MODELING THE TEMPORAL AND SPATIAL VARIABILITY OF SOLAR RADIATION

by

Randall Scott Mullen

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Ecology and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April, 2012

©COPYRIGHT

by

Randall Scott Mullen

2012

All Rights Reserved

APPROVAL

of a dissertation submitted by

Randall Scott Mullen

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr Lucy A. Marshall (Co-Chair)

Dr. Brian L. McGlynn (Co-Chair)

Approved for the Department of Land Resources and Environmental Sciences

Dr. Tracey M. Sterling

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Randall Scott Mullen

April 2012

DEDICATION

This dissertation is dedicated to my parents. Among many other valuable lessons, Mardell Mullen taught me that nothing else matters more than how you treat other people, and Robert Mullen taught me that you are never too old to learn something new, or go back to school.

ACKNOWLEDGEMENTS

I would like to thank my committee co-chairs, Brian McGlynn and Lucy Marshall, who provided constant encouragement and guidance throughout this entire process. I would also like to thank my committee members, Megan Higgs and Paul Stoy who were always willing to provide guidance and advice when needed. The National Resources Conservation Service (NRCS) funded this project. Steve Becker of the Bozeman NRCS office helped focus the project on the needs of the end users. Tim Groove in the Great Plains office and Peter Palmer in the Pacific Northwest office of the Bureau of Reclamation were extremely helpful. Several vendors provided input into the development of Chapter 1 and Appendix A, and I'd like to thank them for providing a unique perspective. They are; Sarah Ray and Sara Biddle of Independent Power Systems in Bozeman, MT and Dwight Patterson of GenPro Energy Solutions in Piedmont, SD. Daily inspiration came to me from fellow students. There are more than I can name, but certainly Jeannette Wolak, Phil Davis, Noelle Orloff and Dana Skorupa all showed me it was possible. David Hoffman, Patrick Lawrence, Melissa Bridges, Zack Miller and Tyler Smith were always willing to engage in thoughtful conversation when needed, and provide sufficient distraction when needed. Others on campus that provided a constant source of encouragement were Amy Chiuchiolo and David Berndt. Lastly, I would like to express my heart felt gratitude to Leslie Overland, without whom none of this would have been possible. Without her support and love, I don't think I would have been able to complete this dissertation.

TABLE OF CONTENTS

1.INTRODUCTION TO DISSERTATION 1

Introduction 1
Dissertation Objectives 2
Site Description 2
Data Description 3
Dissertation Organization 6
Chapter 2 7
Chapter 3 8
Chapter 4 9
Chapter 5 9
Appendix A 10
Literature Cited 11

2.USE OF INTENSITY- DURATION- FREQUENCY CURVES AND EXCEEDANCE- FREQUENCY CURVES FOR QUANTIFYING SOLAR RADIATION VARIABILITY 16

Contribution of Authors 16
Manuscript Information 17
Abstract 18
Introduction 19
Methods 23
2.1 Solar-Intensity-Duration-Frequency Curves 23
Algorithm 1: SIDF Derivation 24
Short-term Solar-Intensity-Duration-Frequency Curves 25
Algorithm 2: SSIDF Derivation 26
Exceedance-Duration Curve 27
Case Study 28
Overview of Case Study 28
Data Collation and Quality Control 29
Results 32
Solar-Intensity-Duration-Frequency (SIDF) Curves 32
ED Curves 32
Case Study Application of IDF and ED Curves 33
Extension to Spatial Solar Radiation Estimation 35
Discussion 35
Conclusion 38
Acknowledgements 39
Tables 40
Figures 41
Literature Cited 50

TABLE OF CONTENTS - CONTINUED

3.A BETA REGRESSION MODEL TO OBTAIN INTERPRETABLE PARAMETERS AND ESTIMATES OF ERROR FOR IMPROVED SOLAR RADIATION PREDICTIONS 53

Contribution of Authors and Co-Authors 53
Manuscript Information 54
Abstract 55
Introduction 56
A Review of ΔT Models for Solar Radiation Prediction 60
A Review of Beta Regression 63
Materials and Methods 67
Data and Site Description 67
Decomposing Global Solar Radiation 68
Prediction Intervals for GSR 72
Model Comparisons 73
Results and Discussion 74
Fitting the Fodor and Mika Model 74
Fitting the Beta Regression Model at the Takini Site. 75
Capture Rates for the Beta Regression Model 78
Model Comparison 79
Combining Strata for the Beta Regression Model 81
Interpolating Between Stations 81
Conclusion 82
Acknowledgements 84
Tables 85
Figures 86
Literature Cited 92

4.MODELING SOLAR RADIATION USING THE SPATIAL AUTO-CORRELATION OF THE DAILY FRACTION OF CLEAR SKY TRANSMISSIVITY 95

Contribution of Authors and Co-Authors 95
Manuscript Information 96
Abstract 97
Introduction 98
Methods 104
Data and Site Description 105
Observed Fraction of Clear Day 106
Model Comparison 107
Beta Regression Models 108

TABLE OF CONTENTS - CONTINUED

Site-Based Models 110
Daily Models 111
Universal Kriging 112
Effect of a Less Dense Monitoring Network 113
Results 114
Model comparison 114
Effect of a Less Dense Monitoring Network 115
Discussion 116
Conclusion 119
Acknowledgements 120
Tables 120
Figures 121
Literature Cited 126

5. EVALUATING A BETA REGRESSION APPROACH FOR ESTIMATING FRACTION OF CLEAR SKY TRANSMISSIVITY IN MOUNTAINOUS TERRAIN 130

Contribution of Authors and Co-Authors 130
Manuscript Information 131
Abstract 132
Introduction 133
Methods 137
Site Description 137
Data 138
Components of Global Solar Radiation (GSR) 139
Determining Observed FCD 140
Applying the Beta Regression Model 141
Using FCD to Estimate GSR 144
Results 144
Predicting at TCEF Using WSSM Station 144
Predicting at TCEF Using Porphyry Station 148
Discussion 150
Conclusion 154
Acknowledgements 155
Tables 156
Figures 158
Literature Cited 165

6. CONCLUSIONS 169

TABLE OF CONTENTS - CONTINUED

Literature Cited 177

LITERATURE CITED 179

APPENDIX A: IDF Curves, ED Curves and Statewide Maps for
Solar Radiation in the State of Montana 189

LIST OF TABLES

Table Page

2.1 Difference in kWh $d^{-1}m^{-2}$ between the complete data sets with a percentage of data missing……..…...40

3.1 Comparisons of the Fodor and Mica model and the beta regression model..... 85

3.2 Comparisons of the Fodor and Mica model and the beta regression model..... 85

4.1 RMSE, MAE and MSD for the site-based beta regression model, the daily beta regression model and universal kriging…….……120

5.1 Seasonal trends in calculated FCD values for the WSSM and Porphyry sites……..156

5.2 RMSE and MSD for the predicted GSR at TCEF……...…….156

5.3 Capture rates for each model constructed from each base site for the year……...157

5.4 Seasonal values for capture rates for each model constructed from the WSSM site data……....157

5.5 Seasonal values for capture rates for each model constructed from the Porphyry site data…….….157

LIST OF FIGURES

Figure Page

2.1 Map showing 64 sites in and around the state of Montana used for analyzing solar radiation……..……. 41

2.2 Example of a typical SIDF curve….……42

2.3 Example of a typical EDF curve……..43

2.4 A series of graphs showing monthly SIDF curves for White Sulphur Springs……..…….…44

2.5 Examples of four different SSIDF curves from four different sites in Montana……...45

2.6. Examples of four different exceedance duration frequency curves from four different sites in Montana for the month of June……...….46

2.7. A series of graphs showing monthly EDF curves for Bozeman, MT and monthly variation……..47

2.8. The relationship between the Moccasin, Montana site and the distribution of solar radiation across the state……..…….…….48

2.9. Average number of hours per day plotted against threshold value…….....48

2.10. Interpolated map of the state of Montana showing the number of hours per day exceeding 600 Watts/m^2…….…….…….49

3.1 The Montana, North Dakota, and South Dakota sites of the AWDN network…..……..……..86

3.2 Transmissivity plotted against day of year for all available years with Fourier series fitted to an envelope curve…..…..…….…….87

3.3 A histogram of the sample sizes for the 99 sites used in the analysis…….....87

3.4 The figure on the left shows data from dry summer days at Redfield, SD. For this data set, the sinusoidal curve is shown fitted to the data. On the right is data from wet winter days at the same site…….....88

LIST OF FIGURES – CONTINUED

Figure Page

3.5 A simple correlation matrix showing how FCD is correlated with the independent variables, and how the independent variable are correlated with each other……..……..89

3.6 Predicted GSR values plotted against observed GSR values with 95% prediction intervals shown……..……..……...90

3.7 Box plots showing the overall distributions of correlations between predicted GSR and observed GSR broken down in seasons and precipitation…….....90

3.8 Predicted GSR plotted against observed GSR for each of the four seasons using data from dry days…….91

4.1 Map showing Montana, North Dakota, and South Dakota sites of the AWDN network …….……..……..121

4.2 Three example variograms are shown……...122

4.3 Predicted FCD vs. observed FCD for the 1000 randomly chosen day – site combinations that were used for the leave-one-out cross-validation…….……..123

4.4 MAE for the universal kriging model is plotted against the number of sites for analysis……..…….…….…..124

4.5 RMSE for the universal kriging model is plotted against the number of sites of analysis…….……125

5.1 Map of the three study sites with reference to their location within the state of Montana. Note the change of terrain complexity between the WSSM site and the TCEF watershed. The Porphyry site is also in the mountains…..…….…..…158

5.2 A scatter plot showing the measured GSR in mega joules by day of year at the WSSM weather station…….…..…159

LIST OF FIGURES – CONTINUED

Figure Page

5.3 A visual representation of an envelope curve for modeling CST. The scatter plot points are daily sky transmissivity values, while the curve is a fitted Fourier series used to model CST…..…....159

5.4 An example of a beta distribution…….……160

5.5 A scatterplot of ΔT vs. observed FCD at the two sites used for prediction…….161

5.6 A scatter plot of predicted FCD vs. observed FCD at TCEF based on the two beta regression models fit at WSSM and Porphyry…..……162

5.7 Predicted versus observed GSR at TCEF for four different models when using the beta regression FCD model fitted at WSSM……...163

5.8 Four scatterplots show the predicted versus observed GSR at TCEF for four different models when using the beta regression model from Porphyry……..164

ABSTRACT

Solar radiation is fundamental to ecological processes and energy production. Despite growing networks of meteorological stations, the spatial and temporal variability of solar radiation remains poorly characterized. Many solar radiation models have been proposed to enhance predictions in areas without measurement instrumentation. However, these models do not fully take advantage of the increasing number of data collection sites, nor are they expandable to incorporate additional metrological information when available. In this dissertation we: 1) developed a method of statistical analysis to summarize and communicate solar radiation reliability, 2) applied a beta regression model to leverage auxiliary meteorological information for enhanced solar radiation prediction, 3) refined the beta regression model and considered spatial auto-correlation to better predict solar radiation across space, 4) extended and evaluated these methods in a mountainous region. These advancements in the characterization and prediction of solar radiation are detailed in the following chapters of this dissertation.

INTRODUCTION TO DISSERTATION

Introduction

The sun imparts approximately 3,850,000 exajoules[1] (EJ) per year into the Earth's atmosphere, oceans and land masses, accounting for 99.97% of the heat energy required for the planet's physical processes (Ogolo 2010). Incoming global solar radiation is fundamental for understanding a broad range of ecological and environmental sciences, including evapotranspiration (Hargreaves and Samani 1982, Allen, Trezza and Tasumi 2006), energy balances of snow cover (Barry et al. 1990, Jin et al. 1999, Marks et al. 1999, Marks and Winstral 2001) vegetation, (Granger and Schulze 1977, Fu and Rich 2002, Pierce, Lookingbill and Urban 2005), net primary productivity (Berterretche et al. 2005, Crabtree et al. 2009) and animal behavior (Zeng et al. 2010, Keating et al. 2007). In anthropogenic terms, the energy from one hour of incoming solar radiation exceeds a year's worth of human energy consumption for the entire planet. Given the obvious importance, it is no surprise that solar radiation was first formally modeled almost 90 years ago (Angstrom 1924). But what does this modeling entail? Angström (1924) described a way to predict the amount of solar radiation, specifically the short-wave component, based on a cloudiness index. Prescott (1940) modified Angstrom's equation, and the Angstrom – Prescott model was born. By 1984, there were at least 120 different papers offering values for the two fitted coefficients (Martínez-Lozano et al. 1984) in the Angström - Prescott model. This seemingly simple procedure of predicting short-wave solar radiation has been the focus of well over 1000 studies since 1924. However, short

[1] 1 exajoule = 10^{18} joules

and long term temporal variability, spatial patterns independent of latitudinal gradients, and summarization techniques that allow for easy access to relevant variables are all areas of ongoing research.

Dissertation Objectives

This dissertation is aimed at further understanding the temporal and spatial variability of solar radiation as well as improvement of prediction methods. The following objectives were undertaken:

1. Develop a new approach for summarizing long-term solar radiation data sets so that end users of photovoltaic technology to easily determine short-term variability;
2. Develop a new approach for predicting global solar radiation using recent developments in multiple beta regression, and compare this model to existing models;
3. Investigate the spatial auto-correlation of the beta regression model residuals, and determine if incorporating auto-correlation improves model fit to observed data; and
4. Evaluate the use of beta regression models in mountainous regions.

Site Description

Research methods were implemented across Montana, North Dakota and South Dakota. However, a few sites in Idaho and Wyoming were additionally used in order to

reduce any edge effect (bias from fewer stations due to a boundary) when interpolating data across the state.

Montana is a large state (381,154 km^2), with highly variable terrain. Elevation ranges from 550 m to nearly 3904 m above sea level. Prairies and badlands dominate the eastern half, while mountains and large intermountain valleys dominate the western half. There are 77 named ranges of the Rocky Mountains in the state of Montana. The continental divide runs north – south through Western Montana, and restricts the flow of warmer, moister air from the Pacific from reaching the eastern side of the divide. In general, east of the divide tends to be drier and cooler; a semi-arid continental climate. Average daytime temperatures vary from about -2 °C in January to about 29.2 °C in July, with extremes reaching -57 °C and 47 °C. Average annual precipitation is 380 mm, but a high amount of variability exists between the wetter, warmer air influenced by the Pacific east of the divide, and the colder drier air west of the divide.

North and South Dakota are located in the north-central United States They each have what is considered a continental climate with very cold winters and hot semi-humid summers, although the western part of North Dakota is considered semi-arid. The highest recorded temperature in either state is 49^o C and the coldest is -51^o C. The average annual precipitation ranges from 35 to 75 cm throughout the study area.

Data Description

All solar radiation models that use empirical data are subject to limitations from the accuracy of observations used for model fitting. Gueymard and Myers (2009)

described three levels of stations that collect solar radiation data. *Solar monitoring sites* use inexpensive and automated instrumentation to provide local data quickly for a minimal cost. *Conventional long-term measurements* use standard techniques and are generally operated by weather service agencies. *Research sites* are typically developed by atmospheric physicists or climatologists to obtain the highest accuracy possible in order to detect trends or test theoretical solar radiation models. Research sites have higher levels of redundancy with respect to instrumentation and power supply. Using (relatively) independent sites with high quality data (such as research sites) in order to formulate predictive equations provides a strong basis for model development and assessment. However, it is a relatively rare situation that research sites will have to estimate solar radiation, given the redundancy in equipment and power supply that these sites maintain. Much of this dissertation focuses on the more likely scenarios of solar monitoring sites needing to infill missing data (Gueymard and Myers 2009) during periods of equipment failure, replacement, calibration, or when power supply's fail, or using nearby solar radiation measurements from a solar monitoring site to predict solar radiation where no sensors exist.

Data from solar monitoring sites present challenges for solar radiation modeling including less precise measurements and potential gaps in long-term data sets. Further, instrumentation calibration requires greater labor and infrastructure costs, and is generally implemented less frequently than at research grade stations. Solar monitoring sites may then produce observations with greater uncertainty or potential bias. Despite this, networks of solar monitoring sites provide vital and valuable information. They

allow for the investigation of spatial characteristics that is simply not possible when inspection is limited to research sites. They also provide information about solar radiation trends outside of research grade sites. Thus there remains a need for flexible modeling frameworks that can be applied to all sites that collect solar radiation data.

Three solar monitoring networks were used for this dissertation. The Bureau of Reclamation (BR) operates 26 weather stations in Montana referred to as the AgriMet (Agricultural Meteorology) monitoring network (Palmer 2011). Twenty-one of these are located east of the continental divide and are operated by the Great Plains regional office. The Pacific Northwest regional office operates five stations west of the divide. An additional five sites in Idaho are used. These sites are, (as their name implies) designed to provide timely meteorological data for agricultural purposes. Therefore, they are found in valleys and plains throughout the state. Care is taken not to place any of these stations in locations with high amounts of topographical shading.

The Western Regional Climate Center (WRCC) serves data collected by the Remote Automated Weather Stations (RAWS) throughout the Western United States (Horel and Dong 2010). This network was put in place largely to monitor forest fire conditions (Reinbold, Roads and Brown 2005), thus many of the sites are located in complex terrain. Approximately 152 RAWS stations have or do exist in Montana, however, many of those were short-term stations, or were long term stations that do not exist anymore. Many of the temporary stations did not collect solar radiation data, and many of the permanent stations that do collect solar radiation data are in locations that are

affected by topographical shading. Twenty-six of these stations were used in this dissertation, 22 in Montana, three in Wyoming and one in Idaho.

High Plains Regional Climate Center (HPRCC) operates about 100 AWDN (Automated Weather Data Network) sites throughout North and South Dakota (Wu, Hubbard and You 2005). They operate about another 120 sites throughout Southern Wyoming, Western Colorado, Nebraska, Kansas and Missouri. Only the sites in North and South Dakota were used in this dissertation. Standard weather variables collected at the AWDN sites include (but are not limited to) daily high temperature, daily low temperature, relative humidity, and precipitation. The goal of these sites is very similar to the AgriMet sites, that is, they are intended to provide timely data for agricultural purposes. These sites tend to be placed in agricultural regions and away from any topographical shading.

Dissertation Organization

Chapter Two and Appendix A introduce and describe a new method of summarizing long time series of solar radiation data such that short-term variability is quantified and easily presented. Chapter Three borrows a framework from historical models that attempts to predict solar radiation based on meteorological variables. Chapter Four inspects the spatial auto-correlation of meteorological effects on final solar radiation estimates, and Chapter Five validates techniques proposed in Chapter Three in mountainous regions.

Chapter 2

Long running continuous solar radiation data sets are becoming increasingly common (Horel and Dong 2010, Reinbold et al. 2005, Palmer 2011, Wu et al. 2005). The published time unit for these data sets can vary from seconds to hours over multiple years. One station recording solar radiation values every 15 minutes for 30 years produces over a million records. These data are typically presented in its raw format, or may be summarized daily. These data often require summaries that are easy to interpret and are meaningful to end users of photovoltaic technology. To this end, monthly and yearly averages may be reported. However, these averages do not convey the short-term variability that can result in periods of low solar radiation detrimental to the daily operation of a device. Battery banks can mitigate much of the short term variability problems, but many photovoltaic applications are implemented without battery banks. For instance, the use of directly coupled photovoltaic water pump systems (DC-PVPS) is increasing (Kolhe, Joshi and Kothari 2004). These DC-PVPS pumps are often the only source of clean drinking water in remote villages throughout third world countries (Lynn 2010, Posorski 1996), and are used for bringing water to the surface for consumption by livestock in around the world (Boutelhig et al. 2012, Boutelhig, Hadjarab and A 2011). Intensity-duration-frequency (IDF) curves (Bernard 1932, Sherman 1931) were adapted in order to develop solar IDF curves, (SIDF) and short-term solar IDF curves, (SSIDF). These summarizations relay return intervals for periods of particularly high and particularly low solar radiation on the order of years (SIDF) and days (SSIDF). Chapter

Two details these adaptations, the construction of the curves, how to intepret them, and how robust they are to incomplete data sets.

Chapter 3

Prediction of solar radiation in the absence of measured solar radiation data remains a concern. Hargreaves and Samani (1982) proposed using the difference between high and low temperature instead of cloud cover data for predicting solar radiation. Bristow and Campbell (1984) formalized that argument, and a family of models known as the B&C models spawned a body of literature that is still growing today (Samani et al. 2011, Ball, Purcell and Carey 2004, Thornton and Running 1999, Thornton, Hasenauer and White 2000, Fodor and Mika 2011). Variations of these B&C models have been undertaken with artificial intelligence (Mellit et al. 2007, Benghanem, Mellit and Alamri 2009, Mellit 2008, Tymvios et al. 2005, Remesan, Shamim and Han 2008, Behrang et al. 2010, Mellit et al. 2005, Dorvlo, Jervase and Al-Lawati 2002) and fuzzy logic (Mellit et al. 2007, Rivington et al. 2005, Santamouris et al. 1999). Chapter Three proposes the use of beta regression (Ferrari and Cribari-Neto 2004) to analyze B&C models. Beta regression is flexible, robust and easy to implement. Additional advantages are that it produces estimates of uncertainty and has a strong theoretical foundation. The beta regression model is shown to perfom well when compared to a previously proposed B&C model.

Chapter 4

Typically, B&C models are parameterized using historical time series data from one site. However, model parameterization could be done using data from numerous sites, but collected on one day. The flexibility of beta regression makes it an appropriate tool for comparing traditional site-based models to these daily models. Chapter Four investigates if analyzing daily data from networks of solar monitoring sites, or if incorporating spatial auto-correlation using universal kriging, leads to more precise and less biased predictions of solar radiation. The results suggest that daily models outperform site-based models, and that the daily universal kriging model slightly outperformed the daily beta regression model when comparing model fit to observed data.

Chapter 5

Chapter 5 validates the use of the beta regression model in mountain regions. While B&C models have experienced great popularity (Fodor and Mika 2011, Grant et al. 2004, Bristow, Campbell and Saxton 1985, Bandyopadhyay et al. 2008, Ball et al. 2004, Winslow, Hunt Jr and Piper 2001, Wu, Liu and Wang 2007, Bechini et al. 2000, Castellvi 2001, Bristow and Campbell 1984) they have only be tested in mountainous regions a few times (Glassy and Running 1994, Thornton et al. 2000). The results presented here compare favorably to previous models when predicting the effects of meteorological variables on daily fluctuations of solar radiation. However, this study elucidated the need for more reliable estimates of the amount of transmissivity on clear, dry days.

