A Model For Determining Drivers of Phenology in Western United States Rangelands

Joseph R. St. Peter
University of Montana - Missoula

Follow this and additional works at: https://scholarworks.umt.edu/etd

Part of the Natural Resources and Conservation Commons, Other Forestry and Forest Sciences Commons, and the Statistical Models Commons

Let us know how access to this document benefits you.

Recommended Citation

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
A MODEL FOR DETERMINING DRIVERS OF PHENOLOGY IN WESTERN
UNITED STATES RANGLANDS

By

JOSEPH ROBERT ST. PETER

Bachelor of Arts, Willamette University, Salem, Oregon, 2007

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
In Resource Conservation

The University of Montana
Missoula, MT

May 2015

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Steve Siebert, Chair
Department of Forest Management

Brady Allred
Department of Forest Management

Matt Reeves
Rocky Mountain Research Center
A Model for Determining Drivers of Phenology in Western United States Rangelands

Chairperson: Steve Siebert

Plant phenology has long been used as an indicator of climate. Recent changes in plant phenology are evidence of the influence of climate change. Modeling plant phenology has become an effective tool to understand the impacts of climate change. Using machine learning techniques I developed a modeling process for accurately predicting phenology across a diverse landscape. This model uses individual site data to set site specific climate thresholds for plant phenology. This model also identifies the limiting factors to vegetation phenology for rangelands in the western United States. NDVI remotely sensed data was used to quantify land surface phenology and DAYMET data was used to quantify climate variables. I found that random forest modeling can predict observed plant phenological dates across western rangelands to within a single day for start of season, end of season and day of max NDVI. The model can also identify the most highly correlated variables for phenological events in the study area and highlight which variables limit growth in different vegetative communities. These results confirm previous work on drivers of temperate phenology. This study’s results show that random forest modeling can accurately identify the most important climate variables for phenological events and use those variables to predict phenological events on a large spatial scale.
Acknowledgements

I am greatly indebted to numerous people for their help and support over the years. I would not have been able to complete this thesis without them. I would like to thank my late committee chair and advisor Don Bedunah for taking a chance and admitting me to be his student, and putting me on a path to earn a Master’s degree. I would also like to acknowledge Matt Reeves who helped me select, design and find funding for this study. He stepped in to advise me at a crucial time and has been a motivating force and collaborator throughout this study. I would be remiss if I did not acknowledge the technical expertise of John Hogland and Rob Langston of the Rocky Mountain Research Station which was invaluable to this study. I would also like to express my gratitude to my committee chair Prof. Steve Siebert, and committee member Prof. Brady Allred for helping me navigate this entire process and provide much needed feedback.

On a personal level I would like to thank my parents Don & Linda St. Peter for their all of their support. Without them I would not have been able to even consider entering a Master’s program let alone complete a thesis, or in fact exist at all. Finally, I would like to thank my wife Hannah St. Peter who took on the roles of editor, graphic designer, personal motivation coach, chauffer, and all around support team in order for me to complete this thesis. Thank you all!
# Table of Contents

Chapter 1: Introduction ............................................................................................................. 1  
  1.1. Vegetation Phenology ................................................................................................. 1  
  1.2. Remotely Sensed Data ............................................................................................... 1  
  1.3. Phenological Mismatches ......................................................................................... 3  
  1.4. Research Objectives .................................................................................................. 4  

Chapter 2: Description of Data and Random Forest Models ................................................. 5  
  2.1. Random Forests ......................................................................................................... 5  
  2.2. NDVI ........................................................................................................................... 8  
  2.3. DAYMET .................................................................................................................... 11  
  2.4. Study Area .................................................................................................................. 12  

Chapter 3: Methodology ...................................................................................................... 15  
  3.1. Climate Variable Selection ....................................................................................... 15  
  3.2. Quantifying Climate Variables .................................................................................. 17  
  3.3. Modeling Variable Importance and Predicting Phenology ...................................... 18  

Chapter 4: Results ............................................................................................................... 20  
  4.1. Variable Importance Models ...................................................................................... 20  
  4.2. Predictive Models ....................................................................................................... 27  

Chapter 5: Discussion ........................................................................................................... 35  
  5.1. Fulfillment of Research Objectives ......................................................................... 35  
  5.2. Error Rate of Predictive Models .............................................................................. 37  
  5.3. Null Values in Predictive Models .............................................................................. 38  
  5.4. Future Research ......................................................................................................... 38  

Chapter 6: Conclusion ......................................................................................................... 39  

References ............................................................................................................................ 41  

Appendix A ............................................................................................................................ 47  

Appendix B ............................................................................................................................ 50
Table of Figures

Figure 1. Smoothed and moving average NDVI taken from Reed et al 1994. .................. 10
Figure 2. The study area with the three largest vegetative communities highlighted. ... 12
Figure 3. SOS Random forest model variable importance .............................................. 21
Figure 4. EOS Random forest model variable importance. ............................................. 22
Figure 5. MAX Random forest model variable importance. ........................................... 23
Figure 6. SOS day variables correlation to Observed SOS day ........................................ 24
Figure 7. Temperature minimum correlated to observed SOS day and elevation............. 26
Figure 8. Observed SOS day compared to model 1 predicted SOS day......................... 29
Figure 9. Observed SOS day compared to model 2 predicted SOS day.......................... 30
Figure 10. Observed EOS day compared to predicted EOS day. .................................... 31
Figure 11. Observed MAX day compared to predicted MAX day................................. 32
Figure 12. Predicted and observed SOS day for the Inter-Mountain Basins Big Sagebrush Shrubland vegetative community................................................................. 33
Figure 13. Predicted and observed SOS day for the Northwestern Great Plains Mixedgrass Prairie vegetative community ................................................................. 33
Figure 14. Predicted and observed SOS day for the Western Great Plains Shortgrass Prairie vegetative community ................................................................. 34
Figure 15. Years 2001-2006 of predicted and observed SOS day.................................... 34
Figure 16. Years 2007-2012 of predicted and observed SOS day.................................... 35

Table of Tables

Table 1. Plant growth variables to inform SOS phenology model................................. 6
Table 2. Vegetative communities and their average phenology in study area............... 13

Equations

\[ \text{NDVI} = \frac{\text{NIR} - \text{VIS}}{\text{NIR} + \text{VIS}} \quad \text{Equation 1} \] .......................................................... 9

\[ \text{GDD} = \text{Tmax} + \text{Tmin2} - Tbase, \text{and GDD} \geq 0 \quad \text{Equation 2} \] .......................... 16
Chapter 1: Introduction

1.1. Vegetation Phenology

Phenology is the study of the timing of biological events and the abiotic and biotic factors that regulate timing (Jolly, et al., 2005). Vegetation phenology specifically is determined by three basic plant growth requirements: energy, water and temperature (Zhao, et al., 2013). These basic requirements for growth vary in their availability throughout the year due to climatic conditions, and the revolution of the earth around the sun. When these resource requirements are met the plant will start to grow. The amount of these three resources that are needed for a plant to grow is referred to as that resources’ ‘threshold’ (Jolly, et al., 2005). Identifying these resource thresholds for vegetation is an important part of building an accurate phenology model (Jolly, et al., 2005).

Vegetation phenology is a biological indicator sensitive to environmental variation and therefore is a useful proxy for climate and weather (Henerby, 2013; Rosenzweig, et al., 2007). This relationship has been utilized to investigate a diverse range of ecological topics (Pettorelli, et al., 2011). The change in the phenology of biological processes is one of the main consequences of global climate change, and many recent studies have used phenology data to identify the impacts of climate change (Badeck, et al., 2004; Brown, et al., 2012; Cleland, et al., 2007; Root, et al., 2003; Zhao, et al., 2013).

1.2. Remotely Sensed Data

Satellite reflectance data has emerged as an increasingly sensitive and valuable resource for quantifying phenology at large spatial and temporal scales. The usefulness of phenology as an indicator of climate change affects and the availability of phenology data has led to a surge in interest in the study of phenology. The subject has been featured in reports such as the IPCC’s Working Group II’s Fourth Assessment Report
which states, “Phenology...is perhaps the simplest process in which to track changes in
the ecology of species in response to climate change” (Rosenzweig, et al., 2007).

Henerby notes, in his 2013 book section, that in the opening chapter of the IPCC report
“the terms ‘phenology’ or ‘phenological’ are mentioned 91 times” (Henerby, 2013). In
2003, Root et al. was able to use the increased work in this field to produce a meta-
alysis of the results of 143 studies on species and climate change (Root, et al., 2003).
In their meta-analysis over 80% of the 1,468 species examined had phenological
changes in the direction expected with climate change (Root, et al., 2003). Studies using
ground based observation, remote sensing data and other measures of growing season,
such as atmospheric CO$_2$ signals, have found an advancement of the start of the growing
season in temperate regions (Badeck, et al., 2004; Cleland, et al., 2007; Schwartz, et al.,
2006; Tucker, et al, 2001; Walther, et al., 2002; White, et al., 2009). These studies have
found that the onset of spring has advanced in temperate latitudes anywhere from 1.2
to 5.5 days per decade over the last half of the twentieth century, with regional and
species specific variance (Root, et al., 2003; Schwartz, et al., 2006). In the decade since
that meta-analysis was performed there has been even more growth in phenology
research, in particular basic and applied research using satellite reflectance for
vegetation and land surface phenology (Henebry & de Beurs, 2013).

