TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITETET I BERGEN

Remote Sensing And Predicting Shifts In Biome Distribution And Resilience Using NDVI Data

Bachelor's Thesis (B.Sc.)

In Partial Fulfilment of the Requirements for the Degree Bachelor of Science (B.Sc.) at the Technical University of Dresden

Primary Reviewer: Dr. Alistair Seddon

Secondary Reviewer: Prof. Dr. Klaus Reinhardt

by

Erik Kusch

born 03/01/1995 in Eilenburg, Germany

Enrolment number: 3990898

10/03/2017

at the Technichal University of Dresden

Abstract

The dependence of humans on stable, intact ecosystems, whose future characteristics and responses to perturbation are understood, has lead to an increasing appreciation of the need for resilience goals to be incorporated into environmental policy catalogues[1,2]. The concept of resilience is not new to ecological sciences. However, the great variety of understandings of its meaning and ensuing confusion of the base-line terminology often impede accurate, quantitative assessments of resilience properties. Nevertheless, such knowledge of resilience of natural ecosystems is paramount for the formulation, implementation and evaluation of policy plans which are at the core of environmental Decision Support Systems (DSSs)[3]. This thesis focussed therefore on presenting a novel framework for (1) identification of alternative stable states, whose existence is a premise to multiple definitions of resilience, (2) determining what drives their distribution patterns and how, as well as (3) assessing the resilient behaviour exerted by those alternative stable states and the greater systems they belong to. The results of the statistical analyses employed within this study proved valuable in adhering to the three corner stones of the proposed framework and resulted in the delineation of ecosystem types which were easily linked to real-world formations of vegetation compositions. Although one of the approaches used to assess resilience within this study had to be dismissed, two other approaches employed for the same task yielded useful information on resilience properties. Additionally, one of these methods was used to identify regions of ecological uncertainty on which to direct future research and policy-making efforts. Consequently, this study may be 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 thank Dr. Alistair Seddon for supervising my thesis, redirecting me when I was steering into dead ends as well as giving me the lee-way and challenging me to really learn how to use R and find solutions to problems in the coding, that might have seemed trivial to him from time to time, by myself.Furthermore, this work couldn't have been conducted without the readiness of Prof. Dr. Klaus Reinhardt to accommodate my thesis in his busy schedule and to act as a supervisor at the Technical University of Dresden. Many of my thanks also go to Prof. Dr. Kathy Willis with the long term ecology lab at Oxford for considering my speculative application to her research group and consequently referring me to Dr. Alistair Seddon. My gratitude also lies with P.E. Dr. John Birks with whom I shared interesting and informative meetings on the vegetation of Minnesota and whom I have to thank for inspiring a good portion of my thesis with thoughtful remarks he gave on preliminary results of my work.

Also, I would like to thank my family for backing me up throughout my academic endeavours so far. Last but not least I wish to thank PD. Dr. Dieter Dörnemann who started me on my path in Biology when he agreed to have me as his protégé to work on an individual research project I had come up with back in 2012.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

AUC Area Under Curve

AVHRR Advanced Very High Resolution Radiometer

BIC Bayesian Information Criterion

CSD Critical Slowing Down

DEM Digital Elevation Model

DSSs Decision Support Systems

EVI Enhanced Vegetation Index

GAM Generalised Additive Model

GAMs Generalised Additive Models

GIMMS Global Inventory Modelling and Mapping Studies

GMTED2010 Global Multi-resolution Terrain Elevation Data 2010

LCCS Land Cover Classification Schemes

NDVI Normalized Difference Vegetation Index

PR Presumed Resilience

ROC Receiver Operating Characteristic

TAC Temporal Autocorrelation

VI Vegetation Index

VIs Vegetation Indices

1. Introduction

Mankind is heavily reliant on goods and services provided by terrestrial ecosystems[4]. Due to an increased effect of humans on their environment[5] and recognition of climate change as an additional influencing factor[6–8], management efforts have been made to protect ecosystems from degradation[3,6,9–11]. The effectiveness of such management endeavours is largely dependent on the accuracy of the fundamental grasp on ecosystem functioning and indicators of its state.

Human dependence on 'intact' ecosystems for resources, recreation and even the regulation of earth's cycles of materials and energy is enforcing the importance of a thorough understanding of terrestrial ecosystems[12,13]. As a consequence of governmental institutions recognizing the necessity for intact natural ecosystems, decisive environmental policy goals have been formulated to avoid degradation and enhance the recovery of perturbed ecosystems. The Aichi targets, for example, establish 20 policy goals to maintain the biodiversity of natural ecosystem on earth through several approaches. Biodiversity target 15 of the Aichi policies catalogue encompasses the obligation for *'maintaining and enhancing the resilience of natural ecosystems to mitigate climate change effects by 2020'*[1].

'Resilience' is a key concept to many ecosystem management efforts and recognized widely as an important component of ecosystem functioning. Especially, in times of global change[14] and increased human influence[12]. It describes a system's capability of responding to a disturbance by resisting the perturbation and recovering quickly from it[15] thus prompting stability of ecosystems in their distributions and functioning despite the pressure exerted by abiotic and biotic stressors. However, the concept of resilience is not a carte blanche for humankind to do as it pleases and count on ecosystems to withstand any perturbation in the long run. In fact, resilience of certain systems can be unfavourable for human needs and requirements of their environment[12]. The importance of the concept of resilience has grown considerably over the past $\sim$ 50 years[15] and culminated in the formulation of resilience policy goals such as Aichi target 15. From this popularity of the resilience concept and importance of detailed knowledge about the resilience exhibited by ecosystems arises one central question:

How can resilience be assessed and measured?

Answering this questions will enable policy makers to further specify their goals, lead to educated decisions on human interaction with the natural environment to be made and researchers being able to assess the extent to which policy goals have been fulfilled thus allowing for further advising of policy makers and the society at large.

1.1 Resilience Theory

Creating strategies to maintain or boost resilience and evaluate their respective efficacies requires a thorough understanding of resilience theory and ways 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[16–18]. First introduced to biological sciences in 1973 by Holling et. al resilience thinking was envisaged to encompass two separate characteristics of ecosystems: *stability* and *resilience*[18]. Stability was 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. 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. 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[18].
Since then, the term 'resilience' itself has undergone a 'greenwash'[2], which has diminished its direct usefulness to researchers. This process has seen the term 'resilience' being introduced originally to ecological sciences but adapted by other branches of research (e.g. social sciences[19]) and society (e.g. policy makers[2,20]), modified for different purposes and re-introduced to ecological sciences with multiple altered definitions[3].
Whilst the state of any given ecosystem can be assessed without entertaining the underlying implications of a resilience framework, the conclusions from such a study may be grossly differing from the actual processes in nature. Assumptions on ecosystem functioning that are made without including resilience thinking, irrespective of this caveat, may result in particularly dire consequences if the systems in question (also referred to as focal systems[12,21]) are production ecosystems[3,12] (ecosystems which are used by humans for resource allocation). A prominent example of environmental mismanagement is represented by the collapse of North Atlantic cod stocks, where abundant populations collapsed after overfishing[22] and never recovered[3].

1.1.1 Terminology

An overview of the major corner stones of the framework to resilience thinking has been arranged in table A.1. Furthermore, a list of recent and frequently used definitions of 'resilience' has been compiled in table A.3 to avoid any confusions.
The two most prominent definitions of resilience, *ecological resilience*[18] and *engineering resilience*[23], have gained considerable traction and it is pivotal to draw attention to the key difference in their underlying assumptions of ecosystem trajectories before being able to make a deduction on which to use:

1. *Ecological resilience* is concerned with ecosystems that have multiple attractors[18,24] thus operating in a non-linear environment.

2. *Engineering resilience* considers ecosystems to be subjected to only one attractor[6,23] therefore assuming a linear development.

These concepts are illustrated in figure 1.1 by means of stability landscapes[14,16] and, in practice, set the scope of hypotheses that can be formulated concerning the functioning of a focal system. According to ecological resilience, a focal system does not freely change state but only does so when pushed from the current attractor domain into that of another, alternate attractor. Once it settles at this attractor, it is in a so-called alternative stable state to its initial condition. Studies which implement ecological resilience are therefore used to make assertions on how much a system can be perturbed from its current state until it shifts into being in an alternative stable state. In engineering resilience, the possibility of alternative stable states is not considered. Because of this drawback, engineering resilience is particularly useful only when considered as a component of ecological resilience. In this case, one can make deductions on how long one has to wait after an intervention for further degrading interactions with the system to be sustainable[3].

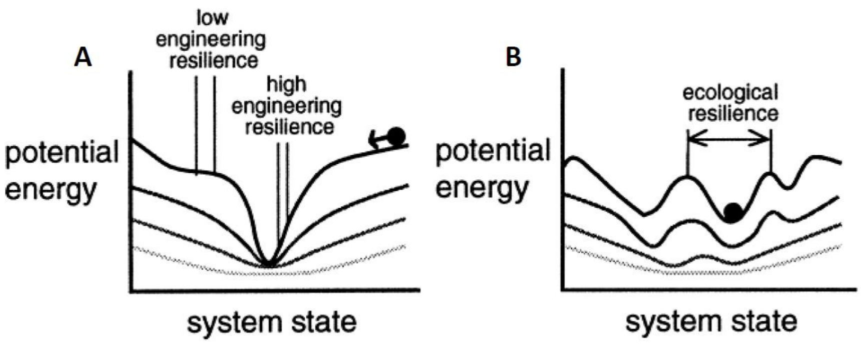


Figure 1.1: Stability Landscapes according to Engineering and Ecological Resilience[24]. Resilience can be illustrated using stability landscapes along which the system state (represented by a ball) can move. A 'valley' in this 'landscape' represents an attractor. If multiple attractors are present one refers to them as alternate attractors. A) *Engineering Resilience* presumes systems to adhere to only one attractor. Irrespective of frequency or intensity of disturbances these systems will always return to their one and only attractor. The magnitude of engineering resilience is defined by the steepness of the stability-slope which indicates the return time from the post-disturbance to the pre-disturbance state and is thus reminiscent of Holling et. al's notion of stability[18]. B) *Ecological Resilience* works on the premise of natural systems being able to comply to several attractors. It describes how intense a disturbance has to be to 'knock' the system out of the reach of one attractor and into an alternate attractor domain thus referring to the width of the attractor domain (the 'valley').

A new understanding of 'resilience' was introduced by Carpenter et. al in 2001 with the notion of *specified resilience*[19]. Specified resilience is a realistic approach for assessing resilience in that it does not aim to incorporate all possible trajectories and characteristics of a focal system but instead characterises the resilience of *what to what*. The first *what* is defined as as a set of specified system characteristics whose resilient nature is to be assessed. The *to what* part describes a set of trajectories that are hypothesized to influence the system characteristics in relation to which the system's resilience is to be measured.

1.1.2 Alternative Stable States

A common goal of environmental policies is to keep natural ecosystems 'as is' and return disturbed ecosystems to their original state. However, it may be more of an realistic approach not to try to confine ecosystems to certain states but incorporate their inherent potential for change and acknowledge the tendency to do so owing to stressors of the anthropocene[12,25] as well as natural shifts in the environment[26]. The view on behaviour of ecosystems that follows from this should encompass the ability of systems to alter their state as this allows for goals to be established and planned action to be taken to turn a system from a degraded state into a favourable one whilst also taking into consideration that human action could cause the inverse. This, in turn, would entail the assumption of alternative stable states to be present.
Stable states of ecological systems have been hypothesized to be self-enforcing through feedback mechanisms[18,27]. Such feedback mechanisms are exclusive to each state the system can be in and thus result in *alternative* stable states with individually exclusive sets of state-governing processes. Working theory suggests that ecosystems which establish stabilizing feedbacks are more resilient than those which do not[28].

A prominent example of feedback mechanisms enforcing unique, alternative stable states can bee seen in fire rates driving the distributions of savannah and forest ecosystems. High fire rates inhibit tree growth which allows for growth of grass which, in turn, serves as fuel for further fires. Past a critical threshold of tree cover however, fires are inhibited due to the lack of grass which is overshadowed by trees leading to furthered tree growth and higher tree cover percentages[29].

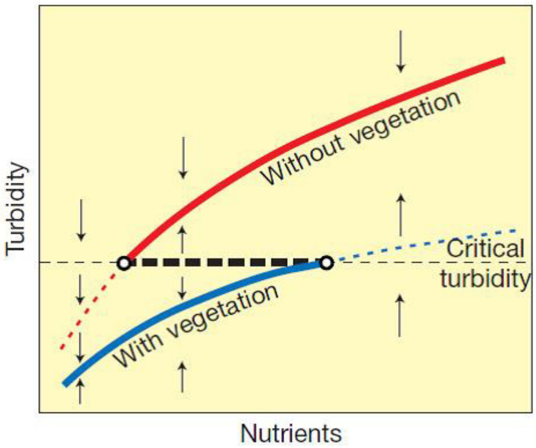
Scheffer et. al proposed vegetation cover as a control parameter for ecosystem state in shallow lakes that unveils two alternative stable states in relation to nutrient level[14] as seen in figure 1.2a. The hypothesis which has been employed to explain this behaviour included three assumptions on aquatic ecosystem functioning:

1. Water turbidity increases with nutrient level.

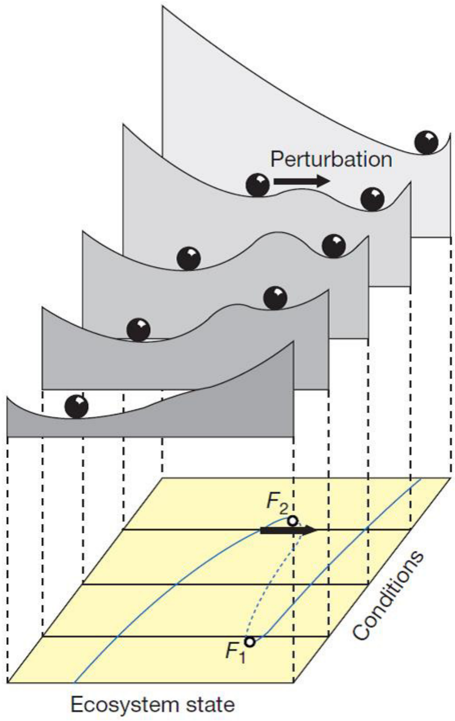
2. Submerged vegetation reduces turbidity.

3. Vegetation disappears when a critical turbidity is exceeded.

Owing to the first two assumptions, equilibrium turbidity can be explained using two different functions of the nutrient level: one for a vegetation-dominated system, and one for an unvegetated system. Above a critical turbidity, vegetation will vanish. In this case, the upper equilibrium line in figure 1.2a is relevant. Below the critical turbidity, the lower equilibrium curve is exerted. As a result, low nutrient levels are characterised by a vegetation-dominated regime whilst high nutrient levels allow only an unvegetated state to persist. At intermediate nutrient levels, the two states exist as alternatives to each other separated by a (dashed) unstable equilibrium. Scheffer et. al also linked alternative stable states to ecological resilience by establishing three-dimensional stability landscapes (as seen in figure 1.2b) which express the resilience of the ecosystem given a particular relation of conditions to the ecosystem state at five different levels of conditions.



(a) Alternative stable states in shallow lakes



(b) Resilience in three dimensions

Figure 1.2: Alternative stable states and resilience[14]: a) The model employed by Scheffer et. al works on the basis of three assumptions which lead a vegetated and an unvegetated state to exist as true alternative stable states. b) The bottom plane presents an equilibrium curve. The stability landscapes depict the attractors at five different conditions. Stable states correspond to valleys and unstable sections translate into hills. The size of the attractor valley is representative of resilience.

Another example of alternative stable states being enforced by feedback mechanisms is the regime shift from 'wet' to 'dry' conditions in the Sahara and Sahel region approximately 5,500 years ago[26] where variations in the Earth's orbit initially lead to the 'wet' state. This state was enhanced through monsoons in the Sahara and Sahel regions which increased the extent of vegetation cover, lakes, and wetlands. The lower albedo and increased ability to re-cycle water, which resulted from this, both helped fuelling the monsoon rains with additional energy and moisture, thus increasing the summer rains and producing a positive feedback on the orbital variations. These feedbacks work in either direction and are hypothesized to establish and maintain two distinct stable states: a wet, 'green' and a dry, 'desert' Sahara[26]. A visual representation of these processes is presented in figure 1.3.

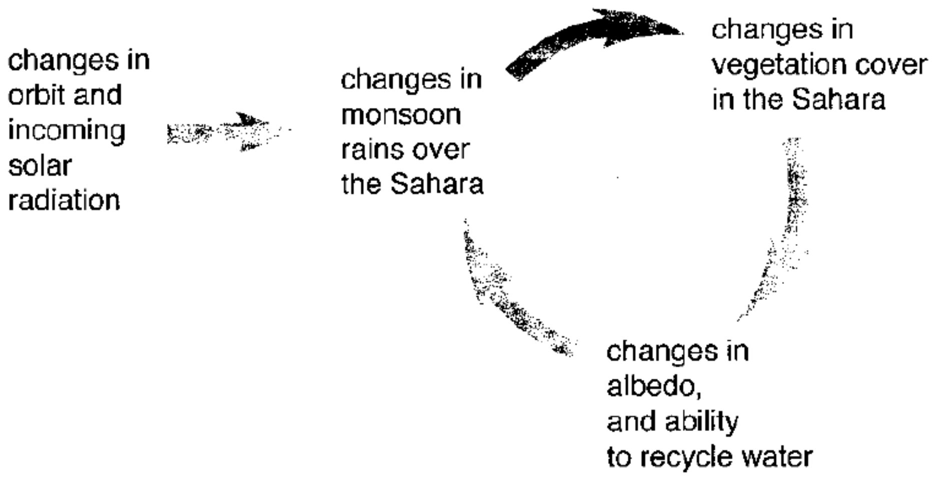


Figure 1.3: Atmosphere-Vegetation Feedbacks:[26] A feedback loop has been proposed for the autonomous intensification of monsoon rains across the Sahra and Sahel. These feedbacks have been hypothesized to have led to the regime shift from the 'wet' to the 'dry' Sahara and Sahel regions approximately 5,500 years ago.

Due to the implication of alternative stable states in *ecological resilience* (see figure 1.1), the resilience framework in this study refers to that of ecological resilience as defined by Holling et. al[18] (see table A.3). Resilience assessments themselves have been carried out as measurements of *specified resilience* (nested within the framework of ecological resilience) as this allowed for a realistic subset of ecosystem descriptors and trajectories to be chosen for the analysis.

1.2 Linking Theory And Observations

The key challenge of any well-established resilience framework is to link the theory with observations thus making resilience of systems and achievement of environmental policies quantifiable. Linking the theory of Holling's resilience framework[18] to observations is a two-step process:

1. **Identifying alternative stable states** will delineate the number and characteristics of attractors whose resilience properties have to be assessed. A variety of methods to separate alternative stable states from a continuum of data points have been proposed by multiple authors. Scheffer & Carpenter established a framework of empirical evidence for the presence of alternative stable states which can be used for this purpose[9].

2. **Measuring the resilience of the alternative stable states** can be undertaken in a multitude of ways. Hodgson et. al established a framework concerned with the measurement of resilience via its components for this purpose[15].

1.2.1 Alternative Stable States

Lengthy arguments have been made for and against the claim that natural ecosystems adhere to multiple attractors. Some ecosystems (such as coral reefs) have been a major focus of this debate. Mumby et. al stated that there is no evidence against the existence of alternate attractors in Caribbean reefs yet there is no definitive proof of their existence either[30]. Hodgson et. al, on the other hand, assume coral reefs to adhere to at least two alternate attractors: coral-dominated and algal-dominated patterns[15].
Scheffer & Carpenter established three unique approaches to confirm or repudiate the presence of alternative stable states in a focal system through analyses of empirical data[9]:

1. *Jumps in time series.* An abrupt shift in control parameters of ecosystem state may be indicative of a stepwise change or the surpassing of a critical threshold. Such observations can especially be triggered by mass colonisation or extinction events and are not indicative of alternative stable states unless the altered control parameters persist for an ecological significant amount of time (in their modified states).

2. *Multimodality* of frequency distribution of states. The frequency distribution of unique ecosystem characteristics are expected to be multimodal if alternative stable states are present. Using this method requires extensive data sets to be available and is not necessarily indicative of the presence of alternative stable states within a focal system due to alternative explanations for multimodality being available.

3. *Dual relationships* of system states to control factors. A dual relationship describes a relationship of system state to its control factors which is best explained not by a single regression function alone but by multiple functions simultaneously. This phenomenon is best accounted for through modelling procedures and model selection processes like

likelihood ratios, the least squares method, etc. to account for the best fitting model. This method is also no definitive proof of the existence of alternative stable states due to alternative explanations. However, a dual relationship can be indicative of an underlying hysteresis curve (see table A.1) which would suggest resilient behaviour[29].

Following the scientific rule of the absence of evidence not being the evidence of absence, not being able to confirm the presence of alternate attractors does not necessarily mean that they don't exist. Especially since the gradual shifts between them can take a long time to even reach noticeable levels[14,31]. Conclusively, the point has been made that, in light of current theory and empirical data, the null model should be that ecosystems have alternate attractors and the burden of proof should be put on verifying their non-existence rather than their presence[9] thus enforcing the assumptions of ecological resilience and setting the scope for resilience assessment.

1.2.2 Assessing Resilience

Assigning practical use to 'resilience' in ecology, thus yielding robust information for environmental DSSs, requires the resilience metric to be as inclusive of processes influencing it as possible as well as quantifiable features of a system to be determined.

Working theory suggests that regime shifts and prominent resilience patterns may result from the forcing of various processes. For example, such shifts can occur due to extrinsic or intrinsic forcing leading to extrinsic and intrinsic regime shifts respectively[32]. A non-exhaustive list of intrinsic and extrinsic factors has been compiled in table A.2. Regime shifts can furthermore be the product of gradually changing environmental conditions that cross a regime-specific threshold[9] or a result of abrupt forcing[33]. Such processes and trajectories are not limited to the focal system itself but can extend far beyond its scales and multiple authors have recently proposed general ways in which to approach resilience measurements[6,21,34].

Quantifying stability landscapes as seen in figure 1.1 requires perfect understanding of the system that is being assessed. Acquiring such holistic and complete insight on a natural ecosystem is immensely laborious and time-consuming reminiscent of a Sisyphean task. Consequently, new methods of measuring resilience have been proposed which assess resilience not directly as a single metric but by identifying and analysing important resilience components[15].

These important components include observable warning signs of 'failing' resilience or approaching regime shifts like Temporal Autocorrelation (TAC) or Critical Slowing Down (CSD) once a system approaches a threshold as expressed by (1) skewness, (2) the coefficient of autocorrelation, (3) the coefficient of spatial correlation, (4) the coefficient of variation which have been described frequently in recent literature[35,36].

A prominent example of CSD has been demonstrated in an experiment on cladoceran zooplankton by Drake et. al in which one group was subjected to a deteriorating environment whilst the control treatment was kept at steady conditions. The results of their study showed that there were periods of CSD preceding the occurrence of regime shifts in the plankton populations with marked increase in all four parameters (skewness, coefficient of autocorrelation,

coefficient of spatial correlation and coefficient of variation)[37]. The concept of CSD as an indicator of shrinking resilience potential of a system has been widely adapted by other authors as well[3,11,16,31,35,36]. TAC has been frequently employed in similar ways[7,16,31,35,38,39]. However, there may be limitations to these approaches since alternative explanations to explain the cause of TAC and CSD patterns could apply.
In a recent paper, Hodgson et. al tackled the problem of resilience measurements by building a framework for resilience assessments by first representing resilience as two separate entities[15]:

1. **Resistance** to perturbation.

2. **Recovery** from perturbation.

Secondly, Hodgson et. al introduced new parameters to consider in assessing relative resilience of ecosystems and whether resilience in particular systems is achieved via resistance or recovery:

- *Latitude.* In terms of resilience landscapes, the latitude is a measurement for the distance from one stable state to the nearest bifurcation point past which an alternative attractor takes over.

- *Precariousness* denotes the distance from the current state to the nearest tipping point.

- *Elasticity* is a measurement for return time, represented by the steepness of the curves in resilience landscape plots and thus equivalent to engineering resilience.

At the heart of these measurements lies the observation and quantification of *change in ecosystem state* which can be done by assessing the initial state through a control variable which is subsequently tracked throughout a time period in which the system is subject to change. Hodgson et. al concluded that a system with high elasticity is mostly resilient through recovery, whilst systems with high precariousness are mostly kept from changing states by means of resistance[15].

1.2.3 Measuring Resilience Via Remote Sensing

Remote sensing reliant studies are a tool especially prone to resilience research in a holistic manner since their coverage is mostly at regional scales[40–42] thus encompassing multiple stable states and utilizing satellite data due to its usefulness in tracking vegetation changes given high temporal resolution[43,44].
Recent remote sensing studies have tackled the challenge of assessing resilience in a multitude of ways which can generally be classified within the framework established by Hodgson et al.
Prominent examples of *elasticity* measurements have been carried out frequently. TAC, for example, has been acknowledged as an indicator of response speed of ecosystems to disturbance in a recent remote sensing study by Seddon et. al. Within their study, vegetation sensitivity was measured by identifying areas of weak resistance to climate variation and slow recovery from a perturbed state as areas vegetated by sensitive plant assemblages and characterised by

low TAC scores[7]. The resulting vegetation sensitivity index was compared to local variance in vegetation reflectance data, air temperature, cloud cover and water availability, both directly and via lag-effects, to identify the role that ecosystem memory effects played in areas of high vegetation sensitivity. TAC was further used in a study on the resilience of tropical forests using a remote sensing approach by Verbesselt et. al. The relationship of TAC of monthly vegetation reflectance records to mean annual precipitation, mean annual temperature and soil quality as well as further confounding variables (tree cover, seasonality, etc.) were assessed using additive regression modelling. The study resulted in tropical forests presumably being resilient at mean annual precipitation levels $> 1{,}500$ mm[35]. Following the same line of thought, de Keersmaecker et. al measured resilience of vegetation assemblages as *engineering resilience* by relating anomalies in vegetation reflectance time series to anomalies in time series of a drought index and mean temperature of the same time frame[45]. The states which were identified as sensitive to drought-impact (strong anomalies in vegetation reflectance due to anomalies in drought index) were further researched for whether they were predominantly bare patches of land, vegetated by non tree plant assemblages or exhibited high tree cover fractions.
Nioti et. al investigated the resilience of pine forests in southern Greece in a post-fire state by assessing *resistance* (precariousness) and *recovery* (elasticity) of vegetation assemblages separately[33]. The observations of pre- and post-fire distribution of vegetation types were assessed for differences thus resulting in a measurement of resistance to fire for each vegetation type. By utilizing a time series of vegetation records, Nioti et. al were able to reconstruct the recovery of certain patches of pine forest and assess the recovery rate of those patches. Both resistance and recovery were then analysed for correlations with altitude, substrate and soil depth.

Recent remote sensing resilience studies have also made use of the *specified resilience* framework by Carpenter et. al[19] (see table A.3), focusing on identifying the resilience of *what to what* which can be used as a way to limit the study parameters to a reasonable number. Tree cover, which was dealt with as a confounding variable by Verbesselt et. al[35], has been a major descriptor of vegetation composition in some of these studies. Hirota et. al evaluated the vulnerability and resilience of three distinct tree cover classifications (tropical rainforest, savannah, treeless) in relation to annual precipitation regimes. The tree cover data was amassed using remote sensed spectral data. As indicators of resilience Hirota et. al proposed to analyse the relation of tree cover (*what*) to rainfall (*to what*) and to assess the fraction of sites of a particular state given a specific annual precipitation level[29].

Scheffer et. al focused on the distribution of boreal forest systems in northern latitudes using tree cover as well[46]. However, they assessed the relation of tree cover to mean July temperature and mean annual precipitation. Scheffer et. al constructed a continuous response function of tree cover as driven by mean July temperature with five distinct attractor regimes. Resilience was then expressed as the distance of a given data point to the next tipping point. The observed distribution of tree cover classes (five in total) displayed sharp boundaries which is not at all surprising given the assumption of gradually changing environmental conditions and systems with alternative attractors[47].

1.2.4 Resilience In The Arctic

Studies that focus on areas of anthropogenic interest (such as the tropical rainforest or agricultural regions) or proposed vegetation vulnerability are particularly useful. Especially areas in northern latitudes have been proposed to be under major influence by climate change[46,48]. The hypothesized influences on northern latitudes include a north facing expansion of boreal forest and loss of tundra regions[49] and thus loss of biodiversity which may lead to yet unforeseeable consequences[50] as well as changes in the dominance hierarchy of established communities with possibly detrimental effects on apex predators[51]. A rapid warming of northern latitudes will additionally transform the abiotic environment making the biological systems more susceptible to invasion by species to which little to none of the resident species may be adapted causing a chain route of small to medium scale extinction events[48,52,53]. Furthermore, change in so-called rain-on-snow (ROS) events are among the most important facets of climate change influencing flora and fauna of northern latitudes. Increased rates of ROS events will result in more severe encapsulation of vegetation in ice, thus preventing herbivores from accessing them and leading to change in herbivore populations as well as vegetation composition[48,54].
Taking these trajectories into account researchers and policy makers have exerted special interest and concern on areas of northern latitudes[55].

1.3 Study Outline

Developing remote sensing methods and refining them will facilitate important knowledge for DSSs on human interactions with the environment and thus induce more tenable guidelines for future human interference with natural ecosystems. This will garner a sophisticated understanding of resilience and, in turn, enable maintenance and enhancement of ecosystems' resilience thus aiding to stabilize natural systems.
It is therefore the encompassing aim of this study to **present a new framework for identifying and mapping alternative stable states at regional scales** and testing it on different study systems. This novel concept is aimed at assessing resilience and identifying thresholds of alternative stable states to serve in future resilience studies and the evaluation of policy actions. In order to achieve this goal, three major research questions have been considered:

1. *What and how many alternative stable states are present in a given focal system and how do their distributions change over time?*

2. *How are the observed alternative stable states related to the climate?*

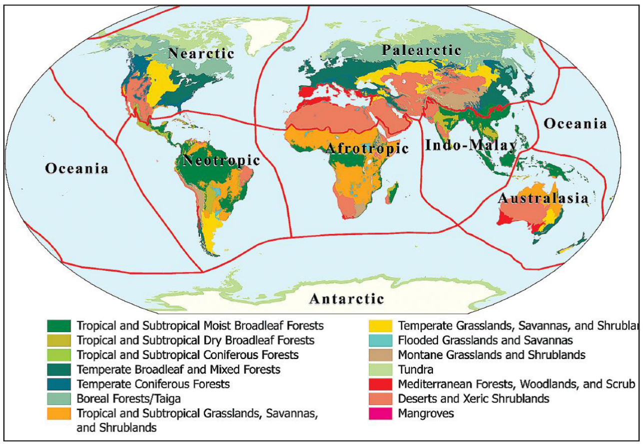
3. *How are the results of the ecosystem classification and vegetation-climate modelling linked to resilience?*

2. Conceptual Framework

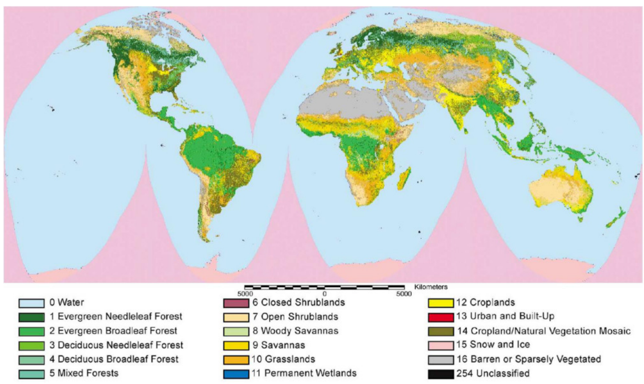
2.1 Quantity And Qualitative Properties Of Alternative Stable States

The first step towards examining the presence of alternative stable states is classifying all possible states respectively thus answering the first major research question of this study (*What and how many alternative stable states are present in a given focal system and how do their distributions change over time?*).
Hirota et. al and Scheffer et. al identified alternative stable states at regional scales as different land cover types or *biomes*[29,46].

A variety of Land Cover Classification Schemes (LCCS) have been proposed for the purpose of separating natural vegetation into several land cover types representing more or less homogeneous and unique vegetation assemblages (see table A.4). Global LCCS usually discriminate between 5[56] to 28[57] distinct terrestrial vegetation states and utilize a wide range of methods for the classification process.
Although two of the more recent LCCS[58,59] (figure 2.1) have been widely accepted by the scientific and non-scientific community alike, they have also been criticised for their subjectiveness[60].



(a) LCCS by Olson et. al, 2001[59]



(b) LCCS by Friedl et. al, 2002[58]

Figure 2.1: Land Cover Classification Schemes (LCCS): The above shown LCCS by Olson et. al and Friedl et. al are the two most widely accepted LCCS so far[60]. Both have been criticised for subjective classification of biome types and rely on different sources of information. a) Olson et. al derive the biome classification from expert knowledge and distinguish 14 individual, terrestrial biomes whereas b) Friedl et. al use preconceived assumptions on biome location and spectral data to establish an algorithm for the classification and arrive at 15 individual, terrestrial biome classification types.

Olson et. al[59] (figure 2.1a), albeit classifying the worlds ecosystems into intuitively distributed and labelled vegetation types, has come under scrutiny for relying on expert knowledge in the classifying process. This assures that the ecosystem classification is done with a sound background knowledge on landscape ecology but brings with it intrinsic subjectiveness in the resulting classification maps.
Friedl et. al's LCCS[58,61](figure 2.1b), on the other hand, rely on trained classification algorithms. This classification procedure is a three-step process. First, grid areas on the earth's surface which are believed to contain a certain vegetation/ecosystem type are selected. These areas are called 'training sites'. Secondly, the pooled spectral data of all the training sites is subjected to an algorithm that separates the different training sites according to their preconceived vegetation types and spectral data. The algorithm in this step is subject to change to generate the best possible fit of data and algorithm classification. Lastly, the optimized algorithm is run on larger sets of spectral data to classify non-training sites[58]. Despite this approach being far more mathematically developed than Olson et al's process, subjectiveness is still introduced to the classification scheme by biased choice of training sites and their respective classification prior to the training step[60]. Frield et. al's method was first introduced in 2002[58] and last updated in 2010[61].

Additionally, many LCCS and studies on change in the delineated clusters of vegetation types suffer greatly from confusion in the underlying terminology. Especially the term 'ecosystem' is very arbitrary and can take on many meanings with the described systems in nature ranging across a wide variety of spatial scales. LCCS are therefore usually employing other terms, with the one most frequently used being *'biome'*, to describe unique kinds of vegetation assemblages. Unfortunately, the term biome has been defined differently by many authors[60,62,63]. It is crucial to be precise about what is meant by 'biome' and an example of how reasonable hypotheses can be thwarted by the lack of a concise definition can be seen in Donoghue & Edwards, 2014[64]. When using remote sensed data, it is useful to accept Higgins et. al's biome definition since it does not require data which would only be obtainable through extensive ground surveys:

'Biomes are global-scale vegetation units defined by structural and functional attributes rather than by species composition.'[60]

Although some LCCS rely on already existing LCCS (see table A.4), this study incorporates two novel, minimalistic classification schemes. To assure accurate depiction of biomes, the classification procedure in this study was carried out in a way so that no preconceived assumptions over biome distribution was introduced to the biome delineation process.

2.2 Vegetation-Climate Relationships

Answering the major research questions of this study requires a set of variables to be chosen from a large pool of accessible parameters to tackle the biome delineation (answering the first research question) and modelling climate-vegetation interactions thus answering the second research question (*How are the observed alternative stable states related to the climate?*).
Remote sensing LCCS rely on satellite derived parameters for biome delineation and further analyses. Higgins et. al conceded that avoiding subjectiveness in LCCS is difficult and especially the choice of parameters which to use for biome identification contains the danger of being arbitrary in itself[60]. In fact, a multitude of variables and parameters have been proposed and used to separate biomes and to unveil the processes that drive their distributions[7,38,60]. These parameters represent two distinct groups known to statistical applications and modelling procedures:

1. **Response variables** are the variables whose variations are being modelled[65] and are represented on the y-axis in conventional two-dimensional graphs.
2. **Explanatory variables** are the variables whose variations are associated to those of the response variables and drive them[65]. They are usually presented on the x-axis in conventional plots.

Special care has to be exerted when settling on a specific set of variables since the usefulness of certain explanatory variables is location-dependent and the same goes for response variables[66].

2.2.1 Response Variables

2.2.1.1 Theory

Response variables carry direct information about the state of a focal system thus aiding to answer the 1st major research question of this study (*What and how many alternative stable states are present in a given focal system and how do their distributions change over time?*).
Various such response variables have been employed in recent studies. These include, but are not limited to: leaf phenology[41], tree cover[45,67], plant area index[68], leaf type[69,70], vegetation height[60], vegetation productivity[60], species composition[35], dominance hierarchy index[71], rooting depth[72], pollen counts[32] and Vegetation Indices (VIs)[43]. Such initial response variables are often processed into what is arguably one of the simplest response variables: a binary depiction of presence and absence of a particular biome.
In terms of this study, indicators of ecosystem characteristics have been referred to as *primary* response variables whereas binary presence descriptors haven been recognized as *secondary* response variables. A masking of the primary variables by creating a secondary response variable is common practice and often entangles clustering processes[73–75], decision trees[76], trained algorithms[58,61] and specifically developed models which incorporate multiple statistical analyses[56,73,77,78].

2.2.1.2 Vegetation Indices

Vegetation has been proposed as a sensitive information criterion of the environment reflecting human impact and climate due to its important role in the energy balance of the ground-atmosphere system of the earth[79].
Conclusively, this study employed a Vegetation Index (VI) as a primary response variable of ecosystem characteristics. VIs have been a recent addition to the arsenal of means readily available to ecologists. Generally, a VI is the result of a computational transformation of spectral data in a given extent of an area.

VIs are particularly useful for ecosystem studies since they offer extensive coverage ranging from regional scales to full global products[43,66,80,81]. However, depending on the region in question, several confounding variables have to be taken into consideration. Especially in Arctic latitudes, the effectiveness of multi-temporal optical sensing can be limited due to short growing season length, solar geometry as well as snow and ice cover[43]. The choice of observation platform is dependent mostly on regional aspects. Very large or remote study regions are not prone to the use of aircraft since these would incur extensive logistic expenditure and also introduce the problem of repeated flyover series of the same region varying in exact positioning of the aircraft in regards to latitude, longitude and elevation above the canopy. Satellites, on the other hand, suffer from lower temporal frequencies in their patterns of passing specific regions when compared to aircraft and usually yield VIs of coarser spatial resolution than aircraft based imagery[43,79]. The primary interest of many established study programs is to study terrestrial vegetation at large scales[39,42,61,68,69] which leads to VIs usually being derived from satellite data with aircraft based VIs playing a minor role at best.
Errors in VI measurements are mostly caused by atmospheric and ground conditions as well as sensor problems[42]. Cloud cover is especially limiting to VI measurements. Therefore, most VIs are available as composites of different time frames[42,68,79] to avoid missing data.

Since the first proposition of VIs in 1974[82] a variety of new VIs have been submitted. These include, among others, the Narrow-band Normalized Difference Vegetation Index (NBNDVI)[83], Photochemical Reflectance Index (PRI)[84], Nitrogen Reflectance Index (NRI)[85], Transformed Chlorophyll Absorption and Reflectance Index (TCARI)[86], Structure-Insensitive Pigment Index (SIPI)[87], Plant Scenescence Reflectance Index (PSRI)[88], Physiological Reflectance Index (PhRI)[89], Normalized Pigment Chlorophyll Ratio Index (NPCI)[90], Anthocyanin Reflectance Index (ARI)[91], Enhanced Vegetation Index (EVI)[7,43,68,72] and Leaf Area Index (LAI)[70,92].
See Devadas et. al[93] and Cammarano et. al[94] for concise reviews of selected VIs .
With such a variety of VIs to choose from it can prove challenging to choose the one that befits the study purpose the most. This study utilized a VI for which especially long time series of monthly data is available: the Normalized Difference Vegetation Index (NDVI).

2.2.2 Explanatory Variables

Once an appropriate set of response variables for biome assignment has been chosen, the observed biome distributions can be linked to a set of explanatory variables. Establishing an understanding of the interactions of response and explanatory variables will yield a more refined grasp on how certain biotic and abiotic factors influence the dispersal and establishment as well as the maintenance of a given biome thus answering the 2nd major research question (*How are the observed alternative stable states related to the climate?*).
A multitude of parameters have been proposed to control biome distribution (see table 2.1) and there has even been the call for biophysical control variables for each biome[66] to be identified. Recent studies have pointed out that the effects of climate extremes may prove more useful for modelling biome distribution than mean indices[67,95]. This study, however, employed mean annual climate indices since indices of climate extremes were not readily available for the entirety of the time period for which VI data was amassed.

Table 2.1: Biome distribution driving parameters: A variety of parameters have been proposed to determine the location of biomes. Some of the most commonly used have been compiled here.

Type of Parameter	Association	Parameter
Climate effects	Temperature	temperature[43,46,96]
		temperature seasonality[97,98]
		growing degree days[43]
		maximum temperature of the warmest month[38]
		minimum temperature of the coldest month[38]
		isothermality[38,95]
		mean diurnal range[38]
	Precipitation	precipitation[29,46,81]
		precipitation seasonality[41,98]
		water availability/moisture indices[7,47]
	Miscellaneous	fire regime[47,67]
		cloud cover[7]
Location effects		soil characteristics[71]
		elevation[33]

When choosing a set of explanatory variables it is key to identify (1) the explanatory variables which do influence the biome distribution, (2) the critical time windows which affect the biome distribution and (3) the aggregate statistics (mean, maximum, minimum, etc.) which best explain the observed biome distribution[38].
Finally, to describe the relation of the distribution of alternative stable states to climate, the explanatory and response variables can be subjected to a modelling procedure that further specifies their interactions.

2.3 Quantifying Resilience

Resilience is most useful when quantifiable because measurable and testable metrics are required in defining objectives, monitoring change, and evaluating management actions as part of DSSs[3]. Multiple ways of quantifying resilience have been proposed which range from analysing indicators of resilience[37] to assessing attributes that make systems resilient[15] as well as measuring the distance of the current system state to the closest bifurcation point on the stability landscape[16]. Such quantification of resilience answers the third and final major research question of this study (*How are the results of the ecosystem classification and vegetation-climate modelling linked to resilience?*).
This study employs the concept of specified resilience (resilience of *what to what*) by Carpenter et. al[19] to answer its third major research question.

The models that define the relations of distribution of stable states and their driving forces can be used to analyse the probability of each stable state occurring, depending on its driving processes. Provided that the models depict actual relationships authentically, the probability of occurrence of a stable state is synonymous with its resilience. The more favourable the environmental characteristics are for the presence of a given attractor, the more resilient the local occurrence of that state will be. The statistical relationships of the occurrence of each stable state (*of what*) and hypothesized influencing factors (*to what*) can consequently be analysed for the resilience of each individual stable state. The higher the probability of occurrence, the higher the resilience of the stable state in question.

Furthermore, the probability of occurrence can not only be used as an indicator of resilience of individual stable states but can also serve as an indicator of resilience of the entire focal system by assessing the possibilities of multiple alternative stable states occurring in the same location under the same circumstances. The current framework of this study incorporates *three approaches* to identify low resilience in empirical analyses and modelling approaches. These are based on the three following assumptions:

1. *Low Resilience manifests in mixed environments.*
   Ecological resilience thinking discloses that resilience will be the highest for an individual alternative stable state of a focal system were only this particular regime comes to dominate. This corresponds to valleys within stability landscapes (figure 1.1). Areas of a focal system which are highly resilient can thus be postulated to show a near homogeneous spatial display of a single alternative stable state. Patchy areas, on the other hand, in which many alternative stable states are present are therefore indicative of low resilience of the focal system. Such areas often coincide with ecotones although not being confined to them. Ecotones are transition areas between ecosystems[75].

2. *Low Resilience is accompanied by fluctuations in state dominance.*
   Regime shifts, which lead to a change in state dominance, are the most prone to occur when a system is close to a bifurcation point on the stability landscape (F_1 and F_2 in the bottom plane of figure 1.2b). Episodes of frequent regime shifts back and forth between multiple adjacent alternative stable states (valleys in the top transects in figure 1.2b) are thus indicative of either (1) low ecological resilience of the alternative stable states in question or (2) the focal system being held constantly in states which are close to a bifurcation point (for example due to anthropogenic actions[12]) and thus at low resilience levels so that low-impact fluctuations of the environment can trigger a regime shift. General trends among driving forces of alternative stable states will manifest in directional regime shifts (shifts from one specific state into another but not in the opposite direction).

3. *Low Resilience allows for multiple alternative stable states to be present under the same circumstances.*
   Ecological resilience theory postulates that the resilience of every system is greatest at the point were one stable state is dominant (in the attractor valley of figure 1.1). As the system nears a bifurcation point and resilience of the current attractor regime is failing, the chances of alternative stable states manifesting within the system increase. The possibility of multiple alternative stable states occurring within the same conditions can thus be understood as indicative of overlap in attractor regimes and consequently low resilience of the focal system with high potential for fluctuations in stable states dominance.

3. Study Sites

The spatial extent of study sites for testing the framework for remote sensing resilience assessment was limited to regions which were suitable for remote sensing studies and of key interest in the face of climate change and anthropogenic impact.

3.1 Minnesota

Minnesota is a region of the United States of America situated at the border to Canada and between latitudes 43°N and 50°N and longitudes 89°W and 97°W thus covering an area of over 225,000 km^2. Despite the relatively small extent when compared to other study sites, Minnesota has been selected as a favourable region for evaluating the biome distributions yielded by the initial clustering process of spectral and climate data for multiple reasons:

1. **A Variety of biomes are present.** A multitude of different vegetation assemblages has been reported to cover Minnesota[47,99] thus allowing for the analysis not only of biome distributions but guaranteeing to include transition zones between biomes. Since a systems resilience is proposed to be lower the closer the system is to a regime shift, the transition zones (also known as ecotones[47]), representative of regime shifts, should be examined to answer the third major research question of this study (*How are the results of the ecosystem classification and vegetation-climate modelling linked to resilience?*).

