A RANGELAND PREDICTIVE PHENOLOGICAL MODEL FOR THE UPPER COLORADO RIVER BASIN AND ITS WEB DELIVERY

by

Yuan Zhang

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Geography

The University of Utah

August 2013

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Yuan Zhang**

has been approved by the following supervisory committee members:

George F. Hepner	, Chair	**06/13/2013** Date Approved
Philip E. Dennison	, Member	**06/13/2013** Date Approved
Richard R. Forster	, Member	**06/13/2013** Date Approved
Ikuho Yamada	, Member	**06/13/2013** Date Approved
Bruce K. Wylie	, Member	**06/13/2013** Date Approved

and by **George F. Hepner** , Chair of

the Department of **Geography**

and by Donna M. White, Interim Dean of The Graduate School.

ABSTRACT

Understanding the spatially and temporally variant phenological responses and cycles can greatly assist the administrative planning, policy making and management in grazing, planting, and ecosystem conservation. The linkages of analysis as a basis for management have received increasing attention in the context of climate change. This research focuses on analyzing phenological responses of vegetation as constrained and moderated by environmental factors, such as landscape and season, in the geographically diverse Upper Colorado River Basin (UCRB). Due to the geographic diversity of phenological forcing in the UCRB, several homogeneous phenological subregions (phenoregions) are delineated, and the phenological responses of vegetation are analyzed on a per phenoregion basis. A multivariate adaptive regression splines (MARS) approach is adopted to model and interpret the regionally and seasonally specific relationships between environmental drivers (temperature, precipitation and solar radiant energy) and vegetation abundance, indicated by a Vegetation Index (VI). Short-term predictions of vegetation abundance are made using the models. Taking into consideration the scale of the study area and the time-step of the models, 1 km 7-day interval eMODIS data and the 1 km NASA AMES Ecocast data are used to articulate the dependent and independent variables. The series of models are integrated into a prototype phenological Decision Support System (DSS) to provide predicted vegetation abundance over the growing season and the trends of climatic variables leading to potential grazing management

strategies. The implementation of the DSS is a unique attempt to integrate phenological theory and GIS technology, the combination of which makes this DSS analytically-based, intuitive and more user-friendly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude, first and foremost, to my advisor, Dr. George Hepner for his continuous guidance and support with his knowledge, experience, patience and motivation throughout my dissertation journey. I would not have been able to hurdle over all the obstacles in the completion of my dissertation without him. I would also like to thank my committee members, Drs. Philip Dennison, Richard Forster, Ikuho Yamada, and Bruce Wylie, for their very valuable insights and suggestions. My thanks also go to the fellow graduate students and other friends who helped, encouraged, and supported me during my five years as a Ph.D. student at the University of Utah. I also would like to thank the Bureau of Land Management for funding the project, and the NASA AMES Ecological Forecasting Lab for providing the essential Ecocast data. Lastly, I am beyond grateful to my parents for their nonstop and selfless support throughout my life.

1 INTRODUCTION

Rangeland mismanagement due to lack of scientific tools and information support in the Upper Colorado River Basin (UCRB) has caused severe problems of rangeland degradation and low productivity. The improvement of rangeland management requires a more robust analytical support system. This system would consist of better, more current information and a predictive modeling capability as part of a Decision Support System (DSS) for managers. The DSS would generate new data, incorporate, analyze and interpret existing data from different data sources, and consequently provide solid information and decision support. The DSS, by assisting with precise, reasonable and reliable rangeland management information and strategies, can help solve the dilemma of achieving a balance between economic output and rangeland health.

Initially, this dissertation delineates phenologically-based units (phenoregions) for the UCRB. A set of phenological models (phenomodels) is developed and validated to make short-term predictions of vegetation abundance in the UCRB, including the rangeland managed by the BLM (US Department of the Interior, Bureau of Land Management). In addition, a prototype DSS is created to present the phenological data, models and supplemental information to land managers in an understandable manner. The set of phenomodels is a crucial component of the DSS, assisting in making reasonable grazing strategies. The DSS must integrate predictive phenological models and a suite of information from different data sources within a Geographic Information

System (GIS). This approach will help land managers in the BLM with improved rangeland management.

1.1 Public Rangeland Management and Issues

The United States has about 3.1 million km^2 of rangeland (National Research Council, 1994). Forty-three percent of rangeland in the US (about 1.4 million km^2) is owned by the federal government (National Research Council, 1994), known as public rangeland. Rangeland in the US can provide forage for livestock and wild grazing animals, habitat for wildlife, as well as commodities and recreation for human. Well-managed rangeland can further help support wildlife diversity, healthy watersheds and carbon sequestration, and thereby sustain rangeland ecosystems (BLM, 2011; National Research Council, 1994).

The BLM, under the Department of the Interior, was formed in 1946 by merging the General Land Office and the US Grazing Service to manage 1 million km^2 of public land with multiple uses, 0.64 million km^2 of which is within grazing districts (BLM, 2011; Dombeck et al., 2003). Most of the BLM managed rangeland (~0.55 million km^2) is in the west (Dombeck et al., 2003).

The BLM developed a series of standards and guidelines in the 1990s to maintain and promote long-term health and productivity of public rangeland (BLM, 2011; National Research Council, 1994). However, with so much effort needed to maintain rangeland health, rangeland conditions under BLM administration are still unsatisfying if not declining.

1.1.1 Limited Rangeland Resources

The area of available rangeland for grazing is decreasing. More than 0.04 million km^2 of rangeland has been lost because of development and crop cultivation (Maczko et al., 2004). Scenarios have been projected that the amount of land available for forage production will continue to decrease from 2000 to 2050 (Van Tassell et al., 2001).

BLM uses the Animal Unit Month (AUM) to provide information on overall livestock use. An AUM is the amount of forage needed to sustain one cow and her calf, one horse, or five sheep or goats for a month (BLM, 2011). Authorized grazing use indicated by AUMs is continuing to decrease, from 18.2 million AUMs in 1953 to 12.4 million AUMs in 2010 due to degrading and shrinking rangeland (BLM, 2011).

1.1.2 Range Degradation

According to the most recent rangeland assessment data in 2010, 8% of BLM managed rangeland was in potential natural community condition, 35% in late seral condition, 41% in midseral condition, and 15% in early seral condition. Potential natural community, late seral, midseral, and early seral are categories of ecological status expressed as the degree of similarity of present vegetation to the potential natural, or climax, plant community and are, respectively, 76 - 100%, 51 - 75%, 26 - 50% and 0 – 25% similarity (BLM, 2011).

The degradation of rangeland in the US has been mainly due to overgrazing of livestock, while drought, erosion, mining, off-road vehicle use and weeds further aggravate this condition (Dombeck at. el., 2003; National Research Council, 1994). Lack of sufficient information makes it difficult for the land managers to decide when to start and stop grazing for a specific pasture. Overgrazing prevents rangeland from fully

recovering or prolongs rangeland recovery. Rangeland degradation can severely affect the economic and ecological values and products provided by rangeland (National Research Council, 1994).

1.1.3 Low Productivity

Land degradation leads to low productivity of public rangelands. In the US, 54% of private rangeland produces 93% of all AUMs, while 46% of public rangeland produces only 7% (National Research Council, 1994). The costs of public rangeland management have been high, far beyond the economic values generated by livestock grazing (Nelson, 1995). The actual situation varies from allotment to allotment and from pasture to pasture within each allotment. However, there is not enough information and scientific insight at high geographic resolution for federal government personnel and land managers to make allotment-specific decisions (Nelson, 1995).

Thus, land degradation resulting in lowered productivity is an outcome of a lack of proper management. Improper rangeland management is the direct consequence of the inability to utilize appropriately scaled climatic, vegetative, land use / cover information integrated with management goals in an effective way at the phenoregion and allotment levels.

Scientific insights and tools are therefore greatly needed to provide necessary and sufficient information to support improved rangeland management. The DSS developed in this research integrates predictive phenomodels, summarizes phenological forcing trends, and refines grazing suggestions and other information. This is believed to be able to contribute to the mitigation of rangeland degradation, boost rangeland health and improve economic outputs.

1.2 Phenology

The word "phenology" originates from the Greek word *phaino*, meaning to show or to appear (Schwartz, 2003). Phenology studies the relationship between periodic biological phenomena and climatic conditions – how organisms grow and behave in response to environmental conditions (Hodges, 1991). Budburst and flowering of vegetation, hibernation of bears and snakes and the migration of birds are all typical phenological events.

Phenological events have been observed and studied since ancient times. The most ancient records and literature have been found in China, written more than 3000 years ago. The longest written phenological record originates in Japan beginning in 705 AD regarding the initiation of cherry flowering (Schwartz, 2003). However, these ancient records of phenology are just simple data collections and applied exclusively in agriculture. It was not until the late 19th century that phenology entered the category of science as a branch of environmental science (Sparks and Jeffree, et al., 2000). At that time, it was realized that phenology is useful in monitoring the environment and understanding the relationships between the environment and phenological events.

Plant phenology (or vegetation phenology) is a main branch of phenology science. It studies how periodic biological phenomena of plants (budburst, flowering, fruiting, senescence and dormancy, which are called "phenological phases" or "phenophases") are influenced and driven by climatic variation at specific locations. Higher temperatures can accelerate plant development leading to earlier onset of phenological events (Badeck et al., 2004; Fitter et al., 1995; Sparks and Carey, 1995; Sparks et al., 2000). Two locations with a mean annual temperature difference of about 5°C can cause the onset of greenup

to differ by as much as a month in the UCRB. Precipitation can affect the timing of different phenophases and accounts for a significant amount of the phenological variation in moisture-limited regions like the UCRB (Reed et al., 1994; Peñuelas et al., 2004; Pickup et al, 1994). Solar radiant energy is a prerequisite for photosynthesis. The variation of solar radiant energy directly affects the production of simple sugars and further the growth of vegetation. The UCRB is one of the regions receiving the highest solar radiation is the US. It is important to consider how solar radiation affects plant phenology in the UCRB.

Since the late 1800s and early 1900s, focus has been put on discovering the factors that can cause different plant phenological events (Blodget, 1857; Garner and Allard, 1920, 1930; Livingston, 1916; Parker and Borthwick 1939; Waldo, 1893; Whitley, 1850). Subsequently, research has focused on phenological modeling for simulation and prediction of phenophases.

In theory, if one can define and measure the relevant factors that affect phenological phase change, prediction of the location and timing of phase change should be possible. This research is an application of phenological theory: phenological theory serves as an important building block for phenomodel development and decision support in the DSS. This research in turn contributes to phenological theory by exploring the unique phenology of the UCRB.

1.3 Research Scope and Research Questions

The outstanding issues associated with BLM managed public rangeland necessitate the achievement of a tradeoff between rangeland health and economic outputs with limited rangeland resources. The key is to use scientific tools and information

support to yield improved rangeland management. Improved rangeland management can not only avoid the impacts of overgrazing on public land resources, it can, on the contrary, support a healthy ecosystem and environment.

Three objectives of the research are as follows:

1. To understand the relationships between vegetation growth and phenological forcing.
2. To develop a set of phenomodels to predict the vegetation abundance 7 days in the future.
3. To design a prototype DSS with phenomodels as a crucial module.

This research uses the Upper Colorado River Basin (UCRB) as the study area. The diversity of climate and terrain of the UCRB enables the extension of phenomodel development framework to other BLM managed lands and other regions with similar ecosystems, such as many rangelands in Australia, Argentina, New Zealand and South Africa.

The UCRB has more than range ecosystems, so this research focuses more than just on rangeland vegetation. A framework for predictive phenological model development is proposed that can be applied to all kinds of vegetation and to other geographically diverse regions.

Satisfactory achievement of research objectives requires knowledge related to four research questions below：

1. How can phenoregions be effectively delineated in the UCRB and how can performance of phenoregions be evaluated?

Due to the geographic diversity in the UCRB, phenoregions (phenologically and climatically self-similar clusters) need to be delineated in order to develop phenoregion-specific models. Several criteria are revealed to evaluate the performance of phenoregions.

2. What are the dependent and independent variables of the phenomodels?

   A Vegetation Index (VI) signifying the abundance of rangeland vegetation is identified as the dependent variable by exploring use of different VIs in the set of phenomodels. The phenological forcing variables affecting the dependent variable, such as temperature, precipitation and light are examined and established as the independent variables in each phenoregion.

3. What are the relationships between the dependent and independent variables identified in question two and how to validate these relationships as represented with a mathematical model?

   The relationships between the dependent and independent variables are explored for model development in each phenoregion. The models are validated to make sure they are effective for practical application, using both cross and independent field validation.

4. What information should be provided and what mechanisms should be adopted in the proposed DSS to most effectively assist land managers in formulating grazing strategies and decisions?

   A DSS is designed and implemented integrating modules of phenomodels, visualization, trend display and analysis, and grazing

information.

1.4 Design of Methodology

1.4.1 Study Area

The UCRB (Figure 1.1) is one of the main locations of large scale livestock grazing activity in the United States. Since the late 1500s, there has been livestock grazing in this area, primarily cattle and sheep. At present, most of the lands in the UCRB are used as rangeland. UCRB has a total area of about 0.29 million km^2 within which the BLM manages about 0.16 million km^2 of rangeland and the US Forest Service manages about 0.05 million km^2 of rangeland. Public rangeland takes up 72% of the entire UCRB. A quarter of the 0.64 million km^2 of BLM managed rangeland is within the UCRB. These figures make the UCRB an ideal study area to understand rangeland phenology and how the DSS could improve BLM rangeland management and health.

The majority of the concerns and issues regarding public rangeland conditions and management apply to the BLM managed rangeland in the UCRB. Despite enormous public administration cost and high input of lands and forages, the output of livestock value is fairly low in the UCRB. This makes grazing information a very important factor to solve the high-input and low-output problem.

The UCRB has highly diverse topography, latitude, soils and climate patterns. Such large tracts of land and extraordinary geographic diversity make the UCRB a worthwhile area for vegetation phenology to be examined and for a phenomodel development framework to be designed. This geographic diversity makes this research distinctive, innovative and challenging.

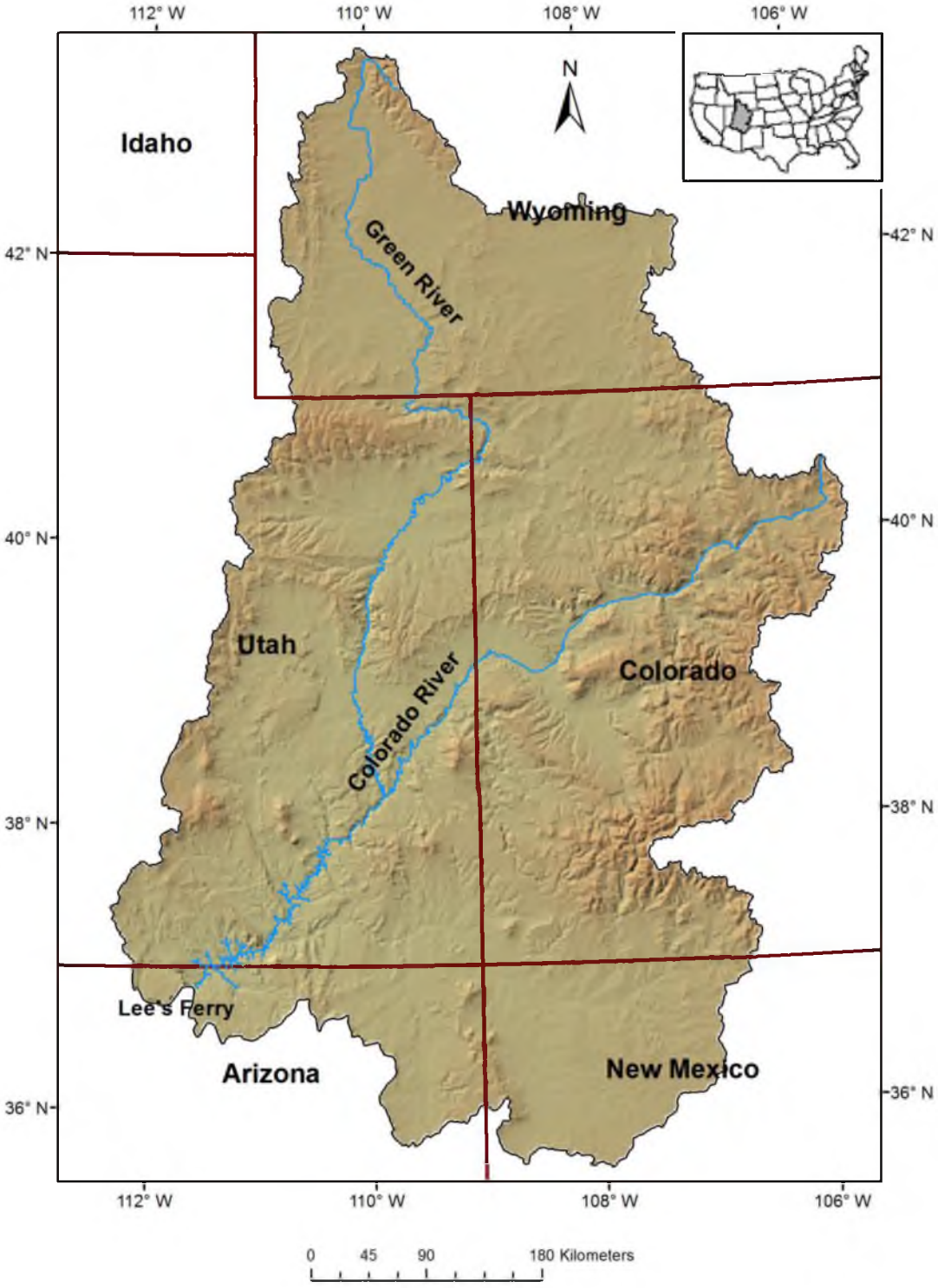


Figure 1.1 The UCRB study area and the location of UCRB within the conterminous US.

1.4.2 Phenoregion Delineation

The geographic diverse nature of the UCRB excluded the possibility of using a single model to depict the relationships between dependent and independent variables, and thus necessitates the delineation of phenology based subregions. Environmental variables related to climate, elevation, topography and vegetative history are identified and analyzed to generate phenoregions having similar phenological forcing. Because the UCRB is a highly diverse region, a quantitative approach (Principal Component Analysis plus *k*-means++ clustering) and high spatial resolution data are used to yield increased homogeneity within each defined phenoregion.

1.4.3 Phenological Models

The Multivariate Adaptive Regression Splines (MARS) approach is adopted in this research to conduct model development in each phenoregion. MARS models are built to represent the relationships between independent and dependent physical environmental variables. The MARS models can depict different relationships within different data intervals, automatically model interactions and are easy to interpret. Therefore, the straightforward, intuitive and more realistic characteristics of MARS models make it much easier to understand how the various environmental variables interact and how the environment as a whole impacts vegetation dynamics at different locations and different dates characterized by different ranges of environmental variables.

1.4.4 Decision Support System

An integrative DSS is developed to provide a suite of information to supplement and present the predicted results from the phenomodels. The implementation of the DSS

in this research is a unique attempt to integrate phenological theory and GIS technology, the combination of which makes DSS vivid, intuitive and more user-friendly.

Grazing management strategies are suggested based on the predicted vegetation abundance. Basic grazing strategies are based on a classification of Vegetation Index values and to evaluate each pasture within an allotment indicating the suitability and capacity of grazing. Historic temperature, precipitation data as well as historic, current and near-future vegetation indices are presented in the DSS in tables, inquiries or plots. The trend and historical situations portrayed in these data can further assist land managers with critical clues on phenophases and current and expected productivity.

1.5 Contributions

1.5.1 Theoretical Contributions

This research provided theoretical advancements in the biogeographical research. Phenoregion delineation in this research identified variables having spatial influence on vegetation phenological patterns and features. Phenoregion identification reveals how different environmental variables, individually and collectively, contribute to vegetation growth. The different models developed for different phenoregions disclose how the relationships between environment and vegetation dynamics differ under varied ranges of phenological forcing.

This research substantiates phenological theory. It helps to enrich this theory by exploring different vegetation responses to the environment in different phenoregions and under different ranges of environmental variables within the same phenoregion. It also examines impacts of variation of single as well as several environmental variables on vegetation dynamics within the study area of the UCRB.

1.5.2 Methodological Contributions

This research developed two analytical frameworks. One framework delineates subregions of similar phenological forcing (phenoregions) in geographically diverse regions. This is accomplished by using principal component analysis and *k*-means++ clustering to decompose the geographic diversity of topography, climatic conditions and vegetation, and finely differentiate the spatial variation in phenological forcing among different geographic locations in the UCRB. This framework has been demonstrated effective in achieving increased homogeneity within each defined phenoregion.

The second is the modeling framework that develops predictive phenological models in geographically diverse regions. The framework contains these elements: phenoregion delineation, variable identification, relationship exploration and phenomodel development / validation. This comprehensive approach to phenological modeling using the spatial analysis and visualization capability of a GIS is an innovative contribution.

1.5.3 Practical Contributions

The set of proposed models are used to predict the vegetation abundance 7 days in the future. Decisions of location and timing for grazing livestock can be suggested based on the predicted results in order to prevent overgrazing and nonsufficient recovery, and thus to help sustain the rangeland ecosystem and maintain or improve rangeland conditions.

Phenomodels in this research uses multiple and frequent data sources for its decision-making. These models provide superiority over current decision basis including mainly expert knowledge and infrequent local, discrete monitoring data. In this sense, the DSS can provide land managers with real-time, abundant and critical historic, current and

near-future information on vegetation abundance, phenoregions, grazing strategies and productivity under the scenario of real-time production of data fed into the phenomodels.

By delivering models and information online, land managers can relatively easily acquire the phenological status at specific locations by providing simple inputs. Using the visualization module to display the predicted results can further assist the land managers with intuitive and vivid overviews about the rangeland phenological situation in the UCRB. The adoption of GIS technology enables the DSS to intuitively interact with land managers.

1.6 Organization of Dissertation

The rest of this dissertation is organized as follows: Section 2 provides readers with a background knowledge of phenology, remote sensing, environmental drivers and factors that influence plant growth and development, and phenological models; Section 3 discusses in detail the methodology of phenoregion delineation, and the model development including selection of dependent and independent variables, data sources and processing, relationship analysis, and MARS model development; Section 4 presents the analysis and results correspondent to each component of the methodology; Section 5 discusses the enhancements of rangeland phenomodels, the possibility of using other models and the influence of the pixelated nature of remote sensing phenology. Section 6 is the conclusions reviewing the whole dissertation, summarizing important consensus and inferences reached and drawn from the preceding chapters and discusses future work.

2 BACKGROUND AND LITERATURE REVIEW

2.1 Remote Sensing and Plant Phenology

2.1.1 Remotely Sensed phenology

Remote sensing is the observation of the physical, chemical, and biological properties of the earth's land and water surface from a distance by means of reflected or emitted electromagnetic energy. The sensor can create remote sensing imagery by recording the electromagnetic energy reflected or emitted by the physical objects on the surface of the earth (Campbell, 2002; Longley et al., 2005).

Remotely sensed plant phenology uses satellite to track gradual phenological changes and phenophases at regional, continental, and global scales. It has been used as a supplement to, and sometimes even a substitute for ground observed phenology for several decades (Reed and Brown, 2005). Remote sensing has its superiority to traditional field phenological methods in phenological monitoring and forecasting, and is indispensable in the broad-scale phenological studies.

The ability of remote sensing to conduct frequent measurements over a vast region for a long period enables regular, broad-scale, and long-term monitoring of phenology (Reed and Brown, 2005; Zhang et al., 2003). Therefore, remote sensing provides a robust means to elevate the monitoring and modeling of plant phenological dynamics from single plant species, and vegetation communities to entire ecosystems; thereby helping to improve the understanding of broad-scale phenological trends, which

are usually very hard or even impossible for traditional ground observation to detect. It has been demonstrated that moderate to good agreement exists between remotely sensed observations and field measurements in terms of vegetation abundance (Jenkins et al., 2002; Schwartz et al., 2002; Fisher and Mustard, 2007; Fisher et al., 2007).

Rather than directly recording, monitoring or estimating specific dates of budburst or flowering used in traditional ground-based phenological research, remote sensing offers a way in which a continuum of processes can be monitored or predicted (White and Nemani, 2006). Remote sensing thus indirectly measures phenophases by monitoring the vegetation status throughout the year, and detecting changes in emitted or reflected electromagnetic radiation caused by these phenophases. Tracking or predicting specific transition dates between phenophases using remotely sensed data has thereby become one of the recent foci of phenological research. For example, Kaduk and Heimann (1996) used the biggest increase in a Vegetation Index to determine the timing of phenophases; White et al. (1997) used a set of thresholds on a Vegetation Index to determine the onset of different phenophases; Zhang et al. (2003) used the minimum and maximum in the rate of change on a Vegetation Index time series as the transition dates between seasons; Fisher et al. (2007) used the half-maximum greenness as a signal of the greenup onset.

Although principles of using remotely sensed data in phenological studies have been well established, close attention needs to be paid to the scaling from individual plants and species to broad-scale vegetation communities and ecosystems. Mixed species and vegetation land cover within a single pixel requires analytical and modeling methods different from those used in ground-based phenology assessment. Further, how the

temporal and spatial resolutions of remotely sensed data affect the environment-vegetation relationship and phenological transition detection remains under exploration.

2.1.2 Vegetation Indices

The use of multispectral sensors to measure the radiation within different spectral bands is the basis of the Vegetation Index (VI), a relatively common approach for monitoring phenology. Using a VI to track the temporal and spatial variations of plant phenology relative to climatic, topographic and edaphic conditions has been very successful and effective (Reed et al., 1994; Zhang et al., 2001; Huete et al., 2002; Ji and Peters, 2004; White and Nemani, 2006; Myneni et al., 1997). The fundamental principle of VIs is to enhance the difference between and normalize the reflectance of two different bands of the electromagnetic spectrum. For example, the differences between the reflectance of red and near infrared band can be used to create a VI because the intense absorption in red band (driven by chlorophyll absorption) and intense reflectance in near infrared band (driven by leaf mesophyll and canopy structure) are unique features of the actively growing vegetation. The unique spectral features make the VIs that use these bands effective indicators of photosynthetically active vegetation.

Popular remote sensing-based vegetation indices include the Normalized Difference Vegetation Index (NDVI), the Soil Adjusted Vegetation Index (SAVI), the Enhanced Vegetation Index (EVI), and the Normalized Difference Water Index (NDWI). Each index is used in different situations dependent on soil background, atmospheric conditions, solar zenith angle and sensor calibration. Soil background and atmospheric conditions are the two main factors that affect vegetation indices (Campbell, 2002).

The NDVI is the most commonly used VI for monitoring vegetation phenology (Ji and Peters, 2004; Reed, et al., 1994; White and Nemani, 2006; Zhang, et al., 2001; Zhang, et al., 2009). NDVI is constructed based on the difference between the reflectance of the red and near infrared band, and is calculated as

$$NDVI = \frac{\rho_{NIR} - \rho_{\mathrm{Red}}}{\rho_{NIR} \_ \rho_{\mathrm{Red}}}$$

where ρ_{NIR} and ρ_{Red} are the reflectance in the near-infrared and red bands, respectively. The NDVI reduces most errors related to sensor calibration, sun zenith angle, landscape, and atmospheric conditions, by canceling out signal variations caused by these situations through ratioing. NDVI values vary from -1 to 1. The NDVI value of vegetation is typically from 0.2 to 0.8 and values increase with denser vegetation canopies. NDVI has been shown to be closely related to Leaf Area Index (LAI), Net Primary Production (NPP), green biomass, and Fraction of Absorbed Photosynthetically Active Radiation (fAPAR) (Running and Nemani, 1988; Tuck et al., 1981). Despite its popularity in phenological research, NDVI has some limitations. It is easily saturated in multilayer canopies, and it is sensitive to atmospheric aerosols and soil background (Huete et al., 2002, Xiao et al., 2003; Xiao et al., 2005; Sirikul, 2007). NDVI is also more sensitive to biophysical parameters. For example, Ji and Peters (2007) demonstrate that NDVI, compared to other VIs, is more sensitive to moderate and low Leaf Area Index (LAI), which is usually defined as one-sided green leaf area per unit ground surface area (Watson et al., 1947).

SAVI has a similar form to NDVI, but reduces sensitivity to soil background:

$$SAVI = \frac{(\rho_{NIR} - \rho_{\mathrm{Red}})(1+L)}{\rho_{NIR} + \rho_{\mathrm{Red}} + L}$$

where L is the soil adjustment factor (Huete, 1988). SAVI values range from -1 to 1. SAVI minimizes the influence of soil background by introducing L into the equation to compensate soil noise. A value of 0.5 for L is reported to work well for a wide range of soil and vegetation amounts in field measured data without the influence of atmosphere (Huete, 1988). However, atmospheric variations can cause the unstable performance of SAVI, and may reinduce soil noise into the index (Liu and Huete, 1995). SAVI has many modified or transformed versions. The TSAVI (Transformed Soil Adjustment Vegetation Index) was proposed by Baret et al. in 1989 to minimize soil brightness. The SAVI2, SAVI3 and SAVI4 were proposed by Major et al. in 1990 based on the theoretical consideration of the effects of wet and dry soils. The MSAVI (Modified Soil Adjustment Vegetation Index) was developed by Qi et al. in 1994 to minimize the effect of bare soil. These series of VIs are all effective for low vegetation cover areas. For example, the NDVI is only useful when the vegetation cover is more than 30% while the SAVI and SAVI-like VIs are still effective when the vegetation cover is as low as 15% (Gibson and Power, 2000).

The EVI considers both the effects of soil background and atmospheric conditions (Liu and Huete, 1995),

$$EVI = G\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + C_1 \times \rho_{Red} - C_2 \times \rho_{Blue} + L}$$

where L is the soil adjustment factor, G is the gain factor, $C_1 = 6.0$ and $C_2 = 7.5$. EVI ranges from -1 to 1. The EVI has an improved sensitivity in high biomass regions by separating the soil background signal and reducing the atmospheric influence (Huete et al., 2002). There are many studies using EVI to indicate the vegetation abundance (Peñuelas et al., 2004; Zhang, et al., 2003; Zhang et al., 2009).

The NDWI is a slightly different VI proposed by Gao (1996) to measure liquid water absorption by vegetation canopies. It uses two narrow channels centered at 0.86 μm and 1.24μm, and is calculated as:

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

Because both channels used in NDWI are hardly affected by atmospheric aerosol scattering, NDWI is less sensitive to atmospheric aerosols as compared with NDVI. It is sensitive to the total amounts of water in vegetation due to the different water absorption features in the two channels. NDWI is still affected by soil background. NDWI values range from -1 to 1 and are positive for green vegetation and negative for dry vegetation.

2.2 Plant Phenology and Geography

Plant phenology, as an important branch of phenological research, studies how periodic biological events in the plant world (such as sprouting and flowering) are influenced by environmental changes driven by weather and climate (Schwartz, 2003). Environmental drivers of these periodic biological events have received increased attention with the ongoing climate change (Dahlgren et al., 2007).

Phenological study, as the definition implies, studies the influence of these drivers on plant development with the focus on the temporal variations of the drivers. At the same time, to a geographer, the spatial variations of the drivers are equally, if not more, important than the temporal variations.

These drivers and their relative importance vary depending on climatic and biome patterns, but it is widely accepted that the main drivers of plant growth and development are temperature, precipitation and light (Lindsey and Newman, 1956; Campbell and

Sugano, 1975; Prins and Loth, 1988; Reed et al., 1994; Fitter et al. 1995; Sparks and Carey 1995; Sparks et al. 1997, 2000; Peñuelas et al., 2004; White et al., 1997). These drivers are referred to as environmental drivers in this dissertation with both of their spatial and temporal variations exerting influence.

The spatio-temporal influence of environmental drivers on plant development is moderated by the spatial variation of some time-invariant factors, notably edaphic conditions and landscape in terms of elevation, slope and aspect (Schwartz, 2003; Sharma, 2005; Batanouny, 2001). These factors are referred to in this dissertation as environmental factors as opposed to environmental drivers. The environmental factors are considered temporally stable, so only their spatial variations matter in the context of plant development. However, the influence of environmental factors on vegetation dynamics and their moderation on the environment-vegetation relationships may not be constant over time.

Under natural conditions, vegetation is not affected by an individual factor alone; rather, it is affected by all environmental drivers and factors at the same time as a whole – the environment. The drivers and factors themselves are interrelated. They influence each other and interact with each other. Therefore, plant growth is the results of not only the sum of, but the interactions between all these environmental drivers and factors.

2.2.1 Environmental Drivers

The spatial variations of environmental drivers are the main determinants of vegetation cover type and its distribution (Walter, 1973; Woodward, 1987; Whittaker, 1970; Suzuki et al., 2000). The temporal variations of phenological drivers are the direct causes of periodic plant life cycle events, as indicated in the definition of phenology.

Both spatial variations and interannual variations can cause the onset and offset of phenophases to differ (White et al., 1997; Dahlgren et al., 2007). The relationships between environmental drivers and status of vegetation abundance vary spatially and temporally (Kawabata et al., 2001; Suzuki et al., 2000; Schultz and Haplert, 1995). The detailed spatio-temporal influences of main environmental drivers on plant development (temperature, precipitation and light) are reviewed below.

2.2.1.1 Temperature

Temperature is one of the major factors that control plant activity, phenological cycles and plant distribution. From the ontogenetic point of view, temperature can affect plant mechanisms and physiology such as respiration, photosynthesis, and transpiration.

Photosynthesis starts at a low temperature, reaches an optimum and decreases at higher temperatures due to the impaired carbon metabolism and substance transportation (Schulze, 2005; Larcher, 2003). At extreme temperatures, photosynthesis reaches a complete standstill. For example, Xu and Baldocchi (2003) found that the photosynthesis processes of trees are prohibited when experiencing air temperatures exceeding 40°C.

Temperature is demonstrated to be a main factor influencing respiration rate (Sirikul, 2007; Schulze, 2005). Similar to photosynthesis, respiration also responds to temperature in the form of an optimum curve (Larcher, 2003). However, respiration is generally activated at a lower temperature as compared with photosynthesis. It increases exponentially over a wider range of temperatures and is only inhibited at very high temperatures (Schulze, 2005). At extremely high temperatures, the heat damage to enzymes and membrane structures causes respiration activity to stop (Larcher, 2003).

Net photosynthesis or net primary production as the balance between photosynthesis and respiration is directly related to the change of biomass over time. The optimum of net photosynthesis is moved slightly to a lower temperature as compared to gross photosynthesis (Schulze, 2005; Larcher, 2003), due to exponentially increasing respiration activity at higher temperatures. The optimal temperature at which net photosynthesis maximizes is not constant. It changes with season and vegetation species (Larcher, 2003; Schwarz, 1997; Sirikul, 2007).

There exist three ranges of temperatures: The optimal temperature range, the rigid (stiff) temperature range, and the lethal temperature range.

Plants grow at a maximal rate within the optimal temperature range. This range is plant specific, and differs among individual tissues and organs within a single plant (Larcher, 2003; Schulze, 2005). Temperatures outside of this optimal range can cause stress in plants and lead to growth anomalies (Schulze, 2005).

The rigid or stiff temperature range is defined by two critical values – the cold limit and the heat limit. Plant activities are limited to a minimum without lethal damage when temperature is around the heat limit and cold limit, because the net photosynthesis declines dramatically at too low or too high temperatures (Schulze, 2005). This situation occurs during the dormancy season, so that plants can survive extreme temperatures (Larcher, 2003). Most plants start to grow when temperature is above 10℃ , and stop growing when temperature is beyond 35℃ (Waugh, 2000).

The lethal temperature range is defined by two cardinal temperature tolerances, and plant development completely ceases and plants suffer lethal and permanent damage once temperature is outside of this range, due to the sudden destruction of cell structures

and cellular functions (Schulze et al., 2005; Larcher, 2003). The lethal limit is also species specific, and organ and tissue specific (Larcher, 2003). The lower lethal temperature limit ranges from less than -40℃ for boreal ecosystems to greater than 10℃ for cold-sensitive species in broadleaf evergreen forests (Smith et al., 1997). The upper threshold of lethal temperature is 42 - 56℃ for most plants (Dahl and Birks, 2007).

Antecedent temperature and cumulative heat supply are also crucial to plant development (Larcher, 2003). Hudson (2009) pointed out that the temperatures of preceding months are also important factors that influences the timing of phenophases in temperate zones (Menzel 2003; Estrella et al., 2007), and in some cases even temperatures from the previous autumn can affect spring phenology. Many measures of heat accumulation have been identified to be directly associated with plant growth and development. Generally, most phenophases are related to certain temperature thresholds and require a certain amount of cumulative heat supply, especially for fruits and seeds to ripen (Larcher, 2003). The heat needed to reach a certain phenophase differs for different species and at different locations (especially at different latitudes) even for the same species (Schwartz, 2003). For example, Strand (1965) found that the cumulative heat needed for the same crop plant variety to reach a phenophase generally decreased with increasing latitude (Schwartz, 2003).

The difference of temperature between day and night has a favorable effect on plant growth (Larcher, 2003). There is an optimal amplitude of diurnal temperature alteration associated with different vegetation species and vegetation in different climatic regions (Larcher, 2003). Larcher (2003) found that vegetation in continental regions has the best growth rate when the day/night temperature fluctuation is about 10 to 15 °C; the

optimal fluctuation for some desert plants, most plants in the temperate zone and in equatorial regions is about 20 °C, 5 to 10 °C, and 3 °C, respectively. This can be partly attributed to the adaptation of plants to the temperature fluctuations in their habitat.

Besides the direct influence of temperature in triggering, accelerating and impeding plant development, temperature can also influence plants indirectly by affecting the timing and duration of snowmelt and soil thaw, which leads to the alteration of the timing of plant activity, and thereby the start or extension of the growing season (Moulin et al., 1997; Myneni et al., 1997).

The relationship between temperature and plant phenology can be detected at different temporal and spatial scales. Fitter (2002) found that there is a significant relationship between mean monthly temperature and 83% of 385 species of interest. Many studies concluded that the onset of many phenophases such as the flowering dates and leaf color changing dates are closely associated with mean monthly temperatures (Park-Ono, et al., 1993; Shigehara et al., 1991; Schwartz, 2003). Although the timing of phenophases is influenced by other environmental conditions as well, temperature is the dominant factor during the dormancy season in winter and during the growing season in spring and summer at mid- and high-latitudes (Schwartz, 2003; Fitter et al., 1995; Sparks et al., 2000; Chmielewski, 2002). The influence of temperature is not as great in autumn (Schwartz, 2003). Similar relationship has been found between the annual mean temperature and onset of phenophases as well. For example, Zhang (1995) estimated that under a CO_2 doubling scenario, a temperature rise of 0.5 to 2.0 °C in annual mean temperature can cause the corresponding phenophases to advance 4 to 6 days in spring and summer, and to delay 4 to 6 days in autumn.

Since remote sensing has become a major means for phenological research, the relationship between vegetation indices computed from remotely sensed images and temperature has been addressed. Generally, temperature has been demonstrated to be strongly related to vegetation indices across different regions in the world (Sirikul, 2007; Nemani and Running, 1989). Cui et al. (2009) found that NDVI responds maximally to the variation of temperature with a lag of about 10 days in eastern China.

Temperature does not only affect time-related plant growth, but it also exerts its effects spatially. Long term average temperature is considered a determinant factor of vegetation distribution, and is widely used in a lot of global vegetative models (Schulze et al., 2005). The advancement or delay of phenophases due to the increase or decrease of temperature varies spatially. Therefore, the temperature change can also change the distribution pattern of onset of phenophases (Schwartz, 2003; Kai et al., 1993).

Besides the mean, maximum and minimum temperature, Growing Degree Days (GDD) is another common temperature measure used in phenological research (Neteler et al., 2011). GDD is a measure of heat accumulation above a plant-specific base temperature. GDD is based on the rationale that plant development will only occur when the temperature exceeds the base temperature (Hudson and Keatley 2009; Neteler et al., 2011), and plants grow in an accumulated manner closely related to the accumulated heat indicated by the GDD. GDD is calculated as:

$$\mathrm{GDD} = \frac{T_{max} + T_{min}}{2} - T_{base}$$

where T_{max} is the maximum temperature, T_{min} is the minimum temperature, and T_{base} is the base temperature; any temperature that is below T_{base} is set to T_{base} before calculation. Some studies use an upper temperature threshold as well because there is also

a heat limit beyond which plant development will stop or become much slower. The Accumulated Growing Degree Days (AGDD) accumulates the GDD from a consistent starting data. Some studies indicate that AGDD is linearly related to plant growth and development.

2.2.1.2 Precipitation

The influence of precipitation can be summarized as the influence of water on plant growth, because precipitation is a critical component of global water cycle and it is how the ecosystem gains water.

Water plays an indispensable role in plant growth because it is crucial for all physiological processes (Lambers et al., 2008). Although photosynthesis requires only a small amount of water to function normally, the resulted decreased cell volume and wilt from water deficiency can cause photosynthetic activity to reduce (Larcher, 2003; Lambers et al., 2008). However, when the water content further drops below the amount required for normal function of photosynthesis, the photosynthetic activity is immediately inhibited. Photorespiration is less sensitive to water deficiency; the rate can only be reduced when the cellular dehydration is severe (Larcher, 2003). Prolonged wilt kills plants (Lambers et al., 2008).

Water is also one of the major components of transpiration through which plants can transport nutrients in water throughout the plant organs from the soil and root, and help balance leaves' energy (e.g., transpiration cools leaves) (Lambers et al., 2008). The process of transpiration requires water. Therefore, if there is not enough water to support the transpiration, nutrient deficiency can result, and reduced cooling effect can cause the microclimate temperatures to rise over the lethal limit.

Therefore, water sometimes functions as a limiting factor. Water stress strongly restricts phenological development and plant production globally, especially under dry climates (Lambers et al., 2008; Schwartz, 2003). Lambers et al, 2008 concluded that water stress has the greatest limitation than other biotic and abiotic factors on crop yield.

The relationship between water and plant growth is complicated, so the relationship between precipitation and plant growth is even more complicated because the amount of water and how water is used is dependent on precipitation forms (rain or snow), soil conditions, vegetation types, phenophases, and evapotranspiration activities influenced by temperature and such.

The influence of precipitation on vegetation and timing of phenophases is thus not as intuitive as that of temperature. Junttila et al. (1983) found that increased precipitation can advance the flowering dates for early flowering species while delaying dates for later flowering species (especially those in plants with high basic/threshold air temperatures for flowering) (Schwartz, 2003). Knapp (1984) found that flowering intensity significantly decreased during drought while it increased during a wetter year generally for three grasses of interest (big blue stem, little bluestem, and switchgrass), although slight differences exist (Schwartz, 2003). Schulze (2005) considered precipitation a decisive factor at the time of budding and during the development of young plants. Zhang et al., (2005) found a threshold of cumulative rainfall beyond which, the onset of vegetation green-up is stimulated in arid and semiarid regions of Africa.

Actually, although temperature has much greater influence on plants and plant phenology during most of the time and at most locations, precipitation is of greater importance in arid and semiarid regions (Schwartz, 2003, Moulin et al., 1997). Moulin et

al. (1997) concluded that the onset of the growing season is more related to the cumulative temperature in temperate deciduous forests and to precipitation in savannas. Schwartz (2003) found that 75% of the species in the dry forest are affected by the precipitation seasonality, compared to only 17% in wet forests. However, Wielgolaski (2003) found that precipitation was still important for plant development even in regions with maritime climates, specifically, he found that germination was accelerated when the number of days with precipitation is higher for *Betula pubescens* in western Norway.

Similar to temperature, the responses of vegetation to precipitation also lagged, and generally with a longer length, because water during rainfall cannot be used by plants until it reaches the soil. An exception is between precipitation and ephemerals where the response is near-instantaneous (Davenport and Nicholson, 1993). Vegetation may also respond to an integrated multimonth or multiyear of rainfall (Davenport and Nicholson, 1993). Tyler (2001) conducted a 12-year phenological study in Swedish temperate deciduous forests, showing that precipitation of the previous year may be more important than temperature for flowering (Schwartz, 2003). Justice et al., 1986 and McMahon et al., 1982 found that the lag of response of monthly NDVI to monthly rainfall is 1 to 2 months, and monthly NDVI is strongly correlated with the cumulative rainfall of the previous 2 months (Chandrasekar et al., 2006). Ji and Peters, 2004 found that different lags exist between NDVI and precipitation in different seasons in grasslands and croplands. The lag is shorter in the early growing season and longer in the mid- to late-growing season generally. Cui et al. (2009) found in eastern China that NDVI maximally responds to the variation of precipitation at a lag of about 30 days, and the response is the most pronounced in autumn. They also concluded that the lag is different in different

seasons, and is longer in summer. Another conclusion of Cui et al. (2009) is that the lag also varies spatially, and it increases from north to south.

Variability in precipitation regimes at seasonal and longer time scales strongly influence ecosystem dynamics in arid and semiarid regions (Lotsch et al., 2003). Long-term average precipitation determines species and ecosystem distribution, and precipitation variation in the growing season has a strong influence on the inter- and intra-annual variation in plant structure and productivity (Lotsch et al., 2003). Tucker and Nicholson (1999) showed that NDVI correlates with annual integrated precipitation, and the interannual variability correlates with variability in precipitation patterns in regions where precipitation is limited.

The long term average precipitation, and the seasonal distribution of precipitation, strongly influences the spatial distribution of vegetation cover and vegetation density. Wet tropics that have abundant and evenly distributed precipitation during the growing season have rich vegetation; grasslands prevail where there are frequent and severe droughts in summer; further reduced precipitation leads to a semidesert, occupied by scattered shrubs; areas with extremely limited precipitation are deserts with no or little vegetation (Lambers et al., 2008). The distribution of two major forests in western Australia (karri and jarrah) are limited mainly by precipitation and soil conditions. Karri is associated with acidic soils, while jarrah has lateritic soils and lower precipitations. In regions where the annual precipitation further decreases, forests change to wandoo woodland. Werger (1983) also arrived at a similar conclusion in west Africa: mean annual rainfall is closely related to vegetation spatial distribution. In areas with 150 to 300mm mean annual rainfall, the major vegetation is ephemeral grasses; in areas with

300-500mm rainfall, grassy savannas prevail; in areas with over 500mm rainfall, woody vegetation appears.

2.2.1.3 Light

Light is heterogeneous both spatially and temporally, thus stationary and photo-autotrophic plants are strongly influenced by the light environment (Timmermans, 2010; Srivastava, 2002).

