USING BAYESIAN REGRESSION TREE MODELS AND REMOTELY SENSED DATA TO
CHARACTERIZE RECENT ENVIRONMENTAL CHANGE IN ALASKA, USA

by

JOANN W. HARVEY

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Ecology and Evolution

Written under the direction of

Edwin J. Green

And approved by

_______________________________________

_______________________________________

_______________________________________

_______________________________________

New Brunswick, New Jersey

OCTOBER, 2011
ABSTRACT OF THE DISSERTATION

Using Bayesian Regression Tree Models and Remotely Sensed Data to Characterize Recent Environmental Change in Alaska, USA

By JOANN W. HARVEY

Dissertation Director:
Edwin J. Green

Remotely sensed Advanced Very High Resolution Radiometer (AVHRR) images, collected between 1995 and 2007, and Bayesian Regression Tree Modeling were brought together to characterize growing season environmental (vegetation, temperature, precipitable water, and cloudiness) change in Alaska. This method highlighted general trends and local variation.

The method was applied in two stages to reduce the effects of cloudiness upon the results and reveal the temporal distribution of cloudiness conditions. A reversible form of tree model “subtree replacement” was included in the Reversible Jump MCMC algorithm. A sensitivity analysis showed that larger values of some hyperprior parameters could increase the number of subsets delineated by the method.

For data collected during 1995 – 2002, the analyses showed local variation and subtle changes. In 2003, conditions of higher precipitable water, higher Normalized Difference Vegetation Index (NDVI), and/or greater cloudiness were highlighted. In 2004, the analyses detected a shift to lower precipitable water and/or lower cloudiness, often accompanied or followed by lower NDVI and higher land surface temperature. In 2007, continued warming was highlighted in the Arctic and northern interior regions, in contrast with a return to earlier conditions and increased cloudiness revealed in regions near the Bering Sea.
Dedication

To Gene
Acknowledgements

I wish to gratefully acknowledge the guidance I received from my advisor, Edwin Green, and my committee members, Richard Lathrop, Peter Morin, Peter Smouse, and William Strawderman.
Table of contents

ABSTRACT OF THE DISSERTATION .................................................................................... ii
Dedication ......................................................................................................................... iii
Acknowledgements ........................................................................................................ iv
List of tables ..................................................................................................................... ix
List of illustrations .......................................................................................................... xv
Foreword ........................................................................................................................ xxii
Introduction ..................................................................................................................... 1
Background ....................................................................................................................... 2
Concepts ........................................................................................................................... 4
Comparison with two time series analyses, one sophisticated and one simple .......... 8
Study Location ................................................................................................................ 10
Methods ........................................................................................................................... 12
Snapshots ......................................................................................................................... 12
Particles moving through state space ............................................................................ 13
Models that explore association .................................................................................... 13
Bayesian Regression Tree Models ................................................................................ 14
Decision trees and tree models ...................................................................................... 14
Cloudiness concerns and a two stage process .............................................................. 15
Non-Bayesian tree modeling methods vs. Bayesian tree modeling methods .......... 18
The Bayesian Regression Tree Model ................................................................. 19

Likelihood of the data given the tree structure ................................................. 24

Process details and a new move .................................................................... 24

Training and validation subsets ...................................................................... 25

Results ............................................................................................................ 26

Cloudiness analysis ....................................................................................... 26

Second stage analysis of Cloudiness Analysis Subset 2 ................................. 31

Discussion ..................................................................................................... 40

Illustrations and discussion of the results of the analyses of the remaining Permafrost Areas ........................................................................................................... 42

Descriptions and illustrations of the remaining Permafrost Areas ............... 42

Discussion of the remaining Permafrost Areas ............................................. 54

Support for the findings of these models from other sources ..................... 55

Model validation ............................................................................................ 56

The sensitivity analysis .................................................................................. 57

Conclusions .................................................................................................... 58

Appendix A ...................................................................................................... 60

Remotely Sensed Indices, Permafrost Areas, and Data Preparation ............. 60

Permafrost Areas ............................................................................................ 62

Data Preparation ............................................................................................. 65

Appendix B ..................................................................................................... 67
Additional Information About the Bayesian Regression Tree Models .................. 67

Reversible Jump MCMC applied to tree models .............................................. 67

Adjusted Likelihoods .......................................................................................... 69

Specifications for these analyses ....................................................................... 69

Appendix C ........................................................................................................... 73

C++ Program ....................................................................................................... 73

Appendix D ........................................................................................................... 128

Additional Results .............................................................................................. 128

Permafrost Area 11 ............................................................................................ 129

Permafrost Area 11, Cloudiness Analysis Subset 2 ............................................. 133

Permafrost Area 12 ............................................................................................ 141

Permafrost Area 12, Cloudiness Analysis Subset 1 ............................................. 145

Permafrost Area 13 ............................................................................................ 154

Permafrost Area 13, Cloudiness Analysis Subset 1 ............................................. 158

Permafrost Area 22 ............................................................................................ 167

Permafrost Area 22, Cloudiness Analysis Subset 1 ............................................. 171

Permafrost Area 23 ............................................................................................ 180

Permafrost Area 23, Cloudiness Analysis Subset 0 ............................................. 184

Permafrost Area 24 ............................................................................................ 193

Permafrost Area 24, Cloudiness Analysis Subset 1 ............................................. 197
List of tables

Table 1. Number of observations in the subsets determined by the cloudiness analysis of Permafrost Area 21 ........................................................................................................................................... 18

Table 2. Tree structure (formulas from Denison et al., 2002) ........................................................................................................ 22

Table 3. Subsets given tree structure (formulas from Denison et al., 2002) ............................................................................. 23

Table 4. Subset boundaries and collection year summary for the validation set of Permafrost Area 21 ........................................................................................................................................... 29

Table 5. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Permafrost Area 21 ........................................................................................................................................... 29

Table 6. Number of observations within subsets by collection year for the validation set of Permafrost Area 21 ........................................................................................................................................... 29

Table 7. Subset boundaries and collection year summary for the validation set of Cloudiness Analysis Subset 2 of Permafrost Area 21 ........................................................................................................................................... 34

Table 8. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Cloudiness Analysis Subset 2 of Permafrost Area 21 ........................................................................................................................................... 35

Table 9. Number of observations within subsets by collection year, for the Validation Set of Cloudiness Subset 2 of Permafrost Area 21 ........................................................................................................................................... 35

Table 10. Number of observations within subsets by collection year, for the Training Set of Cloudiness Subset 2 of Permafrost Area 21 ........................................................................................................................................... 56

Table 11. Number of subsets produced per second stage tree model by Permafrost Area, Cloudiness Subset, and hyperprior values ........................................................................................................................................... 57

Table 12. Permafrost Areas with their associated geographic regions, ecoregions, and climate regions ........................................................................................................................................... 63
Table 13. Subset boundaries and collection year summary, sorted by subset, for the validation set of Permafrost Area 11

Table 14. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Permafrost Area 11

Table 15. Number of observations within subsets by collection year from the validation set of Permafrost Area 11

Table 16. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 2 of Permafrost Area 11

Table 17. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 2 of Permafrost Area 11

Table 18. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 2 of Permafrost Area 11

Table 19. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 12

Table 20. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 12

Table 21. Number of observations within subsets by collection year from validation set of Permafrost Area 12

Table 22. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 12

Table 23. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 12

Table 24. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 12
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 13</td>
</tr>
<tr>
<td>26</td>
<td>Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 13</td>
</tr>
<tr>
<td>27</td>
<td>Number of observations within subsets by collection year for Permafrost Area 13</td>
</tr>
<tr>
<td>28</td>
<td>Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 13</td>
</tr>
<tr>
<td>29</td>
<td>Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 13</td>
</tr>
<tr>
<td>30</td>
<td>Number of observations within subsets by collection year, for cloudiness subset 1 of permafrost area 13</td>
</tr>
<tr>
<td>31</td>
<td>Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 22</td>
</tr>
<tr>
<td>32</td>
<td>Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 22</td>
</tr>
<tr>
<td>33</td>
<td>Number of observations within subsets by collection year from the validation set of Permafrost Area 22</td>
</tr>
<tr>
<td>34</td>
<td>Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 22</td>
</tr>
<tr>
<td>35</td>
<td>Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 22</td>
</tr>
<tr>
<td>36</td>
<td>Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 22</td>
</tr>
</tbody>
</table>
Table 37. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 23 ................................................................. 182

Table 38. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 23................................................................. 182

Table 39. Number of observations within subsets by collection year from validation set of Permafrost Area 23................................................................. 182

Table 40. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 0 of Permafrost Area 23................................................................. 186

Table 41. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 0 of Permafrost Area 23................................................................. 187

Table 42. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 0 of Permafrost Area 23................................................................. 188

Table 43. Subset boundaries and collection year subset, sorted by subset, for Permafrost Area 24 ................................................................................................. 195

Table 44. Subset boundaries and collection year subset, sorted by mean collection year, for Permafrost Area 24 ................................................................................................. 195

Table 45. Number of observations within subsets by collection year from the validation set of Permafrost Area 24 ................................................................................................. 195

Table 46. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 24................................................................. 199

Table 47. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 24................................................................. 199

Table 48. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of permafrost Area 24................................................................. 199
Table 49. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 25 ............................................................................................................................................ 206

Table 50. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 25 ............................................................................................................................................ 206

Table 51. Number of observations within subsets by collection year from validation set of Permafrost Area 25 ............................................................................................................................................ 206

Table 52. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 25 ............................................................................................................................................ 210

Table 53. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 25 ............................................................................................................................................ 211

Table 54. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 25 ............................................................................................................................................ 211

Table 55. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 26 ............................................................................................................................................ 218

Table 56. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 26 ............................................................................................................................................ 218

Table 57. Number of observations within subsets by collection year from the validation set of Permafrost Area 26 ............................................................................................................................................ 218

Table 58. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 26 ............................................................................................................................................ 222

Table 59. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 26 ............................................................................................................................................ 223

Table 60. Number of observations within subsets by collection year, from validation set of Cloudiness Subset 1 of Permafrost Area 26 ............................................................................................................................................ 224
Table 61. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 31
................................................................. 231
Table 62. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 31
........................................................................................................ 231
Table 63. Number of observations within subsets by collection year for Permafrost Area 31
........................................................................................................ 231
Table 64. Subset boundaries and collection year summary, sorted by subset, for Cloudiness analysis Subset 3 of Permafrost Area 31
........................................................................................................ 235
Table 65. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness analysis Subset 3 of Permafrost Area 31
........................................................................................................ 235
Table 66. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 3 of Permafrost Area 31
........................................................................................................ 235
Table 67. Training Data Set Frequencies by subset and year, for Area 11, Cloudiness Subset 2
........................................................................................................ 241
Table 68. Training Data Set Frequencies by subset and year, for Area 12, Cloudiness Subset 1
........................................................................................................ 241
Table 69. Training Data Set Frequencies by subset and year, for Area 13, Cloudiness Subset 1
........................................................................................................ 242
Table 70. Training Data Set Frequencies by subset and year, for Area 22, Cloudiness Subset 1
........................................................................................................ 242
Table 71. Training Data Set Frequencies by subset and year, for Area 23, Cloudiness Subset 0
........................................................................................................ 243
Table 72. Training Data Set Frequencies by subset and year, for Area 24, Cloudiness Subset 1
........................................................................................................ 243
Table 73. Training Data Set Frequencies by subset and year, for Area 25, Cloudiness Subset 1
........................................................................................................ 244
Table 74. Training Data Set Frequencies by subset and year, for Area 26, Cloudiness Subset 1
........................................................................................................ 245
Table 75. Training Data Set Frequencies by subset and year, for Area 31, Cloudiness Subset 3
List of illustrations

Figure 1. Quantile-quantile plot showing approximate normality of a subset’s collection years ... 8
Figure 2. Time series plots of annual mean percent difference from study period mean LST, PWI, and NDVI for Cloudiness Subset 2 of Permafrost Area 21, 1995 – 2007 ....................... 10
Figure 3. Tree model produced by the cloudiness analysis of Permafrost Area 21 .................. 28
Figure 4 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 21. 30
Figure 5. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 2 of Permafrost Area 21 ................................................................................................................. 33
Figure 6. Sequence of maps of Area 21 showing the geographic progression of environmental conditions through the study period (using the Validation Set of Cloudiness Subset 2) ........ 36
Figure 7. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the NDVI (V) by PWI (P) plane, by collection year, with colors indicating the subsets ........................................................................................................................................... 37
Figure 8. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the PWI (P) by LST (T) plane, by collection year, with colors indicating the subsets ........................................................................................................................................... 38
Figure 9. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the LST (T) by NDVI (V) plane, by collection year, with colors indicating the subsets ........................................................................................................................................... 39
Figure 10. Sequence of maps of Permafrost Area 11 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 2) ....................... 45
Figure 11. Sequence of maps of Permafrost Area 12 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1) ....................... 46
Figure 12. Sequence of maps of Permafrost Area 13 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)..........................47

Figure 13. Sequence of maps of Permafrost Area 22 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)..........................48

Figure 14. Sequence of maps of Permafrost Area 23 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 0)..........................49

Figure 15. Sequence of maps of Permafrost Area 24 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)..........................50

Figure 16. Sequence of maps of Permafrost Area 25 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)..........................51

Figure 17. Sequence of maps of Permafrost Area 26 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)..........................52

Figure 18. Sequence of maps of Permafrost Area 31 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 3)..........................53

Figure 19. Map of Training Sample from 1995 showing the permafrost areas..........................63

Figure 20. Tree model produced by the cloudiness analysis of Permafrost Area 11 ..................130

Figure 21 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 11 ..............................................................................................................................................132

Figure 22. Tree model resulting from the second stage analysis of Cloudiness Subset 2 of Permafrost Area 11.......................................................................................................................................134

Figure 23. Sequence of maps of Permafrost Area 11 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 2)..........................137

Figure 24. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space ..................................................................................................................138
Figure 25. Projection of the validation data points onto the precipitable water vs. temperature plane of state space ........................................................................................................................................... 139

Figure 26. Projection of the validation data points onto the temperature vs. vegetation plane of state space ........................................................................................................................................... 140

Figure 27. Tree Model produced by the cloudiness analysis of Permafrost area 12 ............ 142

Figure 28 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 12 ........................................................................................................................................... 144

Figure 29. Tree model resulting from the second stage analysis of Cloudiness Subset 1 of Permafrost Area 12 ........................................................................................................................................... 147

Figure 30. Sequence of maps of Permafrost Area 12 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1) ............... 150

Figure 31. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space ........................................................................................................................................... 151

Figure 32. Projection of the validation data points onto the precipitable water vs. temperature plane of state space ........................................................................................................................................... 152

Figure 33. Projection of the validation data points onto the temperature vs. vegetation plane of state space ........................................................................................................................................... 153

Figure 34. Tree model produced by the cloudiness analysis of Permafrost Area 13 ............ 155

Figure 35 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 13 (validation set) ........................................................................................................................................... 157

Figure 36. Tree model resulting from the second stage analysis of Cloudiness Subset 1 of Permafrost Area 13 ........................................................................................................................................... 160

Figure 37. Sequence of maps of Permafrost Area 13 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1) ............... 163
Figure 38. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space .............................................................. 164

Figure 39. Projection of the validation data points onto the precipitable water vs. temperature plane of state space ......................................................................... 165

Figure 40. Projection of the validation data points onto the temperature vs. vegetation plane of state space ........................................................................................................ 166

Figure 41. Tree Model produced by the cloudiness analysis of Permafrost area 22 ............ 168

Figure 42 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 22 ........................................................................................................................................ 170

Figure 43. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 22 ........................................................................................................ 172

Figure 44. Sequence of maps of Permafrost Area 22 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1).......................... 176

Figure 45. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space ........................................................................................................ 177

Figure 46. Projection of the validation data points onto the precipitable water vs. temperature plane of state space ........................................................................................................ 178

Figure 47. Projection of the validation data points onto the temperature vs. vegetation plane of state space ........................................................................................................ 179

Figure 48. Tree model produced by the cloudiness analysis of Permafrost Area 23 ............ 181

Figure 49 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 23 (Validation Set) ........................................................................................................ 183

Figure 50. Tree model resulting from the second stage analysis of cloudiness analysis subset 0 of Permafrost Area 23 ........................................................................................................ 185
Figure 51. Sequence of maps of Permafrost Area 23 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 0) ........................................ 189

Figure 52. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space .................................................................................................................. 190

Figure 53. Projection of the validation data points onto the precipitable water vs. temperature plane of state space .................................................................................................................. 191

Figure 54. Projection of the validation data points onto the temperature vs. vegetation plane of state space .................................................................................................................. 192

Figure 55. Tree model produced by the cloudiness analysis of Permafrost Area 24 .............. 194

Figure 56 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 24 (Validation Set) .................................................................................................................. 196

Figure 57. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 24 .................................................................................................................. 198

Figure 58. Sequence of maps of Permafrost Area 24 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1) .............. 200

Figure 59. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space .................................................................................................................. 201

Figure 60. Projection of the validation data points onto the precipitable water vs. temperature plane of state space .................................................................................................................. 202

Figure 61. Projection of the validation data points onto the temperature vs. vegetation plane of state space .................................................................................................................. 203

Figure 62. Tree model produced by the cloudiness analysis of Permafrost Area 25 .............. 205

Figure 63 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 25 (validation set) .................................................................................................................. 207
Figure 64. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 25

Figure 65. Sequence of maps of Permafrost Area 25 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)

Figure 66. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space

Figure 67. Projection of the validation data points onto the precipitable water vs. temperature plane of state space

Figure 68. Projection of the validation data points onto the temperature vs. vegetation plane of state space

Figure 69. Tree model produced by the cloudiness analysis of Permafrost Area 26

Figure 70 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 26 (validation set)

Figure 71. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 26

Figure 72. Sequence of maps of Permafrost Area 26 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)

Figure 73. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space

Figure 74. Projection of the validation data points onto the precipitable water vs. temperature plane of state space

Figure 75. Projection of the validation data points onto the temperature vs. vegetation plane of state space

Figure 76. Tree model produced by the cloudiness analysis of Permafrost Area 31
Figure 77 Spatial And Temporal Progression Of The Cloudiness Conditions of Permafrost Area 31 ..............................................................................................................................................................................232

Figure 78. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 3 of Permafrost Area 31..............................................................................................................................................................................234

Figure 79. Sequence of maps of Permafrost Area 31 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 3)..................................................................................236

Figure 80. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space ..............................................................................................................................................................................237

Figure 81. Projection of the validation data points onto the precipitable water vs. temperature plane of state space ..............................................................................................................................................................................238

Figure 82. Projection of the validation data points onto the temperature vs. vegetation plane of state space ..............................................................................................................................................................................239
Foreword

This study applies a novel form of environmental change detection to ten regions of Alaska, USA, that are known as Permafrost Areas. To avoid repetition, the main body of this report presents an analysis of the northernmost Permafrost Area of Alaska, Permafrost Area 21. This is the lowland and upland area underlain by thick permafrost (Ferrians, 1998). It is associated with the North Slope geographic region (National Geographic Society, 2002), the Arctic Foothills Tundra and Arctic Coastal Tundra ecoregions (Ricketts, et al., 1999), and the Arctic climate region (Shulski & Wendler, 2007).

The analyses of the nine other Permafrost Areas of Alaska are briefly summarized in the main body of this report. Details about the nine analyses will be available in Appendix D of this report and also online, on the website of The Grant F. Walton Center for Remote Sensing and Spatial Analysis (CRSSA) at http://crssa.rutgers.edu/projects/.
Introduction

As Earth's climate changes, some people are experiencing severe weather events such as hurricanes and tornadoes. Some are observing early ice breakup, forest fires, and floods. How should scientists document the varied aspects of climate change?

At some point, we will want to piece together the phases of the development of current climate change. We will want to consider its characteristics, consequences, and implications. We will need to view environmental change as a whole, and trace its progress. To accomplish this, we could observe via satellite sensors, interpret using data analysis, and communicate our findings via maps, tables, and graphs.

This study uses information provided by Advanced Very High Resolution (AVHRR) satellite sensors. The information is interpreted by the application of Bayesian Regression Tree Modeling, which will produce subsets of environmental conditions unique to given years or time periods within the study period: the growing seasons of 1995 through 2007. These subsets will be described, tabulated, and mapped. A picture of recent environmental change in Alaska will emerge.
Background

Climate is changing worldwide (Overland et al., 2008; Solomon, et al., 2007). It primarily involves warming, but other aspects of climate are changing, too (Solomon, et al., 2007). Researchers are studying the effects of climate change as it is happening (Francis et al., 2009; Overland J., 2009; Wu et al., 2006). To supplement this, documenting the trajectory of the change in climate would be a valuable undertaking, with the goal of providing retrospective information to researchers who seek to understand its consequences. Monitoring the changes in several environmental quantities simultaneously would provide a more complete picture of climate change, just as presenting statistics about tornado occurrence and severity during the current year in the continental United States, along with statistics on the warming observed during the current year in the continental United States, would provide a more complete picture of the current year's climate change.

Overland and colleagues (Overland et al., 2008) have compared Arctic Surface Air Temperatures and Sea Level Pressures of the 20th century with those of the early part of the 21st century, and they propose that the years 2000 through 2007 be named, “The Recent Arctic Warm Period.” They attribute this warming to two sources: actual warming consistent with the IPCC-AR4 model projections (Solomon, et al., 2007) and a newly observed dipole pattern in Arctic air pressure (named the Arctic Dipole) (Wu et al., 2006). The Arctic Dipole's pressure distribution increased the Arctic's south to north winds, bringing more "heat into the central Arctic Ocean" (Overland et al., 2009, p.8).

Remotely sensed images collected by satellites may be used to document current climate change. Remote sensing has particular value when compared to ground-based information gathering because remote sensing provides frequently updated coverage of difficult to visit areas during relatively recent time periods. It provides information, based upon reflected
and emitted radiation in several bandwidths (ranges of wavelengths), which allows the derivation of several different environmental indicators. For example, three indices, which together describe a range of environmental conditions: vegetation productivity (Tucker & Sellers, 1986), land surface temperature, and atmospheric precipitable water, may be derived from remotely sensed image data.

One can gain insight if one examines each of these indices separately. However, at any given time and place, the conditions represented by these indices are experienced simultaneously by the organisms that live in that place. So, a comprehensive view of environmental conditions, considering all three indices, would be valuable to those that study an ecosystem's response to climate change. This viewpoint is not often seen in the literature.

These three indices are related. One would expect to observe a dependence of vegetation productivity upon temperature and atmospheric precipitable water. As the climate changes rapidly, though, this relationship might itself change qualitatively and quantitatively (for example, consider the progress of succession in response to climate change). Although a model expressing the relationship between productivity and climatic conditions would be a laudable final goal, I propose that a researcher's first task would be to observe and document the changes in these indices, in order to subsequently obtain a more well-informed final model of vegetation dependence upon a changing climate. Therefore, the goal of this research is to support the first task: to provide a more complete documentation of recent climate change by bringing together several remotely sensed indices, and the years during which they were collected, and analyzing the collection of indices based upon their consistencies and inconsistencies from one year to the next.

The study period consisted of the 1995 through 2007 growing seasons. The indices that were used, which together provide a comprehensive view of environmental conditions, and
which may be readily derived from remotely sensed data, are NDVI (Normalized Difference Vegetation Index [Rouse et al., 1974; Tucker & Sellers, 1986]), a simplified version of the LST (Land Surface Temperature [Price, 1984]) index, and a simplified version of PWI (Precipitable Water Index [Eck & Holben, 1994]). The association between the indices and the collection year of the remotely sensed images would be determined.

Of course, cloudiness conditions must be taken into account. Clouds exist in the atmosphere; therefore they appear in many remotely sensed images. Clouds may be considered to be a part of the environmental conditions; however, clouds distort the information collected via remote sensing and can "create the false impression of spectral change between two dates" (Jensen, 2005, p. 471). To avoid spurious results due to cloudiness, an initial analysis stage was performed which partitioned the study data into strata that were relatively homogeneous with respect to cloudiness. Then a second analysis stage was performed in which we aligned collection year with remotely sensed variables within at least one of the strata identified in stage 1.

In the first stage of the analysis, the covariates were the rounded difference from study period mean of the percentage of seasonal cloudy images, early season cloudy images, and late season cloudy images. In the second stage of the analysis, the covariates were the rounded percent difference from study period mean NDVI, percent difference from study period mean (simplified) LST, and percent difference from study period mean (simplified) PWI.

Concepts

The methods used here employ novel concepts that must be discussed before this study is described in detail. To understand these concepts, the reader is asked to consider several metaphors: a collection of snapshots, an illustration of a particle moving through space, and an example of a statistical tree model.
First, consider an ecologist who is examining a set of photographs of a study site. These photographs had been collected yearly over the course of a decade or so. The differences in the photographs lead the ecologist to believe that environmental conditions had changed during the time period spanned by the photographs. The ecologist wonders if some of the photographs might show conditions that were unique to a single year in that timeframe, some might show conditions that existed for several years, and/or some might show conditions that recurred periodically. The ecologist wishes to associate the conditions shown in each photograph with the "date stamp" (in particular the year) of the photograph. *This study is concerned with illustrating the temporal progression of several environmental conditions, using information from remotely sensed snapshots of the conditions, along with the years during which they were taken.*

Next, imagine a particle moving through space or recall how strange attractors are described in discussions of chaos theory. To illustrate the "strange" properties of strange attractors, a trace of a particle, in time, through a coordinate space (which is considered to be the particle's state, or phase, space), is usually shown. Without concerning oneself with chaotic behavior, one may envision combining three time series to create a similar illustration, and, more importantly for this study, *one may consider a location of a particle in its state space and, using its location coordinates, approximate the time period, or time periods, during which the particle was found there.* In this study, we will be exploring the timing of environmental conditions by considering the trajectory through the "state space" of the observed conditions.

Statistical tree models, i.e., statistical models expressed in the form of a decision tree (questions arranged in a directed graph resembling [an upside-down] tree's branches, which one follows according to one's answers to the questions), are integral to this study. Before they are described in detail, consider an early, "real-world" example of a statistical tree model. This
example was provided by Breiman, et al. (1984), and it is a description of a study conducted by Goldman et al. (1982). Goldman et al.'s tree model was created to provide a speedy diagnosis of a heart attack (acute myocardial infarction) in an emergency room setting. The response variable of the model was a variable that indicated whether or not a patient with chest pain was suffering a heart attack, while the covariates were patient characteristics and symptoms. A decision tree containing questions about the patient's EKG (electrocardiogram), pain, and age was the result of the analysis. The collected answers to the questions, following the tree structure, would lead to a diagnosis. Technically, the answers would lead to locations in the model's state space which were associated with either Class 1 patients (without heart attacks) or Class 2 patients (with heart attacks). The investigators produced a useful and interpretable statistical model that was in the form of a decision tree. Also note that, clearly, they were not modeling causality, because patient characteristics and symptoms would not cause heart attacks. In this study, similarly, I use tree models and I am not modeling causality, but, rather, association.

Applying these concepts to this study, we may state that a region's remotely sensed environmental conditions were used as covariates in a tree model to determine their association with the collection years of the study period, i.e., the region's trajectory through environmental state space. In statistical modeling terms, the collection year was the response variable and the environmental conditions were the covariates.

One difficulty may remain. A basic idea of statistical modeling is that the response variable ($Y$) is expressed as the sum of a model term ($f(X)$) and an error term ($\varepsilon$):

$$ Y = f(X) + \varepsilon. $$

How does one reconcile the idea of an error term with the well-defined values of the collection years? Here the error term may be regarded as a proxy for the physical observation that
environmental conditions, such as climate oscillations (Aguado & Burt, 2004), develop and recede gradually over time and space. That is, for a given set of environmental conditions, the statistical distribution of its collection years could be expected to vary over several years and possibly resemble an approximately normal distribution.

Collection years are discrete but ordinal. Although a multinomial classification analysis is most often used when the response variable is discrete, the ordinality of this response variable and the idea of the development and recession of conditions led me to use regression tree analysis and its associated normal prior distribution in this study. Discrete values, such as brightness values obtained via remote sensing, have been assumed to be approximately normal by other investigators (Richards, 1993; Green, Smith, & Strawderman, 1994). The following QQ (quantile-quantile) plot illustrates that a subset from this analysis may have a nearly normal distribution (shown as an almost straight line of steps [although some skewness of the distribution may be indicated at the lower part of this plot]), even when its values are discrete.
Figure 1. Quantile-quantile plot showing approximate normality of a subset’s collection years

**Comparison with two time series analyses, one sophisticated and one simple**

A sophisticated time series analysis method known, in its non-Bayesian version, as Adaptive Spline Threshold Autoregressive (ASTAR, Lewis et al., 1991) or TSMARS (Lewis & Ray, 1997) and, in its Bayesian version, as BAYSTAR (Denison et al., 2002, pp. 121-124) may offer the best way to model the relationship between vegetation productivity and climatic conditions that was mentioned (as a "laudable final goal") earlier in this report. The method allows
autoregressive (i.e., lagged) terms to be included in a time series model in a nonlinear manner. The current study, however, is meant to be more illustrative, concerned with illustrating the change in climatic and vegetative conditions by examining their trajectory through state space.

Alternatively, for simplicity's sake, one may wish to consider a more elementary analysis, such as plotting three time series (NDVI, LST, PWI) of yearly index means, stratified by cloud cover and permafrost area. This is a simple option that generally indicates how the central tendencies of these variables relate to one another over time within different strata (Figure 2). This option may suffice for those who do not desire a greater level of detail about environmental change, but it may not offer a sufficient level of detail for investigators who are tracking the impact of environmental change upon ecosystems, communities, and/or populations in a given region.

In contrast with elementary plots, this report's method offers an algorithm which may be applied consistently to multivariate data. It reveals specific sets of conditions that are associated with specific time periods. The number of sets of conditions, and the combinations within each set, might be challenging to interpret, but a motivated investigator, concerned about climate change in his/her study area, may find the effort worthwhile.

For example, with this report's modeling method, extreme values are not lost as they would be when the data are summarized as time series of mean values. Extreme values, and the associations between them are important in ecosystems. For example, in boreal forests, Hinzman et al. (2006, p. 52) found that "Climate strongly influences fire severity and frequency, with the greatest aerial extent of burning in the hottest, driest years."

This report's modeling method provides greater detail about environmental conditions at the cost of being more complicated than simple time series plots of mean values. This offers investigators a choice. If the simpler plots provide all that is necessary for a particular
investigation, then they should be used. If, however, an investigation would be enhanced by knowing the association between specific sets of conditions and specific time periods, now an investigator has a way to extract this information.

Both of the time series alternatives presented here could be stratified by cloud cover and permafrost area. The stratification method of this report was used to create the cloudiness subset shown in Figure 2. It allows the differences from study period mean environmental indices to be viewed in a cloudiness stratum that includes study period mean cloudiness, providing consistency. Additionally, it excludes other cloudiness strata that exhibit extreme cloudiness values associated with particular time periods. So for either final goal, a sophisticated BAYSTAR time series analyses or simple univariate time series plots, an ecologist could consider using this report's method to create the cloudiness strata for the analysis.

![Figure 2. Time series plots of annual mean percent difference from study period mean LST, PWI, and NDVI for Cloudiness Subset 2 of Permafrost Area 21, 1995 – 2007](image)

Study Location

The climate is changing most rapidly near the earth's poles (Wu et al., 2006; Overland et al., 2008). This is one of the reasons that the region chosen for this study is the state of Alaska,
USA. Other reasons to study environmental change in Alaska include the following: remotely sensed data sets, in the form of AVHRR (Advanced Very High Resolution Radiometer) composite images of Alaska, are freely available for a relatively extended, albeit recent, time period (U.S. Geological Survey, 2009), from the EarthExplorer website (U.S. Geological Survey, 2007); Alaska covers a large area, for a U.S. state (1,593,444 sq km) (National Geographic Society, 2002); it is mostly undeveloped (with a population of 635,000) (National Geographic Society, 2002), which is desirable for environmental change detection because the direct effects of human activity upon this region will presumably be relatively small; it contains a variety of ecosystems: several types of tundra and boreal forests (Ricketts et al., 1999; Chapin III et al., 2006); it contains a variety of topographic features: mountains, lowlands, and uplands (Ferrians, 1998; National Geographic Society, 2002); its regions exhibit a variety of climates: arctic, maritime, and interior (continental) climates (Shulski & Wendler, 2007); and, finally, a classification of the varied areas within Alaska, based upon each area’s extent of permafrost, along with its elevation, is available (Ferrians, 1998).