2. **Established LCCS are available.** Multiple LCCS have been established for Minnesota spanning various time frames[99–101]. The results of these studies can be used for reference in biome identification thus resolving the first major research question *What and how many states are present in a given focal system and how do their distributions change over time?*.

3. **Climate data is very precise.** Due to the high density of weather stations in Minnesota (see figure 4.2), the climate data has undergone less interpolation than in other areas and is thus very precise. Exact data allows for more precise modelling of vegetation-climate interactions thus answering the second major research question (*How are the observed alternative stable states related to the climate?*).

4. **Ecological Relevance.** Anthropogenic impact has a strong influence on biome distribution in Minnesota[99] due to intense agricultural land use and large urban areas. This is a confounding variable in the vegetation-climate interaction modelling which is particularly difficult to account for. It is, however, this intense land use and coverage of Minnesota by production systems that demands detailed information on the resilience potential of natural and agricultural systems to avoid a collapse of either.

3.2 Alaska

Alaska is the northern most state of the United States of America located between latitudes 52°N and 72°N and longitudes 130°W and 170°W thus covering a land area of over 1,000,000 km^2. Despite the logistic complications in high latitudes, which are effective even for remote sensing studies (scarce distribution of weather stations for example, see figure 4.2), it makes for an excellent remote sensing study region for several reasons:

1. **Large extent.** The immense land area of Alaska ensures to include oceanic climate effects as well as continental climate patterns and thus a variety of functional biomes.

2. **Topographical variety.** The terrain of Alaska includes mountain chains, coasts, large river sheds and extensive planes all of which together prompt a diverse assembly of potential biomes.

3. **Low anthropogenic impact.** Alaska is, due to climate conditions, not an agricultural state and is not home to major cities or settlements which would otherwise have to be masked in the data sets[60]. However, several regions have undergone severe deforestation[102].

4. **Established LCCS available.** Due to the aforementioned factors, Alaska makes for a great study site in biome classifications and has been used as such in several studies[43,46]. The results of these studies can be used for reference in biome identification.

5. **Ecological Relevance.** The transition zone between boreal forest and tundra ecosystems, which is present in Alaska[46], has been proposed as a subject to immense change in biome distribution[60,77,103]. Especially since warming rates twice the global average have been reported for the boreal region[46]. Conclusively, climate change has been identified as an imminent potential for alteration of ecosystems in northern regions like Alaska thus making studies on such regions important sources of information on ecosystem functioning under changing conditions.

4. Methodology

4.1 Normalized Difference Vegetation Index

The NDVI has been selected as a primary response variable for biome identification in this study because its target variables are biomass and vegetation cover (indicators of vegetation composition), long time series are available at decent spatial resolutions and due to its demonstrated utility in various ecosystem studies[39].

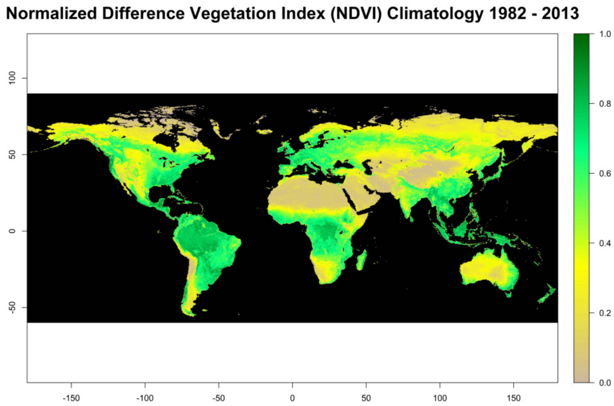
Rouse et. al[82] first introduced the NDVI in 1974 and initially coined it 'Vegetation Index'. The name was subsequently changed to NDVI when other VIs were proposed in the ensuing years and advent of remote sensing studies which relied on VIs[33,80,104].
The NDVI is a composite VI that factors in measurements 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[79]. The calculation formula is as follows[42,79]:

$$NDVI = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}}$$

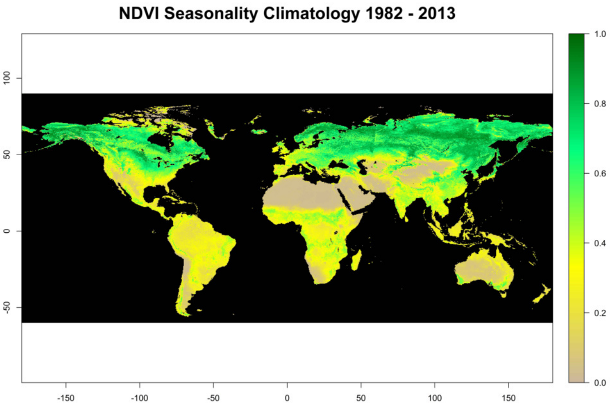
With ρ_{NIR} being the reflection in the NIR band and ρ_{RED} being the reflection in the RED band. The NDVI is bound between 0 and 1 with higher values representing denser vegetation. For a more detailed mathematical background on the calculation of NDVI scores see Matsushita et. al[104].

NDVI data sets are readily available via the Global Inventory Modelling and Mapping Studies (GIMMS). For this study the GIMMS3g data set was used (available at `https://ecocast.arc.nasa.gov/data/pub/gimms/3g.v0/`), which provides Advanced Very High Resolution Radiometer (AVHRR) NDVI data as 15-day maximum-value composites from 1982 to 2013 at 0.083° resolution[60]. The data was processed in R[105], downloaded and compiled into monthly and annual composites using the raster[106] and gimms-package[107]. See A.6.1 for a code that can be run to install all packages needed to reproduce the analysis of this study from a base installation of R.
Annual NDVI seasonality has been calculated as the difference between maximum monthly NDVI score and minimum monthly NDVI score throughout a calendar year.
Averaging the NDVI data over a certain time period to avoid vegetation extremes[61] results in a climatology of said time frame. Relevant for this study are climatologies of NDVI and NDVI

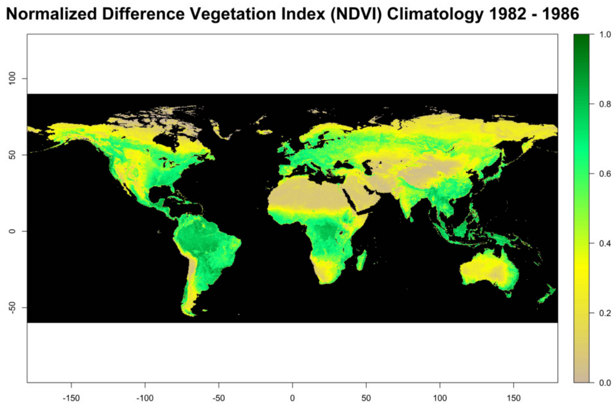
seasonality from 1982-2013, 1982-1986 and 2009-2013 (figure 4.1). These climatologies serve to answer all three major research questions of this study in that they are used in an initial clustering and characterisation of biomes. See A.6.2 for the R-code used to process the data.



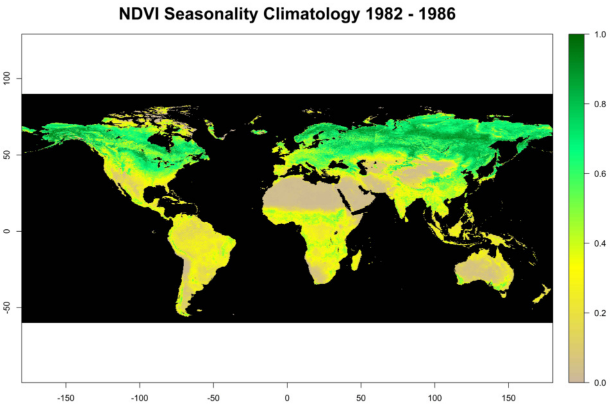
(a) NDVI 1982-2013



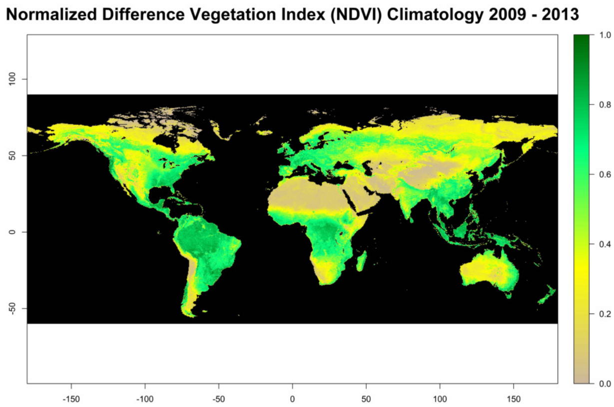
(b) NDVI Seasonality 1982-2013



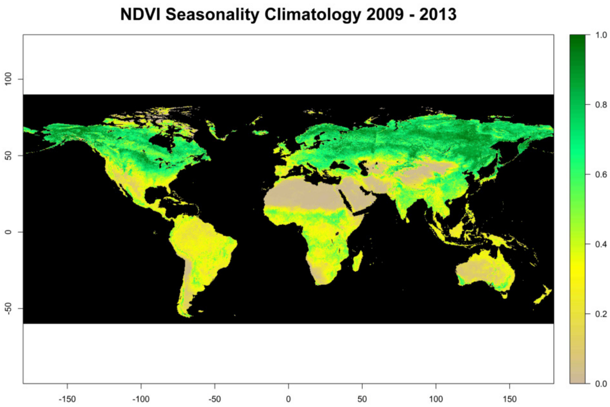
(c) NDVI 1982-1986



(d) NDVI Seasonality 1982-1986



(e) NDVI 2009-2013

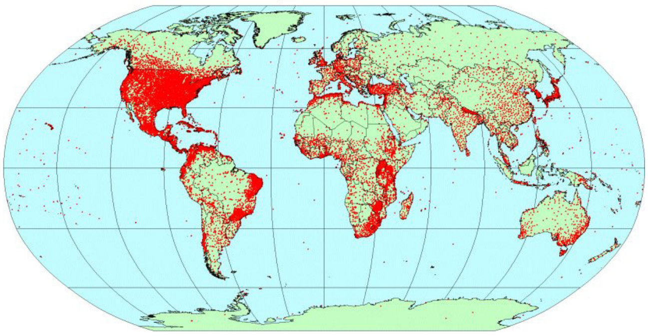


(f) NDVI Seasonality 2009-2013

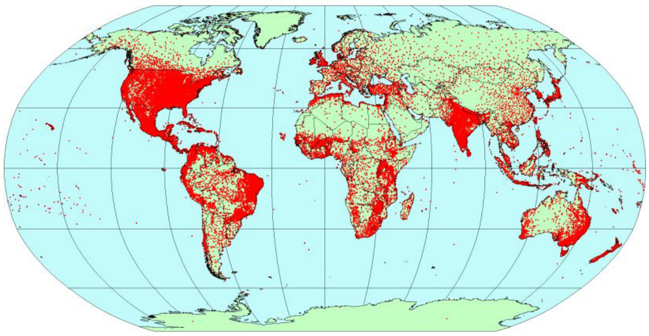
Figure 4.1: NDVI Climatologies: Average composites of mean annual NDVI scores and annual NDVI seasonality scores have been calculated using the R-project for statistical computing.

4.2 Climate Data

Climate parameters were attained using the WorldClim Version 1.4 products for current climate conditions, which can be downloaded at `http://www.worldclim.org/current` and have been employed in this study as explanatory variables.
WorldClim is a set of global gridded climate data at a spatial resolution of about 1 km^2 [108]. The WorldClim data layers have been produced by interpolation of observed average monthly climate data on a 30-arc resolution. Thus the local preciseness of the WorldClim parameters is strongly correlated with the abundance of climate stations across the globe. This can lead to high uncertainty of interpolated records, especially in mountainous and poorly sampled areas[108]. This phenomenon, however, is not limited to the WorldClim data sets but also diminishes the accuracy of other climate data sets such as the Climatic Research Unit (CRU) data sets.
For a representation of the global distribution of weather stations whose records were used in the calculation of current (1960-1990) WorldClim climate products see figure 4.2. Notice especially the sparse availability of weather station records in northern latitudes (Alaska, Siberia).



(a) Mean Temperature Records



(b) Precipitation Records

Figure 4.2: WorldClim Weather Stations. [108] WorldClim data sets have been calculated by interpolation of data from the above presented weather stations. Weather stations are depicted as red dots.

WorldClim data has been used frequently in recent studies[7,109] and, for lack of more precise global climate data sets which are built using data from more weather stations (especially in regions where such are sparsely distributed), this study incorporated two WorldClim variables for the modelling of the climate-vegetation interactions.

The data was downloaded as gridded data sets of a 2.5 minute resolution which have been derived from the original 30 second resolution by aggregation of cells. The data sets were then processed in R using the raster-package and re-sampled to a 0.083° (to match that of the NDVI data) resolution using the *resample()* function provided by the raster-package. See A.6.3 for the R-code used to process the data.

4.2.1 Temperature

Temperature related climate indices have been suggested to drive biome distribution (see table 2.1). Especially mean annual temperature has been identified as a major driving force of biome distribution and functioning. It has been employed extensively as a metric in remote sensing resilience studies, regime shift studies and vegetation modelling[35,57,64,67,80,96].

For this study, it has been calculated as a climatology of WorldClim data spanning 30 years worth of temperature records (1960-1990). The climatology is presented in figure 4.3. It functions as one of the explanatory variables in this study for climate-vegetation modelling and subsequent assessment of resilience.

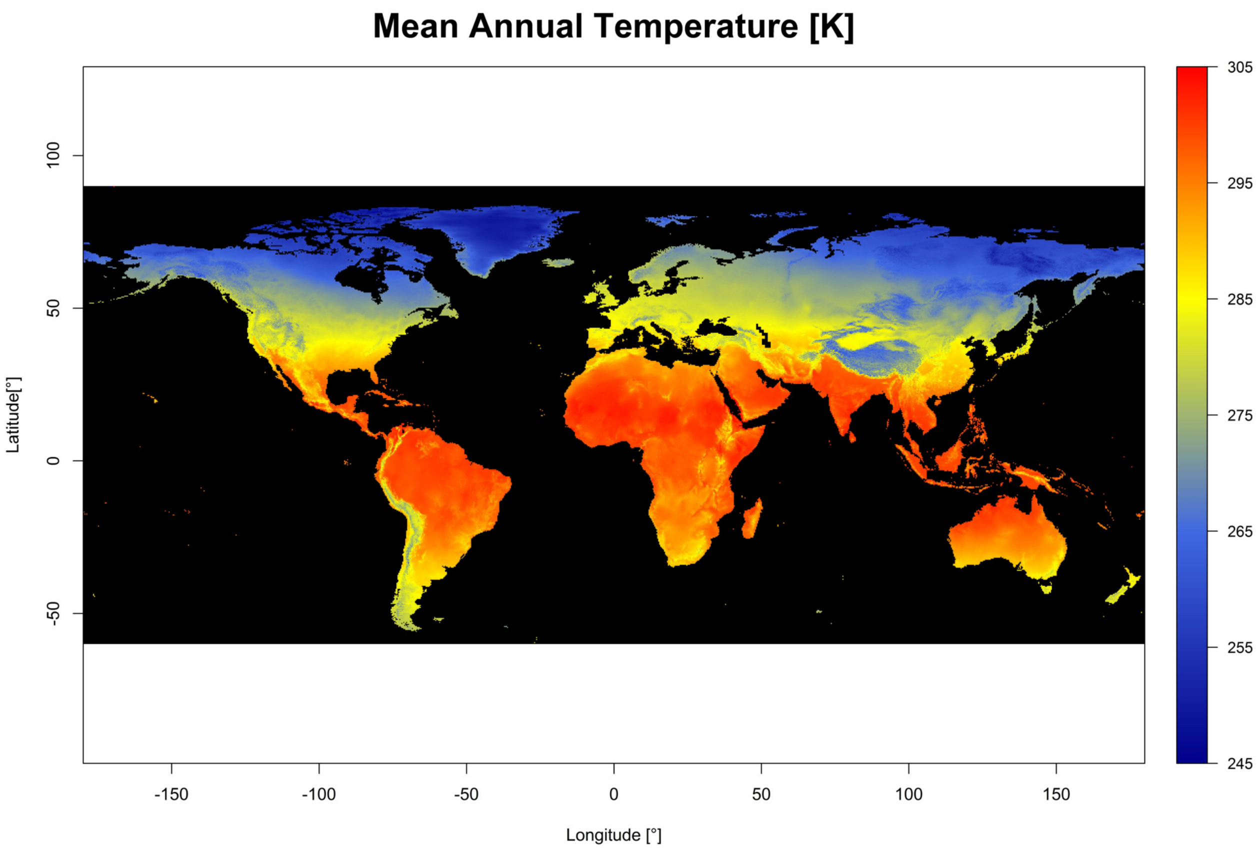


Figure 4.3: Climatology - Mean Annual Temperature: The annual of mean of temperature data per cell has been calculated using the raster-package in R.

4.2.2 Precipitation

Just as temperature related climate indices, precipitation based climate indices have been suggested to affect biome distribution (see table 2.1). Therefore, to explain biome distribution and accurately model climate-vegetation interactions, it is important to include a precipitation metric into biome classification studies[81,95]. Remote sensing studies identified mean annual precipitation as a threshold for biome distributions[49] and functioning of certain ecosystems such as tropical rainforests[35].

This study includes mean annual precipitation (see figure 4.4) for the purpose of serving as an explanatory variable in biome-climate modelling and subsequent resilience assessment.

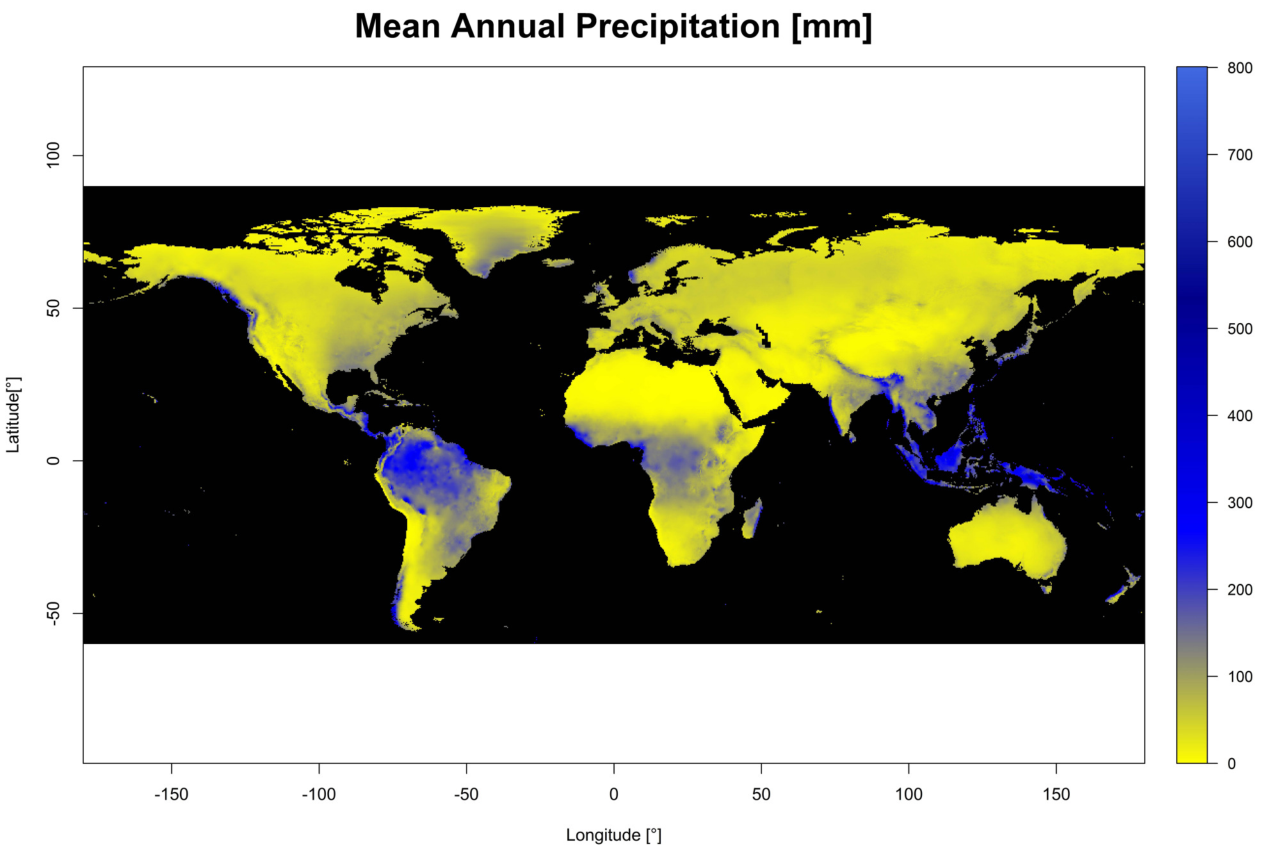


Figure 4.4: Climatology - Mean Annual Precipitation: The annual of mean of precipitation data per cell has been calculated using the raster-package in R.

4.3 Digital Elevation Model

Climate effects are not the only drivers of biome distribution. There are confounding variables such as soil properties[71] and elevation effects[33] (only to name a few).
To account for this, the modelling of climate-vegetation interactions in this study included data derived from a Digital Elevation Model (DEM). The DEM employed was the mean product of the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) which has been developed by the U.S. Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA) and is available for download at `https://lta.cr.usgs.gov/GMTED2010`.

The GMTED2010 has been downloaded as gridded data sets at a spatial resolution of 30-arc-seconds, processed in R using the raster-package and re-sampled to the resolution of 0.083° using the *resample()* function provided by the raster-package (see figure 4.5). Due to the re-sampling from a finer to a coarser spatial resolution, elevation records of the highest summits have been lost, which does not have severe implications due to vegetation being unlikely to occur at altitudes higher than 6,000 m[110]. See A.6.4 for the R-code used to process the data.

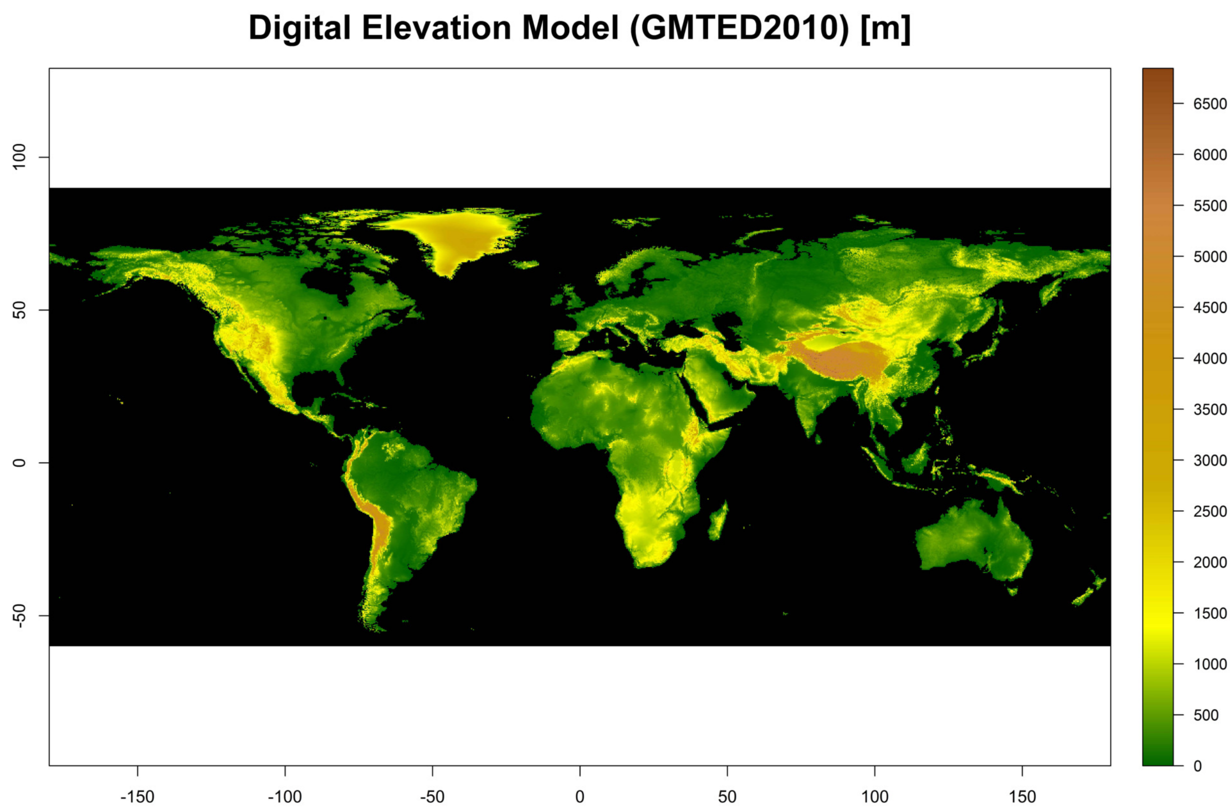


Figure 4.5: Digital Elevation Model: A DEM contains altitude data. This study incorporates the use of the GMTED2010 mean elevation product.

4.4 Statistical Analysis

For the purpose of analysing how climate and elevation influence biome distribution, this study employed general additive modelling[111,112] reliant on the results of a data driven clustering algorithm. The clustering was used to delineate biome types based solely on vegetation characteristics (*1st objective*). The resulting biome classification was subsequently incorporated into Generalised Additive Models (GAMs) using climate and altitude data as explanatory variables to assess biome functioning (*2nd objective*). Resilience of delineated stable states was assessed (*3rd objective*) using clustering uncertainties, biome contingencies and potential distributions of biomes.

The analysis was performed in R after the data had been spatially cropped to the extent and shape of the study sites and masked for lake and urban areas using a set of shapefiles which can be downloaded at `http://www.naturalearthdata.com/downloads/10m-cultural-vectors/` and `http://www.naturalearthdata.com/downloads/10m-physical-vectors/`. The cropped and masked data has been included for visualisation in A.3.

Secondly, response and explanatory variables were checked for collinearity. Collinearity is the term employed by statisticians to describe a linear relationship between multiple explanatory variables. If collinearity occurs, the assumption of independence (variables are not influencing one another) is violated which, in turn, entails the necessity for further parameter selection[113]. See A.6.5 for the R-code used to assess collinearity and A.4 for the collinearity scores of study parameters of each study site.

4.4.1 The Clustering Procedure

4.4.1.1 The mclust Method

The data based clustering process was carried out in R using the mclust-package[114]. Clusters can be thought of as groupings of data in multi-dimensional space. The number of dimensions is equal to the number of clustering components. In mclust, the characteristics of these clusters (orientation, volume, shape) are, if not specified otherwise, estimated from the data. They can be set to vary between clusters or constrained to be the same for all clusters[115]. Depending on cluster characteristics, mclust distinguishes 20 individual models[114] which have been compiled for comparison in table 4.1.

Functions provided by the mclust-package employ mixed effect modelling for the underlying clustering procedure. Mixed effect models combine fixed and random effects of the data in a single model[111] which reduces the number of regression parameters to be estimated. This increases the degrees of freedom and can possibly make for a better fitting model than either fixed or random effect models[111]. For a detailed mathematical explanation on how mclust works see Fraley et. al[115].

Table 4.1: Models in mclust: The r-package mclust distinguishes 20 different models for data clustering based on distribution and cluster characteristics.

Acronym	Distribution	Volume	Shape	Orientation
E	univariate	equal	-	-
V	univariate	variable	-	-
EII	spherical	equal	equal	NA
VII	spherical	variable	equal	NA
EEI	diagonal	equal	equal	coordinate axes
VEI	diagonal	variable	equal	coordinate axes
EVI	diagonal	equal	variable	coordinate axes
VVI	diagonal	variable	variable	coordinate axes
EEE	ellipsodial	equal	equal	equal
EVE	ellipsodial	equal	variable	equal
VEE	ellipsodial	variable	equal	equal
VVE	ellipsodial	variable	variable	equal
EEV	ellipsodial	equal	equal	variable
VEV	ellipsodial	variable	equal	variable
EEV	ellipsodial	equal	variable	variable
VVV	ellipsodial	variable	variable	variable
X	univariate normal			
XII	spherical multivariate normal			
XXI	diagonal multivariate normal			
XXX	ellipsoidal multivariate normal			

Mclust provides the user with a very autonomous process of model calculation and selection. First, if not specified otherwise, mclust calculates all available models for a range of cluster component numbers (by default one to nine clusters). As an example: for a clustering of data with four individual variables, mclust will, by default, calculate 126 individual models (14 model classes $*$ 9 cluster possibilities). It will calculate models from only 14 classes, since E, V, X, XII, XXI and XXX models are only appropriate for single variable clustering[114,115].

Secondly, once the models are established, mclust selects the most appropriate of the models according to their respective Bayesian Information Criterion (BIC) value. Like the Akaike Information Criterion (AIC), the BIC is an indicator of model quality: the lower the BIC, the better the model fits the data[111]. Conclusively, mclust chooses the model with the lowest BIC available for clustering the data.

4.4.1.2 Biome Characterisation

The biome distribution maps used for the vegetation-climate modelling were produced using spectral data (mean NDVI and NDVI seasonality). See A.6.6 for the R-code used to process the data and produce biome distribution maps.
In accordance with established LCCS for Minnesota and Alaska, the number of clusters to be identified for both regions in this clustering step were set to 4[99,100] and 5[46] respectively.
Mclust calculations on data sets from Minnesota were carried out using the entirety of these data sets. For calculations on data sets from Alaska, 20,000 data points were randomly sampled from the total set of data and used in the initial model calculation and selection steps. The model was then exerted on the whole data set using the *predict.Mclust()* function provided by the mclust-package resulting in classifications of the entire data sets.

4.4.2 Vegetation-Climate Modelling

4.4.2.1 Generalised Additive Modelling

Response and explanatory variables do not always exhibit clear linear relationships and Generalised Linear Models (GLMs) did perform poorly when applied to the data sets of this study. In such cases it is advisable to either (1) transform the data, (2) add interaction terms among explanatory variables to the model or (3) chose a Generalised Additive Model (GAM) approach[111]. GAMs use smoothing curves to model the relationship of explanatory and response variables which allow for non-linear relationships[111,113] and have been exerted on the data sets.

To model the vegetation-climate interactions, thus answering the second major research question of this study, the biome distributions resulting from the mclust process were transformed into binary distribution maps for each stable state effectively masking the spectral data used by mclust with a 0,1-representation of absence and presence of a particular vegetation composition. These presence/absence data sets were subsequently treated as response variables and modelled against explanatory variables using the *gam()* function as provided by the mgcv-package[112]. The climate parameters employed as explanatory variables were mean annual precipitation as proposed by Hirota et. al[29] and mean annual temperature which was proposed as a major driver of biome distribution by Woodward et. al[63] and Verbesselt et. al[35]. Altitude data was included utilising the GMTED2010 to account for the effects of higher altitudes and consequently lower temperatures on alternative stable states[33].

4.4.2.2 Assessing Vegetation-Climate Interactions

The vegetation-climate interactions were investigated by plotting response curves for each delineated biome that resulted from the established GAMs. Response functions depict the relationship between the response variable and the explanatory variables. To tease apart which explanatory variable did influence the distribution of a particular stable state, the response function plots do not display the relation of presence and absence of a stable state to all three explanatory variables (which would result in a four-dimensional plot!) but to one explanatory variable at a time. This has been achieved by setting the two explanatory variables that were not assessed to their mean values across the respective study region and formulating the response function on the resulting data frame in which only the binary presence indicator and investigated explanatory variable were allowed to vary. Furthermore, the response functions were calculated to depict not presence/absence but the probability of assigning presence of a stable state to a given level of the explanatory variable. The resulting response functions were additionally treated with a loess smoother to outline the vegetation-climate interactions at play more clearly. The smoother was set to a span of 0.1 to avoid imprecise averaging. See A.6.6 for the R-code used to produce the response functions and A.5 for GAM visualisations including response functions.

4.4.3 Assessing Resilience

In accordance with the three assumptions of the manifestation of low resilience in empirical analyses and modelling approaches described within the conceptual framework of this study, resilience assessments have been carried out using three individual approaches:

1. *Clustering confidence.*
   Low assignment confidence resulting from the mclust algorithm is indicative of patchy environments which coincide with ecotones but are not limited to them.

2. *Biome contingencies.*
   Loss of spatial coverage over time is an indicator of (directional) regime shifts which can be hypothesized to be caused by general trends among driving forces.

3. *Potential for multiple alternative stable states.*
   Potential for multiple alternative stable states to occur in the same location given the current climate regimes have been assessed using GAM predictions and the Receiver Operating Characteristic (ROC) method.

4.4.3.1 Clustering Confidence

Ecological theory determines that ecosystems, as they approach tipping points from one alternative stable state into the domain of a rival alternative stable state, experience a loss of resilience and characteristics distinctive of that particular state. In terms of figure 1.1 the y-axis can be said to depict the inverse of the resilience the focal system exhibits. The closer a system is to any given attractor, the lower the y-score of the system at that point when compared to the y-scores at the two closest bifurcation points.

When concerned with clustering algorithms, resilience can be expressed as the probability of a given data set being associated with a given cluster. In a map of spectral data, for example, every pixel holds a certain value. If the pixels of this map have to be sorted into a given number of clusters every pixel will, aside from being classified into one of the available clusters, also

receive a probability of assignment to each cluster. The higher the probability for any pixel to be assigned to its most likely cluster, the more closely the data of this pixel depicts the data expectation that is predicted for the particular cluster membership. In practical terms this can be re-phrased as follows: If a pixel is a 100% match with the 'boreal forest' cluster the corresponding area on the globe is most likely to be a fully developed boreal forest that is intact and thus fairly resilient. If, however, a pixel is only a partial match with the 'boreal forest'-cluster but no match with any other cluster is any better, the corresponding area is presumed to be vegetated by a boreal forest which is less resilient than the one in the 100%-matching pixel. This may be due to the worse matching pixel being located on the border between the boreal forest region and one of the alternative stable state regions (for example: tundra[46]) or due to large clearings in the boreal forest caused by large scale fires or intense logging.
The third major research question of this study was consequently tackled by analysing the maps of assignment confidence produced by the mclust- and raster-functions.

4.4.3.2 Biome Contingencies

Using the mclust procedure described above, NDVI data was especially useful as a primary response variable due to the long time series available for it (1982-2013) allowing for the calculation of multiple composites (climatologies) and thus for the assessment of vegetation changes whilst ruling out the influence of rapid year-to-year fluctuations[61].
First, an mclust classification model was established on NDVI composite data from 1982-2013. Secondly, the resulting classification scheme was exerted on composites from 1982-1986 and 2009-2013 thus delineating differences in proportion and distribution of stable states as time progressed and allowing for calculations of state specific resilience scores as well as directional regime shifts (which state turned into which other state). See A.6.6 for the R-code used to produce biome distribution maps for three different time frames and calculating a matrix containing the relative changes from one states areas into being classified by other states.

4.4.3.3 Potential For Multiple Alternative Stable States

The assumption of alternative stable states embedded in ecological resilience thinking discloses that the resilience of any given set of alternative stable states on a spatial scale will be the highest for every one of these states were only this particular regime comes to dominate. The resilience of alternative stable states in a spatial distribution can thus be thought of as circular layers revolving around a central area with the highest resilience and dominance of a given state and incrementally decreasing resilience scores with increasing distance from this point of highest resilience. Conclusively, the sum of individual resilience scores of all alternative stable states of a system will be lowest where the maximum amount of alternatives overlap in their spatial extent thus resulting in areas of ecological uncertainty and potentially frequent change of dominant alternative stable state.
Regarding modelling procedures, low resilience of a system can be seen as the potential for

multiple alternative stable states of the same system to be present at the same location at the same time. To investigate this and answer the third major research question of this study, the established GAMs were analysed using the Area Under Curve (AUC) method. This method is currently considered to be the standard in assessing the accuracy of predictive models. It eliminates the subjectiveness of choosing a threshold when transforming probability of assignment scores into presence/absence data[116]. This is done by building a ROC curve for the model performance. In this curve, the rate of false negatives (also referred to as *1-specificity*) and the rate of true positives (also known as *sensitivity*) are depicted on the coordinate axes. The optimal probability cut-off above which to presume presence of a stable state is selected by finding that assignment probability which maximizes sensitivity and specificity. The AUC of the ROC curve is furthermore indicative of how well the model performs overall with higher scores representing better accuracy and scores between 0.7 and 0.8 being acceptable fitting models[33,117].

To obtain a visual depiction of where multiple alternative stable states overlap in their predicted distributions, the assignment probabilities of each pixel to each identified stable state have been turned into binary presence indicators using the probability threshold selected by the ROC method obtained through the *ROC()* function provided by the Epi-package[118].

Finally, the sum of these predicted binary distribution sets have been calculated for each pixel and plotted resulting in a map which depicts the number of alternative stable states predicted for each pixel. See A.6.6 for the R-code used to process the data and produce the ROC predictions.

5. Results

5.1 Minnesota

5.1.1 Initial Biome Delineation

Using NDVI and NDVI seasonality composites for the time period of 1982-2013 as data for the mclust-algorithm, four individual clusters of spectral data were identified in Minnesota. The cluster means around which the data of each stable state have been formed to answer the first major research question (*What and how many alternative stable states are present in a given focal system and how do their distributions change over time?*) have been compiled in table 5.1. Whilst stable state 4 is largely different from the remaining three data clusters through its high mean annual NDVI score of 0.592, teasing apart the remaining clusters by their respective mean annual NDVI data is difficult. NDVI seasonality, on the other hand, is a practical metric for delineating the remaining data into stable states 1, 2 and 3. Stable state 1 is characterized by the lowest NDVI seasonality level of the data set (0.529) and the cluster means of NDVI seasonality data of stable states 2 and 3 are separated by a difference in seasonality scores of 0.136 with data belonging to stable state 2 exhibiting the highest seasonality of monthly NDVI scores (0.891) in the entire spectral data set for Minnesota.

Table 5.1: Mclust Model for vegetation-climate modelling (Minnesota). Clusters have been formed around mean values of the data parameters. See below for model and cluster characteristics.

	Cluster Means	
Stable State	NDVI	NDVI Seasonality
1	0.48	0.529
2	0.459	0.891
3	0.469	0.755
4	0.592	0.674
Model name:	VVE	
BIC:	19611.64	

The 4 two-dimensional clusters of data from Minnesota have been visualized in figure 5.1 as an uncertainty plot. This plot depicts the data colourized in accordance to their cluster group membership (can be found in figure 5.2) with the size of individual dots representing the uncertainty assigned to them in their respective cluster ascription.

Figure 5.1 and table 5.1 show that the model established by mclust consisted of two-dimensional data clusters of variable volume, variable shape and equal orientation (VVE).

In accordance with table 5.1, figure 5.1 depicts stable state 4 as a cluster of the highest mean annual NDVI data situated at intermediate NDVI seasonality levels. Stable states 1, 2 and 3 are located at intermediate mean annual NDVI values but distinguished by high NDVI seasonality levels of stable state 2, low seasonality scores of stable state 1 and intermediate seasonality characterising the cluster of stable state 3.

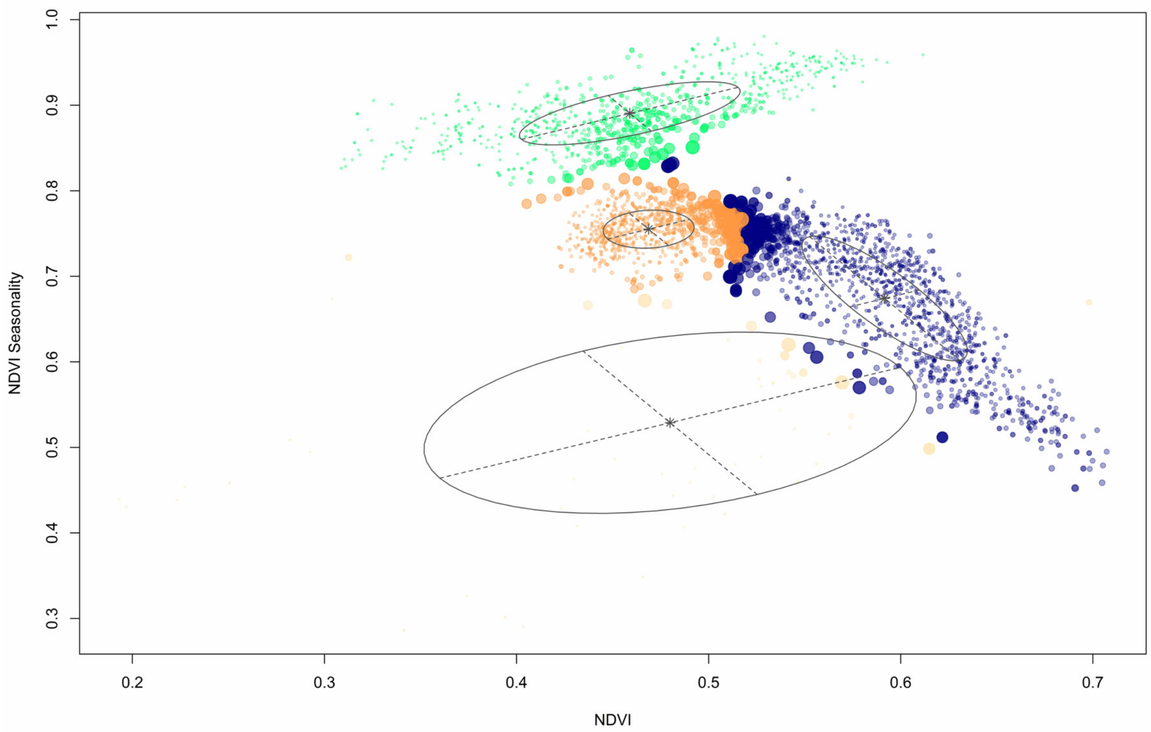
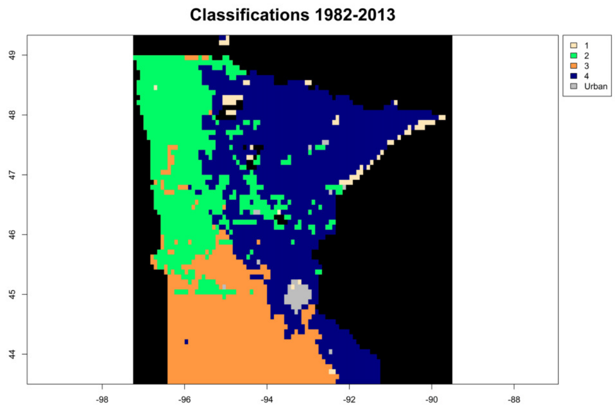


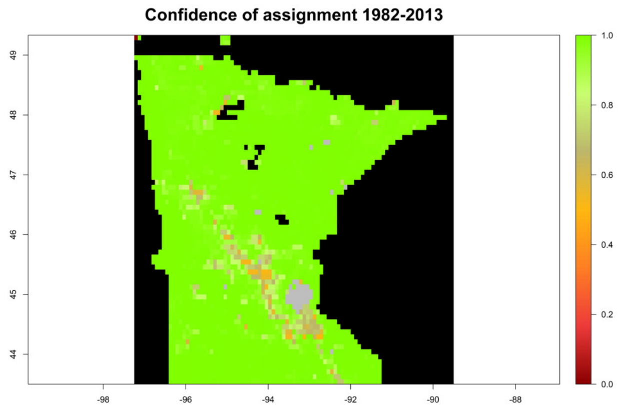
Figure 5.1: Mclust Vegetation Clustering (Minnesota): The data contained within in the identified clusters have been plotted in their respective dimensions as uncertainty plots with larger dots indicating higher uncertainty levels of assignment of the respective data point to its currently assigned cluster. The ellipses correspond to the covariances of the clustering components.

The model built on the full spectral data set (composites of 1982-2013, table 5.1) has been exerted on spectral data composites covering the time frames of 1982-1986 (past) and 2009-2013 (present). The resulting distributions of the clusters identified in table 5.1 and figure 5.2a for the past and present time frames have been mapped out in figure 5.2c and figure 5.2e respectively. Assignment confidences for the classifications of the different time periods have been mapped accordingly in figures 5.2b, 5.2d and 5.2f.

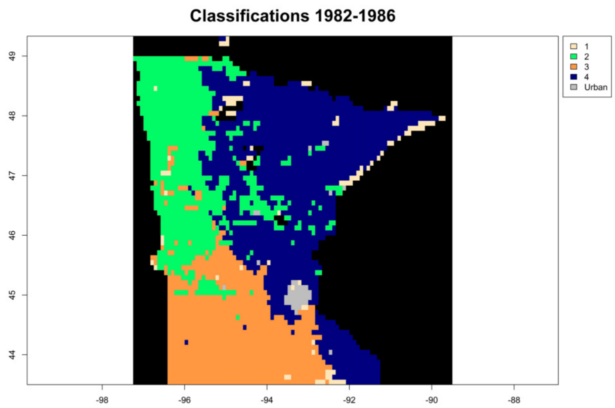
The cluster distributions of the full (figure 5.2a) and past (figure 5.2c) spectral data sets are similar in that cluster 1 is located close to water bodies, cluster 2 is covering the north-western parts of Minnesota, cluster 3 is present in the southern regions and clusters 4 being distributed across the eastern areas of Minnesota. In the spectral data set for 2009-2013 (figure 5.2e), cluster 1 is still confined to the vicinity of water bodies, cluster 4 is still predominant in the eastern regions of Minnesota but the area covered by cluster 2 has increased significantly and taken over large parts of the areas allocated to cluster 3 in the other classification sets.