Light exerts significant influence on plants by participating in a series of basic physiological processes including photosynthesis, transpiration and metabolism (Sharma, 2005; Fitter and Hay, 2001), thereby influencing plant growth and development. Further, many phenological events are associated with light such as germination, seedling development and flowering in which cases light functions as a signal (Sharma, 2005; Srivastava, 2002; Timmermans, 2010). There are several families of photoreceptors that can monitor light across a wide range of bandwidths (Timmermans, 2010). These photoreceptors coordinate to mediate a series of developmental transitions (Timmermans, 2010).

Light affects the photosynthetic process both indirectly and directly. Firstly, light is indispensable for chlorophyll formation (Sharma, 2005); secondly, photosynthesis directly needs the participation of light as the energy source (Fitter, 2001). Only light within the band between 400 and 700 nm, or Photosynthetically Active Radiation (PAR), can influence the photosynthetic process (Crawley, 2009). Light reaches the plants in the form of photons. Pigments in chloroplast capture the photons, and convert them into chemical energy through the photosynthesis process (Pessarakli, 2002; Scott, 2008; Pfafflin, 2006). The energy is stored in the form of carbohydrates and other compounds,

which ultimately are used for plant growth and development (Pessarakli, 2002; Crawley, 2009). The photosynthetic responses to light considerably differ among species and among tissues and organs on the same individual plants (Crawley, 2009). Generally, the rate of photosynthesis increases with increasing PAR. However, for most plants, light is not a limiting factor and photosynthesis easily becomes saturated under usual solar radiant conditions (Fitter, 2001). Therefore, light in these situations is more important as a signal than as an energy source (Srivastava, 2002). In other words, the solar radiation has to reach certain critical values for specific phenophases to occur, and thus thresholds become more important in these cases.

Transpiration is also affected by light. Light can regulate the opening and closing of stomata; it can also exert its influence through the resulted elevation of temperature (Sharma, 2005).

Plants usually tend to adjust themselves to the current light conditions, and this process is called photoacclimation and photoadaptation. Under low light level, they reorient their leaves and chloroplasts to maximize light absorption; under high light level, they fold or drop leaves to avoid interception of excess light and resultant photodamage (Huang, 2006).

Many characteristics of light influence plant development, including the intensity, duration, and quality (or spectral composition).

Light intensity can be quantified by solar radiation, usually in the unit of watts per square meter. Solar radiation has an influence on plant growth and development, but not in a limiting way, because most plants can develop well in less than full sunlight (Pfafflin, 2006). Low solar radiation can lead to the situation that the energy gained from

photosynthesis is not enough for plants to maintain growth and development; when solar radiation increases, metabolism is stimulated to increase the sink capacity of photosynthesis, which can cause the increase of photosynthate production, and thereby increased rate in respiration and growth; high solar radiation can cause the chloroplast to absorb excess energy, and the energy overload may lead to severe oxidative stress, photodamage and even photoinhibition. (Schulze et al., 2005; Huang, 2006).

The responses of plants to similar solar radiation levels differ from species to species (Pugnaire, 1999; Pfafflin, 2006). For instance, under intermediate light levels, light-demanding species grow faster than shade tolerant species (Pugnaire, 1999). Shade tolerant species can still maintain their development under very low solar radiation, such as redwood seedlings, which normally grow in the shade of forest floor (Pfafflin, 2006). Decker (1944) found that photosynthesis in light demanding species such as loblolly pine increased with increasing solar radiation, and reached maximal rate at full sun light; shade tolerant species such as white oak and red oak reached a maximum in photosynthetic activity at only about one-third of full sunlight, and showed a little decrease when solar radiation further increase (Pfafflin, 2006). Photosynthetic responses can be different for the same plants depending on the habitat. For example, *Solidago virgaurea* (golden rod) can be either sun plants or shade plants in different habitats (Pfafflin, 2006).

The duration of light is the day length, or photoperiod. Photoperiod can influence plant growth dependent on the type of plants, and the interactions between photoperiod and plants are among the most complex (Thomas and Vince-Prue, 1996). The responses of plants to photoperiod, or in particular, the timing of light and darkness is known as

photoperiodism (Thomas and Vince-Prue, 1996). It is one of the major means that plants resort to to adjust themselves to seasonal changes in the environment and to avoid unfavorable environments (Thomas and Vince-Prue, 1996). For example, at high latitudes, the shortening photoperiod in autumn acts as a signal to induce dormancy, a response by which plants can survive the following low temperatures in winter; for some species in desert and semidesert areas, dormancy is induced by the increasing photoperiod so that plants can survive water stress in summer (Thomas and Vince-Prue, 1996). Therefore, photoperiodism is widely utilized in breeding and horticulture by artificially manipulating duration of light or darkness (Thomas and Vince-Prue, 1996). Hillman (1969) indicates that total solar radiant energy, as long as it is above a threshold, is relatively not as important as photoperiod.

The first to carry out an experiment demonstrating how photoperiod influences plant development is Henfrey (1852). He found that plant development was faster when the photoperiod is longer. Many other attempts also successfully demonstrated the influence of photoperiod on the acceleration of plant development and onset of some phenophases, especially flowering (Thomas and Vince-Prue, 1996). However, it was Tournois (1912, 1914) and Klebs (1913) who proposed that it is photoperiod, rather than the total quantity of light, that strongly influences plant development. Further, Garner and Allard (1920, 1923) pointed out that a longer photoperiod can actually either accelerate or slow down plant growth rate and some other responses, depending on the plant. They classified plants into the modern photoperiodic groups, and they also introduced the terms “photoperiod” and “photoperiodism.”

Short-Day Plants (SDP) follow a dark dominant mechanism, and they only grow or grow faster when the photoperiod is less than a critical value, i.e., the length of dark period is greater than a critical value; Long-Day Plants (LDP) follow a light dominant mechanism, and they only grow or grow better when the photoperiod exceeds a critical value; Day Neutral Plants (DNP) grow regardless of photoperiodic conditions; there is also a fourth condition in which dark dominant and light dominant mechanisms may both take effect, called Intermediate-Day Plants (IDP) which only grow or grow faster between two critical values of day length (Sharma, 2005; Thomas and Vince-Prue, 1996; Pfafflin, 2006).

Plants that are sensitive to photoperiod can be subdivided into qualitative types and quantitative types. Qualitative types require the photoperiod to be over or under a particular value to grow for LDP and SDP, respectively; quantitative types do not have a critical value of photoperiod to maintain plant growth, yet longer or shorter photoperiods can accelerate the growth rate for LDP and SDP (Thomas and Vince-Prue, 1996). The division between these two kinds is even more obscure, a type may turn into another when a certain environmental condition is satisfied (e.g., a plant can be qualitative at one temperature and quantitative at another) (Thomas and Vince-Prue, 1996). Therefore, Thomas and Vince-Prue (1996) suggests considering the two types as one continuous piecewise function instead with either no change or slight change on one side of critical value and obvious gradual change on the other.

Light quality or spectral composition is also very important to various physiological processes of plants. Solar radiation within the visible range is of most importance (Pfafflin, 2006). All visible light with wavelengths below 680 nm is effective

in chlorophyll synthesis (Sayre, 1928; Pfafflin, 2006). The blue-violet band is responsible for phototrophic responses (Pfafflin, 2006).

Variation in irradiance is one of the major characteristics of habitats dominated by different plants. For example, Blackman and Rutter (1950) found that the distribution of *Scilla nonscripta* (bluebell) in English forests is controlled primarily by solar radiation. Certain plant species need a minimal solar radiation to exist (Tansley, 2010). Local variations also matter, the influence of which is especially important in forest with upper canopy and understory vegetation, because light is the single most limiting factor there (Fitter, 2001; Pugnaire, 1999). Light quality also influences plant distributions, e.g., the increase of ultra-violet light at higher elevation may reduce the number of species (Waugh, 2000).

2.2.2 Environmental Factors

Environmental factors such as topographic and edaphic conditions are temporally stable as compared with environmental drivers, so only their spatial dynamics can influence plant growth and development. The direct influence is on the spatial distribution of vegetation, and indirectly, environmental factors affect environmental drivers and moderate how environmental drivers affect plant growth (Dale, 2000).

2.2.2.1 Topography

The effects of landscape, specifically elevation, slope and aspect, on vegetation are very pronounced in the natural environment.

Plant phenology and distribution are strongly affected by elevation (Campbell, 1974; Schuster et al., 1989) via its influence on abiotic factors. Air temperature on average decreases by 0.6 ℃ per 100 m of increase in elevation (an adiabatic lapse rate of

0.6℃ /100 m). The dry adiabatic lapse rate can reach a maximum of 1℃ /100 m for unsaturated air, and the wet adiabatic lapse rate is about 0.5℃ /100 m for saturated air (Jorgensen, 2009). Also, the lower partial pressure of CO_2 at higher elevation has a direct downward influence on photosynthesis activity, and results in a reduced carbon gain and vegetation growth (Jorgensen, 2009). Due to a thinner atmosphere at higher elevation, an outcome of increased solar radiation is expected (Jorgensen, 2009). Increased elevation is usually associated with increased precipitation due to an orographic effect (Black, 1996). Elevation also affects the state of precipitation (rain or snow), dependent on temperature. Snow is usually associated with higher elevation and rain with lower elevation, due to the lapse rate of temperature (Grayson and Blösch, 2001).

The influences of elevation on these abiotic factors cause a series of phenomena along elevation gradients: when elevation increases, generally there are decreases in the number of species, plant height, plant density, growth rate, and length of growing season (Waugh, 2000). Elevation is also a decisive factor of species distribution, with vegetation cover changing gradually from grassland, shrubland, and deciduous forest, to conifer forest, to barren systems and tundra as elevation increases.

Slope and aspect affects vegetation in a different way as compared to elevation and most other environmental drivers and factors. Their variations across space are more abrupt rather than gradual, thereby creating localized climates or even microclimates with different solar radiation, precipitation, drainage, and exposure to wind (Schwartz, 2003; Sharma, 2005). These microclimates further lead to the creation of microhabitats whose environmental and phenological characteristics can differ a lot even over a small area (Schwartz, 2003; Batanouny, 2001).

Slope can influence soil depth, acidity and drainage through its effect on the rate of water flow (Waugh, 2000; Sharma, 2005). Steeper slopes are usually associated with thinner, less waterlogged and less acidic soils because water flows down more rapidly and soil can be more easily washed down due to gravity (Waugh, 2000; Sharma, 2005). Therefore, in an extreme situation, only certain vegetation or even no vegetation can exist along a very steep slope with little water supply and severe soil erosion, even though there is abundant precipitation (Sharma, 2005). Slope also affects the solar radiation by changing the solar incidence angle, and the different heating causes different atmospheric temperature within the microhabitat (Sharma, 2005).

Aspect is the direction a slope faces. Aspect affects solar radiation, temperature and moisture. Generally, south-facing slopes are more favorable for plant growth than north-facing slopes in the northern hemisphere, because they can receive more solar radiation, and thus have higher temperature and lower humidity (Waugh, 2000). Also, it is common to see different vegetation or vegetation composition on different aspects. For example, the southwest facing Himalayas generally have more abundant rains than the other side due to the rainshadow effect, and thus have flourishing vegetation as compared to much poorer vegetation on the other side (Sharma, 2005).

2.2.2.2 Edaphic Conditions

Variations in soil texture, structure, nutrients, and other characteristics cause plant growth and distribution to vary (Waugh, 2000).

Soil texture is the degree of coarseness or fineness of the soil (Waugh, 2000). Soil can be composed of clay, silt, sand, and gravel. The composition and proportion of them determine the soil texture (Waugh, 2000). Variations in soil texture (i.e., variations in the

size and spacing of soil pores) affects soil water content, water flow and aeration (Waugh, 2000). For example, clay soils tend to hold more water and are less aerated as compared to sandy soils which have larger and more scatted soil pores. Soil texture also influences the ability of soil to hold nutrients. Clay soils can retain more nutrients and make them less easily leached due to the higher absorbability of clay particles compared to sand, thereby making clay soils more fertile. Despite all these disadvantages, it is easier for plant roots to penetrate coarser soils.

Soil nutrients are essential for plant growth, and are mainly released by rock weathering (Waugh, 2000). These nutrients are dissolved in water (soil solution) for plants to absorb. Soil nutrients in soil solution have two forms: cations (positively charged ions) and anions (negatively charged anions). Cations can be absorbed by plants through the cation exchange process, where cations move from the surface of soil particles or soil solution to plant roots (Waugh, 2000). Therefore, the measure of the ability of a soil to retain cations, called Cation Exchange Capacity (CEC) is a crucial measure. Low CEC means that the soil cannot hold enough nutrients for plant use, thus it is less fertile (Waugh, 2000). The CEC level is highly associated with soil texture and organic matter content. Generally, CEC is higher when soil texture is finer and there is more organic matter in the soil.

Anions are usually released from organic matter by the decomposition of microorganisms. Therefore, soil organic matter is another important measure as a major source of nutrients. Also, increase in the soil organic matter leads to the increased water-holding capacity.

Acidity is yet another attribute of soil that influences plant growth. Acidity is actually a measure of the concentration of the cations in the soil on the pH scale. A slightly acidic soil is the optimal condition for plant growth, and overly acidic soil is deleterious to plants, and makes organic matters more soluble and easily leached.

The depth of the water table also causes variations of plant growth. The plant rooting system is confined to the layers above the water table, because roots cannot grow into the ground water zone for lack of aeration. Thus the depth of the water table determines the depth of root system and soil as a nutrient reservoir (Batanouny, 2001).

Based on the review of the attributes above, it can be seen that edaphic condition is also one of the factors that controls vegetation distribution, because plants have different needs for nutrients and different tolerances of acidity and toxic elements (Foth and Ellis, 1997).

2.2.2.3 Latitude

Latitude, similar to elevation, has a gradient that affects abiotic factors. Annual insolation increases with decreasing latitude (Jorgensen, 2009). Latitude is also the only factor determining the spatial variation of photoperiod. Variations in solar radiation and photoperiod are direct controls on temperature. Therefore latitude is considered one of the major factors affecting temperature (Ahrens, 2007). As a consequence, the duration of the growing season varies from about 10 weeks at high latitude to the entire year in the tropics (Jorgensen, 2009).

Latitude influences vegetation distribution and diversity, as well as the environment-vegetation relationship via its influence on light and temperature. Vegetation diversity is higher at low to middle latitudes and lower at high latitudes (Van

Dyke, 2008). The distributions of many species or ecosystems are confined within certain parallels of latitudes. For example, alpine forests are likely to occur at higher altitudes on lower latitudes with strong maritime influences (Jorgensen, 2009). Mangroves are generally distributed between 25° N and 25° S latitude (Jorgensen, 2009). The responses of vegetation to environment vary at different latitudes. For example, different amounts of heat are needed for the same species to reach a certain phenophase at different latitudes (Schwartz, 2003).

2.2.3 Interactions Between and Combined Effects of Environmental Drivers and Factors

As stated above, plant growth is the result of overall influence of all environmental drivers and factors, and their interactions as well. Although it is a common and important analytical and explorative method to study relationships between vegetation and a single environmental driver or factor separately, it is important to bear in mind that the natural environment functions as a whole, and it is meaningless to look into one factor without considering the influence of another.

Firstly, environmental drivers and factors are interrelated and influence one another. Elevation strongly influences temperature, and the vegetation distribution along elevation gradients is largely due to temperature gradients. Generally, temperature decreases by 6.5°C per every 1000 m increase in elevation. Slope and aspect can also cause spatial variations of temperature through different solar radiant conditions. Consistent solar radiation can naturally result in the increase of temperature in the microclimate and in plant organs (Fitter, 2001). Temperature can affect the number of microorganisms in the soil. When temperature is higher, more microorganisms are

involved in the decomposition of organic matters, thus leading to more nutrients released for plant growth (Waugh, 2000). Extremely high precipitation sometimes leads to the decrease in vegetation abundance due to reduced soil nutrients by leaching (Schuur, 2003).

Secondly, variations of one environmental driver or factor can change how vegetation responds to another. How light influences plant growth is dependent on temperature. For example, the moderation of phytochrome, one of the predominant photoreceptors controlling germination, is actually variant at different temperatures (Timmermans, 2010). The saturating solar radiation increases as temperature increases. The solar radiation at which plants grow optimally differs at different elevations as well. Mooney and Billings (1961) found that *Oxyria digyna* reached maximal photosynthetic rate at higher solar radiation level at higher elevation (Pfafflin, 2006). Solar radiant energy needed for plants to maintain growth and development varied on different levels of moisture (Atkinson, 1904). While photoperiod is the most important control of flowering, dormancy is largely induced by variations in both photoperiod and temperature. Photoperiodism can operate when temperature is within an effective range, and when temperature is out of this range (either too high or too low), dormancy cannot be properly induced by photoperiod, and growth may resume and lethal damages may occur due to heat stress, cold stress and water stress (Thomas and Vince-Prue, 1996). Temperature usually has pronounced influence in the spring, in driving leafing, flowering and ripening, because of the extended photoperiod. Other environmental drivers, notably precipitation and light, can also change the response of vegetation to temperature. For example, under the circumstances of lower solar radiation, the temperature optimum

moves to a lower range, which implies that plants can function well during both the morning characterized by lower solar radiation and cooler temperature, and the midday characterized by higher solar radiation and temperature (Schulze et al., 2005).

Lastly, one driver becomes especially important when other drivers are not limiting factors. In the arid and semiarid tropics and subtropics, where temperature is always at its optimum for the vegetation there, precipitation becomes the main driver of plant growth (Schwartz, 2003).

2.3 Rangeland Phenology

Rangeland supports different kinds of vegetation types, such as shrublands, grasslands, steppes, deserts, and woodlands (Heady and Child, 1994). Rangeland phenology is also affected by the environmental drivers and factors enumerated above. However, rangeland vegetation responds more rapidly to the environmental variations as compared to other kinds of vegetation (Reed et al., 1994).

Most rangelands are characterized by low and variable annual rainfall, and are located in dry areas (Grice and Hodgkinson, 2002 and Tussie, 2004). Therefore, among all the environmental drivers and factors, precipitation regime has a much more significant influence on rangeland vegetation, especially on annual grasses and shortgrass prairie at drier locations (Reed et al., 1994). They usually respond to precipitation in a pulsed way, in which they are highly dependent on discrete rainfall events, and their phenology is a steady and rapid reflection of the rainfall events, in terms of productivity, density, and abundance (Rauzi and Dobrenz, 1970).

Another notable factor than constrains rangeland phenology is the livestock grazing. Livestock grazing has considerable impact on rangeland vegetation (Desalew et

al., 2010). The grazing-induced vegetation change is such a big disturbance, especially under high grazing intensity, that it cannot be overlooked when monitoring or predicting rangeland plant phenology. The actual response of rangeland vegetation to grazing is dependent on the types of livestock and the vegetation types and composition.

2.4 Phenological Models and Predictions

A model is a representation of an object or process being studied in the real world (Bender, 2000). Modeling is a process of simplifying and abstracting reality. It is an integration of knowledge already known about reality (Hodges, 1991). A process to be modeled can be divided into three parts – negligible elements, elements that affect the model but whose behavior the model is not designed to study (exogenous variables, parameters, input, or independent variables), and elements whose behavior is designed to be studied using the model (endogenous variables, output or dependent variables) (Bender, 2000).

Phenological models refer to algorithms or quantitative statements which can simulate the response of the major plant physiological processes to environmental variables (Hodges, 1991) in order to predict the phenophases of the plant. Phenological modeling is a method of integrating knowledge about biological and environmental processes into mathematical equations (Hodges, 1991). In a phenological model, the outputs or dependent variables are the phenophases or the vegetation status and the input or independent variables are all of the environmental variables that affect the phenophases.

The earliest effort in phenological modeling was made by Reaumur (1735). He suggests that the spatial and interannual difference in the dates of phenophase onset can

be explained by the degree-day sum. Reaumur (1735) concluded that plant development is proportional to the heat accumulation over time, rather than the temperature around the phenophase onset. This is still the most important assumption in phenomodeling (Schwartz, 2003). Since then, many phenomodels have used different forms of heat accumulation to model many kinds of phenophases, mostly budburst and flowering. The variable being modeled was usually the date of phenophase onset/offset during the early periods of the phenomodeling history, because phenological observations were mostly ground- and field-based, and it is easier to accurately observe the dates rather than the actual vegetation status.

Photoperiod was then added aiming to more accurately model or predict the phenophase onset. The models with photoperiod considered initially, took the form of a multiplication of the heat accumulation variable and a variable based on photoperiod. Photoperiod is still considered a very important independent variable in phenomodels nowadays. For example, Jolly et al., 2005 developed a Growing Season Index (GSI) as the product of three indices related to photoperiod, minimum temperature and vapor pressure deficit.

The advent of remote sensing technology induced great changes in phenomodeling. The phenomodels that utilizes remote sensing to predict phenophase onset focus more on the onsets and offsets of the growing season than on the actual budburst, flowering or leaf color changing dates. A simple way to predict the onset and/or offset of the growing season is to analyze the time series of VIs. There have been many methods to determine the onset/offset dates, including thresholds, maximum rate of change, or a certain percentage of the greatest VI increase.

Remote sensing also provides a way to monitor and record a continuum of the actual vegetation development. A lot of recent attention has been attached to the actual vegetation abundance instead of just the transition dates. One form of this kind of phenomodels is the autoregressive moving average, which uses current and precedent values of VI to predict the future VI values. However, the most popular and successful model is the multiple regression model (Ji and Peters, 2004; Olson et al., 1985). The multiple regression model is a statistical model that considers the underlying physical process by including all environmental drivers and factors, trying to simulate how the environment, as a whole, affects the vegetation abundance.

3 METHODOLOGY

The Upper Colorado River Basin is a highly geographically diverse region. This diversity creates great spatial variation in environmental drivers and factors. Temperature, precipitation and light are identified as three main environmental drivers having spatio-temporal influence on plant development. The influence is moderated by the spatial variation of environmental drivers including topography and edaphic conditions. As reviewed in Section 2.2, the environment-vegetation relationships are dependent on different ranges of environmental drivers and factors; therefore, the great spatial and temporal variations in the environmental drivers and factors lead to the spatial and temporal variations in the environment-vegetation relationships. Both temporal and spatial variations of the environment-vegetation relationship need to be accounted for to achieve the three objectives of the research. The spatial variation of the environment-vegetation relationship in the UCRB is firstly accounted for by phenoregion delineation, i.e., spatially differentiating subregions using environmental drivers and factors, and then by adding the time invariant environmental factors in the phenomodels. The temporal variation of the relationship is conducted by the identification of phenophases, i.e., greenup, maturity, senescence and dormancy and by including the time variant environmental drivers in the phenomodels.

The Multivariate Adaptive Regression Splines (MARS) technique is used as the modeling approach serving both the prediction and interpretation purposes. The onset and

offset of phenological phases can occur within a time window as short as only a few days. Use of data with longer intervals will firstly increase the uncertainty and makes it harder to get accurate predictions necessary for identifying the phenophase onset/offset, and secondly, will mask the accurate onset/offset date since all predictions are made for the next interval. Data with shorter intervals are usually only available at much coarser spatial resolution (Cardot et al., 2008). Therefore, 1-km 7-day eMODIS and Ecocast data are chosen as a trade-off between the two kinds of resolution that satisfies the requirements of rangeland managers. The phenomodels use a large number of samples extracted from remote sensing images, rather than only data derived from sites sporadically distributed in the study area. The NDVI is used as the dependent variable of the phenomodels to represent vegetation abundance, because it has been successfully used in many phenological studies. The saturation issue, as a major limitation of the NDVI, becomes less of a concern in this research in the semiarid and arid UCRB. Different independent variables are selected to be used in different phenoregions based on their influence on the 7-day vegetation dynamics. Models are evaluated and adjusted based on their performance. This modeling process ensures that the interpretation made from these models is representative of the vegetative response to the environment. Two ways of validation are used to assess the model's representation of the reality. Cross validation ensures the model is applicable to other independent regions and years. Field validation ensures the model is applied to the physical environment in a realistic manner. Finally, a DSS is developed to incorporate the models and other information for public use. The DSS is to provide assistance to the land managers and other users. Components of the methodology are presented in Table 3.1.

Table 3.1 Workflow of methodology

Components	Approaches, steps and summaries	Techniques
3.1 Phenoregion Delineation	Select phenological forcing variables; spatially partition the UCRB; evaluate the series of phenoregion maps and select the optimal one used in the following analysis and modeling.	Principal Component Analysis (PCA); *k*-means++ clustering; MapCurves goodness-of-fit score
3.2 Data Sources and Data Processing	Introduce eMODIS and Ecocast data and preprocessing; scrutinize data and assess data quality; data sampling; outlier exclusion; NDVI time series reconstruction	Reprojection and subset; Stratified systematic sampling; random sampling; Savitzky-Golay smoothing filter; iterated Savitzky-Golay fitting
3.3 Time Series Analysis	Mean phenoregional reconstructed NDVI time series generation; identification of phenological phases	Time series smoothing; time series decomposition; NDVI ratio; thresholding
3.4 Premodeling Preparation and Analysis	Calculate candidate variables; pairwise correlation analysis; PCA analysis; lag structure analysis	Image processing and generation through IDL; Pearson's correlation coefficient; PCA
3.5 Model Development	Model development, diagnostics, and adjustments; relationship analysis between NDVI and environmental drivers as moderated by environmental factors	Multivariate Adaptive Regression Splines (MARS)
3.6 Model Validations	Validate phenomodels using both cross validation and field validation	Field site selection; field spectra measurement; spectra processing; bad quality spectrum handling; cross validation; Root Mean Square Error (RMSE); Coefficient of Variation (CV)
3.7 Phenological Decision Support System	Phenological DSS integrating phenomodels and other useful information	Google maps API; KML; C#.net; Web-based GIS

3.1 Phenoregion Delineation

The delineation of subregions based on their climatic, ecologic and geographic characteristics has been increasingly used for planning, policy making, natural resource conservation and management by government agencies and conservation groups (Thompson et al., 2004). Partitioning of regions into functional subregions is dependent on the purpose of the application. Therefore, subregions can take a variety of forms based on the classification logic.

Ecoregions are one type of subregion delineation. The Küchler, Bailey and Omernik systems are three well-known and widely used ecoregion classification systems generated using different data and classification methods (McMahon et al., 2001; Thompson et al., 2004). The concept of Potential Natural Vegetation (PNV) was introduced by Tüxen (1956) as the vegetation that would exist today if human impacts were removed. The Küchler system (Küchler, 1964) is a potential natural vegetation map of the conterminous United States. The Bailey system (Bailey, 1983) adopted maps of climate, topography and vegetation to generate ecoregion maps at nine levels of division while each level is based primarily on one particular map (Omernik, 1987). The Omernik ecoregion system (Omernik, 1987) is based on a combination of four maps: land use, land surface form, potential natural vegetation and soils. These well-known and widely-used ecoregions are delineated by qualitatively analyzing the homogeneity and generality of each adopted map based on the knowledge and experiences of experts.

Extending on the ecoregion concept, the delineation of subregions has become more function-specific. Agroecoregions (a.k.a. agroecozones, crop growth zones or soil productivity zones) are generated by delineating subregions of similar expected crop

performance. They are used for crop suitability analysis and agricultural policy making (Williams et al., 2008). The soil map units used in the Natural Resources Conservation Service, US Department of Agricultural (NRCS) soil survey are mapped by differentiating the properties of natural bodies of soils and serve as the basic map unit for the widely used STATSGO (State Soil Geographic) and SSURGO (Soil Survey Geographic) databases (NRCS, 2011).

Recently, natural area subregion delineation has focused on phenological processes to yield functional phenoregions. The term "phenoregion" was first defined by White et al. (2005) as phenologically and climatically self-similar clusters. The phenoregion system derived by White et al. (2005) served as a global framework for monitoring phenological responses to climate change.

The effective delineation of phenoregions, i.e., the effective spatial partitioning of the vegetation-environment relationship, is a prerequisite to all analyses and phenomodeling. Each phenoregion is an entity in which all pixels in that phenoregion are considered to share similar phenological cycles and vegetation-to-environment responses. Phenoregions are the basic areal units of all analyses. Therefore, phenoregion delineation is a very important first step in this dissertation and the foundation of its methodology.

The geographic diversity of the UCRB necessitates the delineation of phenology based subregions. This dissertation develops 1-km pixel based UCRB-specific phenoregions, which are expected to capture homogeneous environment-vegetation relationships in respective phenoregions by differentiating environmental drivers and factors. Multivariate clustering generates clusters as quantitative subregions based on multiple variables dependent on the function of clusters (Hargrove and Hoffman, 2004;

White et al., 2005). Multivariate clustering has been demonstrated to be effective for subregion delineation at different spatial scales. In this research, normalized principal component analysis (PCA) combined with improved *k*-means clustering (*k*-means++ clustering) is used to generate phenoregion maps of the UCRB (Figure 3.1). The number of clusters is varied, and the results are compared using a set of evaluative criteria to determine the optimal classification from this series of phenoregion maps. The differentiation of environmental drivers and factors can lead to the successful delineation of phenoregions, because:

1. Differentiation of environmental drivers and factors naturally lead to similar ranges of these drivers and factors, i.e., similar phenological forcing within each phenoregion.
2. A literature review indicates that long-term average and seasonal variations of environmental drivers as well as environmental factors determine the vegetation distribution, so a phenoregion is expected to have homogeneous vegetation distribution.
3. As summarized in the section of "Interactions between and Combined Effects of Environmental Drivers and Factors" in Section 2, variations of one environmental driver or factor can change how vegetation responds to another. Phenoregions limit this kind of variations, and can naturally lead to more homogeneous environment-vegetation relationship.

A phenoregion, thereby, is expected to have similar vegetation responses in terms of either one of the environmental drivers and factors and the environment as a whole. The phenoregions in the optimal phenoregion map assured that a unique phenomodel

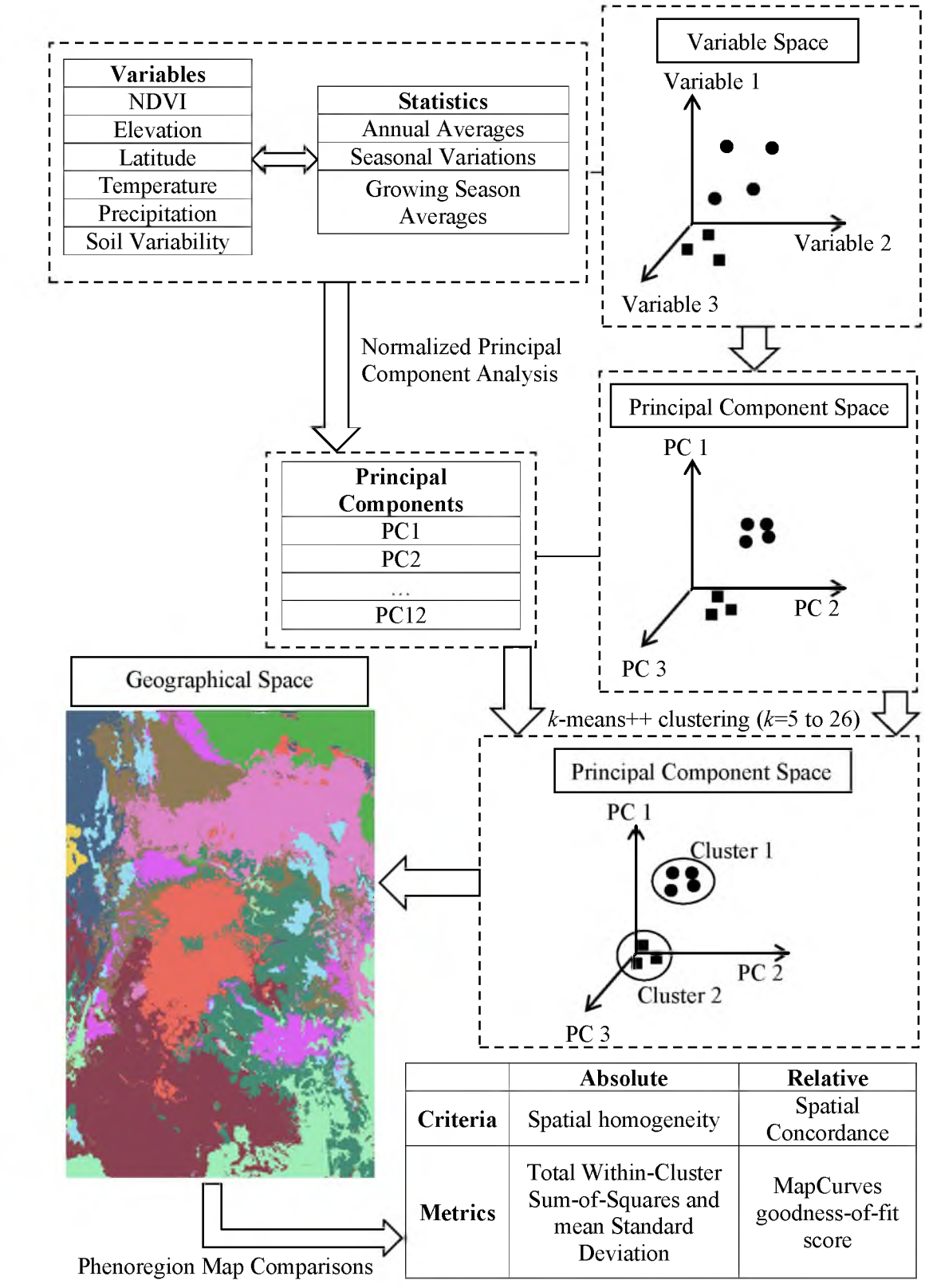


Figure 3.1 Diagram of phenoregion delineation.

with a unique set of independent variables can be developed for each phenoregion, and the common environment-relationship can be explored within each phenoregion by interpreting the respective phenomodel. The diagram of phenoregion delineation is shown in Figure 3.1.

3.1.1 Data Sources and Variables

It has been pointed out that spatial variations of all environmental drivers and factors significantly influence vegetation distribution and spatial variation of environment-vegetation relationship. Theoretically, long-term average and seasonal variations of all environmental drivers and factors should be included in the phenoregion delineation; practically, there are several exceptions and modifications. The variables used in delineation for this research are specified in Table 3.2.

Light conditions, specifically solar radiation and photoperiod, are not included as variables in phenoregion delineation. This is because the spatial variation of solar radiation and photoperiod is largely due to latitude and elevation. As discussed before, the influence of slope and aspect act more like a microhabitat, therefore it is more meaningful to consider the two variables in the phenomodels instead of in the phenoregion delineation. The phenomodels can represent exactly how the variation of landscape can adjust the climate and create microhabitats, by including pixel specific slope and aspect values, and allowing for them to interact with other independent variables.

The annual averages, standard deviation of monthly data and averaged data during growing season are used in phenoregion delineation for the differentiation of phenological forcing and for phenological patterns.

Table 3.2 Summary of variables used in the UCRB phenoregion delineation

Variables	Descriptions	Data Sources	Original Spatial Resolution	Accuracy
Latitude			1km	
Elevation	Mean elevation above sea level	DTED Level 0 by NGA	30 arc seconds	Horizontal accuracy: <60 m Vertical accuracy: <46 m
Mean maximum temperature	Annual maximum temperature averaged for 1971-2000	PRISM dataset by PRISM Climate Group at Oregon State University	0.00833 decimal degrees	130 m
Mean minimum temperature	Annual minimum temperature averaged for 1971-2000			
Mean maximum temperature during growing season	Maximum temperature between May and Oct., averaged for 1971-2000			
Mean minimum temperature during growing season	Minimum temperature between May and Oct., averaged for 1971-2000			
Standard deviation of monthly temperature	Intra-annual variability of temperature			
Mean precipitation	Total annual precipitation averaged for 1971-2000			
Mean precipitation during growing season	Total precipitation between May and Oct., averaged for 1971-2000			
Standard deviation of monthly precipitation	Intra-annual variability of precipitation			
Soil Variability Index	The variable indicating soil variability and was derived by applying PCA on 10 soil characteristics	STATSGO soil characteristics for the conterminous United States by USGS	1 km	N/A
Mean NDVI	Mean annual NDVI averaged for 1990-2005	AVHRR-NDVI by USGS	1 km	<1 km Root Mean Square Error

Table 3.2 lists 12 variables included in phenoregion delineation related to temperature, precipitation, elevation, soil fertility and vegetation. Plant phenology and species distribution patterns have been demonstrated by multiple studies to be strongly affected by elevation (Campbell, 1974; Schuster et al., 1989). Distribution of different species can easily be observed across elevation gradients. The UCRB has an elevation range of about 3000m, making elevation an especially important input variable in this research. The vegetation cover in the UCRB changes gradually from barren systems above 3500m, to conifer- and deciduous-dominated forest at around 2500-3000m, to sagebrush dominated shrubland below 2000m. The elevation variable is derived from the Digital Terrain Elevation Data (DTED) level 0 at 30 arc second (~1km) resolution compiled by the National Geospatial-Intelligence Agency (NGA, 1996) in 2001.

Temperature has long been observed to directly influence phenological phases. A large number of papers have scrutinized the effects of temperature on the phenological timings of plants (Badeck et al., 2004; Fitter et al., 1995; Sparks and Carey, 1995; Sparks et al., 2000). In general, higher temperature accelerates plant development leading to earlier onset of phenological events. Five temperature variables calculated from the PRISM (Parameter-elevation Regressions on Independent Slopes Model, Daly et al., 1994) dataset were included as input variables: mean annual maximum temperature, mean annual minimum temperature and standard deviation of monthly temperature as well as mean maximum and minimum temperature during the growing season, defined as from May to October based on the range of first and last freeze/frost occurrence dates at different stations spread over the UCRB (Koss et al., 1988). PRISM data have a spatial resolution of 0.00833 decimal degrees (~925m) and have a monthly temporal resolution

covering 1971 to 2000 (PRISM Climate Group, 2010). The mean annual maximum and minimum temperature represent the general temperature range. Standard deviation of monthly temperature, and mean maximum and minimum temperature during the growing season were adopted to account for the intra-annual variation of temperature.

Precipitation is another major factor having great effects on vegetation. It affects the timings of different phenophases and accounts for a significant amount of the phenological variation, especially in moisture-limited regions like the UCRB (Reed et al., 1994; Peñuelas et al., 2004). Three variables – annual mean precipitation, standard deviation of monthly precipitation and mean precipitation during the growing season were included. These three precipitation variables were also extracted from the PRISM dataset (PRISM Climate Group, 2010).

Although not as significant as precipitation, soil fertility has close relationship with species distribution pattern (Swaine, 1996). Soil fertility can greatly influence vegetation abundance and species richness (Gentry and Emmons, 1987; Swaine, 1996). Many soil attributes are directly correlated with fertility – such as pH value, organic matter, and cation exchange capacity (Troeh and Thompson, 2005). A principal component (PC) of several soil attributes is used to serve as an index of soil fertility to preserve maximum variability in soil attributes and allow better discrimination of phenoregions. The SSURGO and STATSGO databases are considered to be reliable data sources to derive soil fertility by providing a series of soil attributes directly related to soil fertility. However, the NRCS soil survey project to populate the SSURGO and STATSGO database is still underway in the western US, so this data source is currently unavailable. Instead, the USGS compiled 1 km data set of STATSGO soil characteristics

for the conterminous United States (USGS, 1997) was used, with a full coverage of the study area and limited soil attributes less directly related to soil fertility, yet still greatly influencing vegetation growth. This data set contains 10 soil parameters including the high and low values of the range of organic matter, permeability, available water capacity, bulk density and depth to seasonally high water table. PCA was applied to the 10 soil parameters. The first PC accounted for 99.6% of the total variance, indicating high consistency in the variance of the 10 soil parameters. The first PC was named the soil variability index (SVI) was used as another input variable besides the variables of elevation, temperature and precipitation.

Normalized Difference Vegetation Index (NDVI) is the most commonly used VI for monitoring vegetation phenology. The mean annual NDVI is provided by the 1 km dataset from the USGS by averaging the AVHRR-NDVI from 1990 to 2005, to signify the average vegetation growth and vigor.

Latitude is also demonstrated to have influence on the environment-vegetation relationship, even for the same species.

Twelve variables (Table 3.2) were thus selected and extracted from corresponding data sources. They were all resampled to a common 1 km resolution and pixel alignment, and then subset to the UCRB.

Phenoregions customized by this set of variables captures particular phenological forcing patterns that may be missing in other types of subregion classification (Hargrove and Hoffman, 2004). For example, by considering the annual mean precipitation and the mean precipitation during growing season, delineated phenoregions can provide additional information in discriminating places with desert monsoons and with alpine

climates. Also, the areas with a higher mean temperature in the growing season resulting in an earlier onset of greenup can be distinguished using these variables. More general purpose subregions delineated without these variables might fail to capture these important differences. In this sense, the set of variables can decompose the geographic diversity of topography, climatic conditions and vegetation, and finely differentiate the spatial variation in phenological forcing among different geographic locations in the UCRB.

3.1.2 Principal Component Analysis and *k*-means++ Clustering

Normalized Principal Component Analysis plus iterative *k*-means clustering has been demonstrated to be an effective approach for the delineation of subregions in previous studies (Hargrove and Hoffman, 2004; White et al., 2005). PCA plus *k*-means clustering, as a quantitative method, does not rely on geographic knowledge or familiarity with the data, thus making the delineation of phenoregions more objective. Although this approach is computationally intensive compared with other quantitative methods, its essence of hierarchical nonnestable clustering can lead to independent phenoregion maps with different numbers of clusters, providing a better opportunity to develop an improved classification of phenoregions (Hargrove and Hoffman, 2004).

However, due to the local optima problem associated with ordinary *k*-means algorithm, this research uses *k*-means++ (Arthur and Vassilvitskii, 2007), rather than ordinary *k*-means algorithm, to acquire optimal clustering. The *k*-means++ algorithm is an augmentation of ordinary *k*-means by replacing the random seeding with a careful seeding process. *k*-means++ clustering retains all the advantages of ordinary *k*-means and solves the local optima problem. The combination of PCA and *k*-means++ clustering is

thus used in this research to delineate phenoregions by generating clusters based on all the PCs of the 12 variables in Table 3.2.

Principal component analysis is an essential algorithm before *k*-means clustering can be applied due to potentially strong correlations among input variables. For example, precipitation is correlated with many soil attributes affecting soil fertility; soil fertility and vegetation abundance indicated by NDVI are closely related under higher precipitation (Swaine, 1996); and total annual precipitation and its variability are influenced by the elevation (Prins and Loth, 1988). PCA can effectively reduce the strong correlations between variables by converting the original set of variables into several PCs that are orthogonal in the principal component space (Hargrove and Hoffman, 2004). All of the variables were normalized because PCs are sensitive to scaling. Normalized variables have a mean of 0 and variance of 1. PCA was applied to normalized variables.

The *k*-means++ algorithm was used to cluster pixels that are close in the principal component space formed by all 12 PCs (i.e., similar values in elevation, temperature, precipitation, soil fertility and vegetation history). Ordinary *k*-means clustering has an intrinsic flaw associated with the random seeding: the performance of *k*-means clustering is dependent on the initial selection of cluster centroids. Many researchers have been aware of this problem, that the *k*-means algorithm may terminate at a local optimum instead of a global optimum depending on the initial centroids (Steinley, 2003). *k*-means++ clustering (Arthur and Vassilvitskii, 2007) addresses this problem by introducing a careful seeding process in lieu of random seeding in the ordinary *k*-means algorithm, to ensure the initial centroids are as far away from each other as possible in the multivariate data space. The *k*-means++ algorithm is as follows:

1. Select at random the first centroid from all data points.
2. Calculate a probability statistic using the following equation for each data point,

$$P = \frac{D(i)^2}{\sum_i D(i)^2}$$

where $D(i)$ is the shortest distance (in the principal component space) from a data point *i* to its closet centroid that has already been selected.

3. Select the data point with the largest probability (*P*) as the next centroid.
4. Repeat steps 2 and 3 until all *k* centroids are selected.
5. Proceed with this set of initial centroids as in the ordinary *k*-means algorithm.

Improved seeding can help ensure maximum dissimilarity between phenological forcing clusters and maximum homogeneity within each cluster while avoiding the selection of outliers (the algorithm is insensitive to outliers in seed selection). Adopting the *k*-means++ seeding process can effectively reduce the uncertainty and avoid the sometimes poor clustering that arbitrarily results from the ordinary *k*-means algorithm. In summary, the *k*-means++ clustering improves both speed and accuracy compared to the ordinary *k*-means (Arthur and Vassilvitskii, 2007).

k, the number of clusters, is a priori parameter for each execution of *k*-means++ clustering. An inappropriate selection of *k* could result in a poor classification of phenoregions. Therefore, the clustering was tested with *k* clusters with *k* ranging from 5 to 26. Two types of comparisons were then used to select a phenoregion map with higher homogeneity and spatial concordance (a measure of spatial coincidence and spatial overlap, Hargrove et al., 2006) with other phenoregion maps with different numbers of clusters.

3.1.3 Comparisons of Phenoregion Maps and Selection of the Optimal Phenoregion Map

The optimal phenoregion map should have the following characteristics:

1. The map should be as homogeneous within each phenoregion as possible, i.e., pixels within the same phenoregion should have as low variability or dispersion as possible in terms of latitude, elevation, temperature, precipitation, soil fertility and vegetation history.
2. Since the phenoregion maps with different numbers of clusters are all generated by the same process, maps with higher spatial concordance with other phenoregion maps may indicate consistently more stable phenoregions.

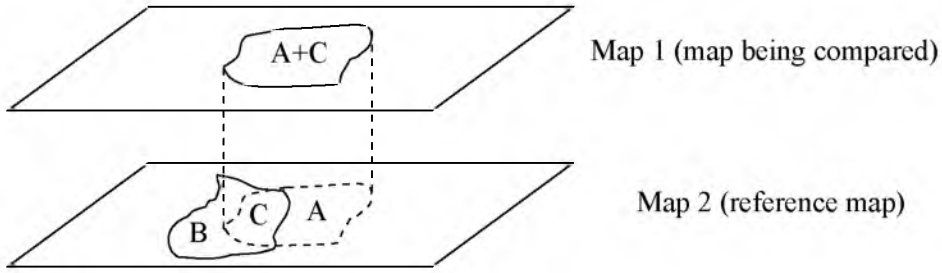
Two methods were adopted to compare phenoregion maps based on these two criteria. The first comparison method calculated the mean standard deviation of the Euclidean distance in principal component space formed by PCs for different phenoregion maps, referred to as "absolute comparison." The second method quantified the spatial concordance between pairs of phenoregion maps, referred to as "relative comparison."