Appendix A

There are an increasing number of DC-PVPS being installed throughout the Western United States, chiefly to provide water the surface for livestock consumption. The Bozeman, MT office of the Natural Resources Conservation Service (NRCS) is tasked with overseeing federal incentive programs that encourage ranchers and farmers to install alternative energy power generation devices in place of nonrenewable tradition power sources. To accomplish this task efficiently, NRCS was in need of maps displaying in higher detail the spatial variability of solar radiation than was currently available for the state of Montana. Further, most DC-PVPS require a minimum amount of power to operate, or they will shut down to prevent overheating. Practitioners of photovoltaic technology can exceed this minimum threshold of operation by installing the appropriate number of solar radiation panels. Predictions of the number of hours per day that solar radiation exceeds certain power thresholds, (i.e. 400 watts m^{-2}, 600 watts m^{-2}) can aid end users in estimating the amount of solar panels needed. These needs were the impetus to developing the summarization methods described in Chapter 2, and Appendix A is the product of these advancements that was delivered to NRCS.

Literature Cited

Allen, R. G., R. Trezza & M. Tasumi (2006) Analytical integrated functions for daily solar radiation on slopes. *Agricultural and Forest Meteorology,* 139**,** 55-73.

Angstrom, A. (1924) Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society,* 50**,** 121-126.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Bandyopadhyay, A., A. Bhadra, N. S. Raghuwanshi & R. Singh (2008) Estimation of monthly solar radiation from measured air temperature extremes. *Agricultural and Forest Meteorology,* 148**,** 1707-1718.

Barry, R., M. Prévost, J. Stein & A. P. Plamondon (1990) Application of a snow cover energy and mass balance model in a balsam fir forest. *Water Resour. Res.,* 26**,** 1079-1092.

Bechini, L., G. Ducco, M. Donatelli & A. Stein (2000) Modelling, interpolation and stochastic simulation in space and time of global solar radiation. *Agriculture, Ecosystems & Environment,* 81**,** 29-42.

Behrang, M. A., E. Assareh, A. Ghanbarzadeh & A. R. Noghrehabadi (2010) The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy,* 84**,** 1468-1480.

Benghanem, M., A. Mellit & S. N. Alamri (2009) ANN-based modelling and estimation of daily global solar radiation data: A case study. *Energy Conversion and Management,* 50**,** 1644-1655.

Bernard, M. M. (1932) Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers,* 96**,** 592-624.

Berterretche, M., A. T. Hudak, W. B. Cohen, T. K. Maiersperger, S. T. Gower & J. Dungan (2005) Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment,* 96**,** 49-61.

Boutelhig, A., Y. Bakelli, I. Hadj Mahammed & A. Hadj Arab (2012) Performances study of different PV powered DC pump configurations for an optimum energy rating at different heads under the outdoor conditions of a desert area. *Energy,* 39**,** 33-39.

Boutelhig, A., A. Hadjarab & B. A (2011) Comparison Study to Select an Optimum Photovoltaic Pumping System (PVPS) Configuration upon Experimental Performances Data of Two Different DC Pumps Tested at Ghardaïa Site. *Energy Procedia,* 6**,** 769 -776.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Bristow, K. L., G. S. Campbell & K. E. Saxton (1985) An equation for separating daily solar irradiation into direct and diffuse components. *Agricultural and Forest Meteorology,* 35**,** 123-131.

Castellvi, F. (2001) A new simple method for estimating monthly and daily solar radiation. Performance and comparison with other methods at Lleida (NE Spain); a semiarid climate. *Theoretical and Applied Climatology,* 69**,** 231-238.

Crabtree, R., C. Potter, R. Mullen, J. Sheldon, S. Huang, J. Harmsen, A. Rodman & C. Jean (2009) A modeling and spatio-temporal analysis framework for monitoring environmental change using NPP as an ecosystem indicator. *Remote Sensing of Environment,* 113**,** 1486-1496.

Dorvlo, A. S. S., J. A. Jervase & A. Al-Lawati (2002) Solar radiation estimation using artificial neural networks. *Applied Energy,* 71**,** 307-319.

Firatoglu, Z. A. & B. Yesilata (2004) New approaches on the optimization of directly coupled PV pumping systems. *Solar Energy,* 77**,** 81-93.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Fu, P. & P. M. Rich (2002) A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture,* 37**,** 25-35.

Glassy, J. M. & S. W. Running (1994) Validating Diurnal Climatology Logic of the MT-CLIM Model Across a Climatic Gradient in Oregon. *Ecological Applications,* 4**,** 248-257.

Granger, J. E. & R. E. Schulze (1977) Incoming Solar Radiation Patterns and Vegetation Response: Examples from the Natal Drakensberg. *Vegetatio,* 35**,** 47-54.

Grant, R. H., S. E. Hollinger, K. G. Hubbard, G. Hoogenboom & R. L. Vanderlip (2004) Ability to predict daily solar radiation values from interpolated climate records for use in crop simulation models. *Agricultural and Forest Meteorology,* 127**,** 65-75.

Gueymard, C. A. & D. R. Myers (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy,* 83**,** 171-185.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Horel, J. D. & X. Dong (2010) An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology,* 49**,** 1563-1578.

Jin, J., X. Gao, S. Sorooshian, Z.-L. Yang, R. Bales, R. E. Dickinson, S.-F. Sun & G.-X. Wu (1999) One-dimensional snow water and energy balance model for vegetated surfaces. *Hydrological Processes,* 13**,** 2467-2482.

Keating, K. A., P. J. P. Gogan, J. M. Vore & L. R. Irby (2007) A Simple Solar Radiation Index for Wildlife Habitat Studies. *The Journal of Wildlife Management,* 71**,** 1344-1348.

Kolhe, M., J. C. Joshi & D. P. Kothari (2004) Performance analysis of a directly coupled photovoltaic water-pumping system. *Energy Conversion, IEEE Transactions on,* 19**,** 613-618.

Lynn, P. A. 2010. *Stand-Alone PV Systems*. John Wiley & Sons, Ltd.

Marks, D., J. Domingo, D. Susong, T. Link & D. Garen (1999) A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes,* 13**,** 1935-1959.

Marks, D. & A. Winstral (2001) Comparison of Snow Deposition, the Snow Cover Energy Balance, and Snowmelt at Two Sites in a Semiarid Mountain Basin. *Journal of Hydrometeorology,* 2**,** 213-227.

Martínez-Lozano, J. A., F. Tena, J. E. Onrubia & J. De La Rubia (1984) The historical evolution of the Ångström formula and its modifications: Review and bibliography. *Agricultural and Forest Meteorology,* 33**,** 109-128.

Mellit, A. (2008) Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review. *International Journal of Artificial Intelligence and Soft Computing,* 1**,** 52-76.

Mellit, A., A. H. Arab, N. Khorissi & H. Salhi. 2007. An ANFIS-based Forecasting for Solar Radiation Data from Sunshine Duration and Ambient Temperature. In *Power Engineering Society General Meeting, 2007. IEEE*, 1-6.

Mellit, A., M. Benghanem, A. H. Arab & A. Guessoum (2005) A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. *Solar Energy,* 79**,** 469-482.

Ogolo, E. O. (2010) Evaluating the performance of some predictive models for estimating global solar radiation across varying climatic conditions in Nigeria. *Indian Journal of Radio and Space Physics,* 39**,** 121 - 131.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

Pierce, K. B., T. Lookingbill & D. Urban (2005) A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology,* 20**,** 137-147.

Posorski, R. (1996) Photovoltaic water pumps, an attractive tool for rural drinking water supply. *Solar Energy,* 58**,** 155-163.

Prescott, J. A. (1940) Evaporation from a water surface in relation to solar radiation. *Trans. R. Soc. Sci. ,* 64**,** 114 - 125.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Remesan, R., M. A. Shamim & D. Han (2008) Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes,* 22**,** 4301-4309.

Rivington, M., G. Bellocchi, K. B. Matthews & K. Buchan (2005) Evaluation of three model estimations of solar radiation at 24 UK stations. *Agricultural and Forest Meteorology,* 132**,** 228-243.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Santamouris, M., G. Mihalakakou, B. Psiloglou, G. Eftaxias & D. N. Asimakopoulos (1999) Modeling the Global Solar Radiation on the Earth's Surface Using Atmospheric Deterministic and Intelligent Data-Driven Techniques. *Journal of Climate,* 12**,** 3105-3116.

Sherman, C. W. (1931) Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers,* 95**,** 951-960.

Thompson, R., M. Ventura & L. Camarero (2009) On the climate and weather of mountain and sub-arctic lakes in Europe and their susceptibility to future climate change. *Freshwater Biology,* 54**,** 2433-2451.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

Thornton, P. E. & S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Tymvios, F. S., C. P. Jacovides, S. C. Michaelides & C. Scouteli (2005) Comparative study of Angström's and artificial neural networks' methodologies in estimating global solar radiation. *Solar Energy,* 78**,** 752-762.

Winslow, J. C., E. R. Hunt Jr & S. C. Piper (2001) A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling,* 143**,** 227-243.

Wu, G., Y. Liu & T. Wang (2007) Methods and strategy for modeling daily global solar radiation with measured meteorological data - A case study in Nanchang station, China. *Energy Conversion and Management,* 48**,** 2447-2452.

Wu, H., K. G. Hubbard & J. You (2005) Some Concerns when Using Data from the Cooperative Weather Station Networks: A Nebraska Case Study. *Journal of Atmospheric and Oceanic Technology,* 22**,** 592-602.

Yesilata, B. & Z. A. Firatoglu (2008) Effect of solar radiation correlations on system sizing: PV pumping case. *Renewable Energy,* 33**,** 155-161.

Zeng, Z. G., P. S. A. Beck, T. Wang, A. K. Skidmore, Y. L. Song, H. S. Gong & H. H. T. Prins (2010) Effects of plant phenology and solar radiation on seasonal movement of golden takin in the Qinling Mountains, China. *Journal of Mammalogy,* 91**,** 92-100.

USE OF INTENSITY- DURATION- FREQUENCY CURVES AND EXCEEDANCE- FREQUENCY CURVES FOR QUANTIFYING SOLAR RADIATION VARIABILITY

Contribution of Authors

Chapter 2: Use of intensity- duration- frequency curves and exceedance- frequency curves for quantifying solar radiation variability.

First Author: Randall S. Mullen

Contributions: I was responsible for all data acquisition, data filtering and quality control. I developed all adaptations and wrote the first draft in its entirety, and was responsible for incorporating all subsequent edits.

Co-Authors: Brian L. McGlynn, Lucy A. Marshall

Contributions: Brian McGlynn and Lucy Marshall contributed significant critique and ideas for development of intellectual content within the paper, and edited successive versions of the manuscript as well as the final version. Lucy Marshall and Brian McGlynn were responsible for securing funding.

Manuscript Information

Mullen, R. S., L. A. Marshall, and B. L. McGlynn. 2012. Use of intensity-duration-frequency curves and exceedance-frequency curves for quantifying solar radiation variability. Manuscript for submittal to *Renewable Energy*

- Journal: Renewable Energy
- Status of manuscript (check one)
  - X Prepared for submission to a peer-reviewed journal
  - ___ Officially submitted to a peer-reviewed journal
  - ___ Accepted by a peer-reviewed journal
  - ___ Published in a peer-reviewed journal

- Publisher: WREN
- Date of submission: Anticipated Submission April 2012

Abstract

Networks of *solar monitoring sites* are increasing in coverage and density worldwide. These networks are producing millions of records of data each year. Typically, these data are summarized as monthly or yearly averages of daily solar radiation totals. However, this does not adequately characterize the short-term temporal variability of solar radiation, nor does it reflect the probabilities of occurrence of periods of low radiation. We suggest using intensity-duration-frequency (IDF) curves to indicate year to year variability and to improve assessment of solar radiation reliability. We modify IDF curves to represent relevant short-termvariability and enhance interpretability. We also propose simple exceedance-duration (ED) curves to represent the number of hours per day exceeding various solar radiation thresholds. These products can be used to determine solar radiation characteristics of relevance to solar collector installation and efficiency

Keywords: *intensity duration frequency, solar intensity duration frequency curves, short-term solar intensity duration frequency curves, Montana, solar radiation, photovoltaic*

Introduction

Effective implementation of off-grid power generation using photovoltaic technology requires accurate site-specific predictions of solar radiation characteristics. This is not limited to summaries of available solar radiation at monthly, daily and hourly scales. Predictions of durations of low solar radiation periods and estimates of the number of hours over a variety of thresholds are important for predicting photovoltaic technology efficacy and efficiency. Gueymard (2009) described three levels of solar radiation data collection sites that maintain solar radiation instrumentation. The level with the least expensive instrumentation and the least amount of redundancy was termed *solar monitoring sites.* Publicaly available long-term data sets from these sites provide affordable characterization of solar radiation at a variety of temporal scales. Networks of automated weather stations typically fall into this category (Wu, Hubbard and You 2005, Palmer 2011, Horel and Dong 2010), thus they can provide a much higher density of measurements than *conventional long-term* sites or *research sites* (Gueymard and Myers 2009). The latter two categories have much more precise instrumentation, but are far less common in North America. Summarizing the spatial characteristics of solar radiation using only these later two sources requires modeling solar radiation at greater distances from the source of data, or producing maps with little to no variability over large geographic areas.

Among the many applications that benefit from maps of projected solar radiation are devices that run directly off of the power generated by a photovoltaic cell, thus circumventing the need for battery banks for energy storage. The simplest directly

coupled photovoltaic water pump systems (DC-PVPS) consist of a DC pump connected directly to a photovoltaic array. Water is pumped to the surface during the day time hours when solar radiation exceeds a minimum threshold needed for operation of the pump. Costly battery banks are eliminated and the storage of energy is replaced by the storage of water (Boutelhig, A.Hadjarab and Bakelli 2011, Boutelhig et al. 2012). Among other uses, DC-PVPS provide affordable access to clean drinking water for many third world communities (Lynn 2010, Posorski 1996) and provide necessary water for livestock (Boutelhig et al. 2011, Boutelhig et al. 2012).Yesilata et. al (2008) concluded that substantial errors in predicted power output for DC-PVPS occur when relying on modeled solar radiation data and that using long-term solar radiation data sets results in smaller discrepancies between predicted and actual output. For these and other photovoltaic technologies that require solar input exceeding a threshold value to operate efficiently, the number of hours per day above that threshold (threshold exceedance value) can be more important than daily cumulative solar radiation values. For these purposes, the variability of solar radiation at a variety of time scales becomes as important as the spatial availability of radiation. The National Renewable Energy Laboratory (NREL) provides monthly and yearly means of available solar radiation ($kWh/m^2/day$) throughout the United States (Myers 2005) but with no indication of daily and weekly variability. For grid-connected users, this information might suffice. However, for off-grid users reliant on battery banks, or users of DC-PVPS, additional information regarding lengths of low radiation (“solar droughts”) and return intervals (probabilities of droughts) is beneficial.

As high density networks of solar monitoring sites expand and data sets grow in time (Reinbold, Roads and Brown 2005, Palmer 2011) there will be an increasing desire to accurately summarize and present this data to photovoltaic practitioners. In addition to seasonal averages, the number of hours per day over a threshold value needed for operation of directly coupled photovoltaic systems (threshold exceedance value) is critical. Graphic summaries should be succinct, easy to read and communicate daily fluctuations (diurnal fluctuations), weather variations (the impact of cloud cover) and seasonal variations (related to the daily path of the sun), and they should do this for both cumulative solar radiation values and threshold exceedance values

Utilizing actual solar radiation data is not without challenges. Long term data sets are sparse compared to modeled or satellite data (Tham, Muneer and Davison 2010, Belcher and DeGaetano 2007). Some require processing before analysis and may not report certain variables in the needed temporal scale. There are studies showing a tendency in commonly used instrumentation to underestimate global solar radiation (Gueymard and Myers 2008). Improvements in data collection, logging and transmission are increasing the reliability of data being collected currently; however, concerns of data availability and quality still exist and need to be addressed in any study using direct solar radiation observations.

In light of this, we propose methods to characterize solar radiation such that inferences can be made regarding periods of high and low daily cumulative radiation and the average number of daily hours over specified thresholds. We adapt intensity-duration-frequency (IDF) curves (Sherman 1931, Bernard 1932) for summarizing daily cumulative

values of global solar radiation. This approach has not been used extensively outside of hydrometric data and (to the best of our knowledge) never for solar radiation. However, IDF curves in their original form provide valuable information for end users of photovoltaic technologies. An example of a traditional IDF curve constructed for solar radiation (SIDF) is included here. Then we extend the SIDF principle to create short-term solar intensity duration frequency (SSIDF) curves.

Short-term variability is often more useful for photovoltaic applications. SSIDF curves characterize periods of high and low solar radiation, but with return intervals on the order of 5 days, 10 days and 25 days. Just like SIDF curves, SSIDF curves can be constructed for any time period of interest, (i.e. weekly, monthly, seasonally, and yearly).

For applications that require exceeding a threshold value in order to operate, hourly solar radiation data are summarized for the purpose of estimating threshold exceedance values. Exceedance–Duration (ED) Curves are presented to aid planners and others that are interested not in daily cumulative solar radiation values but rather events where a threshold is exceeded for a length of time (typically hours). Each of these techniques directly utilizes solar radiation data collected at the earth's surface.

In this paper we investigate the IDF and ED derivation, the ability of the approach to compare multiple solar radiation monitoring sites, and the robustness of the method to missing or sparse data. We present a case study using sites in and near Montana, USA (Figure 2.1)

Methods

2.1 Solar-Intensity-Duration-Frequency Curves

Intensity-Duration-Frequency (IDF) curves were introduced in the 1930's to characterize extreme rainfall and flood events for hydrologic design (Bernard 1932, Sherman 1931, Dingman 2002) and have been in use in water resources engineering ever since (Dingman 2002, Koutsoyiannis, Kozonis and Manetas 1998, Veneziano et al. 2007). IDF curves are typically used to report return intervals for rare events, (e.g. 20 year floods, 100 year floods). As an example, data might be yearly maximums for 1-hr, 6-hr and 24-hour rainfall (Dingman 2002). Here we propose using daily cumulative solar radiation values for the purpose of solar radiation characteristics helpful for photovoltaic technologies that employ battery banks.

The solar radiation IDF (SIDF) approach involves extracting the full run of available data in chronological order. Note that the type of data used is not limited to daily cumulative values, although that is often the case. Realistically, a minimum of 20 years of data is desirable in order to capture the year to year variability that exists over the long term. Shorter data sets are less likely to capture 20 year events, and of course, longer data sets are more likely. There is no definitive length for what makes a data set appropriate for this analysis. Twenty years of daily values would yield 7305 data points, or n = 7305. Span is defined as a time period over which the algorithm will search for periods of continuously high or continuously low solar radiation. These spans presented here are just examples, and can be adjusted to more appropriate ones depending on the

need. In the following algorithm, steps for construction of a monthly SIDF curve are presented.

Algorithm 1: SIDF Derivation

1) Spans of 1, 3, 7 and 10 days are selected as time periods of interest for runs of low or high solar radiation values (note that time spans between 1 and 10 days can be interpolated from the resulting curves).

2) These intervals are denoted as i_1, i_2,…i_p for p different intervals of interest. For intervals of interest of 1 day, 3 days, 7 days and 10 days, then $i_1 = 1$, $i_2 = 3$, $i_3 = 7$and $i_4 = 10$.

3) Moving averages are calculated for each time span over the entire time series for a specific site. There will be n-i moving averages for each time interval of interest.

4) If monthly variability in solar radiation characteristics is desirable, then all of the moving averages are grouped by month to derive monthly SIDF curves. For instance, in order to create SIDF curves for July, only data from the month of July is inspected. For consistency, we assume that spans of days that start in one month and finish in the next are associated with the month in which they start.

5) The minimums and maximums of each of each span for each month in each year are then extracted and combined such that the resulting datasets for each span (one of minimums and one of maximums) will have as many members as there are years of data. To create SIDF curves for the month of July for a data set spanning 1990 to 2009, there will be one maximum and one minimum from July

1990, another from July 1991, July 1992, etc, such that there are 20 maximums and 20 minimums.

6) The 96th, 90th, 80th and 50th percentiles of the list of maximum values and the 4th, 10th, 20th, and 50th percentiles of the minimums are calculated to determine the 25-year, 10-year, 5-year, and 2-year return intervals for the periods of high radiation and the periods of low radiation.

7) These solar radiation percentiles are graphed for each time span, resulting in a separate curve for each return interval.

Short-term Solar-Intensity-Duration-Frequency Curves

The concepts of traditional IDF curves, or solar IDF (SIDF) curves, can be extended to address short-term solar radiation variability at a variety of scales. The same data set described above can be used to characterize daily fluctuations. This concept can be extended to any time step for which data are available.

The data used for constructing short-term solar IDF (SSIDF) curves are identical to the data used to construct SIDF curves, except that the monthly maximum and minimum values for each span are not extracted. Thus, the return intervals are in days instead of years, which leads to less extreme values then those reported for SIDF curves. One thing to consider is the unfiltered data used for SSIDF curves are more prone to auto-correlation problems than the data used for SIDF curves. Recall that a return interval is merely a probability, and that any event can, over the short-term, occur more or less frequently than a return interval would imply. The auto-correlation present in the data

used for SSIDF curves would likely exacerbate this issue. Weather patterns can and will cause some months in some years to be clearer than average, while others will be less clear. Therefore, events are likely to occur more frequently in some years and less in others. It is important to note that IDF curves are not forecasts, but rather representations of past probabilities. We have chosen to stratify by month for the examples in this paper. This does not remove all of the auto-correlation problems, but it does reduce bias. Any stratification desired can be done as dictated by end user needs.

The steps below are nearly identical to those described previously except that step 5, extraction of minimums and maximums, is left out and in step 6, all values are ordered. To obtain a SSIDF curve;

Algorithm 2: SSIDF Derivation

1) Similar to the previous example, spans of 1, 3, 7 and 10 days may be selected as time periods of interest for runs of low or high solar radiation.
2) These intervals are denoted as i_1, i_2,…i_p for *p* different intervals of interest. For intervals of interest of 1 day, 3 days, 7 days and 10 days, then $i_1 = 1$, $i_2 = 3$, $i_3 = 7$ and $i_4 = 10$.
3) Moving averages are calculated for each time span over the entire time series for a specific site. There will be n-i moving averages for each time interval of interest.
4) If monthly SSIDF curves are desired, then all of the moving averages are analyzed by month. For instance, in order to create SSIDF curves for July, only data from the month of July is extracted. As with the previous example, spans of

days that start in one month and finish in the next are associated with the month in which they start.

5) The 4^{th}, 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 96^{th} percentiles are calculated to determine the 25-day, 10-day, 5-day, and 2-day return intervals for the periods of high radiation and the periods of low radiation.
6) These solar radiation percentiles are graphed for each time span, resulting in a separate curve for each return interval. Each curve represents a specific return interval (e.g. 10-year) and connects the points on the graph representing that percentile for each span in days.

Where a SIDF curve denotes *yearly* return intervals, a SSIDF denotes *daily* return intervals. When interpreting a SSIDF curve (Figure 2.2), the x-axis represents the time span of interest. The y-axis represents the average number of kWh per day, however the return intervals are in units of days. For 7.5 kWh per day, the data show that a 'drought' of solar radiation resulting in less than 7.5 kWh lasting 1.75 days has a return interval of 5 days. Similarly, periods of 3.5 days lacking 7.5 kWh have a return interval of 10 days, and periods of 8.5 days have a return interval of 25 days.