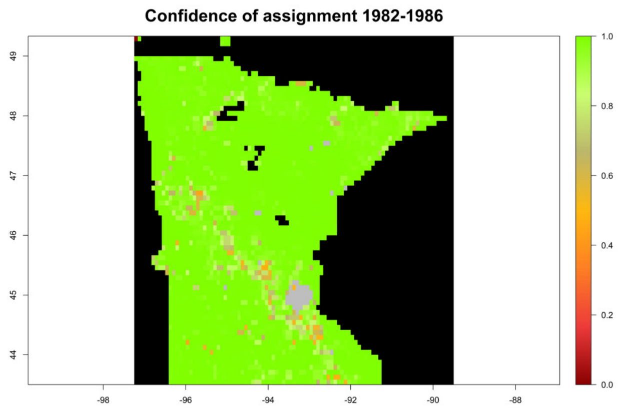
(a) Biome distributions (1982-2013)



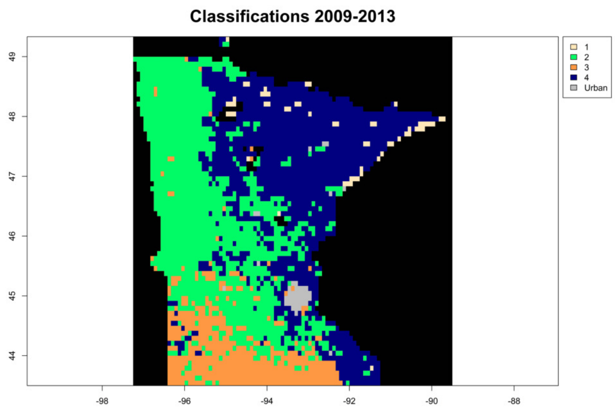
(b) Biome assignment confidence (1982-2013)



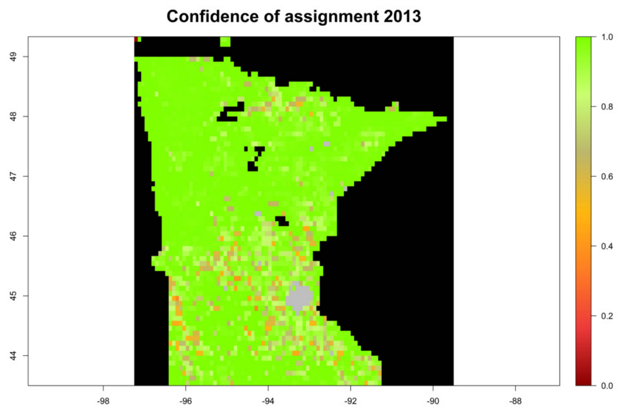
(c) Biome distributions (1982-1986)



(d) Biome assignment confidence (1982-1986)



(e) Biome distributions (2009-2013)



(f) Biome assignment confidence (2009-2013)

Figure 5.2: Biome classifications via mclust (Minnesota): Spectral data sets for 1982-2013 have been clustered using the mclust algorithm. The resulting model has been used to cluster spectral data from 1982-1986 and 2009-2013 respectively.

5.1.2 Biome Distributions And Contingency

Figure 5.3 presents the proportions of alternative stable states of vegetation in Minnesota over the course of 30 years. These proportions are depicted in three stages: full data set (1982-2013), past (1982-1986) and present(2009-2013). The proportions observed in the full data set closely resemble those in the past data set. Obvious change occurred between past and present proportions of alternative stable states, which have been subjected to further assessment. Whilst clusters 1, 3 and 4 shrunk in proportion, cluster 2 massively increased its relative coverage of Minnesota from past to present (25.76% to 42.46% coverage of Minnesota's total land area).

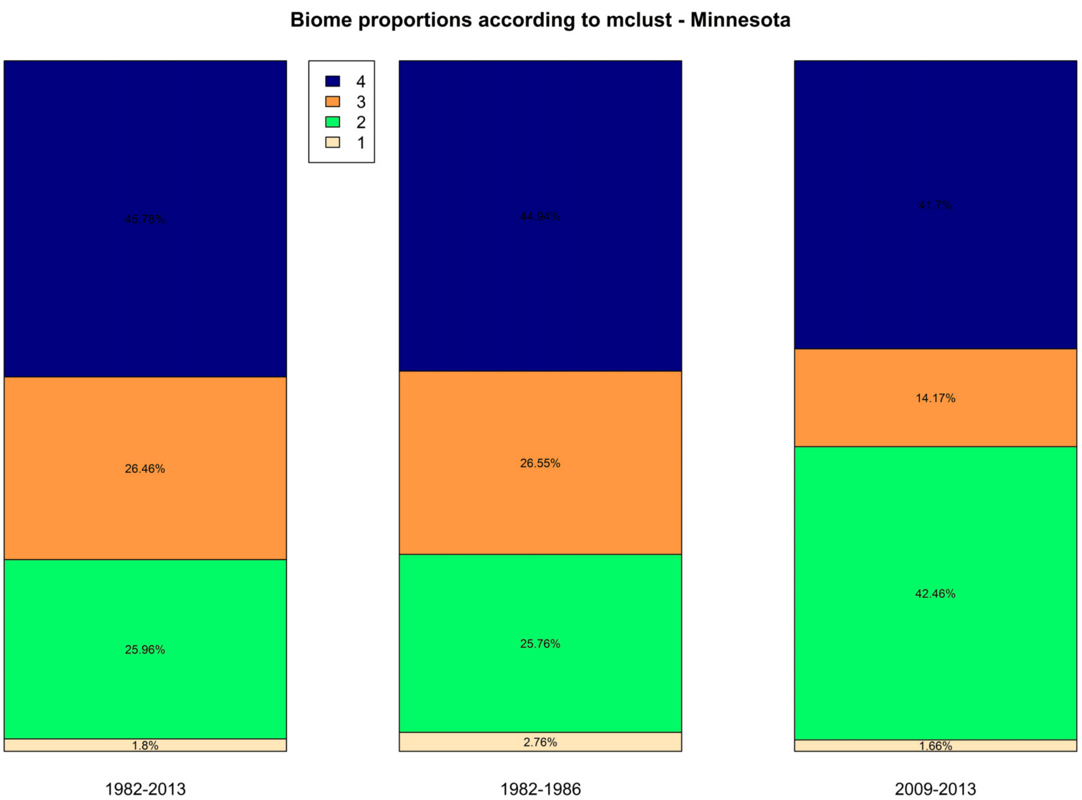


Figure 5.3: Biome Contingency (Minnesota): The change of biome proportions in Minnesota over a time span of 30 years (1982-2013) has been analysed using the mclust method.

Table 5.2 shows the percentages of change in proportion from one state to another relative to the initial proportions in the past data set when contrasted to the proportions in the present data set. Especially remarkable is the fact that stable state 2 lost almost no area and gained nearly half of the area formerly characterised by stable state 3 (41.21%).
Stable state 2 and 4 exhibited stable distributions with 99.89% and 86.05% respectively of their coverage in the past data set remaining unaltered when contrasted with the current data set.

Table 5.2: Biome Contingency (Minnesota): The similarity of stable state distributions obtained by analysing spectral data of 1982-1986 (past) and spectral data of 2009-2013 (present) has been calculated to assess the transitioning of stable states of vegetation in Minnesota. Changes from past to present state have been assessed as percentages relative to the initial proportion of the past state.

	Present State			
Past State	1	2	3	4
1	43.88%	11.22%	25.51%	19.39%
2	0%	99.89%	0.11%	0%
3	0%	41.21%	49.36%	9.43%
4	1%	12.2%	0.75%	86.05%

5.1.3 Presence Of Alternative Stable States

Prediction probabilities of individual stable state GAMs have been treated as presence/absence indicators by applying the respective ROC determined threshold. The resulting binary presence indicators have been mapped for each stable state. Finally, the resulting maps have been projected atop each other resulting in a visual representation of possible distribution of stable states irrespective of the individual biomes but focussing on the possibility of multiple stable states occurring at the same location (figure 5.4).

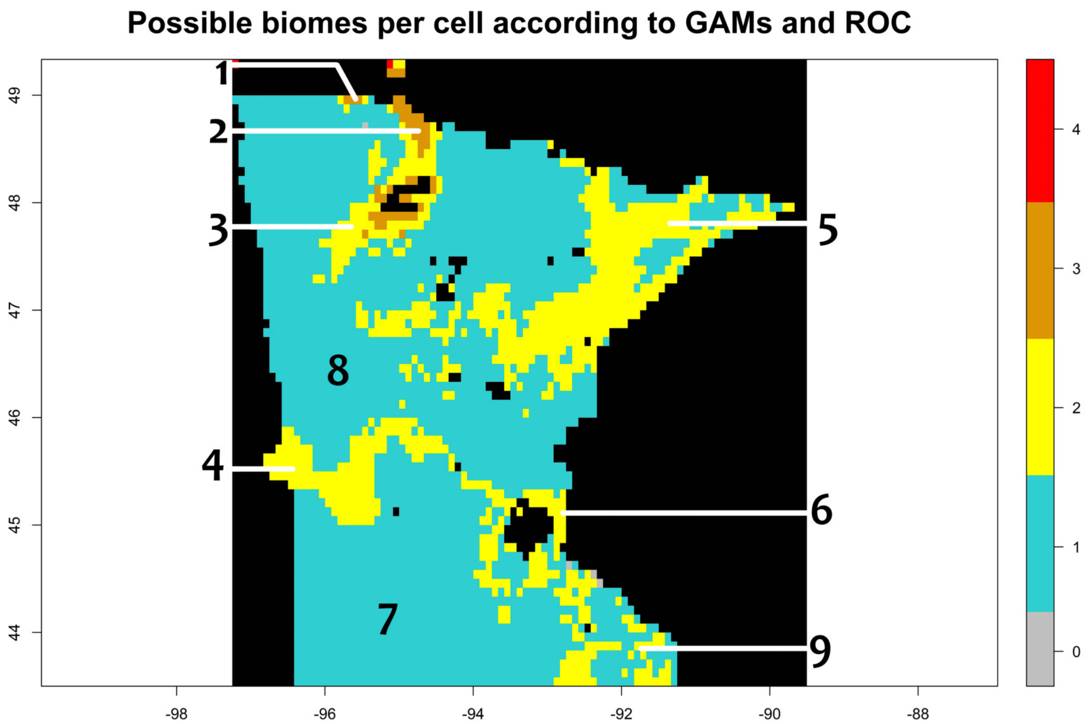


Figure 5.4: ROC prediction of presence of alternative stable states (Minnesota): The ROC thresholding method has been used in combination with GAMs to assess the possibility for multiple stable states to occur in the same place. Numbers have been added for referencing particular areas.

Numbers in square brackets in the following paragraph are used to reference marked areas in figure 5.4.

As figure 5.4 shows, there are few regions in Minnesota for which the potential occurrence of 3 or more alternative stable states has been identified [1,2]. These areas are located mainly in the northern most part of Minnesota around the Red Lakes [2] and Lake of the Woods [1]. Chances for the occurrence of 2 alternative stable states in the same place have been recognized for areas close to large water bodies [3,5], around the Twin Cities (south-east) [6,9] and to the west of the Twin Cities [4]. Whilst most of Minnesota is covered by areas for which at least coverage by one of the individual stable states have been hypothesized [7,8], minute areas for which no cover of any of the stable states could be estimated did occur as well.

5.2 Alaska

5.2.1 Initial Biome Delineation

Five individual clusters of spectral data were identified in Alaska using NDVI and NDVI seasonality composites for the time period of 1982-2013 as data for the mclust-algorithm and following the classifications of Scheffer et. al[46]. The qualitative properties of these stable states have been compiled in table 5.3.

All of the identified clusters are well separated from each other through their mean annual NDVI scores alone with state 4 exerting an extremely low mean cluster value (0.065). The next cluster mean on the mean annual NDVI spectrum is that of stable state 5 (0.236) followed by those around which stable states 2 and 3 have been identified in increments of ~ 0.7. Stable state 1 is defined by the highest mean annual NDVI values occurring in Alaska (0.499). NDVI seasonality is high throughout the range of available data for Alaska with the exception of data belonging to the cluster of stable state 4 which exhibits low seasonality at 0.218.

Table 5.3: Mclust Model for vegetation-climate modelling (Alaska). Clusters have been formed around mean values of the data parameters. See below for model and cluster characteristics.

	Cluster Means	
Stable State	NDVI	NDVI Seasonality
1	0.499	0.728
2	0.301	0.784
3	0.373	0.88
4	0.065	0.218
5	0.236	0.645
Model name:	VVV	
BIC:	87555.91	

The 5 two-dimensional clusters of data from Alaska have been visualized in figure 5.5 as an uncertainty plot depicting the data colourized in accordance to their cluster group membership which can be found in figure 5.6. The size of individual dots is representative of the uncertainty assigned to the corresponding data points in their respective cluster ascription.

Figure 5.5 and table 5.3 show that the model established by mclust consisted of two-dimensional data clusters of variable volume, variable shape and variable orientation (VVV).

Although the separation of clusters in table 5.3 is relatively distinct through mean annual NDVI alone, figure 5.5 depicts a more clumped visualisation of the 5 clusters due to extensive tolerance ranges around clustering means. The data points are arranged in a continuum of increasing NDVI and NDVI seasonality. Only stable state 1 and 4 are distinguishable due to their locations on either end of the data spectrum in NDVI and NDVI seasonality dimensions.

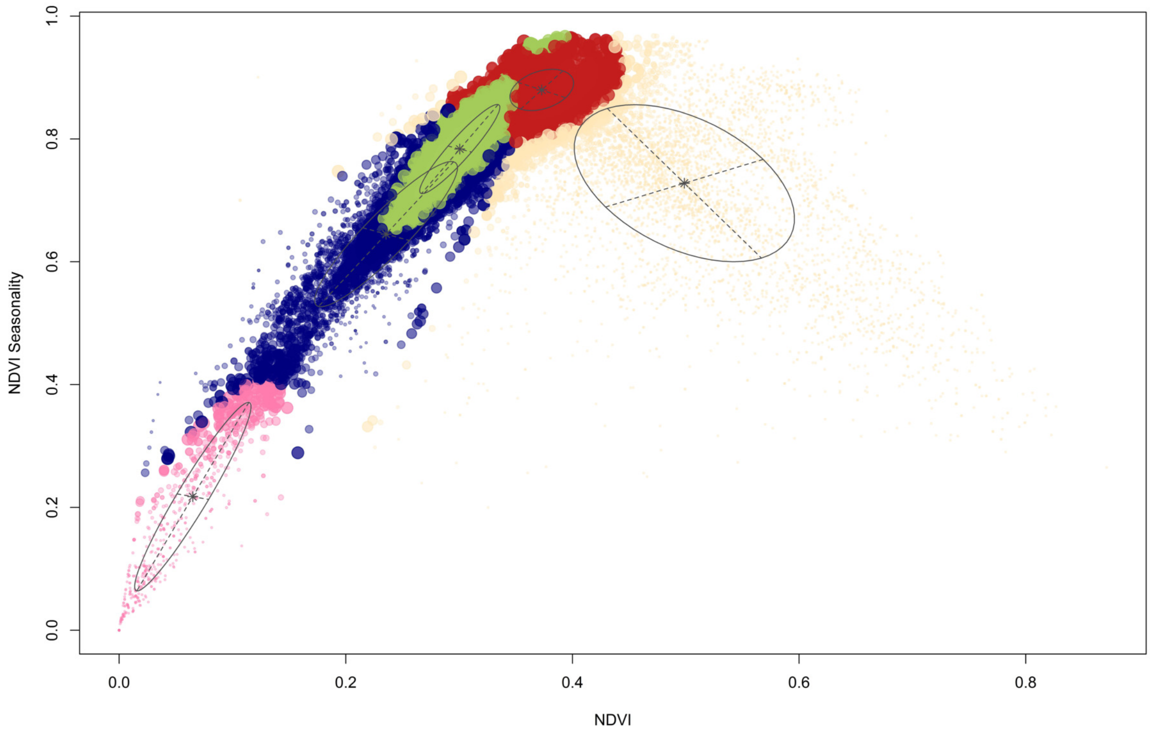
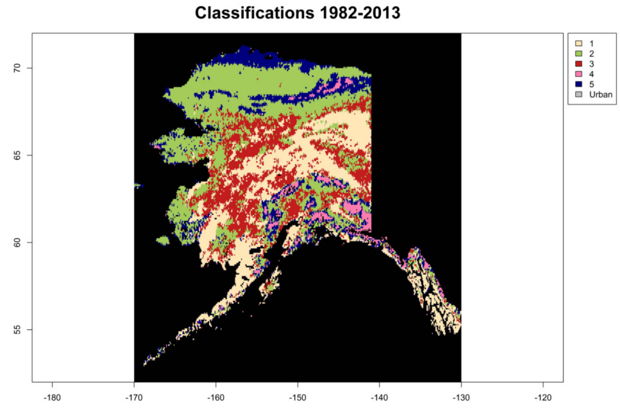


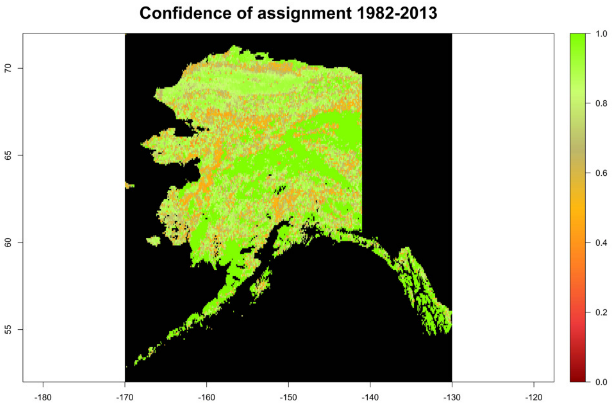
Figure 5.5: Mclust Vegetation Clustering (Alaska): The data contained within in the identified clusters have been plotted in their respective dimensions as uncertainty plots with larger dots indicating higher uncertainty levels of assignment of the respective data point to its currently assigned cluster. The ellipses correspond to the covariances of the clustering components.

The model built on the full spectral data set (composites of 1982-2013, table 5.3) has been exerted on spectral data composites covering the time frames of 1982-1986 (past) and 2009-2013 (present). The resulting distributions of the clusters identified in table 5.3 and figure 5.6a for the past and present time frames have been mapped out in figure 5.6c and figure 5.6e respectively. Assignment confidences for the classifications of the different time periods have been mapped accordingly in figures 5.6b, 5.6d and 5.6f.
The cluster distributions of the full (figure 5.6a) and past (figure 5.6c) spectral data sets were similar in that cluster 1 is confined to central areas of Alaska as well as along the southern coastline. The north of Alaska in these spectral composites is dominated by stable states 2, 4 and 5 with stable state 2 exerting the largest coverage in northern latitudes. Lastly, stable states 4 and 5 occur also in small pockets along the southern regions of Alaska.

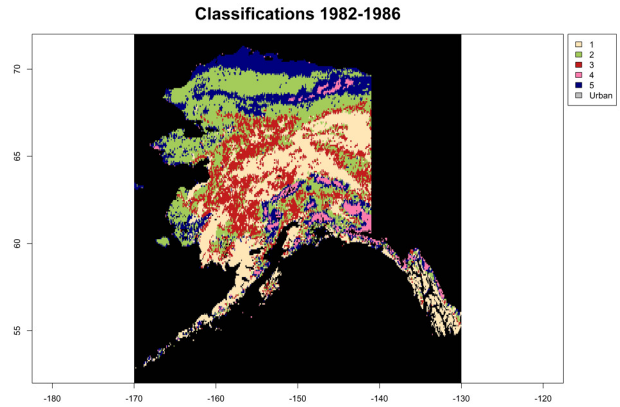
In the spectral data set for 2009-2013 (figure 5.6e), the distributions of stable states 1, 4 and 5 have not shifted much. Cluster 2 however has lost significant parts of its distribution to stable state 3 which expanded to the north and stable state 5 which expanded its northern coastline region southwards as well as to stable state 3.



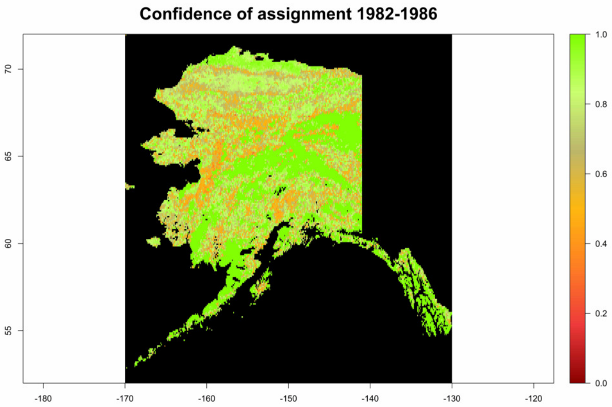
(a) Biome distributions (1982-2013)



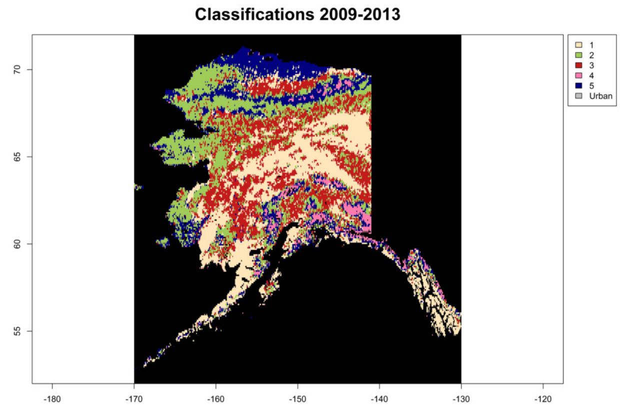
(b) Biome assignment confidence (1982-2013)



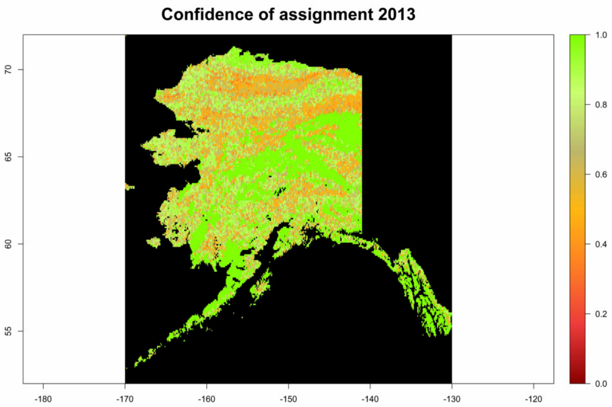
(c) Biome distributions (1982-1986)



(d) Biome assignment confidence (1982-1986)



(e) Biome distributions (2009-2013)



(f) Biome assignment confidence (2009-2013)

Figure 5.6: Biome classifications via mclust (Alaska): Spectral data sets for 1982-2013 have been clustered using the mclust algorithm. The resulting model has been used to cluster spectral data from 1982-1986 and 2009-2013 respectively.

5.2.2 Biome Distributions And Contingency

Figure 5.7 presents the proportions of alternative stable states of vegetation in Alaska over the course of 30 years. These proportions are depicted in three stages: full data set (1982-2013), past (1982-1986) and present(2009-2013). The proportions observed in the full data set closely resemble those in the past data set except for the relative proportion of stable states 2 and 5. Obvious change occurred between past and present proportions of alternative stable states, which have been subjected to further assessment. Whilst clusters 3, 4 and 5 shrunk in proportion, cluster 3 increased its relative coverage of Alaska from past to present (18.96% to 24.27% of Alaska's total land area). There has also been slight increase in relative percentage of stable state 1 coverage across Alaska (26.93% to 30.23%)

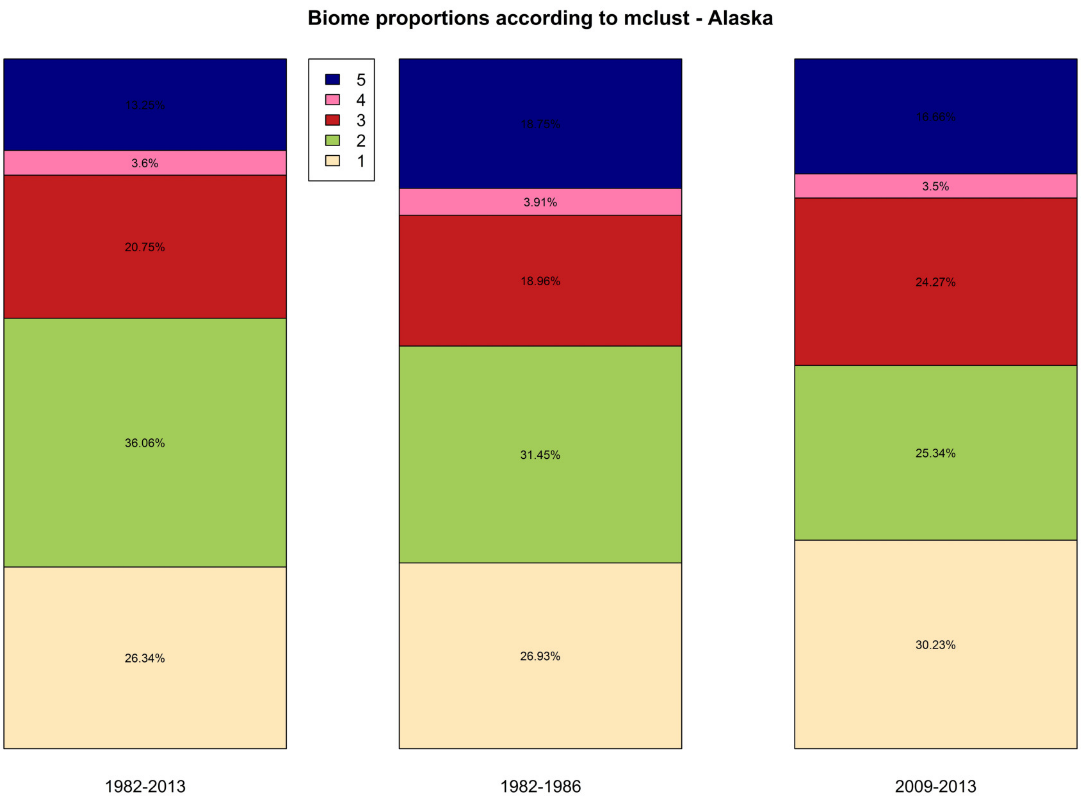


Figure 5.7: Biome Contingency (Alaska): The change of biome proportions in Alaska over a time span of 30 years (1982-2013) has been analysed using the mclust method.

Table 5.4 shows the percentages of change in proportion from one state to another relative to the initial proportions in the past data set when contrasted to the proportions in the present data set which show stable state 1 to be very conservative in its distribution (93.39% of its former area remained the same). Furthermore, stable state 4 has been relatively stable with 76.8% of its distributed area remaining classified as cluster 4. The largest shifts in distributions have occurred from stable states 2 to 3 (29.55%), 4 to 5 (21.77%) and 5 to 2 (26.18%).

Table 5.4: Biome Contingency (Alaska): The similarity of stable state distributions obtained by analysing spectral data of 1982-1986 (past) and spectral data of 2009-2013 (present) has been calculated to assess the transitioning of stable states of vegetation in Alaska. Change from past to present state have been assessed as percentages relative to the initial proportion of the past state.

	Present State				
Past State	1	2	3	4	5
1	93.39%	0.78%	4.56%	0.06%	1.21%
2	5.53%	53.5%	29.55%	0.02%	11.4%
3	12.12%	17.85%	68.51%	0%	1.53%
4	1.23%	0.19%	0%	76.8%	21.77%
5	5.3%	26.18%	4.1%	2.52%	61.89%

5.2.3 Presence Of Alternative Stable States

Prediction probabilities of individual stable state GAMs have been treated as presence/absence indicators by applying the respective ROC determined threshold. A visual representation of possible distribution of stable states (overlap of predicted distributions of stable states) has been compiled in figure 5.8. This figure does not contain information on overlap of individual biomes but focusses on the possibility of multiple stable states occurring at the same location.

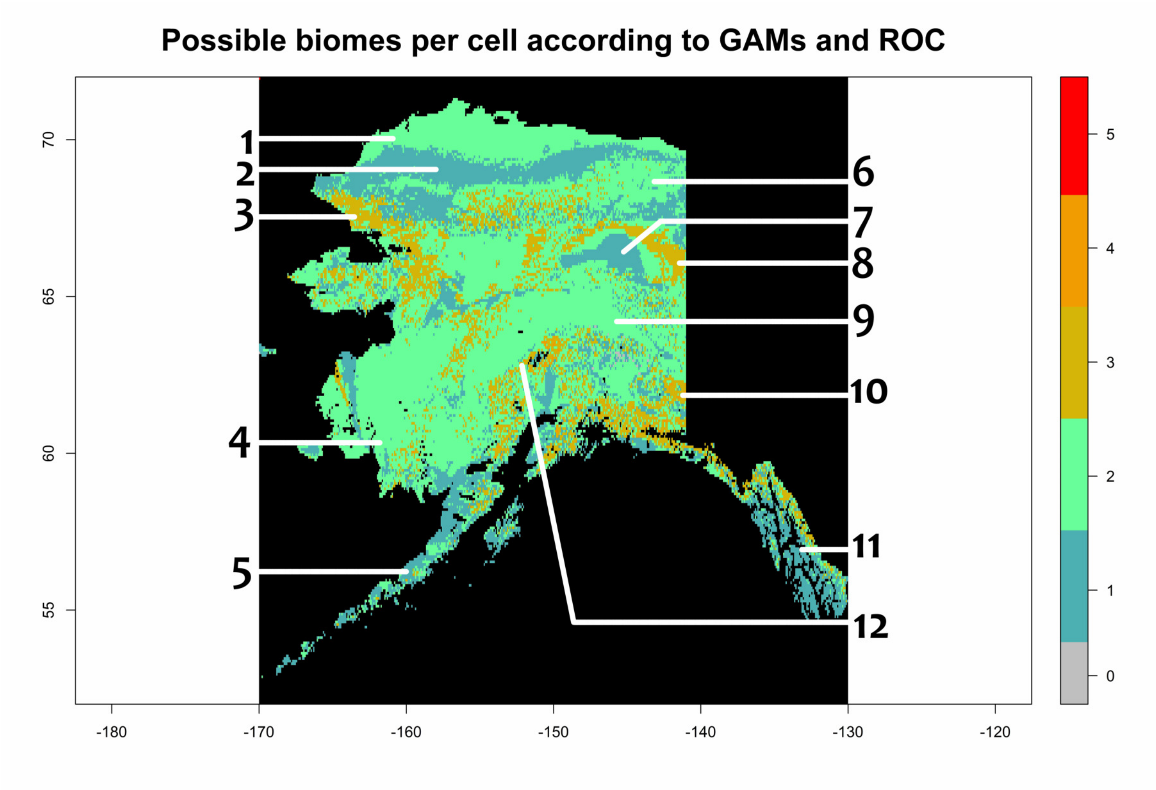


Figure 5.8: ROC prediction of presence of alternative stable states (Alaska): The ROC thresholding method has been used in combination with GAMs to assess the possibility for multiple stable states to occur in the same place. Numbers have been added for referencing particular areas.

Numbers in square brackets in the following paragraph are used to reference marked areas in figure 5.8.
As figure 5.8 shows, there are few regions for which 3 or more alternative stable states have been identified in Alaska. These areas are located mainly near the north-western coastline of Alaska [3], along the southern coastline [10], amidst the central region of Alaska in minor confinements [12] and in a crescent band on the eastern border of Alaska between 65° N and 67° N [8]. Whilst most of Alaska is classified as areas for which at least coverage by two alternatives of the individual stable states have been hypothesized [1,4,6,9], few areas for which cover of only one of the stable states could be estimated did occur as well. These areas are mostly located on the south-eastern appendage of land mass [11], along the south-western chain of peninsulas [5], in a large horizontal band at $\sim 69°$ N [2] and in a tight formation amidst central east Alaska [7].

6. Discussion

6.1 Potential And Realised Biomes

Potential and *realised* are terms which are usually employed in species distribution studies to describe niche types. A *potential* niche (also known as fundamental niche) describes a set of circumstances or a region throughout which a species *can* occur. A *realised* niche, on the other hand, describes the region in which the species *does* occur[119]. This terminology has, due to its usefulness in linking theoretical knowledge to real-world conditions, been incorporated into this study.

The remote sensed biome classifications and distributions - delineated via mclust - of this study have been referred to as *potential biomes*, whilst the real-world manifestations of those biome types have been addressed as *realised biomes*. The latter may deviate from the former in their extents (due to spatial resolution of the study parameters) and characteristics. It is useful to link potential biomes of a remote sensing study to realised biomes for evaluation of its accuracy and sensibility.

Realised biomes are usually not referred to by a numeric system and there is a somewhat loosely defined consensus on biome terminology among ecologists which can be used in identifying real-world biomes. The realised biomes of this study and their denominations were inferred from the LCCS of potential biomes (resulting from the mclust classification process employed in this study) by comparison with established LCCS of other authors. Doing so is a necessary step in evaluating the results of the biome identification process of this study (1st major research question), understanding of real-world biomes in relation to climate forcing (2nd major research question) as well as the analyses of their potential for resilient behaviour (3rd major research question).

For the purpose of linking potential to realised biomes this study employed information from three different established LCCS of Minnesota[99–101] and two established LCCS of vegetation composition of Alaska[46,120] all of which have also been used to infer appropriate numbers of vegetation types to be identified for each study site.

6.2 Minnesota

6.2.1 Realised Biomes

Potential biomes which resulted from the data based clustering of spectral composites from 1982-2013 have been linked to land cover classification types of established LCCS of Minnesota. All time series of monthly NDVI scores exhibited yearly oscillations with lowest values during the winter months (December, January, February). Climate responses have been assessed using the response functions produced by GAMs of each alternative stable state. For a brief summary of the results of this linking see table 6.1.

Table 6.1: Potential and realised biomes of Minnesota: Potential biomes of this study have been renamed according to their realised counterparts. The characteristics of the response functions have been included to this table using a set of abbreviations: (1) (+/+) the higher the climate parameter's level, the more likely the occurrence of the biome type, (2) (−/+) the lower the climate parameter's level, the more likely the occurrence of the biome type, (3) "O" characterizes an optimal level of the climate parameter for biome occurrence and (4) "-" labels cases in which no distinct climate response could be identified.

Biomes		NDVI Time Series		Climate Responses		
Potential	Realised	Lowest	Highest	Temperature	Precipitation	Elevation
1	Peatland	0.2	0.7	-	-	-
2	Prairie	0	0.85	O: 277K	(−/+), O: $<$55mm	-
3	Agricultural	0.2	0.8	(+/+)	-	-
4	Mixed forest	0.3	0.9	(−/+)	(+/+), O: $>$55mm	-

6.2.1.1 Peatland

The first cluster that was identified using the mclust algorithm on spectral data of Minnesota averaged over the time span from 1982 to 2013 (see figures 5.2 and A.7b) has been identified to be a *peatland* biome vegetated by shrubby spruce and *Larix laricina*. This biome classification is not included in the 'province' LCCS by the Minnesota Department of Natural Resources[100] which was used for comparison of potential biomes with realised biomes. However, shrub biomes have been identified for Minnesota within different LCCS[101] and thus included to this study as a possible alternative stable state of realised biomes. Although it can be argued that the section of land on the shore of Lake Superior might much rather be open, sparse forest (characterized by *Abies balsamea*), it is difficult to separate these two states based on NDVI data alone.

The peatland biome is characterised by a moderate seasonality of monthly NDVI scores ($\sim$ 0.2 to $\sim$ 0.7, figure A.7c). The GAM approach did not result in distinct response functions of peatland distribution in relation to mean annual precipitation, mean annual temperature or DEM data. This may be caused by either minute coverage of peatlands in Minnesota and thus small data sets on which to calculate the climate-vegetation relationships or by peatland coverage being influenced by an additional driving force. The peatland biome is, for example, largely confined to the vicinity of water bodies in Minnesota (figure A.7b) which may be a parameter influencing the likelihood of peatland species dispersion and should be considered to be added to future modelling approaches of vegetation-climate relationships.

6.2.1.2 Prairie

The second cluster of the mclust-classification of spectral data composites from 1982-2013 (figure A.7b) translates into a *prairie* classification within a LCCS of realised biome coverage of Minnesota.
Within this study, the prairie biome of Minnesota is characterized by a very strong, yearly seasonality of monthly NDVI scores (~ 0.0 to ~ 0.85, figure A.9c) and appears to be driven by mean annual temperature with an optimum at ~ 277 K (figure A.10d). Furthermore, a distinct response in prairie distribution to mean annual precipitation was delineated which describes the praire biome to be the most likely to occur when mean annual precipitation is low and particularly when its levels are below ~ 55 mm (figure A.10e). Although the response function to altitude data (figure A.10f) does show influence of elevation on prairie distribution, it is arguably not presenting a general trend which can be evaluated.

6.2.1.3 Agricultural land use

Agricultural areas, which have been described as a large portion of land use in Minnesota[99], have been identified as the third cluster within the LCCS established through the mclust-algorithm (figure A.11b).
Their NDVI seasonality function is similar to that of the peatland biome with yearly oscillation of monthly NDVI scores from ~ 0.2 to ~ 0.8 (figure A.11c). Despite the large coverage of Minnesota by agricultural areas (26.46% in the full spectral data set) a distinct response to changes in mean annual precipitation levels across Minnesota could not be calculated (figure A.12e). This may, however, not be a failure of the approach but indicative of agricultural land use in Minnesota relying on artificial irrigation. The distribution of the agricultural areas was successfully linked to mean annual temperature (figure A.12d) with increasing mean annual temperature prompting the presence of agricultural land use. Although there was a weak link of DEM data and presence of agricultural regions (figure A.12f), it can be argued that the elevation changes in Minnesota are not large enough to actually confine agricultural areas.

6.2.1.4 Mixed forest

The last and vastest of the potential biomes identified through the mclust algorithm within Minnesota, the fourth cluster (figure A.13b), is representative of the *Mixed Laurentian Forest* in the north east of Minnesota[100].
Its seasonality, although not very strong, is special for Minnesota because it exerts the highest peaks (~ 0.9) and the low points of its function (figure A.13c) are also the highest among the NDVI time series of biomes in Minnesota at values of ~ 0.3. The presence of the mixed forest

biome in Minnesota is fostered in regions of low mean annual temperatures and high mean annual precipitation (within the spectrum of climate data available for Minnesota, see figures A.14d and A.14e for the response functions). Elevation does not seem to confine the mixed forest biome to specific regions of Minnesota (figure A.14f).

6.2.2 Resilience Assessments

The approaches used within this study to assess resilience required evaluation and could thus be postulated to encompass a systems resilience only to a certain extent. Estimates of resilience identified via these approaches are therefore being referred to as Presumed Resilience (PR) in the following sections and do not claim to represent resilience scores in a holistic sense of the focal system.

6.2.2.1 Clustering Confidence

The results of the mclust analysis of NDVI composite data sets via a model that was established using the full range of available NDVI data exhibit lowered assignment confidence scores and thus lowered PR across transition zones of biomes in Minnesota (figure 5.2). Although this was to be expected and is in concordance with working resilience theory, the ecological transition zones of Minnesota are represented as very distinct but narrow regions.

6.2.2.2 Biome Contingencies

During the time period of 1982-2013, the prairie was the only biome expanding throughout Minnesota, particularly taking over large portions of agricultural regions (figure 5.3, table 5.2). The prairie biome stands out in that it took over $\sim 41\%$ of the past coverage of another biome (agricultural land use) and lost almost no territory to other biome classification types (table 5.2). Both the prairie and the mixed forest biome did not experience heavy loss of relative coverage within Minnesota when contrasting past and present spectral clustering results (figure 5.3). Areas that were identified as either prairie or mixed forest areas in 1982-1986 thus exhibited high levels of PR whilst peatland and agricultural areas were characterized by low PR scores.

6.2.2.3 Potential For Multiple Alternative Stable States

Numbers in square brackets and numbered region references refer, in the following, to the correspondingly numbered areas in figure 5.4.
The prediction of biome distribution throughout Minnesota using the individual GAMs and the ROC method for turning predictions into binary presence indicators resulted in an encompassing coverage of Minnesota of at least one biome in almost every cell of the data grid [7,8] (figure 5.4). Area 7 is predicted to be dominated by agricultural land use whilst the western part of area 8 is characterized as a distribution of the prairie biome. The eastern regions of area 8 are presumed to be characterized as a distribution of the Mixed Laurentian Forest [100].
However, there are areas for which coverage of two or more alternative stable states would be plausible, given the classification of biomes resulting from the first two steps of this study

(clustering spectral data and general additive modelling). These areas have been hypothesized to be linked to lowered resilience of the focal system and are largely distributed in the vicinity of water bodies [1,2,3,5] as well as around the Twin Cities [6,9] and, in a horizontal band, west of the Twin Cities [4]. The southern most regions of lowered PR of Minnesota [6,9] are characterized by the simultaneous prediction of presence of agricultural land use and the mixed forest biome. The possibility for agricultural land use and the prairie biome to be present in the same region and set of circumstances leads to lowered PR levels of Minnesotas vegetative landscape in region 4 and, together with the additional prediction of shrub presence, leads to areas of very low PR in the northern most parts of Minnesota [1]. The large area of lowered PR potential in the vicinity of Lake Superior in the east of Minnesota [5] is caused by an overlap of possible distributions of shrub and mixed forest. The mixed forest biome is also needed to explain the low PR exhibited around the Red Lakes where its predicted distribution coincides with those of shrub, prairie and, in cases, even agricultural land use [2,3].

6.2.3 Ecological Relevance

The rapid decline in agricultural land use fraction of Minnesota may be explained not as a shift from agricultural areas into prairie regions as such but as shifts towards degraded agricultural areas due to recent years of drought [121]. These degraded areas of former thriving agriculture closely resemble those of prairie areas in spectral appearance.

The significant shift from peatland areas to agricultural areas ($\sim 25.51\%$) can be hypothesized to be indicative of a relocation of agricultural land use to the moist areas of the peatlands. However, there are confounding factors, such as (1) human intervention and (2) a very minute fraction of Minnesota being classified as peatland to begin with, which should be considered.

Combining the knowledge on drought forcing of the biomes of Minnesota with the observed directional regime shifts in table 5.2 results in a succession of biomes along a precipitation gradient. Starting at the lower end of the spectrum and increasing the levels incrementally orders the biomes as follows: (1) prairie, (2) agricultural areas, (3) mixed forest and (4) peatland. If the drought effects were to continue on, agricultural areas would degrade further and be forced to move into areas currently classified as peatland. At the same time, the invasion of mixed forest areas by the prairie biome would intensify. This, however, entails the idea of predictable succession of alternative stable states which could only be inferred from a perfect understanding of any given focal system and its alternative stable states and has frequently been refuted [122].

The potential for multiple alternative stable states (figure 5.4) discloses large portions of Minnesota's landscape to exhibit high PR due to an overall coverage of areas for which only one stable state has been postulated to exist. However, there are areas for which the occurrence of multiple alternative stable states has been hypothesized. The presence of the actual dominant biomes in those regions is most likely dependent on a dominance hierarchy among these biomes[123], biological legacies[124], further climate and terrain influences[33,46] and human action[103]. Knowledge about such areas of ecological uncertainty can be used to direct further research and identify areas of special interest for management applications.

It is particularly noteworthy that change in biome proportions and distributions did occur mostly from past (1982-1986) to present (2009-2013) data sets with the classification of the full spectral data set (1982-2013) closely resembling that of the past time frame both in proportions (figure 5.3) and distributions (figure 5.2) of alternative stable states. This may be understood as a sign of very recent changes in biome distributions and proportions linked to rapid change in driving processes across Minnesota such as the recent years of drought conditions.

6.3 Alaska

6.3.1 Realised Biomes

Potential biomes identified for Alaska have been linked to land cover classification types of established LCCS of Alaska. All time series of monthly NDVI scores exhibited yearly oscillations with lowest values during the winter months (December, January, February). Climate responses have been assessed using the response functions produced by GAMs of each alternative stable state. For a brief summary of the results of this linking see table 6.2.

Table 6.2: Potential and realised biomes of Alaska: Potential biomes of this study have been re-named according to their realised counterparts. The characteristics of the response functions have been included to this table using a set of abbreviations: (1) (+/+) the higher the climate parameter's level, the more likely the occurrence of the biome type, (2) (−/+) the lower the climate parameter's level, the more likely the occurrence of the biome type, (3) "O" characterizes an optimal level of the climate parameter for biome occurrence and (4) "-" labels cases in which no distinct climate response could be identified.

Biomes		NDVI Time Series		Climate Responses		
Potential	Realised	Lowest	Highest	Temperature	Precipitation	Elevation
1	Boreal forest	0.2	0.8	(+/+)	(-/+)	-
2	Dwarf shrub	0	0.8	O: 255K/263K	O: 100mm	O: <2,000m
3	Shrubland	0	0.9	O: 265-277K	-	O: <1,700m
4	Bare/Mossy	0	0.2	-	(+/+)	(+/+)
5	Tundra/Sedge	0	0.6	-	(+/+)	O: <2,700m

6.3.1.1 Boreal forest

The first alternative stable state identified for Alaska by the mclust algorithm used in this study (figure A.15b) translates into the *boreal forest* biome which has been frequently described as a large portion of Alaska's vegetation cover[43,46].

The time series of monthly NDVI composites of this biome (figure A.15c) is special amongst the biomes of Alaska because its lowest points do not reach 0.0 but exhibit levels of ~ 0.2 during the winter months. The highest peaks of this time series reach ~ 0.8 thus resulting in a moderate, seasonality of the boreal forest. Although not being influenced by changes in elevation (figure A.16f), the occurrence of boreal forest is favoured by mean annual temperature levels at the upper end of the spectrum across Alaska (figure A.16d) and the lowest mean annual precipitation levels across Alaska (figure A.16e).

6.3.1.2 Dwarf shrub

The *dwarf shrub* community[120] corresponds to the second cluster delineated for Alaska in this study (figure A.17b).

The vegetation within this classification exerts strong seasonality of monthly NDVI scores which range from 0.0 to $\sim$ 0.8 (figure A.17c). Given the location and characteristics of this type of vegetation composition, it is reasonable to assume snow coverage during the winter months leading to the exceptionally low monthly NDVI scores. The response functions of the dwarf shrub biome highlight two separate temperature optima at which the biome can occur ($\sim$ 255 K and $\sim$ 263 K, figure A.18d). Furthermore, an optimum of dwarf shrub emergence in relation to mean annual precipitation levels has been identified at 100 mm (figure A.18e). The increased likelihood of dwarf shrub presence at altitude levels below 2,000 m (A.18f) may be indicative of a temperature response which favours dwarf shrub at higher temperatures.

6.3.1.3 Shrub

The third cluster identified for Alaska by the mclust algorithm employed in this study (figure A.19b) matches the *shrub* biome described by the Alaska Center for Conservation Science[120]. Its NDVI time series exhibits the strongest seasonality across the biomes of Alaska with monthly NDVI scores ranging from 0.0 to $\sim$ 0.9 (figure A.19c). Again, it may be reasonable to assume snow coverage of the vegetation throughout the winter months. Although not being driven by mean annual precipitation (figure A.20e), the distribution of the shrub biome in Alaska is influenced by mean annual temperature (figure A.20d) and elevation levels (figure A.20f) both of which present a positive relationship between temperatures and likelihood of shrub emergence. The inverse relationship of elevation levels and likelihood of shrub presence has been interpreted as masking a direct relation of mean annual temperature and likelihood of shrub emergence.