The absolute comparison uses two metrics to quantify the spatial homogeneity of different phenoregion maps. For both, it is calculated for each pixel the Euclidean distance from itself to the centroid of the cluster it belongs to in the principal component space. This indicates the similarity between a pixel and the mean of the phenoregion (the final centroid after running *k*-means++) in terms of phenological forcing variables. The two metrics are the Total Within-Cluster-Sum-of-Squares (TWCSS) and the Mean Standard Deviation (MSD), calculated as:

$$\text{TWCSS} = \sum_{i} D_i^2 \qquad \text{MSD} = \text{E}\left(\sqrt{\frac{\sum_{i=1}^{N_p} D_i^2}{N_p}}\right)$$

where D_i is the Euclidean distance between the pixel i and the centroid in the principal component space, N_p is the number of pixels in phenoregion p. Both metrics indicate from different perspectives the general degree of homogeneity of a specific phenoregion map. The difference is that the TWCSS is the simple summation of the Euclidean distance over all pixels, and the phenoregions with more pixels naturally are attached with more importance / weight; the MSD, on the contrary, regards the homogeneity (standard deviation of Euclidean distance) of each phenoregion as equally important.

The relative comparison uses Mapcurves Goodness-Of-Fit (GOF) scores (Hargrove et al., 2006) to quantify the degree of spatial concordance between two maps. Each cluster on the map has a GOF score calculated as shown in Figure 3.2, where GOF is the goodness-of-fit score of this cluster, C is the amount of overlapping region, $B+C$ is the total area of the intersected cluster on the reference map and $A+C$ is the total area of the intersected cluster on the map being compared. A Mapcurve, which is a form of cumulative frequency distribution, was plotted for each comparison direction of phenoregion maps (e.g., the map with 5 clusters compared to the map with 6 clusters as a reference). This plot contains an x-axis indicating the GOF score and y-axis indicating the percentage of clusters with a GOF score exceeding the correspondent GOF threshold (Hargrove et al., 2006). Two GOF scores were derived by calculating areas underneath Mapcurves of both comparison directions. The higher score indicates favorable direction of comparison and was selected as the GOF score between these two maps. The Mapcurves GOF scores range from 0 to 1 with the higher value indicating better fit.



$$GOF = \sum\left(\frac{C}{B+C} \times \frac{C}{A+C}\right)$$

Figure 3.2 The algorithm used to calculate GOF score of cluster $A+C$ on Map 1 (modified from Hargrove et al., 2006).

The average Mapcurves GOF score of a phenoregion map was calculated as the sum of scores between this map and each of 5- to 26-phenoregion maps divided by 22 (the number of phenoregion maps). This research used the average Mapcurves GOF score as a measure of the average spatial concordance of a phenoregion map with the whole series of maps with different numbers of clusters.

Absolute and relative comparisons are conducted for the selection of the final phenoregion map used as the basis of all following analyses and modeling. In addition, the number of phenoregions is also an important factor when making the decision because it influences the time and labor consumed for analysis and field work.

3.2 Data Sources and Data Processing

3.2.1 eMODIS Data

MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites of NASA. Terra MODIS and

Aqua MODIS view the entire earth's surface every 1 to 2 days, acquiring data in 36 spectral bands (NASA, 2009). There are many standard MODIS data products for calibration, atmosphere, land, cryosphere and ocean.

The eMODIS product (USGS-EOS MODIS) is generated at the US Geological Survey's (USGS) Earth Resources Observation and Science (EROS) Center. The eMODIS products provide the 7-day interval data at the spatial resolutions of 250 m, 500 m and 1 km, including NDVI and surface reflectance bands over the Continental US. The eMODIS products offer higher temporal frequency (7-day) and faster production output (<24 hours for expedited data and <30 days for historical data) than standard MODIS products (USGS, 2008). While having a high agreement with the standard MODIS NDVI products as indicated by the pixel-by-pixel comparison, the eMODIS NDVI products have much less geometric distortion, especially in high latitude and extreme-longitude regions (Ji et al., 2010). Therefore, eMODIS products are more suitable as the data source from which NDVI can be extracted.

The 1-km eMODIS NDVI product was used in this research. All NDVI images and their associated quality files are reprojected to the NAD83 / UTM zone 12N and subset to the bounding box of the UCRB. The quality files are recoded: the pixels marked as good quality and snow are still kept the same, and the pixels marked as other values including cloudy, bad band quality, negative reflectance and fill values are marked as bad quality in the recoded quality files.

3.2.2 Ecocast Data

Ecocast data is the climate data product produced by the NASA AMES Ecological Forecasting Lab using the Terrestrial Observation and Prediction System

(TOPS) (NASA Ames Ecological Forecasting Lab, 2009). The Ecocast gridded climatic data includes the maximum and minimum temperature, precipitation and solar radiation. The Ecocast data at 1-km over the study area of the UCRB used in this project were generated by the NASA AMES in 2010 at 7-day intervals correspondent with those in the eMODIS data. Similarly, all Ecocast images are reprojected to the NAD83 / UTM zone 12N and subset to the bounding box of the UCRB.

3.2.3 Data Overview, Evaluation and Quality Assessment

A total number of 541 7-day intervals from 2000 to 2010 are used in this dissertation. The first 7-day interval is from the 49th day to the 55th day of 2000 (2000049 - 2000055), and the last 7-day interval is from the 173rd day to 179th day of 2010 (2010173 - 2010179).

However, there exists irregularity in the 541 7-day intervals. Different compositing rationales were applied in different years in the eMODIS data production to match the intervals of the AVHRR product suites: Friday to Thursday were composited in 2000 to 2002, Wednesday to Tuesday were composited in 2003 to 2006, and Tuesday to Monday were composited in 2007 to present. This resulted in the two pairs of neighboring 7-day intervals with overlapping dates. The first one is 2002361 - 2003002 and 2003001 – 2003007 with 2 overlapping days, and the second one is 2006354 - 2006360 and 2006360 to 2007001 with 1 overlapping day. Since the Ecocast data are generated at the same 7-day intervals as eMODIS data, this overlapping interval problem exists in both datasets. Because the number of overlapping days are only limited to 1 and 2 days, they are also considered to be approximately 7 days apart as with other neighboring intervals, for simplicity of analyses. Another issue of missing intervals is

caused by the data anomalies in correspondent intervals. The missing intervals in the eMODIS data set include 2000224 - 2000230, 2001166 - 2001172, 2001173 - 2001179, 2002081 - 2002087, 2003351 - 2003357, and 2004210 – 2004216. The missing interval issue for the Ecocast data only exists in precipitation images: precipitation data during the period of 2007296 - 2007302 are missing over the whole study area.

Aside from the interval irregularity, the data quality in both datasets, especially in the eMODIS dataset is also a problem to be paid attention to.

The eMODIS dataset is pure satellite data, and thereby its quality is easily affected by sensor failure, cloud, and a lot of other possible interferences. Therefore, although the enhanced Maximum Value Compositing (eMVC) used in eMODIS production helps to reduce cloud contamination and to improve the data, the overall eMODIS data quality is still not ideal. There are only 11 out of all 541 images that have 100% good quality. There are also a very small number (also 11) of NDVI images that have more than half of bad quality pixels. Most of the images have a percentage of bad quality pixels below 2%. Pixels identified as snow are also a problem worth attention. Although snow covered pixels can also be considered to have good quality and to represent the actual land surface condition, they are not of research interests in this dissertation. Therefore, pixels with snow are coded as a separate category besides good quality and bad quality.

The quality of the Ecocast data is much more consistent than that of the eMODIS data, because it is fused from multiple data sources such as ground observations, satellite data and ecological modeling. All images of minimum temperature, maximum temperature, and solar radiation are 100% good quality. Precipitation images are the only

type in the Ecocast data that have known quality problems. There are 111 out of 541 precipitation images that have data quality issues and only 5 have more than half bad quality pixels.

Therefore, the data quality issues should be borne in mind and taken care of when doing following analyses, in order to avoid artifacts in model outputs.

The spatial and temporal sampling and the usage of the sampled values are shown in Figure 3.3.

3.2.4 Spatial and Temporal Sampling

Stratified systematic sampling plus random sampling is adopted to extract pixels spatially for modeling and validation. Stratified systematic sampling can effectively reduce spatial autocorrelation, ensure the interdependency of pixel sample one and two, and avoid the infeasible, extremely time-consuming and sometimes destructive exhaustive sampling. About one-twentieth (14693 pixels) of the total number of pixels in the UCRB (293889 pixels) are determined to be extracted. The sample size allocated to each phenoregion is calculated using Neyman allocation, based on the total number of pixels in and the standard deviation of that phenoregion:

$$n_p = n * \frac{N_p * SD_p}{\sum N_i * SD_i}$$

where n_p is the sample size of phenoregion p, n is the total sample size, N_p is the number of pixels of phenoregion p, and SD_p is the standard deviation (heterogeneity) of phenoregion p. To sample n_p pixels in phenoregion p: firstly, systematic sampling is applied in phenoregion p to extract the sample with the size slightly more than n_p; then, n_p pixels were randomly extracted from the systematically sampled pixels. The n_p pixels

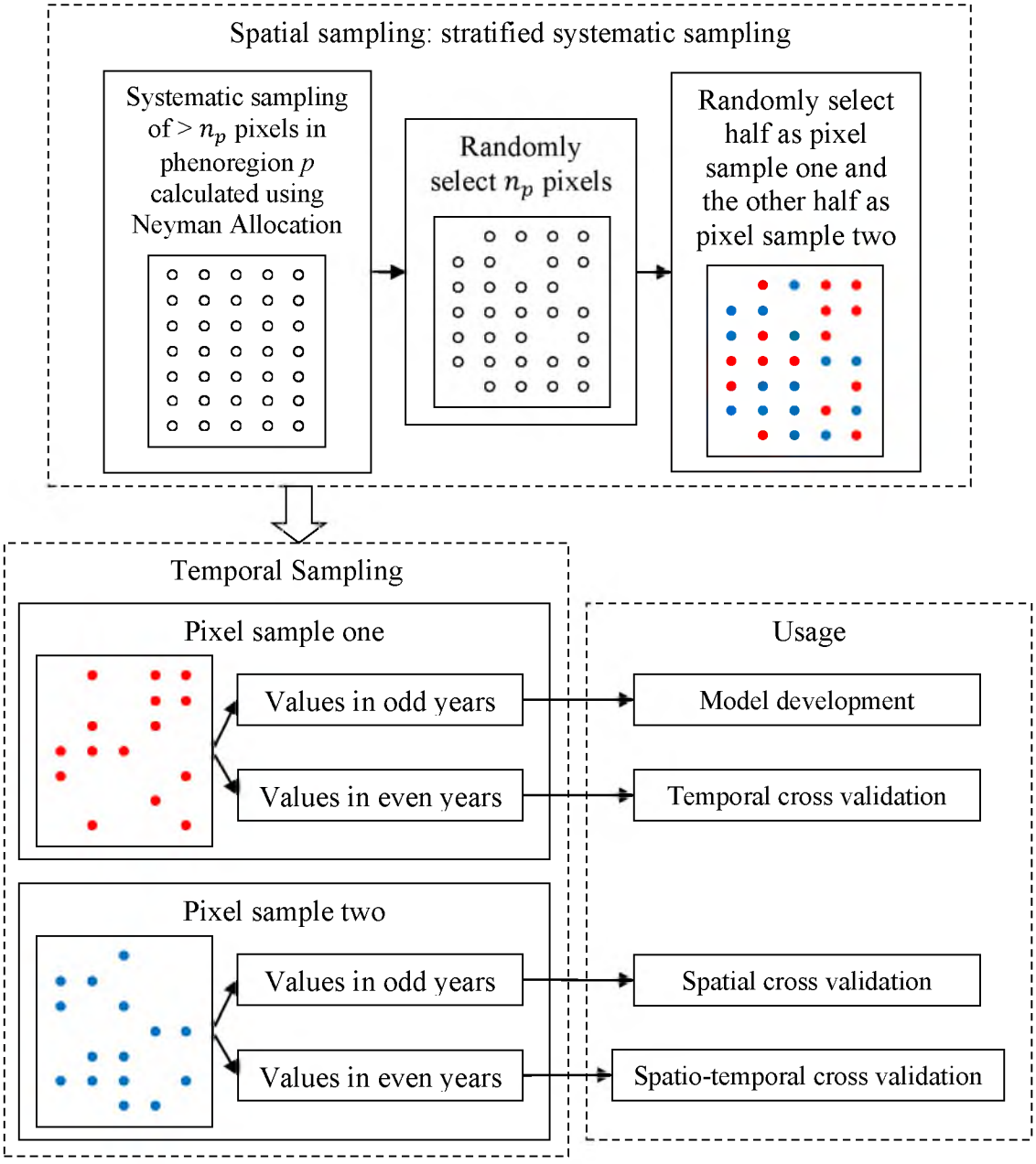


Figure 3.3 Spatial and temporal sampling and the usage of sampled values.

were randomly divided into two sets: pixel sample one and pixel sample two. The NDVI values of pixel sample one in all years (from 2000 to 2010) were used to construct complete NDVI time series necessary for premodeling analysis.

On the temporal dimension, values in odd years and in even years for the two pixel samples were extracted separately. Values of pixel sample one in odd years were used for model development (as training data). Values of pixel sample one in even years are temporally independent, and thereby were used to validate the developed models for their ability to generalize to temporally independent datasets (temporal cross validation). Similarly, values of pixel sample two in odd years were selected as the spatially independent data and used for spatial cross validation; values of pixel sample two in even years were both temporally and spatially independent and were used for spatio-temporal cross validation.

3.2.5 Outlier Exclusion

Although the eMODIS NDVI product is strictly preprocessed and carefully composited, the residual noise after the bad quality values are removed constantly exists in the NDVI time series due to various reasons such as cloud contamination and poor atmospheric conditions.

Inherently, vegetation growth and decline as indicated by NDVI are in a gradual manner. Noise present in the NDVI time series (mainly atmospheric aerosols and snow) tends to depress NDVI values. These reduced NDVI values, if not excluded, will lead to biased vegetation-environment relationship and erroneous phenomodels. An effective and robust technique of high-quality NDVI time series reconstruction is needed to help identify outliers with abnormally reduced NDVI values while it does not mistakenly

exclude NDVI values with normally greater increase or decrease during greenup or senescence.

Commonly used techniques of NDVI time series noise reduction and reconstruction include the Best Index Slope Extraction (BISE) algorithm, Fourier-based fitting method and some iteration based algorithms (Ma and Veroustraete, 2006). An iterated Savitzky-Golay filtering by Chen et al., 2004 is adopted to smooth the original NDVI time series, remove noise and reconstruct high-quality NDVI time series by approaching to the upper NDVI envelope progressively. Chen et al.'s (2004) algorithm is demonstrated to be able to achieve the equally high quality reconstructed time series as the BISE algorithm, while the Fourier-based method has the smoothest but the biggest displacement from the original time series. Also, unlike the BISE algorithm, no parameters need to be determined beforehand to achieve the optimal results. The Savitzky-Golay smoothing filter used in the iteration based algorithms in Chen et al. (2004) also has advantages in preservation of local features (such as relative maxima and minima). The data points in the original NDVI time series that are far below the reconstructed time series are very likely to be less reliable, and are thereby considered outliers and excluded from modeling.

Figure 3.4 illustrates the time series reconstruction and outlier exclusion of one arbitrary sampled pixel (step 1 to 6). The pixel-wise algorithm is as follows:

1. Remove predetermined outliers.
    a) The NDVI values identified as bad quality during the production of eMODIS-NDVI are removed.
    b) The data points along the NDVI time series with an increase of more

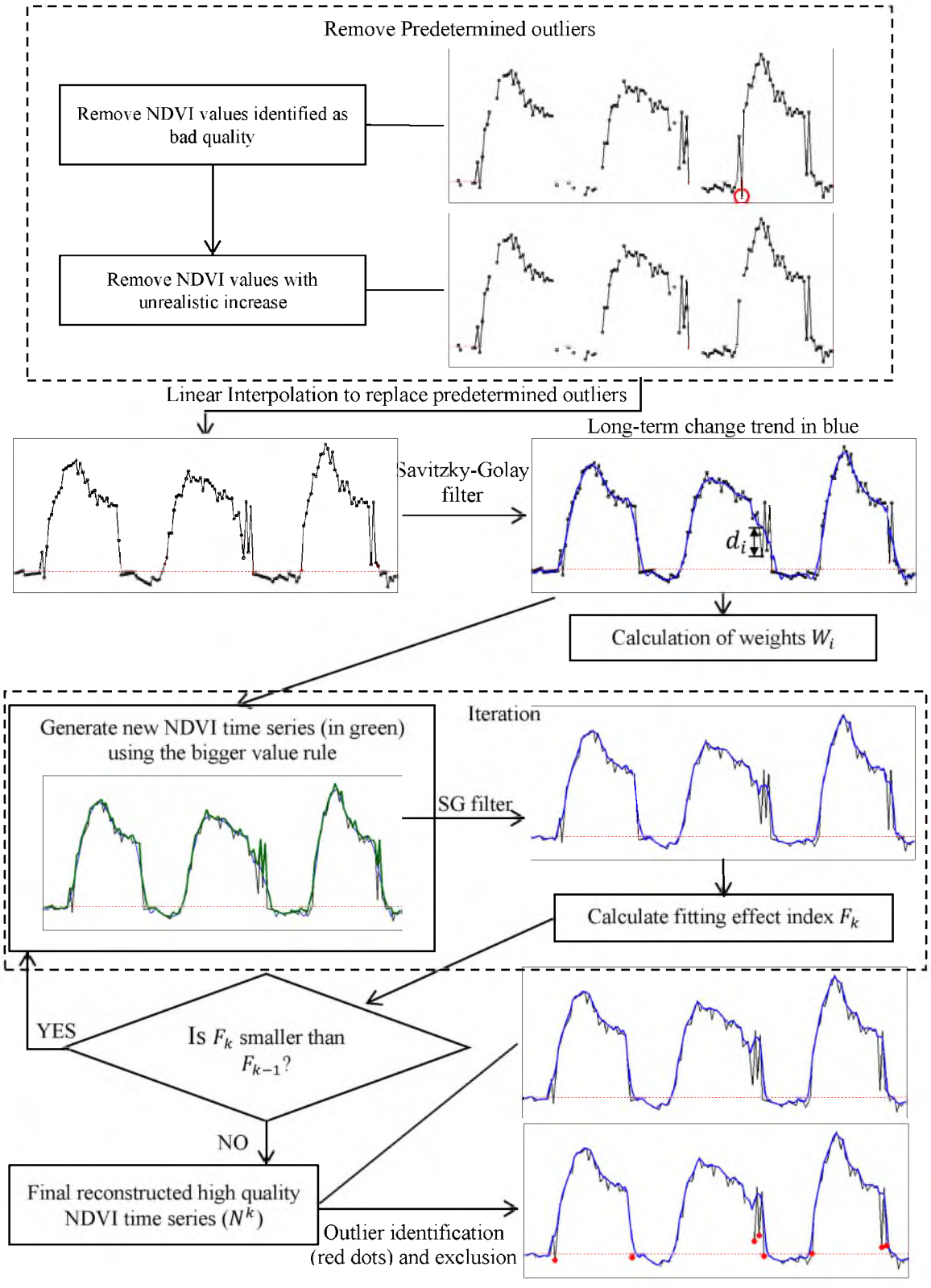


Figure 3.4 Illustration of Chen et al.'s (2004) NDVI time series reconstruction.

than 0.4 within the next three 7-day intervals are considered against the natural vegetation growth cycle, and are also excluded.

2. Replace the predetermined outliers with linearly interpolated values.
3. Apply the Savitzky-Golay filter to generate the long-term change trend.
4. Calculate the weight for each point on the NDVI time series as:

$$W_i = \begin{cases} 1, & \mathrm{N}_i^0 \geq \mathrm{N}_i^{tr} \\ 1 - \dfrac{d_i}{d_{max}}, & \mathrm{N}_i^0 < \mathrm{N}_i^{tr} \end{cases}$$

where W_i is the weight of data point i on the NDVI time series, N^0 is the linearly interpolated NDVI time series, N^{tr} is the long-term change trend generated using the Savitzky-Golay filter, d is the absolute distance between N^0 and N^{tr}. The points on N^0 that are above N^{tr} are considered to be able to represent the actual vegetation cycle, and thereby are given higher weight (1). The points below N^{tr} are likely depressed due to cloud, snow or poor atmospheric conditions, and are thereby assigned weights based on the distance between N^0 and N^{tr}.

5. Iteration:
    a) Generate new NDVI time series from the linearly interpolated time series (N^0) and the Savitzky-Golay fitted time series (N^{tr} for the first time and N^{k+1} during iteration). The Savitzky-Golay filter with the smoothing window size of 9 and degree of the smoothing polynomial of 6 was applied, which was demonstrated by Chen et al. (2004) to have the optimal performance. The new NDVI time series is generated as:

$$N_i^1 = \begin{cases} N_i^0, & N_i^0 \geq N_i^{tr} \\ N_i^{tr}, & N_i^0 < N_i^{tr} \end{cases}$$ for the first time and

$$N_i^{new} = \begin{cases} N_i^0, & N_i^0 \geq N_i^{k+1} \\ N_i^{k+1}, & N_i^0 < N_i^{k+1} \end{cases}$$ during iteration,

where N_i^1 is the initially generated NDVI time series by taking bigger values from using N^0 and N^{tr}, N^{k+1} is the fitted NDVI time series using the Savitzky-Golay filter during the kth iteration, and N^{new} is the new NDVI time series generated by taking bigger values from N^0 and N^{k+1} during each iteration.

b) Apply the Savitzky-Golay filter on N^1 for the first iteration and N^{new} for the rest iterations to gradually approach the upper NDVI envelope.

c) Calculate a fitting effect index as:

$$F_k = \sum_i (|N_i^{k+1} - N_i^0| \times W_i)$$

where F_k is the fitting effect index with smaller value indicating a better fit.

6. Repeat 5 until F_k stops decreasing and choose the N^{k+1} during the kth iteration corresponding to the smallest F_k as the final reconstructed NDVI time series.
7. Outlier exclusion: Outliers can be pinpointed by comparing the original NDVI time series (time series after step 1) and the reconstructed time series.

   a) For each data points on the original time series that are below the reconstructed time series, calculate the distance.

b) Outliers are identified as the data points with distances greater than 0.03. The selection of this threshold directly determines the performance of the phenomodels. A higher threshold value will cause outliers to still remain in the modeling dataset. A lower threshold value will cause the reduction of the data available for modeling and a higher possibility of removal of reliable NDVI values. Several thresholds were tested in the outlier removal process, and 0.03 was found to be a value that can achieve a trade-off between the two situations.

c) Due to the edge effect of the Savitzky–Golay filter, the fitted values of the first four and last four points along the NDVI time series are not reliable. Therefore, the first and last four points are also removed from the time series as outliers, because it is hard to estimate the reliability of the original NDVI values without their correspondent reliable reconstructed values.

3.3 Time Series Analysis

3.3.1 Seven-day Phenoregional Mean NDVI Time Series

Since it is impossible to scrutinize the phenological cycles of all sampled pixels represented by NDVI time series within respective phenoregions, a time series representative of the phenological cycles of all pixels within a phenoregion need to be generated. This representative NDVI time series helps to examine the multiyear vegetation growth cycle and the annual cycle, and to identify phenoregion specific onsets of phenophases.

According to the phenoregion delineation rationale, pixels within the same homogeneous phenoregion should share similar phenological cycles. Further, the sampled pixels using the stratified systematic sampling are representative of all pixels in a phenoregion. Therefore, it can be assumed that the spatially averaged time series over all sampled pixels in this phenoregion can characterize the phenological cycle of that phenoregion as an integrate entity.

However, simply averaging the time series, i.e., using the averaged values of only good quality pixels, can easily lead to artifacts due to the nonnegligible data quality issue summarized above. The fact that the images having a large proportion of bad quality values can result in time series composed of averaged values over different sets of pixels on different dates. What is more, as stated in Section 3.2.5, a lot of residual noise remains in the NDVI time series even after the careful and strict preprocessing during the eMODIS-NDVI production. The phenoregional mean time series computed as the average of all reconstructed NDVI time series, however, can deal with both issues. The action of averaging itself further increases the reliability by reducing the remaining noise associated with time series of single pixels.

3.3.2 Time Series Decomposition

In order to identify a uniform set of phenophase onset dates for multiple years in a phenoregion, the second step is to summarize a uniform seasonal cycle, conducted by time series decomposition. Time series decomposition deconstructs a time series into three components: the trend component (the long term general movements of NDVI values), the seasonal component (the cyclical component that roughly repeat itself at a certain frequency, representing seasonality) and the remainder (the noise or random

fluctuations). The seasonal component is the uniform cycle that can be used to extract a uniform set of phenophase onsets for specific phenoregions.

The time series decomposition requires a fixed frequency (e.g., 365 days, 12 months, or 4 seasons) per year. However, the number of 7-day intervals varies across years. Considering also the interval irregularity discussed before, the multiyear 7-day phenoregional mean time series are extended to daily ones with 365 days in each year. The phenoregional mean NDVI value on a specific day is taken from the 7-day time series at the 7-day interval this day belongs to. For overlapping days, the daily value is the mean value of the two 7-day intervals containing this day. The values on the three 366^{th} days are combined with those on the 365^{th} day of respective years.

A Seasonal Trend decomposition procedure based on Loess (STL) was adopted to decompose the reconstructed daily phenoregional mean NDVI time series. STL (Cleveland et al., 1990) is one of the most popular time series decomposition algorithms. It uses a single smoother, the locally weighted regression (loess) to conduct the trend and seasonal smoothing. Firstly, STL finds the seasonal component by smoothing 365 seasonal subseries, with each one constituting values on the same day of year from 2000 to 2010. Secondly the Loess smoother is applied again on the time series with the seasonal component removed, to find the trend component. The residuals after the removal of seasonal and trend components constitute the remainder. The iterations of the two steps are not necessary for time series that converges fast. Otherwise, the time series is detrended (trend component removed), and a set of robust weights is calculated and used in the next iteration of the above steps. The number of iterations can be determined beforehand, or iterations can continue until convergence. The STL approach, as

compared with other methods of time series decomposition, is flexible (e.g., capable of controlling the degree of smoothness in trend and seasonal smoothing), allows fast computation, can handle any type of seasonality, and is robust to outliers.

The reconstructed daily phenoregional mean NDVI time series are decomposed using STL for each phenoregion as shown in Figures 3.5 and 3.6.

3.3.3 Determination of Phenophase Onsets

As stated in Section 2, the environment-vegetation relationship is temporally variant. While phenoregions are delineated to achieve the spatially relatively homogeneous relationship, the temporal variation also needs to be addressed for the

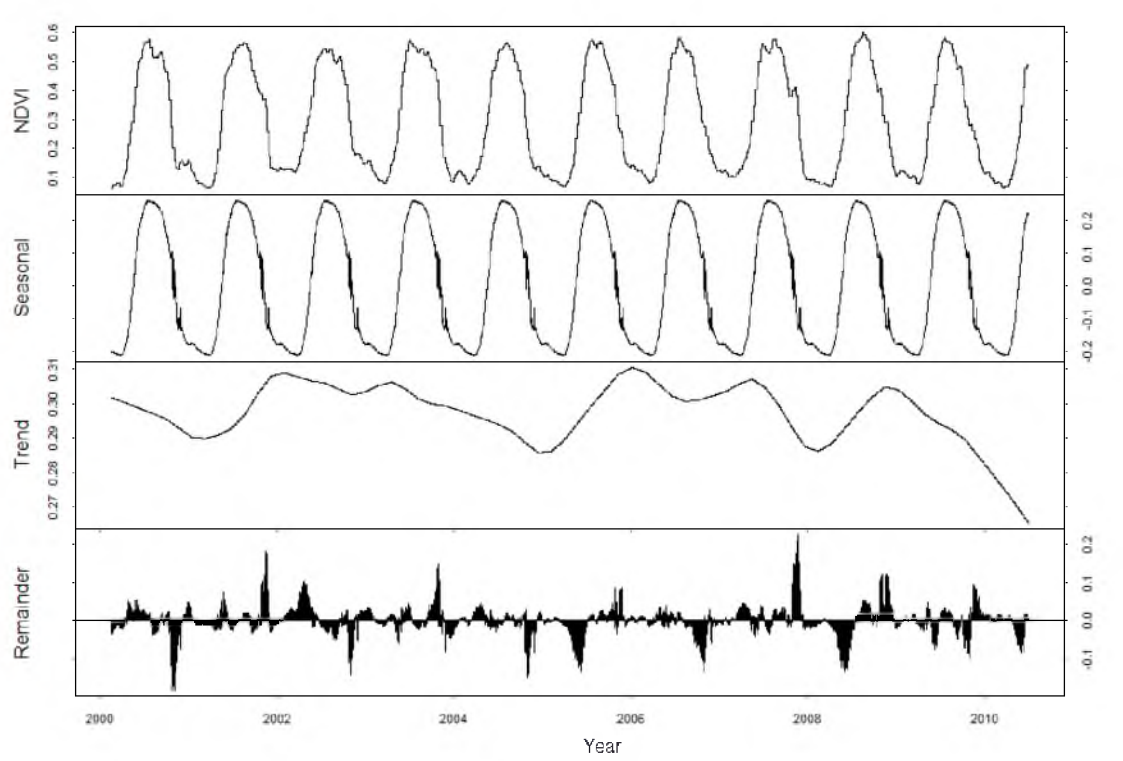


Figure 3.5 NDVI time series decomposed into seasonal, trend and remainder component by STL.

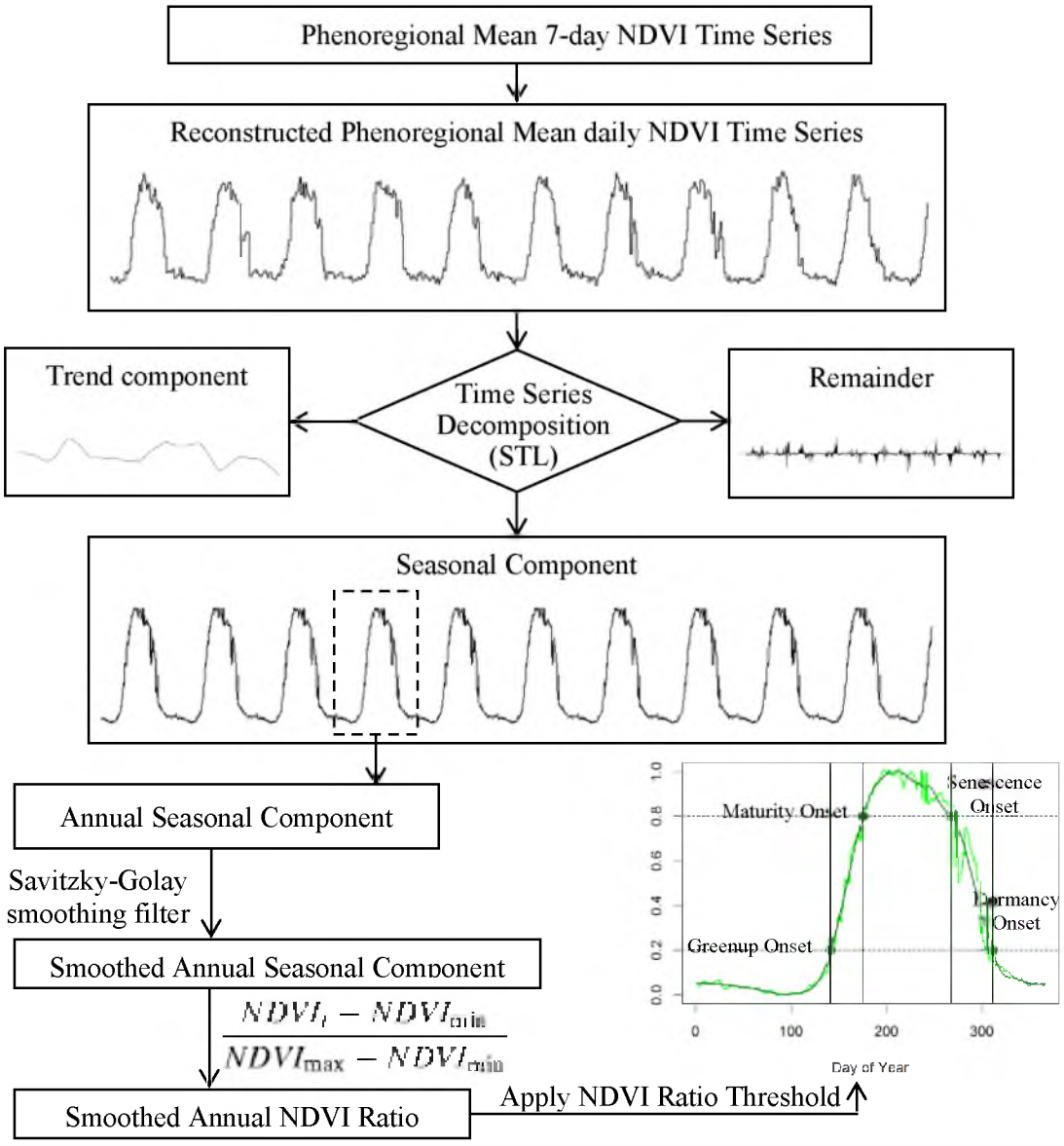


Figure 3.6 Diagram of phenophases identification.

fulfillment of the research objectives. Division of a year into phenological phases is one of the two measures to account for it. It is expected that within the resulted respective period, the vegetation-environment relationship is more homogeneous. A diagram of the phenophases identification in a specific phenoregion is shown in Figure 3.6.

The decomposed seasonal component using STL is uniform throughout all years, and thereby a set of uniform phenophases onset dates can be determined for each phenoregion. Four phenophases described from Zhang et al., 2003 are adopted:

1. Greenup: The NDVI values obviously start to increase;
2. Maturity: The increase of the NDVI values slows down and remains stable during this phenophase;
3. Senescence: The NDVI values obviously start to decrease;
4. Dormancy: The decrease of the NDVI value slows down and remains at a very low level throughout this phenophase.

Generally, as summarized in Section 2.1.1, there are several widely accepted ways for phenophase determination: thresholding, growth rate, and certain percent of the annual amplitude of the VI time series. The thresholding method uses an empirically or experimentally determined VI threshold or a VI threshold that has been successfully applied in other research to determine the phenophase onset. The growth rate method detects the rate of change of VI or curvature, and uses the dates of the maximal and minimal rates as onset dates. The third method is also called the VI ratio method, which uses a level of VI relative to the amplitude to determine the onsets. The purpose of phenophase division here is to partition a year into different temporal segments, during each of which the environment-vegetation relationships are expected to be more

homogeneous. For the first method, the NDVI threshold is hard to determine for different phenoregions at different locations and composed of different vegetation communities. The second method applies best to the quasitrapezoidal shaped time series, and is thus not the most appropriate for this analysis due to the different shapes of NDVI time series in different phenoregions. The most suitable method for this analysis is the percent of annual amplitude due to its versatility. The NDVI ratio threshold can be selected and can be at different levels for different phenophases. Before the NDVI ratio can be applied to the seasonal component, the seasonal component has to be smoothed to further reduce noise, avoid possibly false maximum and minimum NDVI values, and to achieve the local monotonicity to prevent the derivation of two close onset and offset dates for the same phenophases. The Savitzky–Golay smoothing filter was applied for smoothing due to its advantages in preservation of local maxima and minima.

NDVI ratio, developed by Kogan (1995) and Burgan and Hartford (1993), is calculated as normalized NDVI ranging from 0% to 100% as follows:

$$\mathrm{NDVI}_{ratio}(t) = \frac{NDVI_{\boldsymbol{t}} - NDVI_{\boldsymbol{min}}}{NDVI_{\boldsymbol{max}} - NDVI_{min}}$$

where $\mathrm{NDVI}_{ratio}(t)$ is the NDVI ratio at day t, $NDVI_{\boldsymbol{t}}$ is the NDVI value at day t, $NDVI_{\boldsymbol{max}}$and $NDVI_{min}$ are, respectively, the annual maximal and minimal values. The actual NDVI values are thereby transformed to a ratio form. NDVI ratio of 50% is the most commonly used threshold. It is believed that the NDVI increase is the most rapid at this threshold. However, the purpose of the phenophase division is to capture the whole continuous stretch of NDVI increase and decrease, during which it is assumed that the lagged response of vegetation to environment is consistent. Other commonly used thresholds include 10% and 90%, however, to avoid the mistaken use of the snowmelt

onset as the greenup onset, 20% and 80% of NDVI ratio are used as the thresholds to determine phenophase onset (Figure 3.7).

3.4 Premodeling Preparation and Analysis

3.4.1 Exclusion of Nonnaturally-Vegetated Effects

This project focuses on plant phenology, so it is necessary to reduce nonvegetative effects on the data to avoid biased analysis and modeling results. Before modeling, NDVI values are further examined to ensure the data points that are used to build models contain as few nonvegetative effects as possible. Only naturally vegetated

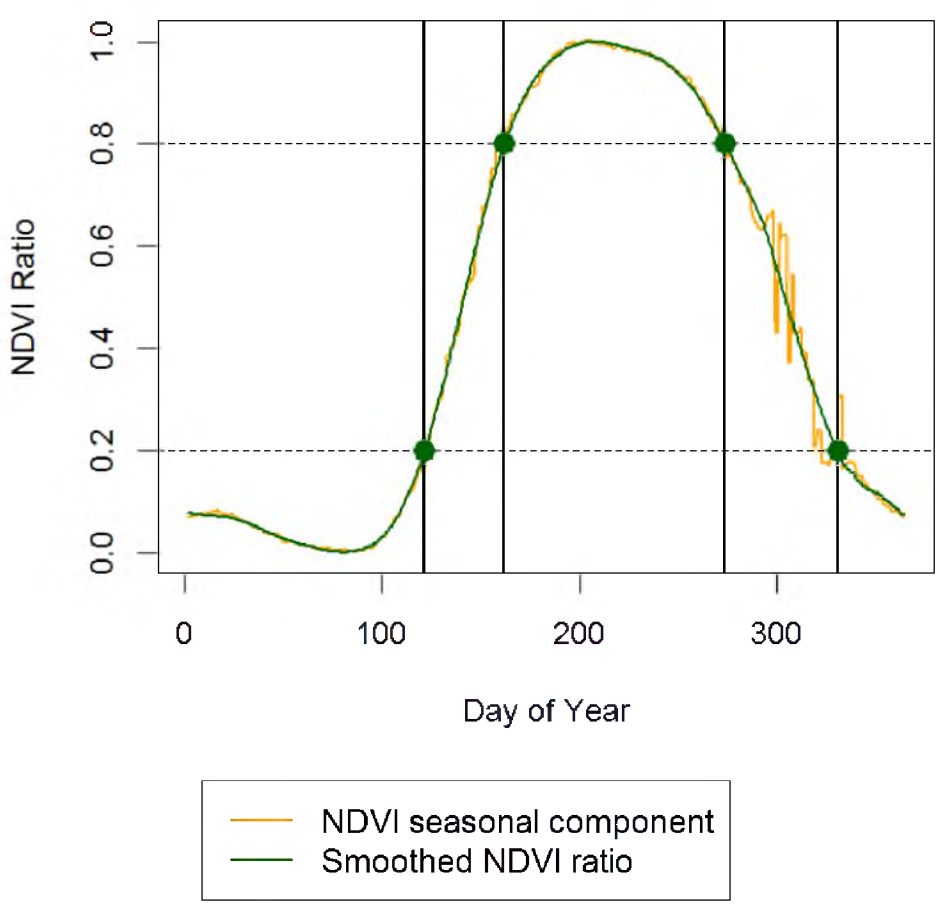


Figure 3.7 Determination of phenophases onset days.

NDVI values are retained. Firstly, snow cover, though representing the actual land surface, is regarded as noise in this research and NDVI values marked as snow in the eMODIS-NDVI quality files are removed. Secondly, the NDVI values below 0.1 are considered to contain no meaningful information in the context of vegetation phenology, and therefore are also removed. Thirdly, the naturally vegetated pixels and nonnaturally-vegetated pixels are identified with the assistance of National Gap Analysis Program (GAP) Land Cover Data (USGS, 2010), and the nonnaturally vegetated pixels are excluded from the following analysis and modeling.

GAP data published in 2010 are used for this identification purpose. GAP land cover map at 30 m resolution is a seamless combination of the work of several different projects representing the ground land cover from 1999 to 2001 (USGS, 2010). GAP data have three levels of details, from the most general of eight classes to the most detailed of 75 classes within the study area of the UCRB (Appendix A). The level-one classification is adopted to assist with the identification of nonnaturally-vegetated pixels. Class one of human land use, class two of aquatic and class three of sparse and barren systems are considered as nonnaturally-vegetated. Since the sampled pixels are 1 km^2, it is highly possible that they are composed of several different land cover types. The layer of sampled pixels were overlaid on the 30-m GAP land cover map to derive the percentage of three classes of human land use, aquatic and sparse and barren systems combined. Sampled pixels with a combined percentage of more than 50% are identified as nonnaturally-vegetated pixels and are thereby excluded.

3.4.2 Preparation of Candidate Variables

The literature review indicates that the spatial and temporal variations of the environmental drivers and factors influence vegetation growth and environment-vegetation relationships. All variables related to environmental drivers should be included in the phenomodeling because their temporal variations directly "drive" plant development. The topographic variables (elevation, slope and aspect) as well as other time invariant variables used in phenoregion delineation, should also be included in phenomodeling to moderate the influence of environmental drivers and other environmental factors on the 7-day vegetation dynamics, and to account for the inter-pixel difference of the environment-vegetation relationships. The GAP level one land cover is also included as one of the candidate variables (GAP) to account for the influence of land cover types on the environment-vegetation relationships.

In the temperature category, the minimum (TMIN) and maximum temperature (TMAX) that can be directly extracted from the Ecocast data are included as candidate variables. The weekly temperature difference (TDIFF), calculated as the difference between TMAX and TMIN is also included, it is argued to have a favorable effect on plant growth (Larcher, 2003). The mean temperature (TMEAN), calculated as the average value of the TMAX and TMIN is also one of the candidate variables. Further, the growing degree days (GDD) and the GDD with the upper threshold (GDDu), as well as the accumulated values of these two variables (AGDD and AGDDu), starting from the first day of each year are also included, because it is indicated in a considerable amount of literature that they are directly associated with vegetation abundance.

As discussed in Section 2.2.1.1, there are two forms of GDD: one calculated with only the base temperature and the other with both the base temperature and the upper temperature threshold. The values of the smoothed daily mean phenoregional time series of TMEAN at the greenup onset dates in different years are averaged to serve as the base temperature. If the calculated base temperature is below zero, then it was set to zero. Based on the physical meaning of the upper temperature threshold (vegetation growth stops or slows down dramatically), the threshold is calculated for each phenoregion as follows (Figure 3.8): firstly, obtain the days for each year with a positive growth rate on the multiyear daily reconstructed phenoregional mean NDVI time series; secondly, obtain the TMEAN values on those days extracting from the smoothed TMEAN time series; thirdly, derive the maximal value from those TMEAN values for each year; finally, average over each years' maximal TMEAN values as the upper temperature threshold for

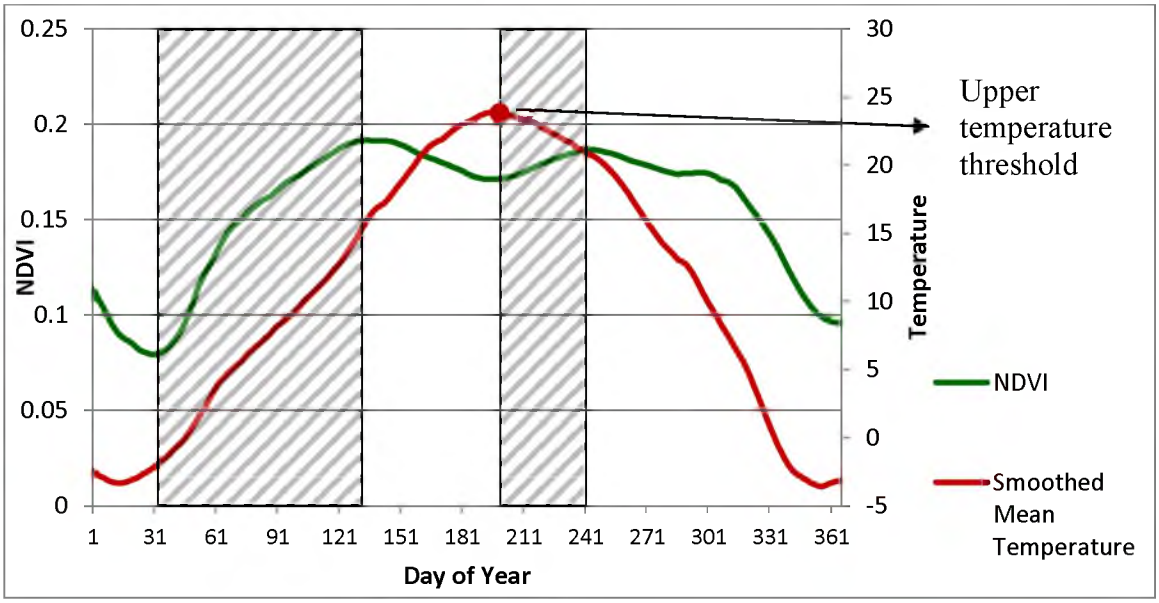


Figure 3.8 Demonstration of upper temperature threshold extraction (hashed regions are the dates with a positive growth rate).

that phenoregion. The phenoregion-specific base temperature and upper temperature threshold are shown in Table 3.3. The TMAX, TMIN, TDIFF, TMEAN, GDD, GDDu, AGDD, AGDDu at the lag of 1 to 10 weeks are extracted or calculated to explain the lagged responses of vegetation to the environment.

In the precipitation category, the variables of precipitation at lags of 1 to 20 weeks are included as candidate variables to account for the variant lagged responses of vegetation to precipitation in different phenoregions.

In the light category, solar radiation (SRAD) from the Ecocast data is one of the candidate variables. Photoperiod (PTPD) also has great influence on plant development (Henfrey, 1852; Tournois, 1912; Tournois, 1914; Klebs, 1913; Garner and Allard, 1920; Garner and Allard, 1923; Thomas and Vince-Prue, 1996). Photoperiod or day length can be calculated from the latitude and Julian day. Solar radiant energy (SRE), calculated as the product of SRAD and PTPD, is yet another candidate variable of the light category.

Last but not the least, the antecedent NDVI (aNDVI), i.e., the correspondent NDVI values 7 days ago, is also included in the models. Vegetation responds to the environment with the exhibition of NDVI increase or decrease on the basis of aNDVI. The antecedent status and developmental stages of vegetation also influence how it responds to the environmental variations.