This initial report will concentrate upon Permafrost Area 21, which is described as the lowland and upland area underlain by thick permafrost (Ferrians, 1998). It is associated with the North Slope geographic region (National Geographic Society, 2002), the Arctic Foothills Tundra and Arctic Coastal Tundra ecoregions (Ricketts, et al., 1999), and the Arctic climate region (Shulski & Wendler, 2007). Similar reports on the remaining areas of the state are available on the CRSSA website.
Methods

Snapshots

AVHRR (Advanced Very High Resolution Radiometer) 14-day composite images of Alaska provided the raw data for this study (U.S. Geological Survey, 2007; U.S. Geological Survey, 2009). Compositing and other means of correcting and calibrating the images (radiometric calibration, atmospheric correction *“for ozone, water vapor absorption, and Rayleigh scattering”*, and geometric registration) have been performed to improve image quality and to allow an investigator to compare images of an area over many years (U.S. Geological Survey, 2009). Each image, like a digital photograph, is divided into many square pixels. Each pixel represents a land area of 1.1 km x 1.1 km in size. This resolution is too large to detect microclimate trends, but it is hoped that it is sufficiently fine to detect general climate and environmental trends.

The raw data provided for each pixel is made up of radiation intensity values. The radiation was emitted by or reflected from the pixel's land area, and it was detected by a satellite's sensor within a set of bandwidths (ranges of wavelengths). Arithmetic combinations of these radiation intensity values result in indices that express the environmental conditions on the ground of, or in the atmosphere above, each pixel.

The study period consisted of the 1995 through 2007 growing seasons. Three indices that together provide a comprehensive view of environmental conditions, and that may be readily derived from AVHRR data, are NDVI (Normalized Difference Vegetation Index), a simplified version of the LST (Land Surface Temperature) index, and a simplified version of the PWI (Precipitable Water Index). These indices are discussed in detail in Appendix A.

For each pixel, seasonal mean and study period mean values were calculated for each index. Then, for each pixel, the percent change of its seasonal mean from its study period mean was calculated. These values were rounded to the nearest integral percentage point. These
values were used to explore the changes of the conditions with respect to time. For example, for the NDVI data from a given pixel that was collected during a given year, the difference between its yearly (growing season) mean and its study period mean, divided by its study period mean, multiplied by 100, and rounded to the nearest integer, would be its percent difference from study period mean NDVI. It is hoped that this normalization of each pixel's data contributes to consistency in the data even when the pixel locations differ with respect to slope, aspect, percent vegetation cover, substrate, percent bare rock or water, etc. Using the study period mean as a reference point, instead of, say, a single year’s mean, reduces the impact of a single year's conditions, which may be anomalous due to random causes, upon the results of the analysis.

**Particles moving through state space**

Assuming for the moment that any cloudiness issues have been handled effectively in an earlier analysis step, the state space of this analysis is the coordinate space determined by the rounded percent difference from study period mean of each pixel's NDVI (Normalized Difference Vegetation Index), simplified LST (Land Surface Temperature), and simplified PWI (Precipitable Water Index). Note that this state space may be considered to be discrete because all the percent differences from study period means have been rounded. Also note that, as time progresses, a pixel's changing set of conditions may be viewed as a path though its state space. Each location on a path in state space may thus be viewed as associated with one or more of the growing seasons (collection years) within the study period. This is the association that is modeled in this study.

**Models that explore association**

Recall the example of the tree model that was used in the medical diagnosis of heart attacks. The condition was associated with its symptoms, and patient symptoms were used as
covariates in a model to determine their association with the presence or absence of heart attacks. In this study, environmental conditions were used as covariates in a tree model to determine their association with the various collection years of the study period. In statistical modeling terms, the collection year was the response variable and the environmental conditions were the covariates. Even though the variables are named that way by convention, they are describing association, and not causality.

**Bayesian Regression Tree Models**

Bayesian Regression Tree Models are statistical models of ordinal data that take the form of a binary decision tree and are derived using Bayesian methodology. Their characteristics are described below and in Appendix B.

**Decision trees and tree models**

Statistical tree models are statistical models that are expressed in the form of a binary decision tree, or "tree models formed by a binary partition of the predictor space" (Denison et al., 2002, p.149). A decision tree may be applied to a data set—a collection of points with a response variable and one or more covariates—by having the questions apply to the values of the covariates of the model. Assuming ordinal covariates, each question in the tree model asks whether a data point’s value, of a specified covariate, is less than (vs. greater than or equal to) a specified value. As Denison, et al. (2002, p. 150) describe, "Each splitting node asks a yes/no question and, depending on the answer, data points are assigned to the left or right branch from this node. The terminal nodes do not have associated questions but just assign all points in them to a common probability density function."

Consequently, the collected terminal nodes (also known as leaves) of a tree model determine a partitioning of the state space (that is, a set of disjoint subsets that covers the state space) of the data set, and the binary decision tree describes a “recursive partitioning” of the
data set (or of the state space of the data set), in which each inner node of the tree represents a subset of the data, along with a question, based on one of the covariates, that splits the node’s subset into two further subsets.

The choice and arrangement of the questions is determined by the response variable. Non-Bayesian tree modeling algorithms determine each splitting question by using what is known as a greedy search algorithm. Such an algorithm searches the data set (or the relevant subset of the data set) until it finds the “best” split of a single variable. It will use one of various mathematical criteria, based upon node purity (i.e., homogeneity of the response variable within the child nodes), to determine the “best” split. An example of a non-Bayesian Regression Tree algorithm is presented by Breiman, et al. (1984). The splitting criterion for tree $T$, with leaves $t$ in the set of terminal nodes $\mathcal{T}$, is to minimize

$$R(T) = \frac{1}{N} \sum_{t \in T} \sum_{x_{n} \in t} (y_{n} - \bar{y}_t)^2$$

where the $x_n$ are the covariates and the $y_n$ are the response variables. Using this, each split is chosen to minimize the mean within node sum of squared deviations of the response variable.

Cloudiness concerns and a two stage process

Even though composite images (images composed of the pixels that contained each pixel area’s maximum NDVI value obtained over a two-week period) were used, cloudiness within pixels was found to be common. As emphasized previously, cloudiness is known to distort remotely sensed indices. Because Alaska often experiences cloudy conditions (Shulski & Wendler, 2007), one could not simply discard the cloudy pixels and hope to retain a sufficient number of clear pixels to produce a valid statistical model.

It would be desirable to “control for” the effect of clouds (i.e., to correctly attribute any effect of clouds to the clouds, and not, rather, to any other variable that happened to be
correlated with cloudiness) in the same manner as is typical in multiple regression models. In
typical regression models, controlling for the effects of nuisance variables is straightforward
(simply including the nuisance variables in the model), because regression coefficients deal with
the entire range of each covariate. In contrast, tree models partition the covariate space, so,
unless a nuisance variable determines the first split (or first splits), the effect of the nuisance
variable will not be controlled for over the entire range of the data. Therefore it was desirable to
force the nuisance variable to appear in the model at its earliest stages, so a two stage process
was designed.

The two stages proceeded as follows: a cloudiness analysis was performed first (in which
the collection year was the response variable and the cloudiness conditions were the
covariates), and the environmental analysis was performed next (in which the collection year
was the response variable and the environmental conditions were the covariates). Because tree
models partition covariate space, the initial analysis stage may be viewed as a division of the
study data into strata that were relatively homogeneous with respect to cloudiness, but one
should note that the initial analysis also took into account the association between cloudiness
and collection year.

For the cloudiness analysis, additional detail was desired (to detect cloudiness extremes
that occurred during years with otherwise "average" cloudiness). Therefore early season, late
season, and seasonal cloudiness proportions, based upon the cloudiness indicators, were used.
The cloudiness covariates were the rounded change from study period (seasonal, early and late
season) mean percent of cloudy images.

Then, a Bayesian regression tree model, using collection year as the response variable
and a state space made up of the three cloudiness covariates, found the subsets that were most
homogeneous with respect to collection year, in the space determined by the cloudiness
covariates. Each of these subsets was associated with a mean collection year ($\hat{\beta}_i$ for subset $i$).

This allowed the computation of the residuals (actual collection year - $\hat{\beta}_i$ for each subset $i$) from the cloudiness analysis. The residuals represent the information contained in the response variable that was not explained by the first stage analysis.

The cloudiness quantities (seasonal change in cloudiness, early season change in cloudiness, and late season change in cloudiness) are linearly dependent. However, the Bayesian Regression Tree algorithm does not require linearly independent covariates, and this combination of quantities was used to allow the algorithm to find the best tree most efficiently, because a single split on seasonal change, if indicated, could take the place of two splits: one on early season change and one on late season change. This is useful because the specifications to the algorithm were chosen to keep the tree models as small and as simple as possible (see the section on the use of the Bayesian Information Criterion as part of the acceptance criteria).

Out of the subsets delineated by the first stage analysis, the subset that included the study period mean cloudiness conditions was chosen to be analyzed in the second stage. This choice would leave out the subsets that contained certain cloudiness extremes: those that were found by the model to be associated with specific collection years in the study period. Because it was the subset that was left over after the extreme cloudiness values had been taken out, the subset that contained the study period mean cloudiness conditions was fairly large. Most of the years were well represented in that subset. Looking ahead to the results of the cloudiness analysis of Permafrost Area 21 (Table 4), we see that six cloudiness subsets were delineated and Cloudiness Subset 2, containing 2019 pixels and with a mean collection year of 2000.532 (according to the validation set), was the one that contained the study period mean cloudiness conditions. The other subsets' (validation set) sizes are shown here in Table 1.
Table 1. Number of observations in the subsets determined by the cloudiness analysis of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>Number of Observations in Subset (Validation Set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>1421</td>
</tr>
<tr>
<td>2</td>
<td>2019</td>
</tr>
<tr>
<td>3</td>
<td>242</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>174</td>
</tr>
</tbody>
</table>

Note that, size permitting, the other cloudiness subsets could also be analyzed in separate second stage steps. Also note that treating each subset separately in the second stage means the residuals from the first stage are simply a shift of the response variable that gives it a mean of zero within each subset.

The residuals (actual collection year - $\hat{\beta}_i$ for each subset $i$) from the chosen subset of the cloudiness analysis were used as the response variable in the second analysis step. In the second step, one finds the most homogeneous second stage subsets with respect to collection year, based upon the rounded percent change from the study period mean of the environmental indices (NDVI, simplified LST, and simplified PWI).

**Non-Bayesian tree modeling methods vs. Bayesian tree modeling methods**

Non-Bayesian tree modeling algorithms are older. A comparison between non-Bayesian methods and Bayesian methods for tree modeling reveals several advantages offered by the Bayesian methods. Classical (i.e., non-Bayesian) tree modeling algorithms partition a data set by using a variant of the “forward selection” method of model determination. This characteristic of the non-Bayesian tree modeling procedure leads to instability; i.e., the form of a tree model may be sensitive to the details of the data set used to build it (Breiman et al., 1984). Another issue is the prevention of overfitting (modeling sporadic characteristics of the particular data set used to derive the model) by a classically derived tree model.
Several investigators have proposed “ensemble” methods to deal with classical tree modeling’s recognized sensitivity and overfitting limitations. Bagging (Breiman, 1996), boosting (Freund & Schapire, 1996), and the random forest method (Brieman, 2001) build a collection of trees that are then combined. This combination acts as a “black box” (i.e., a machine with hidden inner workings) that functions very well as a predictor, when it is given new data points to predict. However, when using an ensemble method, one cannot easily tell which variables, values, and interactions are producing the observed results (Breiman, 1996).

In light of this, several characteristics make Bayesian Regression Tree Models excellent choices for applying statistical models to complex conditions. First, because they are tree models, they detect non-linear trends and they naturally include interactions that may be found within the data. In addition, because they are created using a recursive Bayesian process (Denison et al., 2002; Chipman et al., 1998; Denison et al., 1998), in which proposed trees may be grown and also shrunk, which thereby reduces the influence of a single early split, they are less sensitive to small changes in the data when compared with the other types of tree models. Bayesian model selection procedures also intrinsically guard against overfitting (Denison et al., 2002; Henderson et al., 2010). Finally, the fact that the Bayesian process generates a collection of tree models (i.e., the posterior distribution of models) is also an advantage because, if any instability remains, one may examine several tree models that describe the data well.

The Bayesian Regression Tree Model

The Bayesian Regression Tree Model is a hierarchical model. This terminology does not refer to the tree structure, but it is applied, more generally, to refer to the "hierarchically structured hypothesis space" (Henderson et al., 2010, p. 172) of many Bayesian models. The two levels of this hierarchy are (1) the tree structure, $\theta$, and (2) given a particular tree structure, the
distribution of the response variable within each of the partitions determined by the tree structure.

To determine the tree structure, $\theta$, the model space of tree structures, $\Theta$, is traversed by applying Reversible Jump Markov Chain Monte Carlo (Green P. J., 1995). This procedure may be summarized as follows. In each iteration of the Reversible Jump process, the algorithm will build a candidate tree, $\theta^\prime$, using a single reversible jump (the addition or removal of a question or split located at the bottom [the terminal nodes] of the tree structure), or a single move step (changing a question or split without changing its location in the tree structure), from the previous tree structure, $\theta$. Next, the distribution of the response variable within each of the partitions, given the candidate tree structure (and any hyperpriors for the distribution), is incorporated. For this, the algorithm will examine how well $\theta^\prime$ divides the data into subsets that each have an approximately (because the collection year variable is discrete) normal distribution of the response variable, using an acceptance step based upon a calculated acceptance probability. If the proposed model is accepted, it becomes the new model and is included in posterior space, which is a set of tree models; otherwise the current model is retained.

The basic specifications for this process are shown in the following tables, which summarize the treatment of tree models in Denison et al. (2002). The tree structure determines the partitions of the data set. The tree structure is a function of the locations of the splitting nodes, along with the covariates and values of the splits at the splitting nodes. In the first stage of the analysis, the covariates were the difference from study period mean of the percent of seasonal cloudy images, early season cloudy images, and late season cloudy images. In the second stage of the analysis, the covariates were percent difference from study period mean NDVI, percent difference from study period mean (simplified) LST, and percent difference from study period mean (simplified) PWI.
Consider a tree structure, $\theta$, and suppose that it has divided the covariate space into $k$ subsets (i.e., partitions of the data space that are determined by the tree structure's leaves or terminal nodes). Then, given $\theta$, within each subset, $i, i=1, \ldots, k$, the response variable, $y_{\text{Subset } i}$ (recall that this is the collection year in this study) will be estimated by a single value, $\beta_i$. We will allow the error variance (which here is a proxy for the gradual development and recession of each subset's climate conditions) to vary among the subsets, so we also have the parameter $\sigma_{\epsilon i}$, which is the error variance for the collection year of subset $i$. So, the data is described as follows (Denison et al. 2002):

$$y_{\text{Subset } i} = \beta_i + \epsilon_i,$$

where, approximately here, because $y$ is discrete and not continuous

$$\epsilon_i \sim N(0, \sigma_i).$$
Table 2. Tree structure (θ) from Denison et al., 2002

<table>
<thead>
<tr>
<th>Level</th>
<th>Tree Structure (θ)</th>
</tr>
</thead>
</table>
| Prior         | \[ p(\theta) = \left\{ \prod_{i=1}^{k-1} p(s_i^{\text{rule}} | s_i^{\text{var}}) p(s_i^{\text{var}}) \right\} p \left( \left\{ s_i^{\text{pos}} \right\}^{k-1} \right) \]  

where \( k \) is the number of leaves, so \( k-1 \) is the number of splitting nodes, \( s_i \) is the \( i \)th splitting node, \( pos \) specifies positions of the splitting nodes within the tree, \( var \) specifies the variables used to determine the splits, and \( rule \) specifies the exact rules (i.e., the splitting points) used to make the splits.  

\[ p(\theta) = \left\{ \prod_{i=1}^{k-1} \frac{1}{N(s_i^{\text{var}})} \right\} \frac{1}{S_k K} k! 1 \]  

where \( k \) is the number of leaves, \( N(j) \) is the number of possible splitting rules for the \( j \)th covariate, \( p \) is the number of covariates that may be chosen, \( S_k \) is the number of ways of choosing the positions of the splits for a tree model with \( k \) subsets, and \( K \) is the maximum number of subsets allowed. Finally, the \( k! \) term accounts for the possible permutations of the indices of the basis functions (leaves or terminal nodes) (Denison et al., 2002, p. 157).  

| Posterior     | Determined by Reversible Jump MCMC        |
For this method, the overall mean of $y$ is subtracted from the values of $y$, which simplifies the algorithm by allowing the hyperprior $m$ to be set to zero. Residuals add to zero, so their use in the second stage analysis fits this pattern.

<table>
<thead>
<tr>
<th>Level</th>
<th>$\beta_i$</th>
<th>$\sigma_i^2$</th>
<th>Hyperpriors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>for leaf $i$, given a tree structure ($\theta$) and hyperpriors $m$, $\nu$, $a$, and $b$</td>
<td></td>
<td>$m$, $\nu$, $a$, $b$</td>
</tr>
</tbody>
</table>

Prior

$\beta_i \sim N(0, \sigma_i^2 \nu)$

$\sigma_i^2 \sim \text{Inverse Gamma}(a, b)$

$p(\beta_i, \sigma_i^2|m, \nu, a, b) \sim \text{Normal-Inverse Gamma}(0, \nu, a, b)$

Table 3. Subsets given tree structure (formulas from Denison et al., 2002)

<table>
<thead>
<tr>
<th>Hyperpriors</th>
<th>$m$</th>
<th>$\nu$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>100</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

| Prior | $m_i^* = \left( \frac{\nu}{1 + \nu n_i} \right) \sum_{j=1}^{n_i} y_{ij}$ | $v_i^* = \frac{\nu}{1 + \nu n_i}$ | $a_i^* = a + \frac{n_i}{2}$ | $b_i^* = b + \frac{1}{2} \left[ \sum_{j=1}^{n_i} (y_{ij}^2 - \left( \sum_{j=1}^{n_i} y_{ij} \right)^2 \left( \frac{\nu}{1 + \nu n_i} \right) \right]$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$= b + \frac{1}{2} \left[ \sum_{j=1}^{n_i} y_{ij}^2 - \left( \frac{\nu n_i^2}{1 + \nu n_i} \right) \bar{y}^2 \right]$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Likelihood of the data given the tree structure**

To use these distributions to perform the Bayesian Regression Tree analysis, one more quantity is necessary: the likelihood of the data given the tree structure, $\theta$. This will be used in the acceptance step. In this case, it is (Denison et al., 2002, p. 156; also see Chipman et al., 1998)

$$p(D|\theta) = \prod_{i=1}^{k} \left( \frac{b^a \Gamma(a_i^*)}{(2\pi)^{\frac{d}{2}} (1 + \nu n_i)^{\frac{1}{2}}} \right) b_i^{*a_i^*}.$$

**Process details and a new move**

The reversible jumps and moves used were the Split of a leaf, which is the “birth” step; the Consolidation of a pair of leaves, which is the “death” step; a Change of Split Value of an inner node; a Change of Split Variable, and value, of an inner node; and a new move: the replacement of a subtree with another subtree with the same structure. This last move has not, to this author’s knowledge, been implemented in this way before.

The desirability of a single step that replaces an entire "branch" or subtree of the tree structure has been expressed in the literature (Denison et al., 2002; Chipman et al., 1998). A step of this kind would be desirable because the generally recognized reversible jumps for tree models (and thus, the degree of mixing attained by the reversible jump algorithm) are restricted by the tree structure (Denison et al., 2002).

To envision the idea of branch replacement, the reader may look ahead to Figure 3. An example of a branch or subtree would be the question above Leaf 2, along with all the questions (and leaves) below that question. One could generally envision that any tree structure could replace that subtree. A branch replacement step has been an elusive goal because it would not, in general, be a reversible step. My solution is to keep the proposed subtree (or branch)
arrangement identical to that of the current subtree, thus only changing the split variables and values, as in multiple move steps. This is therefore a reversible step.

**Training and validation subsets**

Two mutually exclusive random samples of 300 pixels per year, for each permafrost area, were obtained in a manner that ensured that any sampled pixel is found only once, *i.e.*, for only one year, in the sample. One of the samples was used as the training data set and one was used as the validation data set.

The training data set was used to derive the tree models and the validation data set was used to evaluate the tree models. Each tree model was evaluated by assigning each pixel in the validation set to the appropriate subset according to the rules expressed by the tree model. Then the distributions of collection years, for each subset, were examined and compared with those of the training data set. If the patterns were similar, the model was finding patterns that may be found consistently in the permafrost area. This is a well-recognized method for objectively evaluating statistical models (Breiman et al., 1984).

The validation set was also used to present the final results. The sample size of 300 pixels per year was somewhat arbitrary, given the exploratory nature of this project, but it appeared to be sufficient for distinguishing subsets that could be validated by the validation data set.
Results

Cloudiness analysis

For the cloudiness analysis, the covariates were the rounded change from study period (seasonal, early and late season) mean percent of cloudy images. Then, a Bayesian regression tree model, using collection year as the response variable and a state space made up of the three cloudiness proportion variables, found the subsets that were most homogeneous with respect to collection year based upon the cloudiness covariates. Each of these subsets was associated with a mean collection year ($\hat{\beta}_i$ for subset $i$).

The tree produced by the cloudiness analysis of Permafrost Area 21 is shown in Figure 3. Six subsets were identified.

The tree diagram conveys a large amount of information. The splits of the tree structure are the lower and/or upper bounds of the leaves' (tree subsets') dimensions in covariate space. However, I found that I was most interested in each subset's differences, if any, from the study period mean. This led to my being interested in discerning, for each subset, which, if any, of the covariates did not include 0 within the bounds determined, by the tree model, for the subset. One way to summarize this information is shown in Table 4. For each subset and each covariate, the subset boundary closest to 0 is illustrated. A positive value of a subset boundary indicates that the subset is located in the positive direction from zero, a negative value or 0 indicates that the subset is located in the negative direction from zero, and a dash indicates that the zero value is contained within the boundaries of the subset.

Additional information may be obtained when one examines data (in this case, an independently sampled, separate, validation set) that is partitioned according to the tree structure. Therefore, for each subset, the mean collection year ($\hat{\beta}_i$ for subset $i$), modal collection year, and number of validation set pixels contained within its boundaries, were included in Table...
4 to describe the distribution of the subset's collection years. To show the trajectory of the changes through time, Table 5 shows the subsets arranged in the order of their mean collection year.

Subset 2 is the largest subset (according to number of pixels). Its three dashes in Table 5 mean that it contains the origin of the cloudiness state space. The next-largest subset, Subset 1, exhibited a decrease from mean late season cloudiness.

Table 6 shows the distributions of the (validation set) data points by subset and year. It shows that the tree model succeeded in delineating subsets with different collection year distributions. Subsets 2 (contains the origin) and 1 (cloudiness decrease) have similar mean collection years but differing distributions. Subset 5 indicates that an increase in cloudiness occurred in 2003.

The spatial component of the cloudiness analysis is illustrated in Figure 4. The shading of the points was determined by the relative mean proportion of seasonal cloudy images in each subset. Black shading indicates the highest mean proportion of seasonal cloudy images, white indicates the lowest mean proportion of seasonal cloudy images. Light cloudiness conditions were prevalent in 1995 and 2004, with geographic patterns that were slightly different from one another. Moderate cloudiness conditions prevailed, fairly uniformly geographically, between 1996 and 2002. Heavier cloudiness conditions covered the region uniformly in 2003. Finally, an approximately even mixture of light and moderate cloudiness conditions was observed between 2005 and 2007, but the geographic arrangement of the conditions varied among those years.
Figure 3. Tree model produced by the cloudiness analysis of Permafrost Area 21.
Table 4. Subset boundaries and collection year summary for the validation set of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-40</td>
<td>27</td>
<td>2001.778</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-8</td>
<td>1421</td>
<td>2001.311</td>
<td>1995</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2019</td>
<td>2000.532</td>
<td>1999</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>-8</td>
<td>38</td>
<td>17</td>
<td>2002</td>
<td>2003</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>-</td>
<td>38</td>
<td>174</td>
<td>2002.408</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 5. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2019</td>
<td>2000.532</td>
<td>1999</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-8</td>
<td>1421</td>
<td>2001.311</td>
<td>1995</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-40</td>
<td>27</td>
<td>2001.778</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>-8</td>
<td>38</td>
<td>17</td>
<td>2002</td>
<td>2003</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>-</td>
<td>38</td>
<td>174</td>
<td>2002.408</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 6. Number of observations within subsets by collection year for the validation set of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7 0 0 2 0 0 0 0 0 10 4 1 3</td>
</tr>
<tr>
<td>1</td>
<td>230 66 110 63 39 56 108 91 18 216 143 118 163</td>
</tr>
<tr>
<td>2</td>
<td>55 218 170 223 229 203 177 185 117 74 128 108 132</td>
</tr>
<tr>
<td>3</td>
<td>6 13 15 6 19 35 13 21 32 0 25 55 2</td>
</tr>
<tr>
<td>4</td>
<td>0 1 0 1 2 0 0 0 12 0 0 1 0</td>
</tr>
<tr>
<td>5</td>
<td>2 2 5 5 11 6 2 3 121 0 0 17 0</td>
</tr>
</tbody>
</table>
Figure 4 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 21
Second stage analysis of Cloudiness Analysis Subset 2

Subset 2, from the first stage analysis, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. This choice would leave out the subsets that contained the cloudiness extremes that were found, in the first stage, to be associated with specific collection years in the study period. The residuals of subset 2 (actual collection year - \( \tilde{\beta}_2 \) for subset 2) from the cloudiness analysis constituted the response variable in the second analysis step. The second step would find the most homogeneous second stage subsets with respect to collection year, based upon the rounded percent change from the study period mean of the environmental indices (NDVI, simplified LST, and simplified PWI).

The tree produced by the second stage analysis of Cloudiness Analysis Subset 2 of Permafrost Area 21 is shown in Figure 5. Fifteen subsets were found.

Subset boundaries, which, as described in the previous section for Table 5, illustrate each subset's covariate's deviation from no change from study period mean, are shown in Table 7. To highlight the trajectory of the changes through time, Table 8 shows the subsets sorted by their study period mean collection year.

The subset containing the point of no change from all three study period mean values, Subset 4, had a mean collection year of approximately 1999, relatively early in the study period. Warmer than study period mean conditions were observed in the later portion of the study period. Higher than study period mean PWI conditions were observed in the early and middle portions of the study period, but lower than study period mean PWI conditions were observed later in the study period. Higher than study period mean NDVI was observed in two early subsets, while lower than study period mean NDVI was observed in the middle portion of the study period. Subsets 0, 1, and 9 had few observations in the validation set, and therefore illustrate the detection of less common conditions.
Table 9 shows the distributions of the data points by subset and year. This illustrates that the tree model succeeded in finding subsets of conditions that were associated with different time periods.

The geographic and temporal progression of conditions through Permafrost Area 21 is shown in Figure 6. The coloration of the points (from the validation set) uses an "RGB" (red-green-blue) scale. Each subset's color was determined by a combination of quantities representing its set of environmental conditions, scaled relative to those of the other subsets. Specifically, the quantities are each subset's relative mean percent difference from study period mean land surface temperature (red), NDVI (green), and precipitable water index (blue). Changes are visible throughout the study period, but an abrupt shift in conditions is apparent between 2003 and 2004, when wetter subsets were replaced by warmer and/or drier subsets.

Finally, the movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates. To indicate the subsets, the same colors used in Figure 6 are used in these figures. Figure 7 shows the data projected onto the NDVI (vegetation, V) by PWI (precipitable water, P) plane, Figure 8 shows the data projected onto the PWI (precipitable water, P) by LST (land surface temperature, T) plane, and Figure 9 shows the data projected onto the LST (land surface temperature, T) by NDVI (vegetation, V) plane. These plots show that the subsets are partitions of the state space, and that the subsets are indicating regions of state space that were visited at different times.
Area 21, Cloudiness Analysis Subset 2

Figure 5. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 2 of Permafrost Area 21
Table 7. Subset boundaries and collection year summary for the validation set of Cloudiness Analysis Subset 2 of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-31</td>
<td>16</td>
<td>2000.438</td>
<td>2001</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-55</td>
<td>-</td>
<td>17</td>
<td>2005</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-23</td>
<td>-</td>
<td>181</td>
<td>2003.221</td>
<td>2005</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-10</td>
<td>125</td>
<td>2001.816</td>
<td>2001</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>372</td>
<td>1999.059</td>
<td>1996</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>2</td>
<td>-6</td>
<td>284</td>
<td>2000.722</td>
<td>2002</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>21</td>
<td>-</td>
<td>138</td>
<td>1999.478</td>
<td>2003</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>74</td>
<td>-</td>
<td>2</td>
<td>2000.5</td>
<td>1998</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>93</td>
<td>1999.301</td>
<td>1999</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>-37</td>
<td>-</td>
<td>48</td>
<td>2005.188</td>
<td>2005</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>144</td>
<td>2001.542</td>
<td>2007</td>
</tr>
</tbody>
</table>
Table 8. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Cloudiness Analysis Subset 2 of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>372</td>
<td>1999.059</td>
<td>1996</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>93</td>
<td>1999.301</td>
<td>1999</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>21</td>
<td>-</td>
<td>138</td>
<td>1999.478</td>
<td>2003</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-31</td>
<td>16</td>
<td>2000.438</td>
<td>2001</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>74</td>
<td>-</td>
<td>2</td>
<td>2000.5</td>
<td>1998</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>2</td>
<td>-6</td>
<td>284</td>
<td>2000.722</td>
<td>2002</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>144</td>
<td>2001.542</td>
<td>2007</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-10</td>
<td>125</td>
<td>2001.816</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-23</td>
<td>-</td>
<td>181</td>
<td>2003.221</td>
<td>2005</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-55</td>
<td>-</td>
<td>17</td>
<td>2005</td>
<td>2005</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>-37</td>
<td>-</td>
<td>48</td>
<td>2005.188</td>
<td>2005</td>
</tr>
</tbody>
</table>

Table 9. Number of observations within subsets by collection year, for the Validation Set of Cloudiness Subset 2 of Permafrost Area 21

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3 2 0 0 0 0 0 5 2 0 1 1 2 0</td>
</tr>
<tr>
<td>1</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td>1 14 5 1 1 26 9 0 0 7 77 40 0</td>
</tr>
<tr>
<td>3</td>
<td>0 5 2 0 0 17 59 14 0 0 2 26 0</td>
</tr>
<tr>
<td>4</td>
<td>7 106 48 11 36 96 16 3 0 6 8 20 15</td>
</tr>
<tr>
<td>5</td>
<td>11 17 17 1 2 32 66 113 20 0 0 4 1</td>
</tr>
<tr>
<td>6</td>
<td>10 42 35 33 28 23 11 14 8 1 0 1 1</td>
</tr>
<tr>
<td>7</td>
<td>12 9 18 23 15 4 6 20 31 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td>1 4 1 4 0 0 0 1 26 0 0 0 0</td>
</tr>
<tr>
<td>9</td>
<td>0 0 0 0 1 0 0 0 1 0 0 0 0 0</td>
</tr>
<tr>
<td>10</td>
<td>0 7 8 20 36 1 0 0 21 0 0 0 0</td>
</tr>
<tr>
<td>11</td>
<td>5 7 25 92 92 0 0 1 6 1 0 1 1</td>
</tr>
<tr>
<td>12</td>
<td>0 0 1 0 0 0 0 0 0 14 16 3 14</td>
</tr>
<tr>
<td>13</td>
<td>0 3 2 0 1 2 2 0 0 39 7 8 59</td>
</tr>
<tr>
<td>14</td>
<td>5 2 8 37 18 2 3 17 3 5 0 3 41</td>
</tr>
</tbody>
</table>
Figure 6. Sequence of maps of Area 21 showing the geographic progression of environmental conditions through the study period (using the Validation Set of Cloudiness Subset 2)
Figure 7. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the NDVI (V) by PWI (P) plane, by collection year, with colors indicating the subsets.
Figure 8. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the PWI (P) by LST (T) plane, by collection year, with colors indicating the subsets.