6.3.1.4 Sparse vegetation

A sparsely vegetated (most likely by mosses) or altogether barren land cover type has been identified to cover minute areas of Alaska[120]. The fourth cluster of spectral data of Alaska within this study (figure A.21b) corresponds to this biome.

It is characterized by a weak yearly oscillation of monthly NDVI composites ranging from 0.0 to $\sim$ 0.2 (figure A.21c) and thus the lowest NDVI values all year-round across Alaska. Snow cover can be assumed with certainty for winter months but can also be hypothesized to be in action throughout the whole year for certain patches (especially summits of high mountains such as Mt. McKinley). Its response to mean annual temperature (figure A.22d) is weak but indicates presence of barren land at low temperatures. Precipitation does not seem to influence the distribution of this biome greatly although the response function in figure A.22e implies a positive correlation between the two. Elevation, however, presented itself as an excellent descriptor of likelihood of barren/mossy vegetation emergence (figure A.22f). The positive trend of this function can reasonably be assumed to mask a temperature response which favours sparse, mossy vegetation (if any vegetation at all) at low mean annual temperatures.

6.3.1.5 Tundra and sedge

The real-world equivalent to the fifth alternative stable state delineated for Alaska in this study (figure A.23b) has been frequently described as a *tundra/sedge* biome[46,120].
Its monthly NDVI scores exhibit moderate seasonality ranging from 0.0 to $\sim$ 0.6 (figure A.23c). Snow-coverage in winter is most likely a certainty. Despite the lack of a distinct relationship of mean annual temperature and presence of this biome (figure A.24d), a masked temperature response can be inferred from the response function to elevation data (figure A.24f) which translates into presence of tundra/sedge vegetation to be favoured at higher temperatures. The positive correlation of mean annual precipitation levels and likelihood of tundra/sedge emergence (figure A.24e) indicates that tundra/sedge distribution is driven by precipitation regimes.

6.3.2 Resilience Assessments

Again, estimates of resilience identified via the following approaches are referred to as PR in the following sections.

6.3.2.1 Clustering Confidence

Assignment confidence scores of the clustering of spectral composite data sets via the mclust-method (figure 5.6) display reduced values at ecological transition zones across Alaska. Such lowered assignment confidences can especially be found in northern regions of the present data set (2009-2013, figure 5.6f), where distinct change in biome dominance has occurred when compared to the past (1982-1986, figure 5.6d) data set. This may be indicative of the 'new' biome dominance in the north not being thoroughly manifested yet with the system not being strongly resilient within the 'new' alternative stable states. The lowered PR scores of north Alaska within the present data set may thus be remnants of a recent shift and ongoing (re)organisation of alternative stable states. At the current stage, however, this is mere speculation. Although the assignment confidence levels across Alaska exhibit more uncertainty than in Minnesota, the areas of assignment uncertainty are, again, relatively confined in their extents, particularly regarding the composite clustering of the full range of available NDVI data (figure 5.6b).

6.3.2.2 Biome Contingencies

Whilst the past three decades saw relative coverage of Alaska by several biome types decreasing (dwarf shrub, barren/mossy and tundra/sedge), these conditions allowed for the expansion the two remaining biomes (boral forest and shrubland, figure 5.7).
Table 5.4 shows that only the boreal biome and the sparsely vegetated barren/mossy biome exerted high levels of PR when contrasting past and present distributions. The boreal forest biome kept to over 93% of its original coverage whilst expanding at the cost of other biomes distributions (especially the shrubland). Although the sparsely vegetated state remained within 76.8% of its original distribution pattern, it lost over 21% of its original area to the

tundra/sedge biome and almost none to any other alternative stable state. This may be indicative of a resilience pattern skewed in favour of regime shifts from bare/mossy biomes to tundra/sedge biomes. Two more examples of such state-specific/directional resilience patterns can be observed: (1) regime shifts from dwarf shrub to shrubland were favoured and (2) large portions of tundra/sedge areas shifted into being defined as shrubland areas. Conjugating these directional regime shifts results in the delineation of a ranking of alternative stable states of Alaska: (1) sparse barren/mossy, (2) tundra/sedge, (3) dwarf shrub, (4) shrub and (5) boreal forest.

6.3.2.3 Potential For Multiple Alternative Stable States

Numbers in square brackets and numbered region references refer, in the following, to the correspondingly numbered areas in figure 5.8.

The ROC predictions of alternative stable states distributions across Alaska exert a rough over-all coverage of at least two alternative stable states being predicted for every pixel within the coverage data map [1,3,4,6,8,9,10,12] (figure 5.8). There are deviations from this concept [2,5,7,11] which will be discussed in more detail shortly.

The northern most areas of Alaska [1,6] are predicted to contain up to two alternative stable states. These states are tundra/sedge and dwarf shrub which also account for areas where two alternative stable states have been predicted as possible coverages on the western coast of Alaska (these regions have not been numbered). Regions 4 and 9 are classified as overlaps of predicted distributions of boreal forest and the shrub biome. Areas of even lower PR (three or more states) have been identified across Alaska and are largely made up of potential for the emergence of shrub, dwarf shrub and boreal forest [3,8] and dwarf shrub, tundra and bare/mossy patches [10,12]. However, there are also areas of high PR for which only the coverage of one stable state at a time has been predicted (the above mentioned deviations). These include: region 2 which is a dwarf shrub patch as well as regions 5,7 and 11 which are potential occurrences of the boreal forest biome.

6.3.3 Ecological Relevance

The directional nature of regime shifts that have been observed among the five biomes of Alaska can be used to establish a ranked hierarchy of adjacent alternative stable states: (1) sparsely vegetated barren/mossy, (2) tundra/sedge, (3) dwarf shrub, (4) shrubland and (5) boreal forest. The results of this study show a marked increase of coverage exerted by boreal forest and shrubland over the last three decades. New areas defined as boreal forest and shrubland zones have been established especially amongst northern regions of Alaska which were originally classified as dwarf shrub land cover classifications. In accordance with a study by Beletov. et al[49] and, together with the reduction of barren/mossy areas, this can be seen as an indicator of global warming. Particularly the boreal forest biome is favoured by increasing mean annual temperatures. Whilst the barren/mossy biome does not exhibit a temperature response as such, the DEM response can be identified to mask a temperature response (higher elevation results in lower mean annual temperature). If the processes inducing vegetation change in Alaska (especially the effects of global warming[46]) over the last 30 years were to continue, the sparsely vegetated barren/mossy biome would be the first to vanish, followed by the tundra/segde state. However, this state would also be eventually superseded by the next alternative stable state of the ranking hierarchy. In the next few decades, this could lead to a marked increase in boreal forest and shrubland across Alaska (the two highest ranking biomes across Alaska). Again, although these successions of alternative stable states may seem intuitive they may be hypothesized as a general trend but should not be taken at face-value for the representation of processes in nature. Following the succession ranking of the biomes of Alaska would result in the boreal forest being the only biome present in Alaska as global warming goes on. Amongst obvious pitfalls such as the neglect of altitude effects, the possibility for biome evolution and niche conservation[64], the idea of ecological succession is also thwarted by its assumption of coming to a halt in Alaska once the boreal forest stage is reached and does not include the possibility for invasion of Alaska's vegetative landscape by other biomes which are not currently present.

The analysis of the predicted biome distributions shows that, with a few exceptions, almost all of Alaska is suitable for two of the five alternative stable states at any given location given the current climate regimes. The presence of the actual dominant biomes in those regions is, again, most likely dependent on a dominance hierarchy among these biomes[123], biological legacies[124], further climate and terrain influences[33,46] and human action[103]. The ROC prediction map (figure 5.8) can be used to direct further research and identify areas of special interest for management applications within Alaska.

Furthermore, a great portion of temperature responses of the set of biomes delineated across Alaska did not show correlation of vegetation occurrence with mean annual temperature or only a weak relationship between the two. Scheffer et. al hypothesized that the vegetation assemblages of Alaska are largely influenced by mean July temperature[46] and it may prove useful to include this metric as an additional explanatory variable to the GAMs.

6.4 Evaluation Of Resilience Assessments

6.4.1 Clustering confidence

Using assignment confidence values for assessing the resilience of a focal system does result in large areas which are classified as being at the maximum of their resilience potential with assignment scores being distributed in a less gradual fashion than theory would indicate. This may be caused by very selective likelihood functions within the mclust algorithm and the values of assignment confidences will change significantly with every clustering parameter that is added or removed from the data set. Additionally, the results of this approach are hugely dependent on the number of clusters that are to be identified. Furthermore, this method may only be used on data that is already available thus offering no predictive capabilities and only being capable of shedding light on resilience patterns of past events instead of being of avail for a priori deductions. Again, however, one has to be vary of predictions of succession in ecological systems since these require a near-perfect understanding of the focal system and any other system it is connected to thus resulting in a snow ball effect of ever increasing information requirements. Assignment confidence, which originates from the initial clustering of response variables, can be dismissed as a potent descriptor of biome resilience.

6.4.2 Biome Contingencies

Assessing biome proportions, especially the relative quantity and quality of regime shifts among them, is a useful tool in delineating stability landscapes by conjugating regime shifts from one state into another based on their quantities of occurring. However, this only yields a succession of adjacent attractor regimes (basins within figure 1.1) but gives little to no information on what actually makes the systems resilient and exactly how resilient they are. Furthermore, this approach does not hold any predictive capabilities. It may, however proof very useful in combination with other approaches which exhibit predictive power and incorporate measurements of resilience type and strength.

6.4.3 Potential For Multiple Alternative Stable States

The analysis for the potential of overlap in hypothesized distributions of alternative stable states results in information on which areas may exhibit instability in that frequent change of the dominant biome type may occur. The analysis furthermore allows for the assessment of which alternative stable states compete in these areas which can, in combination with information on biome functioning (as is yielded via the GAM approach, for example), be used to predict which of the competing alternative stable states will prevail given a certain development of climate regimes. This approach to assessing resilience of a focal system on a regional scale can prove especially useful in directing further research to areas of ecological uncertainty as well as facilitating information for DSSs which are to be used in the formulation of environmental policy goals. Additionally, this approach does hold predictive capabilities due to its underlying process of general additive modelling which can be combined with predicted climate regimes, resulting from climate modelling, to delineate future hotspots of ecological uncertainty and possible biome distributions. GAMs are at the centre of this approach to assessing resilience and it may prove useful to additionally account for spatial autocorrelation to smooth response functions and delineate more general trends of climate-vegetation interaction. On the notion of response functions, it may also be beneficial to incorporate more ecologically appropriate and simple response functions which could be established using the 'plateau' method described by Brewer et. al[119].

6.5 Limitations And Pitfalls

Although aiming to avoid **subjectiveness**, this study did not achieve complete freedom of subjective influence. Two subjective choices within this study influenced the results, especially within the first step of the framework (*identification of alternative stable states*):

1. *Choice of number of alternative stable states.*
   Despite mclust's capability of identifying the most appropriate number of clusters to be classified among the data set handed to its algorithms, the delineation of alternative stable states was set to a pre-conceived number of biomes which had been reported to cover the extent of the two study sites. The reasoning for doing so lies with Friedl et. al's statement that aggregation of clusters depends on user needs[61]. However, the aggregation could have also been approached by first identifying all vegetation types according to the BIC within the mclust algorithm and subsequently aggregating those. This, however, would have required extensive knowledge of the vegetative landscapes of the study sites and, preferably even, ground-proofing of the results, which was beyond the scope of this study.

2. *Choice of parameters.*
   This study employed NDVI measurements which, as sole descriptors of vegetation composition, are somewhat crude because vegetation assemblages of different plant species can produce similar NDVI values[125]. Furthermore, other VIs such as the EVI are available at superior spatial resolutions (NDVI resolution: 0.083°, EVI resolution: 0.05°[7,126]). The choice to use NDVI measurements was made because newer VIs at superior spatial resolutions are not recorded for a time as lengthy as the one for which NDVI measurements have been made (NDVI time frame: 1982-2013, EVI time frame: 2000-2013[7]). This was in accordance with Genkai-Kato's call for the use of long time series due to shifts not occurring frequently and long time records thus increasing the likelihood of encompassing a range of regime shifts on regional scales[127].

Furthermore, the lack of lag effects throughout the vegetation-climate modelling, which has been described by multiple authors[7,128], may be seen as cutting the variety of responses of vegetation to climate forcing short. The GAMs may, furthermore, be refined to incorporate spatial autocorrelation in order to smooth out response functions which are currently very irregular (as seen in figure A.20e, for example). The results of the GAMs (see section A.5) show not only irregular but sometimes even ecologically insensible response functions which should be dealt with in future approaches for example by using the platuea method of Brewer et. al[119].
Lastly, this study worked on the assumption of biomes being the real-world equivalent of alternative stable states on regional scales. Although this has been postulated as the foundation of other remote sensing studies as well, it may not necessarily be appropriate and this study did not identify state-governing feedbacks for every delineated alternative stable state which would specify them as such.

6.6 Future Directions

The identification of additional response parameters as well as a thorough framework concerning their selection and relative importance will enhance the identification of alternative stable states within the first step of the framework this study introduced. It could possibly even lead to fulfilling Mace et. al's request for sets of control variables for each biome to be identified[66]. Such sets of control variables for individual biomes would allow for functions to be built relating the control variable(s) to the likelihood of biome perseverance (also known as 'security score'). Identifying critical thresholds of control parameters within those functions would, in turn, serve as warning signals for ecosystem management as they would be indicative of high chances of strong degradation of the biome in question.

Additional, intermediary climatologies of NDVI data could be used to establish information on regime shifts in a time continuum on a finer temporal resolution and shed more light on the driving factors behind feedback systems of alternative stable states.

Choosing additional explanatory variables to be used within the vegetation-climate modelling will deepen the understanding of what influences biome distributions, especially in cases such as the peatland in Minnesota within this study. Mean July temperature, for example, has been shown to exert strong effects on vegetation in northern latitudes[46].

A new method for calculating response functions of binary presence indicators to climate parameters has been proposed by Brewer et. al[119] and could prove to establish more realistic climate envelopes for the delineated biome classifications.

Combining the method of delineating the succession of alternative stable states through the biome contingency matrices with a process of inferring the width of the resilience basins will enhance the understanding of each individual alternative stable states and its potential for resilient behaviour.

Finally, directing future research efforts towards the areas of ecological uncertainty identified within this study and generating more in-depth analyses of regions in northern latitudes using additional established methodology will make for well-rounded insights on changes in ecosystem structures and the effects of global climate change.

7. Conclusion

Understanding the behaviour of ecosystems through time and sets of conditions on large spatial scales is paramount to many environmental policy decisions. A novel, three-step framework towards examining the climate envelopes and resilience potential of alternative stable states on regional scales has been introduced via this study. It is aimed at answering three major topics of interest in three encompassing steps, which can be regarded as unique buildings blocks serving specific tasks:

1. **Identification of alternative stable states**
   Clustering data sets according to pre-conceived information on how many clusters to identify using only spectral data resulted in sets of potential biomes to be identified for each study site. These were subsequently linked to already described realised biomes.

2. **Assessment of climate envelopes of alternative stable states**
   Relating biome presence to the levels of climate parameters (mean annual temperature and precipitation) using GAMs was done to specify the likelihood of occurrence for each biome given certain climate regimes.

3. **Assessment of resilience of alternative stable states**
   Resilience, which has been identified as a key concept of ecosystem functioning in planning human interference with the environment, has been assessed in this study using three distinct approaches:

   (a) *Clustering confidence* used the assignment confidence for each data point in the original spectral data sets resulting from the mclust algorithm as indicators of resilience.

   (b) *Biome contingencies* were analysed for directional regime shifts and overall confinement of biomes to certain areas.

   (c) *Potential for multiple alternative stable states* was evaluated using GAM predictions for each alternative stable state given the current climate regimes and turned into binary indicators of presence and absence using the ROC thresholding method. Resilience has been hypothesized to be lowered where potential distributions of biomes overlap (given an accurate depiction of realised biomes via potential biomes).

Although, the last two approaches to assessing resilience have been evaluated as yielding useful information (succession of alternative stable states and areas of ecological uncertainty), the results of this study show that the methodology still requires certain degrees of refinement (see sections 6.5 and 6.6). However, due to the flexible nature of the three-step framework within this study it is possible to focus on refining one step at a time without overhauling the entirety of the approach.

The areas of ecological uncertainty identified within in this study should be made subjects to future policy concern and research effort within future applications of these concepts. This further work is encouraged to be based on the coding found in A.6 as R is a free software available to everyone and can, through work and enough time, be handled by everyone.

Bibliography

[1] Target 15 - technical rationale extended (provided in document cop/10/inf/12/rev.1) (2017). URL `https://www.cbd.int/sp/targets/rationale/target-15/`.

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

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

[4] Foley, J. A. *et al.* Global consequences of land use. *Science* (2005).

[5] Dietz, T., Rosa, E. A. & York, R. Driving the human ecological footprint. *Frontiers in Ecology and the Environment* **5**, 13–18 (2007).

[6] 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).

[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] Coté, I. M. & Darling, E. S. Rethinking ecosystem resilience in the face of climate change. *PLoS Biology* **8** (2010).

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

[10] Curtin, C. G. & Parker, J. P. Foundations of resilience thinking. *Conservation Biology* **28**, 912–923 (2014).

[11] Selkoe, K. A. *et al.* Principles for managing marine ecosystems prone to tipping points. *Ecosystem Health and Sustainability* **1**, 17 (2015).

[12] Rist, L., Felton, a., Nyström, M. & Troell, M. Applying resilience thinking to production ecosystems. *Ecosphere* **5**, 1–11 (2014).

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

[14] Scheffer, M., Carpenter, S., Foley, J. a., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–6 (2001).

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

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

[17] B., W., C.S., H., S.R., C. & ., K. A. Resilience, adaptability and transformability in social-ecological systems. *Ecology and Society* **9** (2004).

[18] Holling, C. S. RESILIENCE AND STABILITY OF ECOLOGICAL SYSTEMS. *Annual Review of Ecology and Systematics* **4**, 1–23 (1973).

[19] Carpenter, S., Walker, B., Anderies, J. M. & Abel, N. From Metaphor to Measurement: Resilience of What to What? *Ecosystems* **4**, 765–781 (2001).

[20] Carlson, L. *et al.* Resilience: Theory and Application. *Argonne National Laboratory* 1–64 (2012).

[21] Allen, C. R. *et al.* REVIEW: Quantifying spatial resilience. *Journal of Applied Ecology* **53**, 625–635 (2016).

[22] A.L., B. *et al.* The good(ish), the bad, and the ugly: a tripartite classification of ecosystem trends. *ICES J Mar Sci* (2010).

[23] Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307**, 321–326 (1984). `arXiv:1011.1669v3`.

[24] Peterson, G., Allen, C. R. & Holling, C. S. Ecological Resilience, Biodiversity, and Scale. *Ecosystems* **1**, 6–18 (1997). `9610022`.

[25] Steffen, W. *et al. Global change and the earth system: a planet under pressure* (Springer, UK., 2004).

[26] 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**, 524–539 (2003).

[27] Muthukrishnan, R., Lloyd-Smith, J. O. & Fong, P. Mechanisms of resilience: Empirically quantified positive feedbacks produce alternate stable states dynamics in a model of a tropical reef. *Journal of Ecology* 1662–1672 (2016).

[28] Lenton, T. M. Gaia and natural selection [review]. *Nature* **394**, 439–447 (1998).

[29] Hirota, M., Holmgren, M., van Nes, E. H. & Scheffer, M. Global Resilience of Tropical Forest and Savanna to Critical Transition. *Science* **334**, 232–235 (2011).

[30] Mumby, P. J., Steneck, R. S. & Hastings, A. Evidence for and against the existence of alternate attractors on coral reefs. *Oikos* **122**, 481 (2013).

[31] Bel, G., Hagberg, A. & Meron, E. Gradual regime shifts in spatially extended ecosystems. *Theoretical Ecology* **5**, 591–604 (2012).

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

[33] 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).

[34] Angeler, D. G. & Allen, C. R. Quantifying Resilience. *Applied Ecology* 617–624 (2016).

[35] Verbesselt, J. *et al.* Remotely sensed resilience of tropical forests. *Nature Climate Change* (2016).

[36] Scheffer, M. Foreseeing tipping points. *Nature* **467**, 6–7 (2010).

[37] Drake, J. M. & Griffen, B. D. Early warning signals of extinction in deteriorating environments. *Nature* **467**, 456–459 (2010).

[38] van de Pol, M. *et al.* Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution* 1246–1257 (2016). `arXiv:1011.1669v3`.

[39] 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).

[40] Latifovic, R., Zhu, Z.-L., Cihlar, J., Giri, C. & Olthof, I. Land cover mapping of North and Central America - Global Land Cover 2000. *Remote Sensing of Environment* **89**, 116–127 (2004).

[41] Prasad, V. K., 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).

[42] Evrendilek, F. & Gulbeyaz, O. Deriving vegetation dynamics of natural terrestrial ecosystems from MODIS NDVI/EVI data over Turkey. *Sensors* **8**, 5270–5302 (2008).

[43] 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).

[44] Goward, S., & Dye, D. Evaluating north america net primary productivity with satellite observations. *Advances in Space Research* **7**, 165–174 (1987).

[45] 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).

[46] Scheffer, M., Hirota, M., Holmgren, M., Van Nes, E. H. & Chapin, F. S. Thresholds for boreal biome transitions. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21384–9 (2012).

[47] Danz, N. P., Frelich, L. E., Reich, P. B. & Niemi, G. J. Do vegetation boundaries display smooth or abrupt spatial transitions along environmental gradients? Evidence from the prairie-forest biome boundary of historic Minnesota, USA. *Journal of Vegetation Science* **24**, 1129–1140 (2013).

[48] Descamps, S. *et al.* Climate change impacts on wildlife in a High Arctic archipelago - Svalbard, Norway. *Global Change Biology* 1–13 (2016).

[49] Belotelov, N. V., Bogatyrev, B. G., Kirilenko, A. P. & Venevsky, S. V. Modelling of time-dependent biome shifts under global climate changes. *Ecological Modelling* **87**, 29–40 (1996).

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

[51] Alatalo, J. M., Little, C. J., Jägerbrand, A. K. & Molau, U. Dominance hierarchies, diversity and species richness of vascular plants in an alpine meadow: contrasting short and medium term responses to simulated global change (2014).

[52] Ziska, L. H., Blumenthal, D. M., Runion, G. B., Hunt, E. R. & Diaz-Soltero, H. Invasive species and climate change: an agronomic perspective. *Climatic Change* **105**, 13–42 (2010).

[53] Masters, G. & Norgrove, L. Climate change and invasive alien species. *CABI Working Paper* **1**, 30 pp. (2010).

[54] B.B., H., V., G. & R., A. Climate events synchronize the dynamics of a resident vertebrate community in the high arctic. *Science* **339**, 313–315. (2013).

[55] U.S. Department of the Interior, Invasive Species Advisory Committee (ISAC). Invasive species and climate change.

[56] Hughes, J. K., Valdes, P. J. & Betts, R. Dynamics of a global-scale vegetation model. *Ecological Modelling* **198**, 452–462 (2006).

[57] Salzmann, U., Haywood, A. M., Lunt, D. J., Valdes, P. J. & Hill, D. J. A new global biome reconstruction and data-model comparison for the Middle Pliocene. *Global Ecology and Biogeography* **17**, 432–447 (2008).

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

[59] Olson, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **51**, 933 (2001). `j.1744-7429.2008.00471.x`.

[60] Higgins, S. I., Buitenwerf, R. & Moncrieff, G. R. Defining functional biomes and monitoring their change globally. *Global Change Biology* 1–11 (2016).

[61] 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).

[62] Crisp, M. Biome assembly: What we know and what we need to know. *Journal of Biogeography* **33**, 1332–1333 (2006).

[63] Woodward, F. I., Lomas, M. R. & Kelly, C. K. Global climate and the distribution of plant biomes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1465–76 (2004).

[64] Donoghue, M. J. & Edwards, E. J. Biome Shifts and Niche Evolution in Plants. *Annual*

Review of Ecology, Evolution, and Systematics **45**, 547–572 (2014).

[65] Crawley, M. J. *STATISTICS - An Introduction Using R* (Wiley-Blackwell, 2015), 2 edn.

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

[67] Lasslop, G., Brovkin, V., Reick, C. H., Bathiany, S. & Kloster, S. Multiple stable states of tree cover in a global land surface model due to a fire -vegetation feedback. *GEOPHYSICAL RESEARCH LETTERS* (2016).

[68] Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **83**, 195–213 (2002).

[69] Bartholomé, E. & Belward, A. S. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* **26**, 1959–1977 (2005).

[70] Running, S. W. *et al.* A remote sensing based vegetation classification logic for global land cover analysis. *Remote Sensing of Environment* **51**, 39–48 (1995).

[71] Prentice, I. C. *et al.* A global biome model based on plant physiology and sominance, soil properties and climate. *Journal of Biogeography* **19**, 117–134 (1992).

[72] 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).

[73] Ellis, E. C. & Ramankutty, N. Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and the Environment* **6**, 439–447 (2008).

[74] Fairbanks, D. H. K. Physio-climatic classification of South Africa's woodland biome. *Plant Ecology* **149**, 71–89 (2000).

[75] Walter, H. & Box, E. Global classification of natural terrestrial ecosystems. *Vegetatio* **32**, 75–81 (1973).

[76] 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).

[77] Bergengren, J. C., Thompson, S. L., Pollard, D. & Deconto, R. M. Modeling Global Climate - Vegetation Interactions. *Climatic Change* **50**, 31–75 (2001).

[78] Kutzbach, J. *et al.* Climate and Biome simulations for the past 21,000 years. *Quaternary Science Reviews* **17**, 473–506 (1998).

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

[80] Justice, C. O., Townshend, J. R. G., Holben, B. N. & Tucker, C. J. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* **6**, 1271–1318 (1985).

[81] Roy, P. S. *et al.* Biome mapping in India using vegetation type map derived using temporal satellite data and environmental parameters. *Ecological Modelling* **197**, 148–158 (2006).

[82] 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). This is the first ndvi paper.

[83] Thenkabail, P. S., Smith, R. B. & De Pauw, E. Hyperspectral vegetation indices and their rela- tionships with agricultural crop characteristics. *Remote Sensing of Environment* **71**, 158–182 (2000).

[84] Penuelas, J., Filella, I. & Gamon, J. A. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytologist* **131**, 291–296 (1995).

[85] Aparicio, N., Villegas, D. & Casadesus, J. Spectral vegetation indices as non-destructive tools for determining durum wheat yield. *Agronomy Journal* **92**, 83–91 (2000).

[86] Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J. & Dextraze, L. Integrated narrow- band vegetation indices for prediction of crop chlorophyll content for application to precision agri- culture. *Remote Sensing of Environment* **81**, 416–426 (2002).

[87] Blackburn, G. A. Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing* **19**, 657–675 (1998).

[88] Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B. & Rakitin, V. Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* **106**, 135–141 (1999).

[89] Gamon, J. A., Penuelas, J. & Field, C. B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of the Environment* **41**, 35–44 (1992).

[90] Penuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J. & Field, C. B. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sensing of Environment* **48**, 135–146 (1994).

[91] Gitelson, A. A., Merzlyak, M. N. & Chivkunova, O. B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology* **74**, 38–45 (2001).

[92] Friend, A. D., Stevens, A. K., Knox, R. G. & Cannell, M. G. R. A process-based, terrestrial biosphere model of ecosystem dynamics (Hybrid v3.0). *Ecological Modelling* **95**, 249–287 (1997).

[93] 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).

[94] Cammarano, D., Fitzgerald, G. J., Casa, R. & Basso, B. Assessing the robustness of

vegetation indices to estimate wheat N in mediterranean environments. *Remote Sensing* **6**, 2827–2844 (2014).

[95] 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).

[96] Chuine, I. *et al.* Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Global Change Biology* **22**, 3444–3460 (2016).

[97] Xu, L. *et al.* Temperature and vegetation seasonality diminishment over northern lands. *Nature Climate Change* **3**, 581–586 (2013).

[98] Nightingale, J. M., Fan, W., Coops, N. C. & Waring, R. H. Predicting tree diversity across the united states as a function of modeled gross primary production. *Ecological Applications* **18**, 93 (2008).

[99] Hart, J. F. & Ziegler, S. S. *Landscapes of Minnesota* (Minnesota Historical Society Press, 2008).

[100] Minnesota Department of Natural Resources (2017). URL `http://www.dnr.state.mn.us/ecs/index.html`.

[101] Minnesota Department of Natural Resources. Natural vegetation of minnesota at the time of the public land survey 1847-1907 (1988).

[102] Grünzweig, J. M., Valentine, D. W. & Chapin, F. S. Successional Changes in Carbon Stocks After Logging and Deforestation for Agriculture in Interior Alaska: Implications for Boreal Climate Feedbacks. *Ecosystems* **18**, 132–145 (2014).

[103] Prestele, R. *et al.* Hotspots of uncertainty in land use and land cover change projections: a global scale model comparison. *Global Change Biology* 0–34 (2016).

[104] 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).

[105] 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/`.

[106] Hijmans, R. J. *raster: Geographic Data Analysis and Modeling* (2016). URL `https://CRAN.R-project.org/package=raster`. R package version 2.5-8.

[107] Detsch, F. *gimms: Download and Process GIMMS NDVI3g Data* (2016). URL `https://CRAN.R-project.org/package=gimms`. R package version 0.5.1.

[108] Hijmans, R., Cameron, S., Parra, J., Jones, P. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25** (2005).

[109] Lehmann, C. E. R. *et al.* Savanna Vegetation-Fire-Climate Relationships Differ Among Continents. *Science* **343**, 548–553 (2014).

[110] Dolezal, J. *et al.* Vegetation dynamics at the upper elevational limit of vascular plants in himalaya. *Scientific Reports* (2016).

[111] Zuur, A., Ieno, E. & Smith, G. *Analyzing Ecological Data*. Statistics for Biology and Health (Springer New York, 2007).

[112] Wood, S. *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation* (2016). URL `https://CRAN.R-project.org/package=mgcv`.

[113] Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. *Mixed Effects Models and Extensions in Ecology with R*. Statistics for Biology and Health (Springer New York, 2009).

[114] Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B. & Fop, M. *mclust: Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation* (2016). URL `https://CRAN.R-project.org/package=mclust`.

[115] Fraley, C., Raftery, A., Murphy, T. B. & Scrucca, L. Mclust version 4 for r: Normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report 597, Department of Statistics, University of Washington (2012).

[116] Lobo, J. M., Jimènez-valverde, A. & Real, R. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151 (2008).

[117] Hosmer, D. & Lemeshow, S. *Applied Logistic Regression* (Wiley: New York, NY, USA, 2000), 2 edn.

[118] Carstensen, B., Plummer, M., Laara, E. & Hills, M. *Epi: A Package for Statistical Analysis in Epidemiology* (2016). URL `https://CRAN.R-project.org/package=Epi`.

[119] Brewer, M. J., O'Hara, R. B., Anderson, B. J. & Ohlemüller, R. Plateau: A new method for ecologically plausible climate envelopes for species distribution modelling. *Methods in Ecology and Evolution* 1489–1502 (2016).

[120] Alaska Center for Conversation Science. Vegetation map for northern, western, and interior alaska (2017). URL `http://accs.uaa.alaska.edu/vegetation-ecology/vegetation-map-northern-western-and-interior-alaska/`.

[121] United States Drought Monitor (2017). URL `http://droughtmonitor.unl.edu/MapsAndData/Graph.aspx`.

[122] Johnson, E. A. & Miyanishi, K. Testing the assumptions of chronosequences in succession. *Ecology Letters* **11**, 419–431 (2008).

[123] Iwao, K. *et al.* Creation of New Global Land Cover Map with Map Integration. *Journal of Geographic Information System* **03**, 160–165 (2011).

[124] Moncrieff, G. R., Scheiter, S., Bond, W. J. & Higgins, S. I. Increasing atmospheric CO2

overrides the historical legacy of multiple stable biome states in Africa. *New Phytologist* **201**, 908–915 (2014).

[125] Pettorelli, N. *et al.* Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution* **20**, 503–510 (2005).

[126] Solano, R., Didan, K., Jacobson, A. & Huete, A. *MODIS vegetation index C5 user's guide (MOD13 Series)* (2010).

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

[128] 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).

[129] 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).

[130] Mumby, P. J., Wolff, N. H., Bozec, Y. M. & Chollett, I. Operationalizing the resilience of coral reefs in an era of climate change. *Conservation ...* (2014).

[131] 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.

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

[133] 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).

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

[135] Angelini, C. *et al.* A keystone mutualism underpins resilience of a coastal system to drought. *Nature Communications* **7**, 1–8 (2016).

[136] Andres, N. G. & Rodenhouse, N. L. Resilience of corals to hurricanes: a simulation model. *Coral Reefs* **12**, 167–175 (1993).

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

[138] Holdridge, L. R. Determination of World Plant Formations from Simple Climatic Data. *Science* **105**, 367–368 (1947).

[139] Noever, D. A., Brittain, A., Matsos, H. C., Baskaran, S. & Obenhuber, D. The effects of variable biome distribution on global climate. *BioSystems* **39**, 135–141 (1996).

[140] Monserud, R. A. & Leemans, R. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling* **62**, 275–293 (1992).

[141] Lieng, E. *et al.* Classification of Land Cover Using Decision Trees and Multiple Reference Data Sources (2002).

[142] Lotsch, A. *Biome level classification of land cover at continental scales using decision trees.* Vordiplom, Free University of Berlin (1996). URL `http://cliveg.bu.edu/download/thdis/alotsch.MA.pdf`.

[143] Mayaux, P., Bartholome, E., Fritz, S. & Belward, A. A New Land Cover Map of Africa for the Year 2000. *Journal of Biogeography* **31**, 861–877 (2004).

[144] Xu, G. *et al.* A Bayesian Based Method to Generate a Synergetic Land-Cover Map from Existing Land-Cover Products. *Remote Sensing* **6**, 5589–5613 (2014).

[145] Crisp, M. D. *et al.* Phylogenetic biome conservatism on a global scale. *Nature* **458**, 754–756 (2009).

Appendix

A.1 Terminology

A.1.1 Basic Vocabulary

Table A.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[9,12,14,16,30,34,129,130]. Also known as *stable state*[26], *(stable) equilibrium*[6] or *regime*[129]. Due to ecosystems being subject to slow trends and fluctuations, it seems most appropriate to use the terms **attractor** or **regime**[9].
Disturbance	Any impact that perturbs a systems trajectory from a given attractor[129,131–134].
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[9,11–13,32]. The concept of regime shifts is supported by the presence of alternate attractors maintained by different dominant controlling feedbacks. Also known as *tipping point*[36] or *bifurcation*[16].
Hysteresis	A pattern defined by forward and backward switches between alternate attractors occurring at different critical conditions[9,11,14,16] .

A.1.2 Extrinsic And Intrinsic Factors

Table A.2: Factors influencing resilience. This list is non-exhaustive.

Internal elements	External elements
• Habitat growth nuclei, mutualism and facilitation[135] • Vegetation memory effects[7,45] • Seedling emergence, establishment, survival, growth rates[9,33,135] • Spatial extent of a system and spatial heterogeneity[12,135]	• Gradually changing environmental factors, environmental noise (probably leading to increased resilience over time)[9] • Physical environment, disturbance regime[130,136] • Disturbance frequencies and severities in past and present[136]

A.1.3 Understandings Of Resilience

Table A.3: Definitions of 'resilience': Five of the most popular understandings of the term 'resilience'.

Term	Definition
Ecological Resilience	'[...] measures the magnitude of disturbance that can be absorbed before the ecosystem's structure changes'[45] Ecological resilience is a measure of how much a system can be changed until it shifts from one stable state/attractor to another[3,15,18,24,34,45,132,137]. Also known as *General resilience*[34].
Engineering Resilience	'[...] measures the speed of recovery after the disturbance'[45] 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[3,15,18,24,34,45,129,132,137]. Also known as *Recovery*[132], *Stability*[18] or *Resiliency*[34].
Specified resilience	'The resilience of *what to what*.'[19] The resilience of a system or a specified component of a system to a hypothesized or foreseeable perturbation[34].
Cross-scale resilience	[...] 'explicitly considers the compartmentalization of ecological patterns and processes by spatial and temporal scales.'[34] Therefore, it takes into account that resilience in natural systems is augmented when traits are diverse on temporal and spatial scales. This concept incorporates that of *Response diversity* which 'emphasizes the variation in response to environmental change by species within a functional group within scales'[34].
Coerced resilience	Resilience in an ecosystem that is created and maintained through anthropogenic action[12]. This concept is especially useful in regards to production ecosystems[12]. However, this action of 'resilience maintenance' can lead to catastrophic collapses of ecosystems[3]

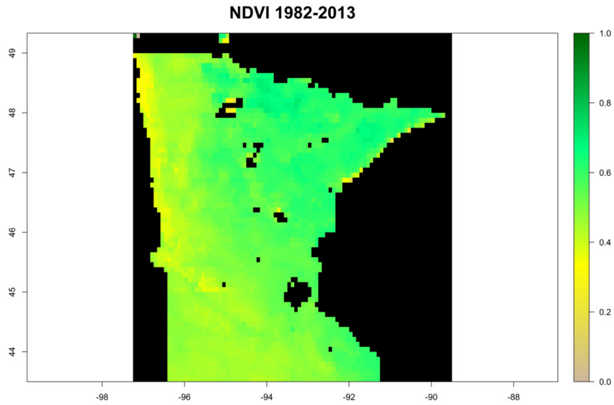
A.2 Biome Classification Schemes

Table A.4: Recent Biome Classification Studies: An overview of 31 biome classification studies. Where stated, the number of biomes specified in the respective study has been included (N_{Biomes}). However, for comparability, this has only been done for global scale studies.

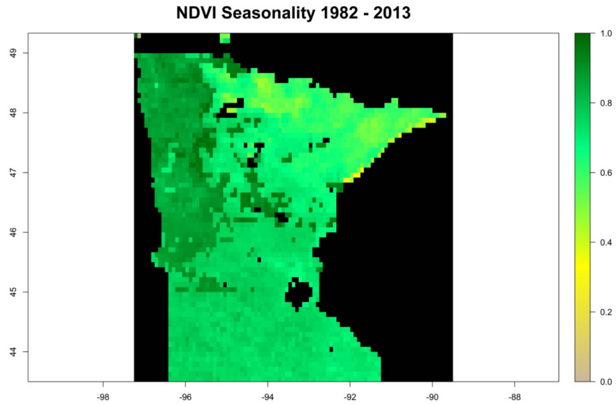
Reference study	Data type	Spatial scale	N_{Biomes}
Huete et. al[68]	climate data	Local	-
Bergengren et. al[77]	climate data	Global	12
Prentice et. al[71]	climate data	Global	15
Belotelov et. al[49]	climate data	Local	7
Friend et. al[92]	climate data	Global	-
Hughes et. al[56]	climate data	Global	5
Woodward et. al[63]	climate data	Global	10
Walter et. al[75]	climate data	Global	9
Holdridge et. al[138]	climate data	Global	19
Noever et.[139]	climate data	Global	18
Monserud et. al[140]	climate data, existing land cover classifications	Global	14
Higgins et. al[60]	climate data, remote sensing	Global	24
Bartholomé et. al[69]	remote sensing	Global	22
Running et. al[70]	remote sensing	Global	6
Lieng et. al[141]	remote sensing	Local	-
Ellis et. al[73]	remote sensing	Global	18
Friedl et. al[58]	remote sensing	Global	17
Roy et. al[81]	remote sensing	Local	-
Lotsch et. al[142]	remote sensing	Local	-
Mayaux et. al[143]	remote sensing	Local	-
Chen et. al[72]	remote sensing	Global	10
Colditz et. al[76]	remote sensing	Local	-
Latifovic et. al[40]	remote sensing	Local	-
Friedl et. al[61]	remote sensing	Global	16
Olson et. al[59]	expert knowledge	Global	14
Fairbanks et. al[74]	field records	Local	-
Xu et. al[144]	existing land cover sets	Global	16
Iwao et. al[123]	existing land cover sets	Global	14
Salzmann et. al[57]	palaeodata	Global	28
Kutzbach et. al[78]	palaeodata	Global	17
Crisp et. al[145]	taxonomic data	Global	7

A.3 Climatologies of Study Parameters for Study Regions

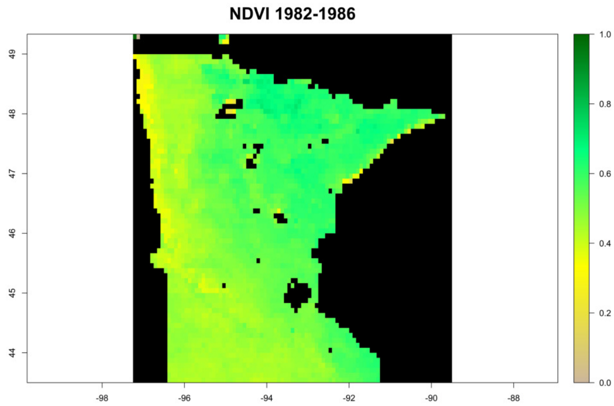
A.3.1 Minnesota



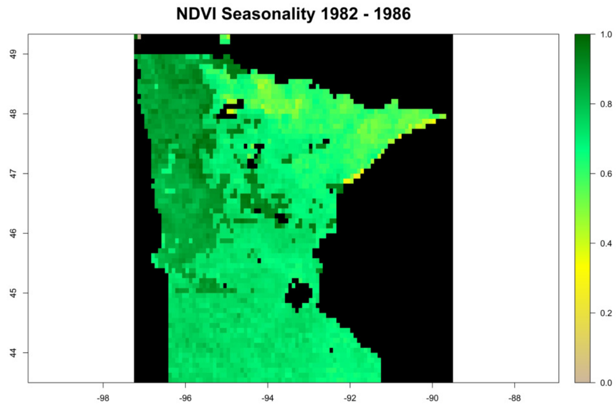
(a) NDVI 1982-2013



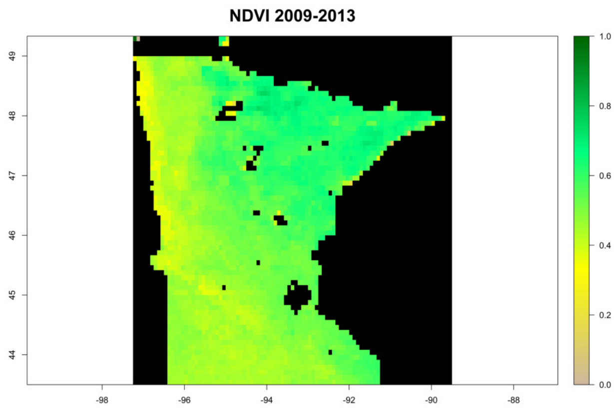
(b) NDVI Seasonality 1982-2013



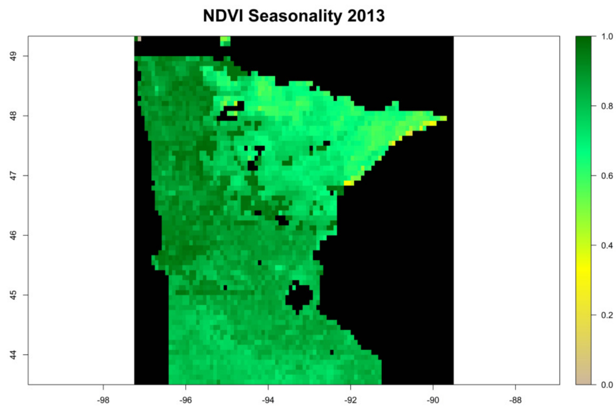
(c) NDVI 1982-1986



(d) NDVI Seasonality 1982-1986

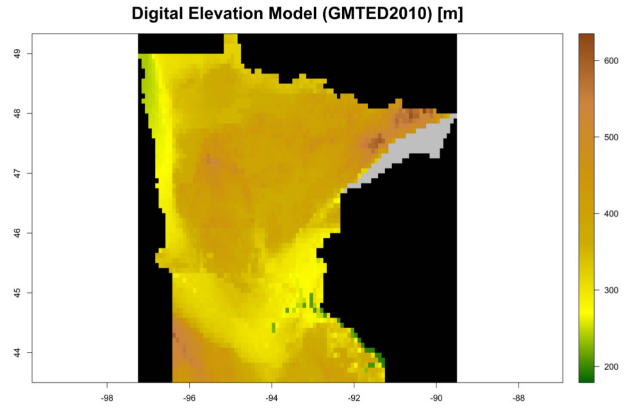


(e) NDVI 2009-2013

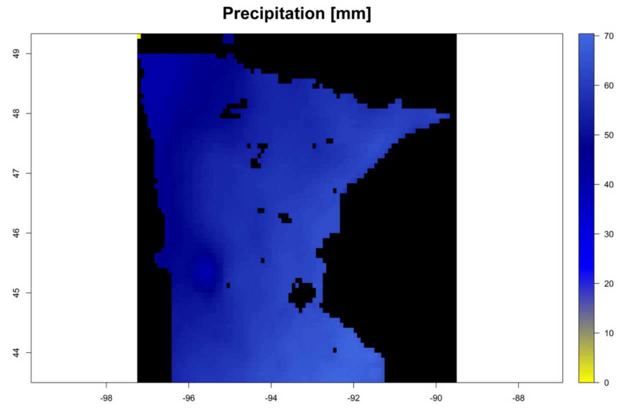


(f) NDVI Seasonality 2009-2013

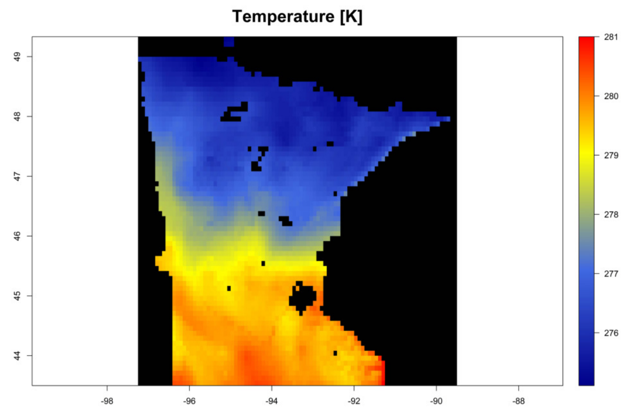
Figure A.1: NDVI Climatologies (Minnesota): Average composites of mean annual NDVI scores and annual NDVI seasonality scores have been calculated using the R-project for statistical computing. The data has been masked for urban areas and water bodies.