Table 3.3 Base temperature and upper temperature threshold for the calculation of GDD and GDDu in the nine phenoregions

Phenoregion	1	2	3	4	5	6	7	8	9
Base temperature (°C)	0	0	2.67	3.71	0	3.01	2.22	0	0
Upper temperature threshold (°C)	24.5	15.9	14.85	15.13	24.42	18.27	15.76	21.75	20.47

Table 3.4 shows all 147 candidate variables and their abbreviations of four categories of environmental drivers and topography. For all variables of environmental drivers, the variables at the lag of 1 week were not given any suffixes (such as TMAX and SRAD), and variables at the lag of more than 1 week were named with suffixes of the number of lags in weeks (such as TMAX_2 through TMAX_10). Categorical variables are marked in bold font in Table 3.4. The variable ASPECT has three categories: north-facing (ASPECT_N), south-facing (ASPECT_S) and west- or east-facing (ASPECT_WE). The SEASON has four categories: SEASON_1 (the greenup phase), SEASON_2 (the maturity phase), SEASON_3 (the senescence phase), and SEASON_4 (the dormancy phase). The GAP variable has eight categories correspondent to the eight classes of the GAP level one classification, with GAP_1 through GAP_9 represents the eight land cover classes of "human land use," "aquatic," "sparse and barren systems," "forest and woodland systems," "shrubland, steppe and savanna systems," "grassland systems," "recently disturbed or modified," and "riparian and wetland systems," and code 6 is not used in the GAP level one coding.

3.4.3 Multicollinearity and Internal Structure

The environment exerts its influence on 7-day vegetation dynamics as an entity, so the influence of each environmental component is entangled. The environmental drivers and factors themselves are also highly correlated, making this situation even more complicated. The correlations between independent variables in the models are referred to as multicollinearity. Although MARS is robust to the violation of many assumptions required by ordinary multivariate regression models, multicollinearity is a serious concern. Multicollinearity can lead to the arbitrariness of variable selection among highly

Table 3.4 Candidate variables and their abbreviations

Category	Lag in Weeks	Candidate Variables	Abbreviations
Vegetation abundance	1	Antecedent NDVI	aNDVI
Temperature	1 - 10	Maximum Temperature	TMAX
		Minimum Temperature	TMIN
		Weekly Temperature Difference	TDIFF
		Mean Temperature	TMEAN
		Growing Degree Days	GDD
		Growing Degree Days with upper threshold	GDDu
		Accumulated Growing Degree Days	AGDD
		Accumulated Growing Degree Days with upper threshold	AGDDu
Precipitation	1 - 20	Precipitation	PRCP
Light	1 - 10	Solar Radiation	SRAD
		Photoperiod	PTPD
		Solar Radiant Energy	SRE
Topography	Time invariant	Elevation	ELEV
		Slope	SLOPE
		Aspect	ASPECT
Other		Phenological Phase	SEASON
		Mean Maximum temperature	MEAN_TMAX
		Mean Minimum Temperature	MEAN_TMIN
		Mean maximum temperature during growing season	GS_TMAX
		Mean minimum temperature during growing season	GS_TMIN
		Standard deviation of monthly temperature	TEMP_STD
		Mean precipitation	MEAN_PRCP
		Mean precipitation during growing season	GS_PRCP
		Standard deviation of monthly precipitation	PRCP_STD
		Soil variability index	SVI
		Mean annual NDVI	MEAN_NDVI
		Latitude	LAT
		The GAP level 1 land cover	GAP

correlated variables, and further affect the selection of variables and knots in the following steps. Multicollinearity can also create problems in interpreting the models and in making predictions for regions other than the sampled pixels if the correlation structure differs (Moilanen et al., 2009). Therefore, to build a robust and practically meaningful MARS model, multicollinearity has to be taken care of in advance.

The reduction of multicollinearity in the context of MARS is not well established in literature. Most solutions use PCA before modeling or PCA related regressions. This approach addresses well the multicollinearity problem, but makes the modeling results harder to interpret and masks the interactions between the original independent variables before the PCA transformation.

Therefore, an approach of variable reduction by iteratively removing variables is adopted. Since TMEAN and TDIFF are the linear combination of TMAX and TMIN, only one of these two sets of variables can be kept to avoid the perfect multicollinearity.

Variance Inflation Factor (VIF) is a commonly used statistic to examine the multicollinearity. The VIF of an independent variable X_i is defined as:

$$VIF_i = \frac{1}{1 - R_i^2}$$

where R_i^2 is the coefficient of determination of the regression with X_i being the dependent variable and all the other independent variables being the independent variables. Two rule-of-thumb values of VIF are proposed as thresholds in the literature: 5 and 10 (Acevedo, 2012). Here the VIF of 10 is used as the cut-off value, which indicates that 90% of the variance in X_i can be explained by the linear combination of other variables. The more conservative value of 5 is not used to avoid the risk that useful information for NDVI prediction may be lost.

The variable with the maximal VIF value is iteratively removed until the VIF values of all remaining variables are less than 10. It is worth noting that the removed variables may still be important to the 7-day vegetation dynamics. The only reason variables are removed is that trends can be explained by other correlated variables. Therefore, the interpretation of the model should also consider the variables that are highly correlated with the variables selected by MARS. In this sense, it is necessary to analyze the entangled relationship within the environment between environmental drivers and factors.

Principal Component Analysis (PCA) provides a way to scrutinize the internal structure of environmental factors and environmental drivers at different lags. The PCA loadings indicate the relative importance of each environmental driver and factor, and the cumulative variance provides insight into the environmental variance accounted for by each PC.

It is also important to examine the lag structure of the time variant dependent and independent variables. The lag structure of a variable calculates the correlation between the current value of the variable and its antecedent values at different lags, and thereby exhibits persistence or stability of a variable over time.

3.5 Model Development

3.5.1 Multivariate Adaptive Regression Splines

Multivariate regression is a widely used and rather mature approach due to its ability to convey the empirical quantitative associations between the independent and dependent variables once the causal relationships have been verified. It is also superior in

implicitly expressing the main and partial effects of independent variables on the dependent variable, as well as how these effects are influenced by the moderate variables.

However, ordinary multivariate regression has the disadvantage that it has to satisfy a lot of assumptions such as a linear relationship, normal distribution and homoscedasticity of residuals, making it much less flexible. Also, the vegetation-environment relationship in reality is complex: it varies in different intervals of environmental drivers and factors, in different phenophases, and the interaction between variables further complicates relationships. The complex relationships between vegetation abundance and all environmental drivers and factors are extremely difficult to quantify using linear regression.

Some other modeling approaches, such as neural network or random forests, have been reported to be able to handle complex and nonlinear relationships and can achieve a high goodness of fit (Francis, 2003; Prasad et al., 2006). However, they either transform independent variables, or conduct the modeling in a "black box," or produce a model that is difficult to understand and interpret.

Considering that the understanding and interpretation of environment-vegetation relationship via the phenomodels is one of the three research objectives in this dissertation, the Multivariate Adaptive Regression Splines introduced by Friedman (1991) is a more reasonable modeling approach.

MARS is a nonparametric regression that has been applied to model complex relationships and deal with high dimensions of independent variables. MARS does not make any assumptions about the underlying relationships between the dependent and independent variables. Instead, MARS is composed of a series of data-driven "basis

functions" that models the relationships within different intervals of data, and automatically identifies the "knots" or breakpoints that separate and connect different relationships. The MARS also adopts the least squares approach in approximating the overdetermined systems, aiming to minimize the sum of squares of the residuals. The MARS models take the form of

$$\mathrm{Y} = \beta_0 + \sum_{i=1}^{k} \beta_i B_i(X)$$

where $B_i(X)$ is a basis function. Essentially, MARS is an extension to the linear regression in the form of piecewise regression. Each subregion in the data space partitioned by the knots has its own regression equation. Therefore, MARS has advantages in flexibly handling nonlinear and nonmonotonic relationships, which are difficult for linear regression to reveal. The piecewise nature of MARS makes it especially appropriate in modeling the complex vegetation-environment relationships that are variant dependent on different levels of environmental drivers and factors.

MARS also automatically models the interactions between independent variables within respective subregions in the data space. As stated in Section 2.2.3, plant development is the results of the overall influence of the environment as a whole. Besides the single environmental drivers and factors, the influence of the interactions between them on vegetation is also important. Therefore, the ability of MARS to model locally existing and variant interactions is another reason that makes it a better candidate modeling approach.

The MARS models are in a format that is simple to understand and interpret as to the effect of each independent variable on the dependent variable. It can be converted to

the form of linear regression models within respective subregions in the data space. This characteristic satisfies the objective of interpretation of the complex lagged vegetation-to-environment responses in this dissertation.

One requirement of the spline-fitting of MARS, especially in the context of high dimensional data, is that a large number of observations are needed. A MARS model with n independent variables and k knots has $(k+1)^n$ subregions. Each subregion needs a certain number of observations to place the knots and estimate the coefficients. Generally, thousands, or hundreds of thousands of observations are required for spline fitting. Although MARS addresses this problem by constructing the model as the summation of basis functions, instead of dealing with each subregion explicitly, still a reasonably large number of observations are required to fit MARS, which is considered a main drawback of the mothod. However, in the context of remote sensing phenological modeling, this requirement is not a problem at all. Each phenoregion has $(\text{number of pixels} \times \text{number of pairs of neighboring } 7-\text{day intervals})$ observations, which are generally much more than enough for the application of MARS even after sampling and outlier exclusion. On the other hand, MARS performs well in dealing with large datasets. The MARS models can be built quickly even with a very large number of independent variables and observations. The development of MARS models include two phases: the forward pass and the backward pass (the pruning pass). The forward pass tends to build an overfit model as with other nonparametric models due to their adaptability. The backward pass then uses pruning techniques to simplify the models and increase the performance of its generalization to the prediction of other new datasets.

The forward pass starts with only the intercept included in the model and adds basis functions gradually to the model. A basis function $B_i(X)$ (Figure 3.9) is a two-sided truncated function or the product of several truncated functions (interaction terms) in the form:

$$\max(0, x - \mathrm{c}) \;\; i.e. \begin{cases} x - c, if x > c \\ 0, otherwise \end{cases}$$

or

$$\max(0, \mathrm{c} - x) \;\; i.e. \begin{cases} c - x, if x < c \\ 0, otherwise \end{cases}$$

or

$$\max(0, x_i - \mathrm{c}_i) * \max(0, x_j - \mathrm{c}_j)$$

where x is an independent variable and c is the knot location. The forward pass searches for each basis function with all possible knots exhaustively, and adds the one to the model that can minimizes the sum of squares of the residuals. This recursive process of

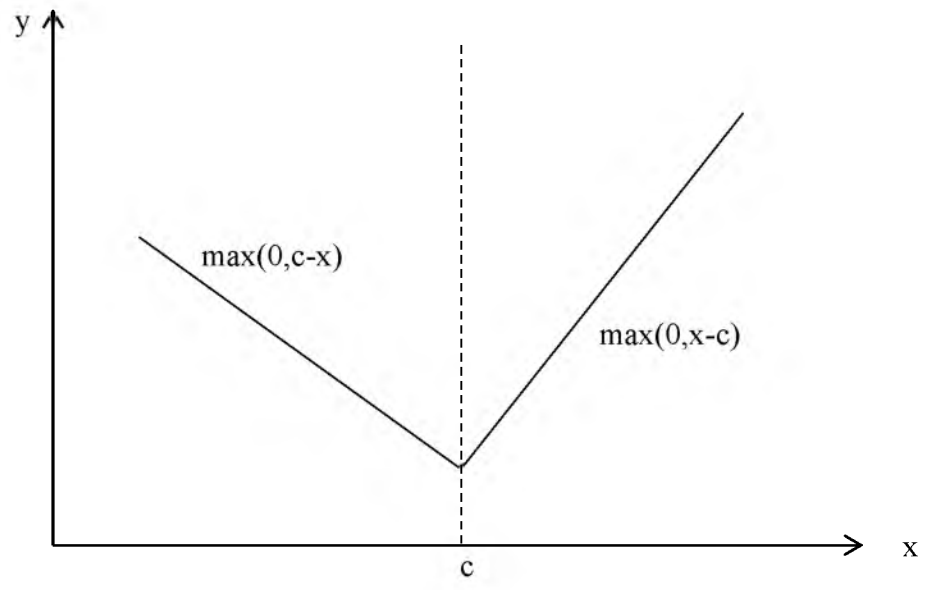


Figure 3.9 Example of an additive MARS model consisting of two basis functions.

adding basis functions stops if the sum of squares does not decrease any more or decreases by only a very small amount, or the model reaches a user-specified maximal complexity.

The backward pass reduces the complexity and increases the generalization ability of the possibly overfitted model built in the forward pass. The terms with the smallest contribution to the decrease of the sum of squares are excluded. The model subsets of the best performance are quantified by the so-called Generalized Cross Validation (GCV) criterion calculated as:

$$\mathrm{GCV} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{(1 - \frac{1 + cd}{N})}$$

where y_i and $\hat{y}_i$ are, respectively, the observed and predicted values of the independent variables, N is the number of cases, c is the penalty that increases with the number of terms, and d is effective degree of freedom. The GCV is a measure based on the trade-off between the goodness-of-fit and the complexity of the model. The model subset with the lowest GCV values is considered the optimal one.

3.5.2 Vegetation Dynamics and the Lagged Dependent Variable

The models in this research aim to disentangle the environment's influence on the 7-day vegetation dynamics. Specifically, the models represent how past NDVI values gradually develop into current NDVI values as influenced by the environment. Therefore, the models should be specified as a dynamic model and the simplest way is to include a Lagged Dependent Variable (LDV) to capture the vegetation dynamics so that the current NDVI value is a function of the antecedent NDVI value as modified by the

environmental drivers in the context of different environmental factors. The specification of dynamic models by including the LDV permits the simulation of the changes of NDVI on a gradual basis over time yielded by the antecedent values of environmental drivers. The coefficients of the environment drivers, factors or interaction terms in the final dynamic MARS models represent the influence of the respective terms on the change of NDVI values independent of the initial level.

There has been controversy about the inclusion of the LDV. If autocorrelation is present in the residuals, the LDV causes the coefficients of the independent variables to be biased downward (Keele and Kelly, 2005). But this is seldom a problem because the inclusion of the LDV can reduce the serial autocorrelation in the residuals. Also, the lagged dependent variable can lead to inconsistent estimates if the model is specified to include fixed effect, i.e., the individual specific effects are included in the residuals. This is because the LDV tends to be correlated with the residuals containing fixed effects. The presence of both LDV and fixed effects is also not a problem in this research, because MARS models with explicitly conveyed individual effects (by environmental factors and the LDV) were used instead of the fixed effects models due to the exploration and interpretation purpose of this research. Essentially, if the answer to the theoretical question "does the past NDVI value impact the current NDVI value in the vegetation development process?" is yes, then it is reasonable to include the lagged dependent variable in the specification of the models (Keele and Kelly, 2005), as long as no fixed effects are included and the residuals are not autocorrelated.

3.5.3 Modeling Procedure

Values of pixel sample one in odd years are used as the training data to build phenomodels. Figure 3.10 shows how values are extracted from images of environmental factors and images of NDVI and environmental drivers on all dates in odd years, and used as dependent and independent variables in the phenomodels.

The vegetation-environment relationship will be modeled using the MARS approach for all nine phenoregions. A unique phenomodel will be built for each of the nine phenoregions so that the unique environment vegetation relationships can be depicted that are different from those in other phenoregions. However, interventions and

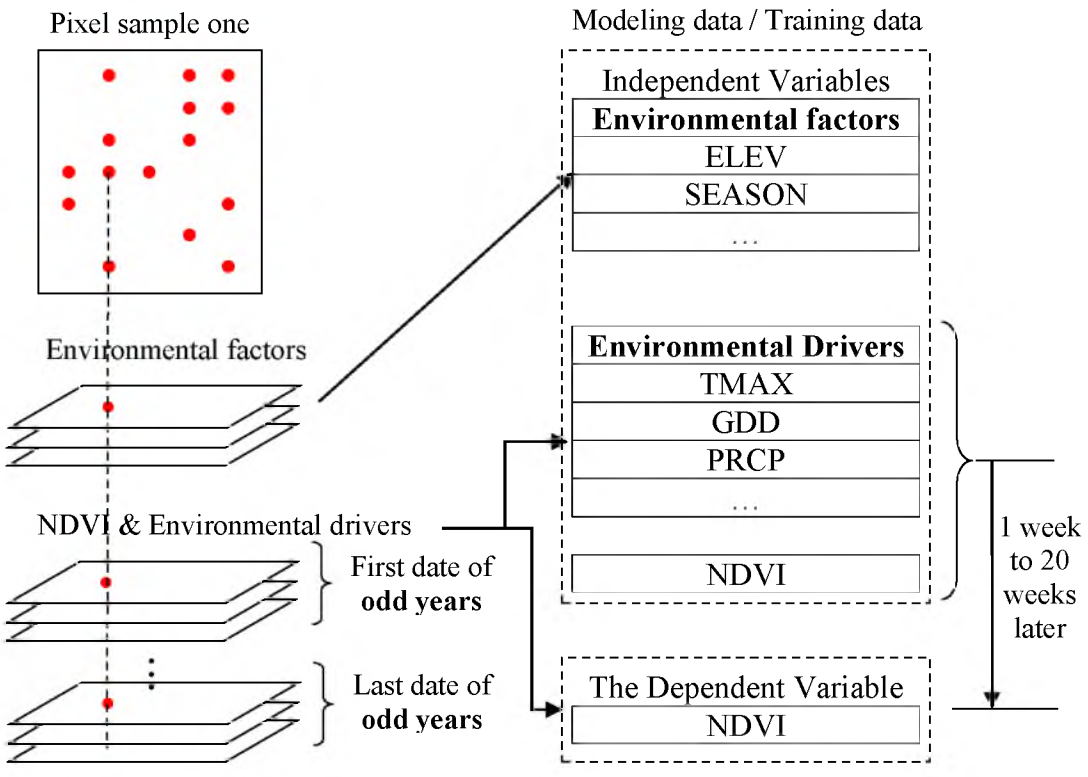


Figure 3.10 From pixel sample to modeling data (training data).

adjustments are needed during modeling in order to make sure the models represent as much of the actual relationship in the physical environment as possible. Models need to be examined repeatedly and adjusted accordingly as to:

1. Whether key variables are missing from the model;
2. Whether the modeled relationship and interactions are consistent with the widely accepted and strictly validated phenological and ecological theories;
3. Whether the locations of the knots are reasonable.

The manual constraints and adjustments assist in avoiding the situation where the completely data driven MARS model is possibly against current understandings of ecological processes.

All remaining variables after the VIF based iterated variable reduction were firstly restricted to enter into MARS models linearly. The second MARS models considers non-linear relationships by allowing the relationship to bend at different knots. The third one adds the interactions. For the sake of interpretability, only two degree interaction terms are allowed. The three models are compared using R^2, GCV and the standard error. The R^2 quantifies the amount of variance in the dependent variable that can be accounted for by the independent variables in the model specification. The GCV is a measure used for the MARS models to select the model with the best performance. The GCV considers both the goodness-of-fit and the complexity of the model and lower values of GCV denote better models. The standard error measures the accuracy of the model, by quantifying the mean deviation of the fitted values from the observed values.

All variables are divided into two categories. The first one directly influences how vegetation develop from aNDVI to NDVI within 1 week. Variables in this category also

influence how vegetation responds to other variables within the same category. This category includes all environmental drivers which are time variant. Variables in the second category include time-invariant environmental factors and aNDVI. They adjust the influence of variables in the first category on vegetation development. Therefore, the models take the form:

$$NDVI = \sum_i a_i ED_i + \sum_{ij} b_i ED_i * EF_j + \sum_{ij}^{i \neq j} c_i ED_i * ED_j + \sum_i d_i ED_i * aNDVI + \sum_i e_i EF_i + f * aNDVI + \varepsilon$$

where *ED* are environmental drivers and *EF* are environmental factors. The practical meaning can be more clearly seen if we change the equation above to:

$$NDVI = \left(\sum_i a_i + \sum_{ij} b_i EF_j + \sum_j c_i ED_j + \sum_i d_i * aNDVI \right) * \sum_i ED_i + \sum_i e_i EF_i + f * aNDVI + \varepsilon$$

The first component in the above equation indicates the influence of a specific environmental driver ED_i on the development of vegetation from the value of *aNDVI* to the value of *NDVI*. This influence is dependent on *aNDVI*, environmental factors and other environmental drivers, explicitly $\left(\sum_i a_i + \sum_{ij} b_i EF_j + \sum_j c_i ED_j + \sum_i d_i * \text{aNDVI}\right)$. Statistically, this term (the influence of ED_i) is composed of the main effect as well as the effect moderated by environmental factors, other environmental drivers and *aNDVI* in the order they appear in the term. The second component ($\sum_i e_i EF_i$), statistically the main effect of the environmental factors, works as the minor adjustment on *NDVI* accounting for the spatial difference of the realtionships.

3.6 Model Validations

3.6.1 Cross Validation

Cross validation applies the models in other independent data sets, in order to assess the models' performance and prediction accuracy. For each phenoregion, three different data sets which are, respectively, spatially independent, temporally independent, and spatio-temporally independent are selected for the conduction of spatial, temporal and spatio-temporal cross validation. The models are applied on values of pixel sample two in odd years for spatial cross validation, on values of pixel sample one in even years for temporal validation, and on values of pixel sample two in even years for spatio-temporal validation. The three kinds of cross validation are adopted to avoid overfitting and to ensure the models can be generalized both temporally and spatially.

For each type of cross validation in a phenoregion, two measures of fit were used to quantify how accurately the phenomodels predict.

Root Mean Square Error (RMSE) measures differences between predicted values and observed values (prediction errors), and is used to quantify the model performance and predicting accuracy. RMSE is calculated as:

$$\text{RMSE} = \sqrt{E((Y_{pred} - Y_{ob})^2)}$$

where Y_{pred} is the predicted value, and Y_{ob} is the observed value.

Coefficient of variation (CV), or the RMSE normalized to the mean of the observed values, is a relative measure of prediction accuracy and independent of the unit of the dependent variable. The CV is calculated as:

$$\text{CV} = \frac{\text{RMSE}}{\overline{Y_{ob}}}$$

The CV is usually expressed as a percentage and makes it easier to compare the accuracy of the models across different phenoregions and when applied to a different data set.

3.6.2 Field Validation

The field validation is to ensure that the phenomodels are capable of practical application. Field site sampling is one critical step determining the time spent in the field and the correctness of the validation. Since phenomodels are built for each phenoregion delineated within the UCRB, it is important to make sure field sites are selected in several phenoregions.

Since the spectra of the selected field sites have to be repeatedly measured once per week for a number of continuous weeks due to the 7-day time step of the models, it is preferred that these sites in different phenoregions are close to each other, i.e., these field sites should be located in a region with many patches belonging to different phenoregions. Also, it was preferred that they are located within an acceptable distance to major roads to reduce traveling time. A region west of Duchesne, Utah was selected based on these criteria. Several field sites were selected in each of the three phenoregions (Phenoregions 2, 8, and 9) located near Duchesne, Utah as candidates before the first trip to ensure that at least one of them was accessible (e.g., not fenced or the road was not in seasonal closure). Three of them were selected as the final field sites to be used in field validation, one for each phenoregion. The field sites in phenoregions two and eight were measured consecutively for 7 weeks, on 04/27, 05/04, 05/13, 05/21, 05/27, 06/03, and 06/10 of year 2011. The field site located in phenoregion nine was covered by snow until 05/21, so it was only measured for 4 weeks, from 05/21 to 06/10 in 2011.

At each field site, a 45 m long transect with randomly extending direction was selected, in order to make sure that the average NDVI measured along this transect could represent the NDVI of the 1 km^2 pixel the field site is in. The transects were marked with nails and stakes during the first measurement to make sure the same points along the same transect were measured each time consistently. These transects of different phenoregions were repeatedly measured each week for detection of NDVI changes. During each measurement, spectra of 16 points equally distributed along the 45 m transect every 3 m was measured five times using an Analytical Spectral Devices (ASD) Full Range field spectrometer. Spectra of a reference panel (spectralon) were recorded at each site to allow correction of the field spectra to reflectance.

In spectra processing, the radiance data were converted to reflectance, averaged over the five measurements at the same point, and further averaged across all points along a transect to derive the mean spectrum for a site. NDVI was calculated using the averaged reflectance over the MODIS spectral response function for the red band (620 – 670 nm) and near infrared band (841-876 nm), consistent with the range of bands one and two of MODIS.

Originally, the Ecocast data should have been used as the independent variables in field validation, however, the Ecocast data in 2011 was not available. Therefore, the currently only comparable dataset, DAYMET (Thornton, et al., 2012), is used in field validation. The DAYMET dataset is a daily climatic dataset generated from the Daymet model by the Numerical Terradynamic Simulation Group (NTSG) at the School of Forestry, University of Montana. It is also a 1-km dataset and contains the variables of maximum and minimum temperature, precipitation and solar radiation. The DAYMET

dataset was either averaged (solar radiation and maximum and minimum temperature) or accumulated (precipitation) over eight 7-day intervals of 2011109 - 2011115 through 2011158 - 2011164. An extra 7-day interval of 2011109 – 2011115 was included to make predictions on the NDVI of the 7-day interval of 2011116 – 2011122 containing the first field trip date. The field measured NDVI serves as the dependent variable of NDVI during these 7-day intervals as well as the independent variable of antecedent NDVI during the corresponding previous 7-day intervals.

The "predicted NDVI values" (the NDVI values predicted using the developed phenomodels with the input of field measured NDVI as the independent variable of antecedent NDVI and DAYMET data as other independent variables) were compared with the "observed values" (field measured NDVI). Similarly with cross validation, the prediction power of the models can also be quantified using the RMSE and CV.

3.6.2.1 Comparison of Ecocast and DAYMET data

Since a different data set from that used in the model development (Ecocast) is used in the field validation (DAYMET), a comparison between the two datasets is necessary to reveal any biases that may affect field validation.

Other than the DAYMET data for field validation (2011109 to 2011164), extra data from 2010005 to 2010179 were also downloaded, processed and composited to create the 7-day interval overlap with Ecocast data (available from 2000049 – 2000055 to 2010173 - 2010179) that enables the comparison.

For each of the 25 7-day intervals within the overlap, 100 pixels were randomly selected from the images of maximum and minimum temperature, precipitation and solar radiation for both Ecocast and DAYMET data. Four scatterplots of the 2500 pairs of

values correspondent to the four variables were generated and correlation coefficients were calculated.

3.7 Phenological Decision Support System

A prototype phenological Decision Support System was developed using C#.net. It aims to provide land managers in the BLM with the predicted vegetation abundance in the next 7-day interval, along with other useful information. It has the following modules.

A spatial visualization module is an essential module in the pheno-DSS. All other functional modules need the cooperation of the spatial visualization module. The spatial visualization module is implemented using the comprehensive, robust and flexible Google Maps API. It provides more intuitive user interactions: users can select locations or regions on the map instead of inputting by hand. It also provides more vivid presentation of the results: different types of data and many layers can be overlaid on Google maps.

A data access module allows a user to display data over Google Maps and download preprocessed eMODIS NDVI and Ecocast climate images on a user selected 7-day interval. Grazing allotments and pastures can be overlaid on the selected image. If an NDVI image is selected to be displayed, users can choose to classify the NDVI values based on the grazing suitability. The layers of grazing allotments and pastures as well as the suitability grades can assist land managers in better understanding the NDVI image and help them in making decisions.

A Time Series module allows users to select either a point, a region, or an allotment or pasture and to display the time series of user-selected variables within a user selected range of dates. This module provides another useful tool for land managers to

arrive at a better grazing strategy by bearing in mind the climatic trends and past phenological cycles.

As the core of the phenoDSS, the prediction module allows the display of the predicted NDVI image for the next 7-day interval generated using the phenomodels developed in this dissertation. As with the data access module, the grazing administration layer can be overlaid and the grazing suitability grades can be calculated to facilitate the land managers' cognition of the vegetation abundance in terms of grazing.

Lastly, a phenological inventory is provided as a useful resource. This is a listing of geospatial and scientific data useful in vegetation (phenology) monitoring and model development in the Upper Colorado River Basin (UCRB). The inventory is designed to provide ecological and hydrological information in support of scientific research and land management. Users can browse or make the specific inquiries to check the relevant records.

4 ANALYSIS AND RESULTS

4.1 Phenoregion Delineation

4.1.1 Principal Component Analysis

Twelve components were generated from the variables listed in Table 3.2 using Principal Component Analysis (PCA). All components were used in the following analysis, and *k*-means clustering were performed in the principal component space constructed by the 12 orthogonal principal components. The component scores imply the dominant variables for each of the PCs, as well as, the correlations between PCs and input variables (Table 4.1). The first PC has the highest score for elevation and all eight climatic variables. It is positively correlated with elevation and three precipitation variables while negatively correlated with five temperature variables. The first PC is moderately correlated with latitude and the mean NDVI. It explained about 61.8% of the total variance by itself. The second PC is highly negatively correlated with latitude. It also represents several of the climatic variables. The soil variability index was highly positively correlated with the third PC while having relatively low correlation coefficients with all other PCs, making the third PC a dominant representation of the soil variability. The fourth through sixth PCs are supplemental in explaining the latitude, the intra-annual climatic variation, and the mean NDVI. The seventh through twelfth PCs are much less correlated with the 12 variables, and these last PCs combined can only explain 2.45% of the total variance.

Table 4.1 Component scores of and variance accounted for by the PCs. (Abbreviations are: LAT, latitude; ELEV, elevation; MEAN_TMAX, mean maximum temperature; MEAN_TMIN, mean minimum temperature; GS_TMAX, mean maximum temperature during growing season; GS_TMIN, mean minimum temperature during growing season; TEMP_STD, standard deviation of monthly temperature; MEAN_PRCP, mean precipitation; GS_PRCP, mean precipitation during growing season; PRCP_STD, standard deviation of monthly precipitation; SVI, soil variability index; MEAN_NDVI, Mean annual NDVI; PC1-PC5, the first to the fifth PCs).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
LAT	**0.383**	**-0.802**	0.034	-0.015	0.243	**-0.385**	0.026	-0.02	0.007	-0.033	0.002	0
ELEV	**0.961**	0.052	0.035	0.019	0.022	0.233	-0.047	-0.098	0.063	-0.044	-0.001	0
MEAN_TMAX	**-0.947**	0.274	-0.047	-0.008	-0.135	-0.018	0.036	0.023	-0.008	-0.069	0.022	-0.001
MEAN_TMIN	**-0.856**	**0.45**	0.044	-0.098	0.197	-0.112	0.025	-0.032	0.014	-0.002	-0.011	-0.005
GS_TMAX	**-0.977**	0.141	-0.044	-0.008	-0.118	-0.082	0.031	-0.002	0.006	-0.042	-0.024	0.003
GS_TMIN	**-0.881**	**0.393**	0.049	-0.098	0.177	-0.133	0.034	-0.072	0.03	0.039	0.012	0.003
TEMP_STD	**-0.69**	**-0.499**	-0.118	0.097	**-0.488**	-0.082	0.041	-0.054	0.031	0.026	0.001	-0.001
MEAN_PRCP	**0.906**	**0.312**	0.02	-0.116	-0.099	-0.129	0.169	0.067	0.089	0.005	0	0
GS_PRCP	**0.92**	0.273	-0.043	0.006	-0.086	-0.041	0.241	-0.058	-0.077	-0.004	-0.001	0
PRCP_STD	**0.74**	**0.323**	0.139	**-0.422**	-0.255	-0.223	-0.186	-0.015	-0.023	-0.003	0	0
SVI	-0.058	0.059	**0.96**	0.253	-0.073	-0.036	0.007	0.001	-0.002	-0.001	0	0
MEAN_NDVI	**0.539**	**0.505**	-0.295	**0.548**	-0.028	-0.232	-0.107	-0.003	0.003	0	0	0
Variance explained	0.618	0.157	0.088	0.049	0.041	0.031	0.012	0.002	0.002	0.001	0.000	0.000
Cumulative Variance explained	0.618	0.775	0.863	0.912	0.952	0.983	0.995	0.997	0.999	1.000	1.000	1.000
Eigenvalue	7.416	1.881	1.056	0.586	0.488	0.369	0.142	0.028	0.020	0.012	0.001	0.000

4.1.2 Phenoregion Maps Generation Using *k*-means++ Clustering

Phenoregion maps with different numbers of clusters (5 to 26) were generated separately following the procedure described in Section 3.1.2. Figure 4.1 illustrates the 5-, 12-, 19-, and 26-phenoregion maps. Visually, the phenoregion maps in Figure 4.1 have similar structures yet with moderate differences. For example, three out of all four maps (the 12-, 19-, and 26-phenoregion maps) delineated the Northern Wyoming Basin (cross sign); all four phenoregion maps delineated the Northern Canyonlands (diamond sign), parks and ranges in northern Utah (triangle sign) and the western White River National Forest in Colorado (donut sign). Phenoregions in one map are not simply the subsets of those in another map with a smaller number of phenoregions because they are not generated as nestable hierarchical clusters. Instead, each pixel was reassigned to a cluster each time the phenoregion map was generated. Mountainous areas tend to be patchier than lower elevations and have more linear shapes following the direction of the elevation contours. A larger number of phenoregions are associated with mountainous areas such as the Southern Rocky Mountains in Colorado (in yellow circles in Figure 4.1) than with flat areas such as the Wyoming Basin (in red circles in Figure 4.1). The patches become smaller and sparser with increased distance from the core area of phenoregions and along the boundaries. This trend is considered to be the representation of gradual instead of abrupt change of phenological forcing in transition areas. These transition areas are called "phenopauses" (first coined by Hargrove and Hoffman (2004) as "ecopauses") (Hargrove and Hoffman, 2004; William et al., 2008). Pixels belonging to the same phenoregions are not necessarily contiguous; instead, they could be distributed

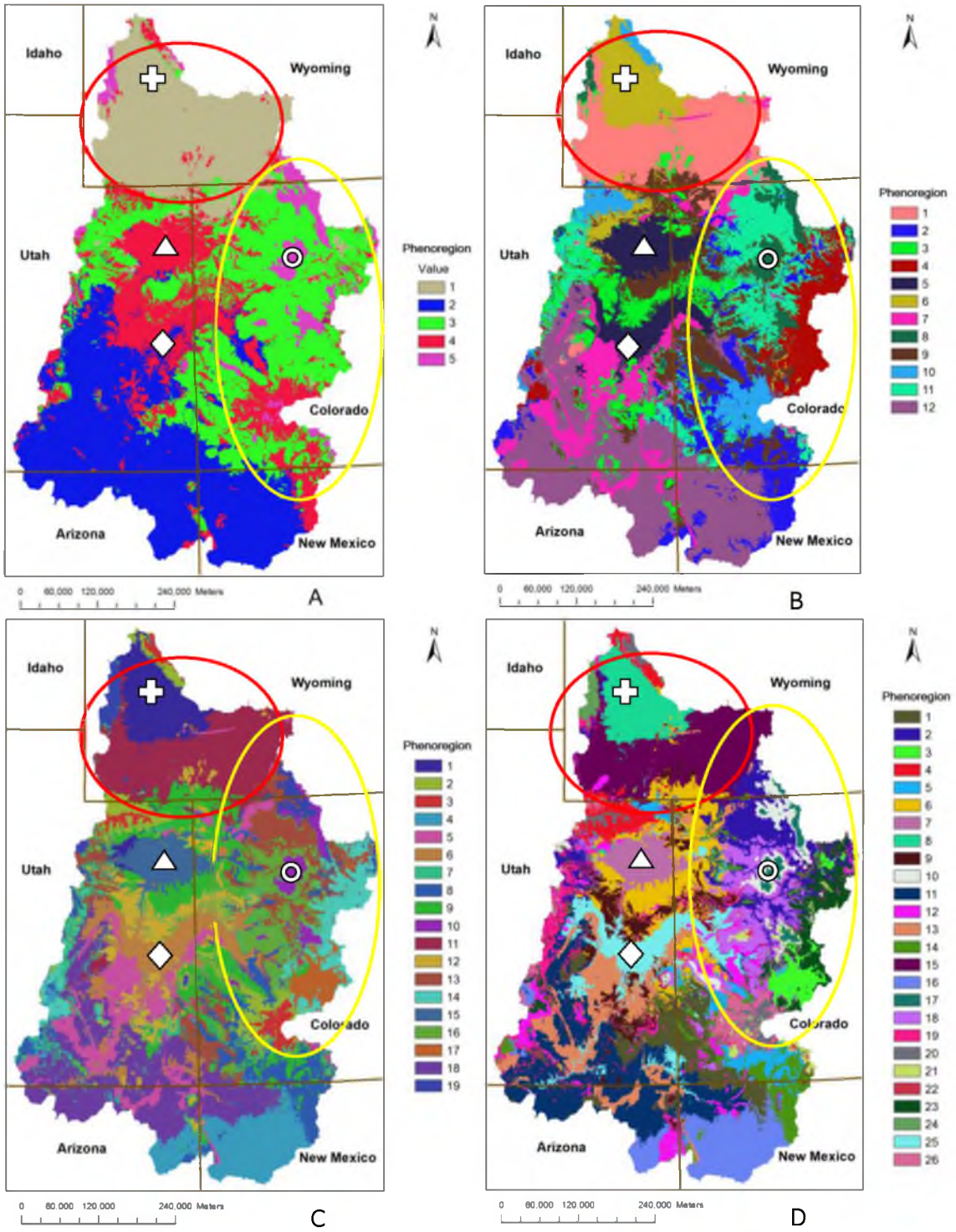


Figure 4.1 5- (A), 12- (B), 19- (C) and 26-phenoregion (D) maps. (Cross sign: Northern Wyoming Basin; diamond sign: Northern Canyonlands; triangle sign: parks and ranges in northern Utah; donut sign: western White River National Forest in Colorado; yellow circle: S).

in either large or small patches far away from each other. This results from similar phenological forcing occurring in different locations within the UCRB.

4.1.3 Selection of Phenoregion Maps

The optimal phenoregion map was selected by absolute and relative comparisons as described in Section 3.1.3. The optimal phenoregion map has higher homogeneity by absolute comparison and spatial concordance with other phenoregion maps by relative comparison.

The total within-cluster sum of squares and mean standard deviation of each phenoregion map are shown in Figure 4.2. Generally, both metrics decrease in a stable manner — the homogeneity becomes higher when the number of phenoregions increases. The MSDs of *k*-means++ clustering from 5- to 14-phenoregion maps decreases, while beyond the 13-phenoregion map, there are some fluctuations. The TWCSS of the *k*-means++ clustering are strictly monotonically decreasing with an increasing number of phenoregions. So judging from the results of absolute comparison, the phenoregion maps with a greater number of phenoregions are generally more favorable.

Figure 4.3 (A) is the matrix of Mapcurves GOF scores between all pairs of phenoregion maps, represented by grayscale values. Brighter tones indicate higher GOF scores. Mapcurves GOF scores between each phenoregion map and itself are always equal to 1, producing the white diagonal from the upper left (5,5) to the lower right (26,26) corner. Except for these perfect fits, most of the phenoregion maps have GOF scores below 0.9, because of the hierarchical yet nonnestable nature of this series of phenoregion maps. A phenoregion map tends to have a higher GOF score when compared to another map with a similar number of clusters. For example, the Mapcurves

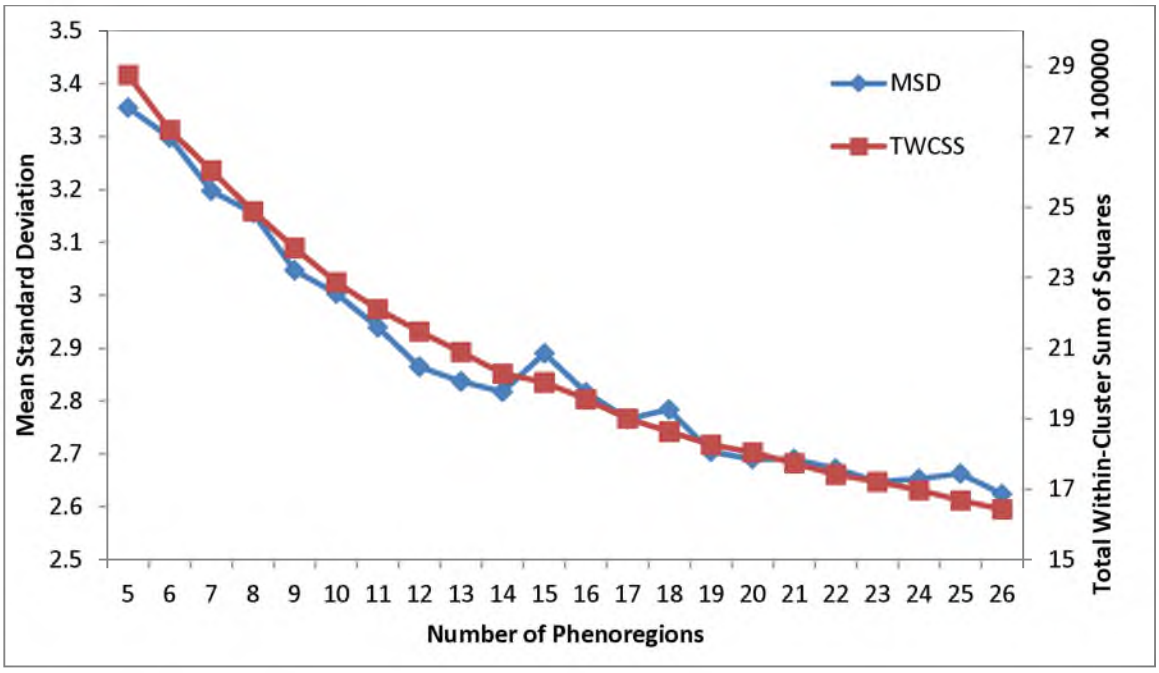


Figure 4.2 Mean standard deviations (MSD) and total within-cluster sum of squares (TWCSS) of phenoregion maps using ordinary *k*-means and *k*-means++ clustering.

GOF score curve of the 15-phenoregion map peaks at 15 clusters and declines on either side of the peak (Figure 4.3 (B)).

The average GOF score shows a general trend that higher scores are associated with intermediate number of clusters. Phenoregion maps with either small (e.g., 5, 6, and 7) or large number of clusters (e.g., 24 and 26) tend to have lower spatial concordance with other maps. Among all phenoregion maps, the 19-phenoregion map has the highest average GOF score (0.781, Table 4.2). The 24-phenoregion map has the lowest average GOF score (0.723). The 19-, 20-, 17-, 16-, 9-, 18-, 22- and 15-phenoregion maps (in decreasing order) have a higher degree of concordance with other maps (Table 4.2), thus are considered to be superior choices for the final phenoregion map.

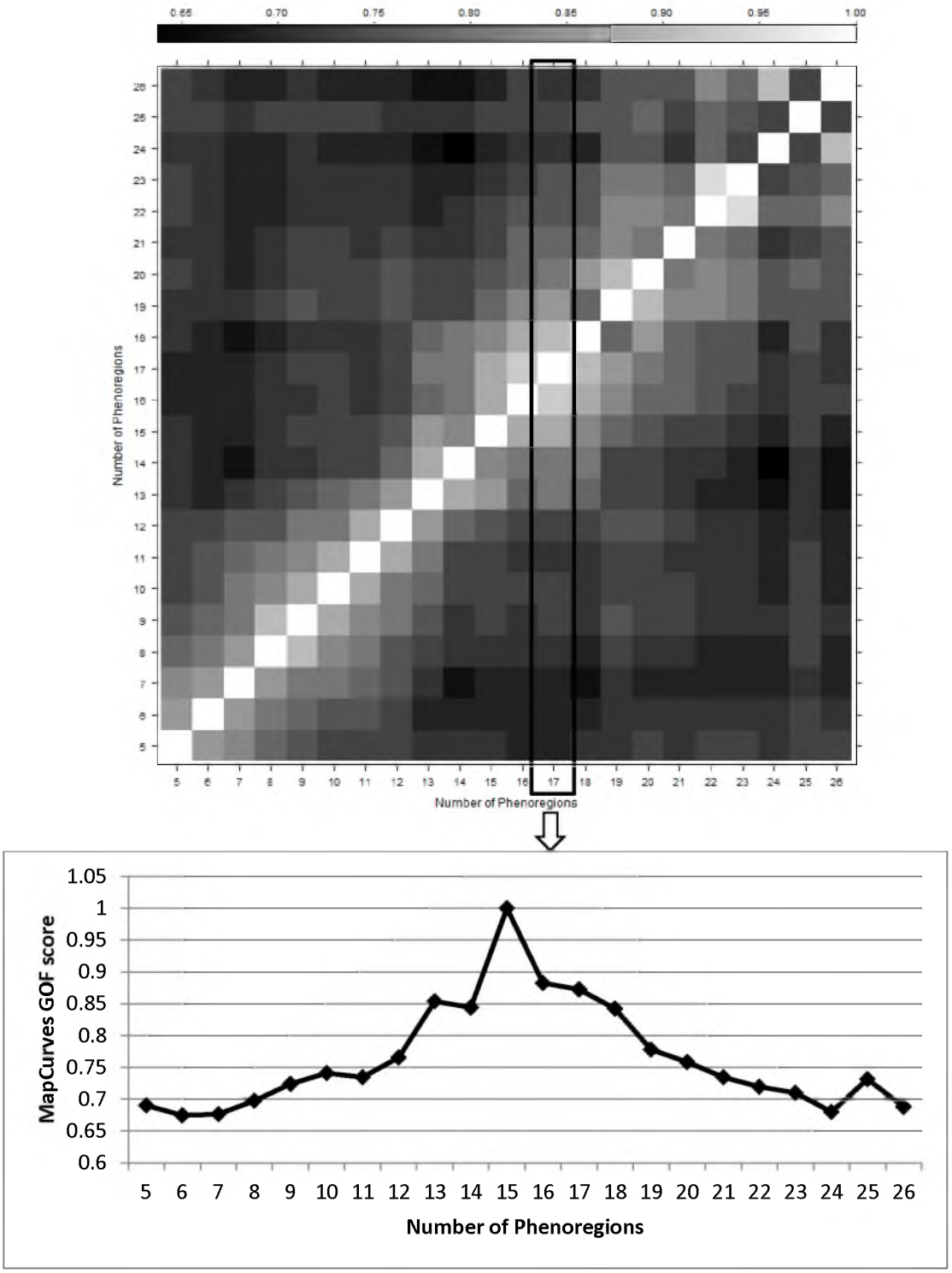


Figure 4.3 (A) Matrix of Mapcurves GOF scores represented by linearly scaled grayscale values with black indicating a score of 0.637 and white indicating 1.00; (B) Mapcurves GOF scores of the 17-phenoregion map.