Advancements in remote sensing have led to the study of ‘land surface
phenology’ a term which is meant to “describe the seasonality of reflectance
characteristics that are associated with stages of vegetation development” (Henebry &
Su, 1995). Land surface phenology depends on continuous satellite reflectance data to
track the changes in ground cover of a landscape associated with vegetation life cycle
events, such as the start, maximum and end of a growing season. This allows
researchers to scale up spatially and spectrally and capture the reflectance associated
with vegetation development for an entire landscape, not just a specific species or
canopy development life cycle (Henebry & de Beurs, 2013).

Many studies in this field have used remotely sensed satellite data from AVHRR
(Advanced Very High Resolution Radiometer), and MODIS (Moderate Resolution
Imaging Spectroradiometer), to measure the absorbed fraction of photosynthetically active radiation. With these tools, researchers are able to produce greenness indices such as NDVI (Normalized Differentiation Vegetation Index) and EVI (Enhanced Vegetation Index). These data have allowed researchers to investigate climate, weather, primary productivity and phenology of landscapes at multiple spatial scales. Much of this research has focused on current and historical shifts in climatological metrics due to climate change (Cleland, et al., 2007; Cook, et al., 2012; Jenerette, et al., 2010; Post, et al., 2008; Reed, et al., 1994; Rosenzweig, et al., 2007; Tucker, et al, 2001; Walther, 2010; White, et al., 2009).

1.3. Phenological Mismatches

Shifts in the timing of the seasonal events due to climate change can alter the phenology of individual species and communities of plants and animals leading to phenological mismatches. An example of which is the snowshoe hare’s (*Lepus americanus*) coat color changing to white despite the later onset and earlier melting of snow in the Northern Rockies, which then results in an increased predation rate for the species (Zimova, et al., 2014). Changing phenology in the rangelands of the United States could have several impacts on wild and domestic animals. For example, if the start of spring begins earlier and a species migration pattern does not also advance then that species would not arrive at the optimal time. Such a mismatch has already been observed in migratory bird populations across the Northern Hemisphere where late arrival due to the advancement of spring has been linked to population declines (Jones & Cresswell, 2010).

The temporal advancement of spring leads to advancement in the timing of peak forage, the time when vegetation is at its most nutritious. Grazing species prefer nutritious vegetation and time their migrations in order to maximize the availability of nutritious vegetation (Hebblewhite, et al., 2008). If vegetation across a landscape matures at different times, due to differing elevation for example, the availability of high quality forage is prolonged, which is an advantage of migration (Post, et al., 2008).
In ungulates the migratory advantage, referred to as the forage maturation hypothesis, occurs when migratory populations gain an advantage over resident populations by following the green wave of plant growth across a landscape, extending the time when peak forage is available (Hebblewhite, et al., 2008). Migrating ungulate populations have been found to use stop over sites in between winter and summer ranges, spending 95% of their migration time in these sites at times of their highest forage quality (Sawyer & Kauffman, 2011). These migratory strategies depend upon the availability of high quality forage varying spatially and temporally. If the time of highest forage quality were to occur more homogenously across the landscape the advantage of the migratory strategy would disappear. Post et al.’s 2008 study has shown that this is already occurring in some areas. Increased warming has decreased the heterogeneous spatial distribution of forage and has been linked to declining recruitment in ungulate species in Greenland (Post, et al., 2008). If the start of spring occurs homogenously across the landscape then animal populations that depend upon migrating to forage will be vulnerable to decline.

### 1.4. Research Objectives

This study was designed to build a phenological model that works on a large scale and has high predictive accuracy. This model will identify the climatic variables that correlate with phenological events in the rangelands of the western United States. The large spatial scale will span a variety of vegetative communities at varied latitudes and elevations. The phenology in these areas will be driven by differing climate variables, some rangelands’ plant growth is water limited, and other rangelands are limited by temperature or energy availability. This study will attempt to reproduce the effect of multiple phenological drivers across a landscape in a single model by using decision tree based statistical modeling. This approach was selected as the most appropriate way to determine what climate variables are affecting diverse rangelands, as it makes no assumptions regarding relationships between climate and phenology. Additionally, statistical, or empirical, models have the benefit of “correct phenological mechanism assumptions” (Zhao, et al., 2013) which leads to higher predictive accuracy.
when compared to mechanistic cause and effect models or theoretical cost-benefit models. A decision tree based model will be able to determine phenological variable thresholds for each site based on that site’s historical data, instead of using a preset threshold that previous phenological models have used (Jolly, et al. 2005; Kovalskyy & Henerby, 2012). Allowing the data for each site to determine the climate drivers and thresholds of phenology at that site will allow for the model to be applicable across a diverse landscape without any additional parameterization and lead to higher accuracy in phenological predictions.

An extensive literature review was conducted to identify the most likely climate variables to influence rangeland phenology (Table 1). All climate variables that could be produced with available data were then combined into a Random Forest (RF) statistical model to develop a ranking of the most important variables for determining rangeland phenology. The values for start of season (SOS), maximum growth (MAX) and end of season (EOS) days, as determined by NDVI data, were used as the dependent variable, while climate variables and two site characteristic variables, elevation and vegetative community, were the independent variables. The most important climate variables for each phenological event were identified and subsequently used to build a second RF model that predicts phenology events. The predictive models were built using a similar method to Piekielek (2012) who used RF modeling to build predictive models for grassland areas in the Great Yellowstone Ecosystem and produced significant results (Piekielek, 2012).

Chapter 2: Description of Random Forest Models and Data

2.1. Random Forests

Random Forests (RF) is a machine learning technique, developed by Breiman (2001). For this study it was preferred to logistic regression-like techniques, used in previous studies of phenology to identify the importance of variables (Piekielek, 2012), because of RF’s “accuracy in determining variable importance and ability to model complex interactions” (Cutler, et al., 2007). This study’s landscape level scale means the
same variables that affect phenology in one vegetative community might not have the same affect in another vegetative community, or even operate in the same way across a vegetative community’s entire range (Prasad, et al., 2006). RF is able to model specialized variables that have great significance in a small area and limited significance in the majority of areas (Berk, 2011). These specialized variables would be overlooked in traditional regression models, but are important to account for when modeling on a large scale.

**Table 1.** Plant growth variables and their quantification to inform the start of season (SOS) phenology model.

<table>
<thead>
<tr>
<th>Growth Requirement</th>
<th>Variable</th>
<th>Quantifiable proxy</th>
<th>Functions*</th>
<th>Temporal Range**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photoperiod</td>
<td>Daylength</td>
<td>Mean</td>
<td></td>
<td>SOS day-14 to SOS day</td>
</tr>
<tr>
<td>Solar Radiation</td>
<td>Incident shortwave</td>
<td>Mean</td>
<td></td>
<td>SOS day-14 to SOS day</td>
</tr>
<tr>
<td></td>
<td>radiation flux</td>
<td>Density</td>
<td></td>
<td>SOS day-21 to SOS day</td>
</tr>
<tr>
<td><strong>Water</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>Daily precipitation</td>
<td>Sum</td>
<td></td>
<td>Day 0 to SOS day</td>
</tr>
<tr>
<td>Water Stress</td>
<td>Water vapor pressure (VP)</td>
<td>Mean</td>
<td></td>
<td>SOS day-14 to SOS day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SOS day-21 to SOS day</td>
</tr>
<tr>
<td><strong>Temperature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>Daily minimum temp</td>
<td>Mean and Minimum</td>
<td></td>
<td>SOS day-14 to SOS day</td>
</tr>
<tr>
<td>Maximum</td>
<td>Daily maximum temp</td>
<td>Mean and Maximum</td>
<td></td>
<td>SOS day-21 to SOS day</td>
</tr>
<tr>
<td>Accumulated Thermal Load</td>
<td>AGDD</td>
<td>See Eqn 2</td>
<td></td>
<td>Day 0 to SOS day</td>
</tr>
<tr>
<td><strong>Site Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>Elevation (m)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Latitude</td>
<td>Degrees</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Vegetative Community</td>
<td>EVT</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

* ‘Functions’ is what function was performed on the quantifier, for example daily minimum temperature was averaged over the temporal range, and the minimum value in that range was identified.

** In temporal range ‘SOS day- x to SOS day’ means the quantifier’s range was from x days before the SOS day to the SOS day. ‘Year-1 EOS to SOS day’ means the range was from the previous year’s end of season day to the SOS day.
RF builds many decision trees and combines their predictions to vote on the predicted value of an observation. Decision tree learning is a way of determining a target’s value by using observations of that target to build decision trees through ‘recursive binary partitioning into regions that are increasingly homogenous’ (Cutler, et al., 2007). The target variable for this model is ‘day of the year’ which is a continuous value, therefore our random forest model will be building regression trees and not classification trees which are built for finite values. The decision trees are made up of nodes, or leaves, which represent target values and branches that represent the splitting based on a predictor variable. The splitting continues until further division no longer increases the accuracy of the prediction. RF selects many bootstrap samples from the data, and fits a decision tree to each sample. Around 63-66% of the original observations occur at least once (Cutler, et al., 2007). Observations that do not occur in this training dataset are held as out-of-bag observations. At each node of the tree a small number of randomly selected variables are used for splitting, typically the square root of the total number of variables (Cutler, et al., 2007). When the trees are fully developed they are then each used to predict the out-of-bag observations. These predictions are used to generate error rates and accuracies. When RF is used to build regression trees, as it was used in this study, the mean of each tree’s predicted values are used as the predictive output. The out-of-bag were not used to build the trees so the out-of-bag estimates act as cross-validation accuracy estimates, with no additional cross-validation techniques necessary (Breiman, 2001; Cutler, et al., 2007).