(a) DEM



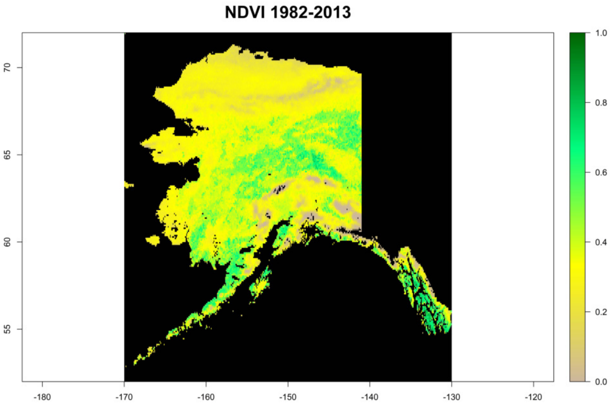
(b) Mean Annual Precipitation 1960-1990



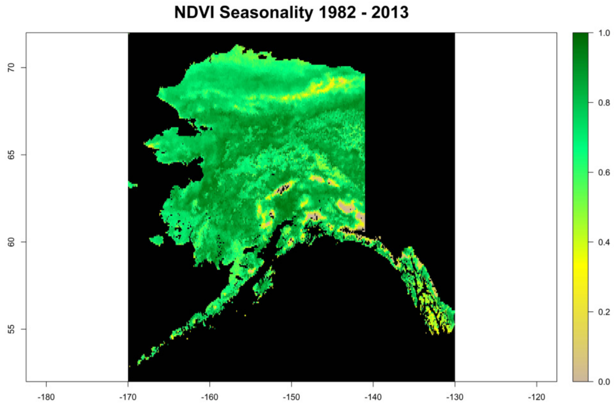
(c) Mean Annual Temperature 1960-1990

Figure A.2: Climate and DEM data (Minnesota): Grey shading in (a) depicts lake areas. The data has been masked for urban areas and water bodies.

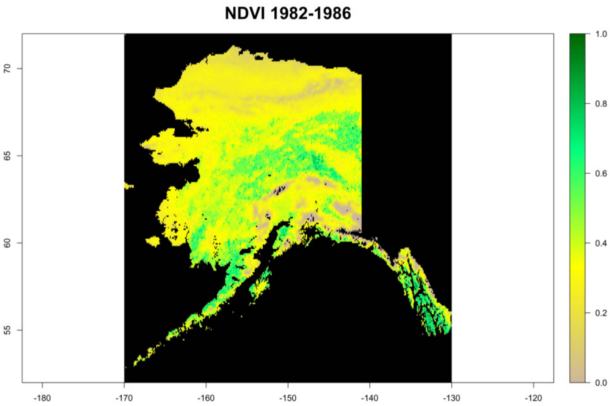
A.3.2 Alaska



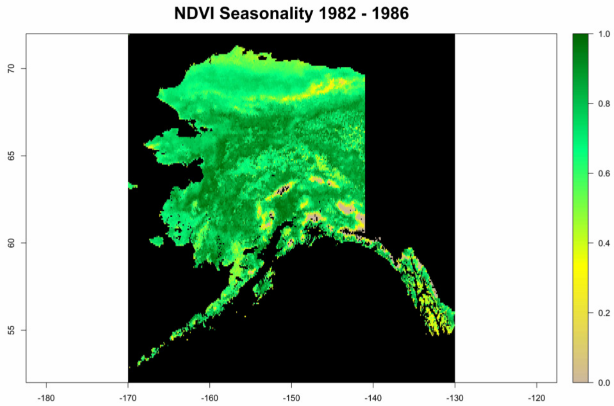
(a) NDVI 1982-2013



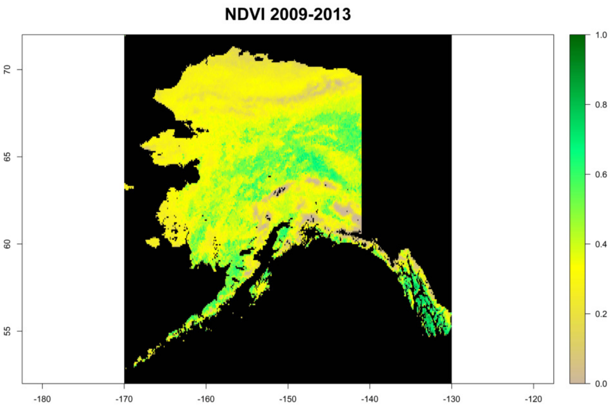
(b) NDVI Seasonality 1982-2013



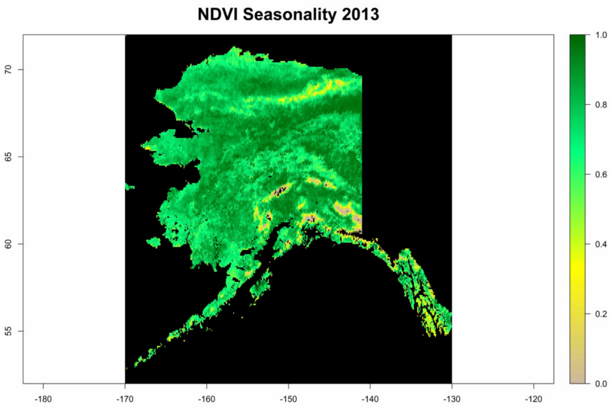
(c) NDVI 1982-1986



(d) NDVI Seasonality 1982-1986

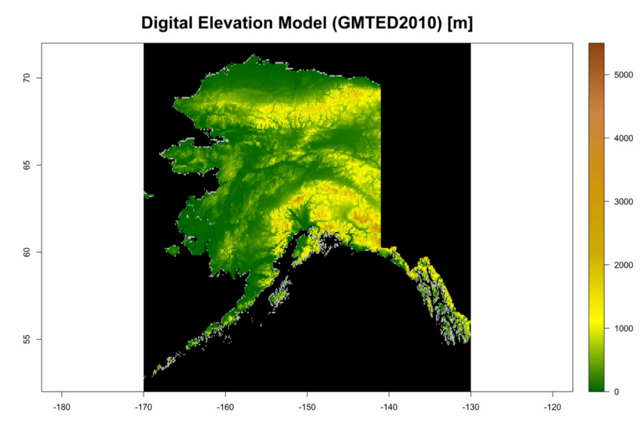


(e) NDVI 2009-2013

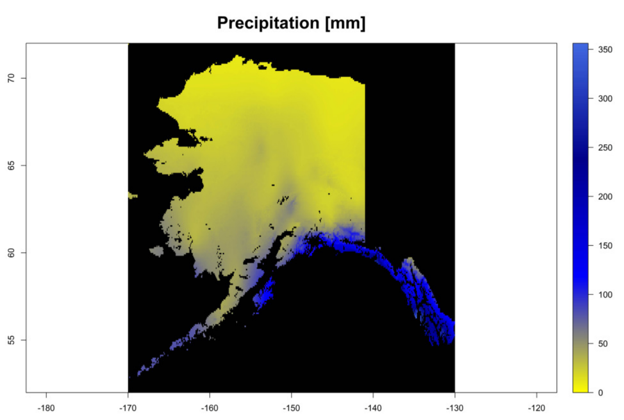


(f) NDVI Seasonality 2009-2013

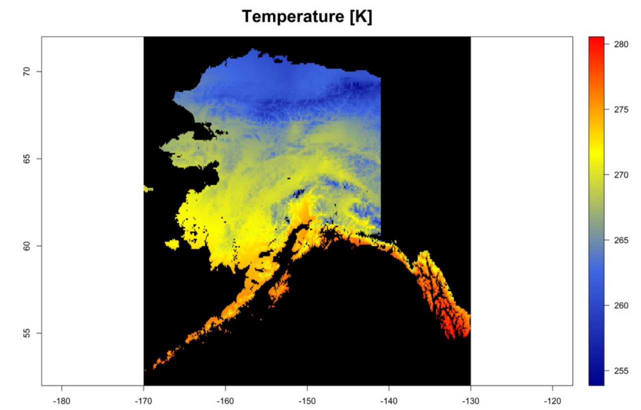
Figure A.3: NDVI Climatologies (Alaska): Average composites of mean annual NDVI scores and annual NDVI seasonality scores have been calculated using the R-project for statistical computing. The data has been masked for urban areas and water bodies.



(a) DEM



(b) Mean Annual Precipitation 1960-1990



(c) Mean Annual Temperature 1960-1990

Figure A.4: Climate and DEM data (Alaska): The data has been masked for urban areas and water bodies.

A.4 Collinearity

A.4.1 Minnesota

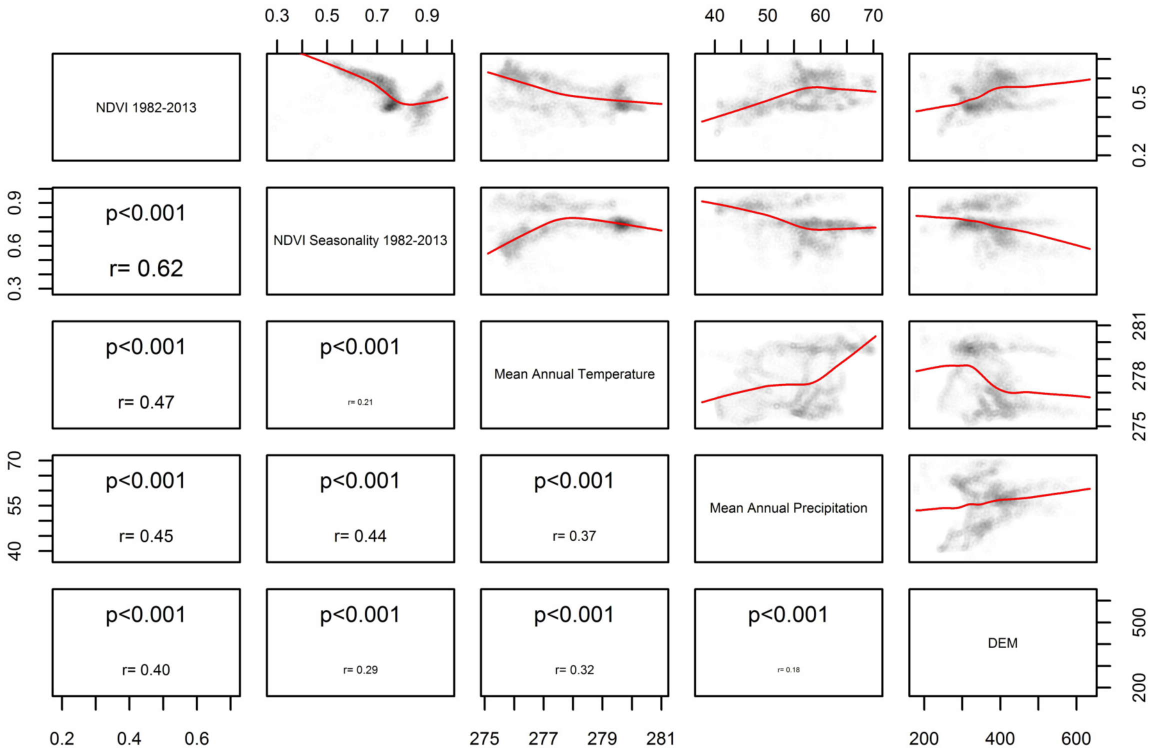


Figure A.5: Collinearity of study parameters (Minnesota): The above variables have been used for vegetation clustering as well as climate-vegetation interaction modelling.

Although mean annual NDVI and seasonality of monthly NDVI scores exerted strong collinearity of 0.62 at a statistically significant level of $p<0.001$ (see figure A.5), both parameters were included as they are important descriptors of vegetation composition and functioning and consequently proved useful in distinguishing stable states 1, 2 and 3 in the vegetation data clustering. All other parameters exhibited collinearity relationships which did not exceed 0.5 at $p<0.001$, which was set as the threshold beyond which the parameters had do be considered for transformation, combination or elimination from the clustering and modelling.
This threshold of $r=0.5$ is, admittedly, arbitrary in nature and an artefact of subjectiveness in the parameter selection process.

A.4.2 Alaska

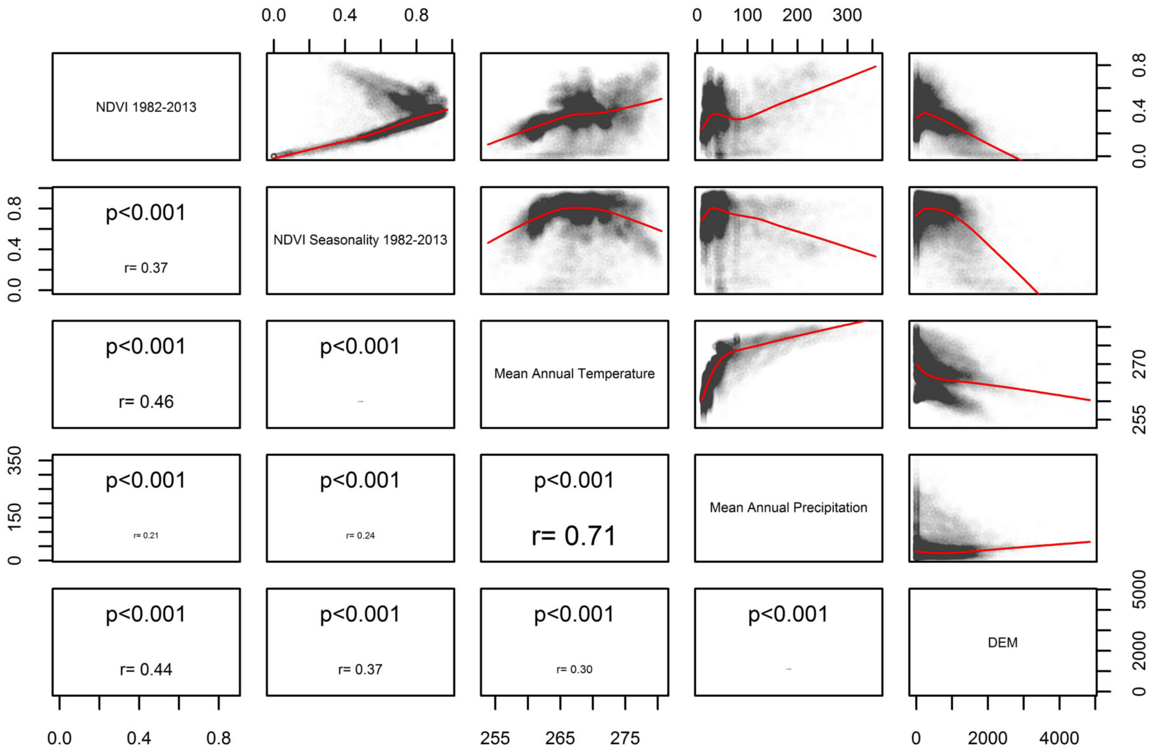


Figure A.6: Collinearity of study parameters (Alaska): The above variables have been used for vegetation clustering as well as climate-vegetation interaction modelling.

Except for mean annual temperature and mean annual precipitation, all study parameters exerted no statistically significant collinearity scores (p<0.001) exceeding the pre-defined threshold of r=0.5 (see figure A.6). Mean annual temperature and mean annual precipitation were included into the approach to the encompassing goal of this thesis due to their hypothesized influence on biome distribution. Conclusively, multiple response functions of biomes in Alaska were dependent on either mean annual precipitation or temperature (see figures A.22d and A.22e for examples).

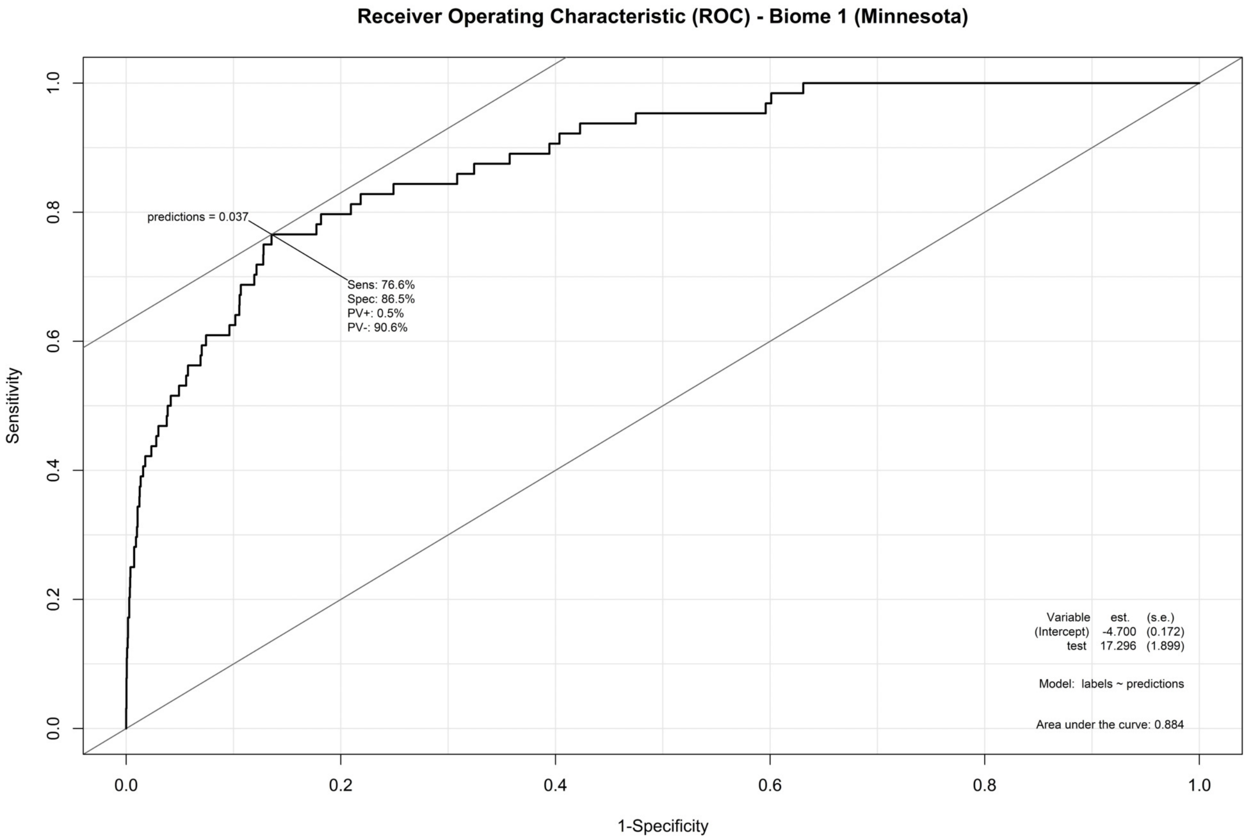
A.5 General Additive Modelling Results

A GAM for each individual stable state delineated in this study was calculated based on binary presence/absence data of the stable state in question (see figure A.7b for an example) and climate as well as altitude data (see figures A.2 and A.4).
The ROC method for identifying model-specific, non subjective probability thresholds above which to assume presence of a particular stable state has been realized using the *ROC* function of the Epi-package in R which resulted in a ROC curve (see figure A.7a for an example) for each GAM. This function was also used to produce the AUC measurement, which is indicative of model performance, for each GAM (included in the bottom right of the ROC plots).
For a deeper understanding of vegetation dynamics, a time series of monthly mean NDVI scores has been plotted for each stable state (see figure A.7c for an example). These plots included blue lines to indicate yearly time periods.

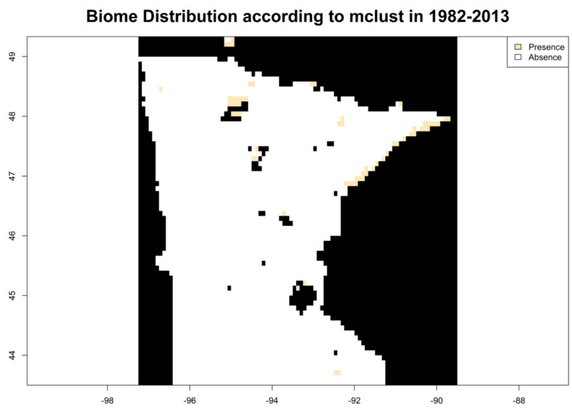
Exerting the state specific GAM on the climate and altitude data of the respective study site resulted in a map of state-specific occurrence probabilities (see figure A.8a for an example). The higher the probability, the more likely the particular stable state is to occur.
The occurrence probabilities have been re-plotted as histograms (see figure A.8b for an example) to which the ROC threshold has been added as a blue, vertical line. This threshold is also included in the upper left of the ROC plot of each individual GAM.
Turning the mapped occurrence probabilities into binary presence indicators by turning every probability $\geq$ ROC threshold into an indicator of state presence resulted in maps of predicted occurrence data (see figure A.8c for an example). These maps were used for the delineation of possible state distribution overlaps within the resilience assessment.
Response functions of individual stable states GAMs have been assessed as seen in figures A.8d, A.8e and A.8f, for example, to identify the ecological traits of individual stable states. Whilst analysing these plots for relationships and thresholds of occurrence probability of a given stable state, most care was exerted on the loess-smoother (green line) which showed the general trend instead of the very detailed response curve itself.

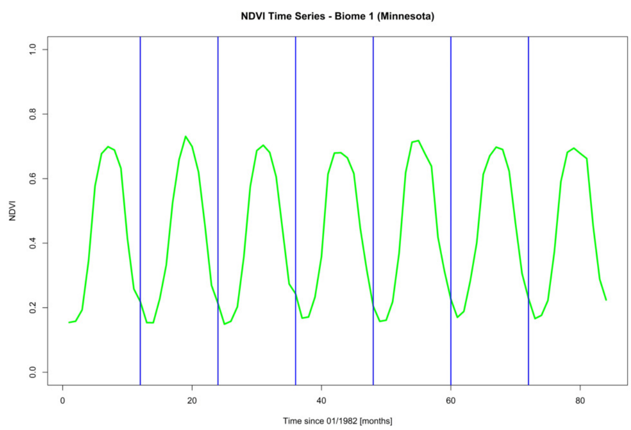
A.5.1 Minnesota

A.5.1.1 Stable State 1 - Shrub



(a) ROC curve of GAM

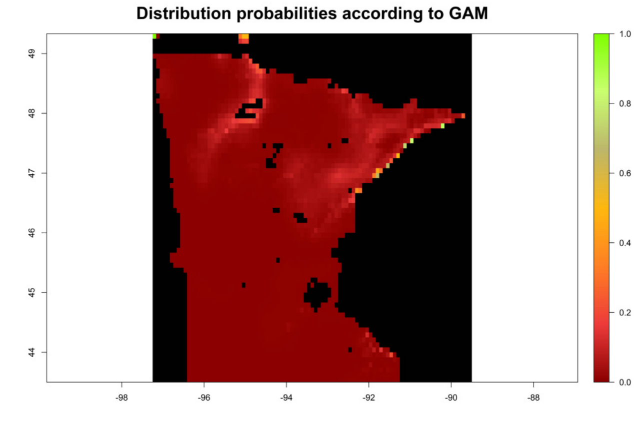




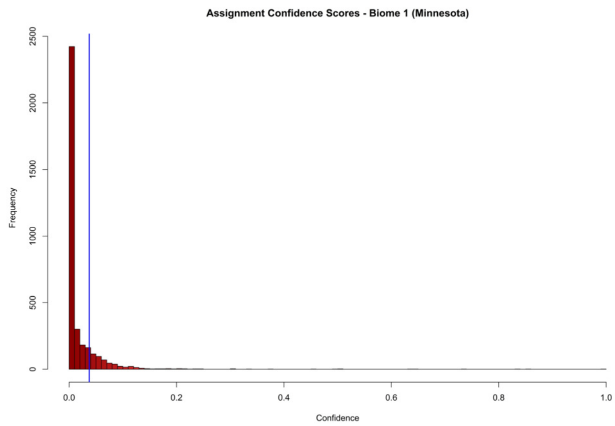
(b) Actual distribution

(c) NDVI time series

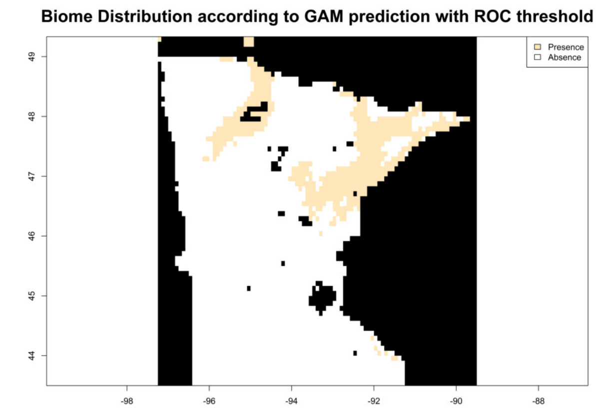
Figure A.7: GAM characteristics of stable state 1 (Minnesota): The ROC curve for the GAM of stable state 1 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



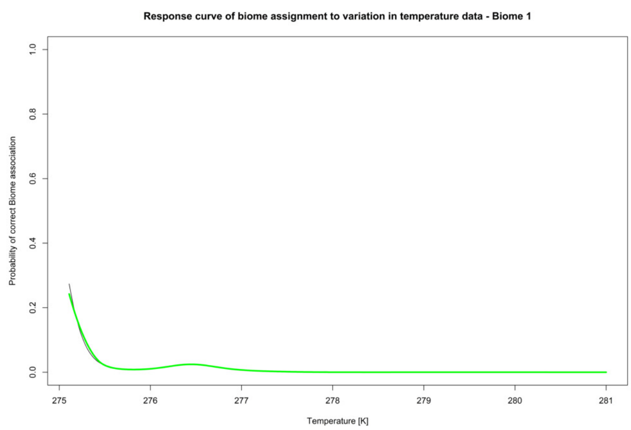
(a) Biome prediction probabilities



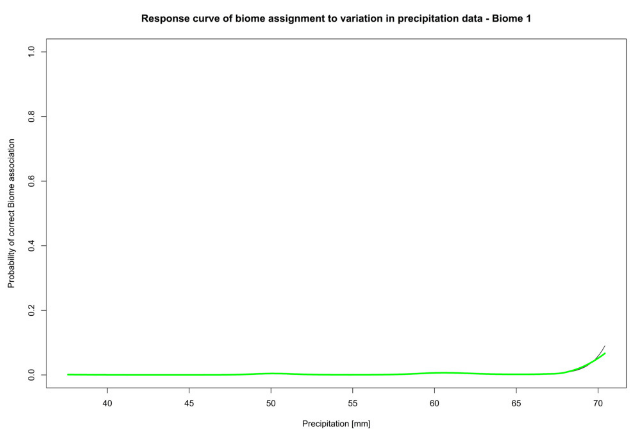
(b) Prediction probabilities with ROC threshold



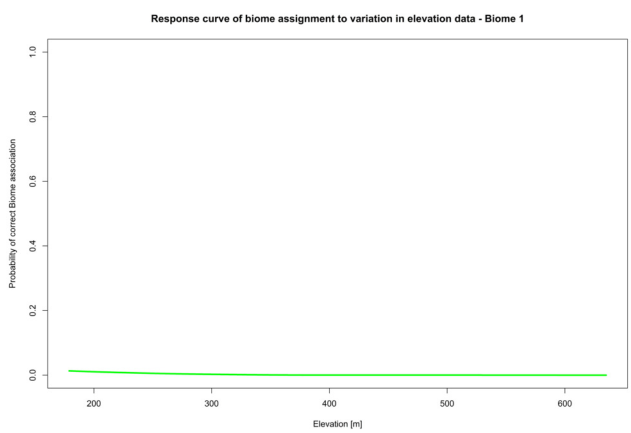
(c) Prediction of occurrence



(d) Response function (temperature)



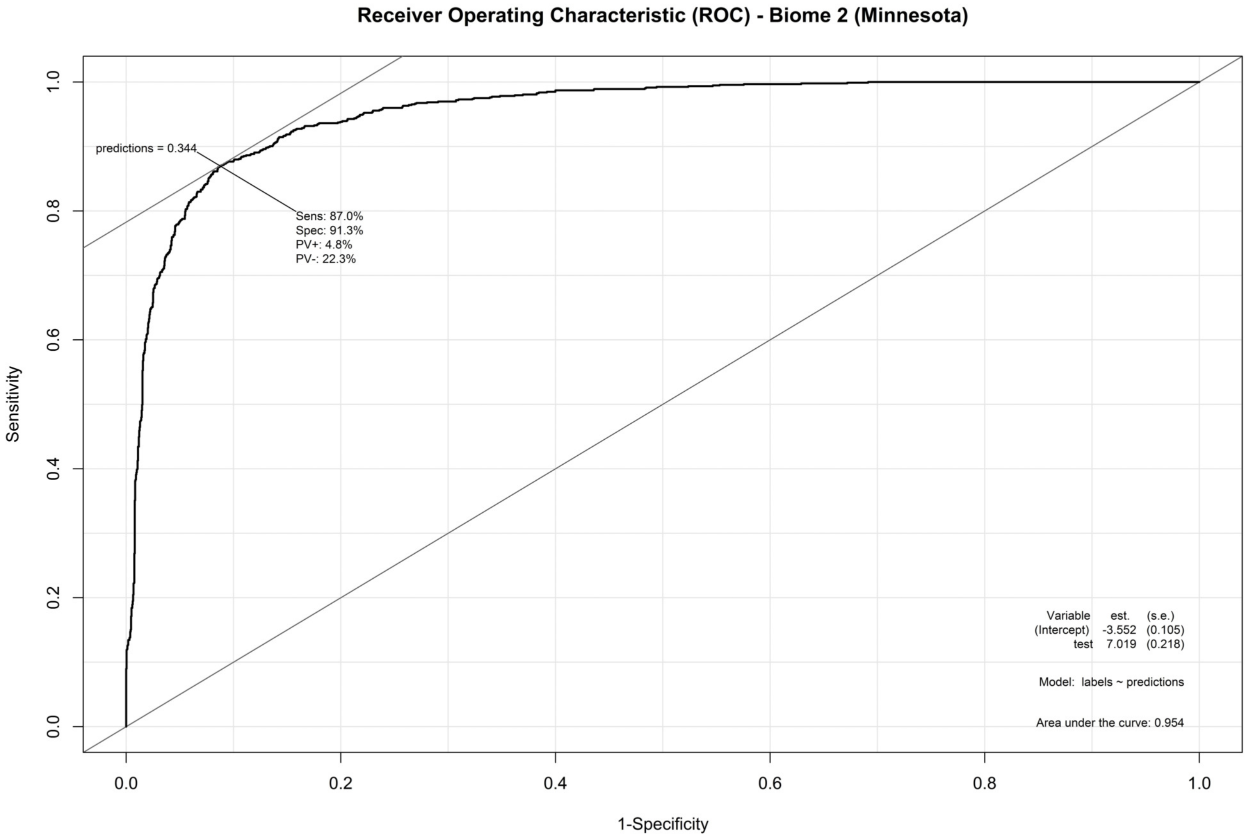
(e) Response function (precipitation)



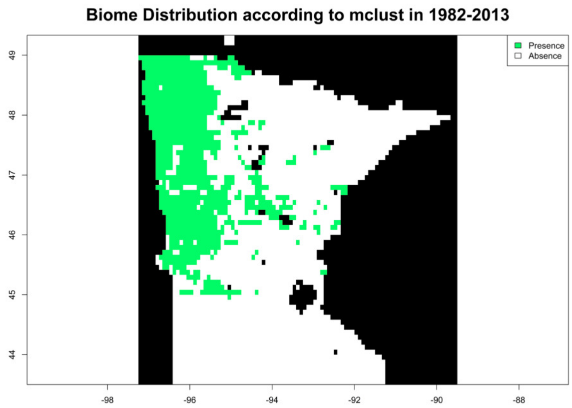
(f) Response function (DEM data)

Figure A.8: GAM characteristics of stable state 1 (Minnesota): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 1 and climate-vegetation interactions.

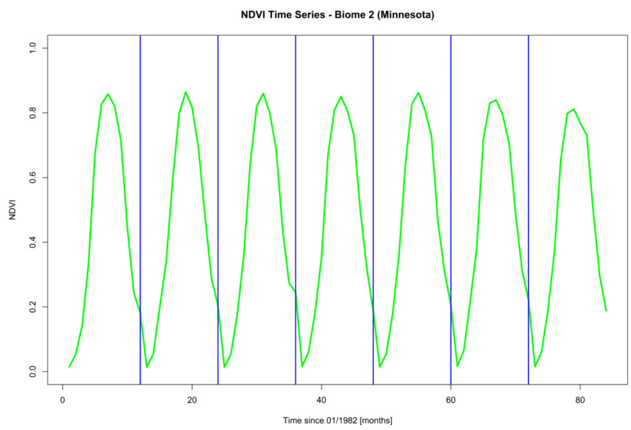
Stable State 2 - Prairie



(a) ROC curve of GAM

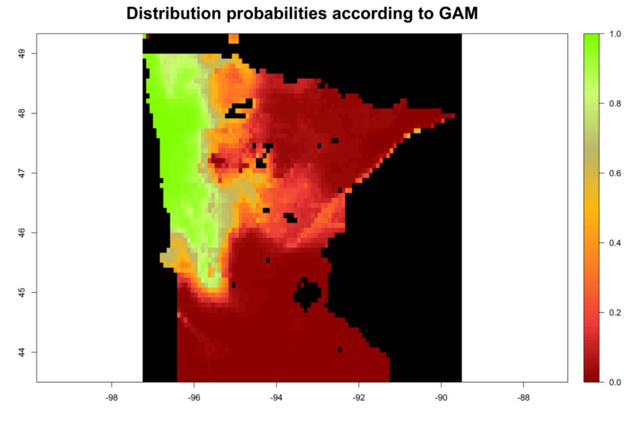


(b) Actual distribution

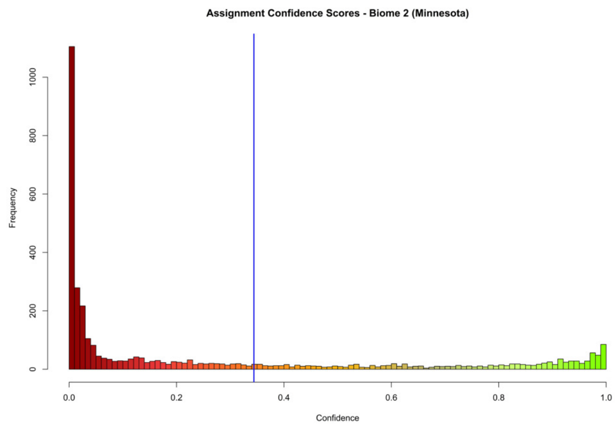


(c) NDVI time series

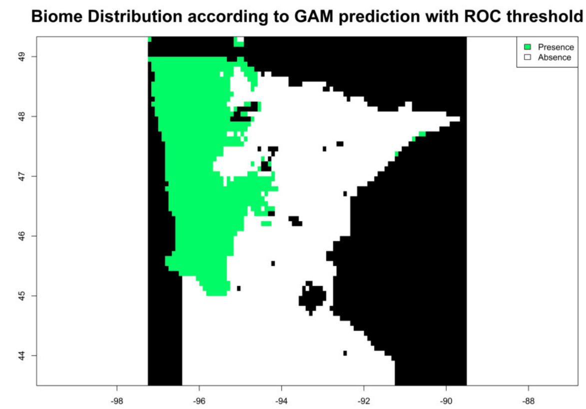
Figure A.9: GAM characteristics of stable state 2 (Minnesota): The ROC curve for the GAM of stable state 2 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



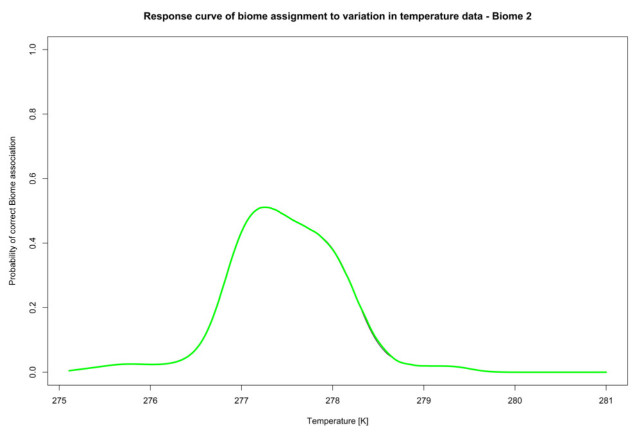
(a) Biome prediction probabilities



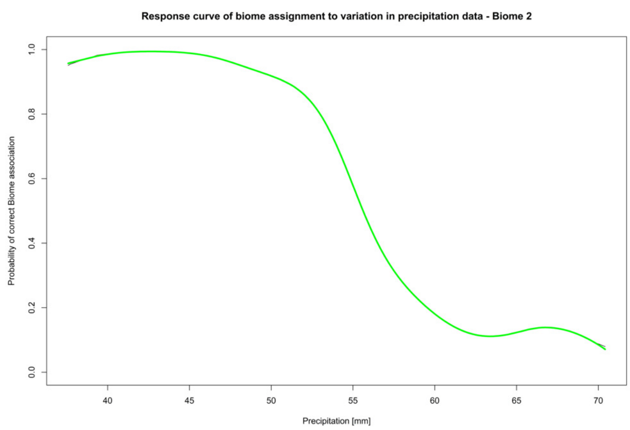
(b) Prediction probabilities with ROC threshold



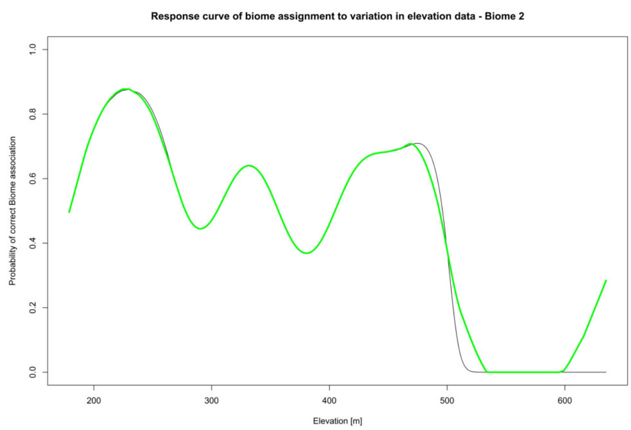
(c) Prediction of occurrence



(d) Response function (temperature)



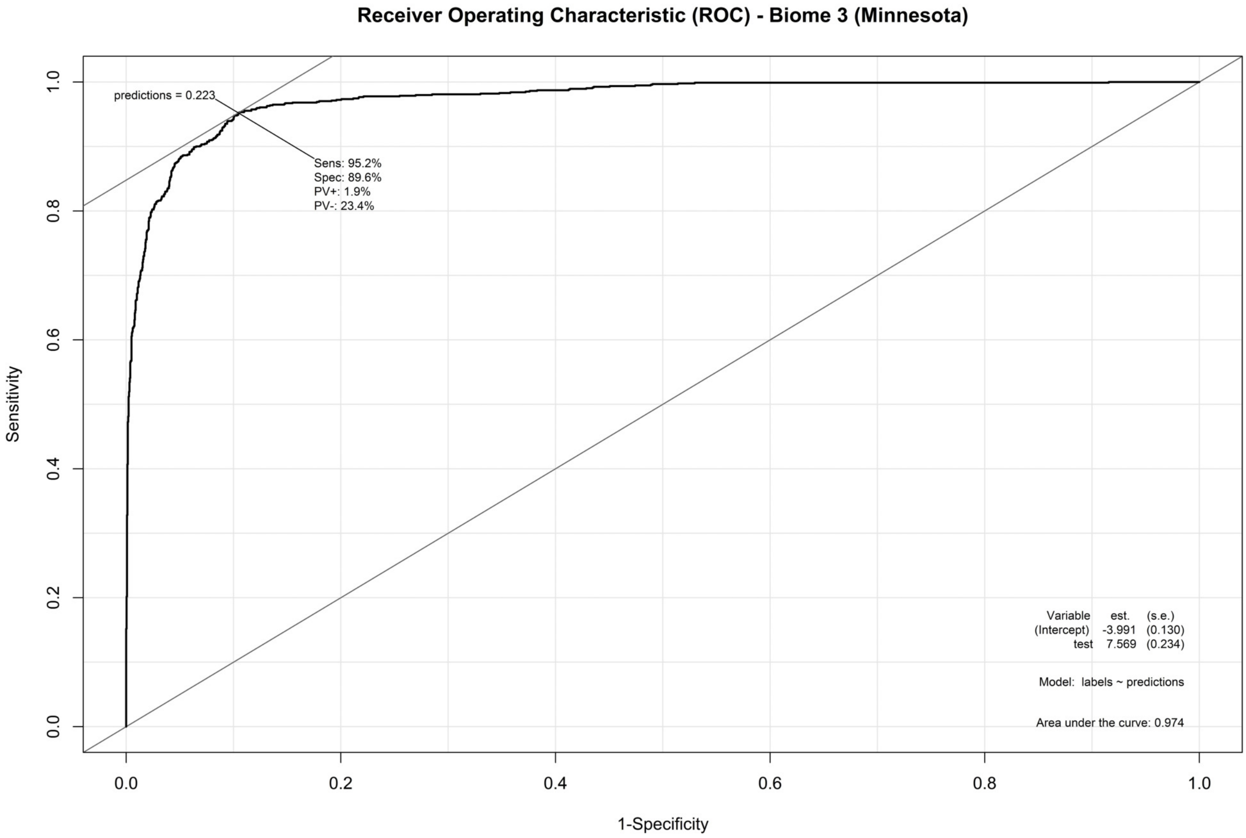
(e) Response function (precipitation)



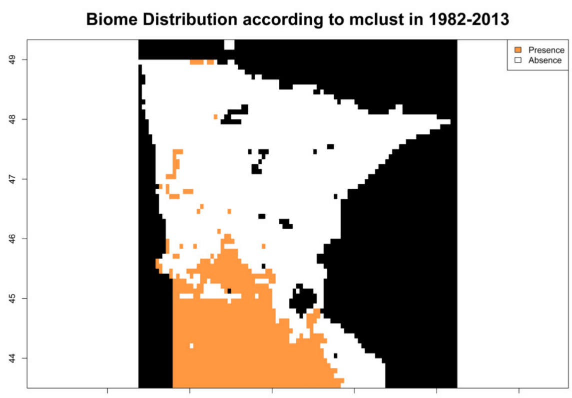
(f) Response function (DEM data)

Figure A.10: GAM characteristics of stable state 2 (Minnesota): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 2 and climate-vegetation interactions.