Exceedance-Duration Curve

For situations where cumulative values are not useful but rather thresholds and the length of time exceeding a given threshold are of relevance, we propose a new concept for data summary and visualization. Exceedance-duration (ED) curves are useful for displaying the average number of hours per day that exceed a given threshold. The construction of ED curves follows from an extension of the IDF concept. The number of

hours (or minutes, seconds, etc) above a threshold are calculated for each day (say, for solar radiation data, watts per meter square). This is done for various thresholds, such as 1000, 900, 800 etc watts per meter square. Then, the average numbers of hours per day that exceed this threshold are calculated for each month, (or week, season, etc). These numbers can be presented in graphic form (Figure 2.3) for easy interpretation and interpolation.

SIDF curves are used to characterize yearly return intervals of periods of low or high solar radiation that last from 1 to 10 days. SSIDF curves characterize daily return intervals for less extreme events lasting 1 to 10 days. The spans of interest are arbitrary, and should be adjusted based on the individual case specific need. The case study will demonstrate how these spans are useful for characterizing solar radiation variability at specific locations.

Case Study

Overview of Case Study

Solar powered water pumps are one of the most popular uses of photovoltaic technology (Firatoglu and Yesilata 2004), due in part to the natural relationship between high solar radiation and lack of water, but also the need for affordable access to clean drinking water in poverty stricken regions of the world (Posorski 1996). In Montana, as well as other western states, DC pumps are used to pump water for livestock on off-grid grazing grounds. The installation of these systems requires summaries of solar radiation data that depends on the nature of the PV technology. For applications that use a battery bank and thus can make use of all available radiation throughout the day, Intensity-

Duration-Frequency (IDF) curves may be used to estimate and convey solar radiation characteristics. For instances where DC-PVPS are employed ED curves are necessary to estimate typical solar radiation thresholds. Here we illustrate the use of IDF and ED curves via case studies considering solar powered water pumps used in sites in Montana. We additionally examine how IDF and ED information may be summarized spatially via solar monitoring networks for improved understanding of regional solar radiation characteristics.

In order to determine potential bias resulting from missing data, a complete data set was analyzed and then reanalyzed with 10, 20 and 30% of the data randomly removed. A data set from Sidney, MT (AWDN) with almost 12 years of complete data was first analyzed to create SSIDF curves for the month of June. Calculating the 5^{th} decile and 3 return intervals for 4 different spans yields 28 calculations. The analysis was repeated after removing 10%, 20% and then 30% of the data. Since each simulated analysis relied on the random removal of data, 10,000 simulations were performed in order to quantify the bias from incomplete data sets. A similar analysis is performed on ED curves with just 10% of the initial data randomly removed.

Data Collation and Quality Control

Solar radiation data were obtained from several sources using similar protocols. The Bureau of Reclamation operates 26 weather stations in Montana referred to as the AgriMet (Agricultural Meteorology) monitoring network (Palmer 2011). Twenty-one of these are east of the continental divide and are operated by the Great Plains regional office. The Pacific Northwest regional office operates 5 stations west of the divide. An

additional 5 sites in Idaho were used for this analysis. The High Plains Regional Climate Center (HPRCC) operates the Automated weather Data Network (AWDN) through North and South Dakota, Wyoming, and other Great Plains states (Wu 2005). This network has two sites in Eastern Montana, and several in North and South Dakota that were used for this study (Figure 2.1). Each station is placed far enough from buildings, trees and topographical obstructions to avoid obstruction of solar radiation. Stations are powered by solar energy. The Western Regional Climate Center (WRCC) serves data collected by the Remote Automated Weather Stations (RAWS) throughout the Western United States (Horel 2010, Reinbold 2005). This network was put in place largely to monitor forest fire conditions, thus many of the sites are affected by topographical and local shading. Twenty-six of these stations were deemed usable (based on their lack of topographical and local shading) for this analysis, 22 in Montana, three in Wyoming and one in Idaho. Measured parameters in addition to solar radiation include air temperature, precipitation, wind speed, relative humidity. Each agency uses a Licor LI-200 (or similar) pyranometer designed primarily for field measurement of global solar radiation in agricultural, meteorological and solar energy studies. This sensor uses a silicon photovoltaic detector. This sensor has been shown to have less than 5% error under natural daylight conditions (Federer and Tanner 1966) or as high as 25% error under adverse conditions (Geuder and Quaschning 2006). A complete list of the station sites, elevation, installation dates and latitude and longitude is included in Appendix 1. The inferences made from the presented data should be extended into mountainous regions with extreme caution. Hill shading and elevation can greatly affect results both directly though the amount of sunlight hitting the

earth's surface, and indirectly through the formation of clouds and thermal inversions and change in aerosols.

Both the Pacific Northwest office and the Great Plains office of the Bureau of Reclamation apply quality control measures when converting hourly (or 15-minute) data to daily cumulative values. Instrument failure or yearly maintenance can result data gaps of a few hours to a few months. These were removed from the final data set for analysis. Single hours of missing data were imputed as the mean of the value before and the value after. All archived data from the Northwest region office were stored and transferred in hourly format. For the Great Plains office, older data were in hourly format, while newer data, (starting in 1997 but varying by station) is available in 15-minute increments. Small negative values, typically instrumentation errors, were converted to zeros.

For the AWDN sites used in this analysis, the data are filtered and flagged, and in some cases imputed for quality control purposes by HPRCC. The two AWDN sites in Montana were installed in 1995. The RAWS data are shipped in hourly format, and were submitted to the same quality assurance methods described above.

The four sites used in example figures are (with longitude, latitude and elevation in meters), Creston (-114.13, 48.19, 899), Dillon (-112.51, 45.33, 1524), Malta (-107.78, 48.37, 692), and Buffalo Rapids-Glendive (-104.80, 46.99, 652).

Results

Solar-Intensity-Duration-Frequency (SIDF) Curves

In order to evaluate SSID bias and robustness when missing data are present, a complete data set was analyzed and then reanalyzed with 10%, 20% and 30% of the data randomly removed. For all durations in all data sets with data removed, the predicted kWh $d^{-1}m^{-2}$ for events of low solar radiation was over estimated. Conversely, the predicted kWh $d^{-1}m^{-2}$ for events of high solar radiation and the median were underestimated, with the exception of events with duration of one day. In general, the recurrence lines are biased towards what would be the center of the graph (Table 2.1) if these values were graphed. The reason the median is biased low is that the distributions of the moving averages across are negatively skewed, with the one day results extremely skewed and the 10 day results lightly skewed. The extreme skewness for the one day results led to the upward bias for even the high radiation events (Table 2.1).

ED Curves

The ED curve allows for easy comparison of various sites (Figure 2.6) or various times of the year for the same site (Figure 2.7). Any time period of interest can be inspected, for instance, weekly intervals throughout a growing season can be calculated, or bi-weekly intervals during grazing period might be more useful. Data from one specific location can be compared to a region, by plotting the site specific ED curve along with the mean and the 0.1 and 0.9 quantiles for the region. This was done for Moccassin, MT curve and the state wide mean in order to show how Moccassin compares to the rest of the state (Figure 2.8).

The robustness of ED curves was tested. A data set from Sidney, MT with almost 12 years of complete data was first analyzed to create ED curves for the month of June. The data set was then reanalyzed with 10% of the data randomly removed in order to investigate bias from missing data. This resulted in a 7.4% bias when 10% of the data missing. However, the bias for higher thresholds was high, 24% for 1000 W/m^2 and 14% for 900 W/m^2, with decreasing bias as threshold decreases. ED curves are not as robust to missing data as IDF curves, and care should be taken when analyzing incomplete data. It is thought that ED curves are less robust since they are essentially counts. For these reasons, regardless of the data set used, ED curves should viewed with caution if a substantial amount of missing data exists.

Case Study Application of IDF and ED Curves

A typical application for a SSIDF curves might be as follows. Assume an off grid pumping system near Dillon, Montana, (latitude 43° 33′,longitude -112° 51′, elevation 1524 meters) has a tank that can store 3 days of water for consumption by livestock and requires 4.5 kWh/m^2 of solar radiation each day to run at full capacity. This system will be utilized in May (Figure 2.5).

To estimate the frequency at which the pumping system will be unable to maintain sufficient water storage, we find the time span of interest, in this case 3 days, (Figure 2.5). For incoming observed solar radiation at 4.5 kWh, the estimated frequency at which this value is not met or exceeded is 10-days. This can be interpreted as, “a 3 day span of 4.5 kWh or less is expected to start about every 10 days during the month of May”. An SIDF curve shows return intervals in years, thus interpretation would differ

and could be stated as such; “a 3 day span of 4.5 kWh or less is expected to occur during the month of May about every 10 years”. Both SIDF and SSIDF curves convey return intervals for periods of high or low solar radiation. Traditional summaries that focus on monthly averages do not convey this type of variability.

ED curves convey monthly summaries, but go beyond simple means. An ED curve near Malta, Montana, (latitude 48° 37′, longitude -107° 78′, elevation 692 meters) is used for demonstration purposes. Assume 700 gallons a day needs to be pumped to the surface for livestock consumption. A total dynamic head of 150 feet is present. A column of water 2.31 feet produces 1 pound per square inch of pressure (psi). A typical fixed position 1 m^2 panel can produce the 100 watts needed to run a typical pump when environmental condition exceed 600 watts per meter square (or about 18% of incoming). The ED curve for the AgriMet site near Malta, Montana is plotted (Figure 2.6). At an incoming radiation value of 0.6 kW it can be seen that an average of 6 hours per day exceeding this value have been observed in the past during the month of May. Since the pump can operate for 6 hours and pump 2.14 gallons per minute, then the total amount of water pumper per day if past observations are indicative of future solar radiation is about 770 gallons per day in May.

For locations not near an established weather station site, it is recommended that one use a spatially interpolated map for the entire state (Figure 2.9).The map shown is for 0.6 kWh, or an average of 600 watts for one hour. The contours represent average number of observed hours per day. A hypothetical location between the contours of 6 and 5 would suggest that for this location, about 5.5 hours a day will produce enough power

to run the water pump, and about 706 gallons of water a day will be pumped. This map was produced using inverse distance weighting (Gopinathan and Soler 1996) with an inverse distance weighting power of 2.

Extension to Spatial Solar Radiation Estimation

We investigate spatial variability in ED curves by plotting all ED curves for all sites within our case study network (Figure 2.10). We examine the dependency of solar radiation spatial variability on latitude alone. A color gradient dependent on the latitude of each site is used to show the correlation between latitude and ED curve metrics. This correlation is evident; however, there is variability in solar radiation that cannot be attributed to latitude alone (Figure 2.10). The effect is most likely atmospheric attenuation since all sites are void of local shading, either from topography or trees. This demonstrates that using an approach that calculates solar radiation using only latitude and ignoring local variation can lead to biased results.

Discussion

This study has described the process for constructing SIDF, SSIDF and ED curves that can be easily interpreted when summary solar radiation information is needed. Furthermore, these concepts are not limited to daily cumulative and hourly data respectively. Return intervals and time spans for SIDF and SSIDF curves can be changed based on needs, and the process adapted to fit any time series data of interest. SIDF, SSIDF and ED curves can be useful for summarizing long data sets and quickly presenting the summaries for planning purposes.

The interpretation of IDF curves such as the SIDF curves presented here are well described in the literature (Dingman 2002, Bernard 1932, Koutsoyiannis et al. 1998, Sherman 1931, Veneziano et al. 2007) with one exception. Where traditional IDF curves conveylong rainfall events, or high intensity events, SIDF curvesconveyperiods of low and high solar radiationAn example interpretation might be; “a 4 day span of less than 4 kWh/day in the month of May has a return interval of 10 years” (Figure 2.2).

SSIDF curves in particular can be useful in quantifying and summarizing short-term variability of solar radiation. Establishing return intervals for periods of low solar radiation is critical for planners and users of photovoltaic technology to better estimate battery bank size and / or periods of inactivity due to insufficient incoming solar radiation. These summaries are meant to enhance and augment monthly averages of daily cumulative solar radiation. SSIDF curves can be adapted to any time frame of interest using the methods described here. The methods described here can be used to characterize high volumes of daily, hourly or even minute-to minute data into usable output.

Despite the strengths and utility of the descriptive methods, there are limitations to these methods, and end users must understand that this approach does not account for temporal non-stationarity. That is, long term trends in the data will not be readily apparent when using the methods described herein. Most all probabilistic forecasting assumes stationarity, so this study is not unique in this fashion. Furthermore, since SSIDF and ED curves are constructed with all available data, auto-correlation is present. Therefore, return intervals should be considered with caution. Even if stationarity did

exist, periods of low or high solar radiation are likely to return at unevenly spaced time intervals based on longer term climate patterns.

Using solar radiation observations for estimation of solar radiation characteristics and variability has definitive advantages over using modeled data. However, is not without limitations. Gueymard (2009) outlines potential biases with pyranomters commonly used in many of today's weather stations. For this reason, it is advisable, when combining data from different sources, to carefully inspect summaries and daily values and determine if proper protocols and upkeep have been employed for the length of time that the data have been collected. This is not trivial, and can involve a lengthy quality control process but biases in the final results can be greatly reduced. Even when solar radiation measurement devices are being properly maintained, data are often corrupted by failures in data loggers, remote power sources, and temporary events that take days to weeks to fix (i.e. dust collecting on sensors, temporary shading due to obstruction, and broken weather stations). Agencies collecting the data provide varying amounts of quality control for hourly and cumulative data. Researchers wishing to use these data need to be familiar with the levels of control and filtering that each agency performs. SIDF and SSIDF curves are robust to missing data, so data filtering can involve simple removal of corrupt data and does not necessarily require that missing values be imputed. Because ED curves are essentially a count, they are less robust to missing data, and it is recommended that with more than 10% missing data, or with any amount of systematically missing data, that a simple imputation method be applied to the missing data before creating ED curve (Badescu et al. 2012, Srivastava, Singh and Pandey 1995).

In exploring variation based on latitude (Figure 2.9), it was readily apparent that while latitude is important in net solar radiation, it is not the only factor. This underscores the importance of using real data from a nearby station and the value of solar monitoring site networks. The case study presented herein provides one example of how an end user might use available data when designing a photovoltaic system.

Very few newly implemented photovoltaic systems will be installed near an existing solar monitoring site. The case study presented here demonstrats one way to interpolate the values from an ED curve across the region using inverse distance weighting (Figure 2.9), but other interpolation methods could be used. There are numerous methods proposed for interpolation of solar radiation values (Ball, Purcell and Carey 2004, Hasenauer et al. 2003, Thornton, Running and White 1997) which may prove valuable for interpolating SIDF, SSIDF and ED values as well. Any simple interpolation approach makes certain assumptions about the new site, and will not account for local variation from shading or localized weather patterns. It is incumbent upon the end user to adjust accordingly.

Conclusion

The increasing number of long term solar radiation data sets, as well as the growing existing data sets require new techniques for summarization that are easily understood by end users, helpful for decision making, and allows for comparisons between sites and time periods. The example of directly coupled photovoltaic water pump systems (DC-PVPS) and available solar radiation is just one illustrative use. To

date, solar radiation summaries have focused on averages (daily, monthly, etc) of total energy or power available. While useful, these averages do not address the short-term variability of solar radiation, and do little to indicate periods where photovoltaic systems that do not have auxiliary power may fail to provide the appropriate amount of energy. We propose the use of SIDF, SSIDF, and ED curves as ways to easily convey the short-term variability of solar radiation at sites where it is measured. IDF curves have a long history of use for hydrometric data, and their interpretation is well understood by climatologists. The adaptation we present has no current analogy in the field of solar radiation monitoring, thus the introduction of these concepts is an important step in furthering the discussion for summarization of long term data sets.

Acknowledgements

The authors would like to thank the Bureau of Reclamation, specifically Tim Groove in the Great Plains office and Peter Palmer in the Pacific Northwest office. We would also like to thank the High Plains Regional Climate and the Western Region Climate Center..

Vendors that provided input are Dwight Patterson of GenPro Energy Solutions, and Sarah Ray and Sara Biddle of Independent Power Systems.

Funding was provided by the Bozeman office of the National Resources Conservation Service (NRCS).

Tables

	Span in Days	Quantile						
		0.04	0.1	0.2	0.5	0.8	0.9	0.96
Missing 10%	1	0.0039	0.0176	0.0071	-0.0002	0.0007	0.0001	0.0009
	3	0.0186	0.0065	0.0033	-0.0050	-0.0017	-0.0021	-0.0008
	7	0.0112	0.0139	0.0164	-0.0069	-0.0054	-0.0016	-0.0011
	10	0.0116	0.0034	0.0096	-0.0015	-0.0014	-0.0069	-0.0078
Missing 20%	1	0.0075	0.0254	0.0100	-0.0006	0.0009	0.0002	0.0012
	3	0.0530	0.0151	0.0052	-0.0096	-0.0048	-0.0034	-0.0019
	7	0.0222	0.0285	0.0280	-0.0125	-0.0096	-0.0057	-0.0043
	10	0.0210	0.0086	0.0201	-0.0041	-0.0056	-0.0120	-0.0127
Missing 30%	1	0.0119	0.0313	0.0118	-0.0009	0.0010	0.0003	0.0013
	3	0.0936	0.0263	0.0061	-0.0133	-0.0077	-0.0048	-0.0031
	7	0.0327	0.0443	0.0377	-0.0166	-0.0140	-0.0108	-0.0091
	10	0.0342	0.0159	0.0277	-0.0064	-0.0105	-0.0171	-0.0170

Table 2.1 The difference in kWh $d^{-1}m^{-2}$ between the complete data sets with a percentage of data missing.

Figures

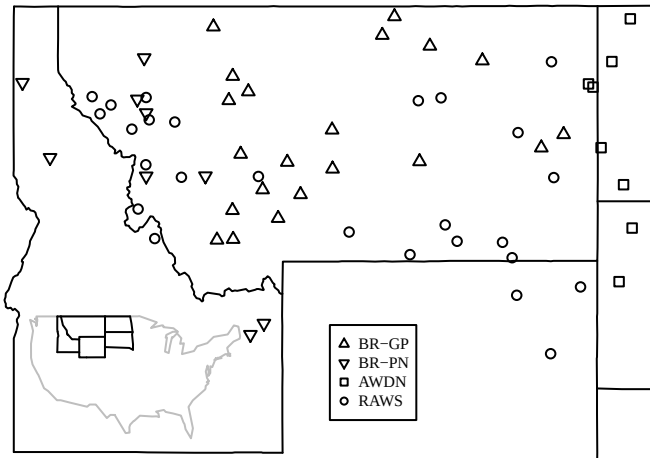


Figure 2.1. Map showing 64 sites in and around the state of Montana used for analyzing solar radiation. Pacific Northwest BOR sites are shown with a triangle pointed down, the Great Plains BOR sites are shown with a triangle pointed up, the AWDN sites are shown with squares, and the RAWS sites are shown with crosses. The four sites that are circled (with longitude, latitude and elevation in meters) are (clockwise from upper left), Creston (-114.13, 48.19, 899), Dillon (-112.51, 45.33, 1524), Malta (-107.78, 48.37, 692), and Buffalo Rapids-Glendive (-104.80, 46.99, 652). These sites are used as examples throughout the text. The five state region in shown in black as part of the Unites States inset shown in gray in the lower left hand corner.

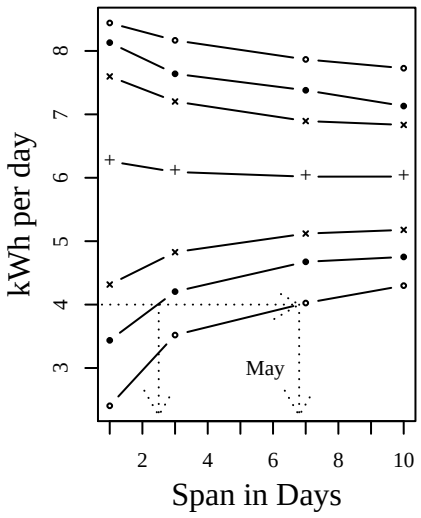


Figure 2.2. Typical SIDF curve for the month of May at Moccasin MT.

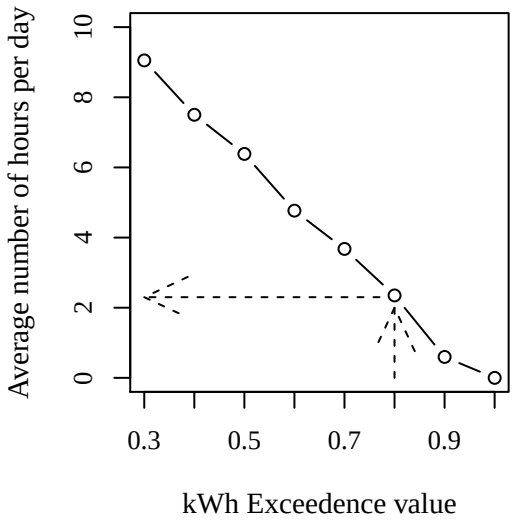


Figure 2.3. Typical EDF curve with arrows indicating hours above 0.8 kWh threshold.