RF also has an internal estimate of the importance of each variable. It produces this estimate by replacing the specific variable with randomly generated values for the out-of-bag observations and these random values are then input to the trees to generate a new prediction. Then the miscalculation rate for the out-of-bag observations of each tree are compared to the new predictions. The difference between the two predictions divided by the standard error is the measure of the importance of the variable (Breiman, 2001; Cutler, et al., 2007). Additional efficiency of random forest
models is found in the internal cross validation, no extra cross validation steps are necessary (Breiman, 2001).

RF has been found to be one of the most accurate statistical modeling techniques for predictive vegetation mapping when compared with other statistical models (Prasad, et al., 2006). RF allows you to grow large numbers of trees, typically 500-2000, without over fitting to the training data set unlike typical decision tree methods. Additionally, bias is kept low through the random selection of predictor variables (Prasad, et al., 2006). Of even greater importance to this study is RF’s ability to determine the relative importance of the variables. RF’s measures of variable importance were used to determine which variables were used to build the predictive phenology model. Because of this ability RF was used to both determine the variables used in the phenology models and to build those models.

The climate variables used in this study are highly correlated (Table 1). In RF modeling this can lead to unstable results when large trees are generated and there is a substantial number of predictor variables correlated with one another (Berk, 2011; Segal, 2003). This multicollinearity problem can be overcome by working with smaller trees and screening out some of these predictors before random forests is applied to the data (Berk, 2011). This study has taken both of these steps; before predictive RF models are built the variables are ranked and correlated variables are screened out leaving only 7-8 variables for building trees in the predictive RF model leading to small trees.

### 2.2. NDVI

Key phenological events were identified using NASA’s MODIS satellite sensors’ Normalized Differentiation Vegetation Index (NDVI) 250m² western United States product provided by the USGS EROS Center (http://phenology.cr.usgs.gov/). NDVI is a simple indicator of plant greenness based on remotely sensed reflectance data that has been found to have a positive correlation to vegetation biomass, and by extension forage availability, in non-forested habitats (Borowik, et al., 2013; Diouf & Lambin, 2001;
Prince & Tucker, 1986; Wessels, et al., 2006). NDVI is calculated by subtracting near infrared light measurement, usually taken by satellites, from the visible red light measurement and dividing that by the addition of the near-infrared red light measurement to the near infrared light measurement, this produces a ratio between -1 and 1 of the relative amount of greenness of vegetation. (Weier & Herring, 2000). Expressed mathematically the formula for NDVI is:

\[
NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad \text{(Equation 1)}
\]

Where NIR represents reflectance in near infrared spectrum and VIS represents reflectance in the visible, red, spectrum.

The USGS EROS Center NDVI data uses the raw NDVI measurements and provides specific days when three phenological events occur, start of growing season (SOS), maximum growth (MAX) and end of growing season (EOS). This data uses an autoregressive moving average method to determine SOS and EOS dates; the methodology is detailed in Reed et al 1994. The NDVI data for each 250m pixel is smoothed over a 14-day period and is compared to the moving average of nine previous 14-day smoothed periods. The point when the smoothed 14-day data crosses the moving average in an upward direction and remains above the moving average is determined to be the date of SOS (Figure 1). The moving average is reversed in time to determine the date of EOS. The moving average is calculated using the nine 14-day time periods after the smoothed data, as opposed to the nine 14-day periods preceding the smoothed data used to determine the SOS. The date when the smoothed data crosses this moving average in a downwards direction is determined to be the date of EOS. The date, and amount, of max NDVI was determined using the raw NDVI data (Reed, et al., 1994). The result of these treatments is a digital map of the contiguous United States with the date for the phenological event of interest in each 250m pixel. These dates can span from day -150 to day 450 (day 215 of the previous year to day 85 of the next year).
Figure 1. A single pixel’s smoothed NDVI and moving average curves used to determine SOS and EOS dates. Taken from Reed et al 1994.

These data were chosen for this study for several reasons. The Reed et al. 1994 methodology for determining dates of phenological events has been found to be consistently related to measured and modeled plant phenology, and has a higher than average retrieval rate than other methodologies (White, et al., 2009). This method was also found to have higher variability than other methodologies for determining phenological dates, which would be expected since each pixel is measured independently and not aggregated with its neighbors (White, et al., 2009). To build a phenology model that identifies climate drivers and thresholds for each site it is important to have phenological data that are independent across sites.

This method has limitations as well, specifically the lack of clear criteria for the period of time used for the moving average. Additionally, this method makes the assumption that phenology is not erratic, occurs multiple times in a year or spans multiple years, and can be captured with a moving average (de Beurs & Henerby, 2010).
These assumptions were not always met in our study area; those exceptions are discussed in the results section of this paper.

The use of remotely sensed data based on the vegetation optical depth (VOD) parameter from satellite passive microwave was also considered as an alternative to the satellite optical-infrared (IR) remote sensing that is the basis for NDVI. These data have several advantages for use as a phenological metric as it sensitive to biomass and water content as opposed to leaf area or greenness that IR records (Jones, et al., 2012). VOD has been used in studies of phenology at the ecoregion scale (Jones, et al., 2011; Jones, et al., 2012). While these data present many benefits for phenology research, the coarse 25km resolution of the data was too large for the data to be of use to this study.

2.3. DAYMET

Climatic variables were defined using Daily Meteorological or DAYMET data (Thornton, et al., 2014). DAYMET daily data provides 1km gridded estimates of weather variables for North America using ground-based meteorological stations. Ground based observations of minimum and maximum temperature, precipitation as well as elevation are used to model the daily weather variables across the landscape (Thornton, et al., 1997). The weather variables DAYMET produces include solar radiation, precipitation, daylength, maximum temperature, minimum temperature, and vapor pressure. This data set was chosen because of its high resolution, the large spatial area it covers and the numerous climatic variables it produces, all of which are necessary in order to inform an accurate phenology model for the western United States.

Both the USGS EROS Center NDVI data and the DAYMET data are based on a standard calendar year, meaning all the years have 365 days. For leap years the values for December 31 are discarded. The data in this study cover the years 2001-2012. This time period is the longest period with both NDVI and DAYMET data at 1km or finer resolution.
2.4. Study Area

The study area was selected to capture a diversity of rangeland types in the western United States. Rangeland vegetative community data was obtained from the Rocky Mountain Research Center (RMRS) at the 250m\(^2\) scale and follows the US ecological systems classification method developed by Comer and Schulz, 2007. The vegetation communities were ranked based on what percentage of the vegetation community was missing NDVI data. This was done so that the problems that arise when working with incomplete datasets would be mostly avoided. The highest ranked communities were selected to use as our study area. Twenty three vegetative communities were selected (Table 2). The three largest vegetative communities selected, Inter-Mountain Basins Big Sagebrush Shrubland, Northwestern Great Plains Mixedgrass Prairie, and Western Great Plains Shortgrass Prairie, cover 53.77\% of the study area (Figure 2), the largest six vegetative communities account for nearly 75\% of the study area (Table 2).

Figure 2. The study area with the three largest vegetative communities highlighted.
In order to gain a general understanding of the phenology in these vegetative communities, average start and end of season days and their standard deviations for the twelve years of NDVI data (2001-2012) were calculated by randomly sampling the NDVI data at 6000 points across the study area (Table 2). The largest vegetation community, Inter-Mountain Basins Big Sagebrush Shrubland, is also the most geographically diverse extending from just below the Canadian border in the state of Washington to southern New Mexico and from California to eastern Colorado (Figure 2). This vegetative community also has one of the earliest average SOS dates at Julian day 75, and a larger SOS day standard deviation of nearly 31 days (Table 2). The second largest vegetative community, Northwestern Great Plains Mixedgrass Prairie, is located primarily in eastern Montana, the Dakotas and northeast Wyoming. It also has one of the least variable SOS days of any of the study area’s vegetative communities, with a standard deviation of only 8.87 days on a Julian day 91 average SOS day.
Table 2. Study area vegetative communities with percent of study area included in the community, the average and standard deviation in days of start of season (SOS) and end of season (EOS) Julian day for the years 2001-2012.