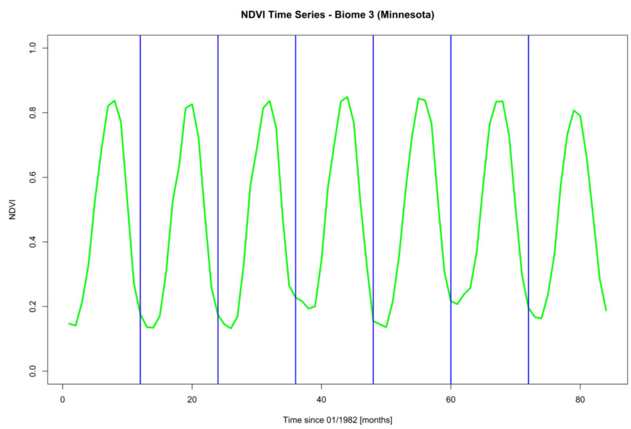
Stable State 3 - Agricultural Land Use



(a) ROC curve of GAM

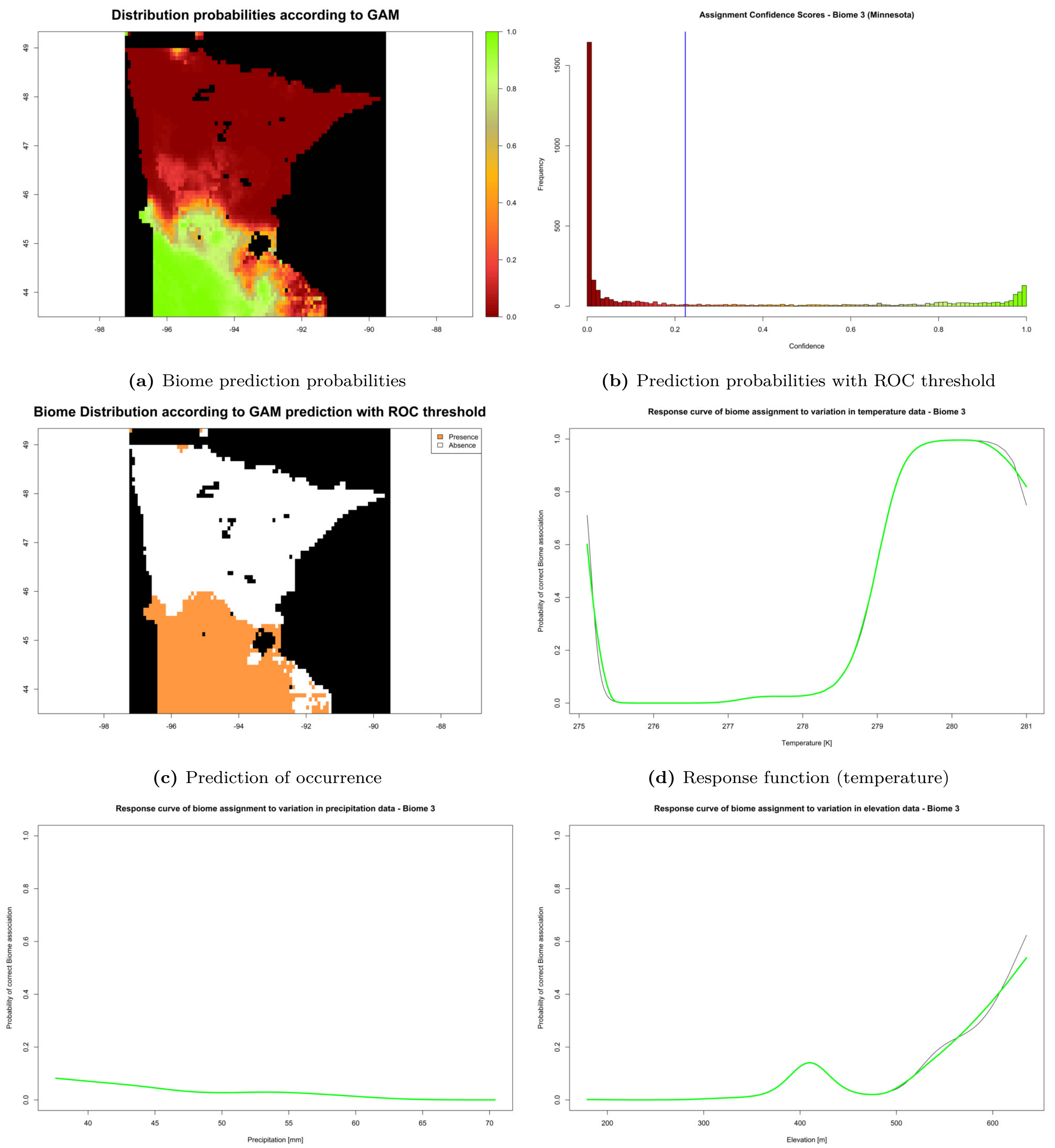


(b) Actual distribution



(c) NDVI time series

Figure A.11: GAM characteristics of stable state 3 (Minnesota): The ROC curve for the GAM of stable state 3 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



(a) Biome prediction probabilities

(b) Prediction probabilities with ROC threshold

(c) Prediction of occurrence

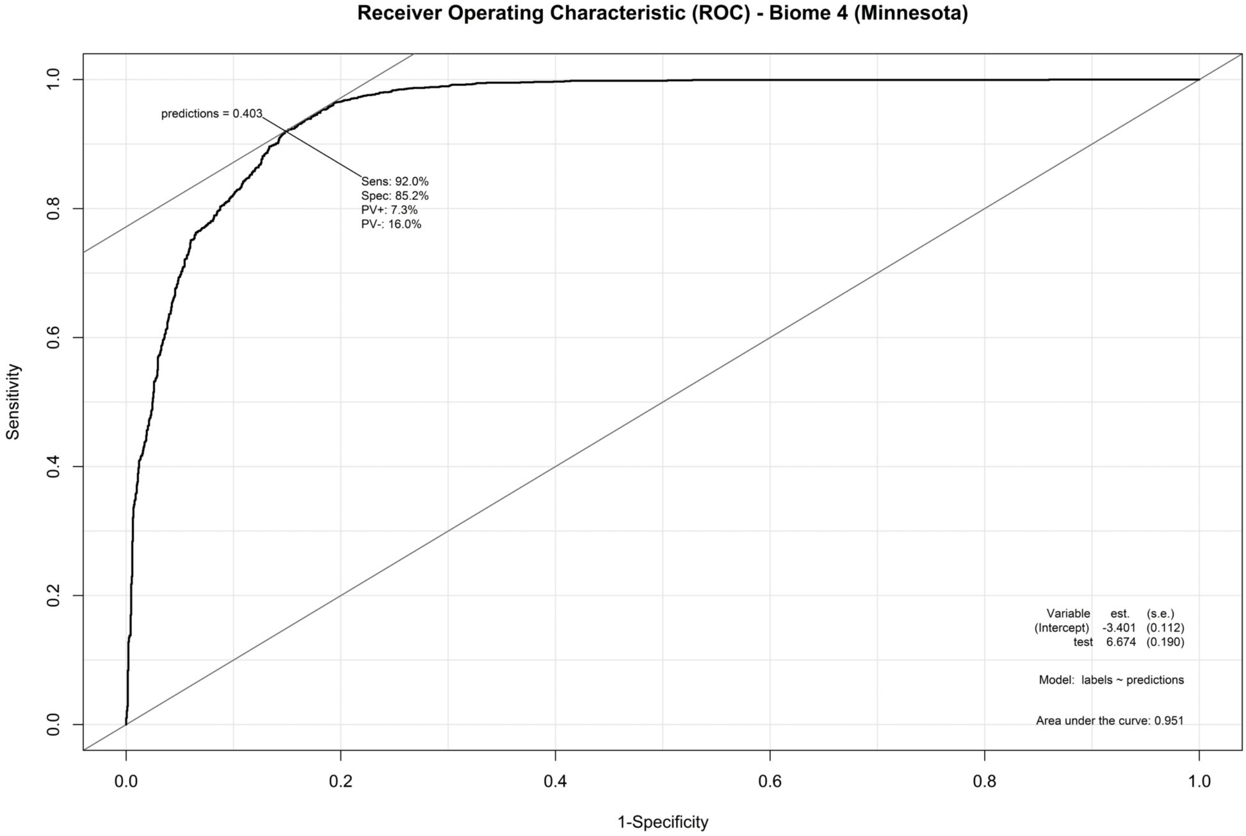
(d) Response function (temperature)

(e) Response function (precipitation)

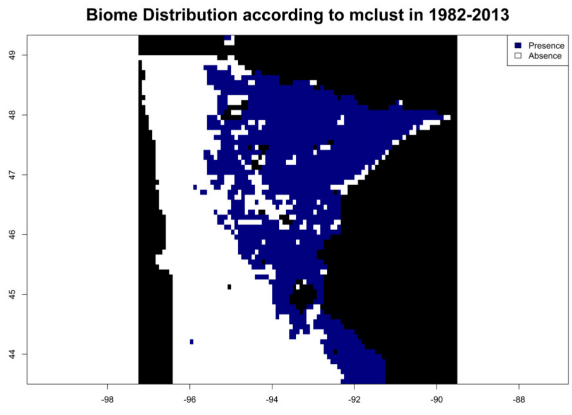
(f) Response function (DEM data)

Figure A.12: GAM characteristics of stable state 3 (Minnesota): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 3 and climate-vegetation interactions.

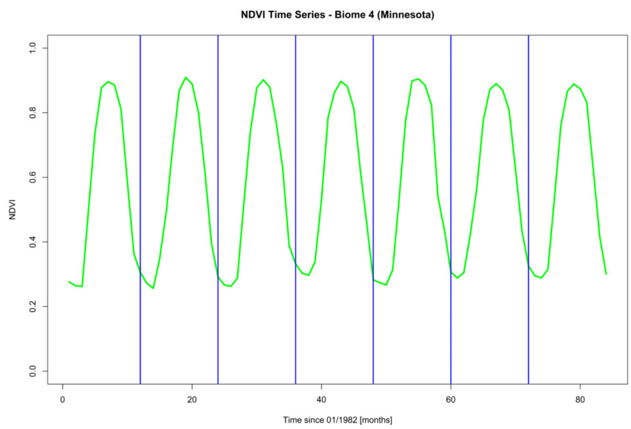
Stable State 4 - Mixed Forest



(a) ROC curve of GAM

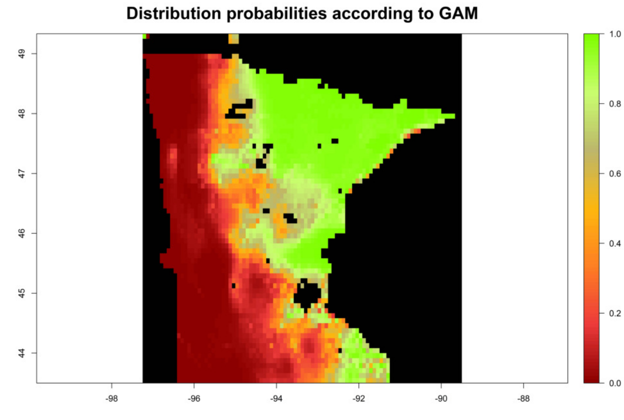


(b) Actual distribution

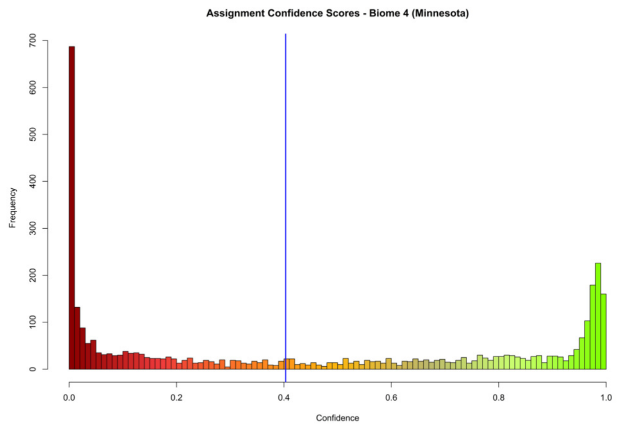


(c) NDVI time series

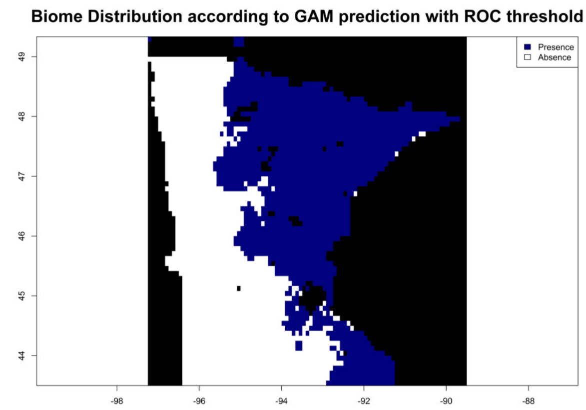
Figure A.13: GAM characteristics of stable state 4 (Minnesota): The ROC curve for the GAM of stable state 4 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



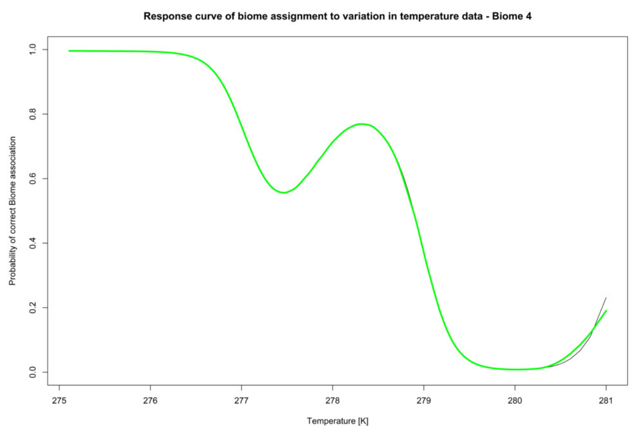
(a) Biome prediction probabilities



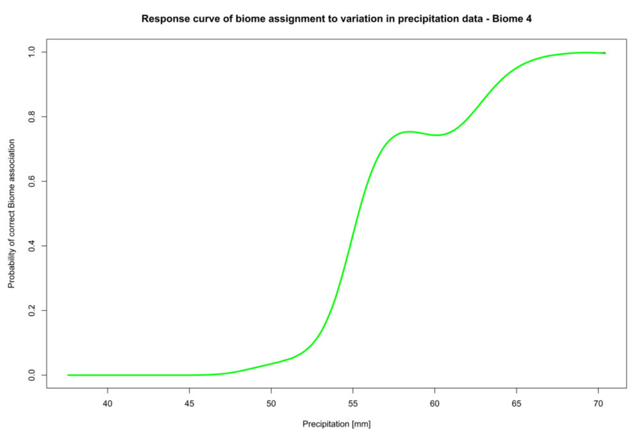
(b) Prediction probabilities with ROC threshold



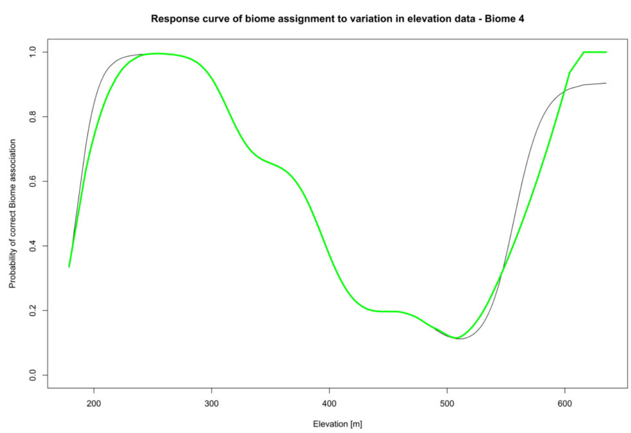
(c) Prediction of occurrence



(d) Response function (temperature)



(e) Response function (precipitation)

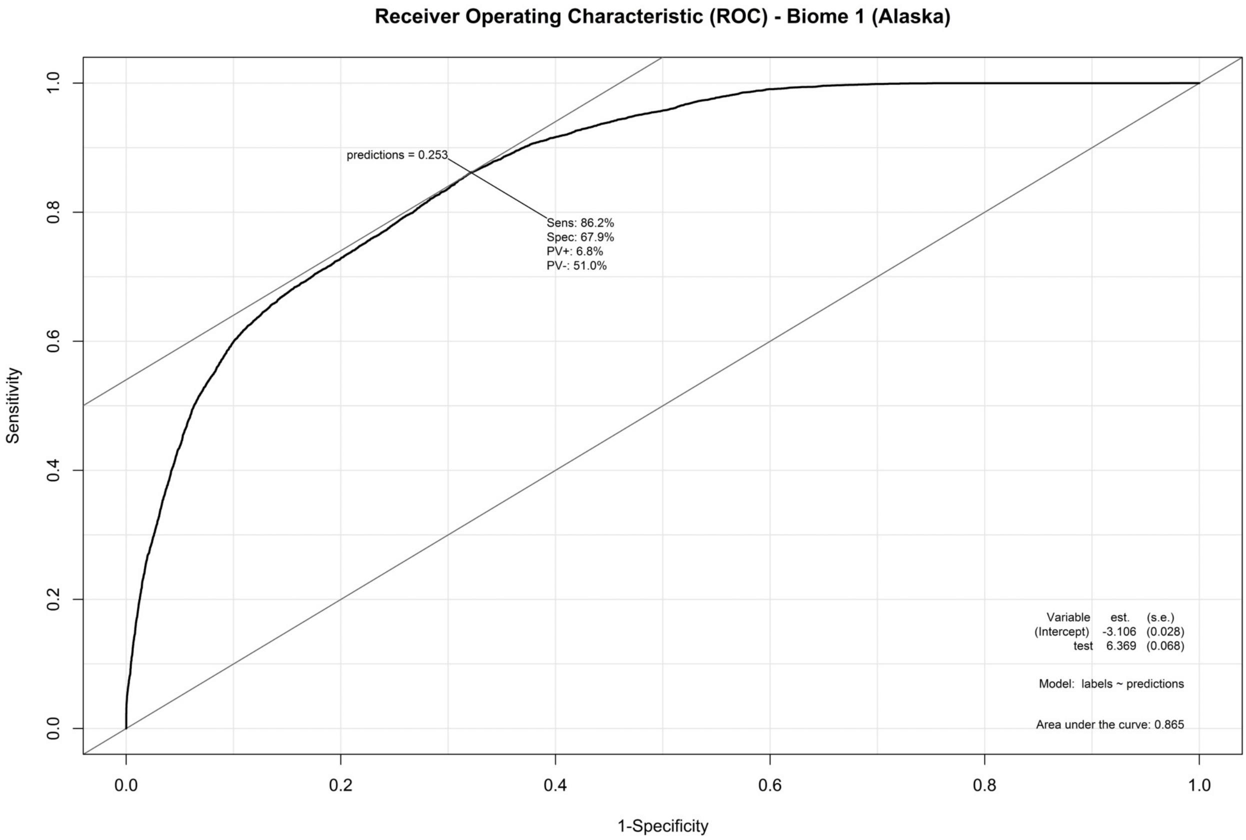


(f) Response function (DEM data)

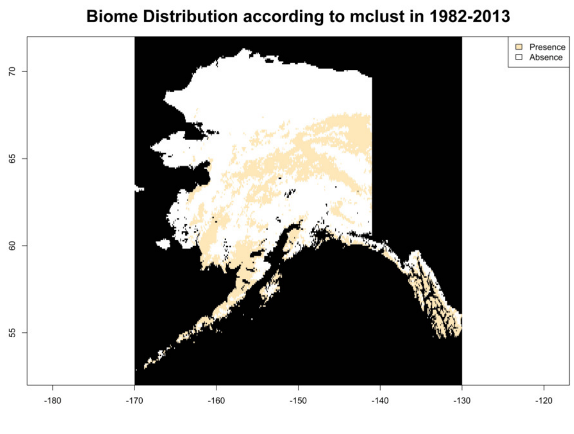
Figure A.14: GAM characteristics of stable state 4 (Minnesota): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 4 and climate-vegetation interactions.

A.5.2 Alaska

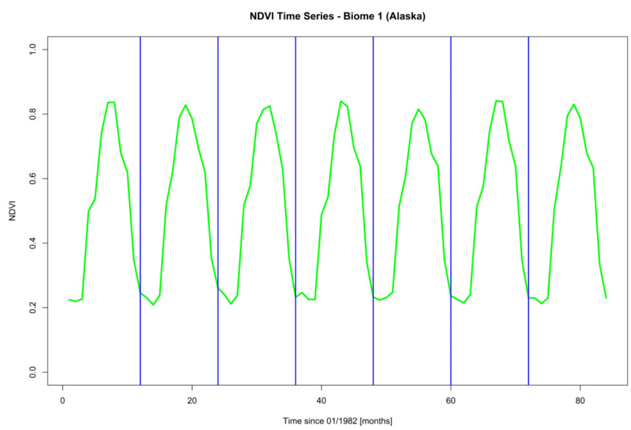
A.5.2.1 Stable State 1 - Boreal Forest



(a) ROC curve of GAM

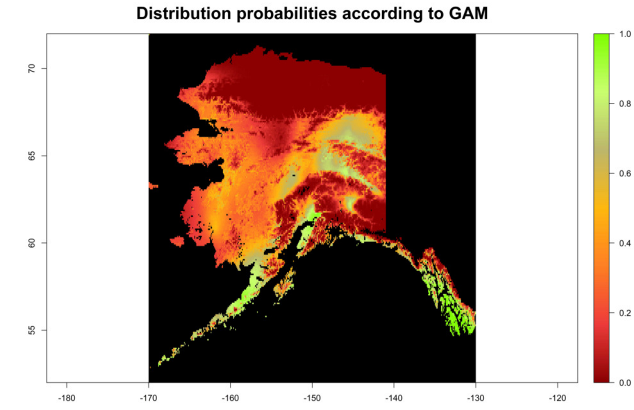


(b) Actual distribution

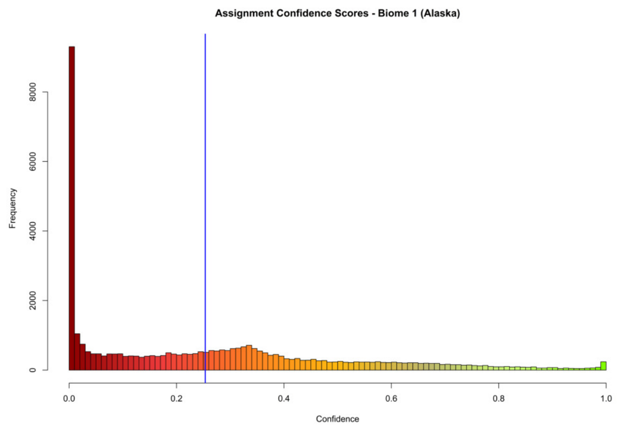


(c) NDVI time series

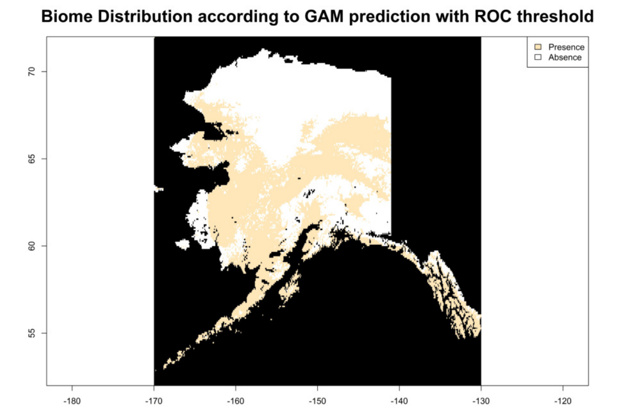
Figure A.15: GAM characteristics of stable state 1 (Alaska): The ROC curve for the GAM of stable state 1 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



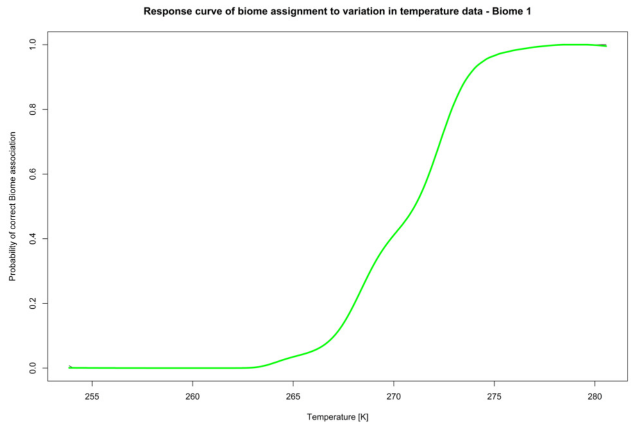
(a) Biome prediction probabilities



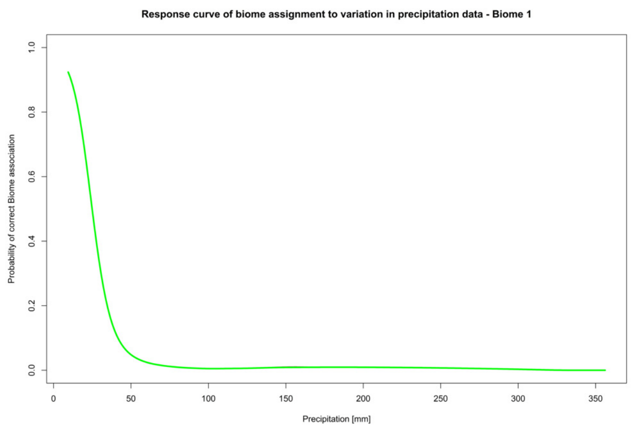
(b) Prediction probabilities with ROC threshold



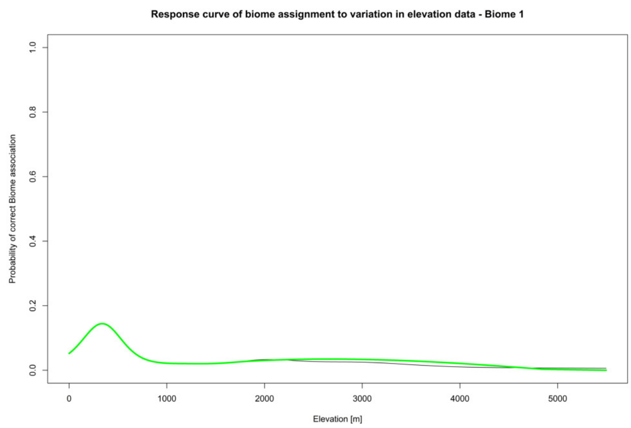
(c) Prediction of occurrence



(d) Response function (temperature)



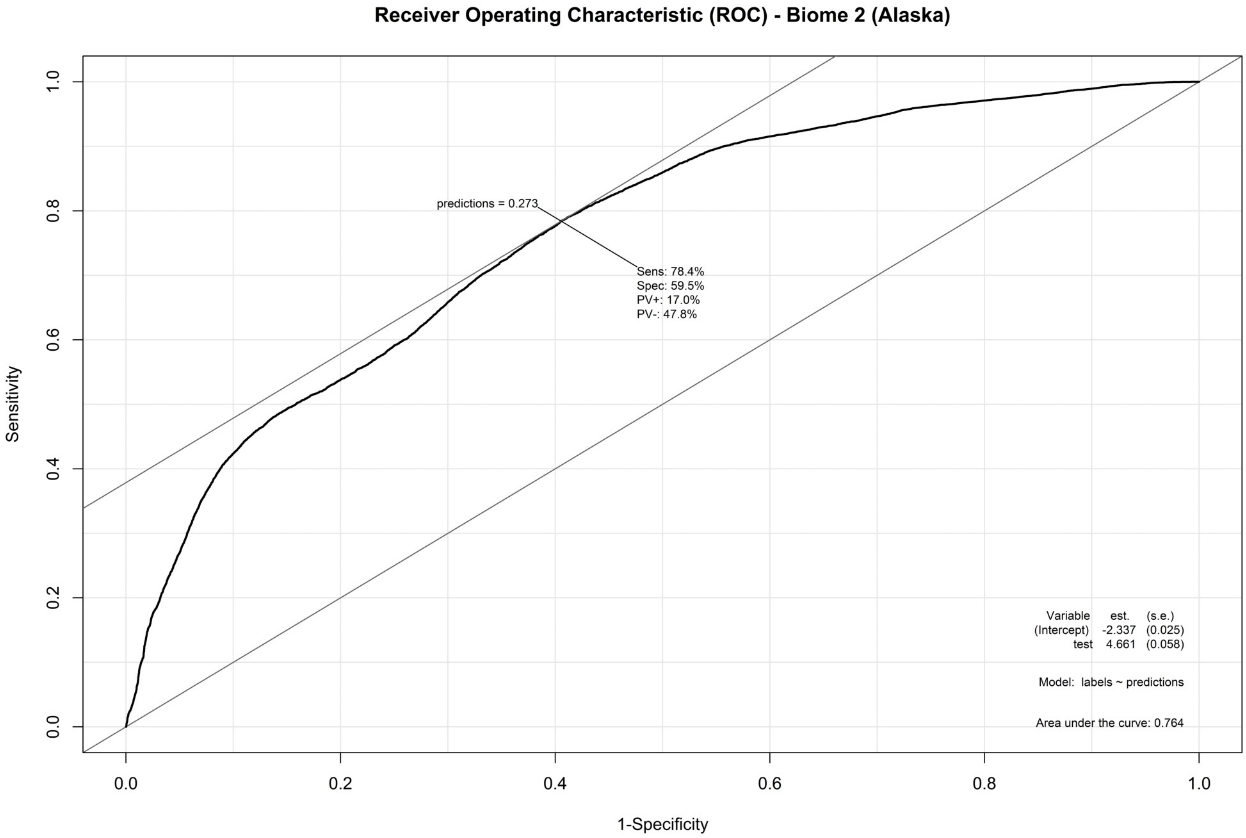
(e) Response function (precipitation)



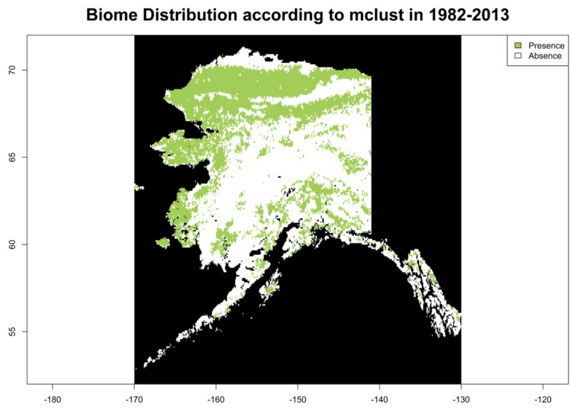
(f) Response function (DEM data)

Figure A.16: GAM characteristics of stable state 1 (Alaska): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 1 and climate-vegetation interactions.

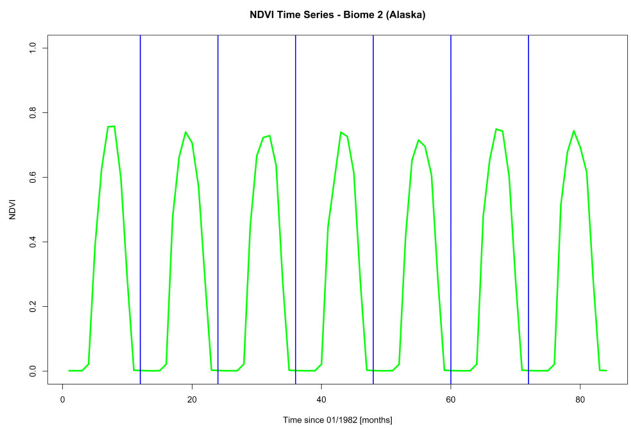
A.5.2.2 Stable State 2 - Dwarf Shrub



(a) ROC curve of GAM

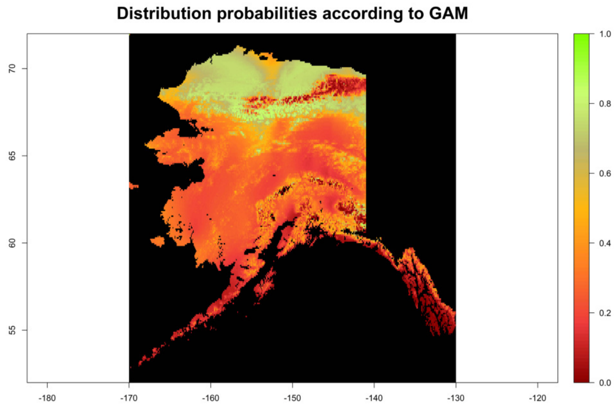


(b) Actual distribution

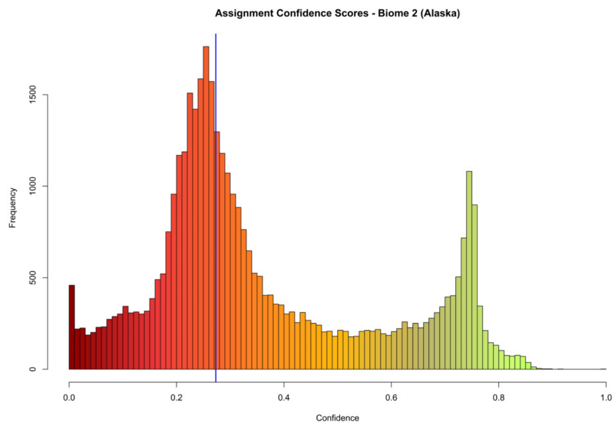


(c) NDVI time series

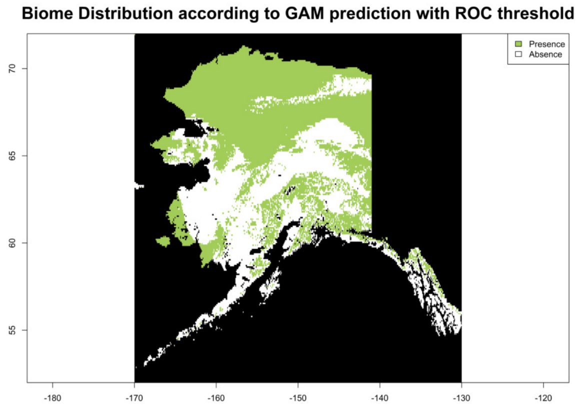
Figure A.17: GAM characteristics of stable state 2 (Alaska): The ROC curve for the GAM of stable state 2 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



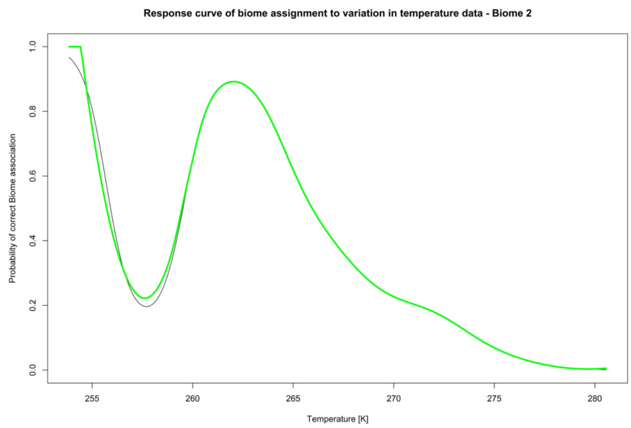
(a) Biome prediction probabilities



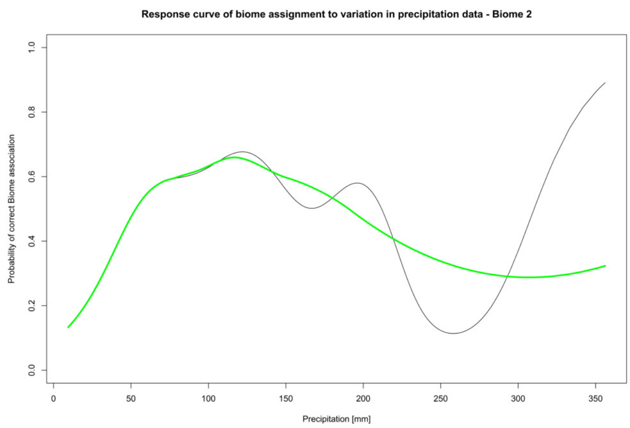
(b) Prediction probabilities with ROC threshold



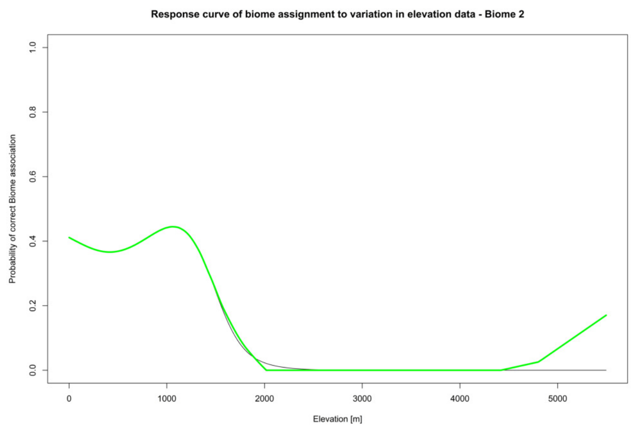
(c) Prediction of occurrence



(d) Response function (temperature)



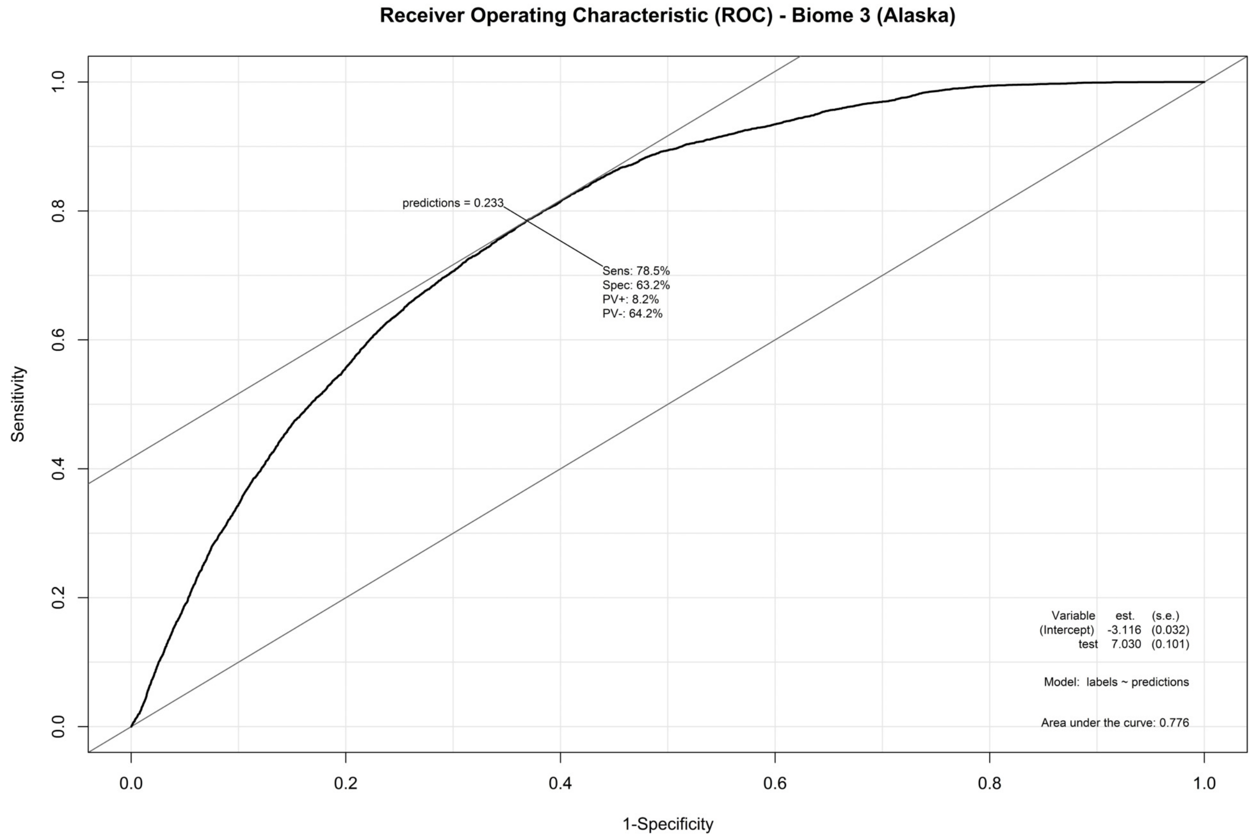
(e) Response function (precipitation)



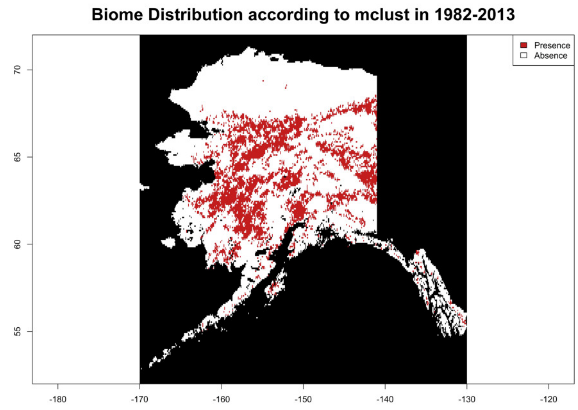
(f) Response function (DEM data)

Figure A.18: GAM characteristics of stable state 2 (Alaska): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 2 and climate-vegetation interactions.

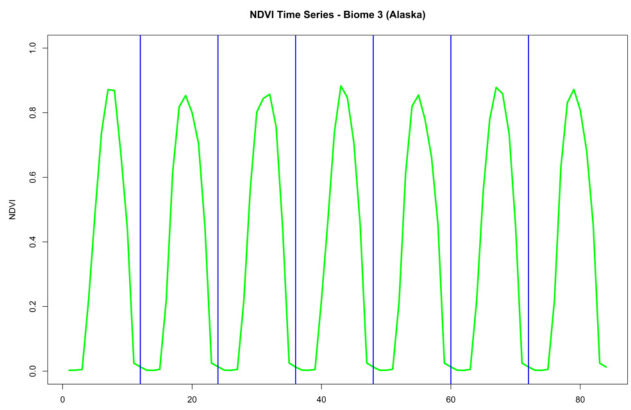
A.5.2.3 Stable State 3 - Shrub



(a) ROC curve of GAM

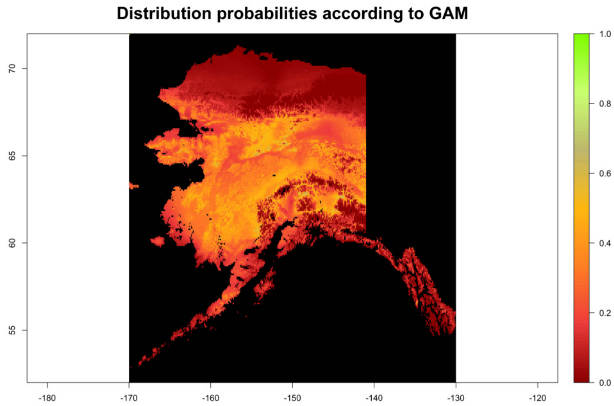


(b) Actual distribution

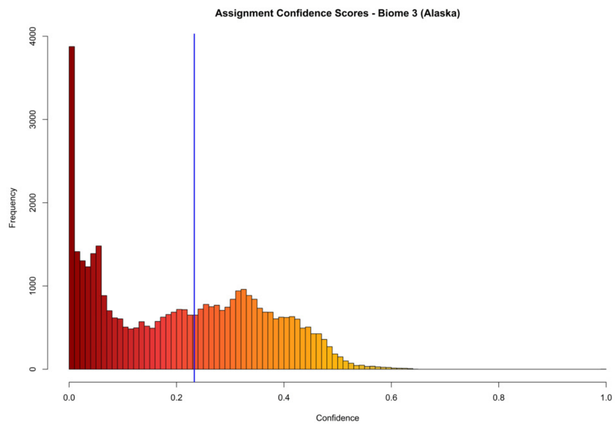


(c) NDVI time series

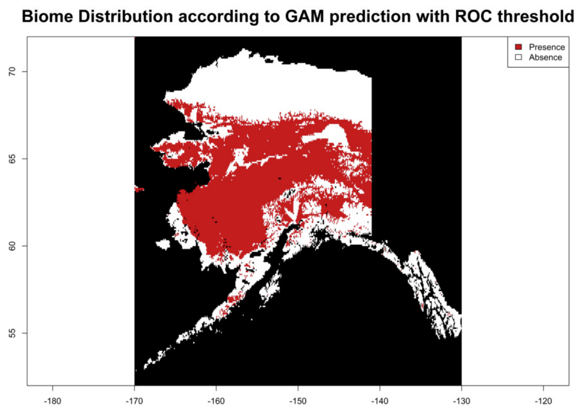
Figure A.19: GAM characteristics of stable state 3 (Alaska): The ROC curve for the GAM of stable state 3 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



(a) Biome prediction probabilities

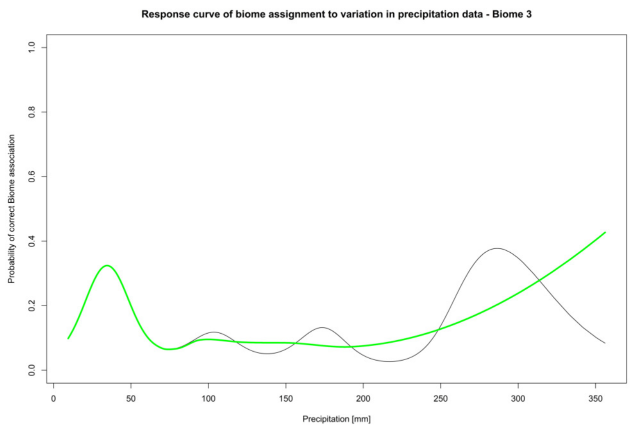


(b) Prediction probabilities with ROC threshold

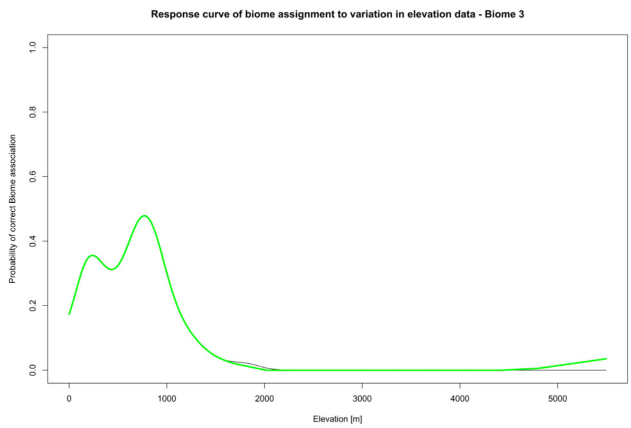


(c) Prediction of occurrence

(d) Response function (temperature)



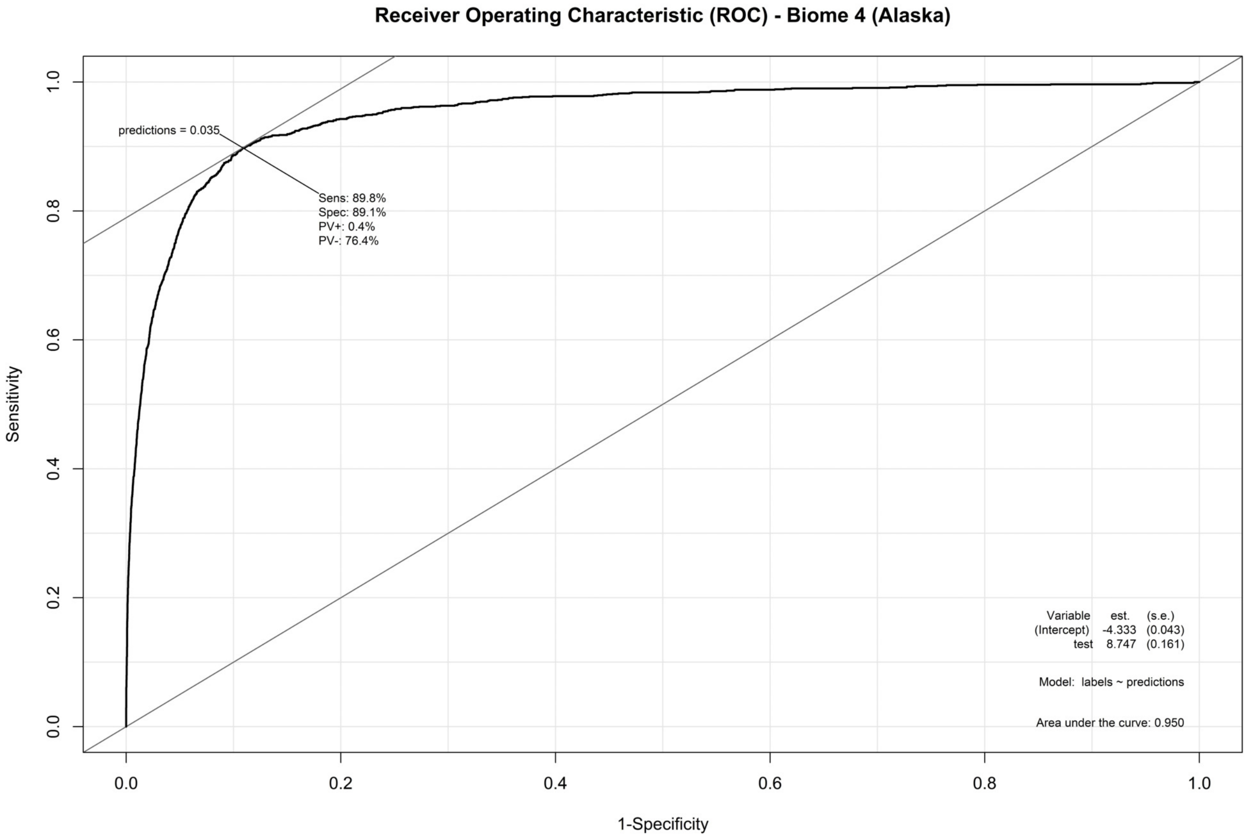
(e) Response function (precipitation)



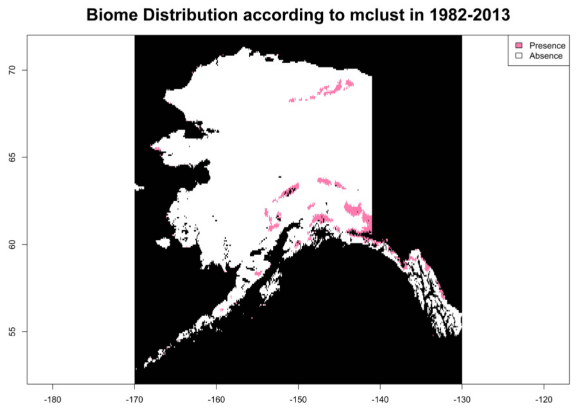
(f) Response function (DEM data)

Figure A.20: GAM characteristics of stable state 3 (Alaska): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 3 and climate-vegetation interactions.