<table>
<thead>
<tr>
<th>Subset</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>blue</td>
</tr>
<tr>
<td>1</td>
<td>yellow</td>
</tr>
<tr>
<td>2</td>
<td>green</td>
</tr>
<tr>
<td>3</td>
<td>light green</td>
</tr>
<tr>
<td>4</td>
<td>light yellow</td>
</tr>
<tr>
<td>5</td>
<td>light blue</td>
</tr>
<tr>
<td>6</td>
<td>light red</td>
</tr>
<tr>
<td>7</td>
<td>light purple</td>
</tr>
<tr>
<td>8</td>
<td>cyan</td>
</tr>
<tr>
<td>9</td>
<td>light cyan</td>
</tr>
<tr>
<td>10</td>
<td>dark green</td>
</tr>
<tr>
<td>11</td>
<td>dark yellow</td>
</tr>
<tr>
<td>12</td>
<td>dark blue</td>
</tr>
<tr>
<td>13</td>
<td>dark red</td>
</tr>
<tr>
<td>14</td>
<td>dark purple</td>
</tr>
</tbody>
</table>
Figure 9. State space of the validation data set of Cloudiness Subset 2 of Permafrost Area 21, projected onto the LST (T) by NDVI (V) plane, by collection year, with colors indicating the subsets.
Discussion

This analysis method summarizes climate and vegetation change—supplemental information relevant to many studies in ecology. Raw climate data contains noise that is difficult to separate from any signal it may contain. For example, McShane and Wyner derived a model that "performs better when using highly autocorrelated noise [as input] rather than proxies to 'predict' temperature" (McShane & Wyner, 2010). In contrast, Bayesian Regression Tree Models find subsets of climate and vegetation conditions that are associated with particular segments of time, which can document the changing climate that the earth is currently experiencing and the concomitant changes in NDVI.

This analysis found recent warming in the most northern area of Alaska (Permafrost Area 21). This demonstrates the degree to which the neighboring terrestrial area was influenced by the observed recent warming of the Arctic Ocean (Overland et al., 2008).

One subset indicates the opposite of expectations: low NDVI occurring with high PWI in Area 21 (subset 5 of Cloudiness Subset 2). Subset 5 could be showing a lag of an NDVI increase in response to an increase in PWI, or perhaps it indicates another influence upon NDVI. Observations such as these would be useful in the design of future Bayesian statistical models for tundra that use NDVI as the response variable.

Ground-based evidence supports the authenticity of the remotely sensed conditions. Statewide drought conditions in 2004 were reported in the book, *The Climate of Alaska* (Shulski & Wendler, 2007), and support the findings of dryness and low levels of cloudiness in 2004. The Arctic Sea has experienced warming and melting that was particularly marked in 2007 (Overland et al., 2008). This analysis shows warming in 2007 in Area 21, which borders the Arctic Ocean.
Additional support for the findings of these models may be obtained by comparing the results of the cloudiness model with those of the second stage model. For example, the initial (cloudiness) analysis showed that greater than study period mean cloudiness was observed in 2003, according to Cloudiness Subset 5 of Permafrost Area 21. From the second stage analysis (the environmental analysis of Cloudiness Subset 2), Subsets 2, 7, and 8 showed distributional peaks in the 2003 collection year. These subsets corresponded to higher than study period mean PWI and are thus likely to be associated with the cloudy conditions also observed at that time.
Illustrations and discussion of the results of the analyses of the remaining Permafrost Areas

The second stage analysis results for the remaining Permafrost Areas are briefly illustrated and discussed in this section. Short descriptions of the areas will be presented, then the maps showing the geographic and temporal progression of the sets of conditions follow. The maps are colored using an RGB (red = temperature, green = vegetation, blue = precipitable water) color scale. Greater detail about the sets of conditions that were derived may be found in the Supplemental Information. The model validation process and a sensitivity analysis are also presented in this section.

Descriptions and illustrations of the remaining Permafrost Areas

Permafrost Area 11 is Alaska's mountainous area underlain by continuous permafrost (Ferrians, 1998). It is associated with the Brooks Range geographic region (National Geographic Society, 2002), the Brooks / British Range Tundra ecoregion (Ricketts, et al., 1999), and it creates the border between the Arctic and Interior climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 10.

Permafrost Area 12 is Alaska's mountainous area underlain by discontinuous permafrost (Ferrians, 1998). It is associated with the Alaska Range and Interior Highlands geographic regions (National Geographic Society, 2002), the Alaska / St. Elias Range Tundra, Interior Yukon / Alaska Alpine Tundra, and Ogilvie / MacKenzie Alpine Tundra ecoregions (Ricketts, et al., 1999), and it borders the Interior and Copper River Basin climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 11.

Permafrost Area 13 is Alaska's mountainous area underlain by isolated masses of permafrost (Ferrians, 1998). It is associated with the Chugach Range and the southwestern
portion of Alaska Range geographic regions (National Geographic Society, 2002), the Alaska / St. Elias Range Tundra ecoregion (Ricketts, et al., 1999), and it borders the Copper River Basin and Cook Inlet climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 12.

Permafrost Area 22 is Alaska’s lowland and upland area underlain by moderately thick to thin permafrost (Ferrians, 1998). It is associated with the Seward Peninsula and Yukon River Delta geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, Interior Alaska / Yukon Lowland Taiga, and Copper Plateau Taiga ecoregions (Ricketts, et al., 1999); and the West Central and Interior climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 13.

Permafrost Area 23 is Alaska’s lowland and upland area underlain by discontinuous permafrost (Ferrians, 1998). It is associated with the Yukon Flats geographic region (National Geographic Society, 2002); the Interior Alaska / Yukon Lowland Taiga ecoregion (Ricketts, et al., 1999); and the Interior climate region (Shulski & Wendler, 2007). Its map sequence is shown in Figure 14.

Permafrost Area 24 is Alaska’s lowland and upland area underlain by numerous isolated masses of permafrost (Ferrians, 1998). It is associated with the Tanana Valley geographic region (National Geographic Society, 2002); the Interior Alaska / Yukon Lowland Taiga and Copper Plateau Taiga ecoregions (Ricketts, et al., 1999); and the Interior and West Central climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 15.

Permafrost Area 25 is Alaska’s lowland and upland area underlain by isolated masses of permafrost (Ferrians, 1998). It is associated with the Bristol Bay and Cook Inlet geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, and Cook Inlet Taiga ecoregions (Ricketts, et al., 1999); and the Bristol Bay and Cook Inlet climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 16.
Permafrost Area 26 is Alaska’s lowland and upland area generally free of permafrost (Ferrians, 1998). It, like Permafrost Area 25, is associated with the Bristol Bay and Cook Inlet geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, and Cook Inlet Taiga ecoregions (Ricketts, et al., 1999); and the Bristol Bay and Cook Inlet climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 17.

Permafrost Area 31 was numbered but not named in the permafrost data set (Ferrians, 1998). With reference to its location, it could be named the Gulf of Alaska Area. It is associated with the Coastal Range, Aleutian Range, Kodiak Island, and Kenai Peninsula geographic regions (National Geographic Society, 2002); the Alaska Peninsula Montane Taiga, Northern Pacific Coastal Forests, and Pacific Coast Mountain Tundra and Ice Fields ecoregions (Ricketts, et al., 1999); and the South Central and Southeast Panhandle climate regions (Shulski & Wendler, 2007). Its map sequence is shown in Figure 18.
Figure 10. Sequence of maps of Permafrost Area 11 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 2)
Figure 11. Sequence of maps of Permafrost Area 12 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 12. Sequence of maps of Permafrost Area 13 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 13. Sequence of maps of Permafrost Area 22 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 14. Sequence of maps of Permafrost Area 23 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 0)
Figure 15. Sequence of maps of Permafrost Area 24 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 16. Sequence of maps of Permafrost Area 25 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 17. Sequence of maps of Permafrost Area 26 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Area 31, Cloudiness Subset 3

Figure 18. Sequence of maps of Permafrost Area 31 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 3)
Discussion of the remaining Permafrost Areas

The following discussion is based upon the analyses that are summarized in the previous subsection of this report. These analyses are presented in more detail in the Supplemental Information.

For every Permafrost Area in Alaska, this analysis method highlighted conditions of lower than study period mean precipitable water that occurred in 2004 and later in the study period. Every Lowland and Upland Permafrost Area experienced higher than study period mean precipitable water during 2003. The Mountainous Permafrost Areas 11 and 13 also experienced higher than study period mean precipitable water in 2003, while Permafrost Area 12 experienced greater cloudiness during 2003. For some of the Permafrost Areas, the higher values of PWI were unique to 2003, compared to the rest of the study period. Note that a strong El Niño occurred in 2003.

The mountainous Permafrost Areas, Areas 11, 12, and 13, along with Area 31 (which is also mountainous, but it is not named as such) and Area 21, exhibit groups of pixels with lower than study period mean NDVI and PWI in 2001. This is an example of conditions that were found in more than one Permafrost Area but not found in the entire state. The similarity of these areas' conditions in 2001 may have been due to their similar climates and ecoregions, which were either mountainous tundra or arctic tundra.

For each Permafrost Area, this analysis method delineated a subset that included the origin of the covariate state space. With two exceptions, the distributions of the response variable (collection year) of these subsets showed oscillations during the early and middle portions of the study period, followed by lower frequencies during the latter portion of the study period. The exceptions were Permafrost Area 13, with an highly multimodal distribution, and 31, with a periodic distribution.
The trends of the temperatures of the Arctic Ocean differed from those of the Bering Sea in 2007. As mentioned in the introduction, the Arctic Ocean has experienced warming and melting that was particularly marked in 2007 (Overland et al., 2008). This analysis shows warming in 2007 in areas neighboring the Arctic Ocean: Areas 21, 11, and 23. However, the Bering Sea experienced warming between 2000 and 2005, but it started to cool in 2006 – 2007 (Overland, 2009; National Oceanic and Atmospheric Administration). These trends are seen for the land areas close to this sea in this report as well.

**Support for the findings of these models from other sources**

It is appropriate to question what is actually being found by these models. Could we be finding relics of the changes made in the actual sensors? The satellites, and therefore, their sensors, were replaced several times during the study period—could the models be reporting the effects of the changes?

Ground-based evidence supports the authenticity of the remotely sensed conditions. Monthly precipitation data is available for Fairbanks, Alaska. Fairbanks is located in the interior of Alaska, within or close to several of the permafrost areas: 12, 23, and 24. In the summer of 2003, wetter than usual conditions were recorded (0.61, 5.96, and 1.89 inches in June, July, and August, respectively), while in the summer of 2004, drier than usual conditions were recorded (0.31, 1.13, and 0.37 inches in June, July, and August) (FAIRBANKS INTL ARPT (PAFA) Monthly Totals/Averages Precipitation (inches)). This information supports this study’s observed contrast between the conditions in the summer of 2003 and those in the summer of 2004. Statewide drought conditions in 2004 were reported in the book, *The Climate of Alaska* (Shulski & Wendler, 2007), and support the findings of dryness and low levels of cloudiness in 2004.
Model validation

Model validation was accomplished by comparing the distributions of collection year, by subset, from the training data sets with those from the validation data sets. The distributions based upon the validation sets are shown in the main body of the report (for Area 21) and in the Supplemental Information (for the remaining Permafrost Areas). The distributions for the training data set of Area 21 are presented here, in Table 10. In general, the model fits appear to be very good, because the subsets' collection year distributions from the training and validation sets appear to be very similar.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>57</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>40</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>109</td>
<td>58</td>
<td>18</td>
<td>34</td>
<td>103</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>15</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>20</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td>29</td>
<td>82</td>
<td>118</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>43</td>
<td>39</td>
<td>29</td>
<td>25</td>
<td>32</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>10</td>
<td>22</td>
<td>35</td>
<td>16</td>
<td>5</td>
<td>3</td>
<td>30</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>18</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>14</td>
<td>28</td>
<td>101</td>
<td>102</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>25</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>12</td>
<td>12</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>20</td>
<td>18</td>
<td>1</td>
<td>7</td>
<td>20</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>31</td>
</tr>
</tbody>
</table>

It is notable that the distributions of the smaller (in terms of number of data points contained) subsets, however, often more markedly differ, relative to their size, between the training and validation sets. This may indicate that these small subsets may be the parts of the models that are most prone to overfitting.
The sensitivity analysis

Setting the hyperprior parameters to \(a=b=1.0\) in the second stage analysis resulted in an appreciably increased number of subsets, compared to \(a=b=0.001\), in most of the permafrost areas studied, as shown in Table 11. Most of the tree models have more subsets when \(a=b=1\). Note that more subsets per tree can make the analysis unwieldy. The exception to this general observation is the comparison for Area 13 (cloudiness subset 1). This may reflect that presence of a fairly unusual split (based upon extremely low NDVI values) found in that area when \(a=b=0.001\).

<table>
<thead>
<tr>
<th>Analysis Hyperpriors</th>
<th>Permafrost Area-Cloudiness Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11-2</td>
</tr>
<tr>
<td>(a=b=1)</td>
<td>19</td>
</tr>
<tr>
<td>(a=b=0.001)</td>
<td>11</td>
</tr>
</tbody>
</table>

We observed more and smaller subsets for the \(a=b=1\) case. This seems to reflect a tendency towards more splitting of subsets when \(a=b=1\), because, in particular, we observed more very small subsets when \(a=b=1\). As we have mentioned, the smaller subsets may be prone to overfitting.

A possible reason for the additional splitting that was observed when the hyperprior values were increased to \(a=b=1\) may be found by referring to the likelihood equation. Specifically, the observed increase in the number of subsets in the models might be due to the influence of the value of the \(b\) term upon the \(b_i^*\) term found in the posterior and likelihood expressions. The data's contribution to likelihood is found in the \(b_i^*\) terms. The formula for \(b_i^*\) shows that it has a lower limit at the value of \(b\). If \(b_i^*s^i\)'s contributions to the likelihood are limited by a large (relative to the incremental contribution of the data) value of \(b\), the contributions of the \(n_i\) terms and the number of subsets may become important influences upon the likelihood. Therefore, the subsets will tend to get smaller to maximize the likelihood, which was observed.
Conclusions

The technique of Bayesian Regression Tree Modeling was applied to thirteen years of remotely sensed environmental data to produce a new method for multi-year multivariable environmental change detection. This method is novel because the remotely sensed images' collection years constituted the response variable of the analysis.

Additionally, this method included an initial step to control for the effect of cloudiness in the remotely sensed images. This analysis method reduced the effects of cloudiness upon the results, by stratification, while also revealing the temporal distribution of cloudiness conditions. Cloudy images were therefore allowed to be included in the analysis, within limits set by the analysis.

In the first stage of the analysis, the covariates were the difference from study period mean of the proportion of seasonal cloudy images, early season cloudy images, and late season cloudy images. In the second stage of the analysis, the covariates were the percent difference from study period mean NDVI, percent difference from study period mean (simplified) LST, and percent difference from study period mean (simplified) PWI.

A new way to replace "tree branches" during the modeling process was proposed and implemented. It is expected to increase the efficiency of the Bayesian tree modeling procedure.

The Bayesian Regression Tree modeling method produced subsets that, taken together, describe the temporal progression of the environmental conditions. Periods of warming and drought, which have been confirmed by local observers, were found by this analysis.

The analyses of all the Permafrost Areas in Alaska showed similar trends. For data collected during 1995 – 2002, the analyses showed local variation and subtle changes. In 2003, conditions of higher precipitable water, higher Normalized Difference Vegetation Index (NDVI), and/or greater cloudiness were highlighted. In 2004, the analyses detected a shift to lower
precipitable water and/or lower cloudiness, often accompanied or followed by lower NDVI and higher land surface temperature. In 2007, continued warming was highlighted in the Arctic and northern interior regions, in contrast with a return to earlier conditions and increased cloudiness revealed in regions near the Bering Sea. These conditions mirrored those of the areas' neighboring large bodies of water (Arctic Ocean or Bering Sea).

The models’ fits appear to be very good, because the subsets' collection year distributions from the training and validation sets appear to be very similar. A sensitivity analysis showed that higher hyperprior values for the within subset variance (a=b=1 vs. a=b=0.001) led to more and smaller subsets in the tree models.

Results of analyses such as this may provide context for modeling (e.g., ecosystem modeling) and field work (e.g., community studies and population studies). They may also be useful for designing, testing and calibrating climate or ecosystem models. These analyses may also provide a starting point for the design of statistical models that would predict NDVI values based upon climate conditions.
Appendix A

Remotely Sensed Indices, Permafrost Areas, and Data Preparation

AVHRR (Advanced Very High Resolution Radiometer) 14-day composite images of Alaska provided the raw data for this study (U.S. Geological Survey, 2007; U.S. Geological Survey, 2009). Compositing and other means of correcting and calibrating the images (radiometric calibration, atmospheric correction “for ozone, water vapor absorption, and Rayleigh scattering”, and geometric registration) have been performed to improve image quality and to allow an investigator to compare images of an area over many years (U.S. Geological Survey, 2009).

Given their derivability from AVHRR images and importance with respect to the study of climate and environmental change, the variables used were (simplified) Land Surface Temperature (Price, 1984), (simplified) Precipitable Water Index (Eck & Holben, 1994), and Normalized Difference Vegetation Index (Rouse et al., 1974). The (simplified) Land Surface Temperature and the (simplified) Precipitable Water Index provide basic climatic information while the NDVI provides biotic information. Together they provide information about basic but also varied aspects of environmental change. These remotely sensed indices were derived from AVHRR composite images as described in the following paragraphs.

NDVI stands for Normalized Difference Vegetation Index (a corrected NDVI is provided for each pixel in the AVHRR composite data set), where \( \text{NDVI} = \frac{\text{Ch}_2 - \text{Ch}_1}{\text{Ch}_2 + \text{Ch}_1} \), and \( \text{Ch}_1 \) and \( \text{Ch}_2 \) are the values from AVHRR Bands 1 and 2. LST stands for Land Surface Temperature, which is also known as the Price Index. In its simplified version, \( \text{LST} = \text{T}_4 + 3.33*(\text{T}_5 - \text{T}_4) \), where \( \text{T}_4 \) and \( \text{T}_5 \) are the brightness temperatures derived from AVHRR Bands 4 and 5. PWI stands for the Precipitable Water Index of Eck and Holben (1994), where, for example, their PWI from Gao, Mali, (in cm, or grams of water vapor per cm\(^2\) land surface area) = 1.337+(0.837*(T4-T5)), and T4
and T5 are the brightness temperatures derived from AVHRR Bands 4 and 5. A simplified version of the PWI, simply set equal to T4-T5, is used here.


The Land Surface Temperature was derived, based on both theoretical and experimental considerations, with a “split window” technique to account for the absorption of radiation by water vapor. The split window corrects for the greater amount of absorption, by water vapor, of the thermal infrared frequencies detected by Band 5 of the AVHRR sensor (wavelengths ranging from approximately 11.5 to approximately 12.5 microns) compared with those detected by Band 4 of the AVHRR sensor (wavelengths ranging from approximately 10.3 to approximately 11.3 microns). The ratio of the absorption coefficients determines the constant term in the formula. The formula used in this analysis was $T_s = [T_{10.8} + 3.33(T_{10.8} - T_{11.9})] (\degree K)$.

The complete formula, which takes the land surface emissivities (ε, "relative ability to emit radiation", according to the Encarta Dictionary: English [North America]), of the different wavelengths of radiation into account, is

$$T_s = [T_{10.8} + 3.33(T_{10.8} - T_{11.9})] \left(\frac{3.5 + \varepsilon_{10.8}}{4.5}\right) + 0.75T_{11.9}(\varepsilon_{10.8} - \varepsilon_{11.9})$$

However, the percent change from mean values that are used in this analysis makes the emissivity factor in the first term irrelevant. In the simplified version of the formula that was used here, the emissivities, at each wavelength, were assumed to be equal, so the added emissivity term is not used here. According to Price, "There is no reason to suspect significant variability" (Price, 1984, p. 7236), but the quantity "may be significantly larger in desert and
mountainous regions where exposed rocks may have significant spectral patterns" (Price, 1984, p. 7236). Using the more complex formula to measure LST in the mountainous regions of Alaska may warrant further research.

The Precipitable Water Index (Eck and Holben, 1994) was also derived using the split window technique, this time to quantify the presence of water vapor. The PWI was derived empirically in several parts of the world. In each case, it has a drawback in that the slope and intercept values in its formula depend upon local surface and climate conditions. It is hoped that a simplified version of the PWI will suffice here. Note that precipitable water is not a direct measure of rainfall, but it is a measure of the amount of water contained in the atmosphere.

**Permafrost Areas**

Due to the size and heterogeneity of the state of Alaska, ten separate analyses were performed. The analyses were separated according to a categorization of ten areas within Alaska, based upon extent of permafrost, along with elevation information, which was available (Ferrians, 1998). These are shown in Figure 19 and Table 12. Some similarities among the results of the ten analyses were observed and noted in the discussion section of the online supplement.

Separate analyses, using these categories, known as Permafrost Areas, provided some control for the effects of long-term temperature conditions. Also, dividing the analysis according to the presence of permafrost implicitly provides some control for hydrology in the analysis. Elevation is also a useful stratification quantity in climate and vegetation studies, and information about elevation is also provided, as the category descriptors of mountainous vs. lowland and upland areas, in the information about the areas.
The following table (Table 12) shows the permafrost categories. It provides the names of the permafrost areas along with the (approximately) associated geographic regions (National Geographic Society, 2002), ecoregions (Ricketts, et al., 1999), and climate regions (Shulski & Wendler, 2007). Note that Permafrost Area 26 is embedded in Permafrost Area 25, so they cover approximately the same geographical area. Therefore, they are combined in a single row of Table 12.

**Table 12. Permafrost Areas with their associated geographic regions, ecoregions, and climate regions**

<table>
<thead>
<tr>
<th>Permafrost Area</th>
<th>Area Name</th>
<th>Geographic Region(s)</th>
<th>Ecoregions</th>
<th>Climate Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permafrost Area 21</td>
<td>Lowland and Upland Area underlain by thick permafrost</td>
<td>North Slope</td>
<td>Arctic Foothills Tundra and Arctic Coastal Tundra</td>
<td>Arctic Climate Region</td>
</tr>
<tr>
<td>Permafrost Area</td>
<td>Area Name</td>
<td>Geographic Region(s)</td>
<td>Ecoregions</td>
<td>Climate Regions</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>----------------------</td>
<td>------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Permafrost Area 22</td>
<td>Lowland and Upland Area underlain by moderately thick to thin permafrost</td>
<td>Seward Peninsula and Yukon River Delta</td>
<td>Beringia Lowland Tundra, Beringia Upland Tundra, Interior Alaska / Yukon Lowland Taiga, Copper Plateau Taiga</td>
<td>West Central and Interior Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area 23</td>
<td>Lowland and Upland Area underlain by discontinuous permafrost</td>
<td>Yukon Flats</td>
<td>Interior Alaska / Yukon Lowland Taiga</td>
<td>Interior Climate Region</td>
</tr>
<tr>
<td>Permafrost Area 24</td>
<td>Lowland and Upland Area underlain by numerous isolated masses of permafrost</td>
<td>Tanana Valley</td>
<td>Interior Alaska / Yukon Lowland Taiga, Copper Plateau Taiga</td>
<td>Interior and West Central Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area 25 and Permafrost Area 26</td>
<td>25: Lowland and Upland Area underlain by isolated masses of permafrost and 26: Lowland and Upland Area generally free of permafrost</td>
<td>Bristol Bay and Cook Inlet</td>
<td>Beringia Lowland Tundra, Beringia Upland Tundra, and Cook Inlet Taiga</td>
<td>Bristol Bay and Cook Inlet Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area 31</td>
<td>(Gulf of Alaska)</td>
<td>Coastal Range, Aleutian Range, Kodiak Island, Kenai Peninsula</td>
<td>Alaska Peninsula Montane Taiga, Northern Pacific Coastal Forests, and Pacific Coast Mountain Tundra and Ice Fields</td>
<td>South Central and Southeast Panhandle Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area 11</td>
<td>Mountainous Area underlain by continuous permafrost</td>
<td>Brooks Range</td>
<td>Brooks / British Range Tundra</td>
<td>Border between the Arctic and Interior Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area 12</td>
<td>Mountainous Area underlain by discontinuous permafrost</td>
<td>Alaska Range and Interior Highlands</td>
<td>Alaska / St. Elias Range Tundra, Interior Yukon / Alaska Alpine Tundra, and Ogilvie / Mackenzie Alpine Tundra</td>
<td>Interior and Copper River Basin Climate Regions</td>
</tr>
<tr>
<td>Permafrost Area</td>
<td>Area Name</td>
<td>Geographic Region(s)</td>
<td>Ecoregions</td>
<td>Climate Regions</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Permafrost Area 13</td>
<td>Mountainous Area underlain by isolated masses of permafrost</td>
<td>Chugach Range and southwestern portion of Alaska Range</td>
<td>Alaska / St. Elias Range Tundra</td>
<td>Copper River Basin and Cook Inlet Climate Regions</td>
</tr>
</tbody>
</table>

**Data Preparation**

The images were prepared for analysis by first assigning each pixel’s location (row and column) to its permafrost area and discarding the pixels not associated with a permafrost area (using ERDAS Imagine and SAS). Next, the indices, the proportion of cloudy images, and their differences (percent and absolute) from study period mean were computed for each pixel in the image at each time point. Finally, two mutually exclusive random samples of 300 pixels per year per permafrost area were obtained, in a manner that ensured that any sampled pixel is found only once, i.e., for only one year, in the sample. To accomplish this, each pixel location was randomly assigned to a sample group. The data for each pixel in each sample group were then sampled for only one of the years in the study period.

One of the samples was used as the training data set and one was used as the validation data set. For each area, its training data set was used to derive the tree model and its validation data set was used to evaluate the tree model. This is a well-recognized method for objectively evaluating statistical models (Breiman et al., 1984). The validation set was also used to present the final results. The sample size of 300 pixels per year was somewhat arbitrary, given the exploratory nature of this project, but it appeared to be sufficient for distinguishing subsets that could be validated by the validation data set.

The sampling method used was designed to minimize the effects of autocorrelation over time and space because it ensured that no pixel was seen more than once in the data set, and it ensured that knowledge of the year of sampling of any pixel in the sample does not imply
knowledge of the year of sampling of any other pixel (close to or far from the chosen pixel).

Topographic heterogeneity exists across many of the regions (Beget et al., 2006), which might decrease spatial autocorrelation. Additional spatial heterogeneity exists because some of the permafrost areas are made up of several discontinuous areas.
Appendix B

Additional Information About the Bayesian Regression Tree Models

Reversible Jump MCMC applied to tree models

In general, for Reversible Jump MCMC, the acceptance probability for each step includes several ratios of probabilities and likelihoods. It includes proposal probabilities (the probability that the particular proposed change or jump will be made during the MCMC process), designated as \( q(\theta' | \theta) \) for the proposed step and as \( q(\theta | \theta') \) for its reversal; the prior probability for both the proposed model, \( p(\theta') \), and the current model, \( p(\theta) \); and the likelihood of the data given the model for both the proposed model, \( p(D | \theta') \), and the previous model, \( p(D | \theta) \).

The acceptance probability is presented by Denison, et al. (2002, p. 54), as,

\[
\min \left\{ 1, \frac{p(D | \theta')}{p(D | \theta)} \frac{p(\theta')}{p(\theta)} \frac{q(\theta | \theta')}{q(\theta' | \theta)} |J| \right\},
\]

"the ratio of marginal likelihoods, prior and proposal distributions together with a Jacobian term \(|J|\), which accounts for the change in scale when moving between models of potentially different dimensions." However, note that in this paper's analysis, the Jacobian term in the acceptance probability is not necessary, because this paper's analysis is engineered to remain discrete (Denison, et al. 2002). The jumps are discrete changes, and, in this analysis, the splits occur solely between integral values of the rounded covariates.

For models that are not tree models, the Bayes factor,

\[
\frac{p(D | \theta')}{p(D | \theta)},
\]

may be the only necessary part of the acceptance probability, but, for tree models, the other terms are also important, and they constitute a factor called \( R \), where
The need for the R factor follows from the geometry of the tree model structure, because, at any given point during the Reversible Jump process, only certain nodes (the leaf nodes) may be split, and only certain nodes (pairs of leaf nodes which are associated with a single splitting node, or "leaf parent") may be consolidated. As Denison et al. (2002, p, 162) state,

In the other models we have met when removing a basis function we have been able to propose any of those currently in the model. However, with a tree structure this is not possible as we only want to remove one terminal node in the DEATH step. To do this we need to find splitting nodes that have two terminal nodes as direct descendants and recombine them.

To view the terms of the acceptance probability, let the proposed step be a birth (split of a leaf node). This implies that the proposed model, \( \theta' \), is based upon the addition of a split, and the existing model, \( \theta \), may be reached again through a consolidation step. Both directions must be considered when deriving the terms of the acceptance probability.

Referring to Denison et al. (2002, pp. 162-163), and noting that a factorial term is not increased because a new leaf index is known to be \( k+1 \), one finds

\[
R = \frac{p(\theta') q(\theta | \theta')}{p(\theta) q(\theta' | \theta)}
\]

(\text{Denison et al., 2002}).

\[
R = \frac{d_{k+1}}{b_k} \frac{k}{S_k} \frac{S_k}{S_{k+1}}
\]

for a birth (split) step, where \( D_{\theta'} \) = the number of leaf parents (nodes with two terminal child nodes which could be consolidated) in the proposed tree structure, \( d_{k+1} \) = the probability of a death (or consolidation) step in the proposed tree structure, and \( b_k = \) the probability of a birth (Denison, et al., 2002). One similarly finds

\[
R = \frac{b_{k-1}}{(k-1)S_{k-1}} \frac{S_k}{d_k}
\]

for a death (consolidation) step.
Note that Catalan numbers are needed to calculate the value of \( R \). They are used to compute \( S_k \), the number of ways of choosing the positions of the splits for a tree model with \( k \) terminal nodes. For \( n=k-1 \) splitting nodes, the Catalan number is

\[
C_n = \frac{1}{n+1} \binom{2n}{n}
\]

(Flajolet & Sedgewick, 2009).

**Adjusted Likelihoods**

To prevent overgrowth of trees during the MCMC process, adjusted likelihoods resembling the BIC (Bayesian Information Criterion)-adjusted likelihood were used in place of the likelihood terms in the acceptance probabilities. The BIC for a model \( m \) with likelihood \( L_m \), number of model parameters \( p_m \), and number of points \( n \) is

\[
BIC = -2 \log L_m + p_m \log n
\]

(Schwarz, 1978). The adjusted Bayes factor was then taken to be

\[
\exp \left\{ \frac{BIC_{\theta} - BIC_{\theta'}}{2} \right\}
\]

(O'Hagan & Forster, 2004).

Including the adjustment may alternatively be considered to be equivalent to adding a prior distribution for the number of leaves to the prior distribution of the tree structure so that the number of leaves (subsets) in each tree model follows a Geometric \((1 - \frac{1}{n})\) distribution, where \( n \) is the number of points in the data set (Poland & Shachter, 1994).

**Specifications for these analyses**

A normal-inverse gamma prior, for the distribution of the response variable within each leaf, was desirable for the sake of retaining the property of conjugacy. This differs from the recommendation, made by Gelman (2006), to avoid small values of inverse gamma hyperprior parameters, and to, instead, use a uniform prior, or something similar, such as a folded, non-
central t distribution. Although I used the conjugate prior, so I could have a closed form for the marginal likelihood, I performed sensitivity analyses to compare the near-zero hyperprior parameter values of $a = 0.001$ and $b = 0.001$, which were often recommended in the past, to the larger parameter values of $a = 1$ and $b = 1$, to see if the smaller values could be influencing the results of the tree analyses. A description of the findings of the sensitivity analysis (based upon all 10 permafrost areas) may be found in the main body of this report.