<table>
<thead>
<tr>
<th>Name</th>
<th>Study Area %</th>
<th>SOS day</th>
<th>STD</th>
<th>EOS day</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Mountain Basins Big Sagebrush Shrubland</td>
<td>20.29%</td>
<td>75</td>
<td>30.99</td>
<td>297</td>
<td>38.76</td>
</tr>
<tr>
<td>Northwestern Great Plains Mixedgrass Prairie</td>
<td>18.27%</td>
<td>91</td>
<td>8.87</td>
<td>306</td>
<td>12.03</td>
</tr>
<tr>
<td>Western Great Plains Shortgrass Prairie</td>
<td>15.21%</td>
<td>113</td>
<td>23.06</td>
<td>310</td>
<td>22.13</td>
</tr>
<tr>
<td>Inter-Mountain Basins Big Sagebrush Steppe</td>
<td>8.56%</td>
<td>72</td>
<td>23.12</td>
<td>293</td>
<td>33.05</td>
</tr>
<tr>
<td>Western Great Plains Sand Prairie</td>
<td>6.52%</td>
<td>100</td>
<td>15.07</td>
<td>304</td>
<td>18.06</td>
</tr>
<tr>
<td>Chihuahuan Mixed Desert and Thornscrub</td>
<td>6.05%</td>
<td>141</td>
<td>34.94</td>
<td>319</td>
<td>35.63</td>
</tr>
<tr>
<td>Inter-Mountain Basins Semi-Desert Grassland</td>
<td>3.19%</td>
<td>127</td>
<td>35.27</td>
<td>322</td>
<td>26.48</td>
</tr>
<tr>
<td>Inter-Mountain Basins Montane Sagebrush Steppe</td>
<td>3.03%</td>
<td>97</td>
<td>28.71</td>
<td>316</td>
<td>27.87</td>
</tr>
<tr>
<td>Artemisia tridentata ssp. vaseyana Shrubland Alliance</td>
<td>2.94%</td>
<td>103</td>
<td>25.38</td>
<td>322</td>
<td>22.33</td>
</tr>
<tr>
<td>Columbia Plateau Low Sagebrush Steppe</td>
<td>2.34%</td>
<td>76</td>
<td>18.71</td>
<td>299</td>
<td>30.34</td>
</tr>
<tr>
<td>Northern Rocky Mountain Lower Montane-Foothill-Valley Grassland</td>
<td>2.30%</td>
<td>95</td>
<td>14.75</td>
<td>315</td>
<td>19.27</td>
</tr>
<tr>
<td>Columbia Plateau Steppe and Grassland</td>
<td>1.81%</td>
<td>43</td>
<td>23.83</td>
<td>246</td>
<td>43.92</td>
</tr>
<tr>
<td>Inter-Mountain Basins Semi-Desert Shrub-Steppe</td>
<td>1.64%</td>
<td>116</td>
<td>42.59</td>
<td>317</td>
<td>37.37</td>
</tr>
<tr>
<td>Rocky Mountain Subalpine-Montane Mesic Meadow</td>
<td>1.47%</td>
<td>122</td>
<td>19.56</td>
<td>319</td>
<td>11.99</td>
</tr>
<tr>
<td>Rocky Mountain Lower Montane-Foothill Shrubland</td>
<td>1.40%</td>
<td>111</td>
<td>28.13</td>
<td>320</td>
<td>19.15</td>
</tr>
<tr>
<td>Rocky Mountain Subalpine/Upper Montane Riparian Systems</td>
<td>1.00%</td>
<td>118</td>
<td>15.70</td>
<td>325</td>
<td>9.54</td>
</tr>
<tr>
<td>Rocky Mountain Alpine Turf</td>
<td>0.85%</td>
<td>153</td>
<td>12.18</td>
<td>314</td>
<td>6.22</td>
</tr>
<tr>
<td>Northern Rocky Mountain Montane-Foothill Deciduous Shrubland</td>
<td>0.67%</td>
<td>107</td>
<td>23.16</td>
<td>320</td>
<td>15.61</td>
</tr>
<tr>
<td>Southern Rocky Mountain Montane-Subalpine Grassland</td>
<td>0.63%</td>
<td>140</td>
<td>21.73</td>
<td>333</td>
<td>14.44</td>
</tr>
<tr>
<td>Northern Rocky Mountain Subalpine-Upper Montane Grassland</td>
<td>0.50%</td>
<td>132</td>
<td>26.40</td>
<td>320</td>
<td>10.55</td>
</tr>
<tr>
<td>Colorado Plateau Mixed Low Sagebrush Shrubland</td>
<td>0.50%</td>
<td>111</td>
<td>26.43</td>
<td>330</td>
<td>15.22</td>
</tr>
<tr>
<td>Northwestern Great Plains Shrubland</td>
<td>0.47%</td>
<td>100</td>
<td>7.50</td>
<td>308</td>
<td>15.68</td>
</tr>
<tr>
<td>Inter-Mountain Basins Curl-leaf Mountain Mahogany Woodland and Shrubland</td>
<td>0.36%</td>
<td>110</td>
<td>10.89</td>
<td>324</td>
<td>24.35</td>
</tr>
</tbody>
</table>
Chapter 3: Methodology

3.1. Climate Variable Selection

An evaluation of relevant literature identified the following climate variables as most likely to influence phenology in our study area: maximum and minimum temperature, accumulated thermal load, accumulated precipitation, soil moisture, vapor pressure deficit, photoperiod, solar radiation, elevation, latitude and vegetative community. These variables represent quantifiable proxies for the three plant growth requirements of energy, water and adequate temperature.

The importance of temperature in the development of plants has been recognized for centuries (McMaster, 2005). Accumulated thermal load, maximum and minimum temperature are the variables most widely used to quantify the effect of temperature on plant phenology (Zhao, et al., 2013). At low temperatures many plants’ bioclimatic processes are adversely affected (Jolly, et al., 2005; Levitt, 1980). Restrictions on root water uptake when soil temperatures are low have been found to be closely related to phenology (Waring, 1969).

At higher temperatures the rate of energy used in respiration may exceed the rate of energy produced through photosynthesis causing the plant to halt growth (Whiting, et al., 2014). Temperature maximum and minimum are widely used variables of phenology models such as the Spring indices model (Schwartz, 1997), or the growing season index (Jolly, et al., 2005), they have also been found to be explanatory variables of phenology, along with precipitation and potential evapotranspiration, in various studies (Kathuroju, et al., 2007; White, et al., 1997; White, et al., 2009).

Accumulated thermal load as a driver of plant development rate is a concept that dates back to de Reaumur’s 1735 publication in which he models plant development based on thermal time, not calendar days. The concept of thermal time evolved and is now quantified through growing degree days, the formula for which follows in Equation 2 (adapted from McMaster 2005):
\[ \text{GDD} = \sum \left( \frac{(T_{\text{max}}+T_{\text{min}})}{2} - T_{\text{base}} \right) \text{[and GDD} \geq 0] \] (Equation 2)

\( T_{\text{max}} \) and \( T_{\text{min}} \) are the daily maximum and minimum temperature, \( T_{\text{base}} \) is the base temperature that growth can occur, below which temperature has no effect, and as such GDD cannot be negative. This quantity is for an individual day and is usually summed over a period of days to derive an accumulated growing degree day (AGDD) value. The time period used to derive AGDD in this study was from day zero to the phenological event. Day zero is a common starting point for calculating AGDD as it usually represents the midpoint of the dormant season in temperate regions.

The importance of water availability to plant phenology is accounted for in the soil moisture and accumulated precipitation variables. Lack of water affects plants’ cell division and expansion causing growth to cease (Palacio, et al., 2014). While soil moisture is a more direct measure of water available to vegetation, precipitation is also used to determine water availability (Flanagan, 2009). Precipitation sums at different temporal ranges have been used in previous prognostic phenology models (Kathuroju, et al., 2007; Kovalskyy & Henerby, 2012).

Vapour pressure deficit is related to both precipitation and temperature; it is a measure of evaporative demand, or water stress, and is used as a surrogate for a complete modelled soil water balance. Water stress can stop cell division, and cause stomatal closure (Jolly, et al., 2005). As such it is an important variable in determining the end of a growing season.

Photoperiod and solar radiation account for the energy, and to some extent temperature, requirement of plant growth (Zhao, et al., 2013). Photoperiod is the period of time each day that a plant receives light, quantified by day length. Solar radiation is the amount of incident shortwave radiation a plant receives a day, it is similar to photoperiod but is ‘generated as a function of Sun-slope geometry and interpolated diurnal temperature range’ (Thornton & Running, 1999). These two
measures of energy available for plant growth are highly correlated with each other and with latitude; therefore latitude was not used in the model.

The two variables of elevation and vegetative community are included to locate the model spatially. The elevation data was at the 30m² scale and was obtained from the Rocky Mountain Research Station. These variables account for the variability in location and type of vegetation in a landscape that can affect phenology and are commonly used in phenology models (Kathuroju, et al., 2007; Jolly, et al., 2005).

3.2. Quantifying Climate Variables

Quantifiable proxies were identified for each variable, as well as the temporal ranges and functions used to quantify the variables, listed in Table 2. Previous studies were reviewed to identify several temporal ranges that could be of significance to rangeland phenology. The ranges used were intended to capture the weather immediately preceding a phenological event, as well as that year’s underlying climate conditions. The short term temporal ranges used were 21 and 14 days preceding phenological events. The 21 day period is used in the GSI model, and Kovalskyy and Henerby’s event driven phenology model among others (Kovalskyy & Henerby, 2012; Jolly, et al., 2005). It is meant to be a short enough time frame to capture weather trends but long enough to avoid the 16 or 14 day data smoothing that is used in many remotely sensed data. The 14 day time frame was also included as it was the period of time the NDVI dataset used in this study is smoothed over, making 14 days the smallest window of time to capture the climate conditions immediately preceding the phenological event (Reed, et al., 1994). Temporal ranges spanning the dormant season, and from midwinter (day zero) were used to capture long term climate conditions leading to phenological events.

These quantifiable proxies were produced by matching DAYMET daily weather data with the NDVI data. To accomplish this a program was developed in the Python 3.2 coding language that selected a user-defined range of DAYMET data and then ran a summarizing function on them, either summing, averaging or finding a maximum or
minimum value of the data in the selected range for each 1km pixel in the western United States. The program output is a raster with the produced value for each pixel.