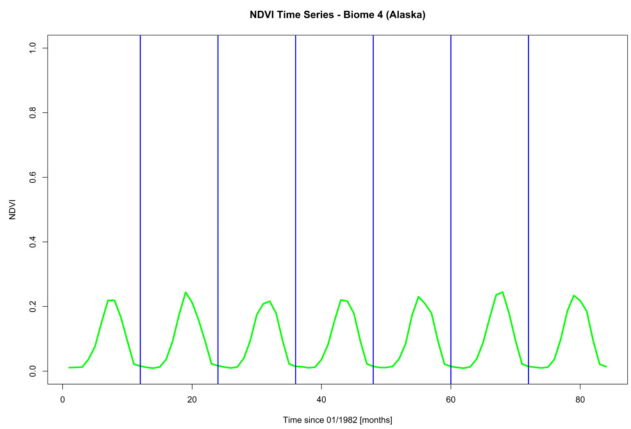
A.5.2.4 Stable State 4 - Sparsely Vegetated/Mossy/Barren



(a) ROC curve of GAM

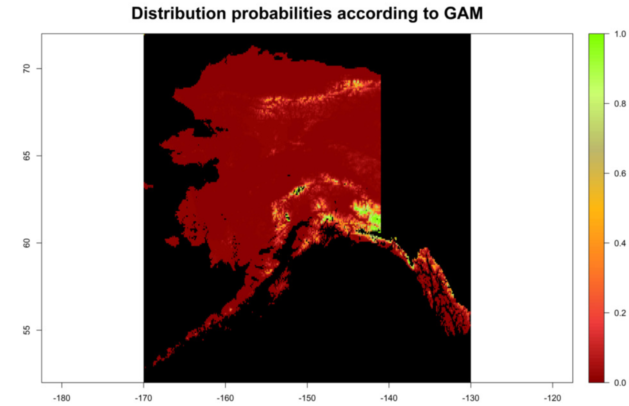


(b) Actual distribution

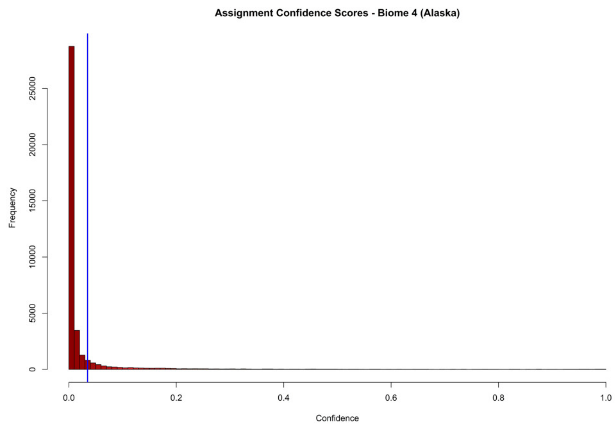


(c) NDVI time series

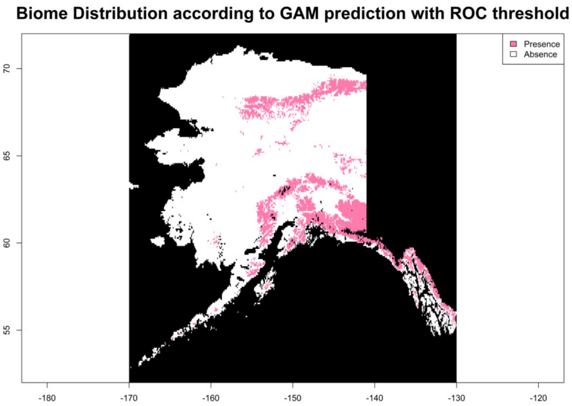
Figure A.21: GAM characteristics of stable state 4 (Alaska): The ROC curve for the GAM of stable state 4 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



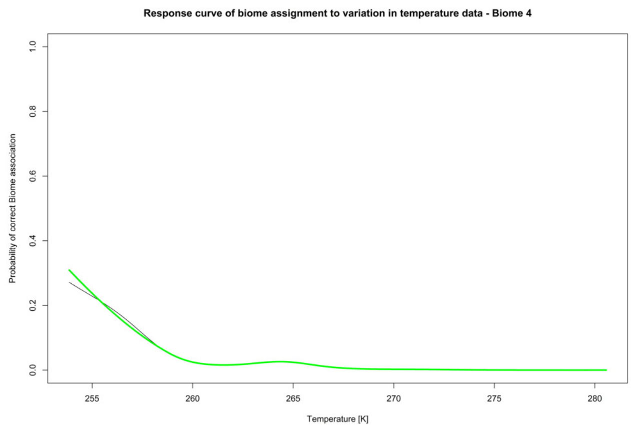
(a) Biome prediction probabilities



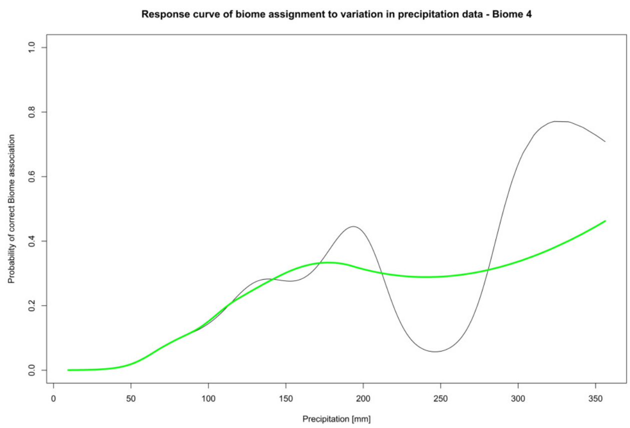
(b) Prediction probabilities with ROC threshold



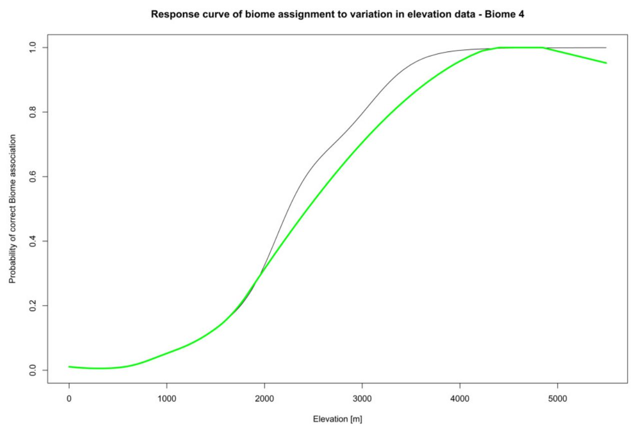
(c) Prediction of occurrence



(d) Response function (temperature)



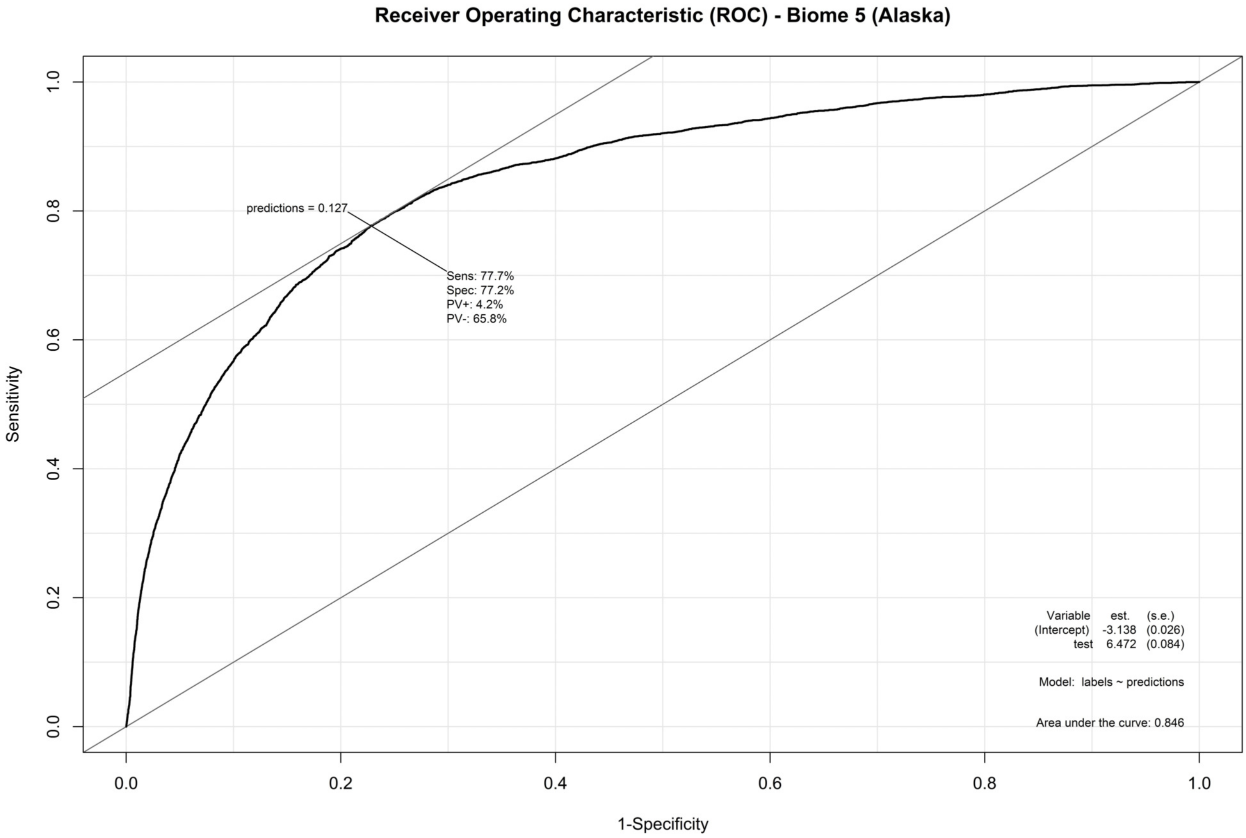
(e) Response function (precipitation)



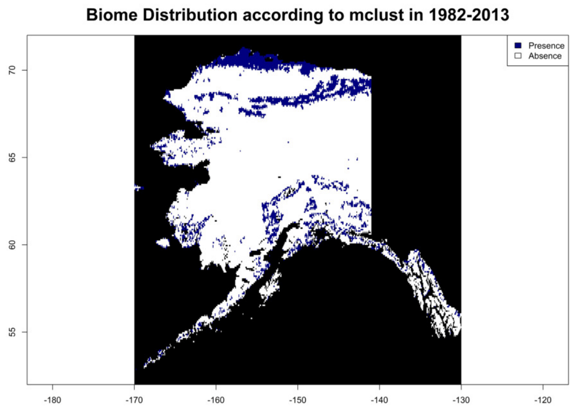
(f) Response function (DEM data)

Figure A.22: GAM characteristics of stable state 4 (Alaska): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 4 and climate-vegetation interactions.

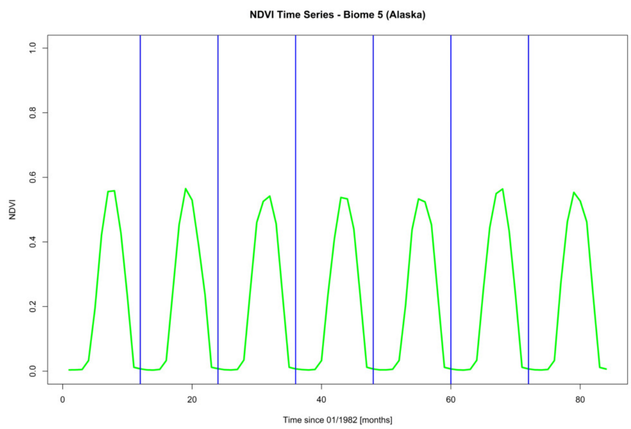
A.5.2.5 Stable State 5 - Tundra/Sedge



(a) ROC curve of GAM

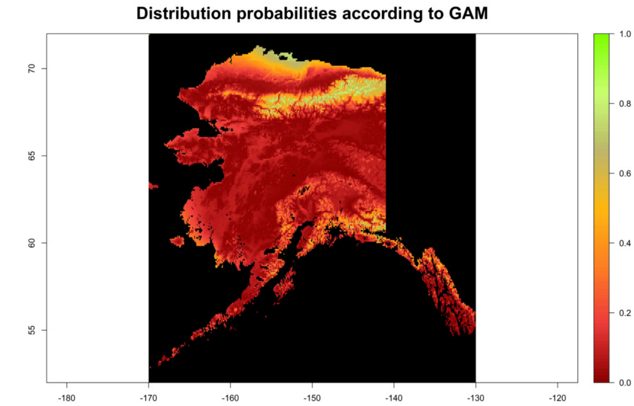


(b) Actual distribution

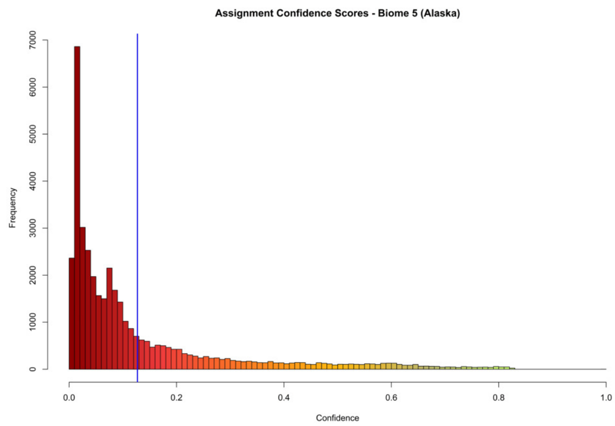


(c) NDVI time series

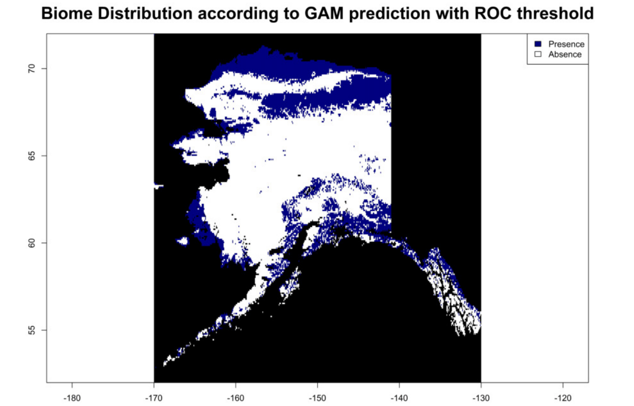
Figure A.23: GAM characteristics of stable state 5 (Alaska): The ROC curve for the GAM of stable state 5 was used to assess the GAM accuracy and produce predictions of occurrence. The NDVI time series has been plotted for 1982-1988 to identify further vegetation dynamics. Blue lines indicate yearly time frames.



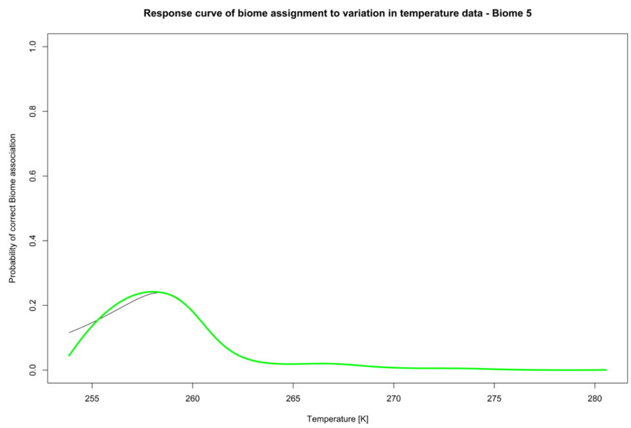
(a) Biome prediction probabilities



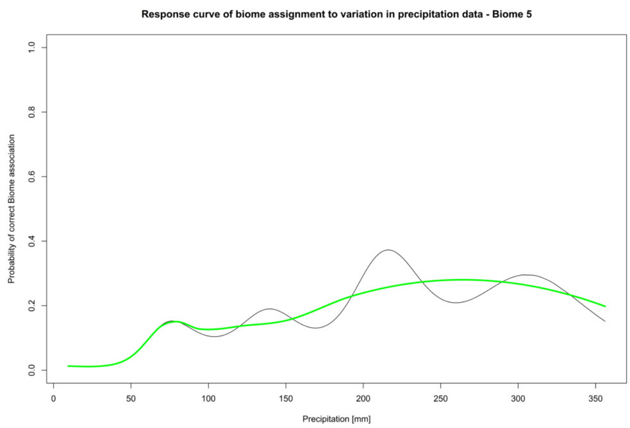
(b) Prediction probabilities with ROC threshold



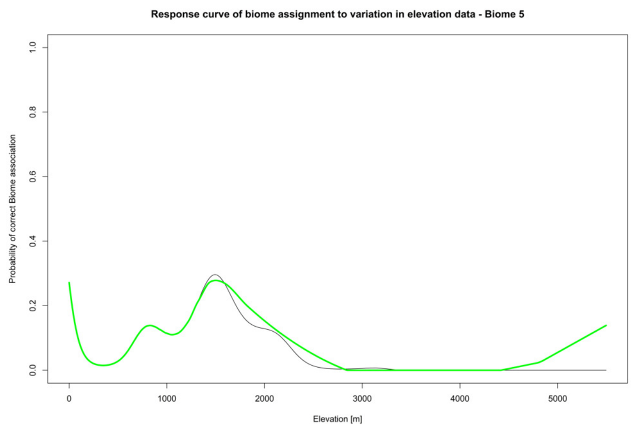
(c) Prediction of occurrence



(d) Response function (temperature)



(e) Response function (precipitation)



(f) Response function (DEM data)

Figure A.24: GAM characteristics of stable state 5 (Alaska): GAM predictions and response functions of explanatory variables have been plotted to assess the probable distribution of stable state 5 and climate-vegetation interactions.

A.6 R Coding

A.6.1 R-Packages

Running the code below in the basic installation of R will install all necessary packages that are required to run the codes used in this study.

```
install.packages("raster")
install.packages("rgdal")
install.packages("sp")

## gimms package in version 0.5.1
install.packages("doParallel") # needed for gimms
install.packages("Kendall") # needed for gimms
install.packages("zyp") # needed for gimms
packageurl <- "https://cran.r-project.org/src/contrib/Archive/gimms/gimms_0.5.1.tar.gz"
install.packages(packageurl, repos=NULL, type="source")

install.packages("gtools")
install.packages("rgeos")
install.packages("mclust")
install.packages("fpc")
install.packages("rgl")
install.packages("rgdal")
install.packages("mgcv")
install.packages("Epi")
install.packages("xlslx")
```

A.6.2 Calculating GIMMS Composites

```r
######################################################################
### LOADING PACKAGES AND SETTING UP DIRECTORIES
######################################################################

## !!!! TO RUN THIS CODE PLEASE KEEP IN MIND THAT A CONNECTION TO THE INTERNET HAS TO BE ESTABLISHED THROUGHOUT THE ENTIRE RUNNING TIME !!!!

library(gimms)
library(raster)
library(sp)

###################################
####### MAINDIRECTORY:
###################################

# Set the main directory, SDD
mainDir <- "D:/Data"

######################################################################

# SET WORKING DIRECTORY FOR RAW DATA
DataDir.Raw <- paste(mainDir, "/X - Raw.Data", sep="")
setwd(DataDir.Raw)

# SET WORKING DIRECTORY FOR RAW GIMMS DATA
Dir.Gimms <- paste(DataDir.Raw, "/1 - Gimms", sep="")
dir.create(Dir.Gimms)
setwd(Dir.Gimms)

# UPDATE FILE LIST IN GIMMS DIRECTORY
gimms_files <- updateInventory

# SET DIRECTORY FOR RASTERS OF ANNUAL NDVI-COMPOSITES
Dir.Gimms.Annual <- paste(DataDir.Raw, "/1 - Gimms-Annual.Mean", sep="")
dir.create(Dir.Gimms.Annual)
setwd(Dir.Gimms.Annual)

# SET DIRECTORY FOR RASTERS OF NDVI SEASONALITY COMPOSITES
Dir.Gimms.Seasonality <- paste(DataDir.Raw, "/1 - Gimms-Seasonality", sep="")
dir.create(Dir.Gimms.Seasonality)
setwd(Dir.Gimms.Seasonality)

######################################################################
### SELECTING DATA, EXPORTING PLOTS AND RASTERS
######################################################################
RasterGIMMs <- function(from, to){
  for(year in from:to){
    # PREPARING DATA
    gimms_files <- downloadGimms(x = as.Date(paste(year,"-01-01",sep="")), y = as.Date(paste(year,"-12-31",sep="")), dsn = Dir.Gimms)
    # Rasterize files
    gimms_raster <- rasterizeGimms(x = gimms_files, remove_header = TRUE)

    # Create monthly maximum value composites
    indices <- monthlyIndices(gimms_files)
    gimms_raster_mvc <- monthlyComposite(gimms_raster, indices = indices)

    gimms_annual <- calc(gimms_raster_mvc, fun=mean, progress='text')
    gimms_annual <- crop(gimms_annual, extent(-180,180,-60,90))
```

```
    gimms_annual[gimms_annual<0] <- 0 # set threshold for barren land (NDVI<0)
    gimms_annual[gimms_annual>1] <- 1 # set threshold for saturated NDVI (NDVI > 1)

    # SETTING UP PLOTS
    col.evi <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000) # Setting the colour gradient for ndvi

    # SAVING DATA
    print(paste("Saving raster of annual means for year", year, sep=" "))
    setwd(Dir.Gimms.Annual)
    writeRaster(gimms_annual, paste("NDVI_",year, ".grd",sep=""), format = "raster", overwrite = TRUE)

    print(paste("Saving rasters of monthly means for year", year, sep=" "))
    dir.create(paste(Dir.Gimms.Annual, "/Monthly",sep=""))
    setwd(paste(Dir.Gimms.Annual, "/Monthly",sep=""))
    gimms_raster_mvc <- crop(gimms_raster_mvc, extent(-180,180,-60,90))
    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)
    for(k in 1:12){
      month <- gimms_raster_mvc[[k]]
      writeRaster(month, paste("NDVI_",year,"_",k, ".grd",sep=""), format = "raster", overwrite = TRUE)
      plot(month, main = paste("Monthly data ", year, " month ", k,sep=""))
    }

    plot(gimms_raster_mvc[[1]])

    # PLOTTING AND SAVING PLOTS
    # MONTHLY PLOTS
    gimms_raster_mvc[gimms_raster_mvc<0] <-0
    gimms_raster_mvc[gimms_raster_mvc>1] <-1

    gimms_raster_mvc[4331] <- 1
    gimms_raster_mvc[4330] <- 0

    names(gimms_raster_mvc) <- paste(month.abb, year)

    print(paste("Saving plots of annual means for year", year, sep=" "))
    jpeg(file=paste(Dir.Gimms.Annual, "/", "Months_NDVI_",year,".jpg", sep = ""), width = 30, height = 30, units = "cm", quality = 100, res = 1000)
    print(spplot(gimms_raster_mvc, main = paste("Normalized Difference Vegetation Index (NDVI)", year, sep=" "),col.regions=col.evi))
    dev.off()

    # ANNUAL PLOTS
    gimms_annual[4330] <- 1
    gimms_annual[4331] <- 0

    jpeg(file=paste(Dir.Gimms.Annual, "/", "NDVI_",year, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
    plot(gimms_annual, main = paste ("Normalized Difference Vegetation Index (NDVI)", year, sep=" "), col=col.evi, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, axis.args=list(at=seq(0, 1, 0.1),labels=seq(0, 1, 0.1),cex.axis=0.9))
    dev.off()
  } # end of year-loop
```

```
}# end of RasterGIMMS-function


################################################################################
### SELECTING DATA, EXPORTING PLOTS AND RASTERS
################################################################################
GIMMsSeasons <- function(from, to){
  for(year in from:to){
    # PREPARING DATA
    gimms_files <- downloadGimms(x = as.Date(paste(year,"-01-01",sep="")), y = as.Date(paste(year,"-12-31",sep="")), dsn = Dir.Gimms)
    # Rasterize files
    gimms_raster <- rasterizeGimms(x = gimms_files, remove_header = TRUE)

    # Create monthly maximum value composites
    indices <- monthlyIndices(gimms_files)
    gimms_raster_mvc <- monthlyComposite(gimms_raster, indices = indices)

    maxi <- calc(gimms_raster_mvc, fun=max, progress = 'text')
    mini <- calc(gimms_raster_mvc, fun=min, progress = 'text')

    gimms_seasonality <- maxi-mini
    gimms_seasonality[gimms_seasonality >1]<-1
    gimms_seasonality[gimms_seasonality <0]<-0
    gimms_seasonality <- crop(gimms_seasonality, extent(-180,180,-60,90))

    # SETTING UP PLOTS
    col.evi <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000) # Setting the colour gradient for ndvi

    # SAVING DATA
    print(paste("Saving raster of ndvi seasonality for year", year, sep=" "))
    setwd(Dir.Gimms.Seasonality)
    writeRaster(gimms_seasonality, paste("NDVI_",year, "_Seasonality.grd",sep=""), format = "raster", overwrite = TRUE)

    # PLOTTING AND SAVING PLOTS
    # ANNUAL PLOTS
    gimms_seasonality[4330] <- 1
    gimms_seasonality[4331] <- 0
    # gimms_annual[gimms_annual<0] <- NA # set threshold for barren land (NDVI <0)

    print(paste("Saving plots of ndvi seasonality for year", year, sep=" "))
    jpeg(file=paste(Dir.Gimms.Seasonality, "/", "NDVI_",year, "_Seasonality.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
    plot(gimms_seasonality, main = paste ("Normalized Difference Vegetation Index (NDVI) Seasonality", year, sep=" "), col=col.evi, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, axis.args=list(at=seq(0, 1, 0.1),labels=seq(0, 1, 0.1),cex.axis=0.9))
    dev.off()
  } # end of year-loop
}# end of GIMMsSeasons-function


################################################################################
### SAVING NDVI BASED CLIMATOLOGIES
################################################################################
GIMMS.Climat <- function(parameter, timespan){
```

```
  year.range <- 1982:2013

  title <- paste(parameter,".Climatology", min(timespan),"-", max(timespan), sep="")

  col.plot <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000) # Setting the colour gradient for evi

  if(parameter == "NDVI"){
    parameter.long <- "Normalized Difference Vegetation Index (NDVI)"
    Dir.GIMMS.NEW <- paste(DataDir.Raw, "/1 - Gimms-Annual.Mean", sep="")
  }else{
    Dir.GIMMS.NEW <- paste(DataDir.Raw, "/1 - Gimms-Seasonality", sep="")
    parameter.long <- "NDVI Seasonality"
  }

  setwd(Dir.GIMMS.NEW)
  # select the actual data rasters needed
  fileNames <- list.files(path = Dir.GIMMS.NEW, pattern = ".grd")[match(timespan, year.range)]

  # CALCULATING THE CLIMATOLOGY
  Stack <- stack(fileNames)
  Climatology <- calc(Stack, mean, na.rm=T)

  Dir.Change <- paste(mainDir,"/1 - Climatology",sep="")
  dir.create(Dir.Change)
  Dir.Change.Climat <- paste(Dir.Change, "/1 - NDVI",sep="")
  dir.create(Dir.Change.Climat)

  # SAVING THE RASTERS
  setwd(Dir.Change.Climat)
  writeRaster(Climatology, paste(title,".grd",sep=""), format = "raster", overwrite = TRUE)

  Climatology[4330] <- 1
  jpeg(file=paste(Dir.Change.Climat, "/", title,".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(Climatology, col = col.plot, main= paste(parameter.long," Climatology ", min(timespan)," - ", max(timespan), sep=""), cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
  dev.off()
} # end of GIMMS.climat-function


##################################
####### RUN THE FUNCTIONS:
##################################
RasterGIMMs(from=1982, to=2013)
GIMMsSeasons(from=1982, to=2013)

GIMMS.Climat(parameter = "NDVI", timespan = 1982:1986)
GIMMS.Climat(parameter = "NDVI.Seasonality", timespan = 1982:1986)
GIMMS.Climat(parameter = "NDVI", timespan = 2009:2013)
GIMMS.Climat(parameter = "NDVI.Seasonality", timespan = 2009:2013)
GIMMS.Climat(parameter = "NDVI", timespan = 1982:2013)
GIMMS.Climat(parameter = "NDVI.Seasonality", timespan = 1982:2013)

print("done")
```

A.6.3 Preparing WorldClim Data

```r
######################################################################
### LOADING PACKAGES AND SETTING UP DIRECTORIES
######################################################################
library("raster")

###################################
####### MAINDIRECTORY:
###################################

# Set the main directory, SDD
mainDir <- "D:/Data"

######################################################################

# SET WORKING DIRECTORY FOR RAW DATA
DataDir.Raw <- paste(mainDir, "/X - Raw.Data", sep="")
setwd(DataDir.Raw)

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES
Dir.Clima <- paste(mainDir, "/1 - Climatology", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF NDVI
Dir.Clima.NDVI <- paste(Dir.Clima, "/1 - NDVI", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF WORLDLCIM
Dir.Clima.WC <- paste(Dir.Clima, "/2 - WorldClim", sep="")
dir.create(Dir.Clima.WC)

# SETTING UP A TEMPORARY DIRECTORY
dirtemporary <- paste(mainDir, "/ZZ - Temporary_Storage", sep="")
dir.create(dirtemporary)

######################################################################


# MOVING NDVI RASTER TO TEMPORARY DIRECTORY FOR RESAMPLING
setwd(Dir.Clima.NDVI)
# moving .gri file
fileNames.ndvi <- list.files(path = Dir.Clima.NDVI, pattern = paste("NDVI.Climatology1982-2013",".gri", sep=""))
print(paste("copying", fileNames.ndvi, sep=" ")) # move data of year in question to temporary directory
file.copy(fileNames.ndvi, dirtemporary)
# moving .grd file
fileNames.ndvi <- list.files(path = Dir.Clima.NDVI, pattern = paste("NDVI.Climatology1982-2013",".grd", sep=""))
print(paste("copying", fileNames.ndvi, sep=" ")) # move data of year in question to temporary directory
file.copy(fileNames.ndvi, dirtemporary)

setwd(dirtemporary)
ndvi <- list.files(path= dirtemporary, pattern = "NDVI.Climatology1982-2013.grd")
ndvi <- raster(ndvi)

######################################################################
### DOWNLOADING, CALCULATING AND SAVING WORLDCLIM-CLIMATOLOGIES
######################################################################
WorldClim <- function(){
```

```
# SET WORKING DIRECTORY FOR RAW WORLDCLIM (WC) DATA
Raw.WC <- paste(DataDir.Raw, "/2 - WorldClim", sep="")
dir.create(Raw.WC)

##############################
###### PRECIPITATION DATA
# DOWNLOAD PRECIPITATION CLIMATOLOGY DATA AT 2.5 MINUTE-RESOLUTION
setwd(Raw.WC)
prec <- getData('worldclim', var='prec', res=2.5)
prec.mean <- calc(prec, mean, na.rm=T) # calculating the annual climatology

# RESAMPLING STEP
setwd(dirtemporary)
writeRaster(prec.mean, "prec.Climatology.grd", format = "raster", overwrite =
    TRUE)
prec.mean.1 <- list.files(path = dirtemporary , pattern = "prec.Climatology.
    grd")
prec.mean.1 <- raster(prec.mean.1)
prec.mean.2 <- resample(prec.mean.1, ndvi, method="ngb")

# PLOT PRECIPITATION CLIMATOLOGIES
col.prec <- colorRampPalette(c("yellow","blue", "darkblue", "royalblue"))
plot(prec.mean.2, col= (col.prec(15000)), main = "Mean annual precipitation [
    mm]", xlab = "Longitude [°]", ylab = "Latitude[°]")

# SAVE THE PLOTS AND RASTER OF CLIMATOLOGIES TO CLIMATOLOGY DIRECTORY
setwd(Dir.Clima.WC)
writeRaster(prec.mean.2, "prec.Climatology.grd", format = "raster", overwrite
    = TRUE)

jpeg(file=paste(Dir.Clima.WC, "/", "prec.Climatology", ".jpg", sep = ""),
    width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(prec.mean.2, main="Mean Annual Precipitation [mm]", col=(col.prec(10000))
    , xlab = "Longitude [°]", ylab = "Latitude[°]",cex.lab=1, cex.axis=1, cex.
    main=2, cex.sub=1, legend.shrink=1, colNA="black",legend.width=2, axis.args
    =list(at=seq(0, 900, 100),labels=seq(0, 900, 100),cex.axis=0.9))
dev.off()

##############################
###### TEMPERATURE DATA
# DOWNLOAD TEMPERATURE CLIMATOLOGY DATA AT 2.5 MINUTE-RESOLUTION
setwd(Raw.WC)
temp <- getData('worldclim', var='tmean', res=2.5)
temp.kelvin <- temp/10 + 273 # worldlcim temperature data is saved as 10-times
    °C, this step converts into kelvin
temp.mean <- calc(temp.kelvin, mean, na.rm=T)

# RESAMPLING STEP
setwd(dirtemporary)
writeRaster(temp.mean, "tmean.Climatology.grd", format = "raster", overwrite =
    TRUE)
temp.mean.1 <- list.files(path = dirtemporary , pattern = "tmean.Climatology.
    grd")
temp.mean.1 <- raster(temp.mean.1)
temp.mean.2 <- resample(temp.mean.1, ndvi, method="ngb")

# PLOT TEMPERATURE CLIMATOLOGIES
col.temp <- colorRampPalette(c("darkblue","royalblue","yellow", "red")) #
    Setting the colour gradient for temp
plot(temp.mean.1, col= (col.temp(15000)), main = "Mean annual temperature [K]"
    , xlab = "Longitude [°]", ylab = "Latitude[°]")
```

```r

  # SAVE THE PLOTS AND RASTER OF CLIMATOLOGIES TO CLIMATOLOGY DIRECTORY
  setwd(Dir.Clima.WC)
  writeRaster(temp.mean.2, "tmean.Climatology.grd", format = "raster", overwrite = TRUE)

  temp.mean.2[4444:4445] <- 305
  temp.mean.2[4432:4433] <- 245
  jpeg(file=paste(Dir.Clima.WC, "/", "tmean.Climatology", ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(temp.mean.2, main="Mean Annual Temperature [K]", col=(col.temp(10000)), xlab = "Longitude [°]", ylab = "Latitude[°]",cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black",legend.width=2, axis.args=list(at=seq(245, 305, 10),labels=seq(245, 305, 10),cex.axis=0.9))
  dev.off()

  # CLEANING THE TEMPORARY DIRECTORY
  setwd(dirtemporary)
  print("Clearing Temporary Directory")
  fileNames.delete <- list.files(path=dirtemporary)
  do.call(file.remove, list(fileNames.delete))
} # end of WorldClim-function


##################################
####### RUN THE FUNCTIONS:
##################################
WorldClim()

print("done")
```

A.6.4 Preparing Digital Elevation Model Data

```r
######################################################################
### LOADING PACKAGES AND SETTING UP DIRECTORIES
######################################################################
library(raster)

###################################
####### MAINDIRECTORY:
###################################

# Set the main directory, SDD
mainDir <- "D:/Data"

######################################################################

# SET WORKING DIRECTORY FOR RAW DATA
DataDir.Raw <- paste(mainDir, "/X - Raw.Data", sep="")
dir.create(DataDir.Raw)
setwd(DataDir.Raw)

# SET DIRECTORY FOR CLIMATOLOGIES
Dir.Clima <- paste(mainDir, "/1 - Climatology", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF DEM DATA
Dir.Clima.DEM <- paste(Dir.Clima, "/3 - DEM", sep="")
dir.create(Dir.Clima.DEM)

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF NDVI
Dir.Clima.NDVI <- paste(Dir.Clima, "/1 - NDVI", sep="")

# SETTING UP A TEMPORARY DIRECTORY
dirtemporary <- paste(mainDir, "/ZZ - Temporary_Storage", sep="")

######################################################################

# MOVING NDVI RASTER TO TEMPORARY DIRECTORY FOR RESAMPLING
setwd(Dir.Clima.NDVI)
# moving .gri file
fileNames.ndvi <- list.files(path = Dir.Clima.NDVI, pattern = paste("NDVI.Climatology1982-2013",".gri", sep=""))
print(paste("copying", fileNames.ndvi, sep=" ")) # move data of year in question to temporary directory
file.copy(fileNames.ndvi, dirtemporary)
# moving .grd file
fileNames.ndvi <- list.files(path = Dir.Clima.NDVI, pattern = paste("NDVI.Climatology1982-2013",".grd", sep=""))
print(paste("copying", fileNames.ndvi, sep=" ")) # move data of year in question to temporary directory
file.copy(fileNames.ndvi, dirtemporary)

setwd(dirtemporary)
ndvi <- list.files(path= dirtemporary, pattern = "NDVI.Climatology1982-2013.grd")
ndvi <- raster(ndvi)

######################################################################
### HANDLING OF DEM DATA
######################################################################
DEM <- function(what){
  print(paste("Caclualtion for DEM product: ", what, sep=""))
```

```r

  # SETTING UP COLOUR FOR PLOT
  col_elevation <- c("black",colorRampPalette(c("darkgreen", "yellow", "gold3",
      "darkgoldenrod3", "peru", "chocolate4"))(10000))

  # GRABBING THE RAW DATA
  Raw <- paste(mainDir,"/X - Raw.Data/3 - GMTED2010/30arc/",what,"30_grd/",what,
      "30_grd",sep="")
  setwd(Raw)
  files <- list.files(path = Raw, pattern = "w001001.adf")
  raster <- raster(files)
  area <- extent(-180, 180, -60, 90)
  raster <- crop(raster, area)
  plot(raster, main ="Digital Elevation Model (GMTED2010) [m]", col= col_
      elevation, breaks=c(minValue(raster),seq(0,maxValue(raster),1)), cex.lab=1,
       cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.
      width=2, axis.args=list(at=c(minValue(raster),seq(0,maxValue(raster),500),
      maxValue(raster)), labels=c(minValue(raster),seq(0,maxValue(raster),500),
      maxValue(raster)), cex.axis=0.9))

  # RESAMPLING AND MASKING ACCORDING TO LANDSEA
  raster.res.bilin <- resample(raster, ndvi, method="ngb") # resample to ndvi
      resolution
  raster.res <- raster.res.bilin
  raster.res[raster.res < 0] <- NA

  # SAVING THE RASTER AND PLOT
  setwd(Dir.Clima.DEM)
  writeRaster(raster.res, paste("DEM_",what,".grd",sep=""), format = "raster",
      overwrite = TRUE)

  jpeg(file=paste(Dir.Clima.DEM, "/DEM_",what,".jpg", sep = ""), width = 32,
      height = 22, units = "cm", quality = 100, res = 1000)
  plot(raster.res, main ="Digital Elevation Model (GMTED2010) [m]", col= col_
      elevation, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1,
      colNA="black", legend.width=2, axis.args=list(at=seq(0,maxValue(raster.res)
      ,500), labels=seq(0,maxValue(raster.res),500), cex.axis=0.9))
  dev.off()
} # end of DEM-function


###################################
####### RUN THE FUNCTIONS:
###################################
DEM(what <- "mn")

print("done")
```