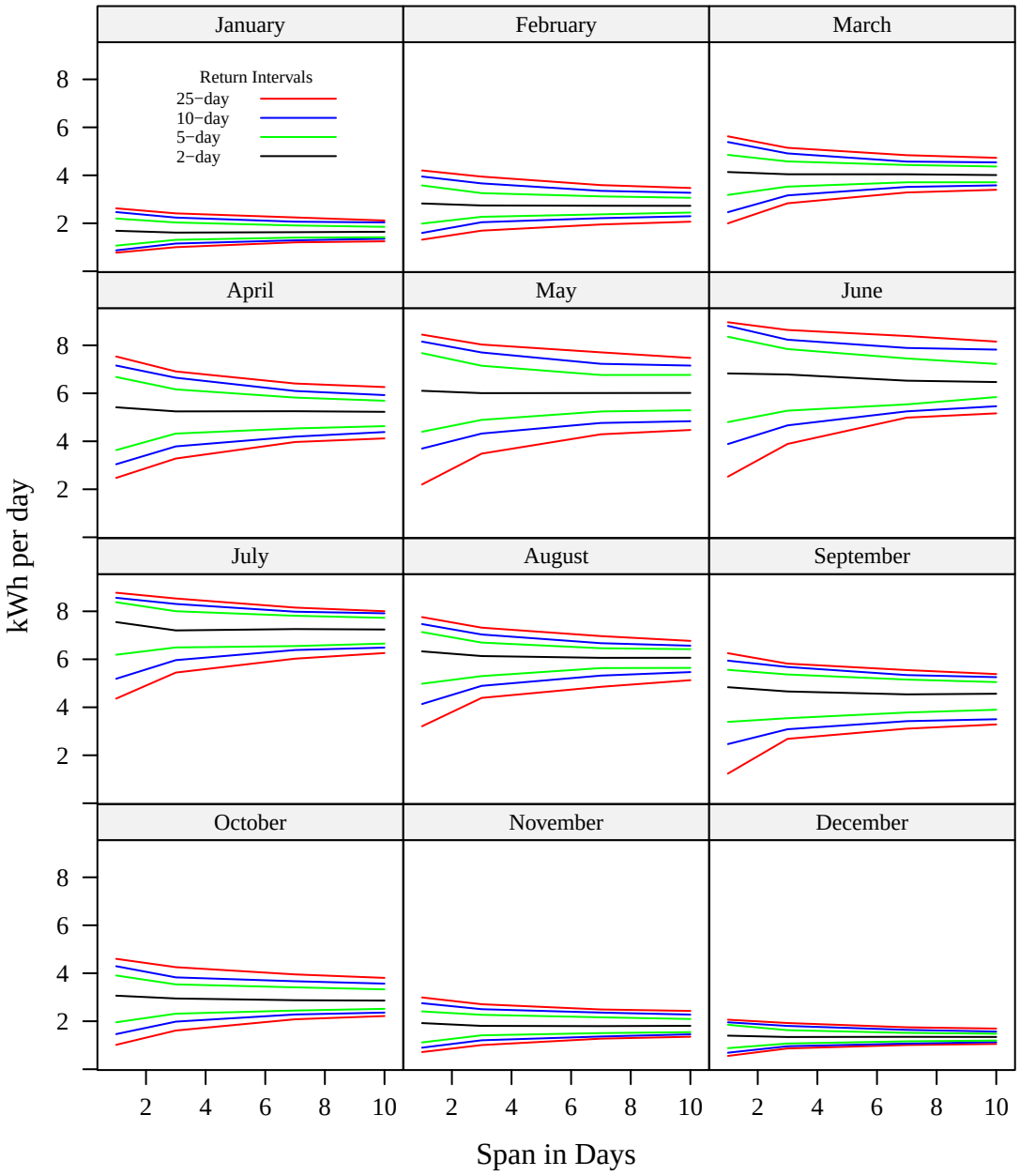


Figure 2.4. Monthly SIDF curves for White Sulphur Springs, MT. The y-axis is kilowatt hours per day, and the x-axis is the number of consecutive days for which that level is observed. The middle black line in each panel is the 2-day recurrence lines. The next line in each direction (above and below the black line) is the 5-day, with the 10-day and 25-day lines shown in blue and red

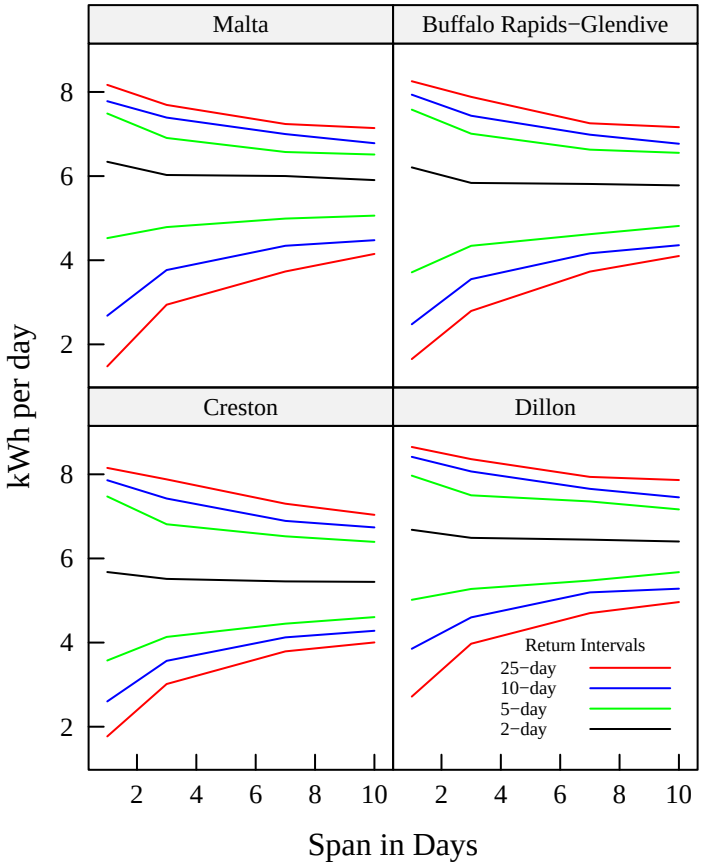


Figure 2.5. Four different SSIDF curves from four different sites in Montana. Data is from the month of May. Shown are 25-day, 10-day, 5-day and 2-day recurrence lines.

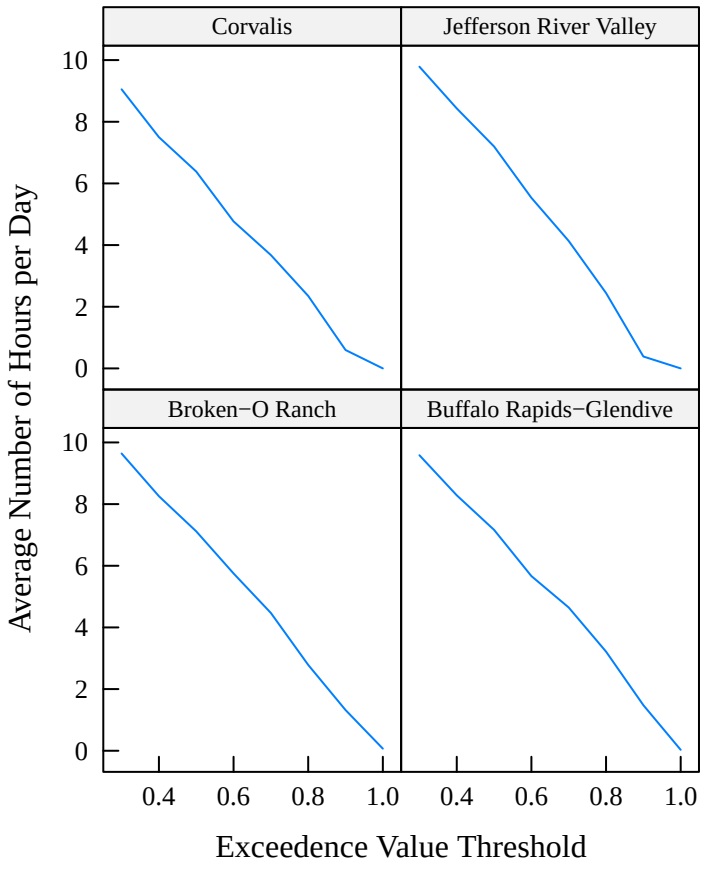


Figure 2.6. Four different Exceedance Duration Frequency curves from four different sites in Montana for the month of June.

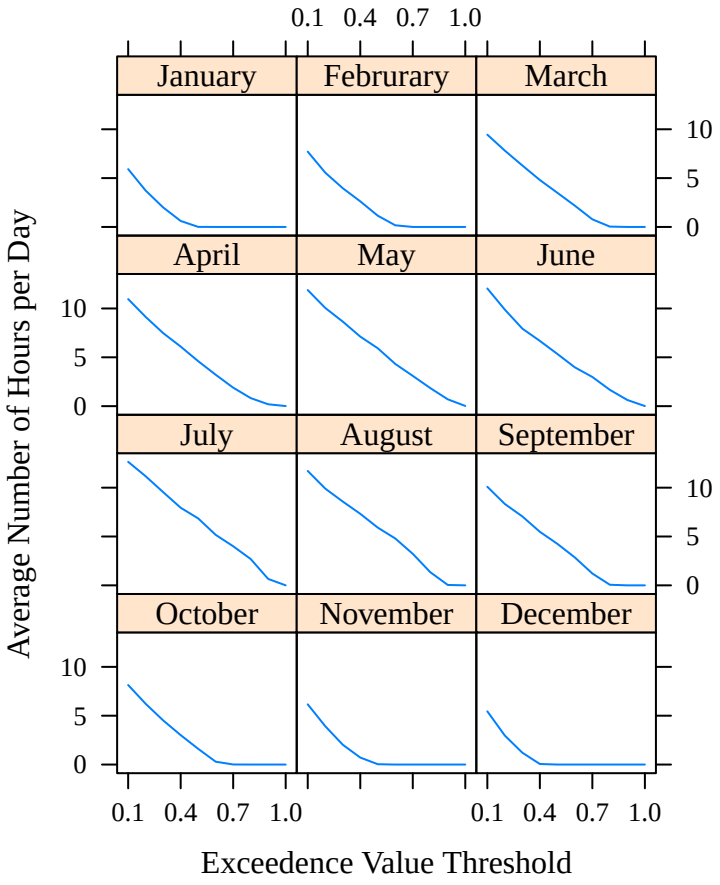


Figure 2.7 Monthly EDF curves for Bozeman, MT and monthly variation. The y-axis is average number of hour per day for the noted month, and the x-axis is the exceedance value threshold.

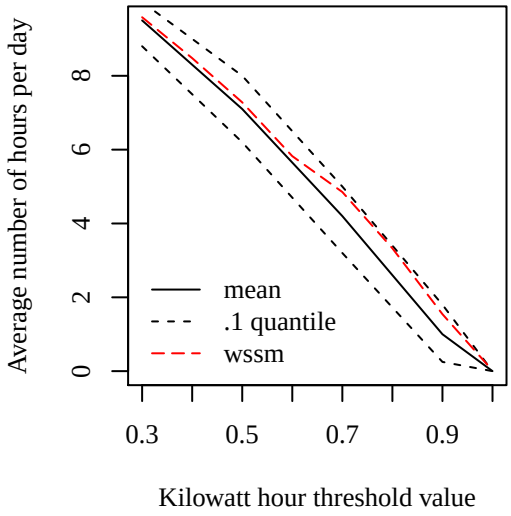


Figure 2.8 The relationship between the Moccasin, Montana site and the distribution of solar radiation across the state. The mean is show as a solid black line, the .1 and .9 quantiles shown in dashed line, and the Moccassin site shown in red dashed line.

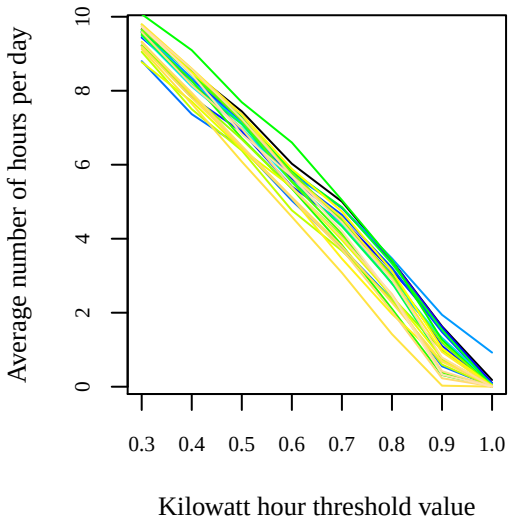


Figure 2.9 Average number of hours per day above radiation threshold values. Topographic color scheme represents latitude of site., If latitude were the only factor for variability, graph would show a continuous color ramp from low to high.

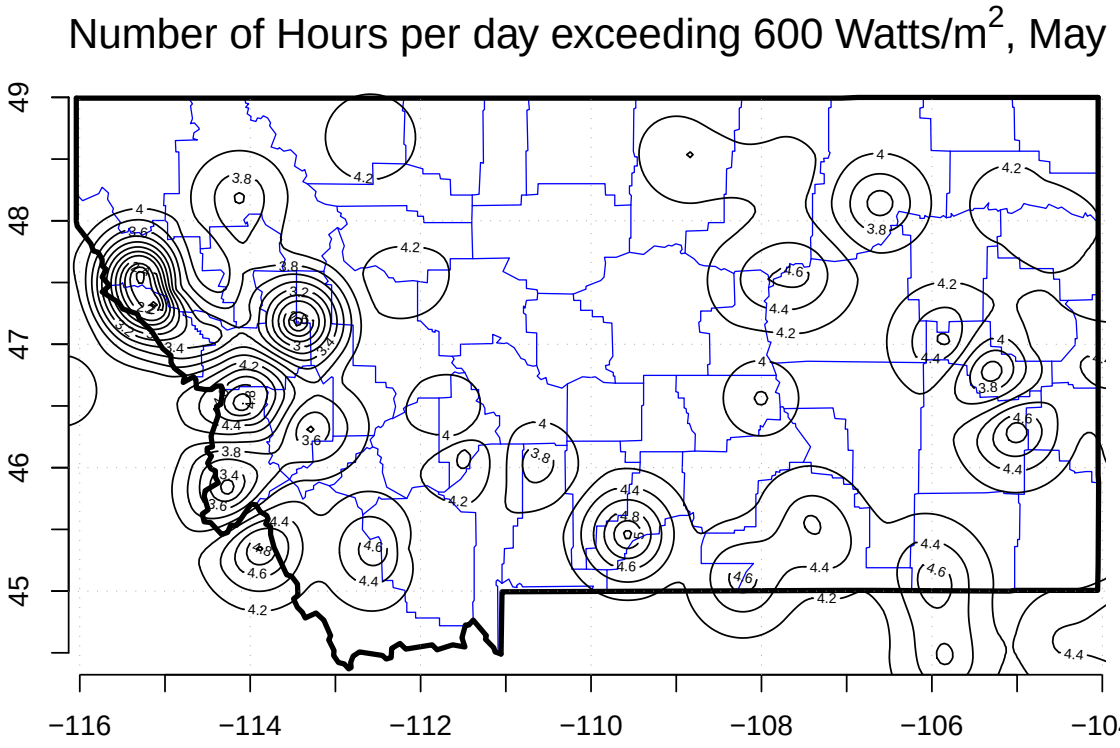


Figure 2.10. Interpolated map of the state of Montana showing the number of hours per day exceeding 600 Watts/m^2. Using a map such as this, used of photovoltaic technology can quickly estimate the number of hours per day in each month above a given threshold.

Literature Cited

Badescu, V., C. A. Gueymard, S. Cheval, C. Oprea, M. Baciu, A. Dumitrescu, F. Iacobescu, I. Milos & C. Rada (2012) Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models. *Renewable and Sustainable Energy Reviews,* 16**,** 1636-1656.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Belcher, B. N. & A. T. DeGaetano (2007) A revised empirical model to estimate solar radiation using automated surface weather observations. *Solar Energy,* 81**,** 329-345.

Bernard, M. M. (1932) Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers,* 96**,** 592-624.

Boutelhig, A., A.Hadjarab & Y. Bakelli (2011) Comparison Study to Select an Optimum Photovoltaic Pumping System (PVPS) Configuration upon Experimental Performances Data of Two Different DC Pumps Tested at Ghardaïa Site. *Energy Procedia,* 6**,** 769-776.

Boutelhig, A., Y. Bakelli, I. Hadj Mahammed & A. Hadj Arab (2012) Performances study of different PV powered DC pump configurations for an optimum energy rating at different heads under the outdoor conditions of a desert area. *Energy,* 39**,** 33-39.

Dingman, S. L. 2002. *Physical Hydrology*. Upper Saddle River: Prentice Hall, Inc.

Federer, C. A. & C. B. Tanner (1966) Sensors for Measuring Light Available for Photosynthesis. *Ecology,* 47**,** 654-657.

Firatoglu, Z. A. & B. Yesilata (2004) New approaches on the optimization of directly coupled PV pumping systems. *Solar Energy,* 77**,** 81-93.

Geuder, N. & V. Quaschning (2006) Soiling of irradiation sensors and methods for soiling correction. *Solar Energy,* 80**,** 1402-1409.

Gopinathan, K. K. & A. Soler (1996) Effect of sunshine and solar declination on the computation of monthly mean daily diffuse solar radiation. *Renewable Energy,* 7**,** 89-93.

Gueymard, C. A. & D. R. Myers. 2008. Solar Radiation Measurement: Progress in Radiometry for Improved Modeling. In *Modeling Solar Radiation at the Earth's Surface*, 1-27.

Gueymard, C. A. & D. R. Myers (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy,* 83**,** 171-185.

Hasenauer, H., K. Merganicova, R. Petritsch, S. A. Pietsch & P. E. Thornton (2003) Validating daily climate interpolations over complex terrain in Austria. *Agricultural and Forest Meteorology,* 119**,** 87-107.

Horel, J. D. & X. Dong (2010) An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology,* 49**,** 1563-1578.

Koutsoyiannis, D., D. Kozonis & A. Manetas (1998) A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology,* 206**,** 118-135.

Lynn, P. A. 2010. *Electricity from sunlight: An introduction to photovoltaics*. John Wiley & Sons, Ltd.

Myers, D. R. (2005) Solar radiation modeling and measurements for renewable energy applications: data and model quality. *Energy,* 30**,** 1517-1531.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

Posorski, R. (1996) Photovoltaic water pumps, an attractive tool for rural drinking water supply. *Solar Energy,* 58**,** 155-163.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Sherman, C. W. (1931) Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers,* 95**,** 951-960.

Srivastava, S. K., O. P. Singh & G. N. Pandey (1995) Correlations for the estimation of hourly global solar radiation. *Applied Energy,* 52**,** 55-64.

Tham, Y., T. Muneer & B. Davison (2010) Estimation of hourly averaged solar irradiation: evaluation of models. *Building Services Engineering Research and Technology,* 31**,** 9-25.

Thornton, P. E., S. W. Running & M. A. White (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology,* 190**,** 214-251.

Veneziano, D., C. Lepore, A. Langousis & P. Furcolo (2007) Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall. *Water Resour. Res.,* 43.

Wu, H., K. G. Hubbard & J. You (2005) Some Concerns when Using Data from the Cooperative Weather Station Networks: A Nebraska Case Study. *Journal of Atmospheric and Oceanic Technology,* 22**,** 592-602.

Yesilata, B. & Z. A. Firatoglu (2008) Effect of solar radiation correlations on system sizing: PV pumping case. *Renewable Energy,* 33**,** 155-161.

A BETA REGRESSION MODEL TO OBTAIN INTERPRETABLE PARAMETERS AND ESTIMATES OF ERROR FOR IMPROVED SOLAR RADIATION PREDICTIONS

Contribution of Authors and Co-Authors

Chapter 3: A beta regression model to obtain interpretable parameters and estimates of error for improved solar radiation predictions

First Author: Randall S. Mullen

Contributions: I was responsible for all data acquisition, data filtering and quality control. I developed all adaptations and wrote the first draft in its entirety

Co-Authors: Brian L. McGlynn, Lucy A. Marshall

Contributions: Lucy Marshall and Brian McGlynn contributed significant critique andideas for development of intellectual content within the paper and edited draft versions of the manuscript as well as the final version. Lucy Marshall and Brian McGlynn were responsible for securing funding.

Manuscript Information

Mullen, R. S., L. A. Marshall, and B. L. McGlynn. 2012 A beta regression model to obtain interpretable parameters and estimates of error for improved solar radiation predictions. Manuscript under review to *Journal of Applied Meteorology and Climatology*

- Journal: *Journal of Applied Meteorology and Climatology*
- Status of manuscript (check one)
  - ___ Prepared for submission to a peer-reviewed journal
  - _X_ Officially submitted to a peer-reviewed journal
  - ___ Accepted by a peer-reviewed journal
  - ___ Published in a peer-reviewed journal
- Publisher: American Meteorological Society
- Date of submission: January 2012

Abstract

Predicting global solar radiation is an integral part of much environmental modeling. There are several approaches for predicting global solar radiation at a site where no instrumentation exists. One popular approach uses the difference between daily high and low temperature, typically using a nonlinear equation to express the relationship between change in temperature and estimated global solar radiation. Additional variables are usually included in successive steps creating a hierarchy of analysis. We propose an alternative beta regression approach to modeling global solar radiation, allowing for the inclusion of multiple environmental predictor variables and strata into one flexible model. We apply the model to several case studies and compare results to recently propose empirical solar radiation models. Beta regression provides a robust, flexible modeling approach for predicating global solar radiation that allows for the addition and removal of independent variables as appropriate and can be interpreted using standard inferential statistics. In addition, the beta regression model provides estimates of uncertainty that be incorporated into subsequent models and calculations.

*Keywords***:** *Global Solar radiation; Beta regression; North Dakota; South Dakota; Automated Weather Data Network; Fraction clear day*

Introduction

Predictions of solar radiation are a requisite to models of soil moisture (Spokas and Forcella 2006), carbon flux and plant growth (van Dijk, Dolman and Schulze 2005), wildlife behavior (Keating et al. 2007), evapotranspiration (Hargreaves and Samani 1982), weed management (Spokas and Forcella 2006) , hydrology (Zhou and Wang 2010), and others. Numerous models have been proposed to predict solar radiation at ungauged locations because of the frequent lack of instrumentation to directly measure it (Thornton and Running 1999). Of particular note are the class of models that predict solar radiation loss due to atmospheric attenuation. These models are typically used to predict solar radiation as an input to evapotranspiration, crop management, or other environmental models (Spokas and Forcella 2006). One common approach is to use the difference between the daily maximum and the daily minimum temperature (ΔT) at a location as a means to predict the fraction of solar radiation that reaches the Earth's surface .

To date, a wide variety of models have been implemented that predict solar radiation based on observations of ΔT. One of the earliest was proposed by Hargreaves and Samani (1982) where the square root of extraterrestrial radiation is multiplied by ΔT and a coefficient, initially fixed at 0.75, but later adjusted by relative humidity. Bristow and Campbell (1984) proposed a model where transmittivity is a function of smoothed ΔT and three fitted parameters that are predicted for an individual site using historical data. Richardson (1985) proposed a simple model where ΔT is a function of two site specific empirical parameters and extraterrestrial radiation. Liu and Scott (2001) compare

nine models that predict solar radiation, three of which use only ΔT, two that use only precipitation and four that use both. Samani et al (2011) propose a modified version of Allen (1997), a model self-calibrated by season and location. A non-linear equation is used in each of these to model the relationship between ΔT and solar radiation. Evrendilek and Ertekin (2008) reviewed 78 empirical models including those based on ΔT, and while some regression models were inspected, they focus on general model suitability for monthly predictions of solar radiation, and not site-specific parameter estimation.

Thornton and Running (1999), proposed a ΔT method enhanced with precipitation and dew point data. Their motivation was to better predict solar radiation for locations where no previously collected data are available. The model uses dew point and precipitation to better predict the maximum atmospheric transmittivity at a location on a given day (CST). ΔT is then used to estimate the fraction of CST on any given day. The model necessitates hourly estimation of potential solar radiation (based on calculation of solar hour and zenith angle amongst other variables) and dew point data to predict surface vapor pressure. However, dew point data is not as widely available as temperature and precipitation data.