Table 4.2 Rank of phenoregion maps by average Mapcurves GOF score

Phenoregion number	Average Mapcurves GOF score	Phenoregion number	Average Mapcurves GOF score
19	0.781	13	0.756
20	0.777	21	0.752
17	0.773	23	0.752
16	0.767	8	0.751
9	0.766	25	0.745
18	0.765	14	0.743
22	0.764	5	0.742
15	0.764	6	0.739
12	0.761	7	0.736
11	0.760	26	0.733
10	0.760	24	0.723

Considering the fact that phenological modeling, validation and ground truthing are time- and labor-consuming, a phenoregion map with a smaller number of phenoregions is preferred. Therefore, the nine-phenoregion map is selected to serve as the basic unit of phenomodeling.

4.2 Characteristics of Nine Phenoregions

It is necessary to explore the characteristics of each of the nine phenoregions in the UCRB before the development of the phenomodels.

Figure 4.4 shows the phenoregion map with the nine phenoregions used in this project as the basic areal unit. The characteristics can be summarized through Figure 4.5, Table 4.3 and Table 4.4.

Phenoregions one, five, and eight share similar characteristics. As compared with other phenoregions, the elevations in these three phenoregions are lower. Phenoregions one and five have almost equally high maximum and minimum monthly temperatures

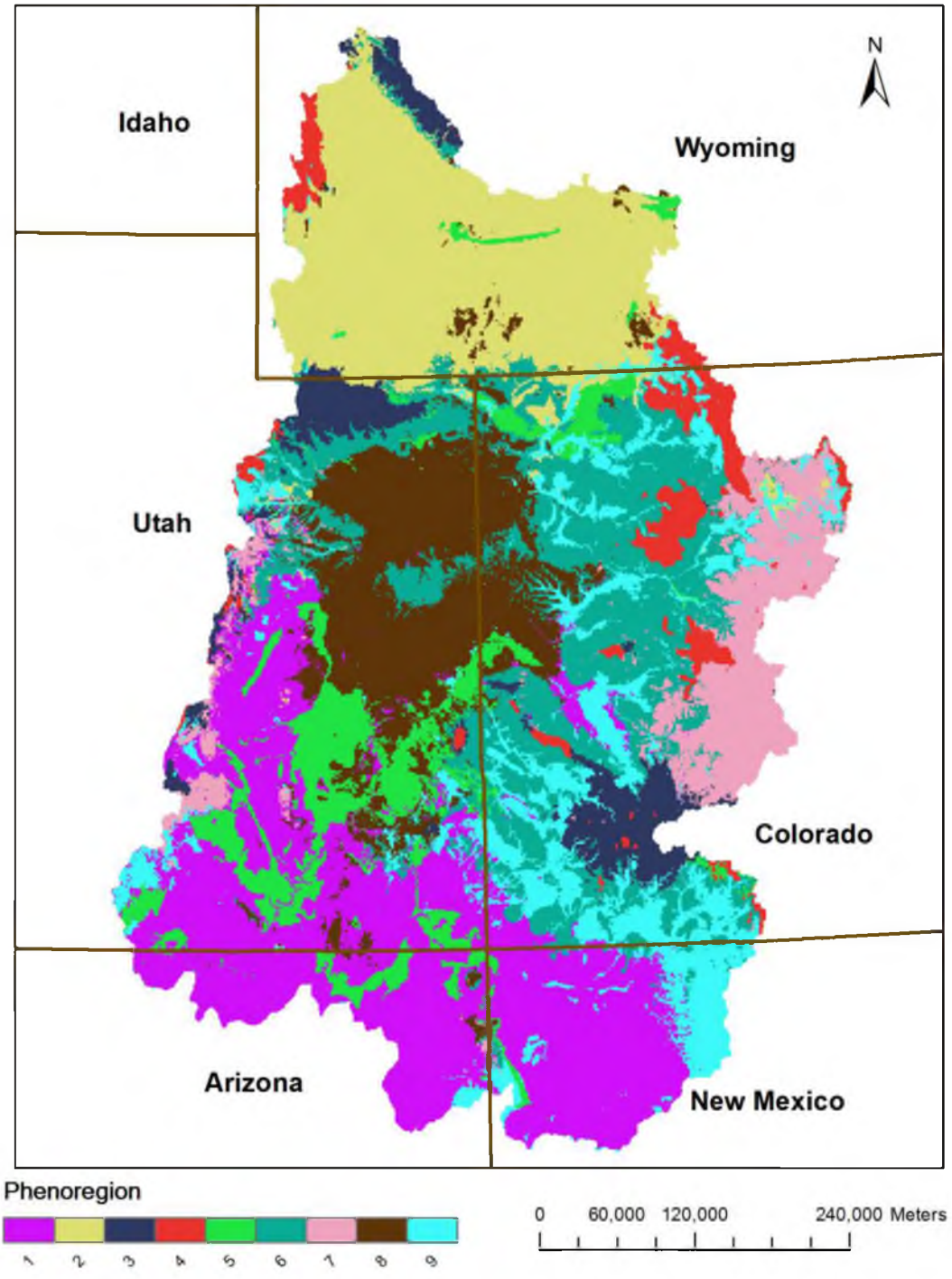


Figure 4.4 The nine-phenoregion map used as the basic unit of phenomodeling.

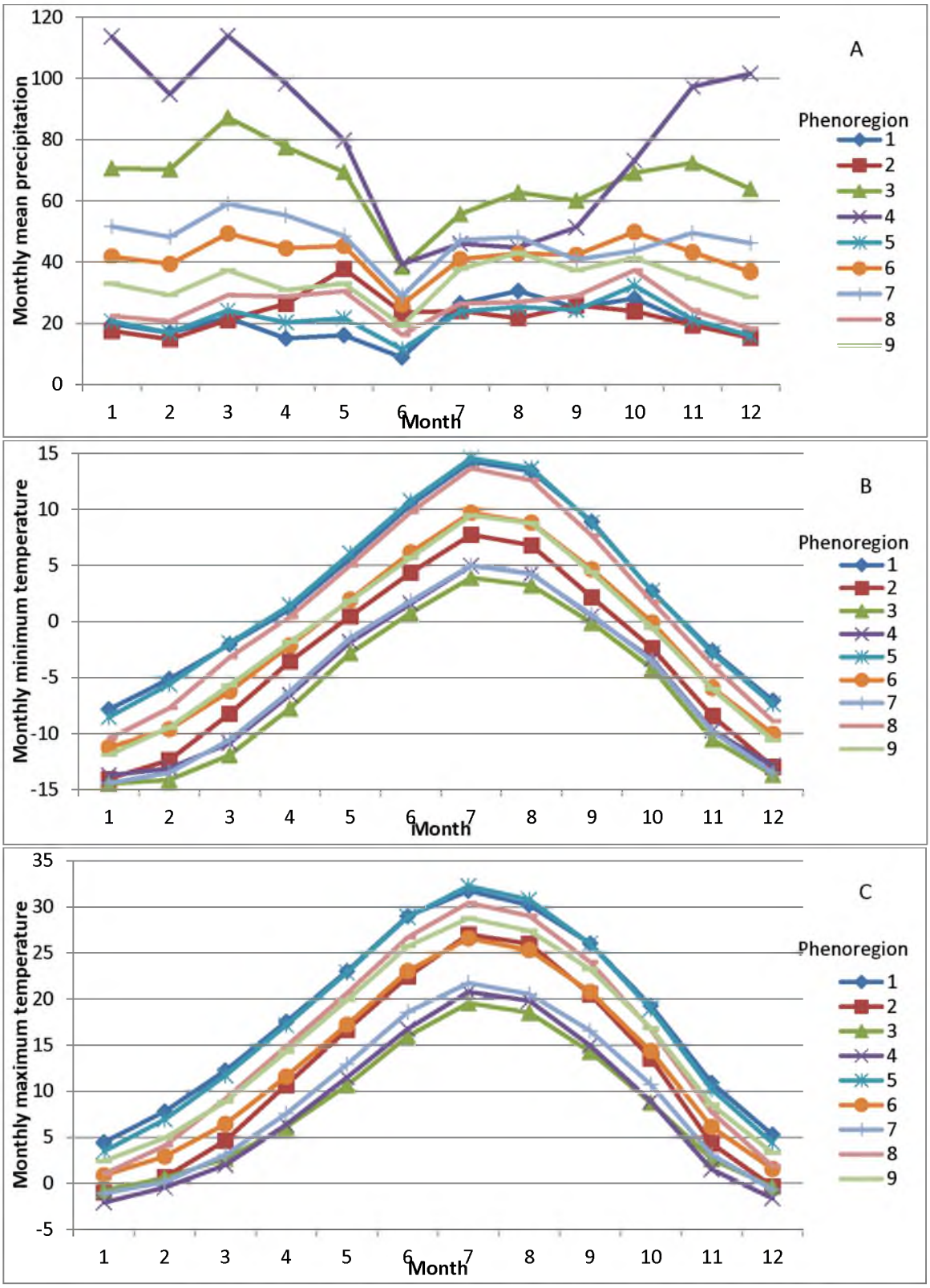


Figure 4.5 Monthly precipitation (A), minimum temperature (B), and maximum temperature (C) in the nine phenoregions.

Table 4.3 Summarization of elevation and climatic conditions in the nine phenoregions

Phenoregion	1	2	3	4	5	6	7	8	9
Mean elevation (m)	1750	2098	3007	2839	1652	2238	2854	1812	2046
Annual cumulative precipitation (mm)	250	277	804	960	264	509	573	316	411
Mean monthly precipitation during growing season (mm)	23	27	60	56	24	42	43	28	36
Mean maximum temperature (°C)	18.6	12.4	8.5	8.5	18.3	13.5	9.7	16.0	15.8
Mean maximum temperature during growing season (°C)	27.0	21.5	15.1	15.9	27.1	21.7	17.3	25.1	24.1
Mean minimum temperature (°C)	2.7	-3.5	-6.2	-5.3	2.7	-1.2	-5.3	1.4	-1.4
Mean minimum temperature during growing season (°C)	9.7	3.5	0.1	1.1	9.9	5.6	1.2	8.9	5.3

Table 4.4 Vegetation composition aggregated from 30-m GAP land cover data

Phenoregion	Vegetation Composition
1	Shrubland: 43.15% Barren: 20.85% Forest: 18.33% Grassland: 10.07%
2	Shrubland: 79.86%
3	Forest: 55.51% Grassland: 16.93%
4	Forest: 63.99% Grassland: 11.86%
5	Shrubland: 48.28% Barren: 31.16% Forest: 13.30%
6	Forest: 49.31% Shrubland: 39.54%
7	Forest: 57.34% Shrubland: 22.52%
8	Shrubland: 54.54% Forest: 22.35% Barren: 10.51%
9	Forest: 42.97% Shrubland: 32.58% Human land use: 14.66%

over the entire year (hot summer and cool winter). The monthly temperatures in phenoregion eight are only slightly lower, but the temperatures in summer are closer to those in phenoregion one and five than the temperatures in winter, indicating a higher intra-annual temperature difference. The three phenoregions are dry throughout the year and the precipitation reaches a minimum in June. It is slightly wetter in late summer and early autumn. The precipitation is also slightly higher in spring in phenoregion eight. All three phenoregions contains almost half shrubland and also a significant amount of barren lands and forestlands.

Phenoregion two is located in the Greater Green River Basin/Wyoming Basin. It is dry and has a warm summer and cold winter. The precipitation is lower in winter and higher in late spring. It is composed predominantly of shrubland.

Phenoregions three, four and seven are located at higher elevation mountains, and contain a large proportion of forestlands. Phenoregions three and four also contain some grassland while phenoregion seven contains some shrubland. As can be obviously observed in Figure 4.5 (A), precipitation is much higher in late fall, winter and early spring in phenoregion four as compared with other phenoregions. Although precipitation is much lower in summer than in other months, it is still higher than that in most of the other phenoregions. Phenoregion three and seven also have higher precipitation in winter and lower precipitation in summer, but the difference is not as high as in phenoregion four.

Phenoregions six and nine have a large amount of both forestland and shrubland on highlands. Precipitation is concentrated in spring and fall, and is lower in summer.

The monthly minimum temperatures are almost the same while the monthly maximum temperature is higher in phenoregion nine.

4.3 Samples and Outlier Exclusion

Spatial sampling was conducted in each phenoregion using stratified systematic plus random sampling as described in Section 3.2.4. The numbers of pixels in pixel sample one and two in each of the nine phenoregions as well as the numbers of pixels identified as nonnaturally-vegetated are listed in Table 4.5 and displayed in Figure 4.6.

The number and percentage of NDVI values of the naturally-vegetated pixels identified as bad quality and snow during eMODIS compositing and as residual outliers by comparing with the reconstructed time series are shown in Table 4.6 for both pixel sample one and two. The numbers and percentages of three kinds of outliers for both samples within the same phenoregion are very close, showing from another perspective the effectiveness of stratified systematic sampling in extracting representative samples, as well as the homogeneity of phenological and environmental characteristics of

Table 4.5 Summary of sampled pixels for modeling and validation

Phenoregion	Number of pixels		Total number of pixels	Percentage of the pixel sample 1 (%)	Standard deviation	Number of nonnaturally-vegetated pixels	
	Pixel sample 1	Pixel sample 2				Pixel sample 1	Pixel sample 2
1	1335	1335	65317	2.0	2.3	242	270
2	1079	1079	48562	2.2	2.5	60	64
3	512	511	13754	3.7	4.2	43	42
4	379	378	10031	3.8	4.2	17	15
5	533	532	22432	2.4	2.7	166	143
6	1218	1217	47382	2.6	2.9	37	43
7	596	596	19865	3.0	3.4	32	19
8	965	965	36772	2.6	2.9	80	85
9	732	731	29774	2.5	2.8	126	108

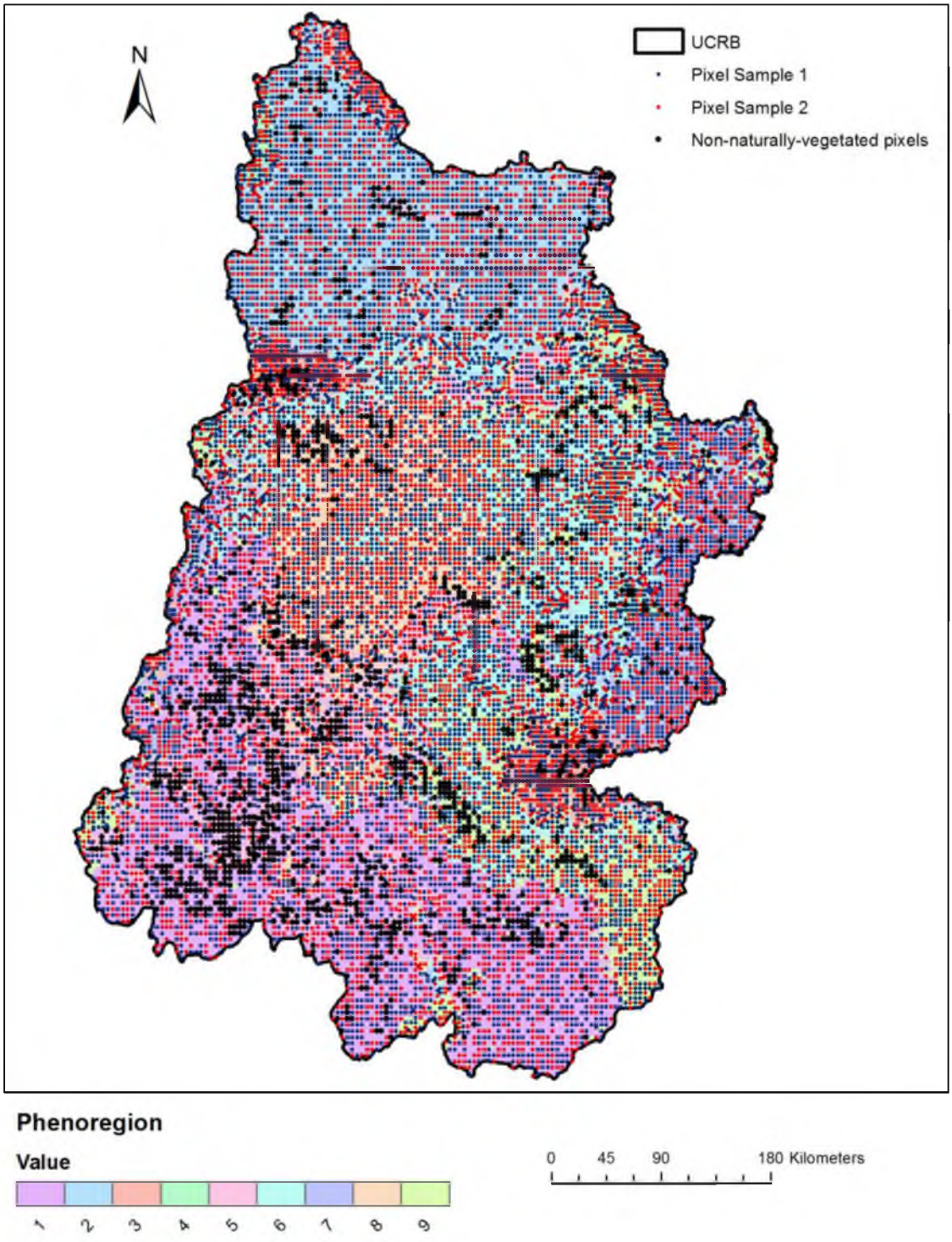


Figure 4.6 Sampled pixels for both modeling and validation: small blue dots are pixel sample one, small red dots are pixel sample two, and bigger black dots are the nonnaturally-vegetated pixels (See Figure 1.1 for the location of UCRB).

Table 4.6 Number of NDVI outlier values in sampled pixels in pixel sample one and two after the removal of nonnaturally-vegetated pixels

Phenoregion		Pixel Sample 1			Pixel Sample 2		
		Bad Quality	Snow	Residual noise	Bad Quality	Snow	Residual noise
1	Count	39746	16332	26388	38866	16564	25747
	Percent	6.72%	2.76%	4.46%	6.75%	2.87%	4.47%
2	Count	52467	62985	31166	52821	61512	31194
	Percent	9.52%	11.43%	5.65%	9.62%	11.20%	5.68%
3	Count	28760	47887	20343	29906	48795	20120
	Percent	11.33%	18.87%	8.02%	11.79%	19.23%	7.93%
4	Count	26682	35234	16741	27153	35329	17040
	Percent	13.62%	17.99%	8.55%	13.83%	17.99%	8.68%
5	Count	14676	8177	9978	15900	8217	10295
	Percent	7.39%	4.12%	5.03%	7.56%	3.90%	4.89%
6	Count	68732	69638	38169	68280	68668	37834
	Percent	10.76%	10.90%	5.97%	10.75%	10.81%	5.96%
7	Count	33173	47659	22143	33373	49349	22649
	Percent	10.87%	15.62%	7.26%	10.69%	15.81%	7.26%
8	Count	43886	33024	23792	43254	33310	23106
	Percent	9.17%	6.90%	4.97%	9.09%	7.00%	4.85%
9	Count	29270	26311	17252	30252	27719	17627
	Percent	8.93%	8.03%	5.26%	8.98%	8.22%	5.23%

phenoregions. On the contrary, the numbers and percentages across phenoregions differ significantly.

Phenoregions three and four have much higher percentage of inconsistent, snowy and noisy (contaminated) NDVI values, due to the constantly higher precipitation across the year as shown in Figure 4.5 (A). Phenoregions one and five, as the two phenoregions with the lowest cumulative precipitation and precipitation during growing season, naturally have the lowest number and percentage of all three kinds of outliers.

4.4 Phenological Cycles

Figure 4.7 displays the phenological cycles as the reconstructed phenoregional mean NDVI time series of pixel sample one and two spaced 0.1 unit apart for visual clarity. The two time series within the same phenoregion are almost exactly the same.

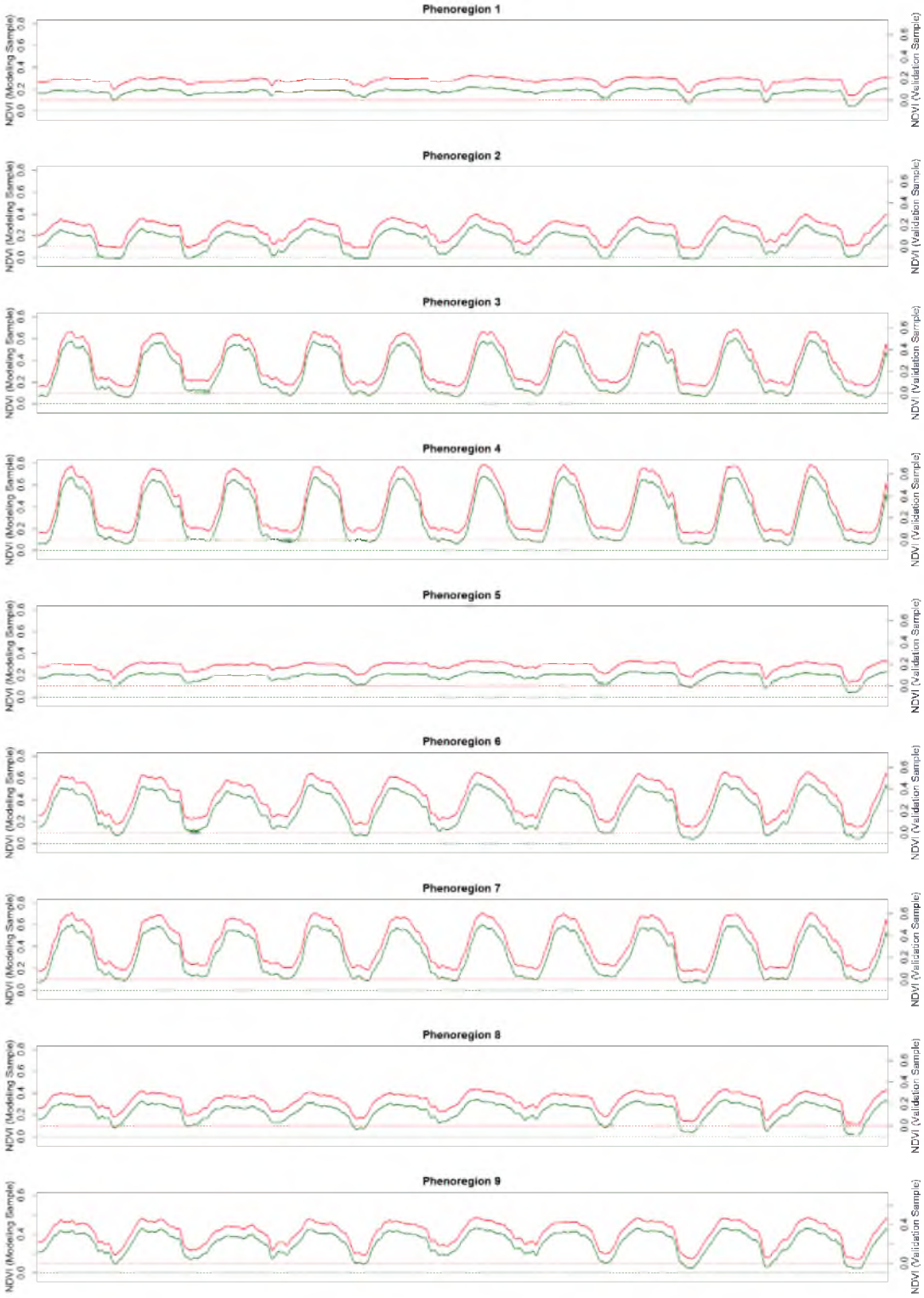


Figure 4.7 Phenoregional mean NDVI time series of pixel sample one (in green) and pixel sample two (in red) spaced 0.1 unit apart for visual clarity.

This further demonstrates the homogeneity of phenoregions and the representativeness of both sets of sample.

Phenoregions one and five show little seasonality: NDVI values remain stable at around 0.2 across multiple years. Phenoregion eight, considered to be similar to phenoregions one and five in the ranges and patterns of environmental variables, shows some seasonality though. This phenomenon of little or some seasonality can be attributed to the high proportion of shrubland and barren land, as well as, the possible vegetation sparsity caused by very limited precipitation during growing season. Phenoregions one and five have the lowest mean monthly precipitation during growing season, only 23mm and 24mm, respectively. The proportion of barren land greatly negatively influences the seasonality, as phenoregion one and five both contain more than 20% of barren land and thus show little seasonality; phenoregion eight shows some seasonality because it only contains around 10.51% of barren lands.

Phenoregion two and nine exhibit a higher degree of seasonality. Phenoregion two contains predominantly shrubland, and thereby has lower values of NDVI ranging from about 0 to 0.2 annually (NDVI values are as low as 0 because snow covered pixels represents the actual land surface and thereby are not removed until before modeling). NDVI values in phenoregion nine have a higher range of about 0.2 to 0.4.

Other phenoregions (three, four, six, and seven) have very obvious seasonality. Phenoregion four has the widest range of NDVI with the lows around 0.1 and highs over 0.6, as it contains the highest proportion of forest.

The annual phenological cycles exhibit different patterns in the nine phenoregions (Figure 4.8). The onset dates of four phenophases are shown in Table 4.7

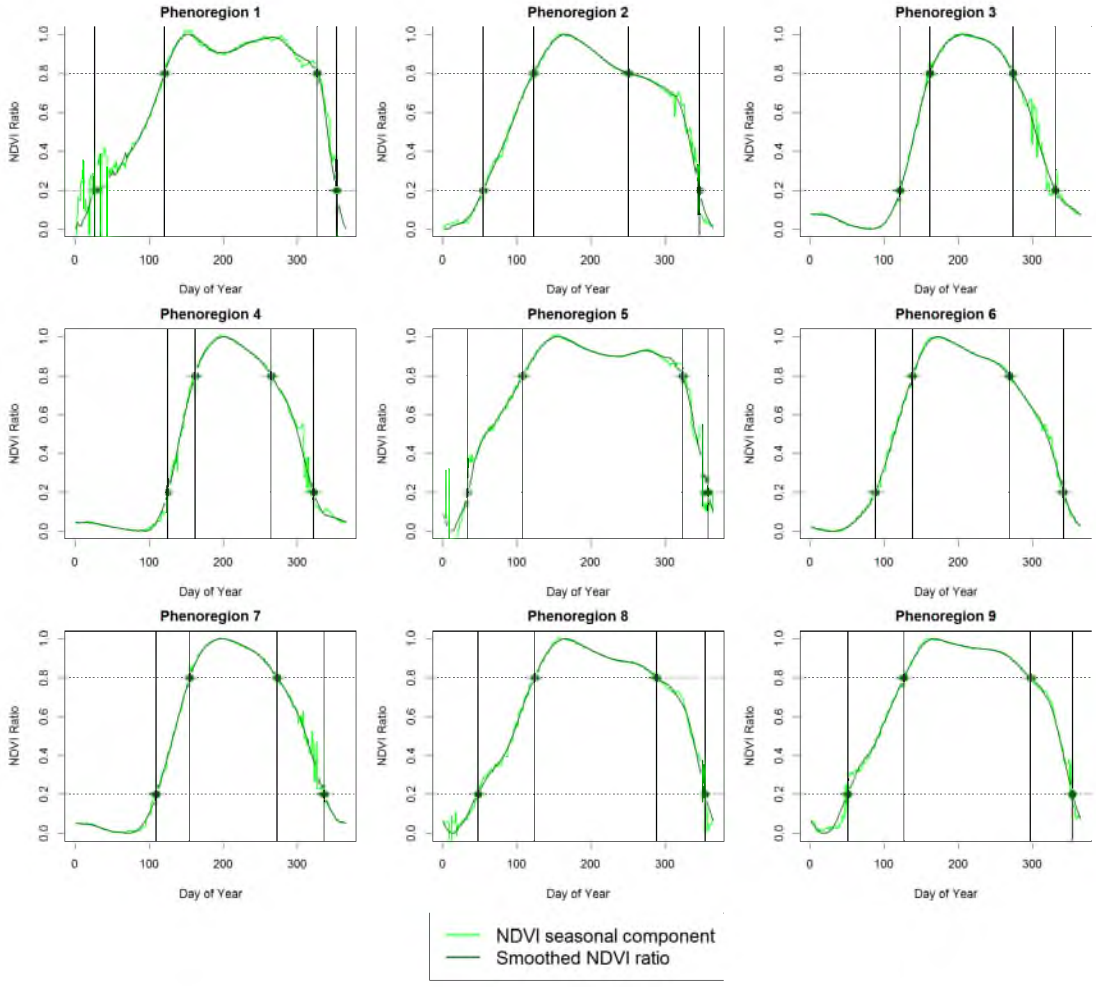


Figure 4.8 Mean annual NDVI time series and phenophases of pixel sample one.

using NDVI ratios and a threshold of 20% and 80% as described in Section 3.3.3. Typical phenological cycles include the bimodal NDVI time series such as phenoregions one and five, pointed peak such as phenoregion four and flat peak such as phenoregion three. It is common and normal that the annual phenological cycles in some phenoregions are composites of several patterns. For example, phenoregion nine shows the combined feature of the bi-modal, pointed peak (the first peak) and the flat peak (the maturity phenophase).

Table 4.7 Phenophase onset in day of year

Phenoregions	Greenup	Maturity	Senescence	Dormancy
1	27	121	327	353
2	55	123	251	347
3	121	162	274	331
4	125	162	265	322
5	34	108	324	358
6	88	138	269	342
7	110	155	273	336
8	48	124	289	354
9	51	127	297	354

4.5 Premodeling Analysis

It is not realistic to present and analyze the results of premodeling analysis and modeling procedure for all phenoregions. Phenoregion one is used as the example case in the premodeling analysis and the modeling procedure. All of the analyses and results presented below are for phenoregion one if not specified otherwise. Phenoregion one contains about 43% shrubland and 10% grassland, and therefore can provide insights into the rangeland phenology and lagged responses of rangeland vegetation to the environmental changes.

4.5.1 Pairwise Correlation Analysis

For phenoregion one, independent variables sharing similar physical meaning naturally are extremely highly correlated, such as the TMEAN, TMAX, TMIN, GDD and GDDu at the same and neighboring lags. TMEAN, TMAX, TMIN, GDD and GDDu at a certain lag are also highly correlated with PTPD, SRE and SRAD after 2 to 5 weeks. There are few causal relationships in these correlations. The high correlations between temperature and the subsequent light condition are most probably the result of their

similar trends over time: the increase of temperature is usually accompanied by the increase of photoperiod, solar radiant energy and solar radiation. TDIFF is most highly correlated with its contemporaneous SRAD, followed by the contemporaneous TMAX and SRE, though the correlation coefficient (0.5~0.8) is lower. The close relationship between weekly temperature difference and light is because the increase of air temperature is largely due to the absorption of radiant energy from the sun. AGDD and AGDDu, as cumulative temperatures, are most highly correlated with themselves at different lags with most of the correlation coefficients greater than 0.8. AGDD and AGDDu are also moderately correlated with different forms of weekly temperatures (TMAX,TMEAN, TMIN, GDD and GDDu) 4 to 5 weeks later.

PTPD, SRAD and SRE are most highly correlated with themselves (light variables) at the same and neighboring lags. Also, as mentioned in the last paragraph, they are also highly correlated with the precedent weekly temperature variables

On the contrary, PRCP is only moderately or weekly correlated with other variables and itself at different lags. The pairwise correlations between PRCP at different lags are only as high as 0.25. The highest correlation of PRCP at a certain lag usually occurs at the two neighboring lags. This is because precipitation is composed of discrete events, unlike temperature and light, whose change is gradual and continuous. The PRCP is moderately negatively correlated with its contemporaneous TDIFF with the correlation coefficient ranging from 0.42 to 0.48. This correlation can be attributed to the complicated relationships between air temperature and precipitation. Precipitation falls from clouds which blocks the sun, greatly decreases the amount of solar radiant energy the air can absorb to increase the temperature (Dai et al., 1999). When precipitation

occurs during daytime, it is usually accompanied with lowered maximum temperature: the air needs to be cooled to its dew point to allow precipitation to develop; precipitation also lowers the air temperature through evaporative cooling.

SLOPE is moderately negatively correlated with TEMP_STD (correlation coefficient being 0.42), indicating that steeper slopes tend to have lower intra-annual variability of temperature. ELEV is highly positively correlated with GS_PRCP and PRCP_STD and moderately positively correlated with MEAN_NDVI, indicating more abundant precipitation and higher intra-annual variability of precipitation at higher elevations. ELEV is also highly negatively correlated with MEAN_TMIN, GS_TMIN, MEAN_TMAX, and GS_TMAX. This is the temperature gradient that can be observed with increased elevation. ELEV is also moderately negatively correlated with TEMP_STD, denoting a mild trend of lower intra-annual variability of temperature at higher elevations. PRCP_STD is positively correlated with ELEV, GS_PRCP, MEAN_PRCP, and negatively correlated with the five temperature environmental factors (MEAN_TMIN, GS_TMIN, MEAN_TMAX, GS_TMAX, and TEMP_STD) as well as LAT. SVI is not as highly correlated with other environmental drivers and factors. It is only weakly positively correlated with MEAN_TMIN and GS_TMIN, and weakly negatively correlated with LAT, TEMP_STD and GS_PRCP.

4.5.2 Principal Component Analysis

The PCA analysis was applied on the 144 continuous independent variables of phenoregion one, and the results show that: among all Principal Components (PCs), the first PC accounts for almost half (46.7%) of the variance in the set of all independent variables. The first PC is almost equally composed of weekly temperature variables

(TMAX, TMIN, TMEAN, GDD, and GDDu), as well as light variables at longer lags (lags of 5 to 10 weeks). The second PC accounts for about another 19.8% of the total variance. This PC is constituted mainly of AGDD, AGDDu and all three light variables at shorter lags (lags of 1 to 4 weeks). These can be understood as that temperature and light variables are the most important in accounting for the environmental variance. All other 142 PCs altogether account for the remaining 33.5% of the total variance, with the highest variance accounted for by a single PC being only 5%.

4.5.3 Lag Structure

The lag structure of NDVI shows apparent difference among phenoregions (Figure 4.9). NDVI in phenoregions one, five and eight has very high persistence, and the correlation is as high as about 0.6 even at the lag of as long as 25 weeks. This is possibly due to the vegetation composition and climatic conditions in these phenoregions, which lead to a much smaller amplitude of NDVI throughout the year.

Phenoregions two, six, and nine have slightly more NDVI instability as compared with phenoregions one, five and eight, though they still exhibit relative small vegetation dynamics.

The difference between the NDVI lag structure of phenoregions three, four, and seven and other phenoregions is quite big. These three phenoregions show very active vegetation dynamics. NDVI is no longer correlated with its antecedent values when the lag is more than 14 weeks, 15 weeks and 17 weeks in phenoregion four, three and seven, respectively.

The difference between lag structures of temperature variables is not as big for different phenoregions. The shape of the lag structures of TMEAN, TMAX, TMIN, GDD

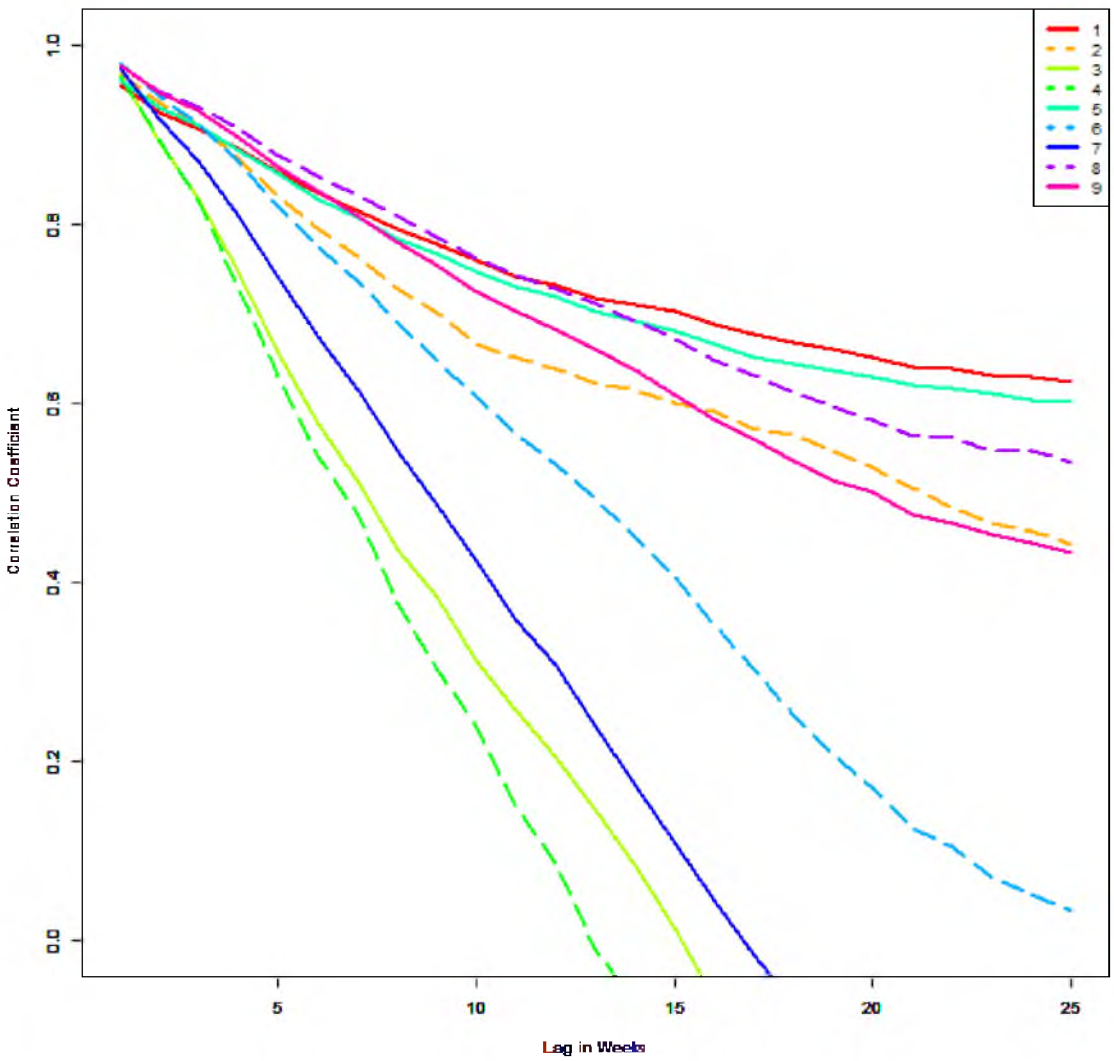


Figure 4.9 NDVI lag structures in phenoregions one to nine.

and GDDu are half parabola with the vertex on the top (Figure 4.10), that is, the correlation coefficient decreases much slower at smaller lags and faster at longer lags. The difference between phenoregions is slightly bigger for GDD and GDDu as compared with other three variables. TMEAN, TMAX and TMIN no longer correlate with their respective lagged values at about 14 weeks in all phenoregions. The pattern shown in the lag structures of GDD and GDDu is roughly consistent with those of NDVI in terms of

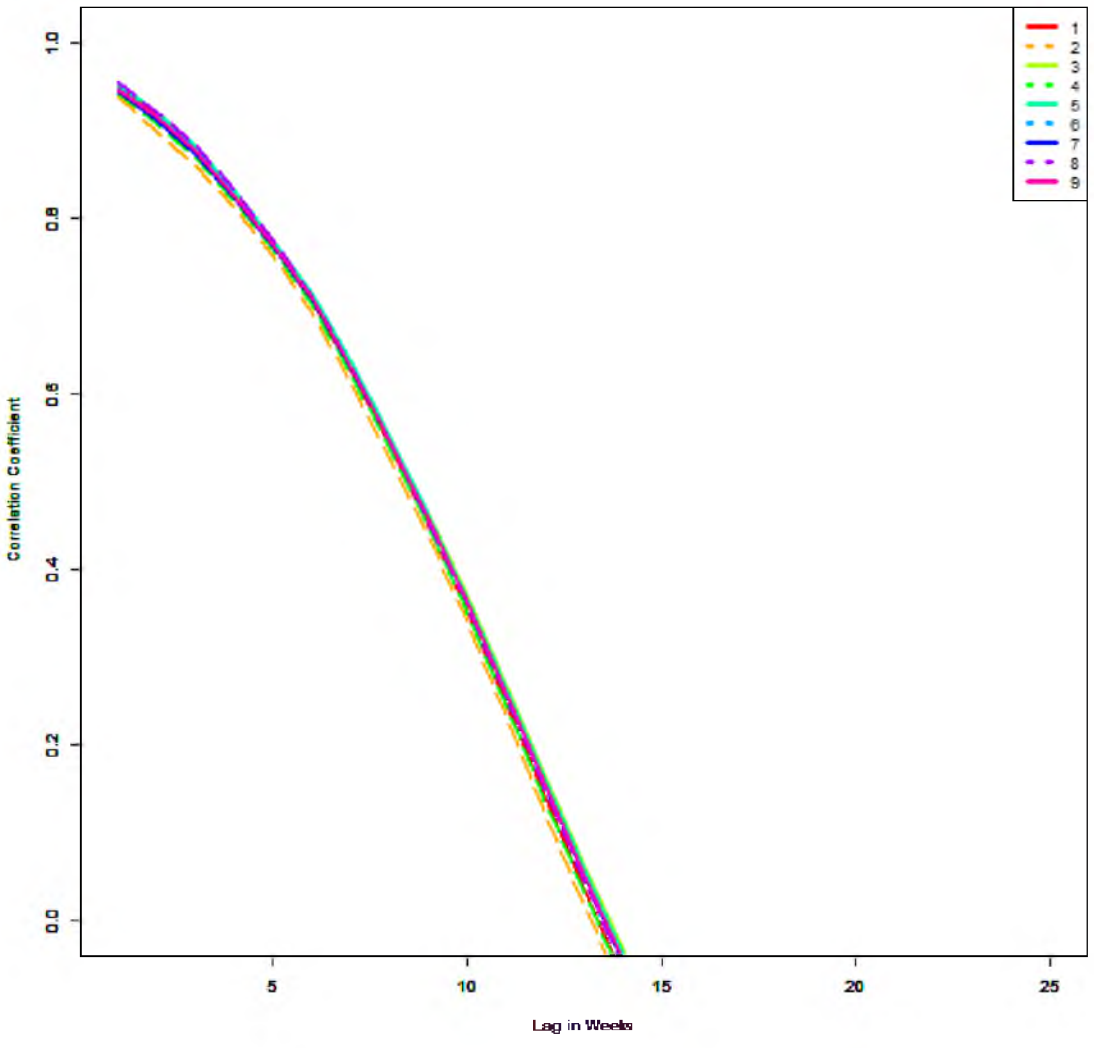


Figure 4.10 TMAX lag structure in phenoregions one to nine.

the variable persistence. The correlation decreases fastest in phenoregion four and slowest in phenoregion five. TDIFF shows a very different pattern. The correlation starts at a lower value and reaches 0 at longer lags. The lag structure of AGDD and AGDDu exhibits a ear-linear decrease pattern. The shape of the lag structure of light variables is very similar to that of TMEAN, TMAX, TMIN, GDD and GDDu. There are also no obvious differences among phenoregions.

Very high instability and low correlation coefficient are associated with the lag structure of PRCP (Figure 4.11), because precipitation consists of discrete events.

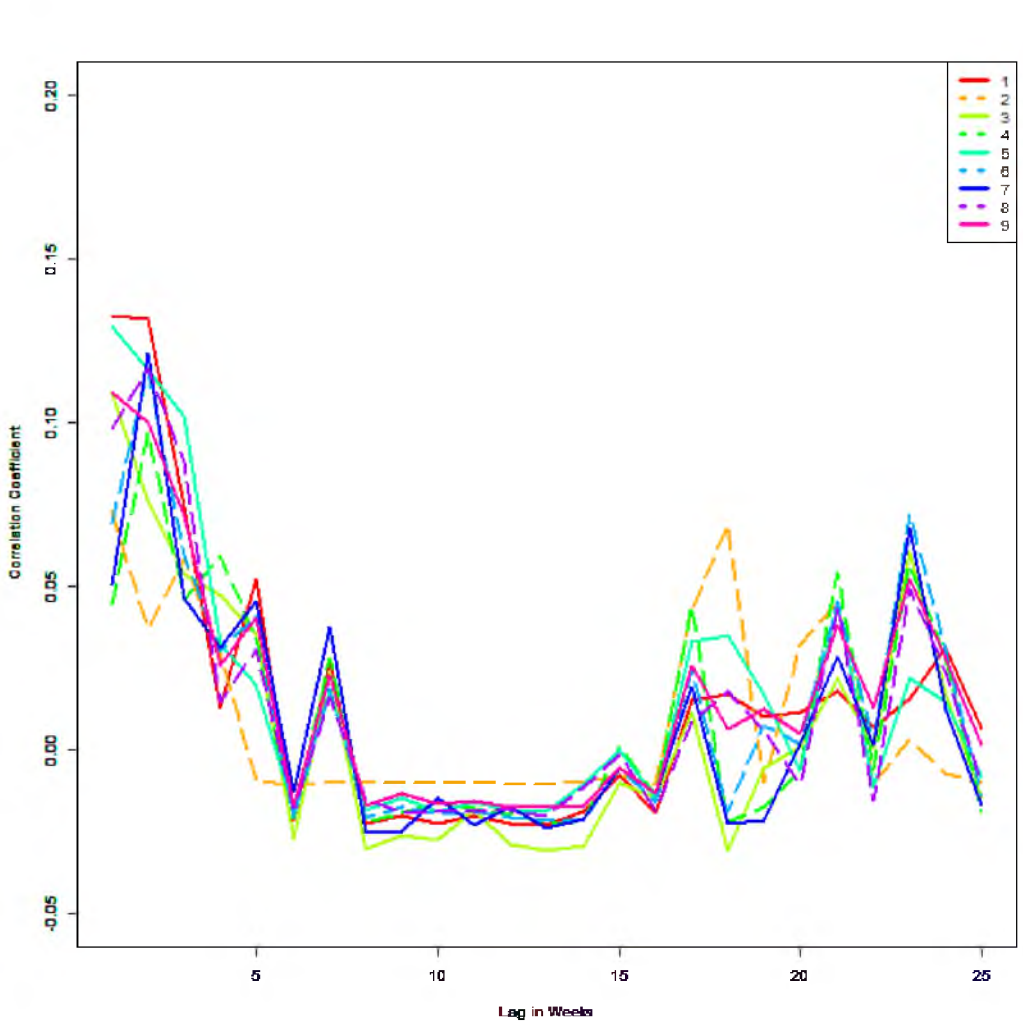


Figure 4.11 PRCP lag structure in phenoregions one to nine.