A.6.5 Assessing Collinearity Amongst Variables

```
#################################################################
### LOADING PACKAGES AND SETTING UP DIRECTORIES
#################################################################
library("raster")
library("rgdal")

################################
####### MAINDIRECTORY:
################################

# Set the main directory, SDD
mainDir <- "D:/Data"

#################################################################

# SET WORKING DIRECTORY FOR RAW DATA
DataDir.Raw <- paste(mainDir, "/X - Raw.Data", sep="")
setwd(DataDir.Raw)

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES
Dir.Clima <- paste(mainDir, "/1 - Climatology", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF WORLDLCIM
Dir.Clima.WC <- paste(Dir.Clima, "/2 - WorldClim", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF NDVI
Dir.Clima.NDVI <- paste(Dir.Clima, "/1 - NDVI", sep="")

# SET DIRECTORY FOR DEM DATA
Dir.DEM <- paste(Dir.Clima, "/3 - DEM", sep="")

# SET UP DIRECTORY FOR SAVING COLINEARITY PLOTS
Dir.PLOTS.COLIN <- paste(Dir.Clima, "/X - Collinearity",sep="")
dir.create(Dir.PLOTS.COLIN)

Dir.Shapes <- paste(mainDir, "/X - Raw.Data/X - ShapeFiles", sep="")

# set up a temporary directory to which the files are moved during the process
    of the calculations
dirtemporary <- paste(mainDir, "/ZZ - Temporary_Storage", sep="")

################################
####### PREPARING THE PROGRAM
################################
setwd(dirtemporary)
print("Clearing temporary directory to avoid remnants in said directory")
fileNames.delete <- list.files(path= dirtemporary)
do.call(file.remove, list(fileNames.delete))

#################################################################
### LOAD DATA, CROP AND ANALYSE FOR COLINEARITY
#################################################################
Colinearity <- function(region, Where, variables, variables.long){
  print(paste("Colinearity for ", paste(variables, collapse = ', '), " in ",
      Where, sep=""))

  ###############################################
  # MOVE CLIMATOOGIES INTO PREDEFINED AND SHARED DIRECTORY
  directories <- c(Dir.Clima.NDVI, Dir.Clima.WC, Dir.DEM)
```

```
file.ending <- c(".grd", ".gri")

# MOVE ALL CLIMATOLOGIES TO TEMPOIRARY STORAGE/DIRECTORY
for(i in 1:length(directories)){
  for(g in 1:2){
    setwd(directories[i])
    fileNames <- list.files(path = directories[i], pattern = file.ending[g])
    print(paste("copying", fileNames, sep=" ")) # move data of year in question to temporary directory
    file.copy(fileNames, dirtemporary)
  } # end of g-loop (file selection)
} # end of i-loop (direcotry selection)

##########################################
# REGION SELECTION WITH SHAPEFILES
print("Reading ShapeFiles:")
setwd(Dir.Shapes)
Urban <- readOGR('.','ne_10m_urban_areas') # reading the shapefiles for urban areas
Lakes <- readOGR('.','ne_10m_lakes') # reading the shapefiles for lakes and streams

if(region == "Country"){
  Shapes=readOGR('.','ne_50m_admin_0_countries') # reading the shapefiles
}else{ if(region == "State"){
  Shapes=readOGR('.','ne_10m_admin_1_states_provinces')
}else{
  print("You have specified no area which to select. No cropping will be done. If you want global data sets ignore this message.")
}
}

##########################################
# IMPORT FURTHER IMPORTANT RASTERS
setwd(dirtemporary)
RasterX <- list.files(path = dirtemporary, pattern = "NDVI.Climatology1982-2013.grd")
RasterX <- raster(RasterX)

##########################################
# DATA PREPARATION
# ESTABLISHING AN EMPTY VECTOR TO BE FILLED WITH INDICES OF SHAPEFILES IN POLYGOMFRAME (SHAPES)
location <- rep(NA, length(Where))

# FILLING THE LOCATIONS VECTOR WITH INDICES
if(region != "Global"){
  for(i in 1:length(Where)){
    location[i] <- which(as.vector(Shapes$name) == Where[i])
  }
}

# EXCEPTIONS FOR RECTANGULAR CROPPING AROUND SHAPES WHERE THE SHAPEFILE ALONE DOESN#T CROP TIGHTLY
if(Where == "Alaska"){
  area <- extent(-170,-130,52,72)
}else{ if(Where == c("United States", "Canada", "Mexico")){
  area <- extent(-170,-50,10,90)
  Where <- "NoAm" # set this here for title management later
}else{ if(Where == "Global"){
  area <- extent(-180,180,-60,90) # this is global data
```

```
    Where <- "Global" # set this here for title management later
  }else{
    area <- extent(Shapes[location,]) # this is for simple state selection
  }
  }
  }

  # CROPPING IMPORTANT RASTERS
  RasterX <- crop(RasterX,area)
  if(region != "Global"){
    RasterX <- mask(RasterX, Shapes[location,])
  }

  Urban <- crop(Urban,area)
  Lakes <- crop(Lakes, area)

  # CREATE RASTERS TO MASK URBAN AND LAKE AREAS
  if(is.null(Urban)){
    Antro <- raster(matrix(rep(NA, length(RasterX)), ncol=RasterX@ncols),
                    xmn=area@xmin, xmx=area@xmax, ymn=area@ymin, ymx=area@ymax)
  }else{
    Antro <- crop(RasterX, area)
    Antro <- mask(Antro, Urban)
    Antro[!is.na(Antro)] <- -8888
  }

  if(is.null(Lakes)){
    Water <- raster(matrix(rep(NA, length(RasterX)), ncol=RasterX@ncols),
                    xmn=area@xmin, xmx=area@xmax, ymn=area@ymin, ymx=area@ymax)
  }else{
    Water <- crop(RasterX, area)
    Water <- mask(Water, Lakes)
    Water[!is.na(Water)] <- -8888
  }

  # make the first parameter column, to determine length which is needed to
      create empty matrix
  setwd(dirtemporary)
  if(variables[1] == "DEM_mn" | variables[1] == "NDVI.Seasonality.
      Climatology1982-1986" | variables[1] == "NDVI.Climatology1982-1986" |
      variables[1] == "NDVI.Seasonality.Climatology1982-2013" | variables[1] == "
      NDVI.Climatology1982-2013"){
    ras <- list.files(path = dirtemporary, pattern = paste(variables[1], sep="")
        )[1]
  }else{
    ras <- list.files(path = dirtemporary, pattern = paste(variables[1],".
        Climatology.grd", sep=""))
  }
  ras <- raster(ras)
  ras <- crop(ras,area)
  if(region != "Global"){
    ras <- mask(ras, Shapes[location,])
  }

  # create empty matrix which is to be filled and make first column into row
      numbers
  Length <- length(as.vector(ras))
  Matrix <- matrix(-8888, nrow = Length, ncol = length(variables)+1)
  Matrix[,1] <- seq(from=1, to=Length, by=1)
  print(paste("Matrix of column length", Length,"created", sep=" "))
```

```
# fill the matrix
for(i in 1:length(variables)){
  if(variables[i] == "DEM_mn" | variables[i] == "NDVI.Seasonality.Climatology1982-1986" | variables[i] == "NDVI.Climatology1982-1986" | variables[i] == "NDVI.Seasonality.Climatology1982-2013" | variables[i] == "NDVI.Climatology1982-2013"){
    ras <- list.files(path = dirtemporary, pattern = paste(variables[i], sep=""))[1]
  }else{
    ras <- list.files(path = dirtemporary, pattern = paste(variables[i],".Climatology.grd", sep=""))
  }
  ras <- raster(ras)
  ras <- crop(ras,area)
  if(region != "Global"){
    ras <- mask(ras, Shapes[location,])
  }
  ras[Antro == -8888] <- NA # masking for urban areas here
  ras[Water == -8888] <- NA # masking for lake areas here
  ras[is.na(RasterX)] <- NA
  Matrix[,i+1] <- as.vector(ras)
  print(paste(variables[i], "masked and fitted into matrix", sep=" "))
  plot(ras, main = variables[i])
}

# finishing up data preparation
print("Matrix Ready!")
colnames(Matrix) <- c("X",variables.long)

# PREPARING THE DATA
data.set <- Matrix[,-1]
data.values <- data.set[!rowSums(!is.finite(data.set)),] # remove all rows with non-finite values

# DEFINING COLOUR FOR PLOTTING SYMBOLS
circles.col=rgb(0,0,0,alpha=0.008)

# DEFINING FUNCTION FOR PRINTING CORRELATION PARAMETERS
panel.cor <- function(x, y, digits=2, prefix="", cex.cor, ...){
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits=digits)[1]
  txt <- paste(prefix, txt, sep="")
  test <- cor.test(x,y)
  Signif <- ifelse(round(test$p.value,3)<0.001,"p<0.001",paste("p=",round(test$p.value,3)))
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.25, paste("r=",txt), cex = cex.cor * r/2.0)
  text(.5, .75, Signif)
}

# DEFINING COLOUR FOR CORRELATION PLOTS AND SMOOTHING LINES
panel.smooth<-function (x, y, col = circles.col, bg = NA, pch = 1,
                        cex = 0.5, col.smooth = "red", span = 2/3, iter = 3, ...) {
  points(x, y, pch = pch, col = col, bg = bg, cex = cex)
  ok <- is.finite(x) & is.finite(y)
  if (any(ok))
    lines(stats::lowess(x[ok], y[ok], f = span, iter = iter), col = col.smooth, ...)
```

```
  }

  # SETTING UP DIRECTORY FOR INDICIDUAL AREAS
  Dir.PLOTS.COLIN.Ind <- paste(Dir.PLOTS.COLIN, "/", Where, sep="")
  dir.create(Dir.PLOTS.COLIN.Ind)

  # SAVING THE PLOT
  jpeg(file=paste(Dir.PLOTS.COLIN.Ind, "/", paste(colnames(data.values[]),
      collapse = '_'), ".jpg", sep = ""), width = 16, height = 11, units = "cm",
      quality = 100, res = 1000)
  pairs(data.values, upper.panel=panel.smooth, lower.panel=panel.cor, cex.labels
      =0.6, cex.axis = 0.8)
  dev.off()

  # CLEAR TEMPORARY DIRECTORY
  setwd(dirtemporary)
  print("Clearing Temporary Directory")
  fileNames.delete <- list.files(path= dirtemporary)
  do.call(file.remove, list(fileNames.delete))
} # end of Colinearity-function


####################################
####### RUN THE FUNCTION:
####################################
# used in modelling
Colinearity(region = "State", Where = "Alaska", variables <- c("NDVI.
    Climatology1982-2013", "NDVI.Seasonality.Climatology1982-2013", "tmean", "
    prec", "DEM_mn"), variables.long <- c("NDVI 1982-2013", "NDVI Seasonality
    1982-2013", "Mean Annual Temperature", "Mean Annual Precipitation", "DEM"))
Colinearity(region = "State", Where = "Minnesota", variables <- c("NDVI.
    Climatology1982-2013", "NDVI.Seasonality.Climatology1982-2013", "tmean", "
    prec", "DEM_mn"), variables.long <- c("NDVI 1982-2013", "NDVI Seasonality
    1982-2013", "Mean Annual Temperature", "Mean Annual Precipitation", "DEM"))
# used in evaluation custering
Colinearity(region = "State", Where = "Alaska", variables <- c("NDVI.
    Climatology1982-2013", "NDVI.Seasonality.Climatology1982-2013", "tmean.JJA",
    "prec"), variables.long <- c("NDVI 1982-2013", "NDVI Seasonality 1982-2013",
    "Mean Summer Temperature", "Mean Annual Precipitation"))
Colinearity(region = "State", Where = "Minnesota", variables <- c("NDVI.
    Climatology1982-2013", "NDVI.Seasonality.Climatology1982-2013", "tmean.JJA",
    "prec"), variables.long <- c("NDVI 1982-2013", "NDVI Seasonality 1982-2013",
    "Mean Summer Temperature", "Mean Annual Precipitation"))

print("Done")
```

A.6.6 Vegetation Clustering And Vegetation-Climate Modelling

```r
########################################################################
### LOADING PACKAGES AND SETTING UP DIRECTORIES
########################################################################
library("raster")
library("mclust")
library("fpc")
library("rgl")
library("rgdal")
library("gtools")
library("mgcv")
library("Epi")
library("xlsx")

##################################
####### MAINDIRECTORY:
##################################

# Set the main directory, SDD
mainDir <- "D:/Data"

########################################################################

# SET WORKING DIRECTORY FOR RAW DATA
DataDir.Raw <- paste(mainDir, "/X - Raw.Data", sep="")

# SET DIRECTORY FOR CHANGE ANALYSIS
Dir.Change.Analysis <- paste(mainDir, "/3 - Change", sep="")
dir.create(Dir.Change.Analysis)

# SET DIRECTORY FOR SHAPE FILES
Dir.Shapes <- paste(mainDir, "/X - Raw.Data/X - ShapeFiles", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES (SET UP HERE FOR FOLLOWING
    DIRECTORY SPECIFICATION)
Dir.Clima <- paste(mainDir,"/1 - Climatology", sep="")

# SET DIRECTORY FOR DEM DATA
Dir.DEM <- paste(Dir.Clima, "/3 - DEM", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF WORLDLCIM
Dir.Clima.WC <- paste(Dir.Clima, "/2 - WorldClim", sep="")

# SET DIRECTORY FOR SAVING THE CLIMATOLOGIES OF EVI
Dir.Clima.NDVI <- paste(Dir.Clima, "/1 - NDVI", sep="")

# set up a temporary directory to which the files are moved during the process
    of the calculations
dirtemporary <- paste(mainDir, "/ZZ - Temporary_Storage", sep="")

##################################
####### PREPARING THE PROGRAM
##################################
setwd(dirtemporary)
print("Clearing temporary directory to avoid remnants in said directory")
fileNames.delete <- list.files(path= dirtemporary)
do.call(file.remove, list(fileNames.delete))

########################################################################
### SELECTING DATA, MASKING IT FOR REGION, RUN ANALYSIS
```

```r
##############################################################################
CHANGE.FUN <- function(region, Where, clusters, sample.num, DEM){
  variables <- c("NDVI", "NDVI.Seasonality", "prec", "tmean")
  variables.long <- c("NDVI", "NDVI Seasonality", "Precipitation [mm]", "Temperature [K]")

  ##############################################
  # REGION SELECTION WITH SHAPEFILES
  print("Reading ShapeFiles:")
  setwd(Dir.Shapes)
  Urban <- readOGR('.','ne_10m_urban_areas') # reading the shapefiles for urban areas
  Lakes <- readOGR('.','ne_10m_lakes') # reading the shapefiles for lakes and streams

  if(region == "Country"){
    Shapes=readOGR('.','ne_50m_admin_0_countries') # reading the shapefiles
  }else{ if(region == "State"){
    Shapes=readOGR('.','ne_10m_admin_1_states_provinces')
  }else{
    print("You have specified no area which to select. No cropping will be done. If you want global data sets ignore this message.")
  }
  }

  ##############################################
  # IMPORT DATA AND FURTHER IMPORTANT RASTERS
  # DEM DATA
  setwd(Dir.DEM)
  files <- list.files(path = Dir.DEM, pattern = paste("DEM_",DEM,".grd",sep=""))
  DEM.ras <- raster(files)

  # TEMPORAL AUTOCORRELATION (NEEDED FOR MASKING PROCESSING LATER)
  setwd(Dir.Clima.NDVI)
  RasterX <- list.files(path = Dir.Clima.NDVI, pattern = "NDVI.Climatology1982-2013.grd")
  RasterX <- raster(RasterX)

  # CLIMATOLOGIES
  file.ending <- c(".grd", ".gri")
  print("copying data")
  for(i in 1:2){
    setwd(paste(mainDir, "/1 - Climatology/2 - WorldClim",sep=""))
    prec.move <- list.files(path=paste(mainDir, "/1 - Climatology/2 - WorldClim",sep=""), pattern=paste("prec.Climatology", file.ending[i], sep=""))
    file.copy(prec.move, dirtemporary)

    tmean.move <- list.files(path=paste(mainDir, "/1 - Climatology/2 - WorldClim",sep=""), pattern=paste("tmean.Climatology", file.ending[i], sep=""))
    file.copy(tmean.move, dirtemporary)

    setwd(Dir.Clima.NDVI)
    NDVI.move <- list.files(path=Dir.Clima.NDVI, pattern=paste(file.ending[i], sep=""))
    file.copy(NDVI.move, dirtemporary)
  } # end of moving data

  files <- list.files(path = dirtemporary, pattern=".grd")[-1][-2][-2][-3] # data spanning 1982-2013
  files.past <- list.files(path = dirtemporary, pattern=".grd")[-2][-2][-3][-3] # data spanning 1982-1986
```

```
files.present <- list.files(path = dirtemporary, pattern=".grd")
    [-1][-1][-2][-2] # data spanning 2009-2013

########################################
# DATA PREPARATION
# ESTABLISHING AN EMPTY VECTOR TO BE FILLED WITH INDICES OF SHAPEFILES IN
    POLYGOMFRAME (SHAPES)
location <- rep(NA, length(Where))

# FILLING THE LOCATIONS VECTOR WITH INDICES
if(region != "Global"){
  for(i in 1:length(Where)){
    location[i] <- which(as.vector(Shapes$name) == Where[i])
  }
}

# EXCEPTIONS FOR RECTANGULAR CROPPING AROUND SHAPES WHERE THE SHAPEFILE ALONE
    DOESN#T CROP TIGHTLY
if(Where == "Alaska"){
  area <- extent(-170,-130,52,72)
}else{ if(Where == c("United States", "Canada", "Mexico")){
  area <- extent(-170,-50,10,90)
  Where <- "NoAm" # set this here for title management later
}else{ if(Where == "Global"){
  area <- extent(-180,180,-60,90) # this is global data
  Where <- "Global" # set this here for title management later
}else{
  area <- extent(Shapes[location,]) # this is for simple state selection
}
}
}

# CROPPING DEM-DATA
DEM.ras <- crop(DEM.ras, area)
if(region != "Global"){
  DEM.ras <- mask(DEM.ras, Shapes[location,])
}

# CROPPING RASTERX DATA
RasterX <- crop(RasterX,area)
if(region != "Global"){
  RasterX <- mask(RasterX, Shapes[location,])
}

# CROPPING MASKING FILES
Urban <- crop(Urban,area)
Lakes <- crop(Lakes, area)

# CREATE RASTERS TO MASK URBAN AND LAKE AREAS
if(is.null(Urban)){
  Antro <- raster(matrix(rep(NA, length(RasterX)), ncol=RasterX@ncols), xmn=
      area@xmin, xmx=area@xmax, ymn=area@ymin, ymx=area@ymax)
}else{
  Antro <- crop(RasterX, area)
  Antro <- mask(Antro, Urban)
  Antro[!is.na(Antro)] <- -8888
}

if(is.null(Lakes)){
  Water <- raster(matrix(rep(NA, length(RasterX)), ncol=RasterX@ncols), xmn=
      area@xmin, xmx=area@xmax, ymn=area@ymin, ymx=area@ymax)
```

```
} else {
  Water <- crop(RasterX, area)
  Water <- mask(Water, Lakes)
  Water[!is.na(Water)] <- -8888
}

# CREATE AN EMPTY MATRIX TO HOLD THE PAST DATA (NDVI; NDVI-SEASONALITY; PREC; TMEAN) AND MAKE FIRST COLUMN INTO ROW NUMBERS
setwd(dirtemporary)
ras.ndvi <- files[1]
ras.ndvi <- raster(ras.ndvi)
ras.ndvi <- crop(ras.ndvi,area)
if(region != "Global"){
  ras.ndvi <- mask(ras.ndvi, Shapes[location,])
}

Length <- length(as.vector(ras.ndvi))
Matrix <- matrix(-8888, nrow = Length, ncol = length(variables)+2)
Matrix[,1] <- seq(from=1, to=Length, by=1)
print(paste("Matrix of column length", Length,"created", sep=" "))

################################ ENTIRE TIME SERIES DATA
Matrix.Full <- Matrix
# FILLING THE MATRIX
for(i in 1:length(variables)){
  ras <- files[i]
  ras <- raster(ras)
  ras <- crop(ras,area)
  if(region != "Global"){
    ras <- mask(ras, Shapes[location,])
  }
  ras[Antro == -8888] <- NA # masking for urban areas here
  ras[Water == -8888] <- NA # masking for lake areas here
  ras[is.na(ras.ndvi)] <- NA # masking for further lake areas here
  Matrix.Full[,i+1] <- as.vector(ras)
  print(paste(variables[i], "masked and fitted into full matrix", sep=" "))
  plot(ras, main = variables[i])
}
Matrix.Full[,length(variables)+2] <- as.vector(DEM.ras)

################################ PAST DATA
Matrix.past <- Matrix
# FILLING THE MATRIX
for(i in 1:length(variables)){
  ras <- files.past[i]
  ras <- raster(ras)
  ras <- crop(ras,area)
  if(region != "Global"){
    ras <- mask(ras, Shapes[location,])
  }
  ras[Antro == -8888] <- NA # masking for urban areas here
  ras[Water == -8888] <- NA # masking for lake areas here
  ras[is.na(ras.ndvi)] <- NA # masking for further lake areas here
  Matrix.past[,i+1] <- as.vector(ras)
  print(paste(variables[i], "masked and fitted into past matrix", sep=" "))
  plot(ras, main = variables[i])
}
Matrix.past[,length(variables)+2] <- as.vector(DEM.ras)

################################ PRESENT DATA
Matrix.present <- Matrix
```

```
# FILLING THE MATRIX
for(i in 1:length(variables)){
  ras <- files.present[i]
  ras <- raster(ras)
  ras <- crop(ras,area)
  if(region != "Global"){
    ras <- mask(ras, Shapes[location,])
  }
  ras[Antro == -8888] <- NA # masking for urban areas here
  ras[Water == -8888] <- NA # masking for lake areas here
  ras[is.na(ras.ndvi)] <- NA # masking for further lake areas here
  Matrix.present[,i+1] <- as.vector(ras)
  print(paste(variables[i], "masked and fitted into present matrix", sep=" "))
  plot(ras, main = variables[i])
}
Matrix.present[,length(variables)+2] <- as.vector(DEM.ras)

# FINISHING UP DATA PREPARATION
colnames(Matrix.Full) <- c("X",variables.long,"DEM")
colnames(Matrix.past) <- c("X",variables.long,"DEM")
colnames(Matrix.present) <- c("X",variables.long, "DEM")
rm(Matrix)
print("Matrices Ready!")

# SAVE SOME RAM
rm(RasterX)
rm(NDVI.move)
rm(prec.move)
rm(tmean.move)

##########################################
# INCLUDING THE ACTUAL DATA
data.set.Full <- Matrix.Full[,1:3]
data.set.past <- Matrix.past[,1:3]
data.set.present <- Matrix.present[,1:3]

# remove all rows which contain NAs because Mclust can't handle these
data.values.Full <- data.set.Full[,-1][!rowSums(!is.finite(data.set.Full[,-1])
    ),]
data.values.past <- data.set.past[,-1][!rowSums(!is.finite(data.set.past[,-1])
    ),]
data.values.present <- data.set.present[,-1][!rowSums(!is.finite(data.set.
    present[,-1])),]

# needed later on in the script, already processed here to be able to empty
    data.set since it is taking up huge amounts of RAM
empty <- as.vector(rep(NA, length(data.set.Full[,1])))

### figure out which cells correspond to rows without NAs in data-matrix
fill.cells.Full <- as.vector(data.set.Full[,1][!is.na(rowSums(data.set.Full
    [,-1]))])
fill.cells.past <- as.vector(data.set.past[,1][!is.na(rowSums(data.set.past
    [,-1]))])
fill.cells.present <- as.vector(data.set.present[,1][!is.na(rowSums(data.set.
    present[,-1]))])

# save yourself some RAM
rm(data.set.Full)
rm(data.set.past)
rm(data.set.present)
```

```
##########################################
# RANDOMLY SAMPLE DATA
if(sample.num == "All"){
  sample <- data.values.Full
}else{
  set.seed(42)
  sample <- data.values.Full[sample(x = 1:nrow(data.values.Full), size= sample.num, replace=FALSE) ,]
}

##########################################
# MCLUST-ANALYSIS
G <- clusters
rm(clusters)

# LETTING MCLUST DETERMINE THE CORRECT AMOUNT OF CLUSTERS; MODEL TO BE USED AND MODEL ITSELF
print("Calculating BIC for Sample")
dataBIC <- mclustBIC(sample, G=G)
print(summary(dataBIC))
plot(dataBIC)

print("Calculating MODEL according to BIC")
mod <- mclustModel(sample, dataBIC, G=G)
mod <-  Mclust(sample, G=G)
print("Calculated MODEL according to BIC")

NClusters <- mod$G # put optimal number of clusters as defined by mclust into environment for later use

##########################################
# MCLUST-PREDICTION
# PREDICTING DATA ASSOCIATION IN LARGE DATA SET USING THE MODEL DEFINED BY SAMPLED DATA
pred.Full <- predict.Mclust(mod, data.values.Full)
pred.past <- predict.Mclust(mod, data.values.past)
pred.present <- predict.Mclust(mod, data.values.present)
print("Prediction done")

##########################################
# PREAPRING DATA FOR PLOTTING
n.col <- ras@ncols
by.row=TRUE
y1 <- ras@extent@ymin
y2 <- ras@extent@ymax
x1 <- ras@extent@xmin
x2 <- ras@extent@xmax

# DATA FOR CLASSIFICATION PLOT, turn classification as defined by mclust into raster
done <- empty
done[fill.cells.Full] <- as.vector(pred.Full$classification)
classes <- done
classes <- matrix(classes, ncol=n.col, byrow=by.row)
classes <- raster(classes, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

done <- empty
done[fill.cells.past] <- as.vector(pred.past$classification)
classes.past <- done
classes.past <- matrix(classes.past, ncol=n.col, byrow=by.row)
classes.past <- raster(classes.past, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
```

```
done <- empty
done[fill.cells.present] <- as.vector(pred.present$classification)
classes.present <- done
classes.present <- matrix(classes.present, ncol=n.col, byrow=by.row)
classes.present <- raster(classes.present, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

# DATA FOR PROBABILITY PLOT, turn porbabilities as defined by mclust into raster
done <- empty
done[fill.cells.Full] <- apply(pred.Full$z, 1, max)
probs <- done
probs <- matrix(probs, ncol=n.col, byrow=by.row)
probs <- raster(probs, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

done <- empty
done[fill.cells.past] <- apply(pred.past$z, 1, max)
probs.past <- done
probs.past <- matrix(probs.past, ncol=n.col, byrow=by.row)
probs.past <- raster(probs.past, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

done <- empty
done[fill.cells.present] <- apply(pred.present$z, 1, max)
probs.present <- done
probs.present <- matrix(probs.present, ncol=n.col, byrow=by.row)
probs.present <- raster(probs.present, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

##########################################
# SETTING UP PLOTS
# COLOURING OF CLASSIFICATION PLOT
my_palette_classes <- colorRampPalette(c("wheat1", "wheat4", "yellow", "yellow4", "lightgoldenrod", "gold", "darkgoldenrod", "darkolivegreen", "darkolivegreen2", "yellowgreen", "mediumspringgreen", "green", "forestgreen", "darkgreen", "indianred1", "indianred4", "red", "red3", "darkred", "orangered3", "chocolate1", "tan1", "lightpink1", "palevioletred1", "hotpink", "violet", "violetred1", "violetred", "magenta", "darkorchid4", "blueviolet", "navyblue"))(n = NClusters)
col_breaks = c(seq(from=0, to=NClusters, by=1)) # breaks for colouring

# COLOURING OF PROBABILITY PLOT
my_palette_probs <- colorRampPalette(c("darkred","brown2","chocolate1","darkgoldenrod1","darkkhaki", "darkolivegreen1", "chartreuse"))(n=1000)

# COLOURING FOR ELEVATION PLOT
col_elevation <- c("grey",colorRampPalette(c("darkgreen", "yellow", "gold3", "darkgoldenrod3", "peru", "chocolate4"))(10000))

# SET WORKING DIRECTORY FOR AREA
dirNew1 <- paste(Dir.Change.Analysis, "/",Where, sep="")
dir.create(dirNew1)

# SAVING THE MEANS OF THE CLUSTERS ACCORDING TO THE MCLUST-MODEL
setwd(dirNew1)
capture.output(mod$parameters, file = paste("MClustModel_Parameters.R", sep=""), append = FALSE, type="output")

##########################################
# CLASSIFICATION PLOT
# SET THE POSITION AND STYLE OF LEGEND CORRESPONDING TO CONTENT OF PLOT
if(NClusters < 30){
  legendpos <- -0.1
```

```r
    legendcol <- 1
  }else{
    legendpos <- 0
    legendcol <- 2
  }

  if(region != "Global"){
    legendposition <- "topright"
  }else{
    legendposition <- "top"
    legendcol <- NClusters/5
  }

  jpeg(file=paste(dirNew1, "/", "1 - Observation_MClust_Full.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(classes, col = my_palette_classes, main="Classifications 1982-2013", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, legend=FALSE)
  par(xpd = TRUE)
  legend(legendposition, inset=c(legendpos,0), legend = c(seq(1,NClusters,1), "Urban"), fill = c(my_palette_classes,"grey"), ncol=legendcol)
  par(new=TRUE)
  plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
  dev.off()

  jpeg(file=paste(dirNew1, "/", "3 - Observation_MClust_Past.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(classes.past, col = my_palette_classes, main="Classifications 1982-1986", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, legend=FALSE)
  par(xpd = TRUE)
  legend(legendposition, inset=c(legendpos,0), legend = c(seq(1,NClusters,1), "Urban"), fill = c(my_palette_classes,"grey"), ncol=legendcol)
  par(new=TRUE)
  plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
  dev.off()

  jpeg(file=paste(dirNew1, "/", "5 - Observation_MClust_Present.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(classes.present, col = my_palette_classes, main="Classifications 2009-2013", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, legend=FALSE)
  par(xpd = TRUE)
  legend(legendposition, inset=c(legendpos,0), legend = c(seq(1,NClusters,1), "Urban"), fill = c(my_palette_classes,"grey"), ncol=legendcol)
  par(new=TRUE)
  plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
  dev.off()

  # SAVING BIOME PROPORTION CHANGES AS EXCEL SHEET
  past <- as.vector(classes.past)
  present <- as.vector(classes.present)

  # this matrix will hold the data, rows will show past state, columns will show present state
  changematrix <- matrix(rep(NA, NClusters^2), nrow=NClusters, ncol=NClusters)
  changevec <- rep(NA, NClusters)

```

```
for(k in 1:NClusters){
  changerun <- changevec
  changeperc <- changevec

  for(m in 1:NClusters){ # fill rates into matrix
    presentcells <- which(present==m) # figure out which cells hold value m
    pastcells <- which(past==k) # figure out which cells hold value k

    rate <- length(Reduce(intersect, list(pastcells,presentcells))) # figure
        out how many of the cell denominators are shared by the two vectors
    changerun[m] <- rate
  }
  changematrix[k,] <- changerun

  for(n in 1:NClusters){ # turn rates into percentages
    changeperc[n] <- changematrix[k,n] / sum(changematrix[k,])
  }
  changematrix[k,] <- changeperc
}
changematrix <- changematrix*100
rownames(changematrix) <- seq(1,NClusters, 1)
colnames(changematrix) <- seq(1,NClusters, 1)

# SETTING UP PERCENTAGE MATRIX FOR STACKED BAR PLOTS OF BIOMES
percentages <- matrix(rep(NA, NClusters*3), ncol=3, byrow=T)
percentages[,1] <- prop.table(table(as.vector(classes)))*100
percentages[,2] <- prop.table(table(as.vector(classes.past)))*100
percentages[,3] <- prop.table(table(as.vector(classes.present)))*100
colnames(percentages) <- c("1982-2013", "1982-1986", "2009-2013")

export <- round(cbind(changematrix, rep(NA, NClusters), percentages[,2],
    percentages[,3]), digits=2)
colnames(export) <- c(seq(1,NClusters,1), "Empty", "Past Proportions", "
    Present Porportions")

setwd(dirNew1)
write.csv(export, "7b - Proportions.csv")

# define plot elements
legendtext <- c(paste(round(as.vector(percentages[,1]), digits=2), rep("%",
    NClusters), sep=""), paste(round(as.vector(percentages[,2]), digits=2), rep
    ("%", NClusters), sep= ""), paste(round(as.vector(percentages[,3]), digits
    =2), rep("%", NClusters), sep= ""))

ypos1 <- rep(NA, NClusters*1)
perc1 <- as.vector(percentages[,1])
ypos2 <- rep(NA, NClusters*1)
perc2 <- as.vector(percentages[,2])
ypos3 <- rep(NA, NClusters*1)
perc3 <- as.vector(percentages[,3])
for(i in 1:NClusters){
  ypos1[i] <- sum(perc1[1:i])-perc1[i]/2
  ypos2[i] <- sum(perc2[1:i])-perc2[i]/2
  ypos3[i] <- sum(perc3[1:i])-perc3[i]/2
}

jpeg(file=paste(dirNew1, "/", "7 - BiomeProportions.jpg", sep = ""), width =
    32, height = 22, units = "cm", quality = 100, res = 1000)
barplot <- barplot(percentages, width = c(10, 10, 10), space=0.4, col=my_
    palette_classes, yaxt='n', main=paste("Biome proportions according to
    mclust - ", Where, sep=""))
```

```
text(x = c(rep(barplot[1], NClusters), rep(barplot[2], NClusters), rep(barplot[3], NClusters)), y = c(ypos1, ypos2, ypos3),labels = legendtext, cex=0.7)
par(xpd=NA,mar=par()$mar+c(0,0,0,6))
legend("topleft", inset=c(0.3,0), legend = rev(seq(1,NClusters,1)), fill = rev(my_palette_classes), ncol=legendcol)
dev.off()

##########################################
# CONFIDENCE PLOT
probs[1] <- 0
probs[2] <- 1

jpeg(file=paste(dirNew1, "/", "2 - Observation_MClust_Probabilities_Full.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(probs, col=my_palette_probs, main = "Confidence of assignment 1982-2013", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
par(new=TRUE)
plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
dev.off()

probs.past[1] <- 0
probs.past[2] <- 1

jpeg(file=paste(dirNew1, "/", "4 - Observation_MClust_Probabilities_Past.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(probs.past, col=my_palette_probs, main = "Confidence of assignment 1982-1986", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
par(new=TRUE)
plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
dev.off()

probs.present[1] <- 0
probs.present[2] <- 1

jpeg(file=paste(dirNew1, "/", "6 - Observation_MClust_Probabilities_Present.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(probs.present, col=my_palette_probs, main = "Confidence of assignment 2013", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
par(new=TRUE)
plot(Antro, col= "grey", legend=FALSE, cex.lab=1, cex.axis=1, cex.main=2.5, cex.sub=1)
dev.off()

##########################################
# CROPPED CLIMATOLOGIES
Dir.Climats.Cropped <- paste(Dir.Change.Analysis, "/", Where, "/X - Climatologies", sep="")
dir.create(Dir.Climats.Cropped)
setwd(Dir.Climats.Cropped)

### NDVI
# full
climats <- matrix(Matrix.Full[,2], ncol=n.col, byrow=by.row)
```

```
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_1982-2013.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[1], "1982-2013", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

# past
climats <- matrix(Matrix.past[,2], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_1982-1986.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[1], "1982-1986", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

# present
climats <- matrix(Matrix.present[,2], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_2009-2013.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[1], "2009-2013", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

### NDVI Seasonality
# full
climats <- matrix(Matrix.Full[,3], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_Seasonality_1982-2013.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste(variables.long[2], "1982 - 2013", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

# past
climats <- matrix(Matrix.past[,3], ncol=n.col, byrow=by.row)
```

```r
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_Seasonality_1982-1986.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[2], "1982 - 1986", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

# present
climats <- matrix(Matrix.present[,3], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("bisque3","yellow","springgreen", "darkgreen"))(10000)
climats[1] <- 1
climats[2] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/NDVI_Seasonality_2013.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[2], "2013", sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black",
  legend.width=2)
dev.off()

### PREC
climats <- matrix(Matrix.Full[,4], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("yellow","blue", "darkblue", "royalblue"))(10000)
climats[1] <- 0

jpeg(file=paste(Dir.Climats.Cropped, "/Prec.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[3], sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

### TEMP
climats <- matrix(Matrix.Full[,5], ncol=n.col, byrow=by.row)
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- colorRampPalette(c("darkblue","royalblue","yellow", "red"))(10000)
if(Where == "Alaksa"){
  climats[1] <- 250
}

jpeg(file=paste(Dir.Climats.Cropped, "/Temp.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(climats, main = paste (variables.long[4], sep=" "), col=col.Climat, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

### DEM
climats <- matrix(Matrix.Full[,6], ncol=n.col, byrow=by.row)
```

```
climats <- raster(climats, xmn=x1, xmx=x2, ymn=y1, ymx=y2)
col.Climat <- col_elevation

if(minValue(climats)<0){
  jpeg(file=paste(Dir.Climats.Cropped, "/DEM.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(climats, main ="Digital Elevation Model (GMTED2010) [m]",col= col_elevation, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2, breaks=seq(0,maxValue(climats),1), axis.args=list(at=c(minValue(climats),seq(0,maxValue(climats),200),maxValue(climats)), labels=c(minValue(climats),seq(0,maxValue(climats),200),maxValue(climats)),cex.axis=0.9))
  dev.off()
}else{
  jpeg(file=paste(Dir.Climats.Cropped, "/DEM.jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(climats, main ="Digital Elevation Model (GMTED2010) [m]", col= col_elevation, cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
  dev.off()
}

##########################################
# GIMMS TIME SERIES
print("Loading the NDVI time series")
Dir.Gimms.Series <- paste(mainDir, "/X - Raw.Data/1 - Gimms-Annual.Mean/Monthly", sep="")
setwd(Dir.Gimms.Series)
TS <- list.files(path = Dir.Gimms.Series, pattern = ".grd")
TS <- mixedsort(TS, decreasing=FALSE) # sort in correct order

TimeSeries <- stack(TS)
TimeSeries <- crop(TimeSeries, area)
TimeSeries <- mask(TimeSeries, Shapes[location,])

##########################################
# CHANGE ANALYSIS
setwd(Dir.Change.Analysis)
rm(DEM) # removing DEM to avoid confusion of the term

ROC.Matrix <- matrix(-8888, nrow = length(climats), ncol = NClusters)

for(j in 1:NClusters){

  # SET WORKING DIRECTORY FOR AREA
  Dir.Biome.Ind <- paste(dirNew1, "/Biome_",j, sep="")
  dir.create(Dir.Biome.Ind)

  binary <- classes
  binary[binary == j] <- 888
  binary[binary != 888] <- 0
  binary[binary == 888] <- 1

  col.binary <- colorRampPalette(c("white", my_palette_classes[j]))(2)

  jpeg(file=paste(Dir.Biome.Ind, "/", "1 - Distribution_Past_Biome_", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(binary, col = col.binary, main = "Biome Distribution according to mclust in 1982-2013", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend=FALSE)
```

```r
par(xpd = TRUE)
legend("topright", inset=c(0,0), legend = c("Presence", "Absence"), fill = c(my_palette_classes[j],"white"), ncol=legendcol)
dev.off()

# DATA FRAME FOR DATA ON WHICH MODEL IS TO BE BUILT
biome.df <- data.frame(biome = as.vector(binary), temp = as.vector(Matrix.Full[,5]), prec = as.vector(Matrix.Full[,4]),DEM = as.vector(Matrix.Full[,6]))

# ESTABLISH MODEL
print(paste("Caclulating GAM for Biome", j, sep=" "))
model <- gam(biome ~ s(temp) + s(prec) + s(DEM), data = biome.df, family = binomial(link = logit))
print("GAM calculated")

setwd(Dir.Biome.Ind)
capture.output(summary(model), file = paste("Model_Biome",j,".R", sep=""), append = FALSE, type="output")

jpeg(file=paste(Dir.Biome.Ind, "/", "x - ModelStatistics", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
op <- par(mfrow = c(2,2), mar=c(5,4,1,2))
gam.check(model)
dev.off()

jpeg(file=paste(Dir.Biome.Ind, "/", "x - ModelSmoothers", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
op <- par(mfrow = c(2,2), mar=c(5,4,1,2))
plot(model)
dev.off()

# PREDICTION ON COMPLETE DATA, RETURNS PROBABILITIES OF ASSIGNING PIXEL X TO BIOME Y
dist <- predict.gam(model, newdata = biome.df, type="response")

dist.curr <- matrix(as.vector(dist), ncol=n.col, byrow=by.row)
dist.curr <- raster(dist.curr, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

dist.curr[1] <- 1
dist.curr[2] <- 0
jpeg(file=paste(Dir.Biome.Ind, "/", "2 - Biome_Probabilites_GAM_", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
plot(dist.curr, col=my_palette_probs, main = "Distribution probabilities according to GAM", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend.width=2)
dev.off()

#######################################
# RESPONSE CURVES
# tmean
response.df <- biome.df[,2:4]
response.df[,2] <- rep(mean(response.df[,2], na.rm = TRUE), length(response.df[,2]))
response.df[,3] <- rep(mean(response.df[,3], na.rm = TRUE), length(response.df[,3]))
response.df <- response.df[order(response.df[,1]),] # order values to avoid spaghetti plots
```

```
response.df <- na.omit(response.df)
response.dist <- predict.gam(model, newdata = response.df, type="response")

loess_fit <- loess(response.dist ~ response.df[,1], response.df, span = 0.1)
loessline <- predict(loess_fit)
loessline[loessline <0] <- 0
loessline[loessline >1] <- 1

jpeg(file=paste(Dir.Biome.Ind, "/", "3 - ResponseCurve_Temp_Biome_", j,"_",
    Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality
    = 100, res = 1000)
plot(x = response.df[,1],y = response.dist, main = paste("Response curve of
    biome assignment to variation in temperature data - Biome", j, sep=" "),
    xlab = "Temperature [K]", ylab="Probability of correct Biome association"
    , pch=20, type="l", ylim = c(0,1))
lines(response.df[,1], loessline, col = "green", lwd=3)
dev.off()

# prec
response.df <- biome.df[,2:4]
response.df[,1] <- rep(mean(response.df[,1], na.rm = TRUE), length(response.
    df[,1]))
response.df[,3] <- rep(mean(response.df[,3], na.rm = TRUE), length(response.
    df[,3]))
response.df <- response.df[order(response.df[,2]),] # order values to avoid
    spaghetti plots
response.df <- na.omit(response.df)
response.dist <- predict.gam(model, newdata = response.df, type="response")

loess_fit <- loess(response.dist ~ response.df[,2], response.df, span = 0.1)
loessline <- predict(loess_fit)
loessline[loessline <0] <- 0
loessline[loessline >1] <- 1

jpeg(file=paste(Dir.Biome.Ind, "/", "5 - ResponseCurve_Prec_Biome_", j,"_",
    Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality
    = 100, res = 1000)
plot(x = response.df[,2],y = response.dist, main = paste("Response curve of
    biome assignment to variation in precipitation data - Biome", j, sep=" ")
    , xlab = "Precipitation [mm]", ylab="Probability of correct Biome
    association", pch=20, type="l", ylim = c(0,1))
lines(response.df[,2], loessline, col = "green", lwd=3)
dev.off()

# DEM
response.df <- biome.df[,2:4]
response.df[,1] <- rep(mean(response.df[,1], na.rm = TRUE), length(response.
    df[,1]))
response.df[,2] <- rep(mean(response.df[,2], na.rm = TRUE), length(response.
    df[,2]))
response.df <- response.df[order(response.df[,3]),] # order values to avoid
    spaghetti plots
response.df <- na.omit(response.df)
response.dist <- predict.gam(model, newdata = response.df, type="response")

loess_fit <- loess(response.dist ~ response.df[,3], response.df, span = 0.1)
loessline <- predict(loess_fit)
loessline[loessline <0] <- 0
loessline[loessline >1] <- 1

jpeg(file=paste(Dir.Biome.Ind, "/", "7 - ResponseCurve_DEM_Biome_", j,"_",
```

```
    Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(x = response.df[,3],y = response.dist, main = paste("Response curve of biome assignment to variation in elevation data - Biome", j, sep=" "), xlab = "Elevation [m]", ylab="Probability of correct Biome association", pch=20, type="l", ylim = c(0,1))
  lines(response.df[,3], loessline, col = "green", lwd=3)
  dev.off()

  ##########################################
  # OBSERVED RELATIONS
  # tmean
  jpeg(file=paste(Dir.Biome.Ind, "/", "4 - ObservationCurve_Temp_Biome_", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(x = biome.df[,2],y = as.vector(binary), main = paste("Observed Biome presence in relation to variation in temperature data - Biome", j, sep=" "), xlab = "Temperature [K]", ylab="Biome presence/absence", yaxt='n')
  axis(2, at=c(0,1),labels=c("Absence", "Presence"), col.axis="black", las=3, pch=20, cex=1)
  dev.off()
  # prec
  jpeg(file=paste(Dir.Biome.Ind, "/", "6 - ObservationCurve_Prec_Biome_", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(x = biome.df[,3],y = as.vector(binary), main = paste("Observed Biome presence in relation to variation in precipitation data - Biome", j, sep=" "), xlab = "Precipitation [mm]", ylab="Biome presence/absence", yaxt='n')
  axis(2, at=c(0,1),labels=c("Absence", "Presence"), col.axis="black", las=3, pch=20, cex=1)
  dev.off()
  # DEM
  jpeg(file=paste(Dir.Biome.Ind, "/", "8 - ObservationCurve_DEM_Biome_", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(x = biome.df[,4],y = as.vector(binary), main = paste("Observed Biome presence in relation to variation in elevation data - Biome", j, sep=" "), xlab = "Elevation [m]", ylab="Biome presence/absence", yaxt='n')
  axis(2, at=c(0,1),labels=c("Absence", "Presence"), col.axis="black", las=3, pch=20, cex=1)
  dev.off()

  ##########################################
  # TIME SERIES
  BiomeTS <- TimeSeries
  BiomeTS[binary == 0] <- NA
  TS.vec <- rep(NA, length(TS))
  for(i in 1: length(TS)){
    TS.vec[i] <- mean(as.vector(BiomeTS[[i]]), na.rm = TRUE)
  }

  jpeg(file=paste(Dir.Biome.Ind, "/", "9 - NDVITimeSeries_FULL_Biome", j,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
  plot(TS.vec, col = "green", main = paste("NDVI Time Series - Biome ", j, " (", Where ,")", sep=""), type="line", ylim = c(0,1), lwd = 3, ylab = "NDVI", xlab = "Time since 01/1982 [months]")
  dev.off()

  jpeg(file=paste(Dir.Biome.Ind, "/", "10 - NDVITimeSeries_1982-1988_Biome", j
```

```r
,"_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm",
    quality = 100, res = 1000)
plot(TS.vec[1:84], col = "green", main = paste("NDVI Time Series - Biome ",
    j, " (", Where ,")", sep=""), type="line", ylim = c(0,1), lwd = 3, ylab =
    "NDVI", xlab = "Time since 01/1982 [months]")
abline(v = 12, col = "blue", lwd=2)
abline(v = 24, col = "blue", lwd=2)
abline(v = 36, col = "blue", lwd=2)
abline(v = 48, col = "blue", lwd=2)
abline(v = 60, col = "blue", lwd=2)
abline(v = 72, col = "blue", lwd=2)
dev.off()

##########################################
# ROC/AUC
dist.curr[1:2] <- NA
labels <- factor(as.vector(binary)[!is.na(as.vector(binary))])
predictions <- as.vector(dist.curr)[!is.na(as.vector(binary))]

jpeg(file=paste(Dir.Biome.Ind, "/", "11 - ROC_Biome", j,"_", Where, ".jpg",
    sep = ""), width = 32, height = 22, units = "cm", quality = 100, res =
    1000)
plot(1, type="n", xlab="", ylab="", xlim=c(0, 1), ylim=c(0, 1) , main =
    paste("Receiver Operating Characteristic (ROC) - Biome ", j, " (", Where,
    ")", sep=""))
par(new = TRUE)
ROC <- ROC(predictions, labels, plot = "ROC")
dev.off()

# sensitivity value at maximizing cutoff point (sensitivity + specifcity =
    MAX)
opt <- which.max(rowSums(ROC$res[, c("sens", "spec")]))
# optimal cut-off point
MaxSens <- ROC$res$predictions[opt]

# re-define my_palette_probs for histograms with colour gradient
my_palette_probs <- colorRampPalette(c("darkred","brown2","chocolate1","
    darkgoldenrod1","darkkhaki","darkolivegreen1", "chartreuse"))(n=100)
dist.curr[1] <- 1

# PLOTTING ASSIGNMENT-PROBABILITIES IN HISTOGRAMM, INCLUDING ROC-THRESHOLD
jpeg(file=paste(Dir.Biome.Ind, "/", "12 - Histogram_ROC_Biome", j,"_", Where
    , ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality =
    100, res = 1000)
hist(dist.curr,main=paste("Assignment Confidence Scores - Biome ", j, " (",
    Where, ")", sep=""),xlab=paste("Confidence", sep=" - "), border="black",
    col=my_palette_probs, breaks=100)
abline(v = MaxSens, col = "blue", lwd=2)
dev.off()

# PLOTTING BIOME DISTRIBUTION WITH ROC-THRESHOLD
ROCras <- dist.curr
ROCras[ROCras >= MaxSens] <- 1
ROCras[ROCras != 1] <- 0

ROCras[1] <- 1
ROCras[0] <- 0

jpeg(file=paste(Dir.Biome.Ind, "/", "13 - Distribution_ROC_Biome", j,"_",
    Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality
    = 100, res = 1000)
```

```r
      plot(ROCras, col = col.binary, main = "Biome Distribution according to GAM prediction with ROC threshold", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, colNA="black", legend=FALSE)
      par(xpd = TRUE)
      legend("topright", inset=c(0,0), legend = c("Presence", "Absence"), fill = c(my_palette_classes[j],"white"), ncol=legendcol)
      dev.off()

      ROC.Matrix[,j] <- as.vector(ROCras)

    } # loop for each biome

    ##########################################
    # COMBINED ROC-ANALYSIS
    # build rowsumms of ROC.Matrix and make resulting vector into raster
    ROC.states <- rowSums(ROC.Matrix)
    ROC.states <- matrix(ROC.states, ncol=n.col, byrow=by.row)
    ROC.states <- raster(ROC.states, xmn=x1, xmx=x2, ymn=y1, ymx=y2)

    jpeg(file=paste(dirNew1, "/", "8 - MultipleStates_ROC_", Where, ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
    plot(ROC.states, main = "Possible biomes per cell according to GAMs and ROC", col = colorRampPalette(c("grey","cadetblue", "cyan", "yellow", "darkgoldenrod","orange","red"))(NClusters+1), breaks = c(0, seq(0.5, NClusters+0.5,1)), colNA="black", cex.lab=1, cex.axis=1, cex.main=2, cex.sub=1, legend.shrink=1, legend.width=2, axis.args=list(at=c(0.25, seq(1, NClusters, 1)),labels=c(0, seq(1, NClusters, 1))))
    dev.off()

    # MCLUST UNCERTAINTY PLOT
    jpeg(file=paste(dirNew1, "/", "X - Mclust-Clusters", ".jpg", sep = ""), width = 32, height = 22, units = "cm", quality = 100, res = 1000)
    plot(mod, what="uncertainty", col=my_palette_classes, main = "Mclust clustering")
    dev.off()

    ##########################################
    # CLEARING TEMPORARY DIRECTORY FOR NEXT RUN
    setwd(dirtemporary)
    print("Clearing Temporary Directory")
    fileNames.delete <- list.files(path= dirtemporary)
    do.call(file.remove, list(fileNames.delete))
} # end of function


##################################
####### RUN THE FUNCTION:
##################################

### Minnesota, provinces
CHANGE.FUN(region <- "State", Where <- "Minnesota", clusters <- 4, sample.num <- "All", DEM <-"mn")
### Alaska, Scheffer
CHANGE.FUN(region <- "State", Where <- "Alaska", clusters <- 5, sample.num <- 20000, DEM <-"mn")

print("done")
```

A.7 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