Accumulated growing degree days (AGDD) were calculated using Equation 2. Average daily temperature data was produced by averaging daily temperature maximum and minimum data for each pixel. The average daily temperature above zero degrees Celsius, a base which McMaster (2005) concludes is “very robust and often sufficient for most purposes,” was then summed from the start of the calendar year, the start of the growing season and the max of growing season day, to the three phenological events of interest, SOS, MAX and EOS, producing an AGDD value for these events at each pixel in our study area.

The climate variables for each year were summarized to produce an average of each variable over the 12 year period or, in the case of precipitation from the dormant season, 11 year period. Means and standard deviations for climate variable were generated spatially using ArcGIS 10.2. The resulting rasters were converted from float to integer data by first rounding to the nearest integer by adding .5 and then truncating using the raster integer tool. Tmax and tmin mean rasters were not rounded; instead these temperature variables were multiplied by 10, to preserve the accuracy of data at one tenth of a degree, and then truncated. To convert mean temperature rasters to integer rasters truncation was preferred over the rounding to nearest method because negative numbers would end up being rounded more than one tenth of a degree using the prior rounding method.

### 3.3. Modeling Variable Importance and Predicting Phenology

All data was aggregate to the 1km\(^2\) scale of the DAYMET data. The 19 mean climate variable rasters plus elevation and vegetative community rasters were banded together using the Rocky Mountain Research Station’s (RMRS) raster utility tool (Hogland & Anderson, 2014)(http://www.fs.fed.us/rm/raster-utility/). A set of 6,000 points was randomly selected from the study area. At each point all the banded variables were sampled, as well as the SOS day value. The values at these points were then used as
inputs into a random forest statistical model to determine their relative importance, with SOS day as the dependent variable and the 21 banded mean rasters as the independent variables. The RF model used had 500 trees, a 66% training ratio, with 5 variables for splitting which is roughly the square root of the total number of variables which is recommended for RF models (Breiman, 2001).

The random forest model produced Root Mean Square Error, Average Error and Average Relative Error scores for each variable (Figure 3). To select the variables of most importance to the phenological event, which would be used to build a predictive model, the variables’ error scores were examined. Because of the random nature of random forests these scores will change slightly each time the model is developed, so it is more useful to consider the ranking of the variables to each other than it is to focus on the error scores of the individual variables. The variables with error scores above saturation for all three error measurements would be selected, unless there was a higher score equivalent variable. For example, if PRCP 14 and PRCP 21 had high error scores only the variable with the highest combined score among the two would be selected as both represent short term precipitation and their temporal ranges overlap. The site characteristic variables of vegetative community (EVT) and elevation would be selected regardless of their error scores. Five climate variables were selected for inclusion in the start of season predictive model, those variables were accumulated growing degree days (AGDD), total precipitation from start of calendar year to SOS day (PRCP 0), average daylength of 14 days prior to SOS day (dayl 14), total precipitation from 14 days prior to SOS day (PRCP 14). These variables were combined with elevation and vegetative community data to inform a random forest model that accurately predicted SOS day in our study area to within a half day.

The SOS predictive model was built using the same methods outlined above, using 500 trees, a 66% training ratio and 3 variables for splitting (∼√7). This model was then informed using the seven selected variables to generate SOS day predictions for every pixel in our study area. A new sample of 6,000 points was taken in the study area to compare the observed SOS day to the model’s predicted SOS day, as well as the
correlation between the seven mean variables and observed SOS day. This process was repeated for the MAX and EOS data sets.

Chapter 4: Results

4.1. Variable Importance Models

For SOS day the variables determined to be of highest relative importance by the Random Forest model (Figure 3) were: accumulated growing degree days (AGDD), total precipitation from start of calendar year to SOS day (PRCP 0), average daylength of 14 days prior to SOS day (dayl 14), average daylength of 21 days prior to SOS day (dayl 21), elevation, total precipitation from 21 days prior to SOS day (PRCP 21), total precipitation from 14 days prior to SOS day (PRCP 14), lowest minimum temperature of the 14 days prior to SOS day (Tmin min 14), and vegetative community (EVT). To avoid giving extra weight to the same climate variable dayl 14 was selected and dayl 21 dropped as dayl 14 was found to be of higher relative importance (Figure 3). PRCP 14 was selected and PRCP 21 was dropped for the same reason. The selection of daylength, AGDD, minimum temperature and two different precipitation sums confirms the findings of previous studies of temperate phenology that found light and temperature as the most important factors with precipitation an important tertiary consideration (Stöckli et al., 2011).

The variables selected as having the highest relative importance to EOS day (Figure 4) were AGDD, total precipitation from start of calendar year to EOS day (PRCP 0), average daylength of 14 days prior to EOS day (Dayl 14), average daylength of 21 days prior to EOS day (Dayl 21), daily average solar radiation of the 14 and 21 days prior to EOS day (SRAD 14 and SRAD 21), the highest maximum daily temperature of the 21 days prior to EOS day (Tmax max), average vapor pressure of the 21 days prior to EOS day (VP avg 21) elevation, and vegetative community (EVT). For the predictive model Dayl 14 and SRAD 14 were selected and Dayl 21 and SRAD 21 due to the higher error scores (when variable was removed) of Dayl 14 and SRAD 14.
The variables selected for MAX day (Figure 5) were AGDD, total precipitation from start of calendar year to MAX day (PRCP 0), average daylength of 21 days prior to MAX day (Dayl 21), average vapor pressure of the 14 days prior to MAX day (VP avg 14), sum of daily average vapor pressure from start of calendar year to MAX day (VP Sum 0) elevation, and vegetative community (EVT).

**Figure 3.** SOS Random forest model root mean square error (RMSE) report showing variable importance of 21 different climate variables. The more important the variable the higher the RMSE Error Value when the variable is removed, the vertical line shows the error when all variables are included.
Figure 4. EOS Random forest model root mean square error (RMSE) report showing variable importance of 23 different climate variables. The more important the variable the higher the RMSE Error Value when the variable is removed, the vertical line shows the error when all variables are included.
Figure 5. MAX Random forest model root mean square error (RMSE) report showing variable importance of 23 different climate variables. The more important the variable the higher the RMSE Error Value when the variable is removed, the vertical line shows the error when all variables are included.

The variables selected to inform the predictive SOS day model were sampled at 6000 new randomly selected points in the study area and compared to SOS day values from the NDVI dataset. The correlation between SOS day and these variables is displayed in Figure 6.
Figure 6. Six of the selected SOS day variables and Observed SOS day at 6000 randomly sampled points.

It is important to recognize the inherent autocorrelation between these variables and SOS day. For example, AGDD is the above zero daily temperature summed to the SOS day, so AGDD will be increase as temperature increases through the year. Therefore SOS days that occur later in the year will have higher AGDD values regardless of AGDD’s affect upon determining when the SOS day occurs. PRCP 0 is the total precipitation from day zero to the SOS day and has the same autocorrelation as AGDD. The later in the year the SOS day occurs the more likely it is that the sum of total precipitation will increase. If temperature does not limit the start of season in any of these sites then the AGDD graph would be expected to have a single horizontal trend of points representing the increase in AGDD as temperatures increase through the year,
which would just represent the variable’s temporal autocorrelation. However, in the
AGDD, Tmin min 14 and PRCP 0 graphs in Figure 6 there appear to be two separate
trends, one vertical and the other horizontal. The horizontal trend is the autocorrelation
of the variables to SOS day that was expected. The vertical trends in contrast represent
the thresholds for the various variables. For AGDD and Tmin min 14, the two variables
representing temperature affects, this means that the sites that fall in this vertical trend
are temperature limited and the sites that fall along the horizontal trend are
representing the autocorrelation of the variable and their start of season is not limited
by temperature. This vertical trend represents the cumulative temperature threshold
that these vegetative communities must pass before growth can begin. This threshold is
roughly between 50°C and 100°C AGDD, and increases with elevation.

The Tmin min 14 graph also reflects a temperature threshold for even more
vegetative communities than the AGDD threshold represents. The Tmin min 14 variable
is the lowest minimum daily temperature in the 14 days before the SOS day. This graph
depicts a vertical trend representing a lower temperature threshold between -12°C and
-5°C and a trend that is likely autocorrelation that represents the vegetative
communities that are not limited by minimum temperature. Note that the units for the
Tmin min 14 graph in Figure 6 are in 1/10°C. Figure 7 includes an elevation classifier to
highlight the relationship between elevation and a temperature threshold. This is a
mixed relationship with various elevations being represented in both the vertical
threshold trend and the autocorrelation trend.

The vertical threshold trends in the AGDD and Tmin min 14 graphs consist of
spatially diverse sites that are located at low to high elevations, low to high latitudes
and consist of many vegetative communities. Many high elevation sites appear along
the vertical temperature threshold trends in the AGDD and Tmin min 14 graphs;
however there are many high elevation sites that do not track this trend. Similarly, many
low and mid elevation sites also track the temperature threshold trend (Figure 7). This
indicates that a single site characteristic, such as elevation, is unable to entirely explain
whether a site’s start of growing season date will be temperature limited.
Figure 7. The lowest minimum temperature of the 14 days before the SOS day and the observed SOS day, with an elevation classifier at 6000 randomly sampled points.