For phenoregion three, after the VIF based variable reduction, TMIN at lags of 3, 6, and 10 weeks, GDD, GDDu at lags of 2, 4, 5, 7, and 9 weeks and AGDD_10 are the remaining temperature variables. SRAD at lags of 2, 6, 8 and 10 weeks are kept as light variables; All precipitation variables at lags 1 through 20 weeks are kept as precipitation variables; ELEV, GS_PRCP, GS_TMIN, LAT, MEAN_NDVI, PRCP_STD, SLOPE and SVI are kept as time invariant variables to account for the interpixel difference.

4.6 Model Development

The MARS approach was adopted in this research to build phenological models for each phenoregion. MARS was chosen due to its ability to deal with high dimensional independent variables, to model different relationships in different data intervals, and to automatically model interactions. MARS models were built using the R package "earth." Consistently, phenoregion one was selected to illustrate the modeling process. After the iterative variable removal process, the remaining variables having obviously reduced multicollinearity are fed into MARS.

At first, all variables are restricted to enter linearly, which ended up with a multiple linear stepwise regression model. In this linear version of the model, AGDD, TMIN, and TMIN_4 in the descending order of importance entered the model as environmental drivers. These three variables belong to two different types in terms of their physical meanings. TMIN and TMIN_4 are the weekly temperature at the lag of 1 and 4 weeks, in the form of minimum temperature. AGDD is the accumulated temperature at the lag of 1 week. MEAN_NDVI and MEAN_PRCP entered the model as the pixel specific spatial adjustments. The model has the R^2 of 91.39%, GCV of 0.000293 and standard error of 0.0171 (Table 4.8).

Table 4.8 Four measurements of model performance (R^2, GCV, standard error and residual autocorrelation) for the three phenomodels (MARS 1, 2 and 3) in phenoregion one

Phenomodels	R^2	GCV	Standard Error	Residual Autocorrelation
MARS 1	91.39%	0.000293	0.0171	-0.178
MARS 2	91.44%	0.000291	0.0171	-0.168
MARS 3	91.61%	0.000286	0.169	-0.166

By allowing the relationship to bend at different knots, a new MARS model was built, with no interaction effects of variables. This second model (MARS 2) yielded an R^2 of 91.44%, GCV of 0.000291, and standard error of 0.0171. These model parameter values indicate that MARS 2 provides a slightly better performance as compared to the linear stepwise model (MARS 1). Besides the five independent variables that MARS 1 included, an extra independent variable (PRCP_10) entered MARS 2. PRCP_10 ranks before TMIN and TMIN_4, indicating that as compared with MARS 1, MARS 2 identified the importance of precipitation over weekly temperature.

The third model (MARS 3) building on the basis of MARS 2, allows for a degree two interaction (product of two variables). The R^2 of the third model increased to 91.61%, the GCV decreased to 0.000286, and the standard error further decreased to 0.169. The three statistics indicate that MARS 3 performs even better when the interaction terms are considered. The variables of AGDD, PRCP, PRCP_10, PRCP_18, AGDDu_5, TMIN, and TMIN_4 are the environmental drivers of MARS 3. The variables of MEAN_NDVI, LAT, ELEV are the environmental factors included in MARS 3.

The correlation between the residuals and lagged residuals is -0.178, 0.168, and 0.166 for MARS 1, MARS 2 and MARS 3, respectively, indicating only minor residual

autocorrelation in all three models and a decreased residual autocorrelation from MARS 1 to MARS 3.

The performance of three models suggested by the measures of R^2, GCV and standard error is that MARS 3 has the best performance and MARS 1 has the worst. MARS 3 has the highest R^2 value, indicating the highest amount of the variation in the dependent variable can be explained by the independent variables. The lowest GCV of MARS 3 means that it is the optimal model considering both goodness of fit and the model complexity. The lowest standard error indicates that MARS 3 has the highest prediction accuracy. Also, based on the importance of precipitation on plant development summarized in Section 2.2.1.2, the inclusion of more precipitation variables and their greater importance over weekly temperature variables in MARS 2 and MARS 3 makes more sense. This further demonstrates the superiority of MARS over linear regression by approximating nonlinear relationships and automatically modeling interactions, in that the vegetation dynamics depicted in MARS models are closer to the actual vegetation dynamics in the physical environment.

4.7 Modeling Results

The MARS models with two degree interaction terms are built separately for nine phenoregions. All nine MARS models have good performance with a range of the largest R^2 of 97.22% in phenoregion six and the lowest R^2 of 91.61% in phenoregion one. Table 4.9 shows the values of R^2 and standard error of the phenomodels as well as the numbers of cases used to build the phenomodels in the nine phenoregions. Appendix B can be referred to for the complete results. Relationships between vegetation and environment can be inferred by interpreting these models.

Table 4.9 The R^2 and standard error of phenomodels and the number of cases used to build the phenomodels in the nine phenoregions

Phenoregion	R^2	Standard Error	Number of cases
1	91.61%	0.0169	85643
2	95.24%	0.0185	85948
3	95.40%	0.0351	14922
4	96.13%	0.0335	13742
5	93.53%	0.0174	31348
6	97.22%	0.0281	80060
7	96.23%	0.0339	23960
8	95.95%	0.0222	69174
9	96.65%	0.0261	53111

4.7.1 Interpretation of Relationships Modeled by MARS

The environment-vegetation relationships can be interpreted from both MARS model specification and the graphic representation of MARS models. The interpretation from the latter one is much easier. The graphic relationships of the MARS model in phenoregion one (Figure 4.12, see Appendix B for the graphic relationships in other phenoregions) is used as an example.

There are two types of plots. The first type represents the main effects of aNDVI, environmental drivers and environmental factors, in the form of connected segments (the first four plots in Figure 4.12). The other type represents the influence of the interaction terms on the vegetation dynamics, or the effect moderated by other variables. This type of plot appears in the form of 3-D surfaces (the lower seven plots in Figure 4.12).

Phenoregion one has four main effect plots and seven interaction term plots. One plot from each type of plot is selected to give an example of interpretation. The fourth main effect plot depicts the relationship with NDVI and PRCP_10. There are two

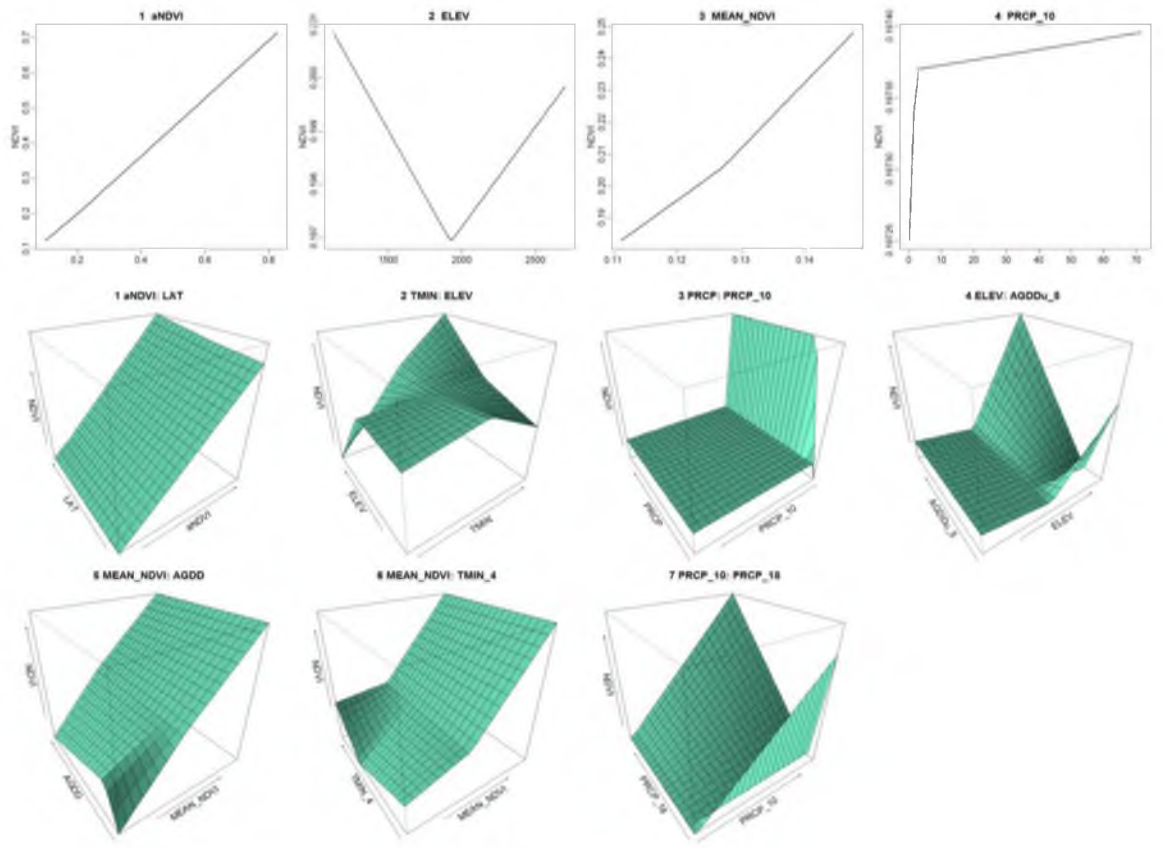


Figure 4.12 The graphic relationships of the MARS model in phenoregion one.

line segments approximating the nonlinear relationship between NDVI and PRCP_10, separated by the knots located at PRCP_10 values of 1.9 mm (Figure 4.12). NDVI increases very fast when the PRCP_10 is below 1.9 mm, and increases much slower at PRCP_10 values beyond 1.9 mm. The second interaction plot shows how the TMIN-NDVI relationship is moderated by elevation. The TMIN-NDVI relationship differs slightly at different elevations. When elevation is below 1925 m, increase in TMIN leads to the increase of NDVI when TMIN is below 2.1 °C and leads to faster decrease of NDVI when TMIN is above 2.1 °C. When elevation is above 1925 m, the increase in

TMIN always results in an increase in NDVI. However, the same amount of increase in TMIN leads to slighly faster vegetation growth when TMIN is less than 3 °C.

4.7.2 Influence of Environmental Variables on Vegetation Dynamics

The temporally variant environmental drivers (temperature, precipitation, and light) directly drives vegetation growth. Environmental factors account for the spatial and temporal differences of the environment-vegetation relationships modeled in the phenomodels, i.e., the differences among different phenophases and among different locations. Based on the variable subset selection rationale of MARS, environmental drivers and factors included in the model are deemed more important than the ones left out of the model within the reduced set of variables that were fed into MARS. However, the variables that are removed beforehand due to high VIF values should not be considered unimportant. The importance of these variables should be analyzed separately based on the pairwise correlation, lag structure and PCA results.

4.7.2.1 Temperature

Three categories of temperature variables at lags of 1 to 10 weeks were analyzed by MARS in terms of their influence on the 7-day vegetation dynamics. The first category is the weekly temperatures, including TMAX, TMIN, TMEAN, GDD and GDDu. The second category is the weekly temperature difference, including TDIFF. The third one is the accumulated temperatures, including AGDD and AGDDu.

Temperature variables are generally included in the models of all nine phenoregions. The first category of weekly temperatures appears in all phenoregions except for phenoregion eight. The third category of temperature variables, the AGDD and

AGDDu at different lags are included in all nine phenoregions, and are of great importance according to the variable importance ranks shown in Appendix B. This indicates that the accumulated temperature is very important in accounting for the 7-day vegetation dynamics.

The temperature variables appear as the main effect terms in only phenoregions of three, four, six and seven where the mean temperature is relatively lower. Phenoregion one is the only phenoregion that does not include temperature variables at longer lags of week 5 to 10. Phenoregions three, four, five, six and seven also contain temperature variables at shorter lags. Phenoregions two, eight, and nine contain only temperature variables at longer lags. Although some of the variables at either shorter or longer lags are included in the phenoregions, they are considered less important, such as TMIN_3 in phenoregion three, and TMIN and GDDu_4 in phenoregion four. From this perspective, phenoregions one, five, and six contain only temperature variables at shorter lags in their more important set of independent variables.

The influence of temperature variables is moderated by aNDVI in eight out of all nine phenoregions. Phenoregion one is the only exception. Except for phenoregion three, five, and six, the temperature-vegetation relationships are affected by MEAN_NDVI in the other phenoregions. The moderation of long term mean temperature is in effect in phenoregion four, six and nine, while the long term mean precipitation affects the temperature-vegetation relationships in phenoregion five, six, seven and eight. The influence of temperature is dependent on the latitude in phenoregion three, four, and five. Phenoregion one is the only phenoregion where temperature's influence is affected by ELEV. In some phenoregions, the influence of temperature on NDVI is also conditioned

by other environmental drivers. For example, in phenoregion four, The influence of AGDD_10 on NDVI is moderated by SRAD_6 and PRCP_16.

4.7.2.2 Precipitation

Precipitation variables include the weekly cumulative precipitation at the lag of 1 to 20 weeks. The lag structure and pairwise correlation matrix reveals that precipitation variables are much less correlated with temperature and light variables and with itself at different lags. Therefore, in all phenoregions, all 20 precipitation variables were retained in the final set of variables after VIF based iterated removing process.

The MARS models of all phenoregions include the precipitation variables except for phenoregion six and eight. This indicates that the temporal variance in precipitation does not have a great influence in driving vegetation growth. However, the environmental factors of GS_PRCP and MEAN_PRCP are included in phenoregion six and eight, respectively. This means that the spatial variance of the mean precipitation helps to explain the spatial difference of environment-vegetation relationships. Among the phenoregions that include the precipitation variables, the phenomodels of phenoregion one and five have the largest number of precipitation variables: phenoregion one has three and phenoregion five has four. Phenoregion two, five, and seven also include the long-term average of precipitation besides the weekly cumulative precipitation variables. The phenomodels of phenoregion one, five, and seven include precipitation variables in the main effect term.

Precipitation variables at much shorter lags (1 to 5 weeks) and much longer lags (16 to 20 weeks) appear to be more important than those at intermediate lags, since only

phenoregions one and five included the latter into their phenomodels (PRCP_10 for phenoregion one and PRCP_7 for phenoregion five).

For phenoregions without long-term average of precipitation (phenoregions one, three, four, and nine), phenoregions one and four contain precipitation variables (PRCP) at shorter lags, but PRCP in phenoregion four is considered much less important. Phenoregion two, five, and seven contain both lagged precipitation variables and long term average precipitation variables, thereby the importance of the lagged precipitation variables are naturally less important than if long-term average precipitation variables are not included, In this sense, phenoregions five and seven, among the three phenoregions, have relatively more important precipitation variables at shorter lags.

The interaction terms that contain the precipitation variables only appear in phenoregions one through five. The precipitation-vegetation relationships are affected by different environmental variables in these five phenoregions. They are affected by precipitation at other lags in phenoregion one, by light in phenoregion two, by aNDVI in phenoregion three, by temperature in phenoregion four and by the long-term average of precipitation and season in phenoregion five.

4.7.2.3 Light

The light variables include photoperiod (PTPD), solar radiation (SRAD) and the product of the two – solar radiant energy (SRE). Light variables are also considered important in vegetation development, but are less commonly included in equivalent models in literature. Analyzing the influence of light on 7-day vegetation dynamics, as with temperature, is harder, due to the high correlation between light and temperature variables and between the same variables at different lags.

The SRAD variables at different lags are included in models of phenoregion two, three, four, six, and seven. Also the SRAD variables are the only kind of light variables that are included in phenomodels. Most of the PTPD and SRE variables are removed due to multicollinearity beforehand. The SRAD variables appear as the main effect term in phenoregion two, three, and seven.

The SRAD variables are included in the interaction terms in phenoregion two, three, six, and seven. The influence of light on vegetation in the maturity phase is different from that in other phases in phenoregion two and seven. The SRAD variables interact with temperature variable of AGDD in phenoregion four, six, and seven.

4.8 Validation Results

The MARS models need to be validated to ensure its ability of generalization to different dates and different locations. This purpose was fulfilled by temporal, spatial and spatio-temporal cross validation. The MARS models also need to be validated for successful application to the physical environmental, and to ensure that they are not only the artifacts of measurement errors and noises in remote sensing data. This purpose of ground truthing is fulfilled by field validation.

4.8.1 Cross Validation

MARS models were built using values of modeling pixels in odd years. The models were applied to values of modeling pixels in even years (temporally independent data), values of validation pixels in odd years (spatially independent data) and values of validation pixels in even years (spatio-temporally independent data). As introduced in Section 3.6.1, the statistic of RMSE is used to quantify the performance of MARS models on different sets of data. RMSE indicates how accurate the model predicts when

applied on different sets of data. The RMSE values were compared with the standard error of the models to indicate if models performs consistently well on different data sets. The results of cross validation are shown in Table 4.10. Phenoregion one, two and five have the lowest RMSE and phenoregion three, four and seven have the highest. The RMSE values of models in all phenoregions are close to their respective standard errors, indicating the successfulness of the models in generalization to temporally and spatially independent data.

The results of CVs are shown in Table 4.11. Considering the mean value of NDVI in each phenoregion, the relative RMSE - CV shows that phenoregion four has the best performance, followed by phenoregion three, six, and seven.

4.8.2 Field Validation

4.8.2.1 Comparison of Ecocast and DAYMET Data

A comparison was made to reveal the similarities between Ecocast and DAYMET data as described in Section 3.6.2.1. The comparison helps to analyze the possible biases

Table 4.10 Root Mean Square Error (RMSE) of the temporal, spatial and spatio-temporal cross validation in nine phenoregions

Phenoregion	Standard Error	RMSE		
		Temporal Cross Validation	Spatial Cross Validation	Spatio-temporal Cross Validation
1	0.017	0.017	0.017	0.018
2	0.019	0.020	0.018	0.020
3	0.035	0.039	0.035	0.039
4	0.034	0.037	0.034	0.037
5	0.017	0.018	0.018	0.018
6	0.028	0.029	0.028	0.029
7	0.034	0.034	0.034	0.035
8	0.022	0.022	0.023	0.023
9	0.026	0.026	0.027	0.028

Table 4.11 Root Mean Square Error (RMSE) normalized to the mean of the observed values of the temporal, spatial and spatio-temporal cross validation in nine phenoregions

Phenoregion	Standard Error	CV		
		Temporal Cross Validation	Spatial Cross Validation	Spatio-temporal Cross Validation
1	8.0%	8.3%	8.1%	8.4%
2	8.7%	9.1%	8.9%	9.3%
3	7.0%	8.1%	7.3%	8.3%
4	5.8%	6.6%	5.9%	6.5%
5	7.5%	7.9%	7.8%	8.1%
6	7.0%	7.9%	7.1%	7.8%
7	7.1%	7.7%	7.2%	7.8%
8	8.3%	8.9%	8.5%	9.1%
9	7.6%	8.2%	7.7%	8.2%

of using different data in field validation.

Figure 4.13 plotted pairs of values extracted from Ecocast and DAYMET data, respectively, for four variables of maximum and minimum temperature, solar radiation and precipitation. All four variables of Ecocast data are highly correlated with their equivalent from DAYMET data with the correlation coefficient being 0.98, 0.97, 0.88 and 0.72. TMAX and TMIN from two data sources are very consistent. The DAYMET SRAD values are generally higher than the Ecocast SRAD values (with most of the points above the y=x line). PRCP is the variable that has the greatest inconsistency between two data sources. It is anticipated that the predicted errors within phenoregions that include PRCP variables may be biased either upwards or downwards.

4.8.2.2 Results of Field Validation

Three sites were selected in phenoregion two, eight, and nine. Consistent per-week spectra measurements were conducted at these three sites from late April to early June, 2011. Figure 4.14 shows the measured NDVI time series at the three field sites

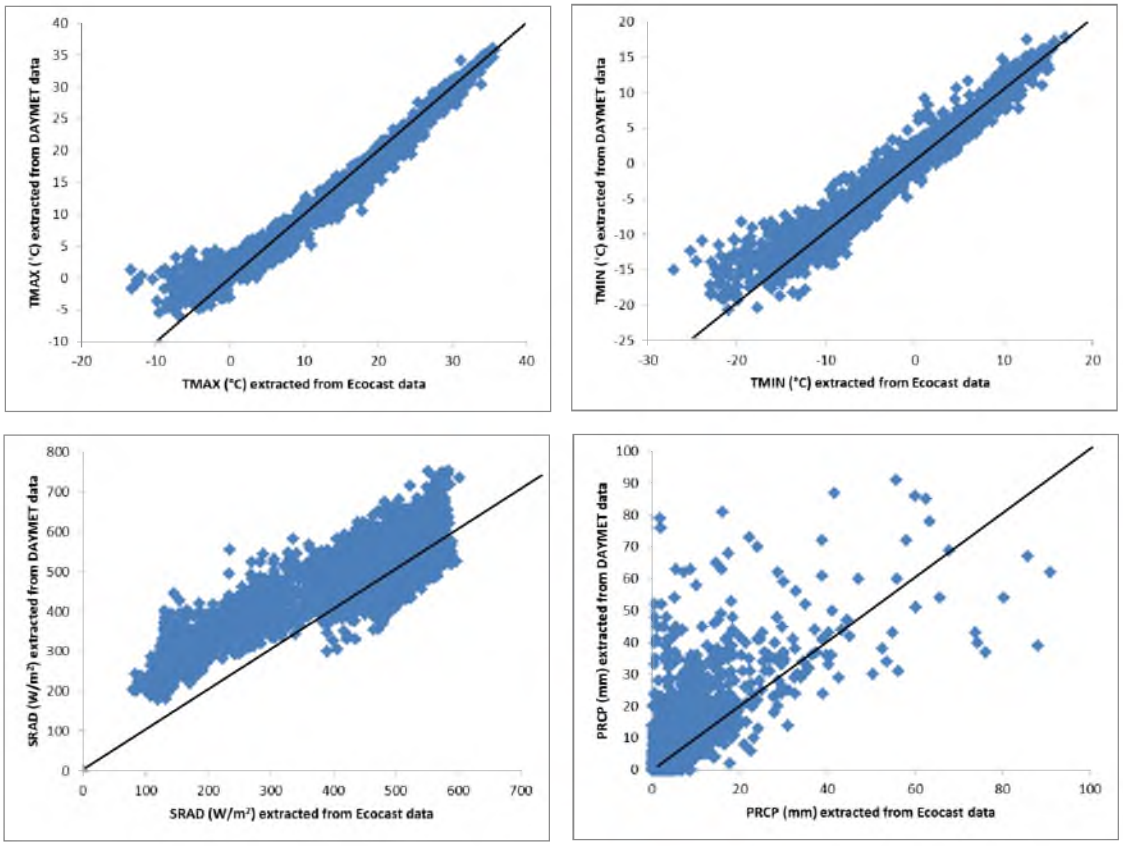


Figure 4.13 Scatter plots of variables extracted from Ecocast and DAYMET data with the black line being the y=x line.

during this period. The measurements on May 21st, 2011 were problematic due to the overcast sky and variant light conditions. The incident light is scattered light instead of direct sunlight under an overcast sky, and the variations of light conditions lead to variations of brightness and high variability of measurements at different points along the transect, the spectra measurements on May 21st, 2011 are considered less reliable and therefore the calculated NDVI on that day should be used with caution.

The RMSE and CV values of the field validation (Table 4.12) tend to be much larger than those of the cross validation in phenoregion two and eight, and smaller in phenoregion nine. The lower accuracy of the phenomodels in field validation is the

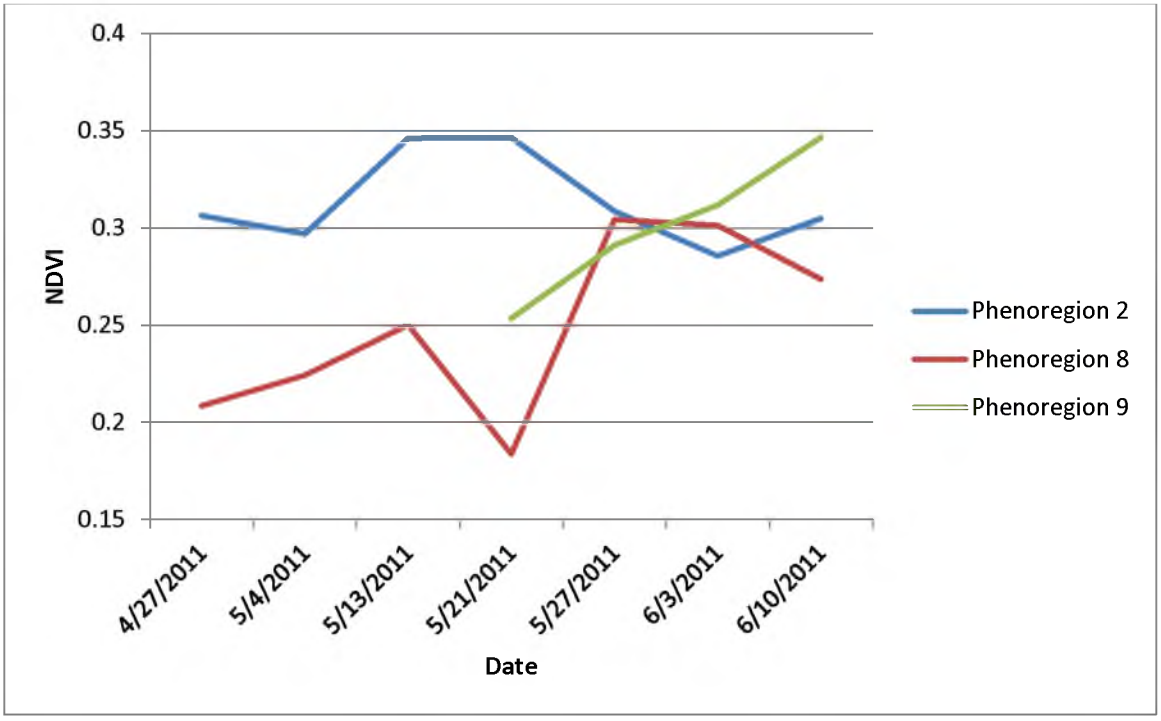


Figure 4.14 NDVI time series of three sites in phenoregion two, eight, and nine in 2011.

Table 4.12 The RMSE and CV of field validation before and after the removal of measurements on May 21st, 2011

Phenoregion	All measurements		Measurements on May 21st, 2011 excluded	
	RMSE	CV	RMSE	CV
2	0.034	10.9%	0.028	9.1%
8	0.052	20.1%	0.015	5.6%
9	0.015	4.7%	0.016	4.8%

combined results of different data sources (DAYMET instead of Ecocast), less reliable measurements (measurements on May 21st, 2011), errors during field measurements (such as transect reconstruction and the spectrometer pistol holding position), and errors associated with pixel alignment and using the 45 m transect to represent the 1 km^2 pixel.

If the less reliable measurements on May 21st, 2011 are excluded from field validation, the RMSE and CV values calculated are shown in the right two columns of Table 4.12. Both RMSE and CV values in phenoregions two and eight decreased significantly and are within reasonable ranges as compared to those in cross validation.

4.9 Phenological Decision Support System

The prototype Pheno DSS implemented in this research contains three main functions:

1. Data access module: Users can display over Google Maps (Figure 4.15 (A) and (B)) or download (Figure 4.15 (B)) the archived preprocessed eMODIS and Ecocast data for the UCRB. All preprocessed eMODIS and Ecocast data were saved on the server side. Users can choose to display or download these historical images by specifying the date and the variable (NDVI for eMODIS data, and maximum and minimum temperature, precipitation, and solar radiation for Ecocast data). Users can also overlay the grazing alloments or grazing pastures on the eMODIS or Ecocast images to examine the values of specific alloments or pastures (Figure 4.15 (B)).
2. Prediction module: The crucial function of the pheno DSS is the prediction of NDVI for the next 7-day interval. In the scenario of the real-time or near real-time production of eMODIS and Ecocast dataset, this DSS can generate the

Figure 4.15 Data access module of the Pheno DSS.

NDVI image of the UCRB for the next available 7-day interval (Figure 4.16 (A)), using the built-in pheno models of the nine phenoregions. The DSS can also generate an NDVI image for user selected historical date (Figure 4.16 (B)). Similar with the functions in the data access module, users can choose to download the image of predicted NDVI, and overlay the layers of grazing allotments and pastures.

3. Time series of NDVI and environmental drivers can be generated (Figure 4.17). Users first need to choose either a point or a region. Users can choose a point by simply clicking on the map or inputting the latitude and longitude of the point by hand. The region can be drawn using the rectangle or polygon tools provided by Google Maps (Figure 4.17 (A)). The rectangle can be either drawn on the map or input the boundary information from the right panel. The region also can be one of the grazing allotments or pastures (Figure 4.17 (B)). When users click on one of the grazing allotments or pastures, it will show on the right panel the information of that allotment or pasture (number and name) (Figure 4.17 (B)). Users will then choose the variable time series which they would like to display and the time range of the time series. The pheno DSS will generate both the graphic time series and a table with exact values (Figure 4.17 (C)). The table can be saved as an Excel file for easy data documentation (Figure 4.17 (C)).

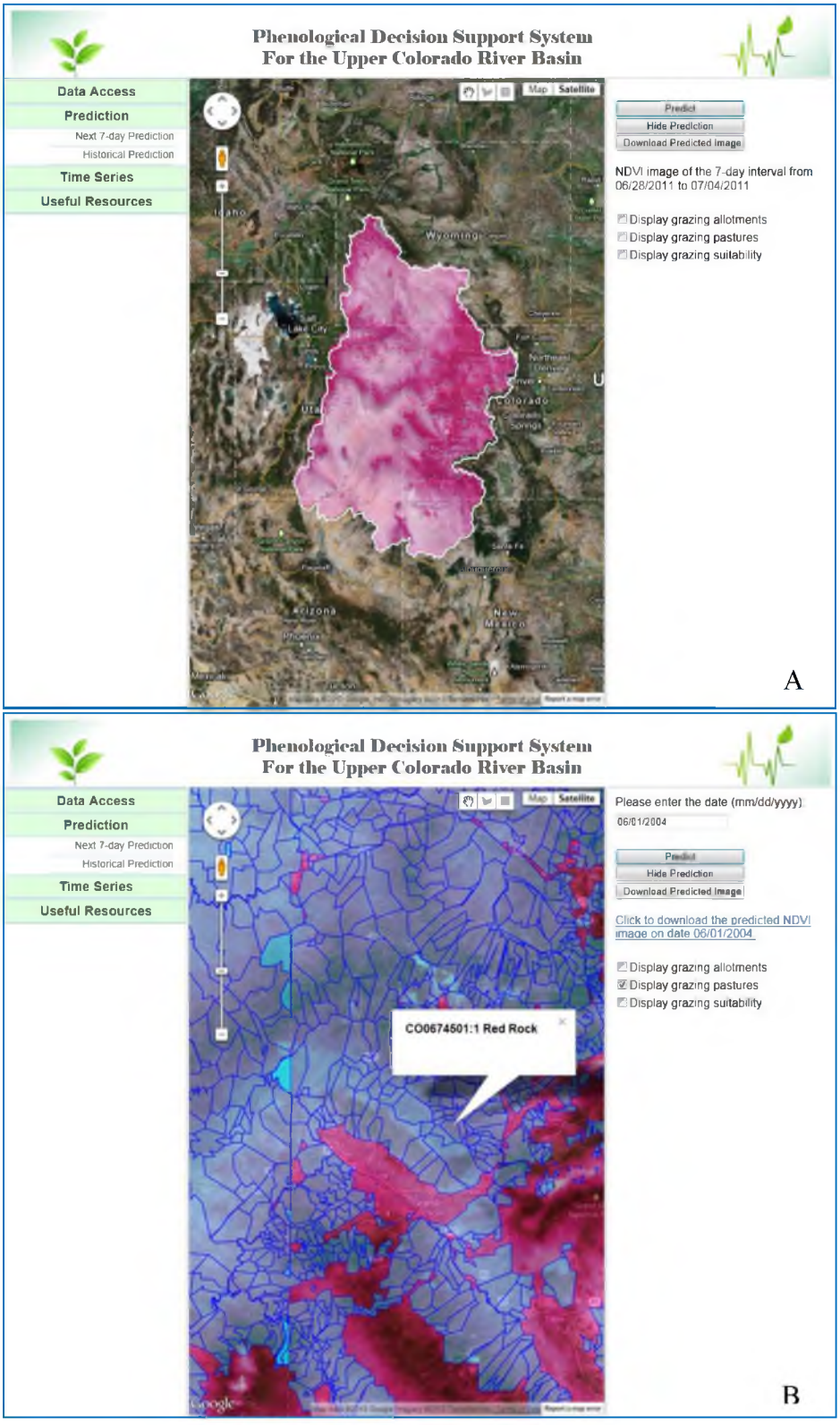


Figure 4.16 Prediction module of the Pheno DSS.

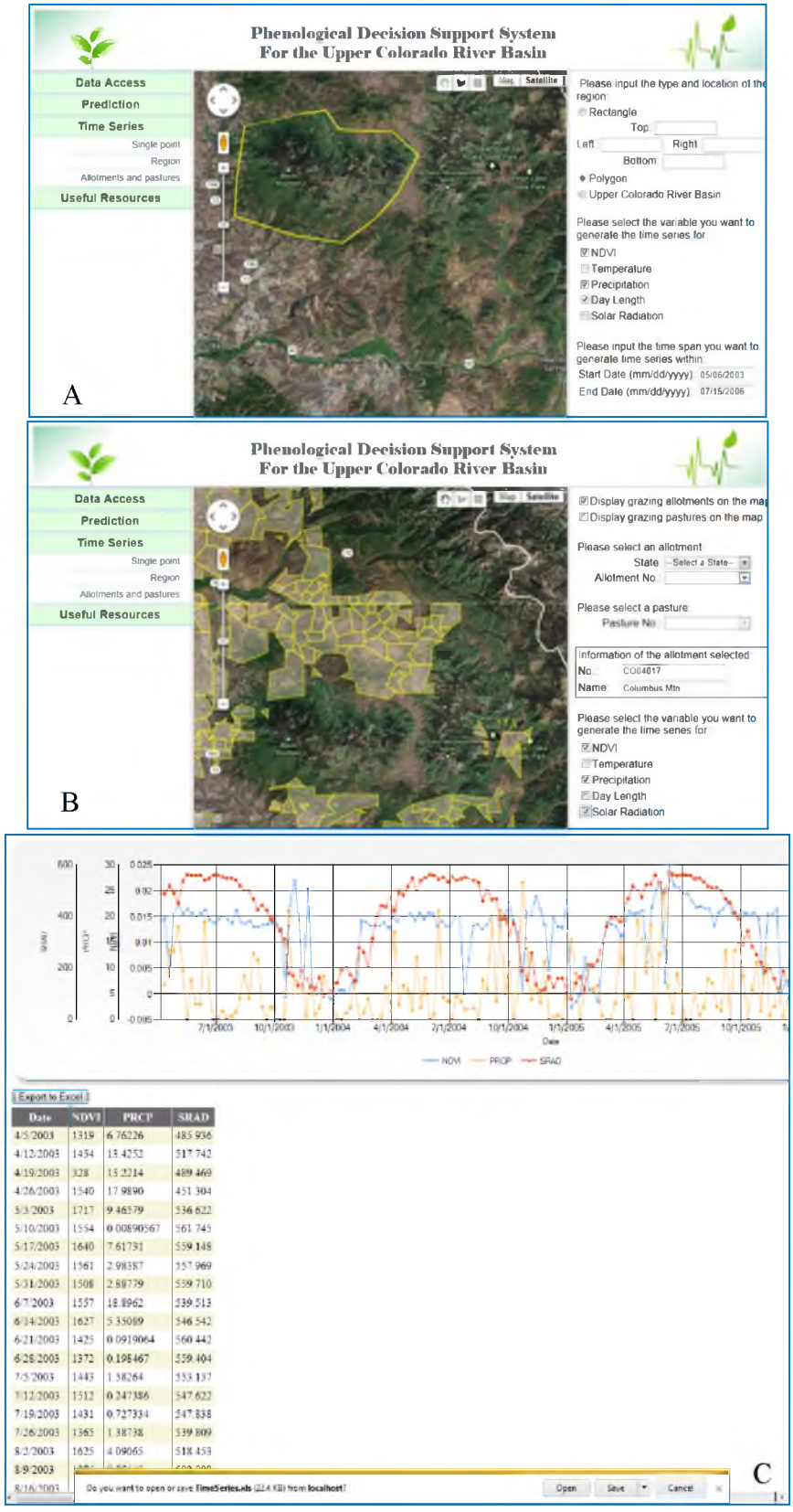


Figure 4.17 Time series module of the Pheno DSS.

5 DISCUSSION / IMPLICATIONS OF RESEARCH

5.1 Research Questions and Research Objectives

The four research questions were answered by this dissertation and three research objectives were fulfilled.

Research question one is, "How can phenoregions be effectively delineated in the UCRB and how can performance of phenoregions be evaluated?" The PCA plus *k*-means++ clustering were adopted to disaggregate the UCRB into self-similar clusters. Twelve variables including environmental factors as well as the long-term average and temporal variations were used in the phenoregion delineation. The spatial homogeneity was quantified using the total within-cluster sum-of squares and the mean standard deviation. The effectiveness of phenoregions and the homogeneity within phenoregions were further proved in the analysis of phenological cycles, outliers and the successful modeling of the environment-vegetation relationship using one MARS model for each phenoregion.

Research question two is, "What are the dependent and independent variables of the phenomodels?" The dependent variable of the phenomodels are NDVI for all phenoregions. The use of NDVI in phenological models and phenological research has been well established in literature, and the major limitation of easy saturation is not a concern in this research. The independent variables differ in each phenoregion. A two-fold independent variable selection was used to select important independent variables from the pool of candidate variables including aNDVI, environmental drivers and

environmental factors, while reducing multicollinearity. The first step reduces multicollinearity using an iterated VIF based variable reduction approach. Then MARS model automatically selects the important variables from the reduced set of variables.

Research question three is, "What are the relationships between the dependent and independent variables identified in question two and how to validate these relationships as represented using a mathematical model?" The relationships between the dependent and independent variables are successfully quantified using MARS. The relationships differ among phenoregions, and differ at different ranges of environmental variables within the same phenoregion. The relationships represented in the MARS models were validated using both cross validation and field validation, ensuring the generalization of the models both temporally and spatially and to practical use in reality.

Research question four is, "What information should be provided and what mechanisms should be adopted in the proposed DSS to most effectively assist land managers in formulating grazing strategies and decisions?" The prototype pheno DSS implemented several modules to provide information about the trend of environmental variables, historical images as a still picture of the environment, predicted NDVI values from the phenomodels, and basic grazing suggestions. The spatial visualization module and the user friendly interaction can greatly assist land mangers in acquiring information needed for their decision making.

The successful answering of the four research questions led to the fulfillment of the three research objectives. Nine phenoregions were delineated to disaggregate spatially the environment-vegetation relationships. Nine MARS models were then built, one for each phenoregion. The interpretation of MARS models in all phenoregions as the

representation of vegetation dynamics reveals the common features in different phenoregions as well as the differences among phenoregions and under different intervals of environmental variables within each phenoregion. The prototype pheno DSS can successfully predict the NDVI values of the UCRB for the next 7-day interval, and has the ability to go real-time or near real-time under the scenario of the real time or near real time production of eMODIS and Ecocast data sets.

5.2 The Applicability of Phenomodels and Modeling Frameworks

The nine phenomodels represent the UCRB specific and phenoregion specific relationships, and are not expected to work equally well in regions outside of the UCRB. However, as mentioned in Section 1.3, UCRB is representative of other BLM managed lands and other regions with similar ecosystems. Therefore, it is possible to apply the phenomodels developed for the UCRB to a new region, as long as this new region is analyzed beforehand and concluded to share similar phenological characteristics with one of the phenoregions in the UCRB.

The phenomodeling framework of phenoregion delineation plus MARS modeling within each phenoregion, though, is a universal approach that can be applied elsewhere. The list of candidate variables was summarized from literature on various regional settings, and therefore makes a comprehensive set that influences the environment-vegetation relationships spatially and temporally. The MARS modeling approach can select the most important variables from the set of environmental drivers and factors. This framework works especially well in regions that are also geographically diverse. It uses its complex mechanism to account for the complex spatially and temporally variant

environment-vegetation relationships. It certainly can be applied to homogeneous regions as well, in which case, the step of phenoregion delineation is not necessary.

However, due to some limitations of the MARS approach, there are some restrictions applying the framework. Firstly, the development of MARS models require a very large number of observations, therefore, it is harder to apply to smaller regions especially when data for only a limited number of years are available. Secondly, the confidence intervals of parameters and other checks on the model cannot be calculated directly for MARS models, and certain validation techniques are required to validate the models. When the number of observations is small, there may not be enough observations that can be used for cross validation. There are also regions where field validation cannot be conducted due to the accessibility issues. In these situations, the MARS modeling is not the optimal approach.

5.3 Phenological Predictive Models for Rangelands

The same modeling framework can be applied to only the pixels within the grazing allotment and pastures in the UCRB. As compared with the modeling for the entire UCRB, the modeling for rangelands is easier in that: the climatic condition and vegetation are not as diverse, which can possibly lead to higher accuracy of prediction; outliers are expected to be less and NDVI time series are expected to be less noisy due to the relatively uniform land cover.

However, there are some difficulties in the NDVI prediction considering the ongoing grazing activity. Vegetation development occurs at the same time as livestock are on the range consuming the vegetation. The current phenomodels do not take this into consideration. When the models are enhanced and adjusted for rangelands, this problem

becomes a serious concern. The possible adjustment to account for this disturbance caused by livestock consumption is to include extra variables (such as the livestock distribution information) into the model.

5.4 Phenoregions / Scale Dependencies and Pixel Nature of Remote Sensing Phenology

The heterogeneity of environmental characteristics and the resulting heterogeneity of environment-vegetation relationship at different locations, has long been a problem in phenological and ecological modeling. This problem requires attention before successful phenomodels can be built for the UCRB. In this research, an approach of disaggregating the UCRB into phenologically homogeneous phenoregions is proposed to address this problem. Successful models were built for each phenoregion with different environmental drivers and factors and with different coefficients, accounting for the influence of environmental drivers on the 7-day vegetation dynamics as moderated by other environmental drivers and factors. The similarities and differences of the environment-vegetation relationships in different phenoregions can be summarized.

Remote sensing of phenology entails a scaling-up process, from field sites to regional and global scales, and from individual plants and species to vegetation communities and ecosystems. Given the outcomes of this research, it is reasonable to use a pixel based spatial partition of the UCRB as the basis of remote sensing phenological modeling. Methods used in ground based phenology should be either adjusted or redesigned to adapt to the broad-scale nature of remote sensing phenology. Therefore, traditional species based phenological monitoring or modeling is not applicable in this research, because:

1. In remote sensing phenology deals with, instead of the single plant, the pixel which is a mixture of signals from everything on the surface including bare ground and different species of vegetation. This mixture of nature is especially true in the geographically diverse UCRB, seen by examining the composition of GAP land covers within 1-km pixels. Even for multispectral sensors that have frequent coverage of the UCRB and the finest spatial resolution (such as ASTER with 15m spatial resolution and 16 days temporal resolution), the mixture problem in a pixel is still inevitable.
2. Within the same ecosystem, such as shrubland, the environment-vegetation relationship can still be very divergent. For example, Knapp (1984) found that the responses of three grasses (big blue stem, little bluestem, and switchgrass) in terms of flowering intensity can be slightly different to increased mean annual precipitation (Schwartz, 2003).
3. Even the same species has different responses given different ranges of environmental drivers and factors. This has been demonstrated by many studies. For example, the same crop plant variety needs different mean annual temperature sums to reach a particular phenophase if cultivated at different latitudes (Schwartz, 2003). Also, the saturating solar radiation at maximal growth rate varies with elevation even for the same species, which is a result of the genetic adaptation to the habitat (Pfafflin, 2006).

This research focus on the phenological cycles and the vegetation-environment relationships of the 1-km pixels, therefore, it is one of the objectives of this research to identify the response mechanism of the pixel with mixed vegetation and species.

5.5 MARS and Other Models

This research uses phenoregion plus MARS modeling to predict the next 7-day interval NDVI values and interpret the influence of environment on the 7-day vegetation dynamics. The phenoregion plus MARS strategy is still regarded as a better choice in this research considering the major objectives. Firstly, a more generic vegetation-environment relationship is desired rather than a pixel specific relationship. The vegetation-environment relationship does vary greatly due to the geographic diversity in the UCRB. However, the reason of the variation is also a question to be answered. Actually, both the spatially and temporally variant relationship is essentially due to the combination, and interaction of different values of environmental factors and drivers. That is also why MARS works better than ordinary multivariate regression models. Secondly, phenoregion accounts for the major difference of the relationships so that more homogeneous relationships can be modeled and interpreted for each phenoregion, though environmental factors still take effect in accounting for the minor interpixel differences or interphase differences.

The complexity of the responses of vegetation to the environment results in the advantages of more flexible MARS over ordinary linear regression. MARS is easily interpretable as with linear regression. MARS does not assume any underlying relationships, instead, it can approximate any kinds of relationships using basis functions driven by data. The basis functions allow different linear relationships in different intervals of data.

This section focuses on the discussion of this strategy as compared with other kinds of models.

Geographically Weighted Regression (GWR), as one of the commonly used spatial regression models, considers spatial dependence and accounts for the spatially variant relationship. GWR, however, lack the flexibility of phenoregion plus MARS in relationship approximation. For example, GWR requires a predetermined relationship between dependent and independent variables, and it does not automatically model interactions.

The fixed and random effect models usually applied on longitudinal data consider the unit or individual effect, and work especially well with datasets with heterogeneous relationships. However, out of the similar reason that it is the purpose of this research to explore what caused the "individual effect," MARS with explicitly expressed individual effect (environmental factors and aNDVI) is preferred.

6 CONCLUSIONS AND FUTURE WORK

This research identified and delineated nine phenoregions in the geographically diverse UCRB. These phenoregions work as the basic areal unit for the phenological modeling. A phenomodel was built for each of the nine phenoregions to predict the vegetation abundance 7 days in the future. Interpretation of the phenomodels contributes to the understanding of the environment with respect to the 7-day vegetation dynamics. A prototype pheno DSS was then developed that integrates phenological theory and GIS technology to provide a suite of information to supplement and present the predicted results from the phenomodels.

The framework of phenoregion delineation plus MARS model development within each phenoregion proposed in this research provides a way to model and understand the complex and variant phenological features and environment-vegetation relationships in the geographically diverse regions.

Some conclusions about the vegetation-environment relationships can be made by examining the phenomodels in all nine phenoregions.

