
    2))
ggplot(data = Extrinsic_df, aes(x = Time, y = Data, color = Variable)) + geom_line() +
    theme_bw() + geom_point() + scale_colour_manual(values = c("blue", "darkgreen")) +
    ggtitle("Extrinsic Memory") + xlab("Time [months]")

## KRIGING SCHEMATIC ----- data
setwd(Dir.ERA.Monthly)
tairf <- brick(list.files()[15])
setwd(Dir.ERA)
tairr <- brick(list.files()[5])
# cropping
Shapes <- readOGR(Dir.Mask, "ne_50m_admin_0_countries", verbose = FALSE)
RegObj <- RegionSelection(Region = c("Portugal", "Spain", "France", "Andorra"), RegionFile = "Iberian Region",
    Extent = extent(-10, 10, 35, 52))
area <- RegObj[[1]]
location <- RegObj[[2]]
RegionFile <- RegObj[[3]]
rasinter <- crop(tairr, area)  # cropping to extent
tairr <- mask(rasinter, Shapes[location, ])  # masking
# plotting
col.tair <- colorRampPalette(c("blue", "turquoise", "yellow", "orange", "red"))(100)
plot(tairr[[1]], colNA = "black", main = paste("Tair January 1981"), cex.main = 1.5,
    legend.width = 2, legend.shrink = 1, axis.args = list(cex.axis = 1.5), cex.axis = 1,
    col = col.tair, axes = TRUE, legend = TRUE)
plot(tairf[[1]], colNA = "black", main = paste("Tair January 1981"), cex.main = 1.5,
    legend.width = 2, legend.shrink = 1, axis.args = list(cex.axis = 1.5), cex.axis = 1,
    col = col.tair, axes = TRUE, legend = TRUE)

## MODEL SCHEMATIC -----
load(paste(Dir.Data, "ModData_df.RData", sep = "/"))  # model data for pixel 35363 in Iberian rasters
## location plot
setwd(Dir.Gimms.Monthly)
IbRas <- mean(brick(list.files()[10], sep = "/"))
cells <- adjacent(IbRas, cells = 35363, directions = 8, include = TRUE)
cells <- cells[10:18]
Cellras <- IbRas
values(Cellras)[cells] <- 8888
values(Cellras)[which(values(Cellras) != 8888)] <- NA
plot(IbRas, colNA = "black", main = "Examplary Raster Cells", cex.main = 1.5, legend = FALSE,
    legend.width = 1, legend.shrink = 1, axis.args = list(cex.axis = 1), cex.axis = 1)
plot(Cellras, col = "red", add = TRUE, legend = FALSE, axis.args = list(cex.axis = 1),
    cex.axis = 1)
## NDVI series
plot_df <- data.frame(Months = rep(1:dim(ModData_df)[1], 2), NDVI = c(ModData_df$`NDVI_anom[1:length(Clim2_anom)]`,
    ModData_df$`NDVI_Lag1[1:length(Clim2_anom)]`), Data = c(rep("Z-Scores", dim(ModData_df)[1]),
    rep("t-1", dim(ModData_df)[1])))
```

```r
ggplot(plot_df, aes(x = Months, y = NDVI, col = Data)) + geom_line(size = 1.2) +
    theme_bw() + labs(title = "NDVI time series")
## Cummulative soil moisture lags
plot_df <- with(ModData_df, data.frame(Months = rep(1:dim(ModData_df)[1], 1), Qsoil = c(ClimCum_6)))
ggplot(plot_df, aes(x = Months, y = Qsoil)) + geom_line(col = "navyblue", size = 1.2) +
    theme_bw() + labs(title = "Cummulative Soil Moisture (0-7cm) Lag 6")
## Tair
plot_df <- with(ModData_df, data.frame(Months = rep(1:dim(ModData_df)[1], 1), Qsoil = Clim2_anom))
ggplot(plot_df, aes(x = Months, y = Qsoil)) + geom_line(col = "red", size = 1.2) +
    theme_bw() + labs(title = "Air Temperature")
## PCA
biplot(princomp(pca_mat))
```

A.6 Declaration Of Authorship

I, Erik Kusch, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Erik Kusch

____

Signature

____

Date