Fodor and Mika (2011) revisited ΔT models, and compared an 'S-shaped' function borrowed from soil science with Donatelli and Campbell's (1998) function for predicting the fraction of solar radiation that hits the Earth's surface. This fraction, called Fraction of Clear Day (FCD), is expressed at the percentage of solar radiation that would be expected to hit the earth's surface on a clear day. This latter value is referred to Clear Sky

transmittivity (CST) and is described in detail, along with FCD, in section two . Both Bristow and Campbell (1985) and Donatelli and Campbell (1998) used equations for FCD that were forced through the origin. While these functions represented FCD observations reasonably well, Fodor and Mika (2011) noted the incongruity with the real world; FCD cannot ever be zero (except perhaps in the polar winters). Fodor and Mika (2011) then proposed a four parameter sinusoidal curve and found it produces smaller prediction errors when compared to Donatelli and Campbell (1998).

All aforementioned models are limited by the available observations for model fitting. Gueymard and Myers (2009) described three levels of stations that collect solar radiation data:(1) *Solar monitoring sites* use inexpensive and automated instrumentation to provide local data quickly for a minimal cost; (2) *Conventional long-term measurements* use proven techniques and are generally operated by weather service agencies; and (3) *Research sites* are typically developed by atmospheric physicists or climatologists to obtain the highest accuracy possible in order to detect trends or test theoretical solar radiation models. These research sites have higher levels of redundancy with respect to instrumentation and power supply. Typically, ΔT models are developed and tested on filtered data collected at research sites. Spokas (2006) used data from 16 research sites throughout North America, Sweden and Australia. Thornton and Running (1999) and Fodor et al (2011) used data from the SAMSON data base (SAMSON, 2009) that included up to 109 stations from around the United States. Liu and Scott (2001) used 39 research sites distributed throughout Australia. Bristow and Campbell (1985) developed their model at three different locations in the northern United States. Using

(relatively) independent sites with high quality data to formulate predictive equations provides a strong basis for model development and assessment. However, it is a relatively rare situation that research sites will have to predict solar radiation, given the redundancy in equipment and power supply that these sites maintain. More likely scenarios are the need to infill missing data from solar monitoring sites (Gueymard and Myers 2009) during periods of equipment or power failure, replacement, and calibration.

Data from solar monitoring is less precise and has more gaps than data from research sites. Further, instrumentation calibration requires greater labor and infrastructure costs, and is generally implemented less frequently. Solar monitoring sites may then produce observations with greater uncertainty or potential bias. Despite this, networks of solar monitoring sites provide vital and valuable information. They allow for the investigation of spatial characteristics that is simply not possible when inspection is limited to research sites. They also provide information about solar radiation trends outside of research grade sites. Thus there remains a need for flexible modeling frameworks that can be applied to all sites that collect solar radiation data.

In this study, we implement a beta regression model to facilitate prediction of incoming solar radiation at ungauged locations. The intent of this study is not to develop a widely transferable model with fixed parameters, but rather establish a flexible method that allows researchers to add or remove variables based on local availability and appropriateness. The model also provides valid estimates of uncertainty, interpretable parameters and accurate predictions. We consider the application of beta regression in the context of solar monitoring networks. As with previous models, the beta regression

model we propose does not directly model global solar radiation but rather FCD. Detailed discussion of ΔT models, the deconstruction of global solar radiation, and beta regression follows.

A Review of ΔT Models for Solar Radiation Prediction

Global solar radiation (GSR) can be broken down into three components. Extraterrestrial radiation (ETR) is the amount of solar radiation that hits the outside of the atmosphere. Clear sky transmittivity (CST), is the amount of ETR that will reach the Earth's surface on a clear day. Fraction of clear day, (FCD), is the fraction of CST that hits the Earth's surface on any given day. ΔT models take advantage of this deconstruction and relate the difference of high and low daily temperature to FCD. The suite of current ΔT models (Fodor and Mika 2011, Bristow and Campbell 1984, Donatelli and Campbell 1998) for predicting FCD and subsequently GSR, can largely be described by the following sequence of analysis.

1. Determine ETR at a given site using geographical location, time of day and time of year (e.g. Gates, 1980).
2. For each day of the year (denoted *yearday*), estimate CST. This can be predicted empirically using historical data or can be modeled (e.g. using Fourier series) with shorter historical data sets.
3. For each day in a given data set, divide measured daily GSR by CST to determine FCD.
4. Calculate ΔT for each day in the given data set. The simple calculation (Hargreaves and Samani 1982) is;

$$\Delta T = T^{i}_{\max} - T^{i}_{\min} \quad (1)$$

where i = the i^{th} day of the dataset.

A smoothed calculation first proposed by Bristow and Campbell (1984) and used frequently is;

$$\Delta T = T^{i}_{\max} - 0.5 \cdot (T^{i}_{\min} + T^{i+1}_{\min}) \quad (2)$$

5. Plot FCD versus ΔT and fit a non-linear curve
6. For any day at this any nearby location, if GSR is unknown and ΔT is known, then GSR can be predicted using the fitted value for FCD and the following relationship.

$$\widehat{GSR} = ETR \cdot CST \cdot FCD_{(fitted)} \quad (3)$$

It is assumed that the procedures in Step 1 are well established (Gates, 1980). For Step 2, if a sufficiently long data set exists then CST can be obtained empirically. Thornton and Running (1999) use a moving window that encompasses 7 days, (3 before and 3 after) for each *yearday* to empirically derive CST while also proposing a way to derive CST with no solar radiation data from a site. Fodor and Mika (2011) suggest using a Fourier series to model CST using the maximum values for each *yearday* in a dataset. For Step 3, the ΔT value (Eq.2) suggested by Bristow and Campbell (1985) has been used by several subsequent studies (Fodor and Mika 2011), however Thornton and Running (1999) found that using non-smoothed values (Eq.1) led to less error.

Step 5, modeling the relationship between FCD and ΔT, is probably the most contested aspect of the above algorithm. Multiple methods have been proposed and tested, with many demonstrating reduced values for root mean squared error (RSME),

mean signed deviance (MSD), or mean absolute error (MAE) over previous studies (Donatelli and Campbell 1998, Fodor and Mika 2011). The traditional justification for fitting ΔT models is that the model is useful for prediction purposes. Little effort is spent interpreting the fitted parameters in part because interpreting the coefficients would not yield better predictions of FCD. Additionally, interpreting parameters of these models is difficult or impossible. Fodor and Mika (2011) make no attempt to interpret parameters using a simplified soil water retention curve. The emphasis here will be on prediction; however, it is important to note that by using a beta regression approach, model interpretation and basic statistical inferences can be made. For example, explnatory variables can be assessed for relevance at different sites under different climatic conditions in order to determine relevance for particular situations. Relevant contributions of each independent variable can be determined. Since these are established statistical principles, we leave the details to the reader, and present the model with a focus on prediction.

Typically, predicted FCD values are inputted at Eq.3 for steps 5 and 6 with no regard for estimates of uncertainty in the predicted values. Resulting GSR predictions are then reported without prediction intervals. Attempts to spatially interpolate parameters and / or final GSR predictions (Step 6) are done as if known measured values are being presented (Ball, Purcell and Carey 2004, Running and Thornton 1999, Thornton, Hasenauer and White 2000, Thornton and Running 1999, Fodor and Mika 2011).

As mentioned, previous methods have not been tested using data from solar monitoring sites, but rather using conventional long-term measurements or research sites

(Fodor and Mika 2011, Bristow and Campbell 1984, Thornton and Running 1999, Spokas and Forcella 2006). While important for developing new models, successful implementation with high quality data does not ensure success with lower quality data. Data from solar monitoring sites is often missing long periods of solar radiation data (greater than two months), has small but notable measurement error (Gueymard and Myers 2008), and might have been collected for a relatively short period of time (< 4 years). However, high density networks of these solar monitoring sites are often placed in agriculturally important regions (e.g., the Automated Weather Data Network or AWDN) (Becker and Smith 1990) or in areas of high snowpack, such as the SNOTEL (SNOwpack TELemetry) sites (Rehman and Ghori 2000). When subsequent soil moisture or hydrological models are analyzed at a regional or watershed scale, data from these monitoring networks (made up of solar monitoring sites) will often be utilized rather than the data from sophisticated yet sparse research sites. In addition, networks of solar monitoring sites allow for detailed inspection of spatial auto-correlation of solar radiation that is simply not possible with sites that are dispersed throughout a continent at a much lower density. Networks of solar monitoring sites then provide an opportunity for more robust model assessment and analysis.

A Review of Beta Regression

Beta regression provides a framework for modeling continuous variables constrained in the standard unit interval (0,1)(Ferrari and Cribari-Neto 2004). A necessary assumption is that the response variable is beta-distributed with a mean that can be related to a set of regressors with estimable coefficients and a link function. The

beta distribution is a continuous probability distribution defined on the interval between 0 and 1 and its probability density function is traditionally expressed as;

$$f(y;p,q)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1},\quad 0<y<1 \tag{4}$$

with shape parameters p and $q > 0$, and where $\Gamma(\cdot)$ is the gamma function. Ferrari and Cribari-Neto (2004) reparameterized the Beta distribution by setting $\mu = p/(p+q)$ and $\phi = p+q$. This yields;

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},\quad 0<y<1 \tag{5}$$

where $0 < \mu < 1$ and $\phi > 0$. As in the original parameterization, $\Gamma(\cdot)$ is the gamma function. The expected value of y is μ, or $E(y) = \mu$. The parameter ϕ is known as the precision parameter since for fixed μ, larger ϕ gives smaller variance for the distribution. A beta distributed variable can be denoted as $y \sim \beta(\mu,\phi)$. In matrix notation, beta regression is then represented as;

$$g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} = \eta_i \tag{6}$$

where $\beta=(\beta_1,...,\beta_k)^T$ is a k x 1 vector of unknown regression parameters, $x_i=(x_{i1},...,x_{ik})^T$ is a vector of k regressors, or independent variables, $g(\mu)$ is a link function (in this case the logit link), and η_i is a linear predictor. This is a naturally heteroscedastic function with;

$$Var(y_i)=\frac{u_i(1-u_i)}{1+\phi} \tag{7}$$

Beta regression provides an effective framework for modeling bounded environmental variables. When a dependent variable, such as FCD, is between 0 and 1 standard regression techniques are likely inappropriate. Assumptions of normality are usually incorrect because truncation of the response value makes even an approximate normal distribution unlikely. Almost by definition they display a large amount of heteroscedasticity with more variation around the mean and less close to 0 or 1. Like most proportion data, FCD distributions tend to be asymmetric, which leads to issues with confidence intervals and hypothesis testing. Beta regression addresses all of these issues (Ferrari and Cribari-Neto 2004). Further, functions to perform beta regression are now readily available in popular software programs (Cribrali-Neto, 2010). The flexibility of beta regression is easily demonstrated by modeling predictions of FCD using a set of climate variables that are regularly collected at weather stations as regressors. Unlike previously proposed methods, beta regression is not limited to one independent variable. Donatelli and Campbell (1988) proposed a two-step approach that allows for the inclusion of more than one independent variable. In this case, the variable *b* is estimated as shown (Eq. 11) then included in the final equation as a known constant.

While non-linear beta regression techniques exist (Simas, Barreto-Souza and Rocha 2010), they are not easily implemented in popular software packages at this time (Cribari-Neto, 2010). However, as has been traditionally done with standard linear regression, the modeling framework can be extended via introduction of a squared and cubic term to account for the specific non-linearity of the relationship between FCD and ΔT.

In past studies, (Fodor and Mika 2011, Thornton and Running 1999, Donatelli and Campbell 1998) ΔT models have focused solely on prediction and prediction errors, with little to no emphasis on parameter interpretation and evaluation of standard errors. Researchers interested in determining statistical significance between various weather variables and solar radiation could not do so in the framework of previously suggested ΔT models. Beta regression (Ferrari and Cribari-Neto 2004) allows for all standard regression inferences to be made when fitting FCD versus ΔT , or any combination of climate variables available to the researcher.

In this study, a new flexible ΔT model using beta regression is compared to the Fodor and Mika (2011) model using data from a network of solar monitoring sites throughout North and South Dakota. Points of analysis include,

1. Comparison of standard indicators of fit such as RMSE, MAE and MSD.
2. Comparison of ease of fitting and assessment of whether all models are equally robust to small data sets..
3. Comparison of reliability of prediction intervals for FCD and demonstration of how to estimate prediction error of GSR using the variance of predicted FCD.
4. Comparison of ease of interpretation of the model parameters.
5. Determination of whether modifications to the standard design of the beta regression model are necessary for improved model predictions, including data stratification.

Materials and Methods

Data and Site Description

The study area is comprised of North and South Dakota in the north-central United States. These states have distinct continental climate with very cold winters and hot semi-humid summers, although the western part of North Dakota is considered semi-arid. The highest recorded temperature in either state is 49^{o} C and the coldest is -51^{o} C. The average annual precipitation ranges from 35 to 75 cm throughout the study area.

Data from 99 AWDN (Automated Weather Data Network) (Figure 3.1) operated by the High Plains Regional Climate Center were inspected for quality and quantity (length of data series and amount of missing data). Standard weather variables collected at the AWDN sites include (but are not limited to) daily high temperature, daily low temperature, relative humidity, and precipitation. Six sites had a sufficient amount of missing data as to prevent the fitting of a Fourier series to derive CST and were dropped from the analysis. Chosen for comparison were 93 sites inside of North and South Dakota and two in Montana very near the border of North Dakota. All data that were flagged as bad, missing, or imputed using regression were deleted. Three sites were chosen to demonstrate a variety of attributes concerning the data, as well as analysis of results. These are the Redfield, Takini, and Brookings sites in South Dakota. The total area that can be reasonably inferred as coverage is approximately 382,843 km^2, yielding a density of 2.5E-4 sites/km^2. It should be noted that this coverage provides an opportunity to assess the model performance over a reasonably dense monitoring network in comparison to relevant previous studies. Fodor and Mika (2011) inspected 109 sites spread across the

contiguous United States and Hawaii (~8,311,200 km^2, 1.3E-5 sites/km^2). Bechini (2000) inspected 29 stations in Northern Italy, (~100,408 km^2, 2.8E-4 sites/km^2). The density of coverage for this study is thus almost 20 times denser than the data set used by previous studies in North America (Fodor and Mika, 2011).

All data used for model comparisons, including minimum temperature, maximum temperature, relative humidity and precipitation, were collected from these sites.

Decomposing Global Solar Radiation

Global solar radiation can be deconstructed into three elements; ETR, CST and FCD. Doing so provides a simple approach for addressing seasonal cycles, effects of elevation and atmospheric attenuation independently. The historical context of this deconstruction is discussed in Section 2, with details of how each component was calculated below. In this study, calculations for ETR and CST are essentially unchanged from past studies. The beta regression model we are proposing is intended to improve upon past methods for predicting FCD.

ETR was calculated using methods described in Gates (1980). In this method, day of year, latitude, distance to sun, and declination (derived using latitude) are determined for each site. The calculation of ETR accounts for seasonal changes in the solar radiation. The solar constant is considered to be 1366 W m^{-2}

The annual course of CST is typically cyclical with relatively small amplitude and asymmetrical peaks (Fodor and Mika, 2011). Daily sky transmissivity (ST) values were determined for every day of the data set. Maximum ST values were extracted for each *yearday* using a 7-day moving window (Thornton and Running, 1999) in case a reliable

maximum cannot be captured in a relatively short data set. These maximums were then fitted with the second order Fourier series shown as Eq. 8, (Fodor and Mika, 2011)

$$y = a + b \cdot \cos(2x) + c \cdot \sin(2x) + d \cdot \cos(4x) + e \cdot \sin(4x) \tag{8}$$

where $x = 2\pi(yearday / 366)$ and *a*, *b*, *c*, and *d* are fitted constants.

This was done for each site individually, resulting in each site having an associated set of values for the Fourier series parameters. CST was modeled yearly, regardless of whether FCD was analyzed in seasonal strata or not. The effects of individual site characteristics on GSR are accounted for in the calculation of CST.

Where GSR data is observed, FCD can be easily calculated by rearranging Eq (3).

$$FCD = GSR \cdot ETR^{-1} \cdot CST^{-1} \tag{9}$$

Where GSR is not observed, FCD can be predicted using temperature and other climate variables. For this step, the proposed beta regression model is implemented. For comparison, other methods are briefly discussed here.

Bristow and Campbell (1984) suggested;

$$FCD = a \cdot (1 - \exp(-b \cdot \Delta T^{c}) \tag{10}$$

where ΔT is calculated as shown in Eq. 2 and *a*, *b*, and *c* are fitted constants.

Donatelli and Campbell (1988) suggested a two-step approach;

$$FCD = 1 - \exp(-a \cdot f_1(T^i_{avg}) \cdot f_2(T^i_{\min}) \cdot \Delta T^2 \tag{11}$$

where $T^i_{avg} = 1/2\,(T^i_{\max} + T^i_{\min})$

$$\Delta T^i = T^i_{\max} - 1/2\,(T^i_{\min} - T^{i+1}_{\min})$$

$$f_1(T^i_{avg}) = 0.017 \cdot \exp(\exp(-0.053 \cdot T^i_{avg}))$$

$$f_2(T^i_{\min}) = \exp(T^i_{\min}/b)$$

and *a* and *b* are fitted constants.

Thornton and Running (1999) suggest an entirely different approach to determining CST which includes calculating hourly ETR and requires dew point data, which is less frequently available (such as the AWDN sites used in this study). Their purpose was to provide a method that allows prediction of GSR with no prior data collected at a given site thus requiring no local calibration. Once CST is determined, they calculate FCD in a similar manner to Bristow and Campbell (1984) except with unsmoothed ΔT values. They provide fitted parameters, which may be used throughout North America with a simple correction for days with measureable precipitation. Fodor and Mika (2011) correctly point out that previous models are inappropriately forced through the origin (Bristow and Campbell 1984, Donatelli and Campbell 1998, Donatelli and Marletto 1994) such that as ΔT approaches zero FCD approaches zero. They then propose a strictly monotonic equation that is not forced through the origin;

$$FCD = 1 - \frac{1-a}{(1+(b\cdot\Delta T)^c)^d} \tag{12}$$

where *a*, *b*, *c*, and *d* are parameters that are empirically fitted at each site.

This model was found to produce smaller errors than a previous study by Donatelli and Campbell (1988). Error was further reduced when separate analyses were perfomred by season (winter, spring, summer, and fall) and precipitation (wet vs. dry). Therefore, eight unique models were required for each site. No interpretation of fitted parameters is provided, and although this is not strictly required for prediction purposes it would be useful for comparison between sites. Estimated parameters are interpolated and

a distance over-which those interpolations are useful is provided, but standard errors of the parameters from individual models are not.

In this study, FCD is predicted using beta regression. For model fitting, observed values of FCD are calculated using data from sites and days where GSR is measured (Eq. 9). Once fitted, the resulting regression equation can be used to predict FCD at locations and on days where explanatory variables are obtained but no measurement of GSR exists. This technique provides locally relevant parameter estimates such that a regression equation that has been fitted using nearby data can be used to predict FCD at a location that does not measure GSR.

There are multiple studies that review the implementation of beta regression models (e.g. Cribari-Neto and Zeileis 2010, Ferrari and Cribari-Neto 2004, Ospina, Cribari-Neto and Vasconcellos 2006, Rocha and Simas 2011, Simas et al. 2010, Smithson and Verkuilen 2006) and related model diagnostics (Espinheira, Ferrari and Cribari-Neto 2008a, Espinheira, Ferrari and Cribari-Neto 2008b). Interested readers are encouraged to consult these for further information on beta regression implementation. Here we construct an example of how the beta regression model may be applied to predictions of daily GSR using data from one station. These results will be compared to the Fodor and Mika model. Data from 2005-2009 are used to fit a beta regression model incorporating multiple climatic predictors The resulting parameters are used to make predictions for the 2010 Takini site data. We then extend the model to a solar monitoring network comprised of 93 stations. GSR is calculated using FCD predictions from both the beta regression model and the Fodor and Mika (2011) model. We assess the performance of the beta

regression model and its ease of use, and make recommendations regarding how it can be implemented. Due to the flexibility of the beta regression model, a binary variable for wet days was created as well as a continuous variable that is simply precipitation in mm. Model selection for the beta regression model was done using AIC. Beta regression was implemented in R (R Development Core Team 2009) using the betareg library (Cribari-Neto and Zeileis 2010).

Prediction Intervals for GSR

Since each predicted value of FCD is a beta distributed value, $y_i \sim \beta(u_i, \phi_i)$, then it distribution can be described with μ and ϕ. The 0.025 quintile can be considered the lower bound, and the 0.975 quintile can be considered the upper bound. These parameters, μ and ϕ, each have an associated uncertainty that is not incorporated into the uncertainty interval of FCD. The failure to account for this uncertainty is what distinguishes this estimate from a true prediction interval, however, it can be used similarly. The lower and upper bounds for the FCD uncertainty interval can be used to predict the upper and lower bounds for GSR.

True prediction intervals can be estimated. One could perform a simulation using predicted μ and ϕ for the new data, or a Bayesian approach could be used. Due to the high number of models run for this study, neither of these methods was used. However, they are an appropriate approach when analyzing one data set using a single model.

Model Comparisons

Root mean square error (RMSE), mean absolute error (MAE) and mean signed deviance (MSD), (an indicator of bias), were used to compare each of the Fodor and Mika models; one for each season and precipitation (wet vs. dry) combination, to a beta regression model using the same subsets of data.. Several studies have shown that performing separate analyses for each subset reduced error and bias (Fodor and Mika 2011, Allen 1997, Samani et al. 2011). However, since beta regression allows for multiple covariates, a single beta regression model that included variables for day of year and precipitation was developed. This eliminated the need for separate analyses of each subset. For instance, season can be entered into the model as a categorical variable or the day of the year can be converted into radians of rotation to account for temporal variability in the model. Since the variable *yearday* is effectively circular data we use the sine and cosine components of the day of year in radians. Traditionally, wet days are analyzed in a separate analysis from dry days. (Allen 1997, Fodor and Mika 2011, Thornton and Running 1999, Bristow and Campbell 1984), since the relationship between FCD and ΔT appears to be different on wet days. Again, due to the flexibility of the beta regression model, both a binary variable for wet days and a continuous variable that is simply precipitation in mm were created.

Ease of fit was determined by comparing the number of times computational efforts to fit each model either failed to converge or produced nonsensical parameter estimates. Non-linear optimization routines are less robust than linear regression with regards to convergence problems, especially when analyzing small data sets void of

distinct structure. The model that Fodor and Mika (2011) proposed cannot be linearized, thus it is susceptible to these issues. It is not unusual for a parsimonious non-linear model to have parameter identifiability issues, especially when no distinct structure in the data exists. In order to determine if subsetting was necessary for the beta regression model, all sites were analyzed with *yearday* converted into sine and cosine components. Precipitation was entered as a continuous variable. In this way, an entire data set can be evaluated at once making individual analyses for each subset of data unnecessary. Total RMSE from the stratified models was compared to the RMSE for the combined model to determine if loss of information occurred.

To test spatial interpolations of the fitted models, each site was analyzed using CST as well as the beta model fitted from the nearest site. The rate at which the observed value was captured by a 95% uncertainty interval was compared to capture rates of uncertainty intervals for each site.