The PRCP 0 graph shows a similar trend of a vertical threshold trend and a horizontal autocorrelation trend. In the PRCP 0 graph the high elevation areas are represented by the horizontal autocorrelation trend, as they are not generally limited by water availability, while the vertical trend represents the water availability threshold the lower elevation vegetative communities must pass before growth can begin. Higher accumulated precipitation is expected at higher elevations so this relationship is not unexpected despite the autocorrelation of the variables (Basist & Bell, 1994).

The day length variable also appears to be responding to two trends: the flatter autocorrelation trend and a steeper trend that could be expressing the threshold of light limited vegetative communities. This relationship is less clear than in the PRCP 0 and AGDD graphs. The day length graph also depicts a lower threshold of 31700 seconds (8.8 hours) of daylight below which no average SOS day occurred in our study area (Figure 7).
6). These separate trends within the climate variables, that represent limiting thresholds and autocorrelation, demonstrate that the RF model is not simply responding to the autocorrelation within these variables but also recognizing the interaction between the climate variables that capture the limiting thresholds in different vegetative communities.

4.2. Predictive Models

The variables selected as most significant for each phenological event using RF’s variable importance output were used to create Random Forest models to predict SOS, EOS and MAX days across the study area. Predictions of the three phenological events were generated for every pixel in our study area and compared visually with the original NDVI values (Figures 8-11). These predictions were also sampled at a newly selected set of 6,000 random points. Each sampled point contained values for the NDVI observation, modeled predictions, and every variable used in the predictive model. Null values were removed from the sample before generating statistics and graphs. The SOS model contained a higher percentage of null values than the MAX or EOS models, approximately 25% of the sampled SOS predictions were null values, while MAX and EOS models had less than 9% of the sampled pixels return null values. A spatial component to production of null values is evident from Figure 8.

The models produced null values whenever any predictor variables that inform the model were missing data for a pixel. The SOS day model contained two variables that had static start dates, precipitation from day zero (PRCP 0) and accumulated growing degree days (AGDD) which also begins summations from day zero. These start dates meant that any negative or low SOS day value would produce no data and the model would be unable to make a prediction. The other two phenological events occurred further in the calendar year and so did not encounter the same rate of null values. In order to reduce the null values a second SOS predictive model was produced without AGDD and PRCP 0 variables. These two variables were replaced with equivalent variables that were more likely to contain values and therefore not produce a null value,
these variables were accumulated growing degree days from the previous dormant season (AGDD dorm) and precipitation from previous dormant season (PRCP dorm). These variables measure the accumulated thermal time (AGDD dorm) and accumulated precipitation (PRCP dorm) from the end of the previous growing season to the start of the current growing season. This removed most of the null values that were generated by AGDD and PRCP 0 in areas with early SOS days (Figure 9). Some areas with especially short or dry dormant seasons still had no precipitation values, these areas contributed to the remaining null values in this second SOS model (Figure 9).

The second SOS day model was able to predict the 12 year average SOS day in the study area with an average error of .43 days, with a standard deviation of .83 days. SOS day predictions were also analyzed within the largest three vegetative communities (Figure 12-14). The model was more accurate in these vegetative communities than the overall study area, indicating that because of their size the model was parameterized towards these communities. The model performed well ($R^2 > 0.7$) in all vegetative communities. The pixels with high error values, where the models were less able to predict observed data, were located in the same areas in all three models. These areas correspond with mountainous areas, such as the Oregon Cascades, which had high error rates in all three models.

The SOS model was then used to build predicted SOS days for our study area using climate data for each year from 2001-2012. The predicted SOS day for each year were then sampled at 6000 points and compared with their observed NDVI dates (Figure 15 & 16). The yearly predictions all follow a consistent shape, generally linear with a curved tail at the highest observed SOS day. The predictions are more widely spaced towards the left side of the x axis which corresponds to early SOS days. The predictions become more accurate for SOS days between days 90 and 150, where the majority of the observations occur, before tailing off around day 210. In the training dataset the SOS day never occurred later than day 210. The points represented in these tails are predominately lower latitude points. The values in this tail are all after the summer solstice and the curve the tail follows roughly follows the day length curve in
this study area. This suggests that the model is using the day length values at these points to predict the start of season days on the earlier side of the solstice without considering that these day length values also occur later in the season in days that the model did not train on.

**Start of Season Day Observed vs. Predicted**

![Map showing observed vs. predicted SOS days](image)

**Figure 8.** Observed mean SOS day for 2001-2012 compared to Random Forest model predicted SOS day. The unit of measurement is days, higher values indicate a greater difference between observed and predicted SOS day. Null values are areas where the model failed to return a prediction.
Figure 9. Observed mean SOS day for 2001-2012 compared to Random Forest model predicted SOS day for the second SOS model. The unit of measurement is days, higher values indicate a greater difference between observed and predicted SOS day. Null values are areas where the model failed to return a prediction.
Figure 10. Observed mean EOS day for 2001-2012 compared to Random Forest model predicted EOS day. The unit of measurement is days, higher values indicate a greater difference between observed and predicted EOS day. Null values are areas where the model failed to return a prediction.
Figure 11. Observed mean MAX day for 2001-2012 compared to Random Forest model predicted MAX day. The unit of measurement is days, higher values indicate a greater difference between observed and predicted MAX day. Null values are areas where the model failed to return a prediction.
Figure 12. Comparison of predicted (RF Model) and observed (NDVI) SOS day values for the mean of the 12 years of study data in the Inter-Mountain Basins Big Sagebrush Shrubland vegetative community.

Figure 13. Comparison of predicted (RF Model) and observed (NDVI) SOS day values for the mean of the 12 years of study data in the Northwestern Great Plains Mixedgrass Prairie vegetative community.
Figure 14. Comparison of predicted (RF Model) and observed (NDVI) SOS day values for the mean of the 12 years of study data in the Western Great Plains Shortgrass Prairie vegetative community.

Figure 15. Comparisons for years 2001-2006 of predicted and observed SOS day at 6000 randomly sampled points.
Figure 16. Comparisons for years 2007-2012 of predicted and observed SOS day at 6000 randomly sampled points.

Chapter 5: Discussion

5.1. Fulfillment of Research Objectives

This study shows that random forest modeling can be used for determining the most important climate variables to phenology as well as predicting phenology on a large spatial scale. The random forest model performed well at selecting the most important variables for predicting SOS, EOS and MAX phenology and was able to predict those phenological events with a high degree of accuracy (Figures 8-11). SOS day was most highly correlated with daylength, AGDD, minimum temperature, long term precipitation, elevation and short term precipitation in that order (Figure 6). However,
the SOS RF model’s variable importance output ranked the variables differently than the linear correlation of the variables to SOS (Figure 3) indicating that a simple linear correlation does not fully capture the importance of each variable or their complex interaction that leads to start of season. The vertical trend observed in the Tmin min 14 graph within Figure 6 representing a lower temperature threshold between -12°C and -5°C is similar to previous models that used a biologically based minimum temperature threshold of -2°C (Jolly, et al., 2005). The lower thresholds found in this graph are due to the variable being the lowest minimum temperature measured covering a two week period prior to SOS day, which will always be a lower value than the static temperature threshold crossed just before plant growth begins on the SOS day.

The importance of AGDD to SOS confirms results from previous studies (Frank & Hofmann, 1989; Henebry & de Beurs, 2013), and its prominence in all three models of variable importance (SOS, EOS and MAX) emphasizes the importance of AGDD as a quantifier of temperature effects on vegetation. Along with AGDD, Precipitation from day zero (PRCP 0) and day length (DAYL 14 and DAYL 21) also ranked highly in the three variable importance models (Figures 3-5). The importance of PRCP 0 for all three phenological phases is consistent with previous studies (Flanagan, 2009; Jolly, et al., 2005; Kathuroju, et al., 2007, Kovalskyy & Henerby, 2012; Piekielek, 2012) and not unexpected as PRCP 0 was one of only two variables, the other being precipitation from dormant season (PRCP dorm), that captured the effect of long term water availability on phenology. Though solar radiation was used in the EOS predictive model, day length always ranked higher than solar radiation in the variable importance models regardless of the temporal ranges for either (Figures 3-5). This result differed from previous studies that found solar radiation to be an important phenological variable (Piekielek, 2012). This difference is likely a result of differences of scale. This study was on a regional to continental scale with over 1.8 million km² pixels used to build the model, while Piekielek’s model was built on the scale of a single watershed of 360 km² (Piekielek, 2012). Both solar radiation and day length are measures of the energy available to plants, the difference is solar radiation is adjusted for atmospheric or land
surface properties, while day length is a measure of the length of each day, a function of latitude and day of the year. The general nature of day length makes it a better proxy for energy at a larger scale than solar radiation, while solar radiation’s adjustments for atmospheric and land surface properties make it a more appropriate measure of energy at smaller local scales. This result confirms the use of day length in general phenology models like the Growing Season Index (Jolly, et al., 2005) and related models (Kovalskyy & Henerby, 2012).