The probability of choosing a move step (a change that would result in a tree of the same dimension) was 0.33. This was the case whether the last accepted tree had only two leaves or if it had more than two. If the last accepted tree had more than two leaves, a same-dimensional move could be a Change Split Value of an inner node, a Change Split Variable (and value) of an inner node, or a Replace Subtree, with equal probabilities. Also, as described in the following paragraph, if the last accepted tree had more than two leaves, the birth (split) and death (consolidate) probabilities were calculated to keep the value of $R$ approximately equal to one, in most cases.

Following (in general, but not specifically) a suggestion from Denison, et al. (2002), the program dynamically set the $d_k$ and $b_k$ proposal probabilities to calculated values for $d_k$ and $b_k^*$, in order to make $R$ nearly equal to 1 most of the time. Given that a same-dimensional move was not proposed, I let

$$
\frac{d_k}{b_k^*} = \frac{d_k}{1 - d_k} = D_k \frac{S_k}{S_{k-1}} \frac{1}{k - 1}
$$

which let the probabilities add to one and kept $R$ close to 1. For $R$ to equal exactly one, we would need to use $b_k$ instead of $b_k^* = b_{k-1}$. However, as the Denison et al. (2002) point out, $b_k$ depends upon $D_k$, which would not be known before the proposed move is made.

For the sake of efficiency, if the current tree structure consisted of only a single split at the root (the top of the binary tree structure), the root split could be changed using a single
iteration (i.e., a move that was a combination of Split and Consolidate steps). The other possible reversible jump at this stage was to split one of the root’s two child nodes (i.e., one of the two subsets of covariate space formed by the first split at the root level).

The algorithm referred to the bounds of each leaf or node to ensure that each proposed split or change would be a “logical” split, i.e., one that does not go outside the existing bounds of the affected leaves or nodes. This would prevent the waste of time and iterations in building and evaluating “illogical” candidate trees. The idea of proposing only the available splits was mentioned by Chipman et al. (2001) in their lecture notes.

In this analysis, the bounds of the subsets were taken into account whenever a birth (split) step was proposed. Also, to exclude illogical splits from the analysis, the bounds of the subsets both above and below a proposed change were taken into account whenever a replacement step was proposed. Because the same bounds existed in both the existing model, \( \theta \), and the proposed model, \( \theta' \), their influence upon the factor, \( p(s_{i}^{rate}|s_{i}^{par}) \), in the priors for \( \theta \) and \( \theta' \) would be the same, and therefore, they would not affect the acceptance probability for a step.

The C++ programming language was used because it is an object-oriented programming language. Many operations on objects (i.e., data structures) with binary tree structure have been extensively studied and described (Deitel & Deitel, 2008).

Borrowing a convention from classical tree modeling, at least five data points per leaf (subset) were required, so a candidate tree structure was not evaluated for acceptance unless it contained at least five data points in each of its leaves. Testing the tree structure for at least five data points per leaf early in the acceptance algorithm was done to increase the efficiency of the program.
Random seeds were explicitly changed for each run of the program. Each run consisted of 200,000 iterations of the algorithm. I considered the following to be evidence of convergence: the observation of five runs of the Markov Chain, starting from different random seeds, that show a leveling off of the adjusted likelihoods within each chain and that reach, approximately, the same adjusted likelihoods. In this dissertation, I am presenting the models, from the posterior space, with the minimum BIC value for each area (which follows Denison et al.’s [2002] example of choosing, as “best”, the model in the posterior with the greatest marginal likelihood).

The graphics were produced using version R2.10.0 of the R language (R Development Core Team, 2009). Map sequences and state space plot sequences were produced using the lattice package for the R language (Sarkar, 2009).
Appendix C

C++ Program

// sensitivity analysis
// a and b now 0.001
// modeling seasonal cloud changes first
// 2nd stage models ndvi, lst, pwi averages for individual leaves
// just use 3 seasonal average variables
// change seed and output file name for a new chain
// adding regression tree functions
// need to enter parameters vreg, breg, areg below
// using BIC for likelihoods

#include <iostream>
using std::cout;
using std::ios;
#include <string.h>
#include <cstdlib>
using std::rand;
using std::srand;
#include <math.h>
#include <fstream>
using std::ifstream;
using std::ofstream;

enum Nodetype { ROOT, TREEROOT, LEAF, INNER, LEAFFARENT, DEFAULT }; 
enum Jumptype { SPLITROOT, SPLITLEAF, CONSOLIDATE, SHIFT, NEWBRANCH, SPLITINNER }

// need these at the top so user can enter variables and bounds;
// number of modelvars must equal nmodelvars;
// number of initbounds must equal 2*nmodelvars;
// also enter a random seed value
// also enter precision of splits
// also enter probability of not changing dimension with jump
// this should have 2 decimal digits
// for now, enter the number of data points as a constant
// note that I should limit the number of tree leaves to no more than
// the number of data points
// enter the number of response classes as a constant
// and the valid values for the response classes
// currently up to 7 characters in model var names
// currently up to 4 characters in resp classes names
// if regression tree enter vreg, areg, breg instead of class info
// pndvsme, plstsme, ppwisme
const int minleaf=4;
const int nmodelvars=3;
const char *modelvars[3]="pndvsme", "plstsme", "ppwisme";
const double initbounds[6] = {-65, 133, -3, 6, -87, 77};
const double dSplitPrecision = 1.0;
const double dPSameDimension = 0.33;
const int ndatapts = 1343;
// note max number of leaves should be <= ndatapts
const int inleaflimit = 200;
// next two constants are for classification tree
const int nrespclasses = 13;
const char *respclasses[13] = {
"-6", "-5", "-4", "-3", "-2", "-1", "0", "1", "2", "3", "4", "5", "6"};
// const char
*respclasses[13] = {

// next three constants are for regression tree
const double vreg = 100;
const double areg = 0.001;
const double breg = 0.001;
// when areg is a fraction, need to enter loggamma of areg and .5+areg;
// as calculated in SAS using lgamma function!;
// also express areg and .5+areg as p/q;
const double lgamareg = 6.9071788854;
const double lgamaregpt5 = 0.5704038975;
const double pareg = 1.0;
const double qareg = 1000.0;
const double paregpt5 = 501.0;
// alpha is for Dirichlet prior;
const int alpha = 1;

double sumloggam(int ncount)
{
  double term1 = lgamareg;
  double term2 = -1 * static_cast<double>(ncount) * log(qareg);
  double term3 = 0.0;
  double loggam;
  for (int iterm = 1; iterm <= ncount; iterm++)
  {
    term3 = term3 + log(pareg + (static_cast<double>(iterm) * qareg) - qareg);
    // std::cout << "term3 " << term3 << "\n";
  }
  loggam = term1 + term2 + term3;
  // std::cout << "loggam " << loggam << "\n";
  return loggam;
}
// this is for loggam(N+(1/2))
double sumloggamNhalf(int N)
{
  double loggam;
  double term3 = 0.0;
  double term1 = lgamaregpt5;
  double term2 = -1 * static_cast<double>(N) * log(qareg);
  // std::cout << "term1 = " << term1 << "\n";
  // std::cout << "term2 = " << term2 << "\n";
  // std::cout << "term3 = " << term3 << "\n";
  // std::cout << "term4 = " << term4 << "\n";
  return loggam;
}
//std::cout << "term2= " << term2 << 
;
for (int iterm=1;iterm<=N;iterm++)
{
    term3=term3+log(paregpt5+(static_cast<double>(iterm)*qareg)-
    qareg);
    //std::cout << "term3= " << term3 << "\n";
}
loggam=term1+term2+term3;
//std::cout << "log N+(1/2) = " << loggam << "\n";
return loggam;
}
//for the sum of the 4 terms that only contain a, astar, b, and n
data points
//fixing so areg is not an integer;
//note that incorporating areg into astar is done in the log gamma
calculations;
double aastarbnterms(int nterm)
{
    double sumterms;
    double astar1;
    double lgastar;
    double lgareg;
    double piterm;
    astar1=(static_cast<double>(nterm)/2);
    //std::cout << "astar = " << astar << "\n";
    if (nterm % 2 == 0)
    { 
        lgastar=sumloggam(static_cast<int>(astar1));
        //std::cout << "astar integer loggamma = " << lgastar << "\n" ;
    }
    else if (nterm % 2 == 1)
    { 
        int Nastar;
        Nastar=static_cast<int>(astar1-0.5);
        lgastar=sumloggamNhalf(Nastar);
        //std::cout << "astar N+(1/2) loggamma = " << lgastar << "\n" ;
    }
    lgareg=lgamareg;
piterm=0.5*nterm*log(acos(static_cast<double>(-1)));
    //std::cout << "pi term " << piterm << "\n";
    sumterms=(areg*log(breg)) + lgastar - piterm - lgareg;
    //std::cout << "sum of 4 terms = " << sumterms << "\n";
    return sumterms;
}
//catalan work
//for k leaves, Ssubk/Ssub(k+1) = (k+1)/(4k-2)
//and Ssubk/Ssub(k-1) = (4k-6)/k
double splitratio(int kleaves)
{
    double Sratio;
    Sratio=(static_cast<double>(kleaves)+1)/((4*static_cast<double>(kleaves))-2);
//std::cout << "split ratio of " << kleaves << "leaves= " << Sratio << "\n";
  return Sratio;
}
double consolratio(int kleaves)
{
  double Sratio;
  Sratio=((4*static_cast<double>(kleaves))-6)/static_cast<double>(kleaves);
  //std::cout << "consol ratio of " << kleaves << "leaves= " << Sratio << "\n";
  return Sratio;
}
//dk and bk are somewhat different than in Denison text
//I am using odds in R term
//this makes them add to one (approximately)
//changes being made here 02/01/10
//not using bkcalc in R calculations
//should use bk actually used (=1-dk)
//also need reciprocal of R in getRd

//in dkcalc kleav and npars both refer to current, pre-consolidation, step
double dkcalc(int kleav, int npars)
{
  double dkodds;
  double dkrecip;
  double dk;

  dkodds=static_cast<double>(npars)*consolratio(kleav)/(static_cast<double>(kleav-1));
  if(dkodds != 0)
    dkrecip=(1.0/dkodds)+1.0;
    dk=1.0/dkrecip;
  else dk=0;
  //std::cout << "dk for " << kleav << "leaves and " << npars << "lf parents= " << dk << "\n";
  return dk;
}
//calculate R term
//here we need the bk*=1-dk that was actually used
//02/01/10 now using passed parameters from keep tree
//and actually used bk*
//for birth step, kleav is k after birth step, npa is npa after birth step
//carefully check these
double getRb(int kleav, int npa, int kleol, int npao)
{
  double Rterm;

  Rterm=(dkcalc(kleav,npa)*static_cast<double>(kleol)*splitratio(kleol))/
  (static_cast<double>(npa)*(1-dkcalc(kleol,npao)));
//std::cout << "R birth= " << Rterm << "\n";
return Rterm;
}
//here klenew and npanew are values after consolidation step
//and I am using (02/01/10) passed parameters from keep tree
//and should use actual bk*=1-dk
double getRd(int klenew, int npanew, int kleold, int npaold)
{
    double Rterm;
    Rterm=(consolratio(kleold)*static_cast<double>(npaold)*(1-
    dkcalk(klenew,npaold)))/
    (dkcalc(kleold,npaold)*static_cast<double>(klenew));
    //old Rterm=(dkcalc(kleold,npaold)*static_cast<double>(klenew))/
    //old (consolratio(kleold)*static_cast<double>(npaold)*bkcalc(klenew,npaold));
    //std::cout << "R consol= " << Rterm << "\n";
return Rterm;
}
//have changed this 02/01/10 to include
//keep tree number of leaves and keep tree number of leaf parents
//to use in calculations
double getR(Jumptype Jmp,int kleaf,int nlp,int kleafold,int nlpold)
{
    double Rfactor;
    if(Jmp==SPLITLEAF)
    {
        Rfactor=getRb(kleaf,nlp,kleafold,nlpold);
    }
    else if(Jmp==CONSOLIDATE)
    {
        Rfactor=getRd(kleaf,nlp,kleafold,nlpold);
    }
    else
    {
        Rfactor=1;
    }
    return Rfactor;
}
int getnewindex(int nvars)
{
    int hiIndex;
    int iChoice;
    hiIndex=nvars;
iChoice = rand() % hiIndex;
    //std::cout << "index choice " << iChoice << "\n";
    //std::cout << "Variable " << modelvars[iChoice] << "\n";
return iChoice;
}
//try truncation in this function
//may have to refine this
double getnewval(double lowBound, double highBound)
{
double drange;
int irange;
int ival;
double dval;

drange=(highBound-lowBound)/dSplitPrecision;
if (drange>1.0)
{
  irange=static_cast<int>(drange);
  ival = rand()%irange;
  dval=lowBound+(static_cast<double>(ival)*dSplitPrecision);
}
if (dval>-1*dSplitPrecision && dval<dSplitPrecision)
{
  dval=0.0;
  //std::cout << "split choice " << dval << "\n";
  return dval;
}
//see Deitels p. 437
void copy1(char * s1, const char * s2)
{
  for (int i=0; (s1[i]=s2[i]) != '\0'; i++)
    ;
}

Jumptype RootJump()
{
  int ipct;
  ipct=static_cast<int>(100*dPSameDimension);
  //std::cout << "ipct " << ipct << "\n";
  int iBinary;
  iBinary = (rand()%100 <= ipct);
  //std::cout << "iBinary " << iBinary << "\n";
  if (iBinary==0)
  {
    return SPLITROOT;
  }
  else
  {
    return SPLITLEAF;
  }
}

Jumptype TreeJump(int kl, int nlp)
{
  int ipct;
  int iBin;
  int iJump;
  ipct=static_cast<int>(100*dPSameDimension);
  //std::cout << "ipct " << ipct << "\n";
  iJump = (rand()%100 <= ipct);
  //std::cout << "iJump " << iJump << "\n";
  //keep same dimension: shift or splitinner or newbranch
  if (iJump==0)
  {
iBin=rand() % 3;
//std::cout << "iBin " << iBin << "\n";
if (iBin==0)
{
    return SHIFT;
}
else if (iBin==1)
{
    return SPLITINNER;
}
else
{
    return NEWBRANCH;
}
//add or subtract dimension: splitleaf or consolidate
else
{
    double pdk;
    int ipctdk;
    pdk=dkcalc(kl,nlp);
    ipctdk=static_cast<int>(100*pdk);
    //std::cout << "ipctdk " << ipctdk << "\n";
    iBin = (rand() % 100 <= ipctdk);
    //std::cout << "iBin " << iBin << "\n";
    if (iBin==0)
    {
        return SPLITLEAF;
    }
    else
    {
        return CONSOLIDATE;
    }
}
//for when we need the index given the variable name
int getvindex(char * vname)
{
    int hiIndex=nmodelvars-1;
    int ikeep;
    int strequal;
    for (int j=0;j<=hiIndex;j++)
    {
        strequal=strcmp(vname,modelvars[j]);
        if (strequal==0)
        {
            ikeep=j;
        }
    }
    return ikeep;
}
//for when we need the response class index given the response class name
int getcindex(char * cname)
{
    int hiIndex=nrespclasses-1;
    int ikeep;
    int strequal;
    for (int j=0; j<=hiIndex; j++)
    {
        strequal=strcmp(cname, respclasses[j]);
        if (strequal==0)
        {
            ikeep=j;
        }
    }
    return ikeep;
}

//rootsplit, leaffsplit
//or
//leaffsplit, parleafconsol, shift, newbranch, splitinner
//can use newbranch on treeroot, even if no other inner nodes

//see Deitels section 8.11
//tested in test1bayes08;
class BoundTable
{
    friend class Node;
    public:
    BoundTable()
    {
        //std::cout << "constructing BoundTable " << "\n";
        pBoundsNode=0;
        int invarindex;
        invarindex=nmodelvars-1;
        for ( int invar=0; invar<=invarindex; invar++)
        {
            for (int ibound=0; ibound<=1; ibound++)
            {
                Nodebounds[invar][ibound]=0.0;
            }
        }
    }
    ~BoundTable()
    {
        //std::cout << "destructing BoundTable " << "\n";
    }
    Node* pBoundsNode;
    double Nodebounds[nmodelvars][2];
    void SetInitBounds(Node* ptoRoot)
    {
        int readcounter=0;
        int invarindex;
        invarindex=nmodelvars-1;
        for ( int invar=0; invar<=invarindex; invar++)


```cpp
{
    for (int ibound=0; ibound<=1; ibound++)
    {
        Nodebounds[invar][ibound]=initbounds[readcounter];
        //std::cout << bounds[ibound] << " of " << modelvars[invar] << " is " << Nodebounds[invar][ibound]<< "\n";
        readcounter++;
    }
}
pBoundsNode=ptoRoot;
}
void SetUpperBound(double inbounds[nmodelvars][2], Node* ptoNode, int varindex, double splitbound)
{
    int invarindex;
    invarindex=nmodelvars-1;
    for ( int invar=0; invar<=invarindex; invar++)
    {
        for (int ibound=0; ibound<=1; ibound++)
        {
            Nodebounds[invar][ibound]=inbounds[invar][ibound];
        }
    }
    Nodebounds[varindex][1]=splitbound;
pBoundsNode=ptoNode;
}
void SetLowerBound(double inbounds[nmodelvars][2], Node* ptoNode, int varindex, double splitbound)
{
    int invarindex;
    invarindex=nmodelvars-1;
    for ( int invar=0; invar<=invarindex; invar++)
    {
        for (int ibound=0; ibound<=1; ibound++)
        {
            Nodebounds[invar][ibound]=inbounds[invar][ibound];
        }
    }
    Nodebounds[varindex][0]=splitbound;
pBoundsNode=ptoNode;
}
void CopyBounds(double inbounds[nmodelvars][2], Node* ptoNode)
{
    int invarindex;
    invarindex=nmodelvars-1;
    for ( int invar=0; invar<=invarindex; invar++)
    {
        for (int ibound=0; ibound<=1; ibound++)
        {
            Nodebounds[invar][ibound]=inbounds[invar][ibound];
        }
    }
    pBoundsNode=ptoNode;
}
```
class Node
{
friend class Tree;
friend class NodeTableRow;
friend class NodeTable;
friend class SimpleNodeList;
friend class SimpleListMember;
friend class TNodeList;
friend class TNode;
friend class PostMember;
public:
Node(Node* pInParent=0,
Nodetype n
= DEFAULT,
Node* pInLeftchild=0,
Node* pInRightchild=0,
char *pSplitvar="nosplit",
double dInSplitval=0.0,
Node* pInRoot=0)
{
//std::cout << "constructing Node " << ntIn << "\n";
iNodecumctr++;
iNodeID = iNodecumctr;
//std::cout << "Node ID is " << iNodeID << "\n";
ntThis=ntIn;
strncpy(cSplitvar,pSplitvar,8);
cSplitvar[8]='\0';
dSplitval=dInSplitval;
pLeftChild=pInLeftchild;
pRightChild=pInRightchild;
pParent=pInParent;
pRootPointer=pInRoot;
BoundTable* pBinbounds=new BoundTable;
pBbounds=pBinbounds;
}
//copy does not increment iNodecumctr or reset iNodeID
Node(bool bcopy=true,
Nodetype ntIn = DEFAULT,
Node* pInParent=0,
Node* pInLeftchild=0,
Node* pInRightchild=0,
char *pSplitvar="nosplit",
double dInSplitval=0.0,
Node* pInRoot=0)
{
//std::cout << "constructing Node " << ntIn << "\n";
iNodeID = 0;
//std::cout << "Node ID is " << iNodeID << "\n";
ntThis=ntIn;
strncpy(cSplitvar,pSplitvar,8);
cSplitvar[8]='\0';
dSplitval=dInSplitval;
pLeftChild=pInLeftchild;
pRightChild=pInRightchild;
pParent=pInParent;
pRootPointer=pInRoot;
BoundTable* pBinbounds=new BoundTable;
  pBbounds=pBinbounds;
}

~Node()
{
    //std::cout << "destructing Node " << "\n";
}

bool IsRoot()
{
    bool rb = (ntThis == ROOT && pParent == 0);
    return rb;
}

bool IsTreeRoot()
{
    bool trb = (ntThis == TREEROOT && pParent == 0);
    return trb;
}

bool IsLeaf()
{
    bool lb = (ntThis == LEAF && pParent != 0 && pLeftChild == 0 && pRightChild == 0);
    return lb;
}

bool IsInner()
{
    bool ib = (ntThis == INNER && pParent != 0 && pLeftChild != 0 && pRightChild != 0);
    return ib;
}

bool IsLeafParent()
{
    bool lpb = (ntThis == LEAFPARENT && pParent != 0 && pLeftChild != 0 && pRightChild != 0);
    return lpb;
}

bool IsLeftChild()
{
    bool IsLeft = (pParent->pLeftChild==this);
    return IsLeft;
}

Nodetype Parenttype()
{
    Nodetype partype=pParent->ntThis;
    return partype;
}

Nodetype LChildtype()
{
    Nodetype lchtype=pLeftChild->ntThis;
return lchtype;
}
Nodetype RChildtype()
{
    Nodetype rchtype=pRightChild->ntThis;
    return rchtype;
}
Nodetype getSiblingtype()
{
    if (IsLeftChild())
    {
        return (pParent->RChildtype());
    }
    else
    {
        return (pParent->LChildtype());
    }
}
double getnewvalue(int ivindex)
{
    double dlow=0;
    double dhigh=0;
    double dsplit=0;
    if (pBbounds !=0)
    {
        dlow=pBbounds->Nodebounds[ivindex][0];
        dhigh=pBbounds->Nodebounds[ivindex][1];
        dsplit=getnewval(dlow,dhigh);
    }
    return dsplit;
}
double getthisupper(int ivindex)
{
    double dhigh=0;
    dhigh=pBbounds->Nodebounds[ivindex][1];
    return dhigh;
}
double getthislower(int ivindex)
{
    double dlow=0;
    dlow=pBbounds->Nodebounds[ivindex][0];
    return dlow;
}
void leafsplit( char *pinsplvar, double dinsplval, int iSplitindex)
{
    if (IsLeaf())
    {
        strncpy(cSplitvar,pinsplvar,8);
        cSplitvar[8]=\0;
        dSplitval=dinsplval;
        //make two leaf nodes
        Node* pnewleft = new Node(this,LEAF);
pLeftChild=pnewleft;
pLeftChild->pRootPointer=pRootPointer;
pLeftChild->pBbounds->SetUpperBound(pBbounds->Nodebounds,iSplitindex,dSplitval);
  //std::cout << "Left Leaf " << pLeftChild->iNodeID << "\n";
  //std::cout << "Left Leaf Upper Bound " << pLeftChild->pBbounds->Nodebounds[iSplitindex][1] << "\n";
Node* pnewright = new Node(this, LEAF);
pRightChild = pnewright;
pRightChild->pRootPointer=pRootPointer;
pRightChild->pBbounds->SetLowerBound(pBbounds->Nodebounds,iSplitindex,dSplitval);
  //std::cout << "Right Leaf Lower Bound " << pRightChild->pBbounds->Nodebounds[iSplitindex][0] << "\n";
// former leaf is now a leaf parent
ntThis = LEAFPARENT;
// former leafs parent (if not root) is now an inner node
if (pParent->IsLeafParent()){
  pParent->ntThis = INNER;
}
if (pRootPointer->IsRoot()){
  pRootPointer->ntThis = TREEROOT;
}

// the following (newsplit) is for resplitting nodes that are not leaf nodes
// to be called when a branch is being replaced
// this does not change child pointers or create new nodes
// this just changes split variable, split value, bounds of child nodes
void newsplit(char *pinsplvar, double dinsplval, int iSplitindex) {
  strncpy(cSplitvar, pinsplvar, 8);
cSplitvar[8] = '\0';
dSplitval = dinsplval;
pLeftChild->pBbounds->SetUpperBound(pBbounds->Nodebounds, iSplitindex, dSplitval);
  //std::cout << "Left Leaf " << pLeftChild->iNodeID << "\n";
  //std::cout << "Left Leaf Upper Bound " << pLeftChild->pBbounds->Nodebounds[iSplitindex][1] << "\n";
  //std::cout << "Right Leaf Lower Bound " << pRightChild->pBbounds->Nodebounds[iSplitindex][0] << "\n";
}
void parleafconsol() {
  Node* ptemp1;
  Node* ptemp2;
  Nodetype Sibling;
  if (IsLeafParent())
Node::iNodecumctr = 0;

class TrimNode
// this class is like node class without parent or child pointers
{
    public:
    TrimNode(Node::ntIn = DEFAULT,
              char *pSplitvar="nosplit",
              double dInSplitval=0.0) 
    {
        // std::cout << "constructing TrimNode " << ntIn << "\n";
        ntThis=ntIn;
    }

    char cSplitvar[8];
    double dSplitval;
    Node* pLeftChild;
    Node* pRightChild;
    Node* pParent;
    Node::ntThis;
    static int iNodecumctr;
    int iNodeID;
    BoundTable* pBbounds;
    Node* pRootPointer;
};
strncpy(cSplitvar, pSplitvar, 8);
cSplitvar[8] = '\0';
dSplitval = dInSplitval;

~TrimNode()
{
    //std::cout << "destructing TrimNode " << "\n";
}

cchar cSplitvar[8];
double dSplitval;
Nodetype ntThis;
protected:
    int iNodeID;
};
class NodeTableRow
{
    //this keeps track of parents and children for copy work
    friend class NodeTable;
    public:
    NodeTableRow(Node* pIn=0)
    {
        pThis = pIn;
pNewThis = 0;
pOldLeftChild = 0;
pOldRightChild = 0;
pNextRow = 0;
pPrevRow = 0;
pOldParent = 0;
pNewParent = 0;
ThisID = 0;
    }

    NodeTableRow* GetNext()
    {
        return (pNextRow);
    }

    NodeTableRow* GetPrev()
    {
        return (pPrevRow);
    }

    int ThisID;

    Node* pOldLeftChild;
    Node* pOldRightChild;
    NodeTableRow* pNextRow;
    NodeTableRow* pPrevRow;
    NodeTableRow* pRowParent;
    Node* pOldParent;
    Node* pNewParent;
    Node* pThis;
    Node* pNewThis;
};

//see Deitels p.952
class NodeTable
{
    public:
    NodeTable()
    {
        pFirstRow=0;
        pLastRow=0;
    }
    bool IsEmpty()
    {
        return pFirstRow==0;
    }
    //first step: new row with children
    //input root pointer
    NodeTableRow* AddRow1(Node* pInNode)
    {
        NodeTableRow* pNewRow = new NodeTableRow(pInNode);
        pNewRow->ThisID=pInNode->iNodeID;
        //std::cout << "constructing Node Table Row " << pNewRow->ThisID << "\n";
        pNewRow->pOldLeftChild=pInNode->pLeftChild;
        pNewRow->pOldRightChild=pInNode->pRightChild;
        if (IsEmpty())
        {
            pFirstRow=pLastRow=pNewRow;
            pNewRow->pPrevRow=0;
            pNewRow->pNextRow=0;
        }
        else
        {
            pNewRow->pPrevRow=pLastRow;
            pLastRow->pNextRow=pNewRow;
            pLastRow=pNewRow;
        }
        return(pNewRow);
    }
    //second step: new row with parent
    //input child pointers, one at a time, from previous step
    NodeTableRow* AddRow2(Node* pInNode2, NodeTableRow* pInRow)
    {
        NodeTableRow* pNewRow2 = new NodeTableRow(pInNode2);
        pNewRow2->ThisID=pInNode2->iNodeID;
        //std::cout << "constructing Node Table Row " << pNewRow2->ThisID << "\n";
        pNewRow2->pOldParent=pInNode2->pParent;
        pNewRow2->pRowParent=pInRow;
        //std::cout << "checking Row Parent " << pNewRow2->pRowParent->ThisID << "\n";
        pNewRow2->pPrevRow=pLastRow;
        pLastRow->pNextRow=pNewRow2;
        pLastRow=pNewRow2;
return(pNewRow2);
}

// third step: add children to row
void AddChildren(NodeTableRow* pThisRow)
{
    if(pThisRow->pThis->ntThis != LEAF)
    {
        pThisRow->pOldLeftChild=pThisRow->pThis->pLeftChild;
        pThisRow->pOldRightChild=pThisRow->pThis->pRightChild;
        //std::cout << "adding children to Node Table Row " << pThisRow->ThisID << "\n";
    }
}

void NodeTableCleanUp()
{
    NodeTableRow* pRowTemp;
    NodeTableRow* pNextTemp;
    bool lastflag=false;
    pRowTemp=pFirstRow;
    while(pRowTemp != 0 && lastflag==false)
    {
        pNextTemp=pRowTemp->GetNext();
        //std::cout << "p Next Node Table Row = " << pNextTemp << "\n";
        if (pRowTemp==pLastRow)
        {
            lastflag=true;
        }
        //std::cout << "destructing Node Table Row" << "\n";
        delete pRowTemp;
        pRowTemp=pNextTemp;
    }
    pFirstRow=0;
    pLastRow=0;
}

NodeTableRow* pFirstRow;
NodeTableRow* pLastRow;

class TNode
{
    friend class TNodeList;
    friend class DataMat;
    public:
    TNode(Node* pinNode=0)
    {
        pThisRoot=0;
        pThisTreeRoot=0;
        if (pinNode->IsLeaf())
        {
            pThisLeaf=pinNode;
            pThisLP=0;
            pThisInner=0;
        }

        void AddChildren(NodeTableRow* pThisRow)
        {
            if(pThisRow->pThis->ntThis != LEAF)
            {
                pThisRow->pOldLeftChild=pThisRow->pThis->pLeftChild;
                pThisRow->pOldRightChild=pThisRow->pThis->pRightChild;
                //std::cout << "adding children to Node Table Row " << pThisRow->ThisID << "\n";
            }
        }
    }

    void NodeTableCleanUp()
    {
        NodeTableRow* pRowTemp;
        NodeTableRow* pNextTemp;
        bool lastflag=false;
        pRowTemp=pFirstRow;
        while(pRowTemp != 0 && lastflag==false)
        {
            pNextTemp=pRowTemp->GetNext();
            //std::cout << "p Next Node Table Row = " << pNextTemp << "\n";
            if (pRowTemp==pLastRow)
            {
                lastflag=true;
            }
            //std::cout << "destructing Node Table Row" << "\n";
            delete pRowTemp;
            pRowTemp=pNextTemp;
        }
        pFirstRow=0;
        pLastRow=0;
    }

    NodeTableRow* pFirstRow;
    NodeTableRow* pLastRow;
};
TNodeType=LEAF;
}
else if (pinNode->IsLeafParent())
{
    pThisLP=pinNode;
    pThisLeaf=0;
    pThisInner=0;
    TNodeType=LEAFPARENT;
}
else if (pinNode->IsInner())
{
    pThisInner=pinNode;
    pThisLeaf=0;
    pThisLP=0;
    TNodeType=INNER;
}
else if (pinNode->IsRoot())
{
    pThisRoot=pinNode;
    pThisInner=0;
    pThisLeaf=0;
    pThisLP=0;
    TNodeType=ROOT;
}
else if (pinNode->IsTreeRoot())
{
    //TreeRoot is also an Inner Node
    pThisTreeRoot=pinNode;
    pThisInner=pinNode;
    pThisLeaf=0;
    pThisLP=0;
    TNodeType=TREEROOT;
}
pNextLeaf=0;
pPrevLeaf=0;
pNextLP=0;
pPrevLP=0;
pNextInner=0;
pPrevInner=0;
}
TNode()
{
    //std::cout << "destructing TNode " << "\n";
}

//get next leaf in list for data work
TNode* GetNextLeaf()
{
    TNode* pLeaf;
    pLeaf=pNextLeaf;
    pLeaf=pNextLeaf;
    return pLeaf;
}

double GetUpperBound(int invarindex)
{ double thisupper; thisupper=pThisLeaf->pBbounds->Nodebounds[invarindex][1]; return thisupper; }

double GetLowerBound(int invarindex)
{
    double thislower;
    thislower=pThisLeaf->pBbounds->Nodebounds[invarindex][0];
    return thislower;
}