Firstly, vegetation responds faster to environmental drivers in dryer phenoregions and phenoregions with more shrubland and grassland. This is inferred from that fact that environmental drivers at shorter lags are more likely to be selected to enter the model in such phenoregions. For example, temperature variables at shorter lags are more important in phenoregions one, five, and six; precipitation variables at shorter lags are considered

more important in phenoregions one, five, and seven. Also, some phenoregions containing more shrubland (such as phenoregion five) include both environmental drivers at shorter and longer lags. This is probably caused by the mixed species of slow-responding forest and deep-rooted shrubs and fast-responding shallow-rooted shrubs and grasses.

Secondly, environmental drivers are found to be more important in phenoregions where they are limiting factors. Precipitation is more important in moisture limited phenoregions such as phenoregions one and five, which have both the largest number of precipitation variables and main-effect precipitation variables. Temperature is more important in cooler phenoregions. For example, the phenoregions three, four, six, and seven with lower temperatures have the main-effect temperature variables. Solar radiation is generally important for all phenoregions, but the relative importance in different phenoregions is hardly observed because solar radiation is seldom a limiting factor in UCRB and can be easily saturated.

Only one or two categories of phenological phases are included in the phenomodels. SEASON variable category maturity is included in seven out of nine phenoregions, indicating the different environment-vegetation relationships in the maturity phase. However, this does not indicate that the vegetation-environment relationship is not temporally variant in other phenophases. Instead, as analyzed in Section 5.3, spatiotemporally-variant relationships are largely due to the different values of environmental drivers and factors in different locations and at different dates. Therefore, the application of MARS, by modeling different relationship within different data intervals, partly accounts for the temporally variant relationship, and weakens the

influence of phenophases as shown in models. Possible improvements of this research in the future include the following:

This research uses NDVI as the indicator of vegetation abundance. However, as indicated in 2.1.2, EVI, SAVI and NDWI are also good candidates. Phenomodels using these VIs are worth trying. The advantages and disadvantages of different VIs in phenomodeling are of interest.

Though currently only 1km data can be acquired frequently, phenomodels based on data of different spatial resolution can be a future direction. The comparison of these phenomodels can shed light on the scale dependency of vegetation-environment relationship.

A few phenomena represented by the phenomodels are hard to explain using the current data and information in this dissertation. Additional information of the UCRB (such as soil types, species) may help make more comprehensive interpretations.

In this dissertation, cross validation is conducted on three randomly selected fixed sets of samples. In the future, the pixels sampled using systematic plus random sampling can be divided into several sets, each set will be subsequentially used as the modeling / training set while others as the validation sets. The final phenoregion will be selected as the one with the best cross validation results.

Due to the limited time, funding and labor, field sites were selected around a location where there are several phenoregions. Field sites located in the center of a large patch are more "pure" and are considered more typical in terms of the vegetation-environment relationships modeled in respective phenomodels, and are therefore more preferable. Also, the 45 m transect may not be representative enough of the 1 km^2 pixel, a

longer transect or other field sampling methods may help to improve the accuracy of field validation.

APPENDIX A

GAP CODES AND NAMES

Table A.1 GAP codes and names

LEVEL1	LEVEL2	LEVEL3	CN_LEVEL1	CN_LEVEL2	CN_LEVEL3
1	12	1201	Human land use	Developed	Developed, Open Space
1	12	1202	Human land use	Developed	Developed, Low Intensity
1	12	1203	Human land use	Developed	Developed, Medium Intensity
1	12	1204	Human land use	Developed	Developed, High Intensity
1	13	1301	Human land use	Mining	Quarries, Mines, Gravel Pits and Oil Wells
1	14	1402	Human land use	Agriculture	Cultivated Cropland
1	14	1403	Human land use	Agriculture	Pasture/Hay
2	21	2102	Aquatic	Open water	Open Water (Fresh)
3	31	3105	Sparse and barren systems	Beach, shore and sand	Undifferentiated Barren Land
3	31	3121	Sparse and barren systems	Beach, shore and sand	Inter-Mountain Basins Active and Stabilized Dune
3	32	3201	Sparse and barren systems	Cliff, canyon and talus	North American Warm Desert Bedrock Cliff and Outcrop
3	32	3202	Sparse and barren systems	Cliff, canyon and talus	Rocky Mountain Cliff, Canyon and Massive Bedrock
3	32	3216	Sparse and barren systems	Cliff, canyon and talus	Inter-Mountain Basins Cliff and Canyon
3	32	3218	Sparse and barren systems	Cliff, canyon and talus	Colorado Plateau Mixed Bedrock Canyon and Tableland
3	33	3304	Sparse and barren systems	Bluff and badland	Inter-Mountain Basins Shale Badland
3	34	3403	Sparse and barren systems	Playa, wash and mudflat	Inter-Mountain Basins Wash
3	35	3502	Sparse and barren systems	Alpine sparse and barren	North American Alpine Ice Field
3	35	3503	Sparse and barren systems	Alpine sparse and barren	Rocky Mountain Alpine Bedrock and Scree
3	36	3603	Sparse and barren systems	Other sparse and barren	Inter-Mountain Basins Volcanic Rock and Cinder Land
4	41	4111	Forest and woodland systems	Deciduous dominated forest and woodland (xeric-mesic)	Rocky Mountain Aspen Forest and Woodland

Table A.1 Continued

LEVEL1	LEVEL2	LEVEL3	CN_LEVEL1	CN_LEVEL2	CN_LEVEL3
4	41	4112	Forest and woodland systems	Deciduous dominated forest and woodland (xeric-mesic)	Rocky Mountain Bigtooth Maple Ravine Woodland
4	41	4147	Forest and woodland systems	Deciduous dominated forest and woodland (xeric-mesic)	Inter-Mountain Basins Curl-leaf Mountain Mahogany Woodland and Shrubland
4	43	4324	Forest and woodland systems	Mixed deciduous/coniferous forest and woodland (xeric-mesic)	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland
4	45	4512	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Colorado Plateau Pinyon-Juniper Woodland
4	45	4526	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Rocky Mountain Foothill Limber Pine-Juniper Woodland
4	45	4527	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Rocky Mountain Lodgepole Pine Forest
4	45	4528	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Southern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest and Woodland
4	45	4530	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Southern Rocky Mountain Ponderosa Pine Woodland
4	45	4531	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland
4	45	4532	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland
4	45	4534	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Southern Rocky Mountain Pinyon-Juniper Woodland
4	45	4543	Forest and woodland systems	Conifer dominated forest and woodland (xeric-mesic)	Middle Rocky Mountain Montane Douglas-fir Forest and Woodland
4	46	4609	Forest and woodland systems	Conifer dominated forest and woodland (mesic-wet)	Northern Rocky Mountain Mesic Montane Mixed Conifer Forest
4	46	4610	Forest and woodland systems	Conifer dominated forest and woodland (mesic-wet)	Southern Rocky Mountain Mesic Montane Mixed Conifer Forest and Woodland

Table A.1 Continued

LEVEL1	LEVEL2	LEVEL3	CN_LEVEL1	CN_LEVEL2	CN_LEVEL3
4	46	4611	Forest and woodland systems	Conifer dominated forest and woodland (mesic-wet)	Rocky Mountain Subalpine Mesic Spruce-Fir Forest and Woodland
5	51	5103	Shrubland, steppe and savanna systems	Alpine and avalanche chute shrubland	Rocky Mountain Alpine Dwarf-Shrubland
5	52	5205	Shrubland, steppe and savanna systems	Scrub shrubland	Inter-Mountain Basins Mixed Salt Desert Scrub
5	53	5307	Shrubland, steppe and savanna systems	Steppe	Inter-Mountain Basins Big Sagebrush Steppe
5	53	5308	Shrubland, steppe and savanna systems	Steppe	Inter-Mountain Basins Montane Sagebrush Steppe
5	53	5309	Shrubland, steppe and savanna systems	Steppe	Inter-Mountain Basins Semi-Desert Shrub Steppe
5	56	5601	Shrubland, steppe and savanna systems	Conifer dominated savanna	Colorado Plateau Pinyon-Juniper Shrubland
5	56	5603	Shrubland, steppe and savanna systems	Conifer dominated savanna	Inter-Mountain Basins Juniper Savanna
5	57	5701	Shrubland, steppe and savanna systems	Sagebrush dominated shrubland	Colorado Plateau Mixed Low Sagebrush Shrubland
5	57	5703	Shrubland, steppe and savanna systems	Sagebrush dominated shrubland	Inter-Mountain Basins Mat Saltbush Shrubland
5	57	5704	Shrubland, steppe and savanna systems	Sagebrush dominated shrubland	Wyoming Basins Dwarf Sagebrush Shrubland and Steppe
5	57	5706	Shrubland, steppe and savanna systems	Sagebrush dominated shrubland	Inter-Mountain Basins Big Sagebrush Shrubland
5	57	5707	Shrubland, steppe and savanna systems	Sagebrush dominated shrubland	Southern Colorado Plateau Sand Shrubland
5	58	5803	Shrubland, steppe and savanna systems	Deciduous dominated shrubland	Colorado Plateau Blackbrush-Mormon-tea Shrubland
5	58	5806	Shrubland, steppe and savanna systems	Deciduous dominated shrubland	Rocky Mountain Lower Montane-Foothill Shrubland

Table A.1 Continued

LEVEL1	LEVEL2	LEVEL3	CN_LEVEL1	CN_LEVEL2	CN_LEVEL3
5	58	5809	Shrubland, steppe and savanna systems	Deciduous dominated shrubland	Rocky Mountain Gambel Oak-Mixed Montane Shrubland
7	71	7102	Grassland systems	Alpine grassland	Rocky Mountain Alpine Fell-Field
7	71	7103	Grassland systems	Alpine grassland	Rocky Mountain Dry Tundra
7	72	7204	Grassland systems	Montane grassland	Northern Rocky Mountain Subalpine-Upper Montane Grassland
7	72	7205	Grassland systems	Montane grassland	Rocky Mountain Subalpine-Montane Mesic Meadow
7	72	7206	Grassland systems	Montane grassland	Southern Rocky Mountain Montane-Subalpine Grassland
7	73	7305	Grassland systems	Lowland grassland and prairie (xeric-mesic)	Inter-Mountain Basins Semi-Desert Grassland
7	73	7306	Grassland systems	Lowland grassland and prairie (xeric-mesic)	Northwestern Great Plains Mixedgrass Prairie
8	81	8101	Recently disturbed or modified	Harvested forest	Recently Logged Areas
8	81	8107	Recently disturbed or modified	Harvested forest	Harvested forest-Shrub Regeneration
8	83	8301	Recently disturbed or modified	Recently burned	Recently Burned
8	84	8404	Recently disturbed or modified	Introduced vegetation	Introduced Upland Vegetation - Annual Grassland
8	84	8406	Recently disturbed or modified	Introduced vegetation	Introduced Riparian and Wetland Vegetation
8	84	8407	Recently disturbed or modified	Introduced vegetation	Introduced Upland Vegetation - Perennial Grassland and Forbland
8	85	8501	Recently disturbed or modified	Other disturbed or modified	Disturbed, Non-specific

Table A.1 Continued

LEVEL1	LEVEL2	LEVEL3	CN_LEVEL1	CN_LEVEL2	CN_LEVEL3
8	85	8502	Recently disturbed or modified	Other disturbed or modified	Disturbed/Successional - Recently Chained Pinyon-Juniper
9	96	9606	Riparian and wetland systems	Wet meadow or prairie	Rocky Mountain Alpine-Montane Wet Meadow
9	97	9707	Riparian and wetland systems	Depressional wetland	Western Great Plains Open Freshwater Depression Wetland
9	97	9711	Riparian and wetland systems	Depressional wetland	Western Great Plains Saline Depression Wetland
9	98	9810	Riparian and wetland systems	Floodplain and riparian	Inter-Mountain Basins Greasewood Flat
9	98	9823	Riparian and wetland systems	Floodplain and riparian	Western Great Plains Floodplain
9	98	9825	Riparian and wetland systems	Floodplain and riparian	Rocky Mountain Lower Montane Riparian Woodland and Shrubland
9	98	9830	Riparian and wetland systems	Floodplain and riparian	Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland
9	98	9832	Riparian and wetland systems	Floodplain and riparian	Rocky Mountain Subalpine-Montane Riparian Woodland
9	98	9837	Riparian and wetland systems	Floodplain and riparian	Rocky Mountain Subalpine-Montane Riparian Shrubland
9	98	9848	Riparian and wetland systems	Floodplain and riparian	Western Great Plains Riparian Woodland and Shrubland

APPENDIX B

DETAILED INFORMATION OF PHENOMODELS

B.1 Phenoregion One

The phenomodel of phenoregion one has an R^2 of 91.61% and standard error of 0.0169. Table B.1 shows the importance of the independent variables included in phenoregion one. Figure B.1 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion one is as follows:

```
NDVI =
  0.22
  +    0.83 * max(0,      aNDVI -      0.21)
  -    0.76 * max(0,      0.21 -     aNDVI)
  - 5.8e-06 * max(0,      ELEV -      1925)
  +   6e-06 * max(0,      1925 -      ELEV)
  +     2.1 * max(0, MEAN_NDVI -      0.13)
  -     1.3 * max(0,      0.13 - MEAN_NDVI)
  +  0.0026 * max(0,       1.9 -   PRCP_10)
  +   0.032 * max(0,     aNDVI -      0.21) * max(0,       LAT -        37)
  -   0.021 * max(0,     aNDVI -      0.21) * max(0,        37 -       LAT)
  - 1.1e-06 * max(0,      TMIN -       2.1) * max(0,      1925 -      ELEV)
  - 3.8e-07 * max(0,       2.1 -      TMIN) * max(0,      1925 -      ELEV)
  + 1.8e-06 * max(0,      TMIN -         3) * max(0,      ELEV -      1925)
  - 2.6e-06 * max(0,         3 -      TMIN) * max(0,      ELEV -      1925)
  - 0.00025 * max(0,        13 -      PRCP) * max(0,       1.9 -   PRCP_10)
  +   1e-07 * max(0,      ELEV -      1925) * max(0,   AGDDu_5 -       121)
  + 3.5e-07 * max(0,      ELEV -      1925) * max(0,       121 -   AGDDu_5)
  -  0.0021 * max(0,      0.13 - MEAN_NDVI) * max(0,      AGDD -       105)
  -   0.015 * max(0,      0.13 - MEAN_NDVI) * max(0,       105 -      AGDD)
  +   0.094 * max(0,      0.13 - MEAN_NDVI) * max(0,    TMIN_4 -       6.9)
  +   0.036 * max(0,      0.13 - MEAN_NDVI) * max(0,       6.9 -    TMIN_4)
  + 5.6e-06 * max(0,   PRCP_10 -       1.9) * max(0,   PRCP_18 -       2.3)
  + 0.00013 * max(0,   PRCP_10 -       1.9) * max(0,       2.3 -   PRCP_18)
```

B.2 Phenoregion Two

The phenomodel of phenoregion two has an R^2 of 95.24% and standard error of 0.0185. Table B.2 and Figure B.2 shows the importance of the independent variables and the geographic relationships. The mathematical equation of the phenomodel in

Table B.1 Importance of entered independent variables in phenoregion one

	nsubsets	gcv	rss
aNDVI	22	100	100
MEAN_NDVI	20	9.7	9.7
LAT	18	6.2	6.3
AGDD	17	5.6	5.6
PRCP	14	4.5	4.6
PRCP_10	14	4.5	4.6
PRCP_18	13	4.1	4.2
ELEV	12	3.6	3.7
AGDDu_5	12	3.6	3.7
TMIN	11	3.3	3.4
TMIN_4	10	2.9	3

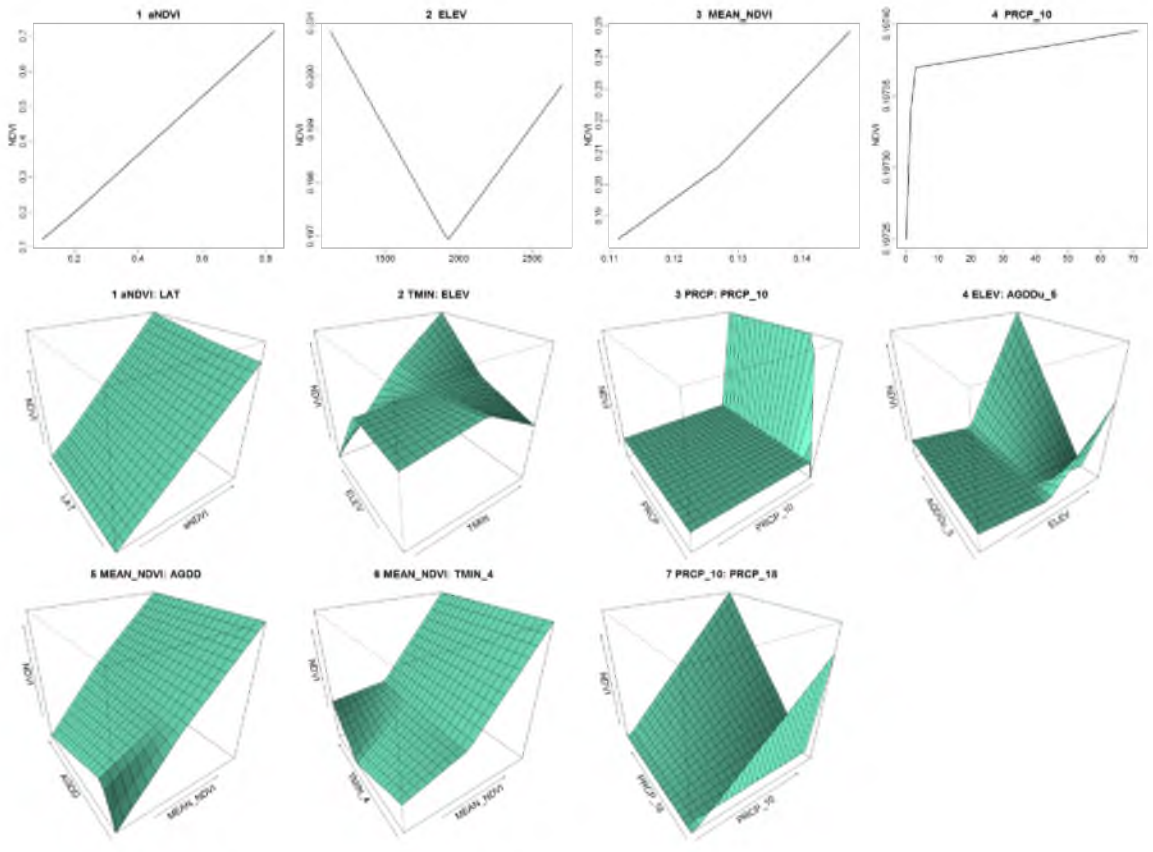


Figure B.1 Geographic relationships in phenoregion one.

Table B.2 Importance of entered independent variables in phenoregion two

	nsubsets	gcv	rss
aNDVI	16	100	100
AGDD_10	15	13.8	13.8
MEAN_NDVI	13	9.3	9.3
SRAD_6	11	6.4	6.4
GDDu_7	11	6.4	6.4
SEASON2	7	3.9	4
GS_PRCP	7	3.5	3.5
PRCP	3	1.7	1.7

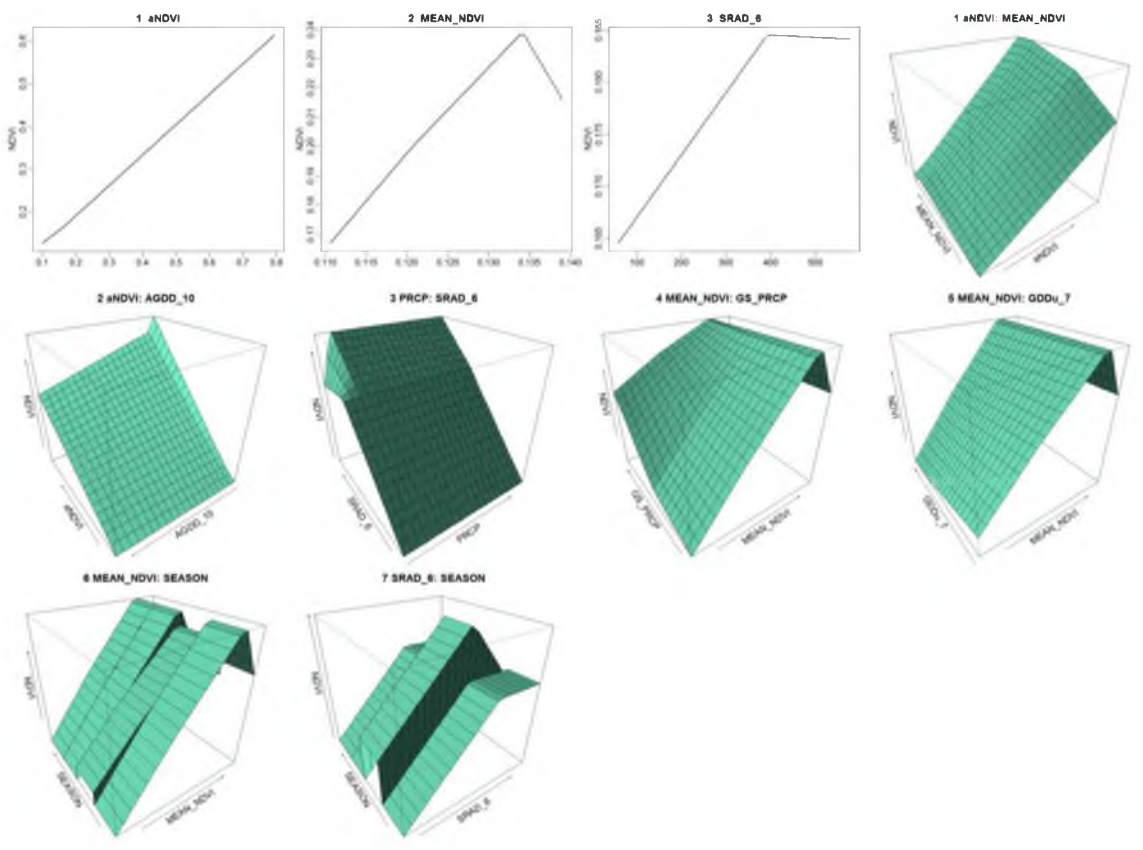


Figure B.2 Geographic relationships in phenoregion two.

phenoregion two is as follows:

```
NDVI =
  0.21
  +    0.82 * max(0,      aNDVI -      0.16)
  -    0.59 * max(0,       0.16 -     aNDVI)
  -     4.9 * max(0, MEAN_NDVI -      0.13)
  -     2.6 * max(0,       0.13 - MEAN_NDVI)
  + 4.6e-05 * max(0,     SRAD_6 -       392)
  -   6e-05 * max(0,        392 -    SRAD_6)
  +     5.3 * max(0,      aNDVI -      0.16) * max(0, MEAN_NDVI -      0.12)
  -      18 * max(0,      aNDVI -      0.16) * max(0,       0.12 - MEAN_NDVI)
  - 0.00052 * max(0,      aNDVI -      0.16) * max(0,    AGDD_10 -        24)
  +  0.0043 * max(0,      aNDVI -      0.16) * max(0,         24 -   AGDD_10)
  -   2e-06 * max(0,         26 -      PRCP) * max(0,     SRAD_6 -       392)
  +     0.1 * max(0,       0.13 - MEAN_NDVI) * max(0,    GS_PRCP -        32)
  -   0.025 * max(0,       0.13 - MEAN_NDVI) * max(0,         32 -   GS_PRCP)
  +   0.037 * max(0,       0.13 - MEAN_NDVI) * max(0,         13 -    GDDu_7)
  +     0.6 * max(0,       0.13 - MEAN_NDVI) * max(0,    SEASON2 -         0)
  - 6.9e-05 * max(0,     SRAD_6 -       392) * max(0,    SEASON2 -         0)
```

B.3 Phenoregion Three

The phenomodel of phenoregion three has an R^2 of 95.40% and standard error of 0.0351. Table B.3 shows the importance of the independent variables. Figure B.3 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion three is as follows:

```
NDVI =
  0.51
  +     0.8 * max(0,      aNDVI -      0.48)
  -    0.81 * max(0,       0.48 -     aNDVI)
  -  0.0034 * max(0,        8.2 -       GDD)
  +     1.3 * max(0, MEAN_NDVI -      0.14)
  -     1.8 * max(0,       0.14 - MEAN_NDVI)
  + 0.00066 * max(0,    SRAD_10 -       570)
  - 3.9e-05 * max(0,        570 -    SRAD_10)
  - 0.00025 * max(0,    AGDD_10 -        22)
  +  0.0045 * max(0,         22 -    AGDD_10)
  -   0.012 * max(0,       0.48 -     aNDVI) * max(0,        GDD -        3.3)
  -   0.019 * max(0,       0.48 -     aNDVI) * max(0,        3.3 -        GDD)
  +   0.015 * max(0,      aNDVI -      0.48) * max(0,     TMIN_3 -        5.9)
  +  0.0079 * max(0,      aNDVI -      0.48) * max(0,        5.9 -     TMIN_3)
  +   0.046 * max(0,      aNDVI -      0.48) * max(0,        1.7 -    PRCP_20)
  + 0.00038 * max(0,        8.2 -       GDD) * max(0,        LAT -         38)
  - 0.00082 * max(0,        8.2 -       GDD) * max(0,         38 -        LAT)
  +   3e-05 * max(0,        GDD -       7.1) * max(0,    AGDD_10 -         22)
  + 4.7e-05 * max(0,        7.1 -       GDD) * max(0,    AGDD_10 -         22)
  -  0.0014 * max(0,        LAT -        40) * max(0,         22 -    AGDD_10)
  +  0.0019 * max(0,        LAT -        41) * max(0,         22 -    AGDD_10)
  - 0.00036 * max(0,         41 -       LAT) * max(0,         22 -    AGDD_10)
  - 0.00016 * max(0,     GDDu_2 -       3.5) * max(0,         22 -    AGDD_10)
  - 0.00032 * max(0,        3.5 -     GDDu_2) * max(0,         22 -    AGDD_10)
  -   0.017 * max(0,     GDDu_7 -        12) * max(0,    AGDD_10 -         22)
  - 1.3e-05 * max(0,         12 -     GDDu_7) * max(0,    AGDD_10 -         22)
  - 3.5e-06 * max(0,     SRAD_8 -       414) * max(0,         22 -    AGDD_10)
  - 9.2e-06 * max(0,        414 -     SRAD_8) * max(0,         22 -    AGDD_10)
  - 0.00096 * max(0,         22 -    AGDD_10) * max(0,    SEASON2 -         0)
```

Table B.3 Importance of entered independent variables in phenoregion three

	nsubsets	gcv	rss
aNDVI	28	100	100
AGDD_10	26	15.4	15.5
SEASON2	25	10.4	10.6
SRAD_8	24	9.4	9.6
MEAN_NDVI	23	8.2	8.4
GDDu_2	21	6.7	7
PRCP_20	19	5.9	6.2
GDD	18	5.5	5.8
GDDu_7	15	4.9	5.2
LAT	14	4.5	4.8
SRAD_10	9	2.9	3.2
TMIN_3	8	2.4	2.7

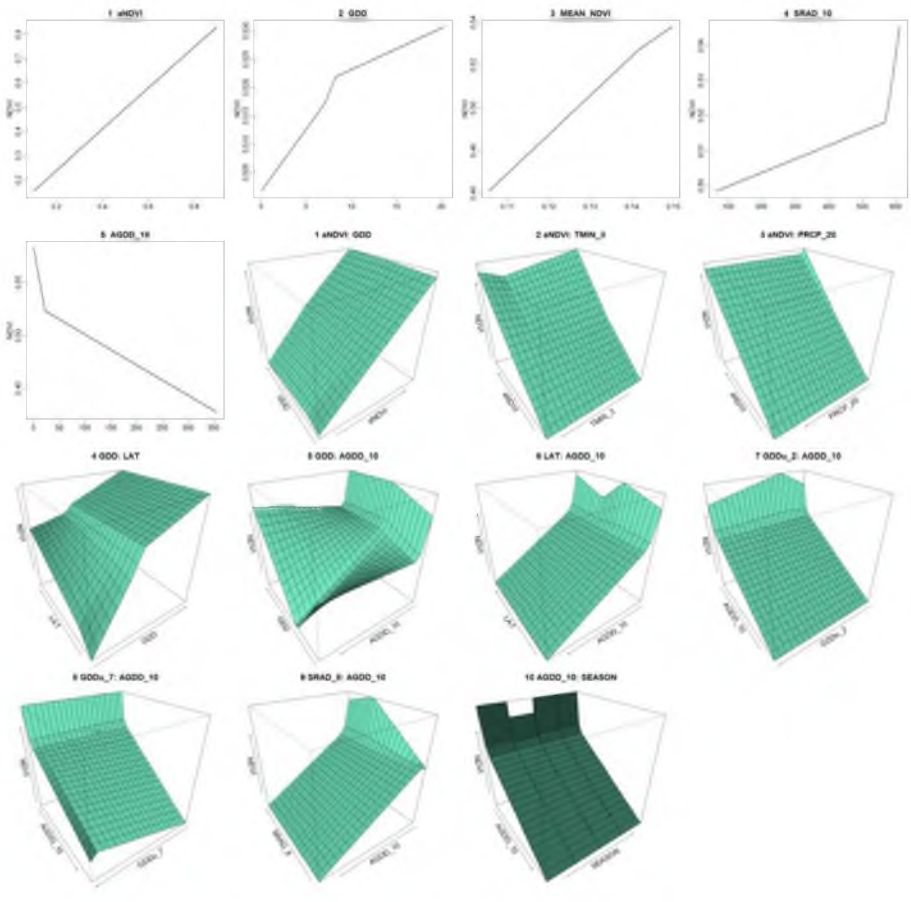


Figure B.3 Geographic relationships in phenoregion three.

B.4 Phenoregion Four

The phenomodel of phenoregion four has an R^2 of 96.13% and standard error of 0.0335. Table B.4 shows the importance of the independent variables. Figure B.4 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion four is as follows:

```
NDVI =
  0.24
  +    0.82 * max(0,      aNDVI -      0.23)
  -    0.91 * max(0,       0.23 -     aNDVI)
  -    0.11 * max(0,      aNDVI -      0.61)
  +  0.0084 * max(0,         16 -   AGDD_10)
  -    0.03 * max(0,    SEASON4 -         0)
  +       5 * max(0,      aNDVI -      0.23) * max(0, MEAN_NDVI -      0.13)
  -      13 * max(0,      aNDVI -      0.23) * max(0,      0.13 - MEAN_NDVI)
  +  0.0029 * max(0,      aNDVI -      0.23) * max(0,        11 -    GDDu_4)
  + 0.00077 * max(0,      aNDVI -      0.23) * max(0,   AGDD_10 -       116)
  +   0.001 * max(0,      aNDVI -      0.23) * max(0,       116 -   AGDD_10)
  -  0.0085 * max(0,      aNDVI -      0.34) * max(0,        16 -   AGDD_10)
  -  0.0064 * max(0,       0.34 -     aNDVI) * max(0,        16 -   AGDD_10)
  + 5.2e-05 * max(0,       TMIN -       1.4) * max(0,   AGDD_10 -        16)
  + 5.6e-06 * max(0,        1.4 -      TMIN) * max(0,   AGDD_10 -        16)
  -   3e-04 * max(0,        2.3 -      PRCP) * max(0,        16 -   AGDD_10)
  +   0.043 * max(0,    GS_TMIN -         5) * max(0,        16 -   AGDD_10)
  - 0.00034 * max(0,          5 -    GS_TMIN) * max(0,        16 -   AGDD_10)
  -    0.12 * max(0, MEAN_NDVI -      0.13) * max(0,        16 -   AGDD_10)
  -  0.0019 * max(0,       LAT -        41) * max(0,        16 -   AGDD_10)
  - 0.00041 * max(0,        41 -       LAT) * max(0,        16 -   AGDD_10)
  +  0.0036 * max(0,       LAT -        42) * max(0,        16 -   AGDD_10)
  - 0.00024 * max(0,    GDDu_2 -         2) * max(0,        16 -   AGDD_10)
  - 0.00051 * max(0,         2 -    GDDu_2) * max(0,        16 -   AGDD_10)
  - 7.7e-06 * max(0,    SRAD_6 -       479) * max(0,        16 -   AGDD_10)
  - 6.2e-06 * max(0,       479 -    SRAD_6) * max(0,        16 -   AGDD_10)
  - 5.6e-05 * max(0,   AGDD_10 -        16) * max(0,       2.4 -   PRCP_16)
```

Table B.4 Importance of entered independent variables in phenoregion four

	nsubsets	gcv	rss
aNDVI	26	100	100
AGDD_10	25	21.3	21.4
GDDu_2	23	11.2	11.3
SRAD_6	21	9.5	9.7
MEAN_NDVI	21	9.3	9.4
LAT	18	7.5	7.6
SEASON4	14	5.7	5.9
PRCP_16	13	5.3	5.4
GS_TMIN	11	4.6	4.7
PRCP	8	3.6	3.7
GDDu_4	7	3.2	3.4
TMIN	6	2.8	3

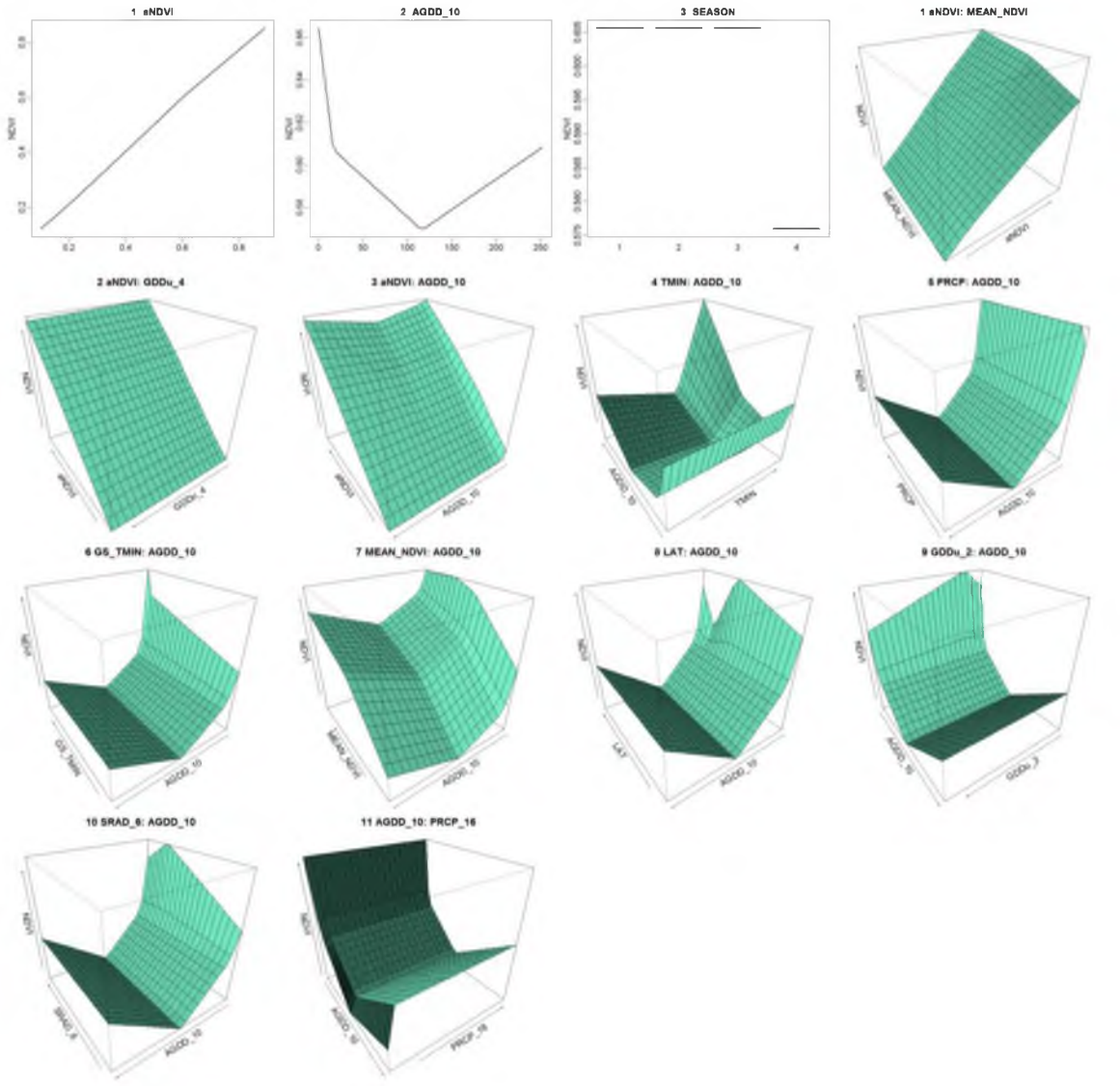


Figure B.4 Geographic relationships in phenoregion four.

B.5 Phenoregion Five

The phenomodel of phenoregion five has an R^2 of 93.53% and standard error of 0.0174. Table B.5 shows the importance of the independent variables. Figure B.5 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion five is as follows:

```
NDVI =  0.21
  +    0.81 * max(0,      aNDVI -      0.18)
  +    0.25 * max(0,       0.18 -     aNDVI)
  +     2.7 * max(0, MEAN_NDVI -      0.13)
  -     2.2 * max(0,       0.13 - MEAN_NDVI)
  - 0.00013 * max(0, MEAN_PRCP -       623)
  - 3.4e-05 * max(0,        623 - MEAN_PRCP)
  + 0.00097 * max(0,        LAT -        38)
  -  0.0061 * max(0,         38 -       LAT)
  - 0.00023 * max(0,     PRCP_7 -       5.9)
  - 0.00067 * max(0,        5.9 -    PRCP_7)
  +  0.0043 * max(0,    SEASON2 -         0)
  -    0.02 * max(0,      aNDVI -      0.18) * max(0, MEAN_TMIN -        -2)
  -     162 * max(0,       0.18 -     aNDVI) * max(0, MEAN_NDVI -      0.12)
  - 0.00094 * max(0,       0.18 -     aNDVI) * max(0, MEAN_PRCP -       463)
  -   0.003 * max(0,       0.18 -     aNDVI) * max(0,       463 - MEAN_PRCP)
  -    0.57 * max(0,       0.18 -     aNDVI) * max(0,       LAT -        41)
  -   0.054 * max(0,       0.18 -     aNDVI) * max(0,        41 -       LAT)
  +   8e-05 * max(0,      aNDVI -      0.18) * max(0,   AGDD_10 -        52)
  +  0.0011 * max(0,      aNDVI -      0.18) * max(0,        52 -   AGDD_10)
  + 7.7e-06 * max(0,       TMIN -      -9.1) * max(0, MEAN_PRCP -       623)
  + 8.2e-06 * max(0,       -9.1 -      TMIN) * max(0, MEAN_PRCP -       623)
  + 2.4e-05 * max(0, MEAN_PRCP -       388) * max(0,        38 -       LAT)
  + 4.5e-05 * max(0,        388 - MEAN_PRCP) * max(0,        38 -       LAT)
  + 5.1e-06 * max(0, MEAN_PRCP -       623) * max(0,    PRCP_5 -        13)
  + 3.6e-06 * max(0, MEAN_PRCP -       623) * max(0,        13 -    PRCP_5)
  + 1.1e-05 * max(0, MEAN_PRCP -       623) * max(0,   PRCP_16 -       7.8)
  + 3.3e-06 * max(0, MEAN_PRCP -       623) * max(0,       7.8 -   PRCP_16)
  + 3.7e-06 * max(0,        LAT -        38) * max(0,    AGDD_4 -        97)
  + 7.7e-05 * max(0,        LAT -        38) * max(0,        97 -    AGDD_4)
  +   4e-04 * max(0,     PRCP_2 -       9.3) * max(0,   SEASON2 -         0)
```

B.6 Phenoregion Six

The phenomodel of phenoregion six has an R^2 of 97.22% and standard error of 0.0281. Table B.6 shows the importance of the independent variables included in the phenomodel of phenoregion six. Figure B.6 illustrates the relationships between NDVI and independent variables in phenoregion six. The mathematical equation of the phenomodel in phenoregion six is as follows:

Table B.5 Importance of entered independent variables in phenoregion five

	nsubsets	gcv	rss
aNDVI	30	100	100
MEAN_PRCP	29	15.2	15.3
LAT	28	11.1	11.3
AGDD_4	28	11.1	11.3
MEAN_NDVI	27	10.2	10.4
MEAN_TMIN	26	9.1	9.2
SEASON2	23	6.4	6.6
TMIN	20	5.2	5.4
AGDD_10	19	4.7	5
PRCP_2	18	4.3	4.6
PRCP_7	16	3.6	3.8
PRCP_16	12	2.6	2.8
PRCP_5	10	2	2.2

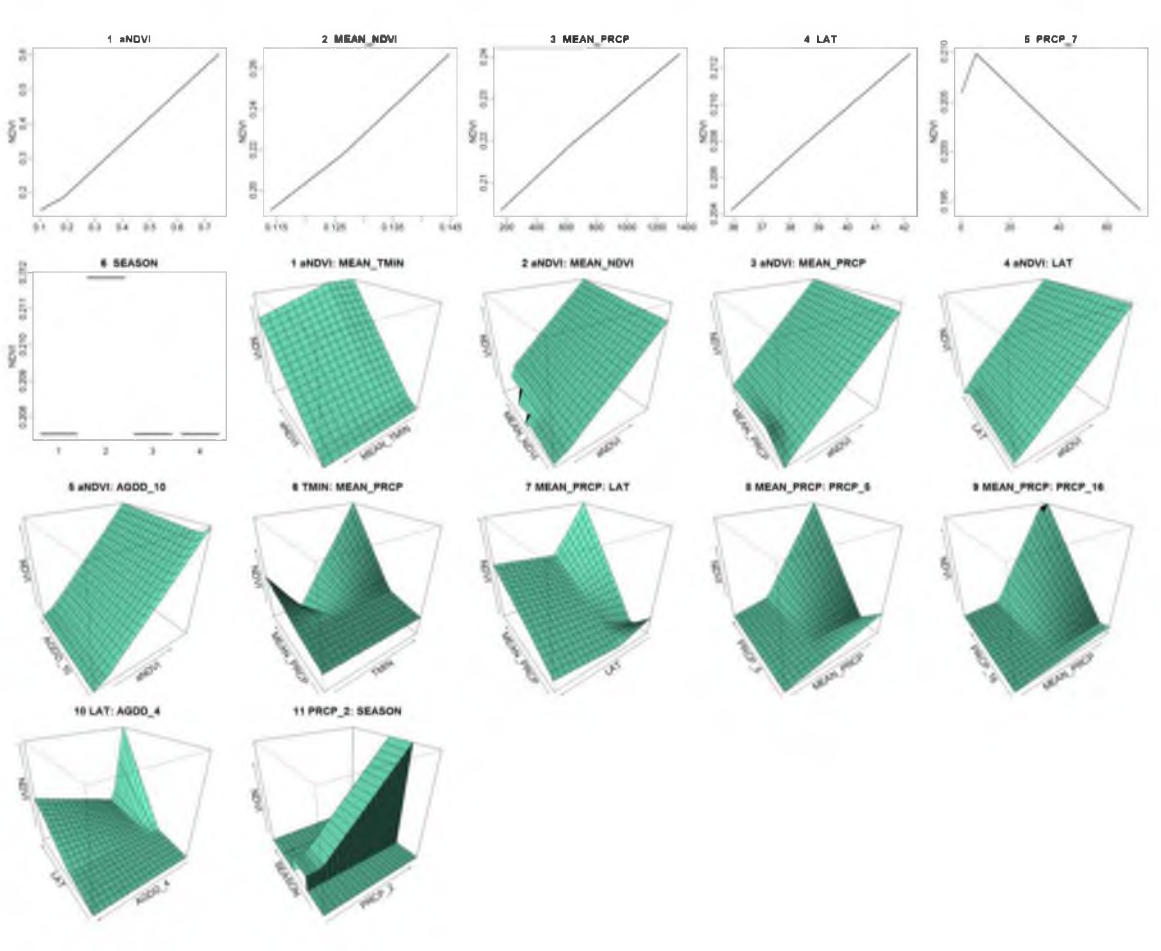


Figure B.5 Geographic relationships in phenoregion five.