Results and Discussion

CST was fitted for all stations (Eq. 8). An envelope curve for the Brookings weather station (Figure 3.2) had the following fitted parameters; $a = 0.7789$, $b = 0.0130$, $c = 0.0193$, $d = -0.0157$, and $d = 0.0067$.

Fitting the Fodor and Mika Model

The data were subseted as recommended (Fodor and Mika 2011). For each season, wet and dry days were split into two groups. Each of the eight resulting groups was analyzed. For one of the eight models, wet winter days, at the Redfield site, the

Fodor and Mika failed to converge. The total number of winter days with precipitation available for analysis was 96; not an uncommonly small sample size when considering sample sizes from the AWDN network (Figure 3.3). Traditionally, it is thought that a sinusoidal curve best represents the relationship between change in temperature and FCD. Most analyses herein support that belief, however, close inspection of the wet winter days stratum for the Redfield site lack this sinusoidal relationship (Figure 3.4). This could be due to the sample size, a different relationship between these two variables at this site during the winter season, or a combination of both. Regardless, forcing a sinusoidal curve through the points shown in Figure 3.4 leads to poor parameter identifiability. The entire data set (93 sites, 4 seasons, and 2 strata for wet and dry days) was analyzed using the Fodor and Mika model. Of the 736 possible models, 236 (32 %) failed to converge when using standard non-linear regression techniques.

Fitting the beta model was not problematic. Due its linear nature, the beta regression model was far more robust to this non-identifiability issue than the Fodor and Mika (2011) model. Additionally, since the model is linear, sound theoretical principles are available that yield estimates of uncertainty surrounding predicted response values (prediction intervals). This is in contrast to non-linear regression, where prediction intervals often rely on asymptotic estimates of variance for parameters (Goh and Pooi 1997).

Fitting the Beta Regression Model at the Takini Site.

Following standard procedures for beta regression, (Cribari-Neto and Zeileis 2010, Espinheira et al. 2008a, Espinheira et al. 2008b, Ferrari and Cribari-Neto 2004,

Ospina et al. 2006, Rocha and Simas 2011, Simas et al. 2010) , the 2005-2009 data set for dry spring days at the Takini station was analyzed. The resulting model parameters were used to construct predictions for the 2010 Takini dry spring days data set. An initial inspection of the explanatory variables (Figure 3.5) suggests there is notable correlation between relative humidity and both ΔT ($r = 0.65$) and adjusted ΔT ($r = 0.72$). This multicollinearity is a concern only if inferences regarding the estimates of coefficients in the final fitted model are desired. The confidence intervals for the fitted parameters will potentially be inflated and either relative humidity or ΔT can be deemed insignificant even though it is important in understanding potential model drivers. One could address this issue with an *a priori* science-based decision as to which variable to leave in the model or the variables can be combined (i.e. principle component analysis, single value decomposition). For prediction purposes, multicollinearity is of little concern.

To fit the sinusoidal relationship between ΔT and FCD, a squared and cubic ΔT term were added to the model. This is a standard approach for fitting non-linear relationships in a linear model. Inspection of the correlation matrix (Figure 3.5) suggests that FCD and subsequently solar radiation might also display a non-linear response to low temperature and relative humidity, therefore squared terms were added for each of those variables. The initial covariates in the beta regression model were ΔT, ΔT^2, ΔT^3, relative humidity (average of the day), relative humidity squared, daily low temperature, and daily low temperature squared. Two-way and three-way interaction terms were allowed between ΔT , relative humidity, and low temperature. AIC values were calculated for each for each possible model that maintained the squared and cubic ΔT

terms. The three models with the lowest AIC value were; 1) the full model with all variables (AIC = -506.74), 2) the full model with the one three way interaction term removed and the two-way interaction between ΔT and low temperature removed (AIC = -506. 05), and 3) the model with all of the single covariates and only one interaction term, relative humidity and ΔT (AIC = -506.14). In addition, there were two other models that had AIC values that were within 3 of the best model, (ΔAIC < 3). For the purpose of interpreting the estimates of the coefficients, model selection techniques can be used to determine the best model (Burnham and Anderson, 2001), but for the purposes of prediction, any of these models may be assumed to work reasonably well. The precision parameter ϕ for the full model was 11.5 (SE = 0.9343). As an example of calculating an uncertainty interval, 22 April 2010, had a low temperature of 5.25° C, a ΔT of 12.00° C, and an average relative humidity of 72.76% at the Takini, SD site. The observed FCD was 0.4104 and the predicted FCD is 0.6497. Predicted CST was 0.813, and ETR was 34.718. The observed GSR value was 11.589 MJ m^{-2} d^{-1}. and the predicted GSR is 18.347 MJ/$m^{-2}d^{-1}$. The uncertainty interval has lower and upper confidence bounds of 10.36 MJ m^{-2} d^{-1} and 18.34 MJ m^{-2} d^{-1} respectively, which capture the observed solar radiation value of 11.589 MJ m^{-2} d^{-1}.

This uncertainty could then be incorporated into all subsequent models that use estimated solar radiation as in input. In the previous example, CST is considered without error; however it is a predicted rather than measured. We incorporated the uncertainty in CST and found a negligible change in the final estimate for $\widehat{GSR}$ and ultimately omitted it from the final analysis.

For this subset, the 95% uncertainty intervals for the Takini 2010 test data set are shown in figure 6. These intervals captured the real value 100% of the time. This is not entirely surprising given the sample size of 46. However, for some purposes, a smaller prediction interval may be required. If smaller prediction intervals are desired, 90% intervals can be calculated by calculating the appropriate quantile from the resulting distribution.

Capture Rates for the Beta Regression Model

The full beta regression model was used to analyze the dry strata for all 93 sites across the four seasons to determine what proportion of the observations were captured by the 95% prediction limits (referred to as the rate of capture). Subsets with less than 15 days available for fitting the model, or less than seven days for testing the model were left out. There were 30 station – season combinations that were omitted for this reason. The average rate of capture of the true value was 95.27%, with a high of 100% and a low of 43.24 %. In this latter case, there were only seven usable days from the dry strata, fall season, 2010 data set (Site = Aurora). Clearly, when dealing with networks of solar monitoring sites, there will be cases such as these that require individual attention. The average capture rates for winter, spring, summer and fall were 97.87%, 94.36%, 94.79%, and 94.33% respectively. In order to assess overall model fit, observed GSR was plotted against predicted GSR and the correlation was calculated (Figure 3.7). This was done for all usable sites and subsets of data. The average correlation of observed GSR and predicted GSR on dry days for winter, spring, summer and fall was 0.89, 0.79, 0.83 and 0.92 respectively. There were more station – season combinations that did not meet the

minimal criteria for testing when inspecting wet days, (n=162). The overall average rate of capture of the true value for wet days was 90.67%, with a high of 100% and a low of 16.77%. In the latter case, there were nine usable days in the 2010 test data set. The average capture rates for winter, spring, summer and fall were 76.32%, 93.26%, 94.71%, and 81.53% respectively. The correlations of observed GSR and predicted GSR on wet days for each season were 0.59, 0.83, 0.87 and 0.85 respectively. The beta regression model tended to underestimate high values of solar radiation (Figure 3.8) and overestimate low values. Overall, this is a smaller problem in winter compared to the other seasons, and is possibly indicative of a missing variable in the model or a bias in instrumentation.

Note that the parameters *a,b,c* and *d* in Fodor and Mika (2011) have very little interpretable value. A particularly high value of *a* does not tell researchers anything about the relationship between FCD and ΔT. All inferential properties of linear models apply to the beta regression model, as long as all standard regression diagnostic criteria are addressed. Standard methods of model selection can be applied to the beta regression approach and model inferences can be made.

Model Comparison

Solar radiation predictions were made for all subsets of data that were successfully fitted using the Fodor and Mika (2011) model and compared to predictions estimated using the beta regression model (Table 3.1) for the Takini site. In each case, CST was derived using a Fourier series (Fodor and Mika 2011). For each data set, the ΔT values were smoothed using Eq. 2. However, as was shown previously, using Eq. 1 led to

better results (Thornton and Running 1999). Therefore, all models were run again using only the change in temperature for the day of interest. This yielded lower errors for all models and has the additional advantage of being less susceptible to erroneous values in the event of missing data (e.g. if day i+1 is missing, then calculation for day i is not jeopardized). The RMSE, MAE and MSD shown (Table 3.1 and Table 3.2) are based on the residuals for actual versus predicted GSR. Similar results can be shown for actual FCD versus predicted FCD, but since the intent of these models is to ultimately predict solar radiation, those results were compared. In all cases the RMSE and the MAE for the beta regression models were smaller than the Fodor and Mika model by an average of 10.28 MJ m^{-2} d^{-1}, with the lowest decrease being 3.34 MJ m^{-2} d^{-1}, and the largest decrease being 16.22 MJ m^{-2} d^{-1}. The MAE decreased for the beta regression model by an average of 3.00 MJ m^{-2} d^{-1}, with the lowest decrease being 0.75 MJ m^{-2} d^{-1}, and the largest decrease being 6.23 MJ m^{-2} d^{-1}. The mean signed deviance was larger for the beta regression model in 5 of the 7 cases. This increase in bias averaged 0.88 MJ m^{-2} d^{-1}, a full order of magnitude less than the decrease in RMSE. Therefore, the relatively small increase in bias is in our opinion negated by the substantial decreases in RMSE and MAE.

Each of the 93 sites was analyzed for each season and precipitation strata in order to determine if this pattern was consistent throughout the study area. The beta regression model outperformed the Fodor and Mika model with reduced RMSE and MAE (Table 3.2) for virtually every strata and every usable site. Overall, the RMSE was reduced an average of 17% and the MAE by 24%. The MSD was generally higher in the

beta regression model but in every case by less than 0.25 MJ m^{-2} d^{-1}. This slight increase in bias should not be a problem for most analyses.

Combining Strata for the Beta Regression Model

When inspecting data output from networks of solar monitoring sites, it is not unusual to have low sample sizes for numerous subsets of data (Figure 3.3). This problem can be alleviated by combining groups. A single beta regression model was used to analyze the Redfield, SD data to determine if seasonal (spring, summer, etc.) and climate (wet vs. dry) grouping is necessary. The y*earday* variable was transformed to radians, (as it is circular data) and the sine and cosine components were entered into the model as covariates. Precipitation was left in the model as a continuous variable. The resulting RMSE was 19.735, which is lower than the RMSE from each of the individual models run on separate strata (19.989). This indicates that indeed one model per site can outperform eight separate models for the same site. The beta regression approach allows for the introduction of numerous continuous variables and is the reason this reduction in subsetting without a loss of information is possible.

Interpolating Between Stations

There are advantages to using data from networks of solar monitoring sites despite less accurate solar radiation measurements. For instance, if site density is sufficient, Thiessen polygons will suffice for spatial interpolation. An analysis was performed using the fitted CST Fourier series and the fitted beta regression model coefficients from the nearest site in order to test if Thiessen polygons were appropriate for spatial interpolation. All available data were used (up through 2010). Predictions

intervals were calculated as previously described and capture rates were recorded. This was done for each site, for each season, and for dry and wet days. The overall mean capture rate was 92.89%. The average capture rate for dry days for winter, spring, summer and fall, were 94.14 %, 93.11%, 9.07%, and 92.210% respectively. The maximum rates were 98.81%, 98.84%, 98.03%, and 99.10% and the minimums were 80.54%, 80.80%, 71.04% and 47.15%. For wet days, the overall mean capture rates was 81.27%. The average capture rate for dry days for winter, spring, summer and fall were 69.83%, 87.46%, 89.57% and 77.05% respectively. The maximum rates were 96.44%, 97.61%, 99.46%, and 100% and the minimums were 69.81%, 87.41%, 89.53% and 76.94%. These capture rates indicate that Thiessen polygons are sufficient for spatial interpolation of beta regression parameters within networks of solar monitoring sites.

Conclusion

We applied a beta regression model to predict global solar radiation and compared results to recently proposed empirical solar radiation (ΔT) models. The beta regression method resulted in a lower RMSE and MAE than recently proposed models (Fodor and Mika 2011) that have outperformed historical models (Bristow and Campbell 1984, Donatelli and Campbell 1998, Donatelli and Marletto 1994). Beta regression can be easily implemented in free software (R Development Core Team 2009) using the betareg library (Cribari-Neto and Zeileis 2010). This allows for a more robust and simpler model fitting method than previously proposed non-linear methods. The parameters obtained using beta regressions are easily interpreted, if all diagnostic criteria is addressed. For

example, certain regions, climate types or strata may show common tendencies towards models with or without certain predictors (e.g. relative humidity, low temperature, etc.). Lower and upper bounds for estimatesof FCD can be used to predict upper and lower bounds for GSR. This is helpful not only as a measurement of uncertainty for GSR, but also for subsequent models that incorporate GSR. The beta regression method is flexible; it can be expanded if additional meteorological variables are available at a specific location, or it can be reduced if some variable are shown to be insignificant or unavailable. Because beta regression allows for a multiple regression analysis, variables such as time and precipitation that have been previously analyzed by subsetting the data can be incorporated into one model, which allows site – season combinations that previously had too few data points to analyze to be analyzed. The distribution parameters that accompany the predictions of a beta regression model can be used to estimate uncertainty in the final prediction of global solar radiation. To determine how well these models could be used at locations where no GSR data exists, each site was analyzed using the nearest neighbor. Predictions made using Thiessen polygons and beta regression parameters have slightly lower capture rates (mean of 93.16%) of the observed value using a 95% prediction interval. We have outlined a flexible modeling approach that allows for the addition and removal of independent variables as appropriate, accompanying measures of uncertainty, and ease of operation.

Acknowledgements

The authors would like to thank the National Resource Conservation Service for funding this project and the High Plains Regional Climate center for collecting this data and making it available for purchase.

Tables

Table 3.1 Comparisons of the Fodor and Mica model (F&M) and the beta regression model. Where NA is shown, the Fodor and Mica model was unable to be fitted. In all cases the RMSE and MAE were lower for the beta regression model. In 5 cases, the bias was lower for the F&M model but note the units are all in MJ m^{-2} d^{-1} and that the increase in bias is very small. This table uses data from the Redfield, SD site.

		RMSE ($MJm^{-2}d^{-1}$)		MAE ($MJm^{-2}d^{-1}$)		MSD ($MJm^{-2}d^{-1}$)	
Season	**Precip**	**Beta**	**F&M**	**Beta**	**F&M**	**Beta**	**F&M**
Winter	Wet	24.791	NA	6.025	NA	-0.069	NA
	Dry	121.160	128.242	6.840	7.664	0.010	0.037
Spring	Wet	125.006	134.704	28.831	33.478	-0.267	0.049
	Dry	207.640	222.446	24.427	28.035	-0.056	-0.015
Summer	Wet	108.674	122.772	22.538	28.765	-0.181	0.058
	Dry	179.974	196.199	18.701	22.225	0.090	0.042
Fall	Wet	44.805	48.143	8.766	10.121	-0.331	-0.036
	Dry	108.659	115.353	5.886	6.633	-0.011	-0.003

Table 3.2 Comparisons of the Fodor and Mica model (F&M) and the beta regression model. In all cases the RMSE and MAE were lower for the beta regression model. In 5 cases, the bias was lower for the F&M model but note the units are all in $MJm^{-2}d^{-1}$ and that the increase in bias is very small. This table uses all data from 92 sites.

		RMSE ($MJm^{-2}d^{-1}$)		MAE ($MJm^{-2}d^{-1}$)		MSD ($MJm^{-2}d^{-1}$)	
Season	**Precip**	**Beta**	**F&M**	**Beta**	**F&M**	**Beta**	**F&M**
Winter	Wet	26.34	31.66	4.72	6.82	0.094	0.012
	Dry	89.08	95.22	4.7	5.37	0.126	0.076
Spring	Wet	96.00	111.2	17.97	24.1	-0.196	0.044
	Dry	145.38	164.36	15.32	19.58	0.022	-0.052
Summer	Wet	86.63	109.3	12.87	20.49	-0.023	0.137
	Dry	110.84	124.57	9.36	11.82	0.06	0.074
Fall	Wet	34.38	43.33	4.19	6.66	-0.077	-0.027
	Dry	78.85	83.82	3.85	4.34	0.045	0.005

Figures

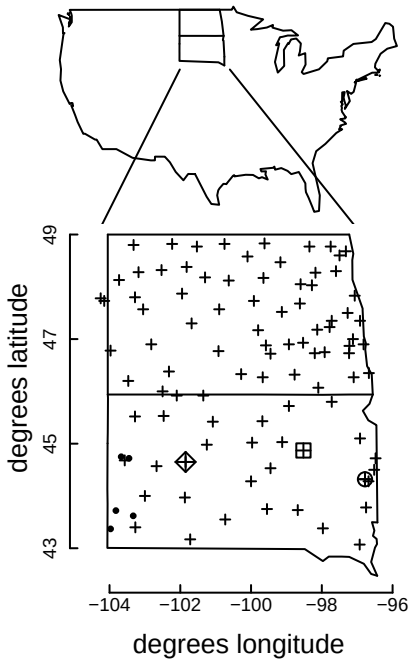


Figure 3.1 The Montana, North Dakota, and South Dakota sites of the AWDN network. Three sites mentioned in the text, Redfield, Takini, and Brookings, are denoted with a square, a diamond and a circle respectively. The bulleted sites are the sites that did not have enough data to create valid CST Fourier series. Top figure shows the location of North and South Dakota in the United States.

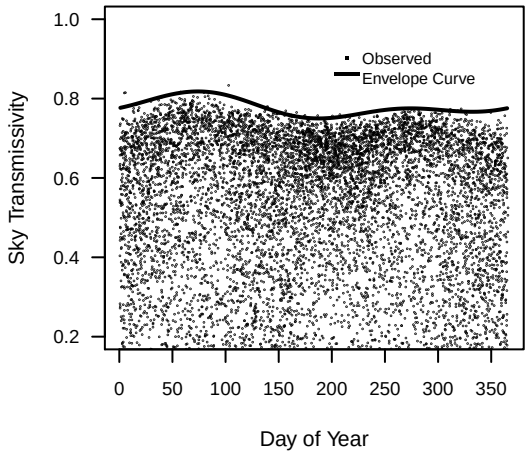


Figure 3.2 Transmissivity if plotted against Day of year (year day) for all available years. The Fourier series fitted envelope curve through the maximums for each year day is considered the fitted CST.

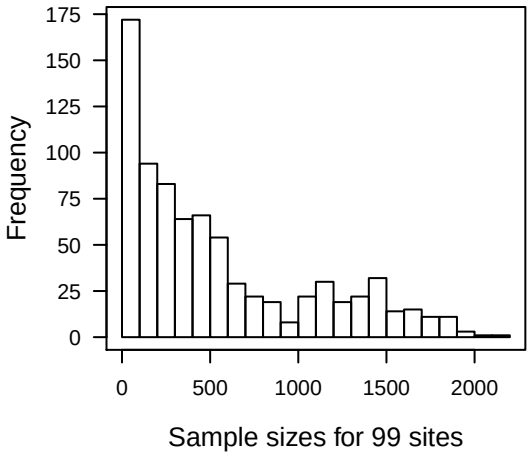


Figure 3.3 A histogram of the sample sizes for the 99 sites. The strata are season and precipitation. Note the high frequency of relatively low sample sizes. This causes problems in fitting models that are limited only to one stratum at a time.

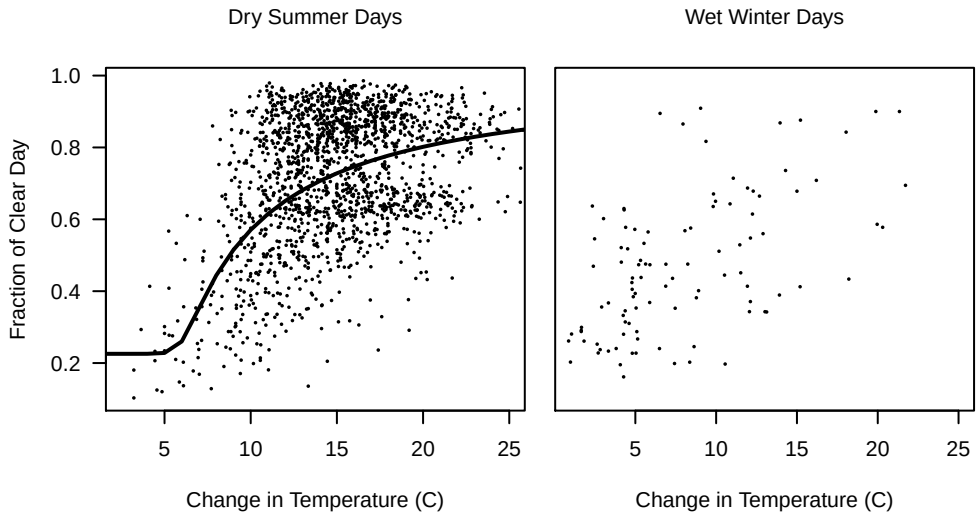


Figure 3.4 The figure on the left shows data from dry summer days at Redfield, SD. For this data set, the sinusoidal curve is shown fitted to the data. On the right is data from wet winter days at the same site. Note the lack of sinusoidal structure to the data. Attempts to fit the data on the right with a four parameter sinusoidal curve led to a variety of possibilities. Non-identifiability of model parameters was an issue.

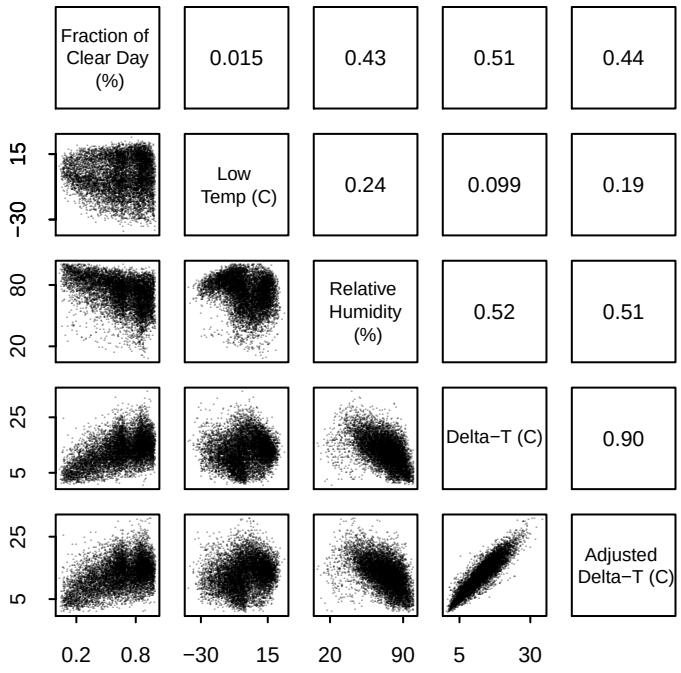


Figure 3.5 A simple correlation matrix showing how FCD is correlated with the independent variables, and how the independent variables are correlated with each other. Numbers in the upper right are the correlation value (rho).