TNode* pNextLeaf;
TNode* pPrevLeaf;
TNode* pNextLP;
TNode* pPrevLP;
TNode* pNextInner;
TNode* pPrevInner;
Node* pThisLeaf;
Node* pThisLP;
Node* pThisInner;
Node* pThisRoot;
Node* pThisTreeRoot;
Nodetype TNodeType;
}

class TNodeList
{
    friend class Tree;
    friend class Node;
    friend class DataMat;
    public:
    TNodeList()
    {
        pFirstLeaf=pLastLeaf=0;
        pFirstLP=pLastLP=0;
        pFirstInner=pLastInner=0;
        pRootTreeRoot=0;
        iLeafCount=0;
        iLPCount=0;
        iInnerCount=0;
        RootType=DEFAULT;
    }
    ~TNodeList()
    {
        //std::cout << "destructing TNodeList " << "\n";
    }
    void CleanUpTree()
    {
        TNode* pTNtemp;
        Node* pNtemp;
        BoundTable* pBTtemp;
        pTNtemp=pFirstLeaf;
        for(int ilf=0;ilf<iLeafCount;ilf++)
{ 
   pNtemp=pTNtemp->pNextLeaf;
   pBTtemp=pTNtemp->pBbounds;
   pTNtemp=pTNtemp->pNextLeaf;
   delete pBTtemp;
   delete pNtemp;
}
pTNtemp=pFirstLP;
for(int ilp=0;ilp<iLPCount;ilp++)
{
   pNtemp=pTNtemp->pNextLP;
   pBTtemp=pNtemp->pBbounds;
   pTNtemp=pTNtemp->pNextLP;
   delete pBTtemp;
   delete pNtemp;
}
pTNtemp=pFirstInner;
for(int iin=0;iin<iInnerCount;iin++)
{
   pNtemp=pTNtemp->pNextInner;
   pBTtemp=pNtemp->pBbounds;
   pTNtemp=pTNtemp->pNextInner;
   delete pBTtemp;
   delete pNtemp;
}
if (RootType==ROOT)
{
   pNtemp=pRootTreeRoot->pThisRoot;
   delete pNtemp;
}
}
```cpp
if (RootType == ROOT) {
    pTNtemp = pRootTreeRoot;
    delete pTNtemp;
}

pFirstLeaf = pLastLeaf = 0;
pFirstLP = pLastLP = 0;
pFirstInner = pLastInner = 0;
pRootTreeRoot = 0;
RootType = DEFAULT;
}

TNode* pFirstLeaf;
TNode* pLastLeaf;
TNode* pFirstLP;
TNode* pLastLP;
TNode* pFirstInner;
TNode* pLastInner;
TNode* pRootTreeRoot;
Nodetype RootType;
int iLeafCount;
int iLPCount;
int iInnerCount;
bool IsEmptyLeaf() {
    return pFirstLeaf == 0;
}

bool IsEmptyLP() {
    return pFirstLP == 0;
}

bool IsEmptyInner() {
    return pFirstInner == 0;
}

void addtoLeaves(Node* pInLeaf) {
    if (pInLeaf->IsLeaf()) {
        TNode* pNewLeaf = new TNode(pInLeaf);
        if (IsEmptyLeaf()) {
            pFirstLeaf = pLastLeaf = pNewLeaf;
            ++iLeafCount;
        } else {
            pNewLeaf->PrevLeaf = pLastLeaf;
            pLastLeaf->NextLeaf = pNewLeaf;
```
pLastLeaf=pNewLeaf;
++iLeafCount;
}
}

void addtoLPs(Node* pInLP)
{
    if (pInLP->IsLeafParent())
    {
        TNode* pNewLP = new TNode(pInLP);
        if (IsEmptyLP())
        {
            pFirstLP=pLastLP=pNewLP;
            ++iLPCount;
        }
        else
        {
            pNewLP->pPrevLP=pLastLP;
pLastLP->pNextLP=pNewLP;
pLastLP=pNewLP;
            ++iLPCount;
        }
    }
}

void addtoInners(Node* pInInner)
{
    if (pInInner->IsInner())
    {
        TNode* pNewInner = new TNode(pInInner);
        if (IsEmptyInner())
        {
            pFirstInner=pLastInner=pNewInner;
            ++iInnerCount;
        }
        else
        {
            pNewInner->pPrevInner=pLastInner;
pLastInner->pNextInner=pNewInner;
pLastInner=pNewInner;
            ++iInnerCount;
        }
    }
}

void addtoRoot(Node* pInRoot)
{
    if (pInRoot->IsRoot())
    {
        TNode* pNewRoot=new TNode(pInRoot);
pRootTreeRoot=pNewRoot;
        RootType=ROOT;
    }
}

void addtoTreeRoot(Node* pInTreeRoot)
{
if (pInTreeRoot->IsTreeRoot())
{
    //std::cout << "will add new TreeRoot TNode" << "\n";
    TNode* pNewTreeRoot=new TNode(pInTreeRoot);
    //std::cout << "new TreeRoot TNode made" << "\n";
    pRootTreeRoot=pNewTreeRoot;
    RootType=TREEROOT;
    //treeroot is also an inner node
    //so we will also have an inner node pointer to tree root
    TNode* pNewInner = new TNode(pInTreeRoot);
    if (IsEmptyInner())
    {
        pFirstInner=pLastInner=pNewInner;
        ++iInnerCount;
    }
    else
    {
        pNewInner->pPrevInner=pLastInner;
        pLastInner->pNextInner=pNewInner;
        pLastInner=pNewInner;
        ++iInnerCount;
    }
}

void addtoList(Node* paddNode)
{
    if(paddNode->IsLeaf())
    {
        addtoLeaves(paddNode);
        //std::cout << "adding a node to Leaf list " << paddNode->iNodeID << "\n";
        //std::cout << "number of leaves" << iLeafCount << "\n";
    }
    else if(paddNode->IsLeafParent())
    {
        addtoLPs(paddNode);
        //std::cout << "adding a node to Leaf Parent list " << paddNode->iNodeID << "\n";
        //std::cout << "number of leaf parents" << iLPCount << "\n";
    }
    else if(paddNode->IsInner())
    {
        addtoInners(paddNode);
        //std::cout << "adding a node to Inner Node list " << paddNode->iNodeID << "\n";
        //std::cout << "number of inner nodes" << iInnerCount << "\n";
    }
    else if (paddNode->IsRoot())
    {
        addtoRoot(paddNode);
//std::cout << "adding a node to Root list " << paddNode->iNodeID << "\n";
}  
else if (paddNode->IsTreeRoot())
{
    addtoTreeRoot(paddNode);
    //std::cout << "adding a node to Tree Root list " << paddNode->iNodeID << "\n";
}

//choose member at random, according to type
TNode* ChooseLeaf()
{
    int LeafChoice;
    TNode* pLeaf=pFirstLeaf;
    LeafChoice = rand() % iLeafCount;
    for (int il=0; il<=LeafChoice;il++)
    {
        if (il<LeafChoice)
        {
            pLeaf=pLeaf->pNextLeaf;
        }
    }
    return pLeaf;
}

//get first leaf to start data work
TNode* GetFirstLeaf()
{
    TNode* pLeaf;
    pLeaf=pFirstLeaf;
    return pLeaf;
}

//get n leaves to start data work
int GetNLeaves()
{
    int iNL;
    iNL=iLeafCount;
    //std::cout << "number of leaves " << iNL << "\n";
    return iNL;
}

TNode* ChooseLP()
{
    int LPChoice;
    TNode* pLP=pFirstLP;
    LPChoice = rand() % iLPCount;
    for (int ilp=0; ilp<=LPChoice;ilp++)
    {
        if (ilp<LPChoice)
        {
            pLP=pLP->pNextLP;
        }
    }
}
TNode* ChooseInner()
{
    int InChoice;
    TNode* pInn=pFirstInner;
    InChoice = rand() % iInnerCount;
    for (int inn=0; inn<=InChoice;inn++)
    {
        if (inn<InChoice)
        {
            pInn=pInn->pNextInner;
        }
    }
    return pInn;
}
};
class SimpleListMember
{
    friend class SimpleNodeList;
    public:
    SimpleListMember(Node* pInputNode)
    {
        pThisTreeNode=pInputNode;
        MemberType=pThisTreeNode->ntThis;
        pNextNode=0;
        pPrevNode=0;
    }
    Node* pThisTreeNode;
    NodeType MemberType;
    SimpleListMember* pNextNode;
    SimpleListMember* pPrevNode;
};
class SimpleNodeList
{
    public:
    bool IsEmptyList()
    {
        return pFirstNode==0;
    }
    void addto(Node* pIn)
    {
        SimpleListMember* pNewMember = new SimpleListMember(pIn);
        if (IsEmptyList())
        {
            pFirstNode=pLastNode=pNewMember;
        }
        else
        {
            pNewMember->pPrevNode=pLastNode;
            pLastNode->pNextNode=pNewMember;
            pLastNode=pNewMember;
        }
```cpp
SimpleListMember* pFirstNode;
SimpleListMember* pLastNode;
};
class PairHighLow
{
friend Tree;
public:
PairHighLow(double iMaxlow=0.0, double iMinhigh=0.0)
{
    //std::cout << "constructing PairHighLow " << "\n";
    Maxlow=iMaxlow;
    Minhigh=iMinhigh;
}
double Maxlow;
double Minhigh;
};
class Tree
{
friend class TNodeList;
public:
Tree(Node* ipRoot=0, bool iNewTree=true)
{
    pRoot=ipRoot;
    NewTree=iNewTree;
    //std::cout << "constructing Tree " << "\n";
    //std::cout << "NewTree is " << NewTree << "\n";
}
~Tree()
{
    //std::cout << "destructing Tree " << "\n";
}

void rootsplit(char *pinsplvar, double dinsplval, int iSplitindex)
{
    if (pRoot != 0 && pRoot->IsRoot())
    {
        strncpy(pRoot->cSplitvar, pinsplvar, 8);
        pRoot->cSplitvar[8] = '\0';
        pRoot->dSplitval = dinsplval;
        // make two leaf nodes
        Node* pnewleft = new Node(pRoot, LEAF);
        pRoot->pLeftChild = pnewleft;
        pRoot->pLeftChild->pRootPointer = pRoot;
        pRoot->pLeftChild->pBbounds->SetUpperBound(pRoot->pBbounds->Nodebounds, pRoot->pLeftChild, iSplitindex, pRoot->dSplitval);
        //std::cout << "Left Leaf of Root " << pRoot->pLeftChild->iNodeID << "\n";
        //std::cout << "Left Leaf of Root Upper Bound " << pRoot->pLeftChild->pBbounds->Nodebounds[iSplitindex][1] << "\n";
    }
```
Node* pnewright = new Node(pRoot,LEAF);
pRoot->pRightChild=pnewright;
pRoot->pRightChild->pRootPointer=pRoot;
pRoot->pRightChild->pBbounds->SetLowerBound(pRoot->pBbounds->Nodebounds, pRoot->pRightChild, iSplitindex, pRoot->dSplitval);
    //std::cout << "Right Leaf of Root Lower Bound " << pRoot->pRightChild->pBbounds->Nodebounds[iSplitindex][0] << "\n";
}

//set up and call root split
void rootsplitsetup()
{
    if (NewTree == true)
    {
        std::cout << "first Root Split" << "\n";
        NewTree=false;
        Node* pNewRoot = new Node(false,ROOT);
pRoot=pNewRoot;
pRoot->pBbounds->SetInitBounds(pRoot);
pRoot->pRootPointer=pRoot;
    }
    else
    {
        if (pRoot != 0 && pRoot->IsRoot())
        {
            Node* ptemp1;
            Node* ptemp2;
            ptemp1=pRoot->pLeftChild;
            ptemp2=pRoot->pRightChild;
pRoot->pLeftChild=0;
pRoot->pRightChild=0;
delete ptemp1;
delete ptemp2;
            strncpy(pRoot->cSplitvar,"nosplit",8);
pRoot->cSplitvar[8]="\0";
pRoot->dSplitval=0.0;
            //std::cout << "Consolidating Leaves of Root" << pRoot->iNodeID << "\n";
        }
        int iSplitind;
        char snew[8];
iSplitind=getnewindex(nmodelvars);
copy1(snew,modelvars[iSplitind]);
double dSplitv;
dSplitv=pRoot->getnewvalue(iSplitind);
    rootsplit(snew,dSplitv,iSplitind);
    }

//set up and call leaf split
void leafsplitsetup(Node* inptr)
{
    if(inptr->IsLeaf())
    {
int iSplitind;
char snew[8];
iSplitind=getnewindex(nmodelvars);
copy1(snew,modelvars[iSplitind]);
double dSplitv;
dSplitv=inptr->getnewvalue(iSplitind);
inptr->leafsplit(snew,dSplitv,iSplitind);
}
//allocate memory off the heap
double* newdouble(void)
{
    double * pdlocal=new double;
    return pdlocal;
}
//preorder traversal to change bounds
//left side look at upper bounds
//right side look at lower bounds
void PreOrderHelperChgBounds(Node* inptr, int invarindex, double oldsplit, double newshift, bool Lside)
{
    double thisupper;
    double thislower;
    if (inptr !=0)
    {
        if (Lside)
        {
            thisupper=inptr->pBbounds->Nodebounds[invarindex][1];
            if (thisupper==oldsplit)
            {
                inptr->pBbounds->Nodebounds[invarindex][1]=newshift;
                //std::cout << "New Upper Bound Shift " << newshift << "\n";
            }
        }
        else
        {
            thislower=inptr->pBbounds->Nodebounds[invarindex][0];
            if (thislower==oldsplit)
            {
                inptr->pBbounds->Nodebounds[invarindex][0]=newshift;
                //std::cout << "New Lower Bound Shift " << newshift << "\n";
            }
        }
    }
    if (inptr->pLeftChild != 0)
    {
        PreOrderHelperChgBounds(inptr->pLeftChild, invarindex, oldsplit, newshift, Lside);
    }
    if (inptr->pRightChild !=0)
    {
        PreOrderHelperChgBounds(inptr->pRightChild, invarindex, oldsplit, newshift, Lside);
//try preorder traversal for getting shift bounds;
//want highest bound of left subtree
//want lowest bound of right subtree
//see Deitels pp. 975-979

void PreOrderHelperShift(Node* inptr, int invarindex, PairHighLow* pHL)
{
    double templow;
    double temphigh;

    if(inptr != 0)
    {
        templow=inptr->pBbounds->Nodebounds[invarindex][0];
        //std::cout << "templow " << templow << "\n";
        temphigh=inptr->pBbounds->Nodebounds[invarindex][1];
        //std::cout << "temphigh " << temphigh << "\n";
        if (templow>pHL->Maxlow)
        {
            pHL->Maxlow=templow;
            //std::cout << "Maxlow " << pHL->Maxlow << "\n";
        }
        if (temphigh<pHL->Minhigh)
        {
            pHL->Minhigh=temphigh;
            //std::cout << "Minhigh " << pHL->Minhigh << "\n";
        }
        if (inptr->pLeftChild != 0)
        {
            PreOrderHelperShift(inptr->pLeftChild,invarindex,pHL);
        }
        if (inptr->pRightChild != 0)
        {
            PreOrderHelperShift(inptr->pRightChild,invarindex,pHL);
        }
    }
}

//note this could be called PreOrderTraversal
//creates pairs of max low and min high bounds
//for shift work
void NewShift(Node* innerptr, int innvarindex)
{
    if (innerptr->IsInner() || innerptr->IsTreeRoot())
    {
        //get split bounds based on left and right subtrees
        double hilow;
        double lowhi;
        double oldsplitval;
        PairHighLow* pHLLeft;
        }
PairHighLow* pHLRight;
pHLLeft=new PairHighLow;
pHLRight=new PairHighLow;
pHLLeft->Maxlow=innerptr->pBbounds->Nodebounds[innvarindex][0];
pHLLeft->Minhigh=innerptr->pBbounds->Nodebounds[innvarindex][1];
pHLRight->Maxlow=innerptr->pBbounds->Nodebounds[innvarindex][0];
pHLRight->Minhigh=innerptr->pBbounds->Nodebounds[innvarindex][1];

//std::cout << "pHLLeft Max Low " << pHLLeft->Maxlow << "\n";
//std::cout << "pHLRight Min High " << pHLRight->Minhigh << "\n";

PreOrderHelperShift(innerptr->pLeftChild, innvarindex, pHLLeft);
hiow=pHLLeft->Maxlow;
//std::cout << "hiow " << hiow << "\n";
PreOrderHelperShift(innerptr->pRightChild, innvarindex, pHLRight);
lowhi=pHLRight->Minhigh;
//std::cout << "lowhi " << lowhi << "\n";
//save old split
oldsplitval=innerptr->dSplitval;
//get split value, as long as there is room
if (lowhi-hiow>2*dSplitPrecision)
{
    innerptr->dSplitval=getnewval(hiow, lowhi);
}
//std::cout << "shift value " << innerptr->dSplitval << "\n";
//change bounds of child nodes
//another traversal
PreOrderHelperChgBounds(innerptr->pLeftChild, innvarindex, oldsplitval, innerptr->dSplitval, true);
PreOrderHelperChgBounds(innerptr->pRightChild, innvarindex, oldsplitval, innerptr->dSplitval, false);

} //Split an Inner Node using conventions used in shift
void NewSplitInner(Node* innerptr, int innvarindex)
{ if (innerptr->IsInner() || innerptr->IsTreeRoot())
{
    double oldsplitval;
    double boundupper;
    double boundlower;
    oldsplitval=innerptr->dSplitval;
    boundupper=innerptr->getthisupper(innvarindex);
    boundlower=innerptr->getthislower(innvarindex);
    //first restore bounds of current split variable to subtrees
    PreOrderHelperChgBounds(innerptr->pLeftChild, innvarindex, oldsplitval, boundupper, true);
    PreOrderHelperChgBounds(innerptr->pRightChild, innvarindex, oldsplitval, boundlower, false);
}
// now choose a new split variable
int iSplitind;
char snew[8];
iSplitind = getnewindex(nmodelvars);
copy1(snew, modelvars[iSplitind]);
strncpy(innerptr->cSplitvar, snew, 8);
innerptr->cSplitvar[8] = '0';

// now high low values as in shift
// get split bounds based on left and right subtrees and new
split variable
double hilow;
    double lowhi;
PairHighLow* pHLLeft;
PairHighLow* pHLRight;
pHLLeft = new PairHighLow;
pHLRight = new PairHighLow;
pHLLeft->Maxlow = innerptr->pBbounds->Nodebounds[iSplitind][0];
pHLLeft->Minhigh = innerptr->pBbounds->Nodebounds[iSplitind][1];
pHLRight->Maxlow = innerptr->pBbounds->Nodebounds[iSplitind][0];
pHLRight->Minhigh = innerptr->pBbounds->Nodebounds[iSplitind][1];
PreOrderHelperShift(innerptr->pLeftChild, iSplitind, pHLLeft);
    hilow = pHLLeft->Maxlow;
PreOrderHelperShift(innerptr->pRightChild, iSplitind, pHLRight);
    lowhi = pHLRight->Minhigh;

    // use bounds to get a new split value
    if (lowhi - hilow > 2 * dSplitPrecision)
    {
        innerptr->dSplitval = getnewval(hilow, lowhi);
    }

    // finally change bounds for new split
    boundupper = innerptr->getthisupper(iSplitind);
    boundlower = innerptr->getthislower(iSplitind);
    PreOrderHelperChgBounds(innerptr->pLeftChild, iSplitind, boundupper, innerptr->dSplitval, true);
    PreOrderHelperChgBounds(innerptr->pRightChild, iSplitind, boundlower, innerptr->dSplitval, false);
}
}

// try preorder traversal for branch replacement;
// see Deitels pp. 975-979
void PreOrderHelper(Node* inptr)
{
    if (inptr->IsLeaf() != true)
    {
        int iSplitind;
        char snew[8];
        iSplitind = getnewindex(nmodelvars);
        ...
copy1(snew,modelvars[iSplitind]);
double dSplitv;
dSplitv=inp->getnewvalue(iSplitind);
inptr->newsplit(snew,dSplitv,iSplitind);
PreOrderHelper(inptr->pLeftChild);
PreOrderHelper(inptr->pRightChild);

//note this could be called PreOrderTraversal
void NewBranch(Node* innerptr)
{
  if (innerptr->IsInner() || innerptr->IsTreeRoot())
  {
    PreOrderHelper(innerptr);
  }
}

//pre order traversal to build node type lists
void NodeTypeListHelper(Node* pThisNode, TNodeList* pNTList)
{
  if (pThisNode != 0)
  {
    pNTList->addtoList(pThisNode);
    //std::cout << "added to NTList"
    NodeTypeListHelper(pThisNode->pLeftChild, pNTList);
    NodeTypeListHelper(pThisNode->pRightChild, pNTList);
  }
}

//pre order traversal to build node type lists
TNodeList* NodeTypeLists(Node* pRootptr)
{
  if (pRootptr->IsRoot() || pRootptr->IsTreeRoot())
  {
    TNodeList* pTypes=new TNodeList;
    //std::cout << "new TNodeList"
    NodeTypeListHelper(pRootptr, pTypes);
    return pTypes;
  }
  else return 0;
}

//assumes we have a work tree and have listed its nodes in a
TNodeList Jumptype NewJump(TNodeList* pWorkNodes)
{
  //choose jump type
  //if root, then either split a leaf or resplit the root
  //otherwise (if tree root) then split a leaf, change a branch,
  //consolidate 2 leaves, or shift a node
  Jumptype Jmp;
  if (pRoot->IsRoot())
  {
    //std::cout << "Root Jumps"
    RootJump RJ;
RJ=RootJump();
if (RJ==SPLITROOT)
{
    //std::cout << "Split Root " << "\n";
    rootsplitsetup();
}
else if (RJ==SPLITLEAF)
{
    //std::cout << "Split Leaf " << "\n";
    TNode* pLeafChoice;
    pLeafChoice=pWorkNodes->ChooseLeaf();
    leafsplitsetup(pLeafChoice->pThisLeaf);
}
    Jmp=RJ;
} else if (pRoot->IsTreeRoot())
{
   //std::cout << "Tree Jumps " << "\n";
    Jumptype TJ;
    TJ=TreeJump(pWorkNodes->iLeafCount,pWorkNodes->iLPCount);
    if (TJ==SPLITLEAF)
    {
        //std::cout << "Split Leaf " << "\n";
        TNode* pLeafChoice;
        if (pWorkNodes->iLeafCount>0)
        {
            pLeafChoice=pWorkNodes->ChooseLeaf();
            leafsplitsetup(pLeafChoice->pThisLeaf);
        }
    }
 else if (TJ==CONSOLIDATE)
    {
        //std::cout << "Consolidate " << "\n";
        TNode* pLPChoice;
        if (pWorkNodes->iLPCount>0)
        {
            pLPChoice=pWorkNodes->ChooseLP();
            pLPChoice->pThisLP->parleafconsol();
        }
    }
 else if (TJ==SHIFT)
    {
        //std::cout << "Shift " << "\n";
        TNode* pInnerChoice;
        if (pWorkNodes->iInnerCount>0)
        {
            pInnerChoice=pWorkNodes->ChooseInner();
            NewShift(pInnerChoice->pThisInner,getvindex(pInnerChoice->pThisInner->cSplitvar));
        }
    }
 else if (TJ==SPLITINNER) {
    TNode* pInnerChoice;
    if (pWorkNodes->iInnerCount>0)
{  
pInnerChoice=pWorkNodes->ChooseInner();  
NewSplitInner(pInnerChoice->pThisInner, getvindex(pInnerChoice->pThisInner->cSplitvar));  
}

}  
else if (TJ==NEWBRANCH)
{
    //std::cout << "New Branch " << "\n";  
    TNode* pInnerChoice;  
    if (pWorkNodes->iInnerCount>0)
    {  
        pInnerChoice=pWorkNodes->ChooseInner();  
        NewBranch(pInnerChoice->pThisInner);  
    }
}

Jmp=TJ;
}
return Jmp;

Node* pRoot;  
bool NewTree;
};

class Counts
{
    friend class DataMat;  
    public:
    Counts(int inleaves,  
           int=ndatapts, int=nrespclasses, int=inleaflimit)  
    {  
        ntreeleaves=inleaves;  
        nleaflimit=inleaflimit;  
        for(int i=0;i<nleaflimit;i++)  
        {  
            iLeafcounts[i]=0;  
            //std::cout << "Leaf counts " << i << " " << iLeafcounts[i] << "\n";  
            for(int j=0;j<nrespclasses;j++)  
            {  
                iLeafRespcounts[i][j]=0;  
                //std::cout << "Leaf Resp counts " << i << " " << j << " " << iLeafRespcounts[i][j] << "\n";
            }
        }
        std::cout << "exiting counts" << "\n";
    }
    void ResetCounts(int inleaves)
    {  
        for(int i=0;i<nleaflimit;i++)  
        {  
            iLeafcounts[i]=0;  
            //std::cout << "Leaf counts " << i << " " << iLeafcounts[i] << "\n";  
            for(int j=0;j<nrespclasses;j++)  
            {  
                iLeafRespcounts[i][j]=0;  
                //std::cout << "Leaf Resp counts " << i << " " << j << " " << iLeafRespcounts[i][j] << "\n";
            }
        }
    }
}
iLeafRespcounts[i][j]=0;
//std::cout << "Leaf Resp counts " << i << " " << j << " " << iLeafRespcounts[i][j] << "\n";
}
}
ntreeleaves=inleaves;

//can change following to test for zero or for less than some integer
//if(iLeafcounts[i]<2)
//if(iLeafcounts[i] == 0)
bool ZeroOK(int=minleaf)
{
  bool okflag=true;
  for(int i=0;i<ntreeleaves;i++)
  {
    if(iLeafcounts[i]<=minleaf)
    {
      okflag=false;
    }
  }
  return okflag;
}
double logppost()
{
  double mter;
double msum;
double nterm;
double nsum;
int nsubi;
int ngroup;
int msubij;
int mgroup;
double preterm;
double ppost;
preterm=ntreeleaves*(sumloggam(nrespclasses*alpha)-
(nrespclasses*sumloggam(alpha)));
//std::cout << "preterm " << preterm << "\n";
nsum=0.0;
for(int i=0;i<ntreeleaves;i++)
{
  nsubi=iLeafcounts[i];
  //std::cout << "nsubi " << i << " " << nsubi << "\n";
  ngroup=nsubi+(nrespclasses*alpha);
  //std::cout << "ngroup " << i << " " << ngroup << "\n";
  nterm=sumloggam(ngroup);
  //std::cout << "nterm " << i << " " << nterm << "\n";
  msum=0.0;
  for (int j=0;j<nrespclasses;j++)
  {
    msubij=iLeafRespcounts[i][j];
  }
}
//std::cout << "msubij " << i << " " << j << " " << msubij
<< "\n";
mgroup=msubij+alpha;
//std::cout << "mgroup " << i << " " << j << " " << mgroup
<< "\n";
mterm=sumloggam(mgroup);
//std::cout << "mterm " << i << " " << j << " " << mterm << "\n";
msum=msum+mterm;
//std::cout << "msum " << i << " " << j << " " << msum << "\n";
}
nsum=nsum+msum-nterm;
//std::cout << "nsum " << i << " " << nsum << "\n";
ppost=preterm+nsum;
//std::cout << "ppost" << ppost << "\n";
return ppost;
}

int ntreeleaves;
int nleaflimit;
in iLeafcounts[inleaflimit];
in iLeafRespcounts[inleaflimit][nrespclasses];
};
class DataMat
{
friend class Counts;
public:
DataMat(int=ndatapts,int=nmodelvars)
{
isizendatapts=ndatapts;
isizenvars=nmodelvars;
for(int i=0;i<ndatapts;i++)
{
char * pctemp=new char[5];
strncpy(pctemp,"0000",5);
pctemp[5]="0";
//std::cout << "pctemp" << pctemp << "\n";
pResp[i]=pctemp;
iLeafDat[i]=0;
for(int j=0;j<nmodelvars;j++)
{
Dat[i][j]=0.0;
}
}
}
int getnrows()
{
return isizendatapts;
}
int getncols()
{
return isizenvars;
}
void resetleaves()
{
    for(int i=0;i<ndatapts;i++)
    {
        iLeafDat[i]=0;
    }
}

void infromfile(double dinvars[nmodelvars], char inResp[5], int inrow)
{
    for(int i=0;i<nmodelvars;i++)
    {
        Dat[inrow][i]=dinvars[i];
        //std::cout << "Dat" << Dat[inrow][i] << "\n";
    }
    char * pctemp=new char[5];
    strncpy(pctemp,inResp,5);
    pctemp[5]='\0';
    //std::cout << "pctemp" << pctemp << "\n";
    pcResp[inrow]=pctemp;
}

void AssignLeaves(TNodeList* InNodeList)
{
    bool assignflag;
    TNode* ThisLeaf;
    int ileaf;
    int iNLeaves;
    double upper[nmodelvars];
    double lower[nmodelvars];
    bool inthisleaf[nmodelvars];
    bool alltrue;
    iNLeaves=InNodeList->GetNLeaves();
    //traverse data rows
    for(int irow=0;irow<isizendatapts;irow++)
    {
        //std::cout << "Test Matrix row " << irow << "\n";
        assignflag=false;
        ThisLeaf=InNodeList->GetFirstLeaf();
        ileaf=0;
        //for each row, traverse leaves
        while (assignflag==false && ileaf<iNLeaves)
        {
            //std::cout << "Evaluating Leaf " << ileaf << "\n";
            //get and test bounds
            for(int jvars=0;jvars<nmodelvars;++jvars)
            {
                upper[jvars]=ThisLeaf->GetUpperBound(jvars);
                lower[jvars]=ThisLeaf->GetLowerBound(jvars);
            }
            for(int kvars=0;kvars<nmodelvars;++kvars)
            {
                inthisleaf[kvars]=(lower[kvars]<=Dat[irow][kvars] &&
                                  Dat[irow][kvars]<upper[kvars]);
            }
//std::cout << "Testing " << lower[kvars] << "<= " << Dat[irow][kvars] << "<" << upper[kvars] << "\n";
}
alltrue=true;
for(int lvars=0;lvars<nmodelvars;++lvars)
{
    if(!inthisleaf[lvars])
    {
        alltrue=false;
    }
}
if (alltrue)
{
    assignflag=true;
    ileafDat[irow]=ileaf;
    //std::cout << " assigned ileaf" << ileaf << "\n";
}
else
{
    //get ready for next iteration
    ThisLeaf=ThisLeaf->GetNextLeaf();
    ++ileaf;
    //std::cout << "next ileaf" << ileaf << "\n";
}
}