5.2. Error Rate of Predictive Models

The error rate for all the predictive models was one day or less which compares well with mechanistic models whose error rates for SOS day can vary from a half day up to a week (Jolly, et al, 2005; Kovalskyy & Henerby, 2012). There are several caveats to this result; the auto correlative nature of the AGDD, PRCP 0 and day length variables which were used in all three predictive models, most likely led to the high accuracy rate of the predictions. Also the model was predicting the dataset it had trained upon, and so the low error rates are to be expected. The selection of the study area also contributed to the low error rate as rangeland temperate areas do not have multiple layered canopies that can reduce the accuracy of remotely sensed data, such as NDVI.

The areas with consistently high errors across all three predictive models correspond with mountainous areas such as in central Colorado (Figures 8-11). In the mountainous areas snow cover is the most likely source of the modeling error as none of the three models include a snow cover variable which has been found to have an important impact on phenology in some areas (Piekielek, 2012). Snow cover will delay the SOS day even if all other energy, water and temperature thresholds are met (Piekielek, 2012). These errors could also be a function of the AGDD threshold observed in Figure 6 not being sufficiently accounted for by the RF model. The results of the yearly predicted SOS days (Figures 15 & 16) suggest that the predictive models are insensitive to extremely late SOS days that did not appear in the model’s training dataset.
5.3. Null Values in Predictive Models

The first SOS predictive model suffered from a significant amount of null values. These results were almost entirely a result of missing data from two variables PRCP 0 and AGDD, which had no values in areas with early start of season dates. When those variables were replaced with variables that would have values regardless of how early the start of season occurred the predictive model was able to produce a prediction. These predictions were extremely accurate at reproducing the NDVI observations (Figures 9-11). This accuracy is to be expected since these models are being compared against their training dataset, however it is a further validation that the variables selected to inform these models are actually correlated to the modeled phenological events.

5.4. Future Research

Further work is necessary to make prediction on future phenology as the current model is dependent on already knowing some SOS days to produce the variables to inform the model. In this study the development of the climate variables depended on a known phenological date in order to determine the temporal range of the climate variables, for example producing values for the fourteen days of minimum temperature preceding the start of season relies upon currently knowing the start of season day. This also led to a high autocorrelation between the variables and the phenological events being predicted. The limited prognostic power of this model is a common limitation of statistically modeling phenology, which is why most phenology models, such as the GSI, that inform larger climate models are mechanistic. While mechanistic models lose accuracy by making more simplified cause-effect assumptions for phenology, they gain the ability to project their models into the future.

Some statistical based models have been able to produce prognostic phenology predictions by separating the year into bi-weekly or 16 day blocks and using those blocks to inform their models (Piekielek, 2012). This approach would also be possible for the predictive models presented in this study; however the accuracy of the predictions
would suffer from the aggregation. This limitation on statistical models can be overcome to produce accurate statistically based future phenology estimates and this study suggests a way forward. Instead of limiting the production of variables based on a single day, or even a block of days, in this case SOS, EOS and MAX days, variables could be produced for every single day in the year. Then every single day and pixel would have a probability value for whether it was the growing season. This method would also eliminate the autocorrelation inherent in the variables used in this study. As demonstrated in this study, machine learning modeling techniques, such as random forests, could handle large amounts of variables interacting in complex ways and select the most important for phenology to produce accurate predictions. In this case the predictions would not be a set SOS, MAX, or EOS day, but the probability that each day was a growing season day. The transferability of machine learning techniques, such as Random Forest, to time periods they were not trained on is questionable and would need to be addressed in order to use these models to predict future phenology (Dobrowski, et al., 2011). Advances in modeling software, such as the Rocky Mountain Research Station’s raster utility (Hogland & Anderson, 2014), make these complex and data intense modeling techniques increasingly accessible.

**Chapter 6: Conclusion**

This study accomplished its main goal of building a highly accurate phenological model that works on a large scale, by individualizing the model to each site using a decision tree modeling technique. The model was able to identify the climate variables important to rangeland phenology and use these variables to create predictive models. The models also showed the value in using random forest modeling to quickly identify the most important variables and demonstrated the potential to build an accurate landscape model using those variables. The identification of limiting climate factors and their thresholds in Figure 6 confirm the importance of temperature and water availability in rangeland vegetative communities. Random forest models are a very efficient and accurate method for determining the value of variables across a landscape
or time period given a dataset on a small section of that landscape, as was
demonstrated in this study. The ability to quickly identify the climate variable limiting
growth in a rangeland is an important management tool that can be used to indicate the
effect changes in that climate variable would have upon a landscape.

The drawback of developing a model at this scale, however, is that the diversity
of smaller vegetative communities can be overlooked. The largest three vegetative
communities appeared to drive the parameterization of the models (Figures 12-14).
Though the model performed well for the vast majority of the study area, and was
responsive to much of the diversity of the modeled landscape, the outlier areas were
not modeled as accurately as the rest of the study area. The models’ spatially located
error rates, as well as the high parameterization to the larger vegetative communities,
indicate that the scale of this model is too large to accurately replicate phenology at the
extreme of rangeland habitats. Instead, a scaled down version of these models, that
includes phenology variables of local importance, would be required at those locations.

The variable importance by-product of the random forest method was
demonstrated to be an efficient means of selecting several variables among many that
are most correlated with the dependent variable, in this case SOS, EOS and MAX days.
However, the current model was not designed to predict future phenology, because the
climate variables were calculated based upon a known phenological date. This is not
surprising considering it is an inherent limitation of statistical modeling, which is why
most forecasting climate models use mechanistic phenology models that set predefined
thresholds on a few environmental variables that represent energy, water and
temperature, such as the growing season index (GSI) (Jolly, et al., 2005).

Precipitation was the limiting factor in SOS day for majority of rangelands in this
study area (Figure 6). This is an important finding to note as this study area consists of
important habitat for elk. When future research considers the impacts of climate change
on these species, changes in precipitation should be considered more important in
these areas. While in the high elevation and latitude areas temperature was found to be
the limiting factor and greatest source of change (Figure 6). These separate climate
drivers of phenology could potentially lead to a more homogenous spatial phenology in elk habitat. If precipitation decreases in the water limited areas growing seasons could start later in the year while increased temperature in high elevation areas could lead to growing seasons starting earlier in the year, leading to less temporal diversity for rangeland migrating species. If this occurs there could be a negative impact on the populations of migratory species.

References


Appendix A.

List of all climate variables used in this study, their code names, temporal range, unit and how they were calculated. All data are for years 2001 through 2012 unless otherwise noted.

sos_agdd        The sum of average daily temperatures in Celsius above zero from the zero day to the start of season day inclusive

sos_dayl14_av   The mean of daylength in seconds per day (s/day) from 14 days prior to start of season to start of season inclusive

sos_dayl21_av   The mean of daylength in seconds per day (s/day) from 21 days prior to start of season to start of season inclusive

sos_prcpdorm   The sum of precipitation in millimeters per day (mm/day) from the prior end of season day to the start of season day inclusive, NOTE: from 2002 to 2012

sos_prcp0_mn    The sum of precipitation in mm/day from day zero to the start of season inclusive

sos_prcp21_mn   The sum of precipitation in mm/day from 21 days prior to start of season to the start of season inclusive

sos_prcp14_mn   The sum of precipitation in mm/day from 14 days prior to start of season to the start of season inclusive

sos_srad_av14   The 14 day mean of the Incident shortwave radiation flux density in watts per square meter, taken as an average over the daylight period of the day, from 14 days prior to start of season to start of season inclusive. NOTE: Daily total radiation (MJ/m2/day) can be calculated as follows: ((srad (W/m2) * dayl (s/day)) / 1,000,000)

sos_srad_av21   The 21 day mean of the Incident shortwave radiation flux density in watts per square meter, taken as an average over the daylight period of the day, from 21 days prior to start of season to start of season inclusive.

sos_tmax_av14   The mean of daily maximum 2-meter air temperature in degrees Celsius from 14 days prior to start of season to start of season inclusive.

sos_tmax_avg    The mean of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to start of season to start of season inclusive.

sos_tmax_max    The maximum of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to start of season to start of season inclusive.

sos_tmax_min    The minimum of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to start of season to start of season inclusive.
sos_tmin_av14  The mean of daily minimum 2-meter air temperature in degrees Celsius from 14 days prior to start of season to start of season inclusive

sos_tmin_avg  The mean of daily minimum 2-meter air temperature in degrees Celsius from 21 days prior to start of season to start of season inclusive

sos_tmin_min  The minimum of daily minimum 2-meter air temperature in degrees Celsius from 21 days prior to start of season to start of season inclusive

sos_tmin_min14  The minimum of daily maximum 2-meter air temperature in degrees Celsius from 14 days prior to start of season to start of season inclusive

sos_vp_avg14  The mean of daily average partial pressure of water vapor pressure in pascals (Pa) from 14 days prior to start of season to start of season inclusive

sos_vp_avg21  The mean of daily average partial pressure of water vapor pressure in pascals (Pa) from 21 days prior to start of season to start of season inclusive

sos_vp_sum0  The sum of daily average partial pressure of water vapor pressure in pascals (Pa) from zero day to start of season inclusive

max_agdd_mean  The sum of average daily temperatures in Celcius above zero from the zero day to the max of season day inclusive

max_agdds_mn  The sum of average daily temperatures in Celcius above zero from the max of season day to the max of season day inclusive

max_dayl14  The mean of daylength in seconds per day (s/day) from 14 days prior to max of season to max of season inclusive

max_dayl21  The mean of daylength in seconds per day (s/day) from 21 days prior to max of season to max of season inclusive