Table B.6 Importance of entered independent variables in phenoregion six

	nsubsets	gcv	rss
aNDVI	19	100	100
AGDD_10	17	10.1	10.1
GS_PRCP	16	8.1	8.1
MEAN_NDVI	15	6	6
MEAN_TMAX	14	5.5	5.5
SEASON2	13	5.1	5.1
SRAD_4	12	4.6	4.6
TMIN_3	10	3.4	3.5
GDDu	7	2.2	2.2

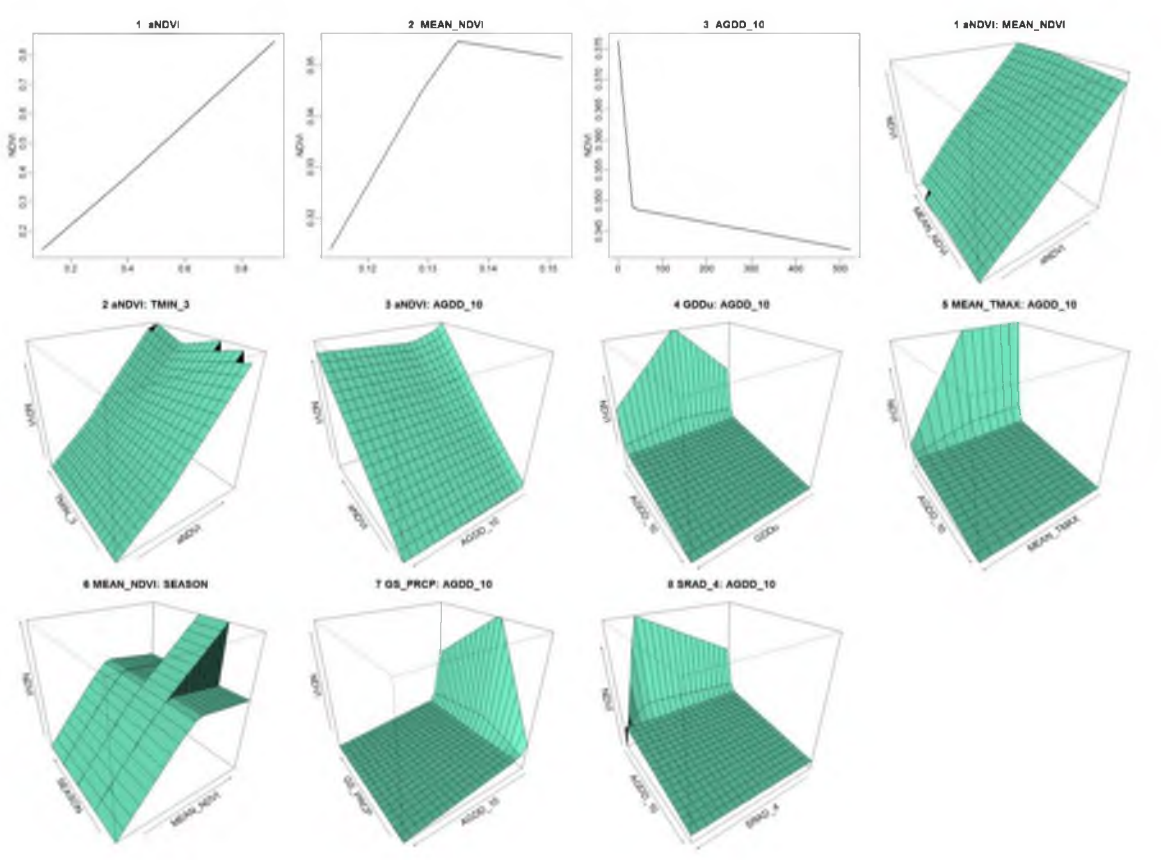


Figure B.6 Geographic relationships in phenoregion six.

```
NDVI =
  0.38
  +    0.81 * max(0,      aNDVI -      0.38)
  -    0.81 * max(0,       0.38 -      aNDVI)
  +    0.19 * max(0, MEAN_NDVI -       0.13)
  -     2.1 * max(0,       0.13 - MEAN_NDVI)
  - 1.4e-05 * max(0,   AGDD_10 -         32)
  +  0.0019 * max(0,         32 -   AGDD_10)
  -      12 * max(0,       0.38 -      aNDVI) * max(0, MEAN_NDVI -       0.13)
  +    0.01 * max(0,      aNDVI -       0.38) * max(0,    TMIN_3 -        3.9)
  +  0.0086 * max(0,      aNDVI -       0.38) * max(0,       3.9 -     TMIN_3)
  + 0.00018 * max(0,      aNDVI -       0.38) * max(0,   AGDD_10 -        150)
  + 0.00077 * max(0,      aNDVI -       0.38) * max(0,       150 -    AGDD_10)
  - 7.6e-05 * max(0,       GDDu -        7.7) * max(0,         32 -    AGDD_10)
  - 6.1e-05 * max(0,        7.7 -       GDDu) * max(0,         32 -    AGDD_10)
  - 0.00019 * max(0, MEAN_TMAX -         12) * max(0,         32 -    AGDD_10)
  +     1.7 * max(0, MEAN_NDVI -       0.13) * max(0,   SEASON2 -          0)
  - 1.9e-05 * max(0,   GS_PRCP -         48) * max(0,         32 -    AGDD_10)
  - 5.6e-05 * max(0,         48 -   GS_PRCP) * max(0,         32 -    AGDD_10)
  -   1e-05 * max(0,    SRAD_4 -        498) * max(0,         32 -    AGDD_10)
  -   1e-06 * max(0,        498 -    SRAD_4) * max(0,         32 -    AGDD_10)
```

B.7 Phenoregion Seven

The phenomodel of phenoregion seven has an R^2 of 96.23% and standard error of 0.0339. Table B.7 shows the importance of the independent variables included in the phenomodel of phenoregion seven. Figure B.7 illustrates the relationships between NDVI and independent variables in phenoregion seven. The mathematical equation of the phenomodel in phenoregion seven is as follows:

```
NDVI =
  0.28
  +     0.9 * max(0,      aNDVI -      0.24)
  -    0.56 * max(0,       0.24 -      aNDVI)
  + 0.00012 * max(0,       PRCP -        16)
  - 0.00053 * max(0,         16 -       PRCP)
  + 0.00055 * max(0,       SRAD -       550)
  - 0.00012 * max(0,        550 -      SRAD)
  +     1.3 * max(0, MEAN_NDVI -      0.14)
  -     1.4 * max(0,       0.14 - MEAN_NDVI)
  - 2.8e-05 * max(0,   AGDD_10 -         26)
  +  0.0023 * max(0,         26 -   AGDD_10)
  -    0.44 * max(0,       0.24 -      aNDVI) * max(0,       GAP4 -          0)
  - 0.00018 * max(0,       -1.6 -       TMIN) * max(0,         26 -    AGDD_10)
  - 1.5e-06 * max(0,       SRAD -        418) * max(0,    AGDD_10 -         26)
  + 3.4e-07 * max(0,        418 -       SRAD) * max(0,    AGDD_10 -         26)
  -   0.001 * max(0,       SRAD -        550) * max(0,    SEASON2 -          0)
  -    0.08 * max(0, MEAN_NDVI -       0.13) * max(0,         26 -    AGDD_10)
  -   0.041 * max(0,       0.13 - MEAN_NDVI) * max(0,         26 -    AGDD_10)
  + 1.2e-05 * max(0,   GS_PRCP -         41) * max(0,         26 -    AGDD_10)
  -   1e-04 * max(0,         41 -   GS_PRCP) * max(0,         26 -    AGDD_10)
  - 0.00014 * max(0,    GDDu_2 -        4.9) * max(0,         26 -    AGDD_10)
  -   1e-04 * max(0,        4.9 -    GDDu_2) * max(0,         26 -    AGDD_10)
```

Table B.7 Importance of entered independent variables in phenoregion seven

	nsubsets	gcv	rss
aNDVI	21	100	100
AGDD_10	20	15.9	15.9
SRAD	17	7.1	7.2
MEAN_NDVI	17	7.1	7.2
TMIN	16	6.6	6.7
GS_PRCP	15	6.2	6.3
GDDu_2	15	6	6.1
PRCP	13	4.3	4.4
GAP4	12	3.9	4
SEASON2	11	3.4	3.5

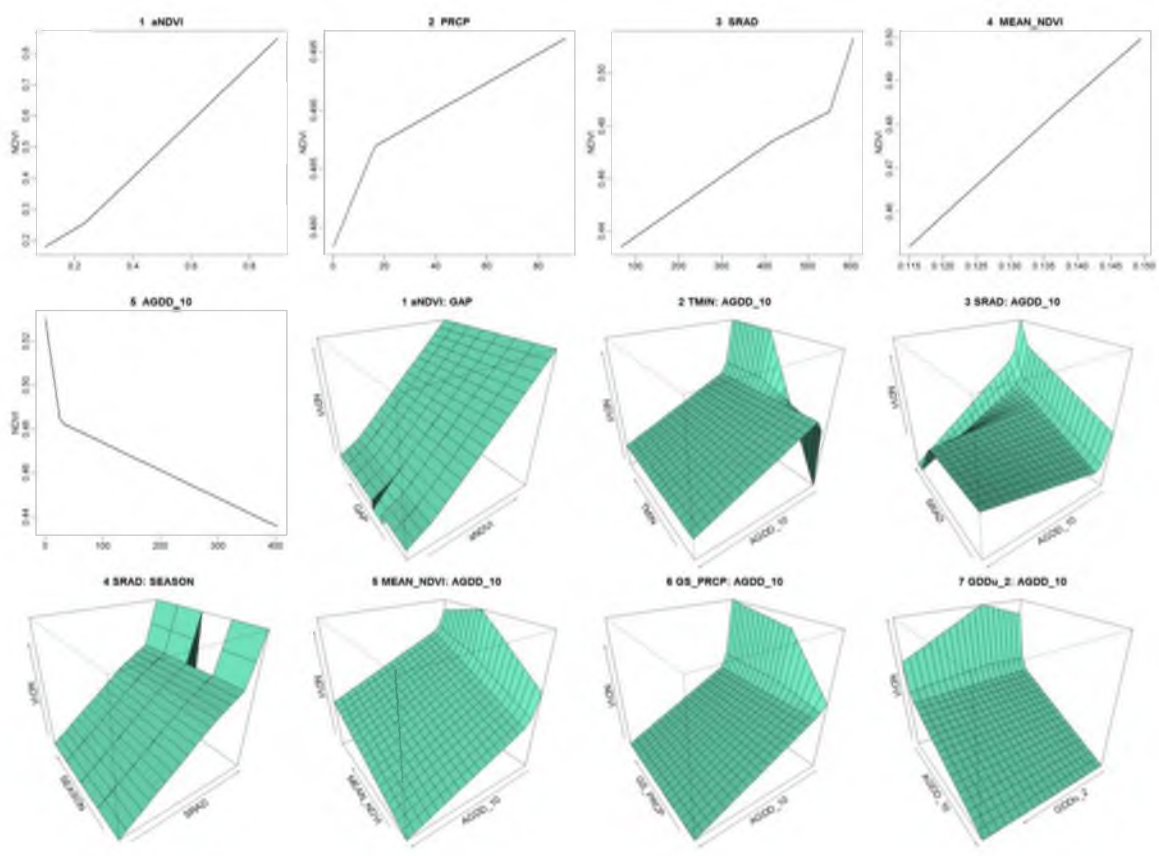


Figure B.7 Geographic relationships in phenoregion seven.

B.8 Phenoregion Eight

The phenomodel of phenoregion eight has an R^2 of 95.95% and standard error of 0.0222. Table B.8 shows the importance of the independent variables. Figure B.8 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion eight is as follows:

```
NDVI =
  0.2
  +     0.8 * max(0,      aNDVI -       0.18)
  -    0.75 * max(0,       0.18 -      aNDVI)
  +    0.85 * max(0, MEAN_NDVI -       0.12)
  -     1.9 * max(0,       0.12 - MEAN_NDVI)
  + 0.00011 * max(0, MEAN_PRCP -        325)
  -   2e-05 * max(0,        325 - MEAN_PRCP)
  -  0.0088 * max(0,        LAT -         42)
  -  0.0025 * max(0,         42 -        LAT)
  +   0.053 * max(0,      aNDVI -       0.18) * max(0,  TEMP_STD -         12)
  +   0.019 * max(0,      aNDVI -       0.18) * max(0,         12 -  TEMP_STD)
  + 3.8e-05 * max(0,      aNDVI -       0.18) * max(0,    AGDD_6 -         65)
  +  0.0017 * max(0,      aNDVI -       0.18) * max(0,        65 -    AGDD_6)
  +   0.016 * max(0,       0.12 - MEAN_NDVI) * max(0,        79 -   AGDD_10)
  +     1.4 * max(0, MEAN_NDVI -       0.12) * max(0,   SEASON2 -         0)
  +     1.2 * max(0, MEAN_NDVI -       0.12) * max(0,   SEASON3 -         0)
  -   5e-07 * max(0, MEAN_PRCP -        325) * max(0,   AGDD_10 -         0)
```

B.9 Phenoregion Nine

The phenomodel of phenoregion nine has an R^2 of 96.65% and standard error of 0.0261. Table B.9 shows the importance of the independent variables. Figure B.9 illustrates the relationships between NDVI and independent variables. The mathematical equation of the phenomodel in phenoregion nine is as follows:

```
NDVI =
  0.24
  +    0.74 * max(0,      aNDVI -       0.22)
  -    0.76 * max(0,       0.22 -      aNDVI)
  + 0.00097 * max(0,         17 - MEAN_TMAX)
  +     2.8 * max(0, MEAN_NDVI -       0.14)
  -     1.6 * max(0,       0.14 - MEAN_NDVI)
  -      45 * max(0,       0.22 -      aNDVI) * max(0, MEAN_NDVI -       0.13)
  + 0.00014 * max(0,      aNDVI -       0.22) * max(0,    AGDD_6 -         65)
  +   8e-04 * max(0,      aNDVI -       0.22) * max(0,        65 -    AGDD_6)
  + 8.6e-05 * max(0,      aNDVI -       0.22) * max(0,  AGDDu_10 -        245)
  + 0.00036 * max(0,      aNDVI -       0.22) * max(0,       245 -  AGDDu_10)
  +   3e-04 * max(0,      aNDVI -       0.22) * max(0,   PRCP_19 -        4.5)
  +  0.0064 * max(0,      aNDVI -       0.22) * max(0,       4.5 -   PRCP_19)
  +   0.066 * max(0,      aNDVI -       0.22) * max(0,   SEASON2 -         0)
  + 8.2e-05 * max(0,         17 - MEAN_TMAX) * max(0,        59 -  AGDDu_10)
  -    0.38 * max(0, MEAN_NDVI -       0.14) * max(0,   TMIN_10 -        6.3)
  -    0.15 * max(0, MEAN_NDVI -       0.14) * max(0,        6.3 -   TMIN_10)
```

Table B.8 Importance of entered independent variables in phenoregion eight

	nsubsets	gcv	rss
aNDVI	16	100	100
AGDD_6	14	8.5	8.5
MEAN_NDVI	13	6.8	6.8
MEAN_PRCP	12	5.5	5.6
SEASON2	11	4.7	4.8
AGDD_10	10	4.6	4.7
SEASON3	10	4.3	4.3
LAT	7	2.7	2.8
TEMP_STD	5	1.6	1.6

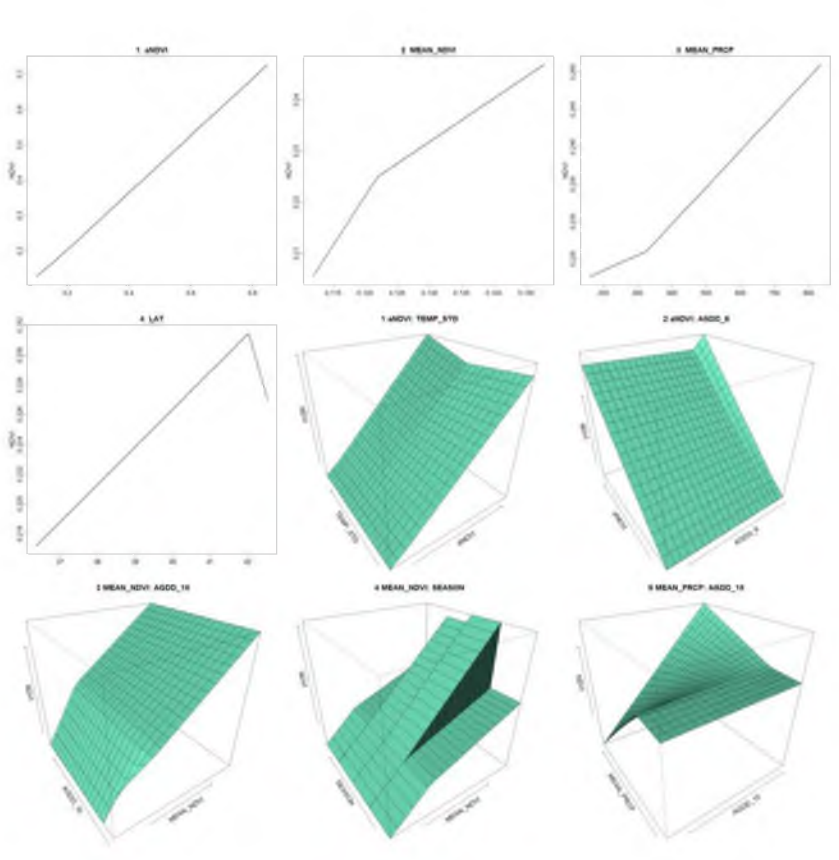


Figure B.8 Geographic relationships in phenoregion eight.

Table B.9 Importance of entered independent variables in phenoregion nine

	nsubsets	gcv	rss
aNDVI	16	100	100
MEAN_TMAX	14	8.4	8.4
AGDDu_10	14	8.4	8.4
MEAN_NDVI	13	5.6	5.6
SEASON2	11	3.5	3.6
PRCP_19	8	2.3	2.3
TMIN_10	7	1.8	1.8
AGDD_6	5	1.3	1.4

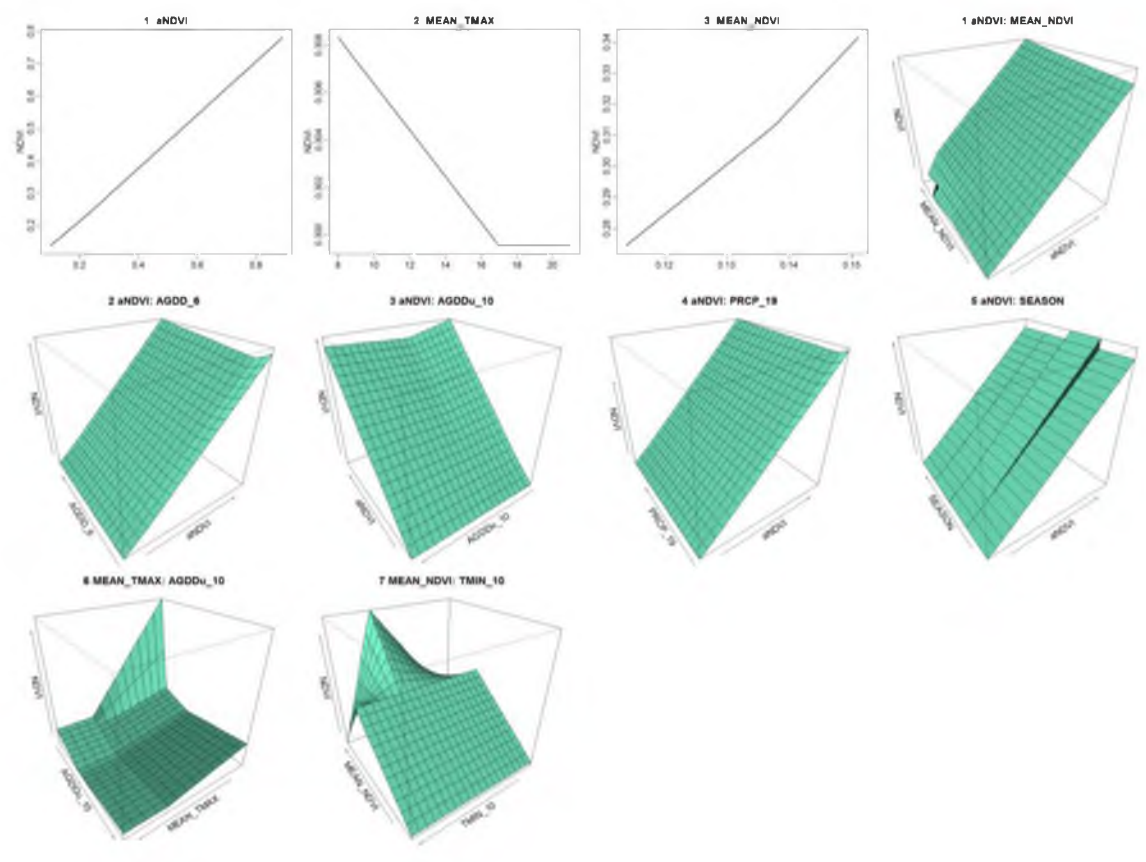


Figure B.9 Geographic relationships in phenoregion nine.

REFERENCES

Acevedo, M.F., 2012, Data analysis and statistics for geography, environmental science, and engineering, CRC PressINC.

Ahrens, C.D., 2007, Meteorology today: An introduction to weather, climate, and the environment, Thomson/Brooks/Cole.

Arthur, D., and Vassilvitskii, S., 2007, *k*-means++: The advantages of careful seeding, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms: New Orleans, Louisiana, Society for Industrial and Applied Mathematics, p. 1027-1035.

Atkinson, G.F., 1904, Relation of plants to environment (or Plant ecology): Outlines of a course of lectures delivered in the summer school of Cornell university 1903 and 1904, The author.

Badeck, F.-W., Bondeau, A., Böttcher, K., Doktor, D., Lucht, W., Schaber, J., and Sitch, S., 2004, Responses of spring phenology to climate change: New Phytologist, v. 162, no. 2, p. 295-309.

Bailey, R.G., 1983, Delineation of ecosystem regions: Environmental Management, v. 7, no. 4, p. 365-373.

Batanouny, K.H., 2001, Plants in the deserts of the Middle East, Springer.

Bender, E.A., 2000, An introduction to mathematical modeling, Courier Dover Publications.

Black, P.E., 1996, Watershed hydrology, Ann Arbor Press.

Blackman, G.E., and Rutter, A.J., 1950, Physiological and ecological studies in the analysis of plant environment: V. An assessment of the factors controlling the distribution of the bluebell (Scilla non-scripta) in different communites: Annals of Botany, v. 14, no. 4, p. 487-520.

BLM (Bureau of Land Management), 2011, Fact sheet on the BLM's management of Livestock grazing, v. 2011, no. July 8.

Blodget, L., 1857, Climatology of the United States: London, J.B. Lippincott and CO.

Campbell, J.B., 2002, Introduction to remote sensing (3rd ed.) The Guilford Press, New York, NY.

Campbell, R.K., 1974, Use of phenology for examining provenance transfers in reforestation of douglas-fir: Journal of Applied Ecology, v. 11, no. 3, p. 1069-1080.

Campbell, R.K., and Sugano, A.I., 1975, Phenology of bud burst in douglas-fir related to provenance, photoperiod, chilling, and flushing temperature: Botanical Gazette, v. 136, no. 3, p. 290-298.

Cardot, H., Maisongrande, P., and Faivre, R., 2008, Varying-time random effects models for longitudinal data: Unmixing and temporal interpolation of remote-sensing data: Journal of Applied Statistics, v. 35, no. 8, p. 827-846.

Chandrasekar, K., Sesha Sai, M.V.R., Jeyaseelan, A.T., Dwivedi, R.S., and Roy, P.S., 2006, Vegetation response to rainfall as monitored by NOAA-AVHRR: Current Science, v. 91, no. 12, p. 1626-1633.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L., 2004, A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter: Remote Sensing of Environment, v. 91, no. 3–4, p. 332-344.

Chmielewski, F.M., and Rotzer, T., 2002, Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes: Climate Research, v. 19.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I., 1990, STL: A seasonal-trend decomposition procedure based on loess: Journal of Official Statistics, v. 6, no. 1, p. 3-73.

Crawley, M., 2009, Plant ecology, Wiley.

Cui, L., Shi, J., Yang, Y., and Fan, W., 2009, Ten-day response of vegetation NDVI to the variations of temperature and precipitation in Eastern China: Acta Geographica Sinica, v. 64, no. 7, p. 850-860.

Dahl, E., and Birks, J., 2007, The phytogeography of northern Europe: British Isles, Fennoscandia, and adjacent areas, Cambridge University Press.

Dahlgren, J., Oksanen, L., Sjödin, M., and Olofsson, J., 2007, Interactions between gray-sided voles (Clethrionomys rufucanus) and bilberry (Vaccinium myrtillus), their main winter food plant: Oecologia, v. 152, no. 3, p. 525-532.

Dai, A., Trenberth, K.E., and Karl, T.R., 1999, Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range: Journal of Climate, v. 12, no. 8, p. 2451-2473.

Dale, M.R.T., 2000, Spatial pattern analysis in plant ecology, Cambridge University Press.

Davenport, M.L., and Nicholson, S.E., 1993, On the relation between rainfall and the normalized difference vegetation index for diverse vegetation types in East Africa: International Journal of Remote Sensing, v. 14, no. 12, p. 2369-2389.

Decker, J.P., 1944, Effect of temperature on photosynthesis and respiration in red and loblolly pines: Plant Physiol., no. 19, p. 679-688.

Desalew, T., Tegegne, A., Nigatu, L., and Teka, W., 2010, Rangeland condition and feed resources in Metema District, North Gondar Zone, Amhara Region, Ethiopia, IPMS (Improving Productivity and Market Success) of Ethiopian Farmers Project Working Paper 25: Nairobi, Kenya.

Dombeck, M.P., Wood, C.A., and Williams, J.E., 2003, From conquest to conservation: Our public lands legacy, Island Press.

Estrella, N., Sparks, T.H., and Menzel, A., 2007, Trends and temperature response in the phenology of crops in Germany: Global Change Biology, v. 13, no. 8, p. 1737-1747.

Feaumur, R.A.E., 1735, Temperature observations in paris during the year 1735, and the climatic analogue studies of l'Isle de France, Algeria and some islands of America: Mem. Acad. Sci.

Fisher, J.I., and Mustard, J.F., 2007, Cross-scalar satellite phenology from ground, Landsat, and MODIS data: Remote Sensing of Environment, v. 109, no. 3, p. 261-273.

Fisher, J.I., Richardson, A.D., and Mustard, J.F., 2007, Phenology model from surface meteorology does not capture satellite-based greenup estimations: Global Change Biology, v. 13, no. 3, p. 707-721.

Fitter, A.H., and Fitter, R.S.R., 2002, Rapid changes in flowering time in British plants: Science, v. 296, no. 5573, p. 1689-1691.

Fitter, A.H., Fitter, R.S.R., Harris, I.T.B., and Williamson, M.H., 1995, Relationships between first flowering date and temperature in the flora of a locality in central England: Functional Ecology, v. 9, no. 1, p. 55-60.

Fitter, A.H., and Hay, R.K.M., 2001, Environmental physiology of plants, Elsevier Science.

Foth, H.D., and Ellis, B.G., 1997, Soil fertility, Lewis Publ.

Francis, L., 2003, Martian Chronicles: Is MARS better than neural networks?, *in* Casualty Actuarial Society Forum, p. 75-102.

Friedman, J.H., 1991, Multivariate adaptive regression splines: The Annals of Statistics, v. 19, no. 1, p. 1-67.

Gao, B., 1996, NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space: Remote Sensing of Environment, v. 58, no. 3, p. 257-266.

Garner, W.W., and Allard, H.A., 1920, Effect of relative length of day and night and other factors of the environment on growth and reproduction in plants: Monthly Weather Review, v. 48, no. 7, p. 415-415.

Garner, W.W., and Allard, H.A., 1923, Further studies in photoperiodism, the response of the plant to relative length of day and night: J. Agric. Res., no. 23, p. 871-920.

Garner, W.W., and Allard, H.A., 1930, Photoperiodic response of soybeans in relation to temperature and other environmental factors: Washington.

Gentry, A.H., and Emmons, L.H., 1987, Geographical variation in fertility, phenology, and composition of the understory of neotropical forests: Biotropica, v. 19, no. 3, p. 216-227.

Gibson, P.J., and Power, C.H., 2000, Introductory remote sensing: Digital image processing and applications, Routledge.

Grayson, R.B., and Blösch, G., 2001, Spatial patterns in catchment hydrology: Observations and modelling, Cambridge University Press.

Grice, A.C., and Hodgkinson, K.C., 2002, Global rangelands [electronic resource]: Progress and prospects, CABI Publishing.

Hargrove, W., and Hoffman, F., 2004, Potential of multivariate quantitative methods for delineation and visualization of ecoregions: Environmental Management, v. 34, no. 0, p. S39-S60.

Hargrove, W., Hoffman, F., and Hessburg, P., 2006, Mapcurves: A quantitative method for comparing categorical maps: Journal of Geographical Systems, v. 8, no. 2, p. 187-208.

Heady, H., and Child, R.D., 1999, Rangeland ecology and management, Westview Press.

Henfrey, A., 1852, The vegetation of Europe, its conditions and causes.: London, J. van Vooret.

Hillman, W.S., 1969, Photoperiodism and vernalization, *in* Wilkins, M.B., ed., Physiology of plant growth and development: Maidenhead, Berkshire, England, McGraw-Hill Publishing Co. Ltd., p. 557-601.

Hodges, T., 1991, Predicting crop phenology, CRC Press.

Hudson, I.L., and Keatley, M.R., 2009, Phenological research: Methods for environmental and climate change analysis, Springer.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G., 2002, Overview of the radiometric and biophysical performance of the MODIS vegetation indices: Remote Sensing of Environment, v. 83, no. 1–2, p. 195-213.

Huete, A.R., 1988, A soil-adjusted vegetation index (SAVI): Remote Sensing of Environment, v. 25, no. 3, p. 295-309.

Jenkins, J.P., Braswell, B.H., Frolking, S.E., and Aber, J.A., 2002, Predicting spatial and interannual patterns of temperate forest springtime phenology in the Eastern U.S.: Geophysical Research Letters, no. 29, p. 54.

Ji, L., and Peters, A.J., 2004, Forecasting vegetation greenness with satellite and climate data: Geoscience and Remote Sensing Letters, IEEE, v. 1, no. 1, p. 3-6.

Ji, L., and Peters, A.J., 2007, Performance evaluation of spectral vegetation indices using a statistical sensitivity function: Remote Sensing of Environment, v. 106, no. 1, p. 59-65.

Ji, L., Wylie, B., Ramachandran, B., and Jenkerson, C., 2010, A comparative analysis of three different MODIS NDVI datasets for Alaska and adjacent Canada: Canadian Journal of Remote Sensing, v. 36, no. S1, p. S149-S167.

Jorgensen, S.E., 2009, Jorgensen's ecosystem ecology, Elsevier Science.

Junttila, O., Stushnoff, C., and Gusta, L.V., 1983, Dehardening in flower buds of saskatoon-berry, Amelanchier alnifolia, in relation to temperature, moisture content, and spring bud development: Canadian Journal of Botany, v. 61, no. 1, p. 164-170.

Justice, C.O., Holben, B.N., and Gwynne, M.D., 1986, Monitoring East African vegetation using AVHRR data: International Journal of Remote Sensing, v. 7, no. 11, p. 1453-1474.

Kaduk, J., and Heimann, M., 1996, A prognostic phenology scheme for global terrestrial carbon cycle models: Climate Research, v. 06, no. 1, p. 1-19.

Kai, K., Kainuma, M., Murakoshi, N., and Omasa, K., 1993, Potential effects on the phenological observations of plants by global warming in Japan: J. Agricult. Meteoral., v. 48, no. 5, p. 771-774.

Kawabata, A., Ichii, K., and Yamaguchi, Y., 2001, Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation: International Journal of Remote Sensing, v. 22, no. 7, p. 1377-1382.

Keele, L., and Kelly, N.J., 2006, Dynamic models for dynamic theories: The ins and outs of lagged dependent variables: Political Analysis, v. 14, no. 2, p. 186-205.

Klebs, G., 1913, Über das Verhåltnis der Aussenwelt zur Entwicklung der Pflanze: Sber. Akad. Wiss. Heidelberg, no. 5, p. 1-47.

Knapp, A.K., 1984, Water relations and growth of three grasses during wet and drought years in a tallgrass prairie: Oecologia, v. 65, no. 1, p. 35-43.

Koss, W.J., Owenby, J.R., Steurer, P.M., and Ezell, D.S., 1988, Freeze/Frost data: National Oceanic and Atmospheric Administration.

Küchler, A.W., 1964, Potential natural vegetation of the Conterminous United States: Soil Science, v. 99, no. 5, p. 356.

Lambers, H., F. Stuart Chapin, I., and Pons, T.L., 2008, Plant physiological ecology, Springer.

Larcher, W., 2003, Physiological plant ecology: Ecophysiology and stress physiology of functional groups, Springer.

Lindsey, A., and Newman, J.E., 1956, Use of official weather data in spring time-temperature analysis of an Indiana phenological record: Ecology, v. 37.

Liu, H., and Huete, A., 1995, A feedback based modification of the NDVI to minimize canopy background and atmospheric noise: Geoscience and Remote Sensing, IEEE Transactions on, v. 33, no. 2, p. 457-465.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., 2005, Geographic information systems and science, Wiley.

Lotsch, A., Friedl, M.A., Anderson, B.T., and Tucker, C.J., 2003, Coupled vegetation-precipitation variability observed from satellite and climate records: Geophysical Research Letters, v. 30, no. 14, p. 1774.

Ma, M., and Veroustraete, F., 2006, Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China: Advances in Space Research, v. 37, no. 4, p. 835-840.

Maczko, K.A., Bryant, L.D., Thompson, D.W., and Borchard, S.J., 2004, Putting the pieces together: Assessing social, ecological, and economic rangeland sustainability: Rangelands, v. 26, no. 3, p. 3-14.

McMahon, G., Gregonis, S.M., Waltman, S.W., Omernik, J.M., Thorson, T.D., Freeouf, J.A., Rorick, A.H., and Keys, J.E., 2001, Developing a spatial framework of common ecological regions for the Conterminous United States: Environmental Management, v. 28, no. 3, p. 293-316.

McMahon, T.A., Arenas, A.D., and Programme, I.H., 1982, Methods of computation of low streamflow: A contribution to the international hydrological programme, UNESCO.

Menzel, A., 2003, Plant phenological anomalies in Germany and their relation to air temperature and NAO: Climatic Change, v. 57, no. 3, p. 243-263.

Moilanen, A., Wilson, K.A., and Possingham, H., 2009, Spatial conservation prioritization: Quantitative methods and computational tools, Oxford University Press.

Mooney, H.A., and Billings, W.D., 1961, Comparative physiological ecology of Arctic and Alpine populations of Oxyria digyna: Ecological Monographs, v. 31, no. 1, p. 1-29.

Moulin, S., Kergoat, L., Viovy, N., and Dedieu, G., 1997, Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements: Journal of Climate, v. 10, no. 6, p. 1154-1170.

Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G., and Nemani, R.R., 1997, Increased plant growth in the northern high latitudes from 1981 to 1991: Nature, v. 386, no. 6626, p. 698-702.

NASA (National Aeronautics and Space Administration), 2009, About MODIS, v. 2009, no. April 10.

NASA Ames Ecological Forecasting Lab, 2009, Ecocast monitoring, modeling and forecasting ecosystem change, v. 2009, no. April 25.

National Geospatial Intelligence Agency (NGIA, f.N.I.a.M.A., 1996, Digital terrain elevation data level 0.

National Research Council: Committee on Rangeland Classification, 1994, Rangeland health: New methods to classify, inventory, and monitor rangelands, National Academies Press.

Nelson, R.H., 1995, Sustainability, efficiency, and god: Economic values and the sustainability debate: Annual Review of Ecology and Systematics, v. 26, p. 135-154.

Nemani, R.R., and Running, S.W., 1989, Estimation of regional surface resistance to evapotranspiration from NDVI and Thermal-IR AVHRR data: Journal of Applied Meteorology, v. 28, no. 4, p. 276-284.

Neteler, M., Roiz, D., Rocchini, D., Castellani, C., and Rizzoli, A., 2011, Terra and Aqua satellites track tiger mosquito invasion: Modelling the potential distribution of Aedes albopictus in north-eastern Italy: International Journal of Health Geographics, v. 10, no. 1, p. 1-14.

Olson, K.C., White, R.S., and Sindelar, B.W., 1985, Response of vegetation of the northern Great Plains to precipitation amount and grazing intensity: Journal of Range Management, v. 38, no. 4, p. 357-361.

Omernik, J.M., 1987, Map supplement: Ecoregions of the Conterminous United States: Annals of the Association of American Geographers, v. 77, no. 1, p. 118-125.

Park-Ono, H.S., Kawamura, T., and Yoshino, M., 1993, Relationships between flowering date of cherry blossom (Prunus yedoensis) and air temperature in East Asia, *in* Proceedings of the 13th International Congress of Biometeorology, Calgary, p. 207-220.

Parker, M.W., and Borthwick, H.A., 1939, Effect of variation in temperature during photoperiodic induction upon initiation of flower primordia in Biloxi soybean: Botanical Gazette, v. 101, no. 1, p. 145-167.

Peñuelas, J., Filella, I., Zhang, X., Llorens, L., Ogaya, R., Lloret, F., Comas, P., Estiarte, M., and Terradas, J., 2004, Complex spatiotemporal phenological shifts as a response to rainfall changes: New Phytologist, v. 161, no. 3, p. 837-846.

Pessarakli, M., 2002, Handbook of plant and crop stress, Second Edition, Taylor & Francis.

Pfafflin, J.R., and Ziegler, E.N., 2006, Encyclopedia of environmental science and engineering, Fifth Edition, Volumes One and Two, Taylor & Francis.

Pickup, G., Bastin, G.N., and Chewings, V.H., 1994, Remote-sensing-based condition assessment for nonequilibrium rangelands under large-scale commercial grazing: Ecological Applications, v. 4, no. 3, p. 497-517.

Prasad, A., Iverson, L., and Liaw, A., 2006, Newer classification and regression tree techniques: Bagging and random forests for ecological prediction: Ecosystems, v. 9, no. 2, p. 181-199.

Prins, H.H.T., and Loth, P.E., 1988, Rainfall patterns as background to plant phenology in Northern Tanzania: Journal of Biogeography, v. 15, no. 3, p. 451-463.

PRISM Climate Group, O.S.U., 2010, PRISM (Parameter-elevation Regressions on Independent Slopes Model) data set, http://prism.oregonstate.edu.

Pugnaire, F.I., and Valladares, F., 1999, Handbook of functional plant ecology, M. Dekker.

Rauzi, F., and Dobrenz, A.K., 1970, Seasonal variation of chlorophyll in Western wheatgrass and blue grama: Journal of Range Management, v. 23, p. 372-373.

Reed, B.C., and Brown, J.F., 2005, Trend analysis of time-series phenology derived from satellite data, *in* Analysis of multi-temporal remote sensing images, 2005 International Workshop on the, p. 166-168.

Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., and Ohlen, D.O., 1994, Measuring phenological variability from satellite imagery: Journal of Vegetation Science, v. 5, no. 5, p. 703-714.

Sayre, J.D., 1928, The development of chlorophyll in seddlings in different ranges of wave lengths of light: Plant Physiol., no. 3, p. 71-77.

Schultz, P.A., and Halpert, M.S., 1995, Global analysis of the relationships among a vegetation index, precipitation and land surface temperature: International Journal of Remote Sensing, v. 16, no. 15, p. 2755-2777.

Schulze, E.D., Beck, E., Müller-Hohenstein, K., Lawlor, D., Lawlor, K., and Lawlor, G., 2005, Plant ecology, Springer.

Schuster, W.S., Alles, D.L., and Mitton, J.B., 1989, Gene flow in limber pine: Evidence from pollination phenology and genetic differentiation along an elevational transect: American Journal of Botany, v. 76, no. 9, p. 1395-1403.

Schuur, E.A.G., 2003, Productivity and global climate revisited: The sensitivity of tropical forest growth to precipitation: Ecology, v. 84, no. 5, p. 1165-1170.

Schwartz, D., 2003, Phenology: An integrative environmental science, Springer.

Schwartz, M.D., Reed, B.C., and White, M.A., 2002, Assessing satellite-derived start-of-season measures in the conterminous USA: International Journal of Climatology, v. 22, no. 14, p. 1793-1805.

Schwarz, P.A., Fahey, T.J., and Dawson, T.E., 1997, Seasonal air and soil temperature effects on photosynthesis in red spruce (Picea rubens) saplings: Tree Physiology, v. 17, p. 187-194.

Service, N.R.C., 2011, Soil survey manual - Chapter One, v. 2011, no. January 10.

Sharma, P.D., 2005, Ecology and environment, Rastogi Publications.

Shigehara, K., Okamura, T., Nakayama, N., and Watanabe, F., 1991, Phenological observation data in Japan to be utilized as an indicator of climatic variation, *in* Proceedings of the International Converence on Climatic Impacts on the Environment and Society, Ibaraki, Japan, p. C1-C6.

Sirikul, N., 2007, Comparisons of MODIS vegetation index products with biophysical and flux tower measurements, The University of Arizona.

Smith, T.T.M., Shugart, H.H., and Woodward, F.I., 1997, Plant functional types: Their relevance to ecosystem properties and global change, Cambridge University Press.

Sparks, T.H., and Carey, P.D., 1995, The responses of species to climate over two centuries: An analysis of the Marsham phenological record, 1736-1947: Journal of Ecology, v. 83, no. 2, p. 321-329.

Sparks, T.H., Carey, P.D., and Combes, J., 1997, First leafing dates of trees in Surrey between 1947 and 1996: The London Naturalist, v. 76, p. 15-20.

Sparks, T.H., Jeffree, E.P., and Jeffree, C.E., 2000, An examination of the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK: International Journal of Biometeorology, v. 44, no. 2, p. 82-87.

Srivastava, L.M., 2002, Plant growth and development: Hormones and environment, Academic Press.

Steinley, D., 2003, Local optima in *k*-means clustering: What you don't know may hurt you: Psychological Methods, v. 8, no. 3, p. 294-304.

Strand, E., 1965, Forelesning i plantekultur: Norges landbrukshogskole, v. 73.

Suzuki, R., Tanaka, S., and Yasunari, T., 2000, Relationships between meridional profiles of satellite-derived vegetation index (NDVI) and climate over Siberia: International Journal of Climatology, v. 20, no. 9, p. 955-967.

Swaine, M.D., 1996, Rainfall and soil fertility as factors limiting forest species distributions in Ghana: Journal of Ecology, v. 84, no. 3, p. 419-428.

Tansley, A.G., 2010, An introduction to plant ecology, Discovery Publishing House Pvt. Limited.

Thomas, B., and Vince-Prue, D., 1996, Photoperiodism in plants, Elsevier Science.

Thompson, R., Shafer, S., Anderson, K., Strickland, L., Pelltier, R., Bartlein, P., and Kerwin, M., 2004, Topographic, bioclimatic, and vegetation characteristics of three ecoregion classification systems in North America: Comparisons along continent-wide transects: Environmental Management, v. 34, no. 0, p. S125-S148.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., and Cook, R.B., 2012, Daymet: Daily surface weather on a 1 km grid for North America,1980 - 2011: Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A., http://daymet.ornl.gov/.

Timmermans, M., 2010, Plant development, Elsevier Science.

Tournois, J., 1912, Influence de la lumière sur la floraison du houblon japonais et du chanvre déterminées par des semis haitifs: C. R. Hebd. Seanc. Acad. Sci. Paris, no. 155, p. 297-300.

Tournois, J., 1914, Études sur la sexualité du houblon: Annis. Sci. Nat., no. 19, p. 49-191.

Troeh, F.R., and Thompson, L.M., 2005, Soils and soil fertility, Blackwell Pub.

Tucker, C.J., and Nicholson, S.E., 1999, Variations in the size of the Sahara Desert from 1980 to 1997: Ambio, v. 28, no. 7, p. 587-591.

Tussie, G.D., 2004, Vegetation ecology, rangeland condition and forage resources evaluation in the Borana Lowlands, Southern Oromia, Ethiopia, Cuvillier.

Tüxen, R., 1956, Die heutige potentielle naturliche Vegetation als Gegenstand der Vegetationskartierung (Contemporary Potential Natural Vegetation as an Object of Vegetation Mapping): Angewandte Pflanzensoziologie, v. 13, p. 5-42.

Tyler, G., 2001, Relationships between climate and flowering of eight herbs in a Swedish deciduous forest: Annals of Botany, v. 87, no. 5, p. 623-630.

USGS (US Geological Survey), 1997, STATSGO soil characteristics for the Conterminous United States, http://water.usgs.gov/lookup/getspatial?muid.

USGS (US Geological Survey), 2008, eMODIS CONUS readme, v. 2009, no. June 19.

USGS (US Geological Survey), 2010, National Gap Analysis Program Land Cover Data.

Van Dyke, F., 2008, Conservation biology: Foundations, concepts, applications, Springer.

Van Tassell, L.W., Bartlett, E. T., Mitchell, John E., 2001, Projected use of grazed forages in the United States, 2000 to 2050: A technical document supporting the 2000 USDA Forest Service RPA assessment: Fort Collins, CO, US Dept. of Agriculture, Forest Service, Rocky Mountain Research Station.

Waldo, F., 1893, Modern meteorology: London, Walter Scott, Ltd.

Walter, H., and Wieser, J., 1973, Vegetation of the earth in relation to climate and the eco-physiological conditions, English Universities Press.

Watson, D.J., 1947, Comparative physiological studies on the growth of field crops: I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years: Annals of Botany, v. 11, no. 1, p. 41-76.

Waugh, D., 2000, Geography: An integrated approach, Nelson.

Werger, M.J.A., 1983, Tropical grasslands, savannas, woodlands: Natural and manmade, Man's impact on vegetation: Boston, Dr. W. Junk Publishers, The Hague, p. 107-137.

White, M.A., Hoffman, F., Hargrove, W.W., and Nemani, R.R., 2005, A global framework for monitoring phenological responses to climate change: Geophys. Res. Lett., v. 32, no. 4, p. L04705.

White, M.A., and Nemani, R.R., 2006, Real-time monitoring and short-term forecasting of land surface phenology: Remote Sensing of Environment, v. 104, no. 1, p. 43-49.

White, M.A., Thornton, P.E., and Running, S.W., 1997, A continental phenology model for monitoring vegetation responses to interannual climatic variability: Global Biogeochem. Cycles, v. 11, no. 2, p. 217-234.

Whitley, N., 1850, On the climate of the British Islands in its effects on cultivation: Journal of the Royal Agricultural Society of England, no. 11, p. 1-65.

Whittaker, R.H., 1970, Communities and ecosystems, Macmillan, London: Collier Macmillan.

Wielgolaski, F., and Inouye, D., 2003, High latitude climates, *in* Schwartz, M., ed., Phenology: An integrative environmental science: Springer Netherlands, p. 175-194.

Williams, C.L., Hargrove, W.W., Liebman, M., and James, D.E., 2008, Agro-ecoregionalization of Iowa using multivariate geographical clustering: Agriculture, Ecosystems & Environment, v. 123, no. 1-3, p. 161-174.

Woodward, F.I., 1987, Climate and plant distribution, Cambridge University Press.

Xiao, X., Braswell, B., Zhang, Q., Boles, S., Frolking, S., and Moore Iii, B., 2003, Sensitivity of vegetation indices to atmospheric aerosols: Continental-scale observations in Northern Asia: Remote Sensing of Environment, v. 84, no. 3, p. 385-392.

Xiao, X., Zhang, Q., Saleska, S., Hutyra, L., De Camargo, P., Wofsy, S., Frolking, S., Boles, S., Keller, M., and Moore Iii, B., 2005, Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest: Remote Sensing of Environment, v. 94, no. 1, p. 105-122.

Xu, L., and Baldocchi, D., 2003, Seasonal trend of photosynthetic parameters and stomatal conductance of blue oak (Quercus douglasii) under prolonged summer drought and high temperature: Tree Physiology, v. 23, no. 13, p. 865-877.

Zhang, F., 1995, Effects of global warming on plant phenological events in China: Acta Geographica Sinica, v. 50, no. 5, p. 402-410.

Zhang, X., Friedl, M.A., and Schaaf, C.B., 2009, Sensitivity of vegetation phenology detection to the temporal resolution of satellite data: International Journal of Remote Sensing, v. 30, no. 8, p. 2061-2074.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., and Huete, A., 2003, Monitoring vegetation phenology using MODIS: Remote Sensing of Environment, v. 84, no. 3, p. 471-475.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., and Liu, Z., 2005, Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments: Journal of Geophysical Research: Atmospheres, v. 110, no. D12, p. D12103.

Zhang, X., Hodges, J.C.F., Schaaf, C.B., Friedl, M.A., Strahler, A.H., and Feng, G., 2001, Global vegetation phenology from AVHRR and MODIS data, *in* Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International, p. 2262-2264, vol.2265.