};

void Countresps(Counts * pinCounter)
{
    //std::cout << "entered Countresps function" << "\n";
    int irespindex;
    for(int i=0;i<ndatapts;i++)
    {
        //std::cout << "Leaf" << i << " " << ileafDat[i] << "\n";
        //std::cout << "index" << i << " " << pcResp[i] << "\n";
        irespindex=getcindex(pcResp[i]);
        //std::cout << "irespindex " << irespindex << "\n";
        ++pinCounter->iLeafRespcounts[ileafDat[i]][irespindex];
        ++pinCounter->iLeafcounts[ileafDat[i]];
        //std::cout << "Respcounts " << i << " " << irespindex << " " << inCounter.iLeafRespcounts[ileafDat[i]][irespindex] << "\n";
        //std::cout << "Leafcounts " << i << " " << inCounter.iLeafcounts[ileafDat[i]] << "\n";
    }
}
double Dat[ndatapts][nmodelvars];
char *pcResp[ndatapts];
int ileafDat[ndatapts];
private:
    int isizendatapts;
    int isizenvars;
};
class DataMatReg
friend class LeafYSums;
public:
DataMatReg(int=ndatapts,int=nmodelvars)
{
    isizendatapts=ndatapts;
    isizenvars=nmodelvars;
    for(int i=0;i<ndatapts;i++)
    {
        YResp[i]=0.0;
        iLeafDat[i]=0;
        for(int j=0;j<nmodelvars;j++)
        {
            Dat[i][j]=0.0;
        }
    }
}
int getnrows()
{
    return isizendatapts;
}
int getncols()
{
    return isizenvars;
}
void resetleaves()
{
    for(int i=0;i<ndatapts;i++)
    {
        iLeafDat[i]=0;
    }
}
void infromfile(double dinvars[nmodelvars], double inYResp, int inrow)
{
    for(int i=0;i<nmodelvars;i++)
    {
        Dat[inrow][i]=dinvars[i];
    }
    YResp[inrow]=inYResp;
}
void AssignLeaves(TNodeList* InNodeList)
{
    bool assignflag;
    TNode* ThisLeaf;
    int ileaf;
    int iNLeaves;
    double upper[nmodelvars];
    double lower[nmodelvars];
    bool inithisleaf[nmodelvars];
    bool alltrue;
    iNLeaves=InNodeList->GetNLeaves();
    //traverse data rows
    for(int irow=0;irow<isizendatapts;irow++)
    {
//std::cout << "Test Matrix row " << irow << "\n";
assignflag=false;
ThisLeaf=InNodeList->GetFirstLeaf();
ileaf=0;
//for each row, traverse leaves
while (assignflag==false && ileaf<iNLeaves)
{
    //std::cout << "Evaluating Leaf " << ileaf << "\n";
    //get and test bounds
    for(int jvars=0;jvars<nmodelvars;++jvars)
    {
        upper[jvars]=ThisLeaf->GetUpperBound(jvars);
        lower[jvars]=ThisLeaf->GetLowerBound(jvars);
    }
    for(int kvars=0;kvars<nmodelvars;++kvars)
    {
        inthisleaf[kvars]=(lower[kvars]<=Dat[irow][kvars] &&
                          Dat[irow][kvars]<upper[kvars]);
        //std::cout << "Testing " << lower[kvars] << "<= "
                   "<" << upper[kvars] << "\n";
        alltrue=true;
        for(int lvars=0;lvars<nmodelvars;++lvars)
        {
            if(!inthisleaf[lvars])
            {
                alltrue=false;
            }
        }
        if (alltrue)
        {
            assignflag=true;
            iLeafDat[irow]=ileaf;
        }
        else
        {
            //get ready for next iteration
            ThisLeaf=ThisLeaf->GetNextLeaf();
            ++ileaf;
            //std::cout << "ileaf" << ileaf << "\n";
        }
    }
    double Dat[ndatapts][nmodelvars];
    double YResp[ndatapts];
    int iLeafDat[ndatapts];
private:
    int isizendatapts;
    int isizenvars;
};

class LeafYSums
public:
LeafYSums()
{
    nleaf=0;
    sumyij=0.0;
    sumofyijsquares=0.0;
    ymean=0.0;
}
int nleaf;
double sumyij;
double sumofyijsquares;
double ymean;
};
class AllLeafYs
{
public:
AllLeafYs(int inleaves, int=ndatapts)
{
    LeafYSums * pLeafY;
    ntreeleaves=inleaves;
    for(int i=0;i<ndatapts;i++)
    {
        pLeafY=new LeafYSums;
        pLeafYs[i]=pLeafY;
    }
}
void newInitYs(int inleaves)
{
    //reset-initialize
    ntreeleaves=inleaves;
    for(int j=0;j<ndatapts;j++)
    {
        pLeafYs[j]->nleaf=0;
        pLeafYs[j]->sumyij=0.0;
        pLeafYs[j]->sumofyijsquares=0.0;
        pLeafYs[j]->ymean=0.0;
    }
}
void SumYs(DataMatReg * pinDat)
{
    for(int i=0;i<ndatapts;i++)
    {
        //std::cout << "Leaf" << i << " " << inDat.iLeafDat[i] << "\n";
        //std::cout << "YResp" << i << " " << inDat.YResp[i] << "\n";
        pLeafYs[pinDat->iLeafDat[i]]->nleaf=pLeafYs[pinDat->
iLeafDat[i]]->nleaf+1;
        //std::cout << "nleaf" << inDat.iLeafDat[i] << " " <<
        pLeafYs[inDat.iLeafDat[i]]->sumyij=pLeafYs[pinDat->
iLeafDat[i]]->sumyij+pinDat->YResp[i];
        pLeafYs[pinDat->iLeafDat[i]]->sumofyijsquares=pLeafYs[pinDat->
iLeafDat[i]]->sumofyijsquares+pow(pinDat->YResp[i],2);
for(int k=0;k<ntreeleaves;k++)
{
    if(pLeafYs[k]->nleaf !=0)
    {
        pLeafYs[k]->ymean=pLeafYs[k]->sumyij/pLeafYs[k]->nleaf;
    }
}

double calcBstar(int nowleaf)
{
    double leafterm;
    double bstar;
    leafterm=pLeafYs[nowleaf]->sumofyijsquares-
    ((vreg*pow(static_cast<double>(pLeafYs[nowleaf]->nleaf),2)*
    pow(static_cast<double>(pLeafYs[nowleaf]->ymean),2))/
    (1+(vreg*pLeafYs[nowleaf]->nleaf)));
    bstar=breg+.5*leafterm;
    return bstar;
}

double reglogmarg()
{
    double first4;
    double astar;
    double nextterm;
    double lastterm;
    double logmarginall;
    double logmarginallleaves=0.0;
    for(int i=0;i<ntreeleaves;i++)
    {
        first4=aastarbnterms(pLeafYs[i]->nleaf);
        astar=areg+(static_cast<double>(pLeafYs[i]->nleaf)/2);
        //std::cout << "astar = " << astar << "\n";
        nextterm=-1*astar*log(calcBstar(i));
        lastterm=.5*log((pLeafYs[i]->nleaf*vreg)+1);
        logmarginall=first4+nextterm-lastterm;
        logmarginallleaves=logmarginallleaves+logmarginall;
    }
    //std::cout << "log marginal likelihood= " << logmarginallleaves
    << "\n";
    return logmarginallleaves;
}
//can modify following to test for zero or for less than some integer
//if(pLeafYs[i]->nleaf < 2)
//if(pLeafYs[i]->nleaf == 0)
bool ZeroOK(int=minleaf)
{
    bool okflag;
    okflag=true;
    for(int i=0;i<ntreeleaves;i++)
    {
        if(pLeafYs[i]->nleaf <= minleaf)


```c++
{
    okflag=false;
}
//std::cout << "leaf " << i << " ok " << okflag << "\n";
return okflag;
}

//note max number of leaves should be <=ndatapts
int ntreeleaves;
LeafYSums * pLeafYs[ndatapts];
};

//leaves of a tree accepted into the posterior distribution
class PostMember
{
    friend class PostList;
    friend class Node;
public:
    PostMember(Node* pTreeLeaf)
    {
        for(int i=0;i<nmodelvars;i++)
        {
            lowerb[i]=pTreeLeaf->pBbounds->Nodebounds[i][0];
            upperb[i]=pTreeLeaf->pBbounds->Nodebounds[i][1];
            //std::cout << "saving leaf bounds " << lowerb[i] << " , " << upperb[i] << "\n";
        }
        pNextLf=0;
        pPrevLf=0;
    }
    double upperb[nmodelvars];
    double lowerb[nmodelvars];
    PostMember* pNextLf;
    PostMember* pPrevLf;
};
//tree accepted into the posterior distribution
//contains leaf boundaries and log marginal likelihood
class PostList
{
    public:
    PostList(double llike)
    {
        pFirstLf=0;
        pLastLf=0;
        nleaf=0;
        pNextTree=0;
        pPrevTree=0;
        logmarginallike=llike;
    }
    bool IsEmptyLf()
    {
        return pFirstLf==0;
    }
```
void addtoLves(Node* pInLf)
{
    if (pInLf->IsLeaf())
    {
        PostMember* pNewLf = new PostMember(pInLf);
        if (IsEmptyLf())
        {
            pFirstLf=pLastLf=pNewLf;
            ++nleaf;
        }
        else
        {
            pNewLf->pPrevLf=pLastLf;
            pLastLf->pNextLf=pNewLf;
            pLastLf=pNewLf;
            ++nleaf;
        }
    }
}

void CopyLeafList(TNodeList* pNodeList)
{
    int ntotal;
    ntotal=pNodeList->GetNLeaves();
    TNode* ptraverse;
    ptraverse=pNodeList->GetFirstLeaf();
    while (ptraverse && nleaf<=ntotal)
    {
        addtoLves(ptraverse->pThisLeaf);
        ptraverse=ptraverse->pNextLeaf;
    }
}

int nleaf;
double logmarginallike;
PostMember* pFirstLf;
PostMember* pLastLf;
PostList* pNextTree;
PostList* pPrevTree;
};
class PostCollection
{
public:
    PostCollection(int nmodelvars)
    {
        nmvars=nmodelvars;
        for(int j=0; j<nmodelvars; j++)
        {
            char * ptemp=new char[8];
            strncpy(ptemp,modelvars[j],8);
            ptemp[8]=\0';
            modvars[j]=ptemp;
        }
    }
    // Methods...
int readcounter=0;
for ( int invar=0; invar<nmodelvars; invar++)
{
    initlowerb[invar]=initbounds[readcounter];
    readcounter++;
    //std::cout << "saving init lower b " << invar << "=" << initlowerb[invar] << "\n";
    initupperb[invar]=initbounds[readcounter];
    readcounter++;
    //std::cout << "saving init upper b " << invar << "=" << initupperb[invar] << "\n";
}
pFirstTree=0;
plastTree=0;
nposttrees=0;
}
bool IsEmptyCollection()
{
    return pFirstTree==0;
}

void AddPostTree(TNodeList* pAcceptTree, double logmlikeli)
{
    PostList* pNewTree = new PostList(logmlikeli);
pNewTree->CopyLeafList(pAcceptTree);
    if (IsEmptyCollection())
    {
        pFirstTree=pLastTree=pNewTree;
        ++nposttrees;
    }
    else
    {
        pNewTree->pPrevTree=pLastTree;
        pLastTree->pNextTree=pNewTree;
        pLastTree=pNewTree;
        ++nposttrees;
    }
    //std::cout << "Added Posterior Tree # " << nposttrees << "\n";
    //std::cout << "with Log Marginal Likelihood of " << pNewTree->logmarginallike << "\n";
}

void PostOut(ofstream &outfile)
{
    PostList* pOutTree;
    PostMember* pOutLeaf;
pOutTree=pFirstTree;
    for (int i=0;i<nposttrees;i++)
    {
        pOutLeaf=pOutTree->pFirstLf;
        for(int j=0;j<pOutTree->nleaf;j++)
outfile << i << " " << pOutTree->logmarginallike << " " << j << " ";
  for(int k=0;k<nmodelvars;k++)
  {
    outfile << pOutLeaf->lowerb[k] << " " << pOutLeaf->upperb[k] << " ";
  }
  outfile << "\n";
  pOutLeaf=pOutLeaf->pNextLf;
  pOutTree=pOutTree->pNextTree;
}
}

void CleanPost()
{
  PostList* pPtemp;
  PostList* pPnext;
  PostMember* pPLtemp;
  PostMember* pPLnext;
  pPtemp=pFirstTree;
  for (int i=0;i<nposttrees;i++)
  {
    pPLtemp=pPtemp->pFirstLf;
    for(int j=0;j<pPtemp->nleaf;j++)
    {
      pPLnext=pPLtemp->pNextLf;
      delete pPLtemp;
      pPLtemp=pPLnext;
    }
    pPnext=pPtemp->pNextTree;
    delete pPtemp;
    pPtemp=pPnext;
  }
  PostList* pFirstTree;
  PostList* pLastTree;
  int nposttrees;
  int nmvars;
  char *modvars[nmodelvars];
  double initupperb[nmodelvars];
  double initlowerb[nmodelvars];
};

bool AcceptTree(bool NoZeroLeaf, double ThisLogMarg, PostCollection* pPostSpace, int numleav, double ThisR=1)
{
  bool Accept;
  int AcceptProb;
  int AcceptRand;
  if(!NoZeroLeaf)
  {

Accept=false;
   //std::cout << "Zero Leaf" << "\n";
}
else
{
    double bfR;
    double thisterm;
    double lastterm;
    double testterm;
    //use BIC
    //but change order because lower BIC is better
    thisterm=(-2*ThisLogMarg)+
        (static_cast<double>(numleav)*log(static_cast<double>(ndatapts )));
    lastterm=(-2*pPostSpace->pLastTree->logmarginallike)+
        (static_cast<double>(pPostSpace->pLastTree->nleaf)*log(static_cast<double>(ndatapts )));
    testterm=0.5*(lastterm - thisterm);
    if (testterm>10.0) {
      testterm=10.0;
      std::cout << "testterm= " << testterm << "\n";
    }
    bfR=ThisR*exp(testterm);
    //bfR=ThisR*exp(ThisLogMarg - pPostSpace->pLastTree->logmarginallike);
    //std::cout << "This Log Marginal = " << ThisLogMarg << "\n";
    //std::cout << "Last Accepted Log Marginal = "
    // << pPostSpace->pLastTree->logmarginallike << "\n";
    //std::cout << "bfR= " << bfR << "\n";
    if(bfR>1.0)
    {
      bfR=1.0;
      Accept=true;
    }
    else
    {
      AcceptProb=static_cast<int>(bfR*100);
      AcceptRand = rand() % 100;
      //std::cout << "random accept" << AcceptRand << "\n";
      if(AcceptRand<AcceptProb)
      {
        Accept=true;
      }
      else
      {
        Accept=false;
      }
    }
}
return Accept;

Tree* CopyTree(Tree* pOldTree)
{
//each tree has a root and at least two leaves
Tree* pCPTree = new Tree(0, false);
NodeTable* pTable = new NodeTable;
NodeTableRow* pCurrent = 0;
NodeTableRow* pTabLeft = 0;
NodeTableRow* pTabRight = 0;
pCurrent = pTable->AddRow1(pOldTree->pRoot);

//std::cout << "pCurrent " << pCurrent->ThisID << "\n";
//create a new node without a new node id;
Node* tempnode = new Node(true);
pCPTree->pRoot = tempnode;

//info
pCurrent->pNewThis = pCPTree->pRoot;
pCPTree->pRoot->iNodeID = pCurrent->pThis->iNodeID;
pCPTree->pRoot->ntThis = pCurrent->pThis->ntThis;
strncpy(pCPTree->pRoot->cSplitvar, pCurrent->pThis->cSplitvar, 8);
pCPTree->pRoot->cSplitvar[8] = '\0';
pCPTree->pRoot->dSplitval = pCurrent->pThis->dSplitval;
pCPTree->pRoot->pRootPointer = pCPTree->pRoot;
pCPTree->pRoot->pBbounds->CopyBounds(pCurrent->pThis->pBbounds->Nodebounds, pCPTree->pRoot);

//std::cout << "new copy root node " << pCPTree->pRoot->iNodeID << "\n";
//std::cout << "new copy root node " << pCPTree->pRoot->ntThis << "\n";
//std::cout << "new copy root node " << pCPTree->pRoot->cSplitvar << "\n";
//std::cout << "new copy root node " << pCPTree->pRoot->dSplitval << "\n";
//std::cout << "new copy root node 0,0 " << pCPTree->pRoot->pBbounds->Nodebounds[0][0] << "\n";
//std::cout << "new copy root node 1,1 " << pCPTree->pRoot->pBbounds->Nodebounds[1][1] << "\n";
Node* pLeft = new Node(true);
Node* pRight = new Node(true);
pCurrent->pNewThis->pLeftChild = pLeft;
pCurrent->pNewThis->pRightChild = pRight;
pLeft->pParent = pCurrent->pNewThis;
pLeft->pRootPointer = pCurrent->pNewThis->pRootPointer;
pRight->pParent = pCurrent->pNewThis;
pRight->pRootPointer = pCurrent->pNewThis->pRootPointer;
pTabLeft = pTable->AddRow2(pCurrent->pOldLeftChild, pCurrent);
pTabLeft->pNewThis = pLeft;

//std::cout << "old left node " << pTabLeft->pThis->iNodeID << "\n";
//std::cout << "old left node " << pTabLeft->pThis->ntThis << "\n";
//std::cout << "old left node " << pTabLeft->pThis->cSplitvar << "\n";
//std::cout << "old left node " << pTabLeft->pThis->dSplitval << "\n";
//info
pTabLeft->pNewThis->iNodeID = pTabLeft->pThis->iNodeID;
pTabLeft->pNewThis->ntThis = pTabLeft->pThis->ntThis;
strncpy(pTableLeft->pNewThis->cSplitvar,pTabLeft->pThis->cSplitvar,8);
pTabLeft->pNewThis->cSplitval[8] = '\0';
pTabLeft->pNewThis->dSplitval = pTabLeft->pThis->dSplitval;
pTabLeft->pNewThis->pBbounds->CopyBounds(pTabLeft->pThis->pBbounds->Nodebounds, pTabLeft->pNewThis);
    //std::cout << "new copy left node " << pCPTree->pRoot->pLeftChild->iNodeID << "\n";
    //std::cout << "new copy left node " << pCPTree->pRoot->pLeftChild->ntThis << "\n";
    //std::cout << "new copy left node " << pCPTree->pRoot->pLeftChild->cSplitvar << "\n";
    //std::cout << "new copy left node " << pCPTree->pRoot->pLeftChild->dSplitval << "\n";
    pLeft=pTable->AddRow2(pCurrent->pOldLeftChild,pCurrent);
pTabRight=pTable->AddRow2(pCurrent->pOldRightChild,pCurrent);
    //std::cout << "old Right node " << pTabRight->pThis->iNodeID << "\n";
    //std::cout << "old Right node " << pTabRight->pThis->ntThis << "\n";
    //std::cout << "old Right node " << pTabRight->pThis->cSplitvar << "\n";
    //std::cout << "old Right node " << pTabRight->pThis->dSplitval << "\n";
    //info
    pTabRight->pNewThis->iNodeID=pTabRight->pThis->iNodeID;
pTabRight->pNewThis->ntThis=pTabRight->pThis->ntThis;
    strncpy(pTabRight->pNewThis->cSplitvar,pTabRight->pThis->cSplitvar,8);
pTabRight->pNewThis->cSplitval[8] = '\0';
pTabRight->pNewThis->dSplitval = pTabRight->pThis->dSplitval;
pTabRight->pNewThis->pBbounds->CopyBounds(pTabRight->pThis->pBbounds->Nodebounds, pTabRight->pNewThis);
    //std::cout << "new copy Right node " << pCPTree->pRoot->pRightChild->iNodeID << "\n";
    //std::cout << "new copy Right node " << pCPTree->pRoot->pRightChild->ntThis << "\n";
    //std::cout << "new copy Right node " << pCPTree->pRoot->pRightChild->cSplitvar << "\n";
    //std::cout << "new copy Right node " << pCPTree->pRoot->pRightChild->dSplitval << "\n";
PCurrent=pCurrent->GetNext();
    //std::cout << "Current " << pCurrent->ThisID << "\n";
while (pCurrent !=0)
    {if ( !(pCurrent->pThis->IsLeaf()) )
    {
pTable->AddChildren(pCurrent);
pLeft=new Node(true);
pRight=new Node(true);
pCurrent->pNewThis->pLeftChild=pLeft;
pCurrent->pNewThis->pRightChild=pRight;
pLeft->pParent=pCurrent->pNewThis;
pLeft->pRootPointer=pCurrent->pNewThis->pRootPointer;
pRight->pParent=pCurrent->pNewThis;
pRight->pRootPointer=pCurrent->pNewThis->pRootPointer;
pTabLeft=pTable->AddRow2(pCurrent->pOldLeftChild, pCurrent);
pTabLeft->pRight=pNewThis->pLeft;
//std::cout << "old left node " << pTabLeft->pThis->iNodeID << 
\n";
//std::cout << "old left node " << pTabLeft->pThis->ntThis << 
\n";
//std::cout << "old left node " << pTabLeft->pThis->cSplitvar << 
\n";
//std::cout << "old left node " << pTabLeft->pThis->dSplitval << 
\n";
//info
pTabLeft->pNewThis->iNodeID=pTabLeft->pThis->iNodeID;
pTabLeft->pNewThis->ntThis=pTabLeft->pThis->ntThis;
strncpy(pTabLeft->pNewThis->cSplitvar, pTabLeft->pThis->cSplitvar, 8);
pTabLeft->pNewThis->cSplitvar[8]=\0;
pTabLeft->pNewThis->dSplitval=pTabLeft->pThis->dSplitval;
pTabLeft->pNewThis->pBbounds->CopyBounds(pTabLeft->pThis->pBbounds->Nodebounds, pTabLeft->pNewThis);
//std::cout << "new copy left node " << pTabLeft->pNewThis->iNodeID << "\n";
//std::cout << "new copy left node " << pTabLeft->pNewThis->ntThis << "\n";
//std::cout << "new copy left node " << pTabLeft->pNewThis->cSplitvar << "\n";
//std::cout << "new copy left node " << pTabLeft->pNewThis->dSplitval << "\n";
//info
pTabRight->pNewThis->iNodeID=pTabRight->pThis->iNodeID;
pTabRight->pNewThis->ntThis=pTabRight->pThis->ntThis;
strncpy(pTabRight->pNewThis->cSplitvar, pTabRight->pThis->cSplitvar, 8);
pTabRight->pNewThis->cSplitvar[8]=\0;
pTabRight->pNewThis->dSplitval=pTabRight->pThis->dSplitval;
pTabRight->pNewThis->pBbounds->CopyBounds(pTabRight->pThis->pBbounds->Nodebounds, pTabRight->pNewThis);
//std::cout << "new copy Right node " << pTabRight->pNewThis->iNodeID << "\n";
//std::cout << "old Right node " << pTabRight->pThis->iNodeID << "\n";
//std::cout << "old Right node " << pTabRight->pThis->ntThis << "\n";
//std::cout << "old Right node " << pTabRight->pThis->cSplitvar << "\n";
//std::cout << "old Right node " << pTabRight->pThis->dSplitval << "\n";
//info
pTabRight->pNewThis->iNodeID=pTabRight->pThis->iNodeID;
pTabRight->pNewThis->ntThis=pTabRight->pThis->ntThis;
strncpy(pTabRight->pNewThis->cSplitvar, pTabRight->pThis->cSplitvar, 8);
pTabRight->pNewThis->cSplitvar[8]=\0;
pTabRight->pNewThis->dSplitval=pTabRight->pThis->dSplitval;
pTabRight->pNewThis->pBbounds->CopyBounds(pTabRight->pThis->pBbounds->Nodebounds, pTabRight->pNewThis);
void MCMCmain(int niter)
{
    // srand(seed);
    bool Okleaves;
    bool Oktree;
    TNodeList* pNodeTypes1;
    TNodeList* pKeepTypes;
    // following lists help with cleaning up
    TNodeList* pNodeKeepOut;
    TNodeList* pNodeWorkOut;
    int NumLeaves;
    DataMat Test1;
    double dvars[nmodelvars];
    double dintemp;
    char Resp[5];
    double lmarginal;
    Jumptype Jump;
    double R;
    ifstream
    inClientFile("classinputfilepath\classinputfilenamea.txt", ios::in);
    if (!inClientFile)
    {
        // std::cout << "Cannot Open File" << "\n";
    }
    for(int iin=0;iin<ndatapts;iin++)
    {
        for(int j=0;j<nmodelvars;j++)
        {
            inClientFile >> dintemp;
            dvars[j]=dintemp;
        }
        inClientFile >> Resp;
        Test1.infromfile(dvars,Resp,iin);
    }
    PostCollection* pPosteriorSpace=new PostCollection;
    Tree* pWorkTree=new Tree;
    Tree* pKeepTree;
    Tree* ptemp;
// first tree
// std::cout << "first tree new " << pWorkTree->NewTree << "\n";
pWorkTree->rootSplitSetup();
pNodeTypes1=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
NumLeaves=pNodeTypes1->GetNLeaves();
std::cout << "NumLeaves" << NumLeaves << "\n";
Test1.AssignLeaves(pNodeTypes1);
Counts CountTest(NumLeaves);
// std::cout << "Test1check00" << Test1.Dat[0][0] << "\n";
// std::cout << "Test1checkmax" << Test1.Dat[2599][31] << "\n";
Test1.CountResp(&CountTest);
lmarginal=CountTest.logppost();
std::cout << "lmarginal" << lmarginal << "\n";
pPosteriorSpace->AddPostTree(pNodeTypes1, lmarginal);

// copy first tree to keep tree
// work tree remains as itself
pKeepTree=CopyTree(pWorkTree);
// more trees
for (int itree=0; itree<niter; itree++)
{
    Jump=pWorkTree->NewJump(pNodeTypes1);
pNodeTypes1->TNListCleanUp();
pNodeTypes1=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
NumLeaves=pNodeTypes1->GetNLeaves();
pKeepTypes=pKeepTree->NodeTypeLists(pKeepTree->pRoot);
R=getR(Jump, &pNodeType1->iLeafCount, pNodeTypes1->iLPCount, pKeepTypes->iLeafCount, pKeepTypes->iLPCount);
pKeepTypes->TNListCleanUp();
Test1.resetleaves();
Test1.AssignLeaves(pNodeTypes1);
CountTest.ResetCounts(NumLeaves);
Test1.CountResp(&CountTest);
for(int i=0; i<NumLeaves; i++)
{
    // std::cout << "Leafcounts " << i << " " << 
    CountTest.iLeafcounts[i] << "\n";
    for (int j=0; j<nrespclasses; j++)
    {
        // std::cout << "Leaf Resp counts " << i << " " << j << " 
        CountTest.iLeafRespcounts[i][j] << "\n";
    }
}
lmarginal=CountTest.logppost();
Okleaves=CountTest.ZeroOK();
Oktree=AcceptTree(Okleaves, lmarginal, pPosteriorSpace, R);
if(Oktree)
{
    std::cout << "Accepting Tree" << "\n";
pPosteriorSpace->AddPostTree(pNodeTypes1, lmarginal);
    // clean out old keep tree
    pNodeKeepOut=pKeepTree->NodeTypeLists(pKeepTree->pRoot);
pNodeKeepOut->CleanUpTree();
pNodeKeepOut->TNListCleanUp();
    // copy accepted tree to keep tree
pKeepTree=CopyTree(pWorkTree);
}
else
{
    //std::cout << "Rejecting Tree" << "\n";
    //clean out old work tree
    pNodeWorkOut=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
    pNodeWorkOut->CleanUpTree();
    pNodeWorkOut->TNListCleanUp();
    //copy keep tree to work tree
    pWorkTree=CopyTree(pKeepTree);
    pNodeTypes1->TNListCleanUp();
    pNodeTypes1=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
    NumLeaves=pNodeTypes1->GetNLeaves();
}
ofstream
outPostFile("classoutputfilepath\classoutputfilenamexa.txt",
ios::out );
pPosteriorSpace->PostOut(outPostFile);
void MCMCmainreg(int niter, unsigned seed, ofstream &outPostFile)
{
    srand(seed);
    bool Okleaves;
    bool Oktree;
    TNodeList* pNodeTypes1;
    TNodeList* pKeepTypes;
    //following lists help with cleaning up
    TNodeList* pNodeKeepOut;
    TNodeList* pNodeWorkOut;
    int NumLeaves;
    DataMatReg Test1;
    double dvars[nmodelvars];
    double dintemp;
    double dinyResp;
    double lmarginal;
    Jumptype Jump;
    double R;
    ifstream inClientFile("inputfilepath\inputfilenamexa.txt", ios::in );
    if (!inClientFile)
    {
        std::cout << "Cannot Open File" << "\n"
    }
    for(int iin=0;iin<ndatapts;iin++)
    {
        for(int j=0;j<nmodelvars;j++)
        {
            inClientFile >> dintemp;
            dvars[j]=dintemp;
        }
inClientFile >> dinyResp;
Test1.infromfile(dvars,dinyResp,iin);
}
PostCollection* pPosteriorSpace=new PostCollection;
Tree* pWorkTree=new Tree;
Tree* pKeepTree;
Tree* ptemp;
//first tree
std::cout << "first tree new " << pWorkTree->NewTree << "\n";
pWorkTree->rootsplitsetup();
pNodeType1=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
NumLeaves=pNodeType1->GetNLeaves();
Test1.AssignLeaves(pNodeType1);
AllLeafYs YTest(NumLeaves);
YTest.SumYs(&Test1);
lmarginal=YTest.reglogmarg();
pPosteriorSpace->AddPostTree(pNodeType1, lmarginal);
//copy first tree to keep tree
//work tree remains as itself
pKeepTree=CopyTree(pWorkTree);
//more trees
for (int itree=0; itree<niter; itree++)
{
  Jump=pWorkTree->NewJump(pNodeType1);
  pNodeType1->TNListCleanUp();
pNodeType1=pWorkTree->NodeTypeLists(pWorkTree->pRoot);
  NumLeaves=pNodeType1->GetNLeaves();

  //this should simplify things
  pKeepTypes=pKeepTree->NodeTypeLists(pKeepTree->pRoot);
  R=getR(Jump, pNodeType1->iLeafCount, pNodeType1->iLPCount,
        pKeepTypes->iLeafCount, pKeepTypes->iLPCount);
pKeepTypes->TNListCleanUp();
  Test1.resetleaves();
  Test1.AssignLeaves(pNodeType1);
  YTest.newInitYs(NumLeaves);
  YTest.SumYs(&Test1);
  lmarginal=YTest.reglogmarg();
  Okleaves=YTest.ZeroOK();
  Oktree=AcceptTree(Okleaves, lmarginal, pPosteriorSpace, NumLeaves, R);
  if(Oktree)
  {
    //std::cout << "Accepting Tree" << "\n";
    pPosteriorSpace->AddPostTree(pNodeType1, lmarginal);
    //clean out old keep tree
    pNodeKeepOut=pKeepTree->NodeTypeLists(pKeepTree->pRoot);
    pNodeKeepOut->CleanUpTree();
pNodeKeepOut->TNListCleanUp();
    //copy accepted tree to keep tree
    pKeepTree=CopyTree(pWorkTree);
  }
}
else
void std::cout
// Rejection Tree
// Clean out old work tree
pNodeWorkOut = pWorkTree->NodeTypeLists(pWorkTree->pRoot);
    pNodeWorkOut->CleanUpTree();
    pNodeWorkOut->TNListCleanUp();
// Copy keep tree to work tree
pWorkTree = CopyTree(pKeepTree);
    pNodeTypes1->TNListCleanUp();
    pNodeTypes1 = pWorkTree->NodeTypeLists(pWorkTree->pRoot);
    NumLeaves = pNodeTypes1->GetNLeaves();
}

pPosteriorSpace->PostOut(outPostFile);
    pPosteriorSpace->CleanPost();
}

int main(int argc, char * argv[])
{
    unsigned seed1 = 1510213410;
    ofstream outPostFile1("outputfilepath\outputfilenamet.txt",
        ios::out);
    MCMCMainreg(200000, seed1, outPostFile1);
    unsigned seed2 = 85195397;
    ofstream outPostFile2("outputfilepath\outputfilenameb.txt",
        ios::out);
    MCMCMainreg(200000, seed2, outPostFile2);
    unsigned seed3 = 204139659;
    ofstream outPostFile3("outputfilepath\outputfilenamec.txt",
        ios::out);
    MCMCMainreg(200000, seed3, outPostFile3);
    unsigned seed4 = 1830852445;
    ofstream outPostFile4("outputfilepath\outputfilenamed.txt",
        ios::out);
    MCMCMainreg(200000, seed4, outPostFile4);
    unsigned seed5 = 1991176727;
    ofstream outPostFile5("outputfilepath\outputfilenamee.txt",
        ios::out);
    MCMCMainreg(200000, seed5, outPostFile5);
    return 0;
}
Appendix D

Additional Results

This is a supplement to a report about an application of Bayesian Regression Tree modeling to environmental conditions in 10 Permafrost Areas of Alaska, USA. The main report includes results and discussion based upon an analysis of the northernmost Permafrost Area of Alaska. The results of the analyses of the other nine Permafrost Areas of Alaska may be found in this report, which will be available online, on the website of The Grant F. Walton Center for Remote Sensing and Spatial Analysis (CRSSA) at http://crssa.rutgers.edu/projects/.
Permafrost Area 11

This section describes the analysis of Permafrost Area 11, which is Alaska’s mountainous area underlain by continuous permafrost (Ferrians, 1998). It is associated with the Brooks Range geographic region (National Geographic Society, 2002), the Brooks / British Range Tundra ecoregion (Ricketts, et al., 1999), and it creates the border between the Arctic and Interior climate regions (Shulski & Wendler, 2007).

The tree resulting from the cloudiness analysis of Permafrost Area 11 is shown in Figure 20. Five subsets were delineated. The early season cloudiness variable was not used in this tree model.