max_prcp_sum0  The sum of precipitation in mm/day from day zero to the max of season inclusive

max_prcp21  The sum of precipitation in mm/day from 21 days prior to start of season to the max of season inclusive

max_prcps  The sum of precipitation in mm/day from from the max of season day to the max of season day inclusive

max_srad14_mn  The 14 day mean of the Incident shortwave radiation flux density in watts per square meter, taken as an average over the daylight period of the day, from 14 days prior to max of season to max of season inclusive. NOTE: Daily total radiation (MJ/m2/day) can be calculated as follows: ((srad (W/m2) * dayl (s/day)) / 1,000,000)
max_srad21_mn  The 21 day mean of the Incident shortwave radiation flux density in watts per square meter, taken as an average over the daylight period of the day, from 21 days prior to max of season to max of season inclusive

max_tmax_avg  The mean of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to max of season to max of season inclusive

max_tmaxa14  The mean of daily maximum 2-meter air temperature in degrees Celsius from 14 days prior to max of season to max of season inclusive

max_tmax_max  The maximum of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to max of season to max of season inclusive

max_tmax_min  The minimum of daily maximum 2-meter air temperature in degrees Celsius from 21 days prior to max of season to max of season inclusive

max_tmin_av14  The mean of daily minimum 2-meter air temperature in degrees Celsius from 14 days prior to max of season to max of season inclusive

max_tmin_avg  The mean of daily minimum 2-meter air temperature in degrees Celsius from 21 days prior to max of season to max of season inclusive

max_tmin_min  The minimum of daily minimum 2-meter air temperature in degrees Celsius from 21 days prior to max of season to max of season inclusive

max_tmina14  The mean of daily maximum 2-meter air temperature in degrees Celsius from 14 days prior to max of season to max of season inclusive

max_vp_avg14  The mean of daily average partial pressure of water vapor pressure in pascals (Pa) from 14 days prior to max of season to max of season inclusive

max_vp_avg21  The mean of daily average partial pressure of water vapor pressure in pascals (Pa) from 21 days prior to max of season to max of season inclusive

max_vp_sum0  The sum of daily average partial pressure of water vapor pressure in pascals (Pa) from zero day to max of season inclusive

max_vp_sums  The sum of daily average partial pressure of water vapor pressure in pascals (Pa) from start of season to max of season inclusive
Appendix B.

Table of the random forest variable importance error values and rank for each climate variable.

<table>
<thead>
<tr>
<th>SOS Model</th>
<th>RMSE</th>
<th>Average Error</th>
<th>Average Relative Error</th>
<th>Variable name</th>
<th>RMSE Rank</th>
<th>Average Error Rank</th>
<th>Average Relative Error Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.31820</td>
<td>3.2810</td>
<td>0.1197</td>
<td>PRCP 0</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6.26008</td>
<td>3.4624</td>
<td>0.1154</td>
<td>dayl 14</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6.44282</td>
<td>3.4608</td>
<td>0.1124</td>
<td>AGDD</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6.22805</td>
<td>3.4893</td>
<td>0.1137</td>
<td>dayl 21</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6.06860</td>
<td>3.1959</td>
<td>0.1089</td>
<td>elevation</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6.03177</td>
<td>3.1372</td>
<td>0.1094</td>
<td>Tmax avg 21</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6.03024</td>
<td>3.1230</td>
<td>0.1086</td>
<td>PRCP 14</td>
<td>8</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>5.99815</td>
<td>3.1344</td>
<td>0.1086</td>
<td>Tmax max 21</td>
<td>12</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6.01321</td>
<td>3.1331</td>
<td>0.1079</td>
<td>EVT</td>
<td>11</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>5.99169</td>
<td>3.1383</td>
<td>0.1085</td>
<td>Tmax min 21</td>
<td>16</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>6.02395</td>
<td>3.1228</td>
<td>0.1081</td>
<td>PRCP 21</td>
<td>9</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>6.04250</td>
<td>3.1177</td>
<td>0.1077</td>
<td>Tmin min 14</td>
<td>6</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>6.02190</td>
<td>3.1230</td>
<td>0.1075</td>
<td>PRCP dorm</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>5.99439</td>
<td>3.1139</td>
<td>0.1076</td>
<td>Tmax avg 14</td>
<td>15</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>5.97945</td>
<td>3.1262</td>
<td>0.1066</td>
<td>VP avg 14</td>
<td>18</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>5.99488</td>
<td>3.1122</td>
<td>0.1069</td>
<td>Tmin avg 21</td>
<td>14</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5.97517</td>
<td>3.1223</td>
<td>0.1074</td>
<td>VP avg 21</td>
<td>20</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>5.99707</td>
<td>3.1049</td>
<td>0.1066</td>
<td>SRAD avg 14</td>
<td>13</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>5.98057</td>
<td>3.1140</td>
<td>0.1064</td>
<td>Tmin min 21</td>
<td>17</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>5.97927</td>
<td>3.1066</td>
<td>0.1061</td>
<td>Tmin avg 14</td>
<td>19</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>5.92583</td>
<td>3.0810</td>
<td>0.1055</td>
<td>SRAD avg 21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EOS Model</th>
<th>RMSE</th>
<th>Average Error</th>
<th>Average Relative Error</th>
<th>Variable name</th>
<th>RMSE Rank</th>
<th>Average Error Rank</th>
<th>Average Relative Error Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.801945</td>
<td>3.955762333</td>
<td>0.013768602</td>
<td>PRCP M</td>
<td>23</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>6.813446</td>
<td>3.971082182</td>
<td>0.013838768</td>
<td>PRCP21</td>
<td>22</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>6.816253</td>
<td>3.952169023</td>
<td>0.013767777</td>
<td>Tmax min</td>
<td>21</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>6.816764</td>
<td>3.962512878</td>
<td>0.013806078</td>
<td>PRCP 14</td>
<td>20</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>6.817797</td>
<td>3.975317954</td>
<td>0.013853781</td>
<td>PRCP S</td>
<td>19</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>6.827216</td>
<td>3.964821444</td>
<td>0.013810208</td>
<td>VP avg 14</td>
<td>18</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>6.829649</td>
<td>3.972457358</td>
<td>0.013822658</td>
<td>Tmax 14</td>
<td>17</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>6.835357</td>
<td>3.974691871</td>
<td>0.013852124</td>
<td>Tmin min</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Variable name</td>
<td>RMSE</td>
<td>Average Error</td>
<td>Average Relative Error</td>
<td>RMSE Rank</td>
<td>Average Error Rank</td>
<td>Average Relative Error Rank</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>---------------</td>
<td>------------------------</td>
<td>-----------</td>
<td>--------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>elevation</td>
<td>10.76523</td>
<td>6.483946027</td>
<td>0.036257424</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AGDD</td>
<td>9.774022</td>
<td>5.961262346</td>
<td>0.033257423</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>AGDD S</td>
<td>9.25708</td>
<td>5.71786401</td>
<td>0.031994138</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Dayl 21</td>
<td>9.226001</td>
<td>5.689686694</td>
<td>0.031976</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Dayl 14</td>
<td>9.210946</td>
<td>5.688294544</td>
<td>0.031926</td>
<td>9</td>
<td>11</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>VP Sum 0</td>
<td>9.440645</td>
<td>5.703984845</td>
<td>0.031767788</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>PRCP 0</td>
<td>9.210568</td>
<td>5.607542193</td>
<td>0.031195565</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>EVTU</td>
<td>9.288978</td>
<td>5.589974046</td>
<td>0.031179209</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>PRCP S</td>
<td>9.23501</td>
<td>5.571941408</td>
<td>0.031110735</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>SRAD 21</td>
<td>9.116677</td>
<td>5.557483043</td>
<td>0.031084216</td>
<td>13</td>
<td>13</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>VP avg 14</td>
<td>9.140469</td>
<td>5.559791846</td>
<td>0.031036466</td>
<td>11</td>
<td>5</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>PRPC 14</td>
<td>9.346462</td>
<td>5.553954011</td>
<td>0.031009165</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>SRAD 14</td>
<td>9.082403</td>
<td>5.540361929</td>
<td>0.030936121</td>
<td>18</td>
<td>14</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>VP avg 21</td>
<td>9.119838</td>
<td>5.532481508</td>
<td>0.030920517</td>
<td>12</td>
<td>18</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Tmin min</td>
<td>9.050398</td>
<td>5.528448418</td>
<td>0.030898544</td>
<td>22</td>
<td>15</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Tmax 14</td>
<td>9.093785</td>
<td>5.528508185</td>
<td>0.030886109</td>
<td>15</td>
<td>19</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Tmax max</td>
<td>9.087095</td>
<td>5.535445799</td>
<td>0.030865717</td>
<td>16</td>
<td>17</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>VP Sum S</td>
<td>9.065611</td>
<td>5.52102375</td>
<td>0.03085044</td>
<td>19</td>
<td>21</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Tmax min</td>
<td>9.083381</td>
<td>5.519103423</td>
<td>0.03081683</td>
<td>17</td>
<td>12</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Tmin 21</td>
<td>9.095979</td>
<td>5.519142399</td>
<td>0.030790204</td>
<td>14</td>
<td>16</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Tmax max</td>
<td>9.064918</td>
<td>5.511967542</td>
<td>0.030778899</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>PRCP 21</td>
<td>9.039047</td>
<td>5.495892927</td>
<td>0.030699584</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

51