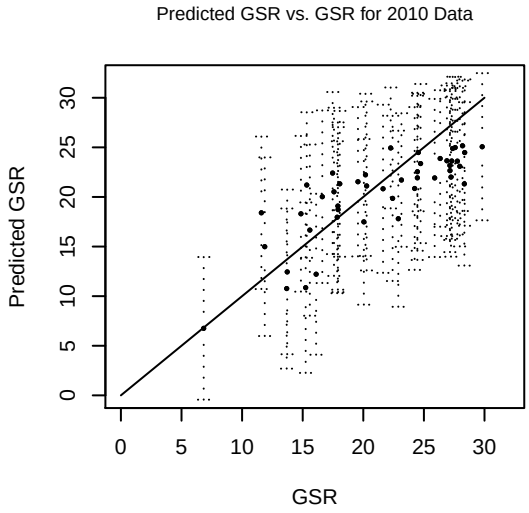


Figure 3.6 Predicted GSR values plotted against observed GSR values with 95% prediction intervals shown. Data used was from dry spring days in Takini, SD.

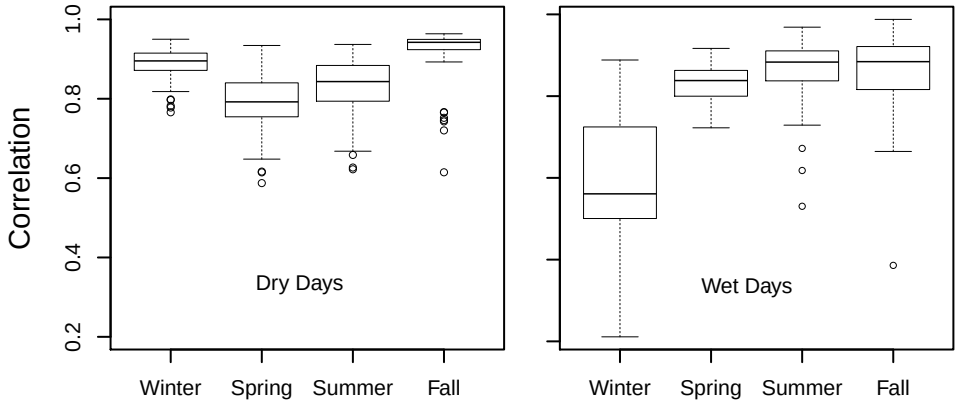


Figure 3.7 Box plots showing the overall distributions of correlations between predicted GSR and observed GSR broken down in seasons and precipitation strata. Dry days shown on the left graph and wet days are shown on the right graph.

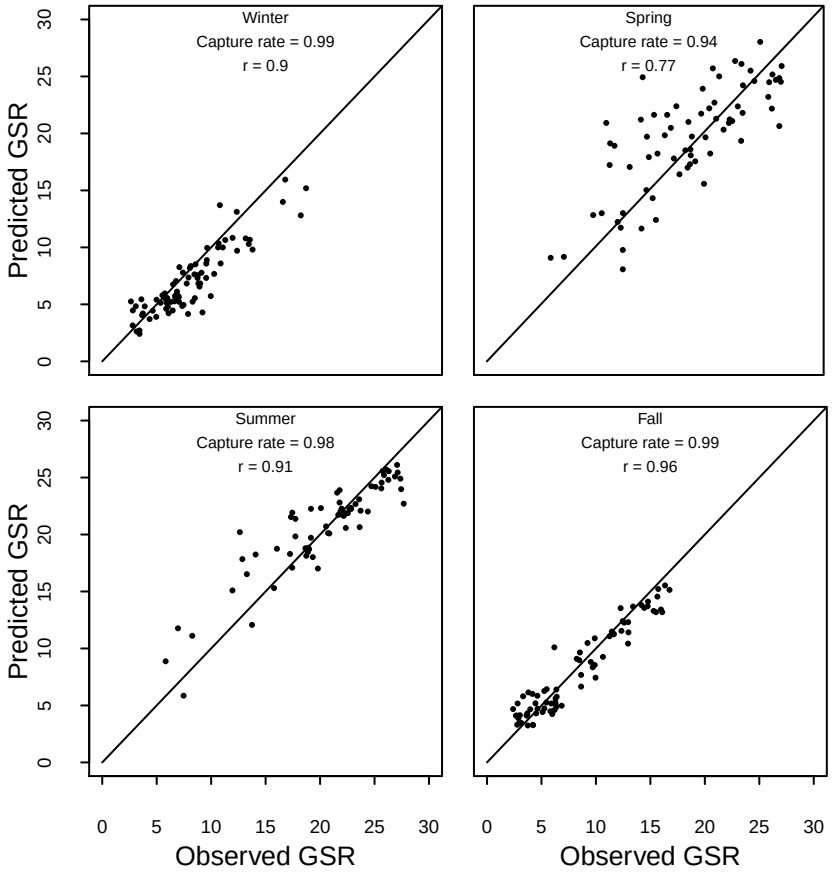


Figure 3.8 Predicted GSR plotted against observed GSR for each of the four seasons using data from dry days. The tendency to underestimate days of high GSR is prevalent throughout all 92 sites, although in general, this is a bigger problem in the summer and spring and less of a problem in winter. Capture rates are the rates at which the 95% prediction interval captured the observed value.

Literature Cited

Allen, R. G. (1997) Self-Calibrating Method for Estimating Solar Radiation from Air Temperature. *Journal of Hydrologic Engineering,* 2**,** 56-67.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Becker, P. & A. P. Smith (1990) Spatial autocorrelation of solar radiation in a tropical moist forest understory. *Agricultural and Forest Meteorology,* 52**,** 373-379.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Cribari-Neto, F. & A. Zeileis (2010) Beta Regression in R. *Journal of Statistical Software,* 34**,** 1-24.

Donatelli, M. & G. S. Campbell. 1998. A simple model to estimate global solar radiation. In *Proceedings of Fifth European Society of Agronomy Congress, vol 2. The Slovac Agric. univ. Nitra, Slovak Republic*, 133 - 134.

Donatelli, M. & V. Marletto. 1994. Estimating surface solar radiation by means of air temperature. In *Proceedings of third annual ESA Congress, September 18-22. Abano, Italy*, 352-353.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008a) Influence diagnostics in beta regression. *Computational Statistics & Data Analysis,* 52**,** 4417-4431.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008b) On beta regression residuals. *Journal of Applied Statistics,* 35**,** 407 - 419.

Evrendilek, F. & C. Ertekin (2008) Assessing solar radiation models using multiple variables over Turkey. *Climate Dynamics,* 31**,** 131-149.

Ferrari, S. & F. Cribari-Neto (2004) Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics,* 31**,** 799 - 815.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Goh, K. L. & A. H. Pooi (1997) Adjustment of Prediction Intervals in Nonlinear Regression. *Biometrical Journal,* 39**,** 719-731.

Gueymard, C. A. & D. R. Myers. 2008. Solar Radiation Measurement: Progress in Radiometry for Improved Modeling. In *Modeling Solar Radiation at the Earth's Surface*, 1-27.

Gueymard, C. A. & D. R. Myers (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy,* 83**,** 171-185.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Keating, K. A., P. J. P. Gogan, J. M. Vore & L. R. Irby (2007) A Simple Solar Radiation Index for Wildlife Habitat Studies. *The Journal of Wildlife Management,* 71**,** 1344-1348.

Liu, D. L. & B. J. Scott (2001) Estimation of solar radiation in Australia from rainfall and temperature observations. *Agricultural and Forest Meteorology,* 106**,** 41-59.

Ospina, R., F. Cribari-Neto & K. L. P. Vasconcellos (2006) Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis,* 51**,** 960-981.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. Vienna, Austria}: R Foundation for Statistical Computing.

Rehman, S. & S. G. Ghori (2000) Spatial estimation of global solar radiation using geostatistics. *Renewable Energy,* 21**,** 583-605.

Richardson, C. W. (1985) Weather simulation for crop management models. *Transactions of the ASAE,* 28**,** 1602.

Rocha, A. & A. Simas (2011) Influence diagnostics in a general class of beta regression models. *TEST,* 20**,** 95-119.

Running, S. W. & P. E. Thornton. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Simas, A. B., W. Barreto-Souza & A. V. Rocha (2010) Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis,* 54**,** 348-366.

Smithson, M. & J. Verkuilen (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods,* 11**,** 54-71.

Spiess, A.-N. & N. Neumeyer (2010) An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology,* 10**,** 6.

Spokas, K. & F. Forcella (2006) Estimating Hourly Incoming Solar Radiation from Limited Meteorological Data. *Weed Science,* 54**,** 182-189.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

Thornton, P. E. & S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

van Dijk, A. I. J. M., A. J. Dolman & E.-D. Schulze (2005) Radiation, temperature, and leaf area explain ecosystem carbon fluxes in boreal and temperate European forests. *Global Biogeochem. Cycles,* 19**,** GB2029.

Zhou, Z.-m. & X.-c. Wang. 2010. Distributed Hydrological Model Based on Topography Trend. In *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on*, 1-4.

MODELING SOLAR RADIATION USING THE SPATIAL AUTO-CORRELATION OF THE DAILY FRACTION OF CLEAR SKY TRANSMISSIVITY

Contribution of Authors and Co-Authors

Chapter 4: Modeling solar radiation using the spatial auto-correlation of the daily fraction of clear sky transmissivity

First Author: Randall S. Mullen

Contributions: I was responsible for all data acquisition, data filtering and quality control. I developed all new concepts and wrote the first draft in its entirety, including all graphics, necessary computer code and analysis.

Co-Authors: Brian L. McGlynn, Lucy A. Marshall

Contributions: Lucy Marshall and Brian McGlynn contributed significant critique and ideas for development of intellectual content within the paper. Lucy Marshall, Megan Higgs and Brian McGlynn edited draft versions of the manuscript. Lucy Marshall and Brian McGlynn were responsible for securing funding.

Manuscript Information

Mullen, R. S., L. A. Marshall, and B. L. McGlynn. 2012 Modeling solar radiation using the spatial auto-correlation of the daily fraction of clear sky transmissivity to *Theoretical and Applied Climatology*

- Journal: *Theoretical and Applied Climatology*
- Status of manuscript (check one)
  - X Prepared for submission to a peer-reviewed journal
  - ____ Officially submitted to a peer-reviewed journal
  - ____ Accepted by a peer-reviewed journal
  - ____ Published in a peer-reviewed journal

- Publisher: Springer link
- Expected date of submission: April 2012

Abstract

Traditional methods for predicting global solar radiation typically analyze a time series of data from one site. Estimated parameters from this site-based model can then be extrapolated make predictions at nearby locations or to locations of similar climate. In this study, we demonstrate a method to obtain robust predictions of global solar radiation using a daily model that incorporates data from many locations on any given day. We compared daily models to traditional site-based models using beta regression and a suite of explanatory climate variables. We inspected the residuals from these daily models for the presence of spatial auto-correlation, and incorporated the auto-correlation using a universal kriging model. Three models are compared, a site-based beta regression model, a daily beta regression model, and a universal kriging model. The site-based model incorporated all available historic data from one site to predict solar radiation at that site. The latter two incorporated data collected at many sites on the same day, to predict solar radiation at one site. Model fit was compared using 1000 permutations of leave-one-out cross-validation. We determine that daily beta regression model outperform site-based models, and that universal kriging outperformed them both. However, the difference between the daily beta regression model and universal kriging was not practically significant. We suggest that incorporating daily data from networks of solar monitoring sites leads to more precise and less biased estimates and yields additional insight into the structure of the spatial auto-correlation of global solar radiation.

Keywords: Spatial auto-correlation, Global Solar radiation, Beta regression, Fraction of clear day, North Dakota, South Dakota, Automated Weather Data Network,

Introduction

Global solar radiation is an essential variable in the study and prediction of ecologic, hydrologic, and geophysical systems. As such, the development of new models of solar radiation and expansion of existing solar monitoring networks are receiving increased attention. Gueymard and Myers (2009) described three levels of stations that typically collect solar radiation data. *Solar monitoring sites* use inexpensive and automated instrumentation to provide local data quickly for a minimal cost. *Conventional long-term measurements* use proven techniques and are generally operated by weather service agencies. *Research sites* are typically developed by atmospheric physicists or climatologists to obtain the highest accuracy possible in order to detect trends or test theoretical solar radiation models. Solar monitoring sites are often attractive to government and research agencies due to their low cost and thus there are growing networks of solar monitoring sites around the United States (Palmer 2011, Horel and Dong 2010) and Europe (Thompson, Ventura and Camarero 2009). For example the University of Wisconsin – Madison operates a network of automatic weather stations (AWS) in Antarctica (Reusch and Alley 2002) that has grown from 6 sites in 1980 to 68 sites in 2011. The expansion of the Automated Weather Data Network (AWDN) (Becker and Smith 1990), one such network operated by the High Plains Regional Climate Center (HPRCC) (Fig 1), has been extensive in the last 25 years. Since 1983, 97 sites have been installed throughout North and South Dakota. The Remote Automated Weather Stations (RAWS) network (Horel and Dong 2010, Reinbold, Roads and Brown 2005) has 2200 sites throughout the United States. The Northwest office of the Bureau of Reclamation

maintains 72 Agricultural Meteorology (AgriMet) stations in Washington, Oregon, Idaho, Western Montana, and Northern California. Each is equipped with global solar radiation monitoring instruments. These examples highlight that the number of remote weather sites has and likely will continue to increase, and with it, the density of solar radiation measurements.

The increasing density of solar monitoring sites makes new models for the prediction of solar radiation possible. Spatial interpolation of solar radiation is not a new concept (Grant et al. 2004, Haberlandt 2007, Lloyd and Atkinson 2004, Aguilar, Herrero and Polo 2010), however, the increasing density of sites allows for better modeling of the spatial auto-correlation structure. Additionally, the increasing density of measured explanatory climate variables means that better predictions of solar radiation can be made in the absence of observed solar radiation data. Traditional site-based models do not take advantage of the additional information that networks of meteorological stations offer such as the spatial auto-correlation structure of solar radiation or other meteorological variables. While a site-based model can take advantage of temporal auto-correlation, this temporal auto-correlation becomes far less relevant than spatial auto-correlation when predicting solar radiation at a site where no observations of global solar radiation exist. That is, the spatial auto-correlation directly relates to the interpolation of solar radiation across the landscape.

Ozone, oxygen, water vapor, pollutants and clouds are responsible for most of the variation in atmospheric attenuation. The interaction between incoming solar radiation and these atmospheric elements is complex, especially on cloudy days. However,

relationships between variables can be exploited to construct useful models of incident solar radiation without the physical processes of these interactions being modeled directly. For example, if one assumes a linear relationship between net radiation and total incoming solar radiation (Chang 1968). Ignoring soil heat flux, net radiation can be partitioned into sensible and latent heat. Sensible heat is that which is responsible for the day time heating of air above night time temperatures. Latent heat is that energy that does not raise the air temperature, but contributes to a phase change of water. If one assumes that the ratio of sensible heat to latent heat is constant, then additional solar radiation results in more heating of air. This then leads to higher differences between daily maximum and minimum temperatures (ΔT) (Campbell and Norman 1998). Additionally, atmospheric water vapor can limit atmospheric transmissivity. Greater atmospheric water vapor is associated with an increased dew point temperature. When an air mass cools to the dew point, the latent heat of condensation can act as a minimum temperature buffer. Therefore, nighty minimum temperatures can often be associated with atmospheric water vapor content. Bristow and Campbell (1984) demonstrated the usefulness of these relationships with site-based models that predict global solar radiation using ΔT as an indicator of atmospheric transmittance.

Global solar radiation must be decomposed into three components in order to relate ΔT to daily fluctuations in attenuation, Extraterrestrial radiation (ETR) is the amount of solar radiation that hits the outside of the atmosphere. Clear sky transmittivity (CST), is the amount of ETR that will reach the Earth's surface on a clear day. Fraction of clear day, (FCD), is the fraction of CST that hits the Earth's surface on any given day.

This decomposition reduces season effects (ETR), elevation and other local effects (CST), and daily fluctuations in atmospheric attenuation (FCD) such that daily changes of the explanatory climate variables can be related to the daily changes in atmospheric attenuation, or FCD. This approach has been applied widely (Fodor and Mika 2011, Bristow and Campbell 1984, Donatelli and Campbell 1998, Thornton, Hasenauer and White 2000, Thornton and Running 1999, Bandyopadhyay et al. 2008, Samani et al. 2011, Ball, Purcell and Carey 2004) and forms the basis for our analyses. FCD is the response variable in all of the models compared in this study. For simplicity, our predictions are not transformed back into global solar radiation. However, comparison results would remain unchanged since ETR and CST are calculated in the same way for all models.

ΔT is not the only climate variable that can be used to predict FCD. At a given temperature, and independent of barometric pressure, dew point is a function of absolute humidity. Night time low temperatures do tend towards dew point, however, for particularly dry nights the amount of heat released when the water vapor undergoes a phase change to frost is minimal, and thus temperatures can drop well below dew point. For this reason, low temperature can be included in any model that predicts FCD (inherent in ΔT models). Humidity and precipitation have also been found to be useful when modeling FCD (Thornton and Running 1999, Thornton et al. 2000) and are included among the four climate variables used in this study to model the process by which solar radiation is allowed to penetrate the atmosphere; ΔT, minimum temperature, relative humidity, and precipitation.

The relationship between the explanatory climate variables and FCD can be modeled in a variety of ways. Traditionally, when more than one explanatory climate variable is used to model FCD, a multi-step approach is followed (Samani et al. 2011, Thornton and Running 1999, Thornton et al. 2000, Donatelli and Campbell 1998), resulting in a univariate non-linear model that has one or more coefficients that have been fitted in a separate regression. The sole explanatory variable in the non-linear model is ΔT, and dew point, precipitation, or humidity can be used to determine the values of the coefficients. More recently, beta regression has been introduced as a suitable statistical alternative for modeling FCD (Mullen, In review). A beta regression model provides an estimate of variance for all FCD predictions (Ferrari and Cribari-Neto 2004, Smithson and Verkuilen 2006, Simas, Barreto-Souza and Rocha 2010, Ferrari and Pinheiro 2011, Ospina, Cribari-Neto and Vasconcellos 2006) such that estimates of uncertainty can be carried through to predictions of global solar radiation (Mullen et al, In review. The beta regression method has been used extensively in ecology (Eskelson et al. 2011, Korhonen et al. 2007) and psychology (Smithson and Verkuilen 2006), and site-based beta regression models have been shown to outperform traditional site-based models (Mullen et al, In review).

The flexibility of beta regression models makes them an obvious choice for comparing site-based models to daily models. A site-based model uses historical data from one site and relates FCD to explanatory climate variables. However, the same model can relate FCD at numerous sites in one day to the explanatory climate variables collected at each of those sites on that day. The only difference in the model formulation

is the input data: long series of daily observations from one site versus observations from one day across a region of interest.

Considering spatial auto-correlation when modeling solar radiation can improve the robustness of model predictions and interpretations. Becker (1990) analyzed spatial auto-correlation of direct and diffuse solar radiation at scales of 1 – 125 m under a tropical canopy and found improved predictions when auto-correlation was included. Evrendilek and Ertekin (2007) inspected monthly averages of global solar radiation in Turkey, however, longer-term averages can lead to smoothing that can change the intensity or relevance of the spatial auto-correlation. They compared standard multiple linear regression to universal kriging to compare estimates of spatial solar radiation. They found that universal kriging outperformed a wide variety of best fit regression models for various regions within Turkey for certain times of the year. Rehman and Ghori (2000) also applied ordinary kriging to predict global solar radiation in Egypt, and found that it appropriately represented solar radiation spatial characteristics. Residual kriging was further shown to improve monthly estimates of solar radiation in other studies (Alsamamra et al. 2009). These studies did not model FCD directly; however, they provide ample evidence that incorporating spatial auto-correlation can improve predictions of FCD. Universal kriging allows the same explanatory variables used in the beta regression models (both the site-based and daily) to model the trend simultaneously with spatial auto-correlation.

We investigate the utility of daily beta regression models and also daily universal kriging for predicting FCD by comparing predictions of FCD from these two models to

those obtained using a site-based beta regression model. Both the daily and site-based beta regression models use the relationship between the explanatory variables and solar radiation. The universal kriging model does the same, but additionally makes direct use of the spatial auto-correlation structure of the residuals of solar radiation after accounting for trend to improve predictions. Both the daily beta regression model and the universal kriging model can be used to infill gaps in data at weather station sites when solar radiation values are not obtained due to equipment failure. These types of failures are common (e.g., power outage, instrumentation fails or becomes dirty, recording devices are corrupted) and can substantially impact the usefulness of long-term data sets. More importantly, both models can also be used to predict solar radiation at locations where no solar radiation data are collected but where data for the more commonly measured explanatory climate variables are available. We further conducted simulations to determine the role of weather station density on the error of the daily beta regression model and the universal kriging model.

Methods

In this study, we compare two fundamentally different approaches for predicting FCD at a location when measurements of solar radiation are not available but necessary explanatory climate variable measurements are available. The first is the traditional approach (Ball et al. 2004, Bristow and Campbell 1984, Hargreaves and Samani 1982, Samani et al. 2011, White et al. 1998, Running and Thornton 1999, Thornton et al. 2000) that estimates FCD on a given day at a given site using data collected at that site over the

time period of three or more years (referred to here as site-based models). The second approach predicts FCD on a given day using data collected at other sites on that day (referred to as daily models). Two different daily models are tested, a beta regression model that simply uses data from one day and many sites, and a universal kriging model that uses the same data as the daily beta regression model but also incorporates spatial auto-correlation. Aside from the obviously advantage of being able to use the spatial auto-correlation, these daily models capitalize on the fact that the relationship between FCD and the explanatory climate variables can vary day to day. This information is used implicitly in the daily models.

Data and Site Description

We implement our methods for a study area that encompasses North and South Dakota in the north-central United States. The predominate climate in these states is distinctly continental with very cold winters and hot semi-humid summers. However, the western part of North Dakota is considered semi-arid. The highest recorded temperature in either state is 49 ^{o}C and the coldest is -51 ^{o}C. The average annual precipitation ranges from 35 to 75 cm throughout the study area. The highest point in either state is 2208 m, and the lowest is 229 m.

In 1983, the AWDN network installed three sites that had solar radiation sensors in South Dakota. By the year 2000, there were 49 sites in North and South Dakota, and two in eastern Montana very near the North Dakota border. By 2011 there were 99 sites throughout North and South Dakota (including the two in eastern Montana) with more

site installations planned (Figure 4.1). North and South Dakota are approximately 382,843 km^2, yielding a density of 2.5E-4 sites/km^2.