The subset boundaries that describe the distance between each subset and the point of no difference from study period mean are listed in Table 13. Subset 2 contains the origin (indicated by dashes in each dimension), so its conditions include the point of no change from study period mean. Table 14 shows that the subsets with greater than study period mean seasonal cloudiness tended to occur earlier in the study period and those with lower than study period mean seasonal cloudiness tended to occur later.

Table 15 shows the distributions of the data points by subset and year. Subset 3, with cloudier than study period mean conditions, displays a multimodal distribution with its largest mode occurring in 2003. The spatial aspect of the cloudiness analysis is illustrated in Figure 21. Black shading implies the highest mean seasonal proportion of cloudy images; white implies the lowest mean seasonal proportion of cloudy images.
Figure 20. Tree model produced by the cloudiness analysis of Permafrost Area 11
Table 13. Subset boundaries and collection year summary, sorted by subset, for the validation set of Permafrost Area 11

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-20</td>
<td>-</td>
<td>-36</td>
<td>78</td>
<td>2004.987</td>
<td>2007</td>
</tr>
<tr>
<td>1</td>
<td>-20</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>2003.456</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2220</td>
<td>2001.098</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>1347</td>
<td>2000.682</td>
<td>2003</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>165</td>
<td>1999.048</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 14. Subset boundaries and collection year summary, sorted by mean collection year, for the validation set of Permafrost Area 11

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>165</td>
<td>1999.048</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>1347</td>
<td>2000.682</td>
<td>2003</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2220</td>
<td>2001.098</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-20</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>2003.456</td>
<td>2004</td>
</tr>
<tr>
<td>0</td>
<td>-20</td>
<td>-</td>
<td>-36</td>
<td>78</td>
<td>2004.987</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 15. Number of observations within subsets by collection year from the validation set of Permafrost Area 11

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>21</td>
<td>20</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>24</td>
<td>17</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>245</td>
<td>155</td>
<td>216</td>
<td>97</td>
<td>144</td>
<td>67</td>
<td>194</td>
<td>205</td>
<td>85</td>
<td>246</td>
<td>185</td>
<td>166</td>
<td>215</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>35</td>
<td>132</td>
<td>66</td>
<td>159</td>
<td>146</td>
<td>178</td>
<td>99</td>
<td>88</td>
<td>200</td>
<td>9</td>
<td>78</td>
<td>116</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>44</td>
<td>7</td>
<td>55</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 21 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 11
Permafrost Area 11, Cloudiness Analysis Subset 2

Subset 2, from the first stage analysis of Permafrost Area 11, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 22. Eleven subsets were found.

The subset boundaries that describe the distance between each subset and the point of no difference from study period mean are shown in Table 16 and 5. The subset that contains the point of no change from study period mean, Subset 4, had the earliest mean collection year. Temperatures increased from study period mean, by at least 1 or 2 percentage points, in the later subsets, though a temperature increase was also seen in one subset with a mid-study-period mean collection year. Precipitable water increased in some subsets in the early- to mid-study-period and decreased in several later subsets. NDVI decreased in the middle of the study period.

Table 18 shows the details of the distributions of the data points by subset and year. Periodicity (Subsets 0 and 3), unimodal distributions, and multimodal distributions are evident. The geographic progression of conditions through Permafrost Area 11 is illustrated in Figure 10, using an RGB (red = temperature, green = vegetation, blue = precipitable water) color scale. A marked change in conditions appears between 2003 and 2004. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 24, precipitation vs. temperature in Figure 25, and temperature vs. vegetation in Figure 26.
Figure 22. Tree model resulting from the second stage analysis of Cloudiness Subset 2 of Permafrost Area 11
Table 16. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 2 of Permafrost Area

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-20</td>
<td>-</td>
<td>147</td>
<td>2001.952</td>
<td>2006</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-37</td>
<td>-</td>
<td>141</td>
<td>2004.34</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-20</td>
<td>-</td>
<td>225</td>
<td>2003.667</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-5</td>
<td>93</td>
<td>2001.333</td>
<td>2001</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>278</td>
<td>1998.478</td>
<td>1997</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>244</td>
<td>2000.066</td>
<td>1995</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>3</td>
<td>-17</td>
<td>31</td>
<td>2001.71</td>
<td>2002</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>3</td>
<td>-6</td>
<td>135</td>
<td>2001.259</td>
<td>2002</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>657</td>
<td>1999.183</td>
<td>1995</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-23</td>
<td>-</td>
<td>204</td>
<td>2004.912</td>
<td>2004</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>58</td>
<td>2004.828</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 17. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 2 of Permafrost Area

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>278</td>
<td>1998.478</td>
<td>1997</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>657</td>
<td>1999.183</td>
<td>1995</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>244</td>
<td>2000.066</td>
<td>1995</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>3</td>
<td>-6</td>
<td>135</td>
<td>2001.259</td>
<td>2002</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-5</td>
<td>93</td>
<td>2001.333</td>
<td>2001</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>3</td>
<td>-17</td>
<td>31</td>
<td>2001.71</td>
<td>2002</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-20</td>
<td>-</td>
<td>147</td>
<td>2001.952</td>
<td>2006</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-20</td>
<td>-</td>
<td>225</td>
<td>2003.667</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-37</td>
<td>-</td>
<td>141</td>
<td>2004.34</td>
<td>2004</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>58</td>
<td>2004.828</td>
<td>2007</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-23</td>
<td>-</td>
<td>204</td>
<td>2004.912</td>
<td>2004</td>
</tr>
</tbody>
</table>
Table 18. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 2 of Permafrost Area 11

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>36</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>25</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>40</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>4</td>
<td>25</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>64</td>
<td>42</td>
<td>30</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>34</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>59</td>
<td>68</td>
<td>12</td>
<td>14</td>
<td>34</td>
<td>16</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>72</td>
<td>41</td>
<td>68</td>
<td>12</td>
<td>14</td>
<td>34</td>
<td>16</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>18</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>36</td>
<td>65</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>107</td>
<td>33</td>
<td>65</td>
<td>78</td>
<td>107</td>
<td>17</td>
<td>75</td>
<td>88</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>46</td>
<td>6</td>
<td>55</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>32</td>
</tr>
</tbody>
</table>
Figure 23. Sequence of maps of Permafrost Area 11 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 2)
Figure 24. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space.
Area 11, Cloudiness Subset 2

Figure 25. Projection of the validation data points onto the precipitable water vs. temperature plane of state space
Figure 26. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
Permafrost Area 12

This section describes the analysis of Permafrost Area 12, which is Alaska's mountainous area underlain by discontinuous permafrost (Ferrians, 1998). It is associated with the Alaska Range and Interior Highlands geographic regions (National Geographic Society, 2002), the Alaska / St. Elias Range Tundra, Interior Yukon / Alaska Alpine Tundra, and Ogilvie / MacKenzie Alpine Tundra ecoregions (Ricketts, et al., 1999), and it borders the Interior and Copper River Basin climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 12 is shown in Figure 27. Five subsets were defined. The seasonal cloudiness variable was not used in the tree model.

Table 19 shows that Subset 1 contains the origin (indicated by dashes in each dimension). Table 20 shows that the subsets with greater cloudiness occurred earlier in the study period and that the one with less cloudiness occurred later.

Table 21 shows the distributions of the data points by subset and year. Subset 2, with greater late season cloudiness, displays a maximum in 2003. The spatial aspect of the cloudiness analysis is illustrated in Figure 28. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Area 12, Cloudiness Analysis

diff btn % early cloudy
images and % early
1995-2007 < 16

Leaf 0

Leaf 1

Leaf 2

diff btn % late cloudy
images and % late
1995-2007 < -17

Leaf 3

diff btn % early cloudy
images and % early
1995-2007 < 63

Leaf 4

1995-2007 < 6
Table 19. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 12

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-17</td>
<td>419</td>
<td>2002.85</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1543</td>
<td>2001.349</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>1302</td>
<td>2000.23</td>
<td>2003</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>625</td>
<td>2000.52</td>
<td>2000</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>63</td>
<td>-</td>
<td>11</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 20. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 12

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-</td>
<td>63</td>
<td>-</td>
<td>11</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>1302</td>
<td>2000.23</td>
<td>2003</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>625</td>
<td>2000.52</td>
<td>2000</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1543</td>
<td>2001.349</td>
<td>2001</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-17</td>
<td>419</td>
<td>2002.85</td>
<td>2004</td>
</tr>
</tbody>
</table>

Table 21. Number of observations within subsets by collection year from validation set of Permafrost Area 12

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>15</td>
<td>26</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>48</td>
<td>17</td>
<td>31</td>
<td>83</td>
<td>60</td>
<td>31</td>
<td>61</td>
</tr>
<tr>
<td>1</td>
<td>107</td>
<td>123</td>
<td>152</td>
<td>86</td>
<td>100</td>
<td>50</td>
<td>172</td>
<td>74</td>
<td>77</td>
<td>161</td>
<td>149</td>
<td>148</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>155</td>
<td>143</td>
<td>93</td>
<td>109</td>
<td>134</td>
<td>66</td>
<td>44</td>
<td>127</td>
<td>173</td>
<td>30</td>
<td>81</td>
<td>93</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>19</td>
<td>29</td>
<td>87</td>
<td>53</td>
<td>167</td>
<td>36</td>
<td>82</td>
<td>19</td>
<td>26</td>
<td>10</td>
<td>27</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 28 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 12
Permafrost Area 12, Cloudiness Analysis Subset 1

Subset 1, from the first stage analysis of Permafrost Area 12, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 29. Ten subsets were found.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 22 and 11. The subset containing the point of no change from study period mean, Subset 7, had the earliest mean collection year. Temperatures were greater than study period mean in one subset with an early mean collection year (Subset 8) and in one with a late mean collection year (Subset 5). Several subsets with mid-study-period to late mean collection years exhibited lower than study period mean precipitable water values, with the subsets with the later mean collection years showing the greatest decreases. Decreases from study period mean NDVI accompanied decreases in precipitable water in some small subsets: Subsets 0, 1, and 2. Increases from study period mean NDVI were observed in Subset 9 and, non-intuitively, paired with a decrease in precipitable water, in Subset 4.
Table 24 shows the distributions of the data points by subset and year. The geographic progression of conditions through Permafrost Area 12 is illustrated in Figure 11, using an RGB (red = temperature, green = vegetation, blue = precipitable water) color scale. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 31, precipitation vs. temperature in Figure 32, and temperature vs. vegetation in Figure 33.
Figure 29. Tree model resulting from the second stage analysis of Cloudiness Subset 1 of Permafrost Area 12
### Table 22. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 12

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-12</td>
<td>-96</td>
<td>10</td>
<td>2001.6</td>
<td>2001</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-12</td>
<td>-43</td>
<td>16</td>
<td>2000.625</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-25</td>
<td>-12</td>
<td>15</td>
<td>2005.4</td>
<td>2006</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-25</td>
<td>-</td>
<td>92</td>
<td>2004.478</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-25</td>
<td>2</td>
<td>162</td>
<td>2004.784</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-25</td>
<td>-</td>
<td>102</td>
<td>2004.814</td>
<td>2005</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-12</td>
<td>-</td>
<td>233</td>
<td>2003.197</td>
<td>2007</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>614</td>
<td>1999.096</td>
<td>2001</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>280</td>
<td>2000.246</td>
<td>1997</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>59</td>
<td>19</td>
<td>2001.947</td>
<td>2004</td>
</tr>
</tbody>
</table>

### Table 23. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 12

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>614</td>
<td>1999.096</td>
<td>2001</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>280</td>
<td>2000.246</td>
<td>1997</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-12</td>
<td>-43</td>
<td>16</td>
<td>2000.625</td>
<td>2001</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-12</td>
<td>-96</td>
<td>10</td>
<td>2001.6</td>
<td>2001</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>59</td>
<td>19</td>
<td>2001.947</td>
<td>2004</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-12</td>
<td>-</td>
<td>233</td>
<td>2003.197</td>
<td>2007</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-25</td>
<td>-</td>
<td>92</td>
<td>2004.478</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-25</td>
<td>2</td>
<td>162</td>
<td>2004.784</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-25</td>
<td>-</td>
<td>102</td>
<td>2004.814</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-25</td>
<td>-12</td>
<td>15</td>
<td>2005.4</td>
<td>2006</td>
</tr>
</tbody>
</table>
Table 24. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 12

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 30. Sequence of maps of Pemafrost Area 12 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 31. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space.
Figure 32. Projection of the validation data points onto the precipitable water vs. temperature plane of state space.
Figure 33. Projection of the validation data points onto the temperature vs. vegetation plane of state space
**Permafrost Area 13**

This section describes the analysis of Permafrost Area 13, which is Alaska's mountainous area underlain by isolated masses of permafrost (Ferrians, 1998). It is associated with the Chugach Range and the southwestern portion of Alaska Range geographic regions (National Geographic Society, 2002), the Alaska / St. Elias Range Tundra ecoregion (Ricketts, et al., 1999), and it borders the Copper River Basin and Cook Inlet climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 13 is shown in Figure 34. Seven subset are evident.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 25 and 14. Subset 1 contains the origin (indicated by dashes in each dimension). Also, the subsets with greater cloudiness tended to occur earlier in the study period and that the one with less cloudiness occurred later.

Table 27 shows the distributions of the data points by subset and year. The spatial aspect of the cloudiness analysis is shown in Figure 35. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Figure 34. Tree model produced by the cloudiness analysis of Permafrost Area 13.
### Table 25. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 13

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-32</td>
<td>202</td>
<td>2003.337</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1603</td>
<td>2001.596</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>1671</td>
<td>2000.575</td>
<td>2006</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>51</td>
<td>2</td>
<td>18</td>
<td>2000.556</td>
<td>2000</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>386</td>
<td>1999.168</td>
<td>1998</td>
</tr>
</tbody>
</table>

### Table 26. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 13

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>386</td>
<td>1999.168</td>
<td>1998</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>51</td>
<td>2</td>
<td>18</td>
<td>2000.556</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>1671</td>
<td>2000.575</td>
<td>2006</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1603</td>
<td>2001.596</td>
<td>2004</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-32</td>
<td>202</td>
<td>2003.337</td>
<td>2004</td>
</tr>
</tbody>
</table>

### Table 27. Number of observations within subsets by collection year for Permafrost Area 13

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2 0 6 7 0 4 12 21 16 84 33 6 11</td>
</tr>
<tr>
<td>1</td>
<td>74 114 138 67 108 75 162 138 132 178 173 79 165</td>
</tr>
<tr>
<td>2</td>
<td>0 0 0 2 2 5 0 0 0 0 0 1 1</td>
</tr>
<tr>
<td>3</td>
<td>166 149 144 130 153 134 111 127 139 37 94 177 110</td>
</tr>
<tr>
<td>4</td>
<td>1 0 0 3 0 1 0 1 0 0 0 1 2</td>
</tr>
<tr>
<td>5</td>
<td>0 0 0 2 0 13 1 0 0 0 0 1 1</td>
</tr>
<tr>
<td>6</td>
<td>57 37 12 89 37 68 14 13 13 1 0 35 10</td>
</tr>
</tbody>
</table>
Figure 35 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 13 (validation set)
Permafrost Area 13, Cloudiness Analysis Subset 1

Subset 1, from the first stage analysis of Permafrost Area 13, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 36. Twelve subsets were found.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 28 and 17. The subset containing the point of no change from study period mean was Subset 10. Two earlier subsets showed a decrease from study period mean temperature, while two later subsets showed an increase from study period mean temperature. Subset 11 exhibited an increase, from study period mean, in precipitable water. Its maximum frequency was observed in 2003. Decreases from study period mean precipitable water were observed in the subsets with later mean collection years. Subsets 0, 1, 2, and 3, each containing a small number of observations, displayed an NDVI decrease that may indicate a loss of vegetation or a replacement of vegetation by another substance. Differences in all three variables were observed in Subset 9.
Table 30 shows the distributions of the data points by subset and year. The geographic progression of conditions through Permafrost Area 13 is shown in Figure 12, using an RGB color scale. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 38, precipitation vs. temperature in Figure 39, and temperature vs. vegetation in Figure 40. One may observe the locations of the points exhibiting much lower NDVI conditions (Subsets 0, 1, 2, and 3). One sees a marked change in conditions between 2003 and 2004, but also a possible return to previous conditions in 2007.
Figure 36. Tree model resulting from the second stage analysis of Cloudiness Subset 1 of Permafrost Area 13
Table 28. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 13

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1</td>
<td>-</td>
<td>-122</td>
<td>4</td>
<td>2001.25</td>
<td>2001</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-52</td>
<td>-122</td>
<td>6</td>
<td>2000.667</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-14</td>
<td>-122</td>
<td>27</td>
<td>2001.333</td>
<td>2001</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-122</td>
<td>7</td>
<td>2001.143</td>
<td>2002</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>145</td>
<td>2002.083</td>
<td>2001</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-24</td>
<td>-24</td>
<td>38</td>
<td>2003.868</td>
<td>2005</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-38</td>
<td>-</td>
<td>197</td>
<td>2004.924</td>
<td>2005</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-24</td>
<td>-</td>
<td>100</td>
<td>2003.47</td>
<td>2004</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>-24</td>
<td>-</td>
<td>33</td>
<td>2004.273</td>
<td>2004</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-24</td>
<td>12</td>
<td>61</td>
<td>2004.639</td>
<td>2005</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>419</td>
<td>2000.721</td>
<td>2001</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>564</td>
<td>2000.032</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 29. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 13

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>564</td>
<td>2000.032</td>
<td>2003</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-52</td>
<td>-122</td>
<td>6</td>
<td>2000.667</td>
<td>2001</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-52</td>
<td>-</td>
<td>419</td>
<td>2000.721</td>
<td>2001</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-122</td>
<td>7</td>
<td>2001.143</td>
<td>2002</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>-</td>
<td>-122</td>
<td>4</td>
<td>2001.25</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-14</td>
<td>-122</td>
<td>27</td>
<td>2001.333</td>
<td>2001</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>145</td>
<td>2002.083</td>
<td>2001</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-24</td>
<td>-</td>
<td>100</td>
<td>2003.47</td>
<td>2004</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-24</td>
<td>-24</td>
<td>38</td>
<td>2003.868</td>
<td>2005</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>-24</td>
<td>-</td>
<td>33</td>
<td>2004.273</td>
<td>2004</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-24</td>
<td>12</td>
<td>61</td>
<td>2004.639</td>
<td>2005</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-38</td>
<td>-</td>
<td>197</td>
<td>2004.924</td>
<td>2005</td>
</tr>
</tbody>
</table>
Table 30. number of observations within subsets by collection year, for cloudiness subset 1 of permafrost area 13

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>26</td>
<td>27</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>57</td>
<td>71</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>30</td>
<td>13</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>6</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>25</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>65</td>
<td>35</td>
<td>17</td>
<td>34</td>
<td>17</td>
<td>77</td>
<td>29</td>
<td>2</td>
<td>42</td>
<td>39</td>
<td>11</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>47</td>
<td>29</td>
<td>96</td>
<td>37</td>
<td>58</td>
<td>27</td>
<td>23</td>
<td>86</td>
<td>123</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>21</td>
</tr>
</tbody>
</table>
Figure 37. Sequence of maps of Permafrost Area 13 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 38. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space
Figure 39. Projection of the validation data points onto the precipitable water vs. temperature plane of state space
Figure 40. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
Permafrost Area 22

This section describes the analysis of Permafrost Area 22, which is Alaska's lowland and upland area underlain by moderately thick to thin permafrost (Ferrians, 1998). It is associated with the Seward Peninsula and Yukon River Delta geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, Interior Alaska / Yukon Lowland Taiga, and Copper Plateau Taiga ecoregions (Ricketts, et al., 1999); and the West Central and Interior climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 22 is shown in Figure 41. Five subsets were identified.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 31 and 20. Subset 1 includes the origin. Subsets with greater cloudiness were observed in the early portion of the study period.

Table 33 shows the distributions of the data points by subset and year. The spatial component of the cloudiness analysis is shown in Figure 42. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Area 22, Cloudiness Analysis

- Leaf 0: 
  - diff btn % late cloudy images and % early images 1995-2007 < 36
  - diff btn % early cloudy images and % early 1995-2007 < 6
  - diff btn % seas cloudy images and % seas 1995-2007 < -28

- Leaf 1: 
- Leaf 2: 
- Leaf 3: 
- Leaf 4: 
  - diff btn % late cloudy images and % late 1995-2007 < -45
Table 31. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 22

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2445</td>
<td>2001.157</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>6</td>
<td>-45</td>
<td>11</td>
<td>2002.727</td>
<td>2001</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>1191</td>
<td>2001.024</td>
<td>1999</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>36</td>
<td>234</td>
<td>1998.987</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 32. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 22

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>36</td>
<td>234</td>
<td>1998.987</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>1191</td>
<td>2001.024</td>
<td>1999</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2445</td>
<td>2001.157</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>6</td>
<td>-45</td>
<td>11</td>
<td>2002.727</td>
<td>2001</td>
</tr>
</tbody>
</table>

Table 33. Number of observations within subsets by collection year from the validation set of Permafrost Area 22

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>245</td>
<td>225</td>
<td>198</td>
<td>122</td>
<td>128</td>
<td>66</td>
<td>183</td>
<td>169</td>
<td>200</td>
<td>279</td>
<td>266</td>
<td>196</td>
<td>168</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>68</td>
<td>33</td>
<td>144</td>
<td>170</td>
<td>155</td>
<td>111</td>
<td>121</td>
<td>88</td>
<td>18</td>
<td>24</td>
<td>93</td>
<td>126</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>4</td>
<td>69</td>
<td>34</td>
<td>2</td>
<td>79</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 42 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 22
**Permafrost Area 22, Cloudiness Analysis Subset 1**

Subset 1, from the first stage analysis of Permafrost Area 22, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 43. Seventeen subsets were found.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 34 and 23. The subset containing the point of no change from study period mean values, Subset 7, had the earliest mean collection year. All subsets with an increase from study period mean PWI exhibited their maximum frequencies in 2003, and most of these subsets also exhibited increases from study period mean NDVI. Subsets with later mean collection years (greater than 2003.0) exhibited lower than study period mean PWI values, occasionally accompanied by higher than study period mean LST values.

Table 36 shows the distributions of the data points by subset and year. The geographic progression of conditions within Permafrost Area 22 is shown in Figure 13, using an RGB color scale. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 45, precipitation vs. temperature in Figure 46, and temperature vs. vegetation in Figure 47. Changes in conditions are apparent throughout the study period. Some examples of notable changes are those that occurred 1) between 2002 and 2003, 2) between 2003 and 2004, and 3) between 2006 and 2007.
Figure 43. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 22
Table 34. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-</td>
<td>-35</td>
<td>-</td>
<td>72</td>
<td>2004.514</td>
<td>2005</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-20</td>
<td>-</td>
<td>360</td>
<td>2003.267</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-28</td>
<td>-</td>
<td>53</td>
<td>2004.434</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-20</td>
<td>-</td>
<td>38</td>
<td>2004.553</td>
<td>2004</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-4</td>
<td>-</td>
<td>475</td>
<td>2000.495</td>
<td>1996</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>788</td>
<td>1998.853</td>
<td>1997</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>168</td>
<td>2000.262</td>
<td>1999</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>22</td>
<td>7</td>
<td>20</td>
<td>2001.3</td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>22</td>
<td>11</td>
<td>8</td>
<td>2002.25</td>
<td>2003</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>32</td>
<td>7</td>
<td>23</td>
<td>2002.391</td>
<td>2003</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>21</td>
<td>13</td>
<td>47</td>
<td>2002.702</td>
<td>2003</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>57</td>
<td>-</td>
<td>13</td>
<td>2002.923</td>
<td>2003</td>
</tr>
</tbody>
</table>
Table 35. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 22

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>788</td>
<td>1998.853</td>
<td>1997</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>168</td>
<td>2000.262</td>
<td>1999</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-4</td>
<td>-</td>
<td>475</td>
<td>2000.495</td>
<td>1996</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>22</td>
<td>7</td>
<td>20</td>
<td>2001.3</td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>22</td>
<td>11</td>
<td>8</td>
<td>2002.25</td>
<td>2003</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>32</td>
<td>7</td>
<td>23</td>
<td>2002.391</td>
<td>2003</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>21</td>
<td>13</td>
<td>47</td>
<td>2002.702</td>
<td>2003</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>57</td>
<td>-</td>
<td>13</td>
<td>2002.923</td>
<td>2003</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-20</td>
<td>-</td>
<td>360</td>
<td>2003.267</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-28</td>
<td>-</td>
<td>53</td>
<td>2004.434</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-35</td>
<td>-</td>
<td>72</td>
<td>2004.514</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-20</td>
<td>-</td>
<td>38</td>
<td>2004.553</td>
<td>2004</td>
</tr>
</tbody>
</table>
Table 36. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 22

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>97</td>
<td>32</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>26</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>53</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>95</td>
<td>77</td>
<td>83</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>23</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>99</td>
<td>19</td>
<td>13</td>
<td>15</td>
<td>31</td>
<td>56</td>
<td>10</td>
<td>0</td>
<td>47</td>
<td>19</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>140</td>
<td>42</td>
<td>161</td>
<td>81</td>
<td>57</td>
<td>30</td>
<td>76</td>
<td>127</td>
<td>36</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>8</td>
<td>4</td>
<td>16</td>
<td>37</td>
<td>0</td>
<td>21</td>
<td>19</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>9</td>
<td>13</td>
<td>0</td>
<td>17</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Figure 44. Sequence of maps of Permafrost Area 22 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 45. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space.
Figure 46. Projection of the validation data points onto the precipitable water vs. temperature plane of state space
Figure 47. Projection of the validation data points onto the temperature vs. vegetation plane of state space
Permafrost Area 23

This section describes the analysis of Permafrost Area 23, which is Alaska's lowland and upland area underlain by discontinuous permafrost (Ferrians, 1998). It is associated with the Yukon Flats geographic region (National Geographic Society, 2002); the Interior Alaska / Yukon Lowland Taiga ecoregion (Ricketts, et al., 1999); and the Interior climate region (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 23 is shown in Figure 48. Four subsets were found. The seasonal cloudiness variable was not included in this model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 37 and 26. Subset 0, which includes the origin, had approximately 86% of the observations. The small early subsets exhibited an increase in late season cloudiness, while the later subset exhibited an increase in early season cloudiness.

Table 39 shows the distribution of collection year by subset. The spatial component of the cloudiness analysis is illustrated in Figure 49. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Figure 48. Tree model produced by the cloudiness analysis of Permafrost Area 23, Cloudiness Analysis

<table>
<thead>
<tr>
<th>Leaf 0</th>
<th>Leaf 1</th>
<th>Leaf 2</th>
<th>Leaf 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>diff btn % late cloudy images and % late 1995-2007 &lt; 53</td>
<td>diff btn % late cloudy images and % late 1995-2007 &lt; 64</td>
<td>diff btn % early cloudy images and % early 1995-2007 &lt; 16</td>
</tr>
</tbody>
</table>

1995-2007 < 53
1995-2007 < 64
1995-2007 < 16
Table 37. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 23

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3372</td>
<td>2000.998</td>
<td>1995</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>503</td>
<td>2001.05</td>
<td>2000</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>53</td>
<td>20</td>
<td>2000.25</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>5</td>
<td>2000.4</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 38. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 23

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>53</td>
<td>20</td>
<td>2000.25</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>5</td>
<td>2000.4</td>
<td>2000</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3372</td>
<td>2000.998</td>
<td>1995</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>503</td>
<td>2001.05</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 39. Number of observations within subsets by collection year from validation set of Permafrost Area 23

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>294 275 288 218 271 186 267 238 251 242 290 272 280</td>
</tr>
<tr>
<td>1</td>
<td>6 25 12 80 25 100 33 62 44 58 10 28 20</td>
</tr>
<tr>
<td>2</td>
<td>0 0 0 2 3 11 0 0 4 0 0 0 0</td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 0 1 3 0 0 1 0 0 0 0</td>
</tr>
</tbody>
</table>
Figure 49 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 23 (Validation Set)
Permafrost Area 23, Cloudiness Analysis Subset 0

Subset 0, from the first stage analysis of Permafrost Area 23, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 50. Twenty four subsets were delineated.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 40 and 29. The subset containing the point of no change from study period mean values, Subset 6, had the earliest mean collection year. The subsets that exhibited their maximum frequency in 2003 were characterized by at least one of the following conditions: greater than study period mean PWI values, cooler then study period mean LST values, and/or greater then study period mean NDVI values. Subsets with mean collection years greater than 2003.0 exhibited greater than study period mean LST and/or lower than study period mean PWI. Notably, Subsets 20, 22, and 23 exhibited maximum frequencies in 2007 and greater than study period mean LST.

Table 42 shows the details of the distributions of the data points by subset and year. The geographic progression of conditions within Permafrost Area 23 is shown in Figure 14, using an RGB color scale. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 52, precipitation vs. temperature in Figure 53, and temperature vs. vegetation in Figure 54. While conditions are seen to have varied throughout the study period, the most marked change seems to have occurred between 2003 and 2004.
Area 23, Cloudiness Analysis Subset 0

Figure S0. Tree model resulting from the second stage analysis of cloudiness analysis subset 0 of Permafrost Area 23.
Table 40. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 0 of Permafrost Area

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-31</td>
<td>-</td>
<td>212</td>
<td>2004.514</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-31</td>
<td>3</td>
<td>4</td>
<td>2003.25</td>
<td>2006</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-31</td>
<td>3</td>
<td>107</td>
<td>2004.869</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-23</td>
<td>-5</td>
<td>27</td>
<td>2004.778</td>
<td>2005</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-10</td>
<td>-5</td>
<td>71</td>
<td>2003.465</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-10</td>
<td>-</td>
<td>269</td>
<td>1999.963</td>
<td>1996</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>737</td>
<td>1998.198</td>
<td>1996</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>1999.569</td>
<td>2001</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>11</td>
<td>-</td>
<td>452</td>
<td>1999.971</td>
<td>2002</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>71</td>
<td>2001.761</td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>11</td>
<td>7</td>
<td>35</td>
<td>2002.6</td>
<td>2003</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>33</td>
<td>3</td>
<td>12</td>
<td>2002.417</td>
<td>2003</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>11</td>
<td>10</td>
<td>25</td>
<td>2002.4</td>
<td>2003</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>11</td>
<td>3</td>
<td>226</td>
<td>2000.611</td>
<td>1999</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>-26</td>
<td>-</td>
<td>109</td>
<td>2004.606</td>
<td>2005</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>-15</td>
<td>-2</td>
<td>50</td>
<td>2003.58</td>
<td>2005</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>-2</td>
<td>40</td>
<td>1999.45</td>
<td>1995</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>128</td>
<td>2001.844</td>
<td>2006</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>272</td>
<td>2004.327</td>
<td>2007</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>162</td>
<td>1999.833</td>
<td>1999</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>36</td>
<td>2006.5</td>
<td>2007</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>-</td>
<td>12</td>
<td>7</td>
<td>2005.429</td>
<td>2007</td>
</tr>
<tr>
<td>Subset</td>
<td>LST Boundary</td>
<td>PWI Boundary</td>
<td>NDVI Boundary</td>
<td>Number of Observations in Subset (Validation Set)</td>
<td>Mean Collection Year</td>
<td>Collection Year with Maximum Frequency</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------------------------------------------</td>
<td>----------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>737</td>
<td>1998.198</td>
<td>1996</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>-2</td>
<td>40</td>
<td>1999.45</td>
<td>1995</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>209</td>
<td>1999.569</td>
<td>2001</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>162</td>
<td>1999.833</td>
<td>1999</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-10</td>
<td>-</td>
<td>269</td>
<td>1999.963</td>
<td>1996</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>11</td>
<td>-</td>
<td>452</td>
<td>1999.971</td>
<td>2002</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>11</td>
<td>3</td>
<td>226</td>
<td>2000.611</td>
<td>1999</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>71</td>
<td>2001.761</td>
<td>2003</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>128</td>
<td>2001.844</td>
<td>2006</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>11</td>
<td>10</td>
<td>25</td>
<td>2002.4</td>
<td>2003</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>33</td>
<td>3</td>
<td>12</td>
<td>2002.417</td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>11</td>
<td>7</td>
<td>35</td>
<td>2002.6</td>
<td>2003</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-31</td>
<td>3</td>
<td>4</td>
<td>2003.25</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-10</td>
<td>-5</td>
<td>71</td>
<td>2003.465</td>
<td>2005</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>-15</td>
<td>-2</td>
<td>50</td>
<td>2003.58</td>
<td>2005</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>272</td>
<td>2004.327</td>
<td>2007</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-31</td>
<td>-</td>
<td>212</td>
<td>2004.514</td>
<td>2004</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>-26</td>
<td>-</td>
<td>109</td>
<td>2004.606</td>
<td>2005</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-23</td>
<td>-5</td>
<td>27</td>
<td>2004.778</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-31</td>
<td>3</td>
<td>107</td>
<td>2004.869</td>
<td>2004</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>-</td>
<td>12</td>
<td>7</td>
<td>2005.429</td>
<td>2007</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>36</td>
<td>2006.5</td>
<td>2007</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>69</td>
<td>27</td>
<td>8</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>93</td>
<td>157</td>
<td>137</td>
<td>84</td>
<td>41</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>6</td>
<td>24</td>
<td>15</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>22</td>
<td>44</td>
<td>82</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>74</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>39</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>29</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 51. Sequence of maps of Permafrost Area 23 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 0)
Figure 52. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space
Figure 53. Projection of the validation data points onto the precipitable water vs. temperature plane of state space
Figure 54. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
Permafrost Area 24

This section describes the analysis of Permafrost Area 24, which is Alaska's lowland and upland area underlain by numerous isolated masses of permafrost (Ferrians, 1998). It is associated with the Tanana Valley geographic region (National Geographic Society, 2002); the Interior Alaska / Yukon Lowland Taiga and Copper Plateau Taiga ecoregions (Ricketts, et al., 1999); and the Interior and West Central climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 24 is shown in Figure 55. Six subsets were found. The seasonal cloudiness variable was not used in this model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 43 and 32. Subset 1 includes the origin. Two subsets, Subsets 0 and 3, contained more observations than Subset 1: Subset 3 had an increase in late season cloudiness and Subset 0 had a decrease in late season cloudiness.

Table 45 shows the distribution of collection year for each subset. The spatial aspect of the cloudiness analysis is illustrated in Figure 56. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Figure 55. Tree model produced by the cloudiness analysis of Permafrost Area 24

Area 24, Cloudiness Analysis

- Leaf 0: diff btn % late cloudy images and % late 1995-2007 < -10
- Leaf 1: diff btn % late cloudy images and % late 1995-2007 < 4
- Leaf 3: diff btn % early cloudy images and % early 1995-2007 < -45
- Leaf 4: diff btn % late cloudy images and % late 1995-2007 < -28
- Leaf 5: diff btn % early cloudy images and % early 1995-2007 < 45
Table 43. Subset boundaries and collection year subset, sorted by subset, for Permafrost Area 24

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Later Season Boundary</th>
<th>Number of Observations in Subset</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-10</td>
<td>1172</td>
<td>2001.864</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1014</td>
<td>2001.147</td>
<td>2005</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>45</td>
<td>28</td>
<td>1996.786</td>
<td>1995</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>17</td>
<td>-</td>
<td>541</td>
<td>2000.769</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 44. Subset boundaries and collection year subset, sorted by mean collection year, for Permafrost Area 24

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Later Season Boundary</th>
<th>Number of Observations in Subset</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>45</td>
<td>28</td>
<td>1996.786</td>
<td>1995</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>45</td>
<td>28</td>
<td>1996.786</td>
<td>1995</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>17</td>
<td>-</td>
<td>541</td>
<td>2000.769</td>
<td>2000</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-10</td>
<td>1172</td>
<td>2001.864</td>
<td>2004</td>
</tr>
</tbody>
</table>

Table 45. Number of observations within subsets by collection year from the validation set of Permafrost Area 24

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58 67 80 28 98 38 156 56 110 163 122 95 101</td>
</tr>
<tr>
<td>1</td>
<td>99 78 88 60 95 19 101 30 68 82 106 90 98</td>
</tr>
<tr>
<td>2</td>
<td>7 1 3 0 0 0 0 0 0 0 0 1 1 0</td>
</tr>
<tr>
<td>3</td>
<td>120 146 105 129 55 35 27 158 102 39 69 104 43</td>
</tr>
<tr>
<td>4</td>
<td>12 0 5 9 0 1 0 0 1 0 0 0 0</td>
</tr>
<tr>
<td>5</td>
<td>4 8 19 74 52 207 16 56 19 16 2 10 58</td>
</tr>
</tbody>
</table>
Figure 56 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 24 (Validation Set)
Permafrost Area 24, Cloudiness Analysis Subset 1

Subset 1, from the first stage analysis of Permafrost Area 24, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 57. Seven subsets were delineated. Two variables, PWI and NDVI, were used by the model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 46 and 35. The subset containing the point of no change from study period mean values, Subset 3, had the earliest mean collection year. Increases from study period mean NDVI occurred along with increases from study period mean PWI in subsets with a maximum collection year of 2003. The two latest subsets, according to mean collection year, exhibited the greatest decreases from study period mean PWI.

Table 48 presents the distributions of observations by subset and collection year. The geographic progression of conditions through Permafrost Area 24 is shown in Figure 15, using an RGB color scale. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 59, precipitation vs. temperature in Figure 60, and temperature vs. vegetation in Figure 61. A marked change in conditions occurred between 2003 and 2004.
Figure 57. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 24, Cloudiness Analysis Subset 1

Area 24, Cloudiness Analysis Subset 1

- %diff PWI, 1995-2007, mean < -30
- %diff PWI, 1995-2007, mean < -15
- %diff PWI, 1995-2007, mean < -6
- %diff PWI, 1995-2007, mean < 18
- %diff PWI, 1995-2007, mean < 6
- %diff PWI, 1995-2007, mean < 13
- %diff NDVI, 1995-2007, mean < 13

1995-2007
Table 46. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area 24

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-30</td>
<td>-</td>
<td>141</td>
<td>2005.298</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-6</td>
<td>-</td>
<td>139</td>
<td>2001.626</td>
<td>2007</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>387</td>
<td>1998.416</td>
<td>1995</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>18</td>
<td>6</td>
<td>42</td>
<td>2002.381</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 47. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 24

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>387</td>
<td>1998.416</td>
<td>1995</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-6</td>
<td>-</td>
<td>139</td>
<td>2001.626</td>
<td>2007</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>18</td>
<td>6</td>
<td>42</td>
<td>2002.381</td>
<td>2003</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-30</td>
<td>-</td>
<td>141</td>
<td>2005.298</td>
<td>2005</td>
</tr>
</tbody>
</table>

Table 48. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of permafrost Area 24

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>55</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>43</td>
<td>42</td>
<td>33</td>
<td>44</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>23</td>
<td>6</td>
<td>0</td>
<td>22</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>7</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>33</td>
<td>65</td>
<td>37</td>
<td>66</td>
<td>8</td>
<td>70</td>
<td>13</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>1</td>
<td>16</td>
<td>22</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 58. Sequence of maps of Permafrost Area 24 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 59. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space
Figure 60. Projection of the validation data points onto the precipitable water vs. temperature plane of state space.
Figure 61. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
Permafrost Area 25

This section describes the analysis of Permafrost Area 25, which is Alaska’s lowland and upland area underlain by isolated masses of permafrost (Ferrians, 1998). It is associated with the Bristol Bay and Cook Inlet geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, and Cook Inlet Taiga ecoregions (Ricketts, et al., 1999); and the Bristol Bay and Cook Inlet climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 25 is shown in Figure 62. Five subsets were found. Seasonal cloudiness was not included in this model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 49 and 38. Subset 1 includes the origin, but it is not the largest subset in terms of number of observations. Subset 0 contains the most observations, had a lower than study period mean proportion of late season cloudy images, and it particularly dominated the 2004 collection year. This will have an impact upon the second stage analysis, as few observations from 2004 will be in it.

Table 51 shows the distributions of collection year by subset. The distribution of collection year in Subset 4, which exhibits an increase in early season cloudiness, suggests that these conditions may occur cyclically. The spatio-temporal aspect of the cloudiness analysis is illustrated in Figure 63. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Area 25, Cloudiness Analysis

diff btn % early cloudy
images and % early
1995-2007 < 31

Leaf 0

diff btn % late cloudy
images and % late
1995-2007 < -3

Leaf 1

diff btn % early cloudy
images and % early
1995-2007 < 27

Leaf 2

Leaf 3

Leaf 4
Table 49. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 25

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-3</td>
<td>1836</td>
<td>2001.385</td>
<td>1996</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1225</td>
<td>2000.637</td>
<td>1995</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-30</td>
<td>27</td>
<td>14</td>
<td>1996.429</td>
<td>1997</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-30</td>
<td>27</td>
<td>411</td>
<td>1999.039</td>
<td>1998</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>31</td>
<td>-</td>
<td>414</td>
<td>2002.471</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 50. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 25

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-</td>
<td>-30</td>
<td>27</td>
<td>14</td>
<td>1996.429</td>
<td>1997</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-30</td>
<td>27</td>
<td>411</td>
<td>1999.039</td>
<td>1998</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-30</td>
<td>-</td>
<td>1225</td>
<td>2000.637</td>
<td>1995</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-30</td>
<td>-3</td>
<td>1836</td>
<td>2001.385</td>
<td>1996</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>31</td>
<td>-</td>
<td>414</td>
<td>2002.471</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 51. Number of observations within subsets by collection year from validation set of Permafrost Area 25

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>63</td>
<td>262</td>
<td>92</td>
<td>44</td>
<td>97</td>
<td>41</td>
<td>196</td>
<td>215</td>
<td>195</td>
<td>249</td>
<td>174</td>
<td>167</td>
<td>41</td>
</tr>
<tr>
<td>1</td>
<td>158</td>
<td>32</td>
<td>142</td>
<td>121</td>
<td>94</td>
<td>76</td>
<td>96</td>
<td>67</td>
<td>95</td>
<td>37</td>
<td>125</td>
<td>102</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>5</td>
<td>54</td>
<td>101</td>
<td>39</td>
<td>84</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>34</td>
<td>70</td>
<td>99</td>
<td>4</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>158</td>
</tr>
</tbody>
</table>
Figure 63 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 25 (validation set)
**Permafrost Area 25, Cloudiness Analysis Subset 1**

Subset 1, from the first stage analysis of Permafrost Area 25, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 64. Thirteen subsets were differentiated.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 52 and 41. The subset containing the point of no change from study period mean values (Subset 9) had the earliest mean collection year. Higher than study period mean PWI and NDVI were seen in Subset 11 and Subset 12. These subsets exhibited maximum frequencies in 2003. A few subsets with lower than study period mean land surface temperature were found. All subsets with mean collection years greater than 2003.0 had lower than study period mean PWI.

Table 54 shows the distributions of the response variable (collection year) for each subset. Periodicity may be visible in the distribution of Subset 6, which had lower than study period mean temperature and precipitable water. The geographic progression of conditions through Permafrost Area 25 is shown, using an RGB color scale, in Figure 16. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 66, precipitation vs. temperature in Figure 67, and temperature vs. vegetation in Figure 68.
Figure 64. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 25

Area 25, Cloudiness Analysis Subset 1

%diff PWI, 1995-2007, mean < -37

%diff NDVI, 1995-2007, mean < -15

%diff LST, 1995-2007, mean < 0

%diff PWI, 1995-2007, mean < -26

%diff NDVI, 1995-2007, mean < -16

%diff LST, 1995-2007, mean < 0

%diff PWI, 1995-2007, mean < -16

%diff PWI, 1995-2007, mean < -2

%diff PWI, 1995-2007, mean < 0

%diff PWI, 1995-2007, mean < 1

%diff PWI, 1995-2007, mean < 20

%diff PWI, 1995-2007, mean < 50
Table 52. Subset boundaries and collection year summary, sorted by subset, for Cloudiness Analysis Subset 1 of Permafrost Area

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-26</td>
<td>-15</td>
<td>7</td>
<td>2002.286</td>
<td>2000</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-26</td>
<td>0</td>
<td>5</td>
<td>2004.2</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-26</td>
<td>0</td>
<td>23</td>
<td>2004.304</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-26</td>
<td>-</td>
<td>61</td>
<td>2005.164</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-16</td>
<td>-</td>
<td>115</td>
<td>2004.043</td>
<td>2005</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-2</td>
<td>-</td>
<td>167</td>
<td>2001.012</td>
<td>2001</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>185</td>
<td>1999.178</td>
<td>1998</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>259</td>
<td>1997.282</td>
<td>1997</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>155</td>
<td>1999.69</td>
<td>2002</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>20</td>
<td>1</td>
<td>112</td>
<td>2001.563</td>
<td>2003</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>50</td>
<td>1</td>
<td>12</td>
<td>2003</td>
<td>2003</td>
</tr>
</tbody>
</table>
Table 53. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 25

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>259</td>
<td>1997.282</td>
<td>1997</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>185</td>
<td>1999.178</td>
<td>1998</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>155</td>
<td>1999.69</td>
<td>2002</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-2</td>
<td>-</td>
<td>167</td>
<td>2001.012</td>
<td>2001</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>20</td>
<td>1</td>
<td>112</td>
<td>2001.563</td>
<td>2003</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-26</td>
<td>-15</td>
<td>7</td>
<td>2002.286</td>
<td>2000</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>0</td>
<td>-2</td>
<td>94</td>
<td>2002.606</td>
<td>2007</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>50</td>
<td>1</td>
<td>12</td>
<td>2003</td>
<td>2003</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-16</td>
<td>-</td>
<td>115</td>
<td>2004.043</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-26</td>
<td>0</td>
<td>5</td>
<td>2004.2</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-26</td>
<td>0</td>
<td>23</td>
<td>2004.304</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-26</td>
<td>-</td>
<td>61</td>
<td>2005.164</td>
<td>2005</td>
</tr>
</tbody>
</table>

Table 54. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 1 of Permafrost Area 25

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>46</td>
<td>28</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>12</td>
<td>25</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>49</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>23</td>
<td>17</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>38</td>
<td>1</td>
<td>16</td>
<td>43</td>
<td>30</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>9</td>
<td>86</td>
<td>3</td>
<td>92</td>
<td>25</td>
<td>13</td>
<td>5</td>
<td>9</td>
<td>19</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>10</td>
<td>17</td>
<td>21</td>
<td>22</td>
<td>1</td>
<td>26</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>71</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Figure 65. Sequence of maps of Permafrost Area 25 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1).
Figure 66. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space
Figure 67. Projection of the validation data points onto the precipitable water vs. temperature plane of state space.
Figure 68. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
**Permafrost Area 26**

This section describes the analysis of Permafrost Area 26, which is Alaska's lowland and upland area generally free of permafrost (Ferrians, 1998). It, like Permafrost Area 25, is associated with the Bristol Bay and Cook Inlet geographic regions (National Geographic Society, 2002); the Beringia Lowland Tundra, Beringia Upland Tundra, and Cook Inlet Taiga ecoregions (Ricketts, et al., 1999); and the Bristol Bay and Cook Inlet climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 26 is shown in Figure 69. Five subset were identified.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 55 and 44. Subset 1 includes the origin.

Table 57 shows that the cloudiness model for Permafrost Area 26 is very similar to that of Permafrost Area 25, which is to be expected, because they cover essentially the same area. The spatio-temporal aspect of the cloudiness analysis is illustrated in Figure 70. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Figure 69. Tree model produced by the cloudiness analysis of Permafrost Area 26.
Table 55. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 26

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-9</td>
<td>1613</td>
<td>2001.399</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1425</td>
<td>2000.538</td>
<td>1995</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>27</td>
<td>387</td>
<td>1999.455</td>
<td>1998</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>1999.154</td>
<td>1998</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>28</td>
<td>-</td>
<td>449</td>
<td>2002.472</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 56. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 26

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>1999.154</td>
<td>1998</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>27</td>
<td>387</td>
<td>1999.455</td>
<td>1998</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1425</td>
<td>2000.538</td>
<td>1995</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-9</td>
<td>1613</td>
<td>2001.399</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>28</td>
<td>-</td>
<td>449</td>
<td>2002.472</td>
<td>2007</td>
</tr>
</tbody>
</table>

Table 57. Number of observations within subsets by collection year from the validation set of Permafrost Area 26

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>1</td>
<td>183</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>
Figure 70 Spatial and temporal progression of the cloudiness conditions of Permafrost Area 26 (validation set)
Permafrost Area 26, Cloudiness Analysis Subset 1

Subset 1, from the first stage analysis of Permafrost Area 26, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 71. Twenty two subsets were derived.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 58 and 47. The subset containing the origin was Subset 16. Subset 11, with a decrease from study period mean LST, had its maximum observed frequency in 2007. Subsets with an increase from study period mean LST had few observations and were seen throughout the study period (Subsets 14, 15, 20, and 3). NDVI and PWI varied in the early part of the study period, but the subsets with maximum frequencies observed in 2003 exhibited increases from study period mean NDVI and PWI. Subsets with mean collection years greater than 2003.0 exhibited decreases from study period mean PWI. Some of these also exhibited increases from study period mean NDVI.

Table 60 shows the distribution of the observations by subset and year. Unimodal and multimodal distributions, along with apparently periodic behavior (of Subsets 7 and 11), may be observed. The geographic progression of conditions through Permafrost Area 26 is shown, using an RGB color scale, in Figure 17. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 73, precipitation vs. temperature in Figure 74, and temperature vs. vegetation in Figure 75.
Figure 71. Tree model resulting from the second stage analysis of Cloudiness Analysis Subset 1 of Permafrost Area 26

Area 26, Cloudiness Analysis Subset 1

%diff PWI, 1995-2007
mean < -27

%diff NDVI, 1995-2007
mean < -36

%diff PWI, 1995-2007
mean < -14

%diff NDVI, 1995-2007
mean < -21

%diff NDVI, 1995-2007
mean < -13

%diff PWI, 1995-2007
mean < -25

%diff NDVI, 1995-2007
mean < -21

%diff PWI, 1995-2007
mean < -19

%diff LST, 1995-2007
mean < 0

%diff LST, 1995-2007
mean < 2

%diff PWI, 1995-2007
mean < 8

%diff PWI, 1995-2007
mean < -21

%diff NDVI, 1995-2007
mean < 1

%diff PWI, 1995-2007
mean < 12

%diff PWI, 1995-2007
mean < 1

%diff NDVI, 1995-2007
mean < 5

%diff NDVI, 1995-2007
mean < 12

%diff LST, 1995-2007
mean < 1

%diff LST, 1995-2007
mean < 34

%diff PWI, 1995-2007
mean < 1

%diff PWI, 1995-2007
mean < -5

%diff LST, 1995-2007
mean < 1
<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-27</td>
<td>-14</td>
<td>10</td>
<td>2002</td>
<td>2000</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-36</td>
<td>-</td>
<td>19</td>
<td>2005.895</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-27</td>
<td>2</td>
<td>42</td>
<td>2005.048</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-27</td>
<td>12</td>
<td>27</td>
<td>2004.704</td>
<td>2005</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-13</td>
<td>-</td>
<td>146</td>
<td>2002.281</td>
<td>2006</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-25</td>
<td>8</td>
<td>13</td>
<td>2005.077</td>
<td>2005</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-19</td>
<td>8</td>
<td>38</td>
<td>2004.684</td>
<td>2005</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-13</td>
<td>8</td>
<td>25</td>
<td>2004.08</td>
<td>2005</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>5</td>
<td>-</td>
<td>151</td>
<td>1999.325</td>
<td>1998</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>-5</td>
<td>141</td>
<td>1996.645</td>
<td>1995</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>-</td>
<td>-5</td>
<td>5</td>
<td>1995.8</td>
<td>1995</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>-5</td>
<td>24</td>
<td>1997.25</td>
<td>1997</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>322</td>
<td>1998.329</td>
<td>1995</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>173</td>
<td>2000.335</td>
<td>2001</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>17</td>
<td>4</td>
<td>35</td>
<td>2000.4</td>
<td>2003</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>17</td>
<td>8</td>
<td>13</td>
<td>2002.538</td>
<td>2003</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>34</td>
<td>4</td>
<td>63</td>
<td>2002.921</td>
<td>2003</td>
</tr>
</tbody>
</table>
Table 59. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness Analysis Subset 1 of Permafrost Area 26

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1</td>
<td>-</td>
<td>-5</td>
<td>5</td>
<td>1995.8</td>
<td>1995</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>-5</td>
<td>141</td>
<td>1996.645</td>
<td>1995</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>-5</td>
<td>24</td>
<td>1997.25</td>
<td>1997</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>322</td>
<td>1998.329</td>
<td>1995</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>5</td>
<td>-</td>
<td>151</td>
<td>1999.325</td>
<td>1998</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>2000.335</td>
<td>2001</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>2000.4</td>
<td>2003</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-13</td>
<td>6</td>
<td>2001.167</td>
<td>2000</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-27</td>
<td>10</td>
<td>2002</td>
<td>2000</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-13</td>
<td>146</td>
<td>2002.281</td>
<td>2006</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>-</td>
<td>17</td>
<td>8</td>
<td>2002.406</td>
<td>2003</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>17</td>
<td>8</td>
<td>13</td>
<td>2002.538</td>
<td>2003</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>-</td>
<td>34</td>
<td>4</td>
<td>2002.921</td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-13</td>
<td>8</td>
<td>25</td>
<td>2004.08</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>-19</td>
<td>8</td>
<td>38</td>
<td>2004.684</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-27</td>
<td>12</td>
<td>27</td>
<td>2004.704</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-27</td>
<td>2</td>
<td>42</td>
<td>2005.048</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>-25</td>
<td>8</td>
<td>13</td>
<td>2005.077</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-27</td>
<td>21</td>
<td>2005.095</td>
<td>2006</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-36</td>
<td>19</td>
<td>2005.895</td>
<td>2006</td>
</tr>
</tbody>
</table>
Table 60. Number of observations within subsets by collection year, from validation set of Cloudiness Subset 1 of Permafrost Area

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>20</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>7</td>
<td>20</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>33</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>17</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>53</td>
<td>21</td>
<td>13</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 72. Sequence of maps of Permafrost Area 26 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 1)
Figure 73. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space.
Figure 74. Projection of the validation data points onto the precipitable water vs. temperature plane of state space
Figure 75. Projection of the validation data points onto the temperature vs. vegetation plane of state space.
**Permafrost Area 31**

This section describes the analysis of Permafrost Area 31, which was numbered but not named in the permafrost data set (Ferrians, 1998). With reference to its location, it could be named the Gulf of Alaska Area. It is associated with the Coastal Range, Aleutian Range, Kodiak Island, and Kenai Peninsula geographic regions (National Geographic Society, 2002); the Alaska Peninsula Montane Taiga, Northern Pacific Coastal Forests, and Pacific Coast Mountain Tundra and Ice Fields ecoregions (Ricketts, et al., 1999); and the South Central and Southeast Panhandle climate regions (Shulski & Wendler, 2007).

The tree produced by the cloudiness analysis of Permafrost Area 31 is shown in Figure 76. Five subsets were identified. Seasonal cloudiness was not used in this model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 61 and 50. Subset 3 includes the origin. Subsets exhibiting decreases in late season cloudiness (Subsets 0 and 2) had later mean collection years.

Table 63 shows the distributions of collection years for each subset. The spatial aspect of the cloudiness analysis is illustrated in Figure 77. Black shading implies the highest mean seasonal proportion of cloudy images, while white implies the lowest mean seasonal proportion of cloudy images.
Figure 76. Tree model produced by the cloudiness analysis of Permafrost Area 31.
Table 61. Subset boundaries and collection year summary, sorted by subset, for Permafrost Area 31

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-22</td>
<td>587</td>
<td>2001.964</td>
<td>2004</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-10</td>
<td>-</td>
<td>879</td>
<td>1999.811</td>
<td>1996</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-11</td>
<td>-</td>
<td>483</td>
<td>2002.337</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1594</td>
<td>2001.117</td>
<td>2000</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td>357</td>
<td>2000.008</td>
<td>1999</td>
</tr>
</tbody>
</table>

Table 62. Subset boundaries and collection year summary, sorted by mean collection year, for Permafrost Area 31

<table>
<thead>
<tr>
<th>Subset</th>
<th>Seasonal Boundary</th>
<th>Early Season Boundary</th>
<th>Late Season Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-10</td>
<td>-</td>
<td>879</td>
<td>1999.811</td>
<td>1996</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td>357</td>
<td>2000.008</td>
<td>1999</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1594</td>
<td>2001.117</td>
<td>2000</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-22</td>
<td>587</td>
<td>2001.964</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-11</td>
<td>483</td>
<td>2002.337</td>
<td>2004</td>
</tr>
</tbody>
</table>

Table 63. Number of observations within subsets by collection year for Permafrost Area 31

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>24</td>
<td>34</td>
<td>79</td>
<td>2</td>
<td>8</td>
<td>26</td>
<td>40</td>
<td>45</td>
<td>37</td>
<td>138</td>
<td>89</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>112</td>
<td>135</td>
<td>78</td>
<td>125</td>
<td>42</td>
<td>30</td>
<td>47</td>
<td>29</td>
<td>64</td>
<td>50</td>
<td>72</td>
<td>62</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>14</td>
<td>16</td>
<td>38</td>
<td>21</td>
<td>14</td>
<td>30</td>
<td>39</td>
<td>46</td>
<td>54</td>
<td>62</td>
<td>60</td>
<td>31</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>126</td>
<td>90</td>
<td>78</td>
<td>110</td>
<td>140</td>
<td>176</td>
<td>167</td>
<td>152</td>
<td>139</td>
<td>47</td>
<td>77</td>
<td>140</td>
<td>152</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>24</td>
<td>25</td>
<td>27</td>
<td>42</td>
<td>96</td>
<td>38</td>
<td>7</td>
<td>28</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>32</td>
<td>27</td>
</tr>
</tbody>
</table>
Figure 77 Spatial And Temporal Progression Of The Cloudiness Conditions of Permafrost Area 31
Permafrost Area 31, Cloudiness Analysis Subset 3

Subset 3, from the first stage analysis of Permafrost Area 31, which included the study period mean cloudiness conditions, was chosen to be analyzed in the second stage. The tree produced by this analysis is shown in Figure 78. Six subsets were derived by the model.

The subset boundaries that describe the minimum distance between each subset and the point of no difference from study period mean are shown in Table 64 and 53. The subset containing the origin was Subset 3. Subsets with earlier mean collection years, Subsets 4 and 5, exhibited an increase from study period mean PWI. Subset 4 additionally displayed a decrease from study period mean LST. Its maximum frequency was observed in 2003. Subsets 0, 1, and 2, with mean collection years that occurred later in the study period, exhibited a decrease from study period mean PWI. Subset 0 additionally displayed a decrease from study period mean NDVI.

Table 66 shows the distribution of collection years by subset. A notable example of periodicity is observed in Subset 3, which contained the origin in state space. The geographic progression of conditions through Permafrost Area 31 is shown, using RGB colors, in Figure 18. The movement of the data through state space is illustrated in three series of projections onto the planes determined by the covariates: vegetation vs. precipitation in Figure 80, precipitation vs. temperature in Figure 81, and temperature vs. vegetation in Figure 82.
Area 31, Cloudiness Analysis Subset 3

%diff PWI, 1995-2007
mean < -15

%diff NDVI, 1995-2007
mean < -25

%diff PWI, 1995-2007
mean < -37

%diff PWI, 1995-2007
mean < 3

%diff LST, 1995-2007
mean < 0

0
1
2
3
4
5
Table 64. Subset boundaries and collection year summary, sorted by subset, for Cloudiness analysis Subset 3 of Permafrost Area 31

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>349</td>
<td>2001.011</td>
<td>2001</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>-</td>
<td>472</td>
<td>2000.222</td>
<td>2003</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>315</td>
<td>1998.908</td>
<td>1995</td>
</tr>
</tbody>
</table>

Table 65. Subset boundaries and collection year summary, sorted by mean collection year, for Cloudiness analysis Subset 3 of Permafrost Area 31

<table>
<thead>
<tr>
<th>Subset</th>
<th>LST Boundary</th>
<th>PWI Boundary</th>
<th>NDVI Boundary</th>
<th>Number of Observations in Subset (Validation Set)</th>
<th>Mean Collection Year</th>
<th>Collection Year with Maximum Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>315</td>
<td>1998.908</td>
<td>1995</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>-</td>
<td>472</td>
<td>2000.222</td>
<td>2003</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>349</td>
<td>2001.011</td>
<td>2001</td>
</tr>
</tbody>
</table>

Table 66. Number of observations within subsets by collection year, from the validation set of Cloudiness Subset 3 of Permafrost Area 31

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>6</td>
<td>16</td>
<td>31</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>21</td>
<td>37</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>13</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>21</td>
<td>29</td>
<td>49</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>18</td>
<td>14</td>
<td>20</td>
<td>33</td>
<td>51</td>
<td>56</td>
<td>44</td>
<td>0</td>
<td>9</td>
<td>13</td>
<td>18</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>15</td>
<td>18</td>
<td>42</td>
<td>71</td>
<td>61</td>
<td>49</td>
<td>62</td>
<td>91</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
<td>45</td>
<td>37</td>
<td>25</td>
<td>20</td>
<td>32</td>
<td>27</td>
<td>25</td>
<td>48</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 79. Sequence of maps of Permafrost Area 31 showing the geographic progression of conditions through the study period (validation set of Cloudiness Subset 3)
Figure 80. Projection of the validation data points onto the vegetation vs. precipitable water plane of state space.
Figure 81. Projection of the validation data points onto the precipitable water vs. temperature plane of state space.
Figure 82. Projection of the validation data points onto the temperature vs. vegetation plane of state space
Model validation

Model validation was accomplished by comparing the distributions of collection year, by subset, from the training data sets with those from the validation data sets. The distributions for the training data sets are shown in Table 67 through Table 75. In general, the model fits appear to be very good, because the subsets' collection year distributions from the training and validation sets appear to be very similar.

It is more notable that the distributions of the smaller (in terms of number of data points contained) subsets, however, often more markedly differ, relative to their size, between the training and validation sets. This may indicate that these small subsets may be the parts of the models that are most prone to overfitting.
### Training Data Set Frequencies, by Subset and Year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>2</td>
<td>39</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>20</td>
<td>45</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>49</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>8</td>
<td>5</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>51</td>
<td>47</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>6</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>35</td>
<td>73</td>
<td>57</td>
<td>6</td>
<td>14</td>
<td>28</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>76</td>
<td>2</td>
<td>59</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>16</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>38</td>
<td>50</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>105</td>
<td>24</td>
<td>60</td>
<td>73</td>
<td>122</td>
<td>23</td>
<td>96</td>
<td>101</td>
<td>95</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>58</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>14</td>
<td>1</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

### Table 67. Training Data Set Frequencies by subset and year, for Area 12, Cloudiness Subset 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>46</td>
<td>14</td>
<td>19</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>37</td>
<td>49</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>36</td>
<td>11</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>5</td>
<td>36</td>
<td>13</td>
<td>6</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>46</td>
<td>37</td>
<td>29</td>
<td>54</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>54</td>
<td>80</td>
<td>97</td>
<td>73</td>
<td>37</td>
<td>24</td>
<td>84</td>
<td>66</td>
<td>48</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>39</td>
<td>19</td>
<td>41</td>
<td>7</td>
<td>42</td>
<td>1</td>
<td>41</td>
<td>20</td>
<td>19</td>
<td>6</td>
<td>6</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Table 69. Training Data Set Frequencies by subset and year, for Area 13, Cloudiness Subset 1

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>47</td>
</tr>
</tbody>
</table>

Table 70. Training Data Set Frequencies by subset and year, for Area 22, Cloudiness Subset 1

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 71. Training Data Set Frequencies by subset and year, for Area 23, Cloudiness Subset 0

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>114</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>106</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>21</td>
<td>38</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 72. Training Data Set Frequencies by subset and year, for Area 24, Cloudiness Subset 1

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 73. Training Data Set Frequencies by subset and year, for Area 25, Cloudiness Subset 1

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>
**Table 74. Training Data Set Frequencies by subset and year, for Area 26, Cloudiness Subset 1**

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>59</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>71</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 75. Training Data Set Frequencies by subset and year, for Area 31, Cloudiness Subset 3**

<table>
<thead>
<tr>
<th>Subset</th>
<th>Collection Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>
Bibliography