Data from these 99 AWDN sites (Figure 4.1) were inspected for quality and quantity (length of data series and amount of missing data). Standard weather variables collected at the AWDN sites include (but are not limited to) daily high temperature, daily low temperature, relative humidity, precipitation, evaporation, soil temperature, snowfall, snow depth, wind speed and wind direction, . Chosen for comparison were 93 sites inside of North and South Dakota and two in Montana very near the border of North Dakota. All data that were flagged as bad, missing, or imputed using regression were deleted. The total area that can be reasonably inferred as coverage is approximately 382,843 km^2, yielding a density of 2.5E-4 sites/km^2. All data used for model comparisons, including the explanatory variables minimum temperature, maximum temperature, relative humidity and precipitation, were collected from these sites.

Observed Fraction of Clear Day

Extraterrestrial radiation (ETR) was determined for a given site using geographical location, time of day and time of year (e.g. Gates, 1980). For each site on each day of the year, the percentage of solar radiation that would reach the Earth's surface on a clear day was estimated. This is referred to as clear sky transmittivity (CST), and can be determined empirically (Thornton and Running 1999) using historical data or can be modeled using Fourier series (Fodor and Mika 2011). Mechanistic models have additionally been developed for predicting maximum possible daily global solar radiation (Meek 1997) where no previous data exists but these were not considered in this study.

CST was modeled using a Fourier polynomial series and empirically obtained maximum values for each day. The Fourier polynomial series had the form;

$$CST(yearday) = \left[a_1 \cos(2x) + b_1 \sin(2x)\right] + \left[a_2 \cos(4x) + b_2 \sin(4x)\right]$$

$$x = \pi(yearday / 365)$$

Six sites had an insufficient amount of data as to prevent the fitting of a Fourier series necessary for the site-based modeling approach and were dropped entirely from the analysis, leaving 92 sites. Measured daily global solar radiation (GSR) data was divided by CST to determine the fraction of CST that occurred on a given day. This was then specified as the observed fraction of clear day, (FCD).

Model Comparison

The three models are compared using 1000 permutations of leave-one-out cross-validation. Each permutation removes the observed value for one randomly chosen combination of day and site from 2010. Each model is used to make a prediction for that day – site combination and the predicted value is compared to the observed value in order to test the predictive abilities of each model in the absence of measured data. For this data set, one complete years' worth of test data would involve 33,580 combinations of day and site (365 days x 92 sites = 33,850 observed FCD response values). A little over 10 % of the 2010 data were missing so about 30,000 response values remained. The cross validation was performed on 1000 randomly chosen site -day combinations from these 30,000 available response values. This reduction from 30,000 to 1,000 randomly chosen site-day combinations reduced computational requirements. After these two predictions of FCD were obtained (one from the daily model and one from the site-based model) for

a given site on a given day, they were compared to the observed value of FCD held out from analysis. Root mean squared error (RSME), mean signed deviance (MSD), and mean absolute error (MAE), were used to compare the results for 1000 randomly selected day-site combinations.

The minimum number of years of operation for the 92 sites left in this analysis was three. The maximum number of years available at any one site was 27. All sites were currently operational as of December 31, 2010.

Beta Regression Models

Beta regression is a statistical model used for modeling beta distributed dependent variables (Ferrari and Cribari-Neto 2004). The beta distribution is a continuous probability distribution defined on the interval between 0 and 1 Beta regression is thus appropriate for rates, proportions, and other continuous variables constrained in the standard unit interval (0,1) with a mean that can be related to a set of regressors with estimable coefficients and a link function.. Ferrari and Cribari-Neto (2004) reparameterized the traditional form of the beta distribution by setting $u = p/(p+q)$ and $\phi = p+q$. This yields:

$$f(y;u,\phi) = \frac{\Gamma(\phi)}{\Gamma(u\phi)\Gamma((1-u)\phi)} y^{u\phi-1}(1-y)^{(1-u)\phi-1}, 0 < y < 1$$

where $\Gamma(\cdot)$ is the gamma function, $0 < u < 1$ and $\phi > 0$. The parameter ϕ is the precision parameter where for fixed u, a larger ϕ gives a smaller variance. A beta distributed variable is denoted as $y \sim \beta(u,\phi)$. The expected value of y is u, or $E(y) = u$. In matrix notation, beta regression is then represented as;

$$g(u_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} = \eta_i$$

where $\beta = (\beta_1, ..., \beta_k)^T$ is a k x 1 vector of unknown regression parameters, $x_i = (x_{i1}, ..., x_{ik})^T$ is a vector of k regressors, or independent variables, and η_i is a linear predictor. This is a naturally heteroscedastic function with;

$$Var(y_i) = \frac{u_i(1-u_i)}{1+\phi}$$

Beta regression is especially effective for modeling bounded environmental variables (Eskelson et al. 2011). When a dependent variable, such as FCD, is bounded between 0 and 1 standard regression techniques are likely inappropriate. Assumptions of normality are usually incorrect because truncation of the response value makes even an approximate normal distribution unlikely. Almost by definition these variables display a large amount of heteroscedasticity with more variation around the variable mean and less close to the variable boundaries. Like most proportion data, FCD distributions tend to be asymmetric, which leads to issues with confidence intervals and hypothesis testing. Beta regression addresses all of these issues (Ferrari and Cribari-Neto 2004, Ferrari and Pinheiro 2011). Further, functions to perform beta regression are now readily available in popular software programs (Cribrali-Neto, 2010). For the interested reader, a growing body of literature on beta regression is available (Branscum, Johnson and Thurmond 2007, Chien 2010, Cribari-Neto and Zeileis 2010, Eskelson et al. 2011, Espinheira, Ferrari and Cribari-Neto 2008a, Espinheira, Ferrari and Cribari-Neto 2008b, Ferrari and Cribari-Neto 2004, Ferrari and Pinheiro 2011, Hunger, Baumert and Holle 2011, Matsuda et al. 2006, Ospina et al. 2006, Rocha and Simas 2011, Simas et al. 2010, Smithson and

Verkuilen 2006) that includes overinflated beta regression (Ospina and Ferrari 2010, Ospina and Ferrari 2011), mixed beta regression models, (Grün, Kosmidis and Zeileis 2011) and autoregressive beta regression models (Rocha and Cribari-Neto 2009).

Site-Based Models

Site-based models were fit using beta regression (Mullen et al, 2012). All data collected at the specific site before the chosen day were used for model fitting. Data that were collected after the randomly selected test day were omitted in order to simulate a real world situation. Note that CST from nearby sites could be extrapolated to these sites but this does lead to less precise results for the site-based models(Fodor and Mika 2011)(Mullen et al, 2012). Relative humidity, daily low temperature, change in temperature, and precipitation depth were the explanatory variables used for fitting the beta regression models. These variables are commonly collected at weather stations and have been shown to adequately predict solar radiation in the absence of measured solar radiation data (Mullen et al, 2012). Day of year was additionally included in the model using the sine and cosine components to represent seasonal variability (Mullen et al, 2012). These additions were simply to negate the need to subset the data by month and / or season, as past studies have demonstrated that subsetting improved models fit (Fodor and Mika 2011) by addressing the seasonal interactions of temperature and water vapor. It was shown that by including day of year in the model as cyclical data (thus using the sine and cosine components), stratification was unnecessary (Mullen, In review).

Daily Models

Daily FCD models were fitted using available data from the 92 sites described above for a specific day. Not every randomly selected day – site combination had complete data for the remaining 91 sites used for prediction. However, days with missing data from the sites used for prediction were retained in the analysis to simulate a real world situation whereby a monitoring network may have data gaps or missing data. The explanatory climate variables selected were the same as those used for the site-based beta regression model, but taken from each station on the specified test day as opposed to being taken from past days at the test station. Day of year was not added to the daily model since this model is using data from one day. Beta regression for both the daily model and the site-based model was carried out in R (R Development Core Team 2009) using the betareg library (Cribari-Neto and Zeileis 2010).

We looked for potential spatial auto-correlation in the daily model for values estimated in the year 2010. Variograms were constructed for the daily model residuals obtained from the 365 possible daily beta regression models from 2010. Visual inspection of the variograms was done in order to determine if spatial auto-correlation existed (Banerjee, Carlin and Gelfand 2004). The existence of spatial auto-correlation (Figure 4.2) suggested incorporating the auto-correlation into the daily model in order to obtain more precise and unbiased estimates of FCD.

Universal Kriging

Universal kriging is a geostatistical technique that is used to make predictions across a region using measured values with known locations. . It can be written as (Schabenberger, Gotway 2005);

$$\hat{Z}(s_0) = X(s_0)\hat{\beta}_{gls} + \Sigma_{0Z}\Sigma_{ZZ}^{-1}\left[Z(s) - X(s)\hat{\beta}_{gls}\right]$$

where X(s_0) is a (p x 1) vector of explanatory variables, Σ_{zz} is the n x n variance-covariance matrix of the data, Σ_{oz} is the r x n variance- covariance matrix between the data and the unobservables, Z(s) is data, or the observed values, and $\hat{\beta}_{gls}$ is;

$$\left[X(s)'\Sigma^{-1}X(s)\right]^{-1}X'\Sigma^{-1}Z(s)$$

In our case, for each location that a prediction of FCD is desired, the necessary explanatory climate variables must be present. The universal kriging model uses the same explanatory variables as the daily beta regression model. However, the response variable was logit transformed and treated as normally distributed. These explanatory variables are used to describe the underlying trend in the data. The trend is simultaneously estimated as part of the prediction process, and not made explicit in the results (Bailey and Gatrell 1995) In order to perform universal kriging, starting values for computational methods can be found by first fitting a model to the spatial auto-correlation structure of the residuals of an initial linear model with just the explanatory variables included. Traditionally, variograms are fit by eye, and the need for analytical methods distinguishing a better fit are not necessary (Banerjee et al. 2004). However, given the high number of models that needed to be fit, this process was automated and several

candidate variogram models were analyzed for each day. Tested variogram models included the spherical, exponential, Gaussian, and Matern models (Banerjee et al. 2004). The maximum distance used for fitting variograms was set to 250 km. This was done for 2 reasons; 1) visual inspection of a sub-sample of days indicated 250 km was sufficient to represent the spatial auto-correlation for most days, and 2) it allowed for better automated selection and fitting of variograms. The best fitting variogram model for each day was determined by weighted least-squares, and then used in the universal kriging model to predict FCD and compared to the daily beta regression model. There is no guarantee that the variogram model (i.e. spherical, exponential, etc) initially chosen is the best fit for the final universal kriging model. Further model selection criteria could be applied. However, for the automated approach used described herein, only the variogram initially chosen was used.

Effect of a Less Dense Monitoring Network

A simulation was performed where stations were randomly removed from the analysis described above comparing predictions made using a daily beta regression model to estimates obtained using a universal kriging approach. The number of stations left in the analysis was varied from 50 to 98. Stations were randomly chosen for removal for each simulation, such that the pattern of missing stations varied for each simulation. The intent of this simulation was to provide network managers with a demonstration of the effect of station density on model performance, realizing of course that stations are implemented for a variety of climate measurements. The RMSE was plotted against the number of stations used in order to demonstrate how station density may affect the

respective spatial and non-spatial models. A non-zero slope would suggest a relationship between the number of sites and RMSE, and, that additional sites lead to more precise estimates.

Results

Model Comparison

The universal kriging model had the lowest RMSE (0.0883) and MAE (0.0610), although, the daily beta regression model had only a slightly higher RMSE (0.0962) and MAE (0.0697. The site-based model had a substantially higher RMSE (0.1551) and MAE (0.1225) (Table 4.1). The daily beta regression model had the lowest, MSD (0.0002). The universal kriging model had a higher, although still small) MSD (0.0063). The site-based model had the highest MSD (-0.0121) (Table 4.1). A 15% difference in FCD can amount to a difference of about 1.5 MJ per day in the winter when total MJ per day ranges from less than 1 MJ to about 7.5 MJ per day. Similarly, a 15% difference in FCD can amount to a difference of about 5 MJ per day in the summer, when total MJ per day ranges from less than 5 to almost 30 MJ per day.

The site-based model tended to overestimate FCD in the low to mid-range, (0.2 – 0.5), and underestimate FCD in the upper range (> 0.8) (Figure 4.3). The daily beta regression model did this to a much lesser degree, and the universal kriging model did not seem to have any change in bias across the range of FCD. The MSD, generally thought of as an indicator of bias, was smaller for the daily beta regression, however, the difference (0.0061) is small, and not considered practically meaningful. Furthermore, MSD refers to

the entire range of FCD [0, 1], and does not account for bias that is specific to isolated certain circumstance, (i.e. low FCD, high FCD).

In all cases, the variance of the prediction was smaller for the daily beta regression model compared to the variance for the kriged model. The mean of the differences was 0.619 with a high of 0.956 and a low of 0.484. The mean percent improvement was 95% (high was 97%, low was 92%)

Effect of a Less Dense Monitoring Network

The error (RMSE, MAE, MSD) of the daily beta regression model without spatial auto-correlation did not display any trends as sites were removed from the analysis (Figure 4.1) RMSE and MAE for universal kriging had non-zero ($p < 0.05$) negative slopes when plotted against number of sites used. . This is intuitive, since the universal kriging process incorporates spatial auto-correlation, and as sites are removed, there is a loss of information. When regressing MAE of kriging against the number of sites, as number of sites increases, MAE decreases (slope = -0.0005, p-value < 0.00001, $r^2 = 0.005$) (Figure 4.4). When regressing the RMSE for kriging against the number of sites, as the number of sites increases, the RMSE decreases (slope = -0.0002, p-value < 0.00001, $r^2 = 0.0027$) (Figure 4.5). While these results are statistically significant ($p < 0.05$), the actual amount of improvement is very small and would likely not be considered practically significant.

Discussion

When comparing the FCD predictions from the daily beta regression model to those from the site-based beta regression models, the daily model outperformed the site-based model in all categories. We are not suggesting that the daily model can take the place of the site-based models in all cases where predictions of global solar radiation are needed. Indeed, the original motivation for developing these site-based models was that they could be used in lieu of solar radiation measurements (Bristow and Campbell 1984) given the extreme dearth of solar radiation sensors (Thornton and Running 1999). There are still places around the world and likely in North America where networks of solar radiation monitoring sites do not exist, and will likely not exist in the near future, thus making it impossible to implement the daily model. . Rather, our intent was to demonstrate that changing the general approach of predicting fraction of clear day, and subsequently, global solar radiation can be helpful and insightful, and that final predictions using daily models can be more precise and less biased than traditional site-based methods (Figure 4.3). In both the daily beta regression model, and the universal kriging model, RMSE, MAE and MSD were smaller than for the site-based model (Table 4.1).

Residuals from the daily beta regression model were inspected for spatial auto-correlation. Spatial auto-correlation was found, but with varying degrees of magnitude (Figure 4.2). Since each model was individually fitted with a different theoretical variogram, it is difficult to draw broad conclusions about seasonal variation in spatial structure, or variation on certain types of days, (e.g., dry days versus wet days). However,

sufficient auto correlation was present to suggest that modeling the spatial dependency of FCD could improve FCD predictions, as well as variance estimates. The universal kriging model is able to incorporate both the underlying trend and the spatial auto-correlation structure simultaneously, thus is a natural model to utilize in this case.

When comparing the daily beta regression to a universal kriging, it was shown that while predictions improved using the universal kriging model, the improvement was not nearly as much as moving from a site-based model to a daily model. Additionally, the daily beta regression model had lower prediction variances than the universal kriging model. In other words, for the data used herein, the simpler daily beta regression model would probably suffice for most research needs for infilling missing data or predicting at a location that has adequate explanatory climate variables but does not have solar radiation sensors. However, there are advantages to using the universal kriging model, we found less overestimation in the middle values of FCD, and less underestimation in the high values (Figure 4.3), and ultimately, use of the final predictions will probably dictate which level of modeling is necessary. For prediction purposes, it is thought that the explanatory variables explained sufficient variation in FCD such that incorporating spatial auto-correlation did not improve RMSE, MAE and MSD in a practically meaningful way. It is important to consider that the explanatory variables themselves are auto-correlated, therefore, using them to predict FCD could suffice in modeling the spatial auto-correlation of FCD. Another consideration is that if predicting for a limited number of days, then individual fitting of variograms would undoubtedly improve the universal kriging model. For the leave-one-out cross-validation approach used herein, it

was not considered plausible to fit 1000 variograms by eye. Also, it is likely that future studies will have to make a large number of predictions, and the desire was to provide examples of realistically derived models that depend on auto-fitting variograms. Two previous studies have indicated that incorporation of spatial auto-correlation improves models predictions (Evrendilek and Ertekin 2007, Rehman and Ghori 2000). However, both of these studies used monthly averages and modeled daily cumulative global solar radiation directly. Decomposing GSR into the three parts, ETR, CST and FCD, allows the daily fluctuations of atmospheric attenuation (FCD) to be directly related to daily changes in the explanatory variables. ETR and CST are highly spatially auto-correlated, but those components of GSR are well described independently of the daily changes in climate variables, therefore should be left out of any model that relates daily fluctuations in GSR to daily climate variables.

The study area was originally chosen because of the relative homogeneity of the landscape in North and South Dakota. It is thought that as one moved into areas of higher topographical variety, (e.g., the intermountain west) where nearby locations are less related, then modeling spatial-autocorrelation would require denser networks. If the range of the spatial auto-correlation decreases, then in general, sites need to be closer to one another to take advantage of the auto-correlation in a modeling framework. An example of this would be two valleys separated by a narrow mountain range, (not uncommon in places through the Rocky Mountain West). While the Euclidian distance between the two sites may be small, the daily auto-correlation of FCD may also be, thus reducing the effectiveness modeling the spatial auto-correlation. Another way to think about this issue

is that, if the minimum distance between sites exceeds the range of the auto-correlation structure, the response variables start to become independent. Further investigation would be needed to ascertain how much of an issue this would be in the intermountain west and how dense weather stations would have to be to overcome it.

Conclusion

We have presented two different daily model approaches, one using the beta regression and one using universal kriging, that have a lower RMSE, MAE and MSD than the site-based approach that has been historically used for modeling FCD (Ball et al. 2004, Bristow and Campbell 1984, White et al. 1998, Samani et al. 2011, Thornton and Running 1999, Thornton et al. 2000, Hasenauer et al. 2003)(Mullen, in review). The approaches discussed here are appropriate for infilling missing data from networks of solar radiation monitoring sites, or for predicting fraction of clear day, and subsequently global solar radiation, at a site that does not have solar radiation monitoring sensors. Both of these models assume a suite of explanatory climate variables are present, however, as networks of remote weather stations increase in both number and size (Gueymard and Myers 2009, Horel and Dong 2010, Palmer 2011, Reusch and Alley 2002), these explanatory climate variables will become more readily available as well. It was shown that for prediction purposes, it might be sufficient to leave spatial auto-correlation out of the model, (Table 4.1), however, when comparing predicted FCD to observe FCD using both the daily beta regression model and the universal kriging model, the universal kriging model did show less bias in the mid and high range of FCD values. Which model

would suffice in a future studies would probably depend on the use of the predictions. Future work should focus on what the best applicable model is across a variety of situations, as well as focusing on the integration of both spatial and temporal auto-correlation.

Acknowledgements

The authors would like to thank the National Resource Conservation Service for funding this project and the High Plains Regional Climate center for collecting this data and making it available for purchase.

Tables

Table 4.1 RMSE, MAE and MSD for the site-based beta regression model, the daily beta regression model and universal kriging. The kriging model slightly outperformed the daily beta regression model with respect to RMSE and MAE. The Daily beta regression had a lower MSD. Both outperformed the site specific beta regression model in all three criteria.

	RMSE	**MAE**	**MSD**
	Units are Fraction of Clear Day		
Site based beta regression model	0.1551	0.1225	-0.0121
Daily beta regression model	0.0962	0.0697	0.0002
Daily Universal kriging	0.0883	0.0610	0.0063

Figures

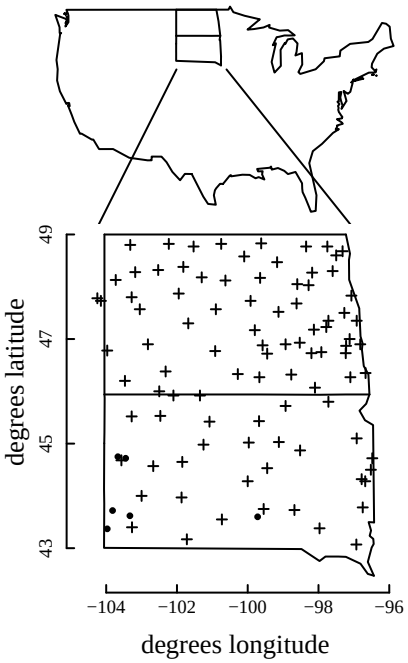


Figure 4.1 Shown is the Montana, North Dakota, and South Dakota sites of the AWDN network. The bulleted sites are the sites that did not have enough data to create valid CST Fourier series. For the daily model, CST from the nearest site with enough data was used. Top figure shows the location of North and South Dakota in the United States.

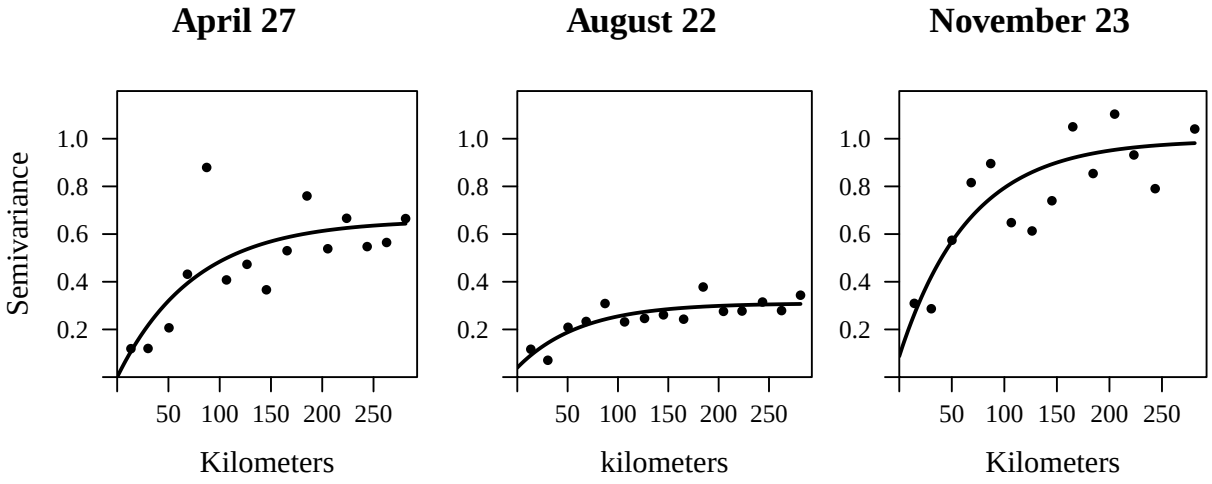


Figure 4.2 Three example variograms are shown. The shape of each fitted variogram indicates the presence of spatial auto-correlation.

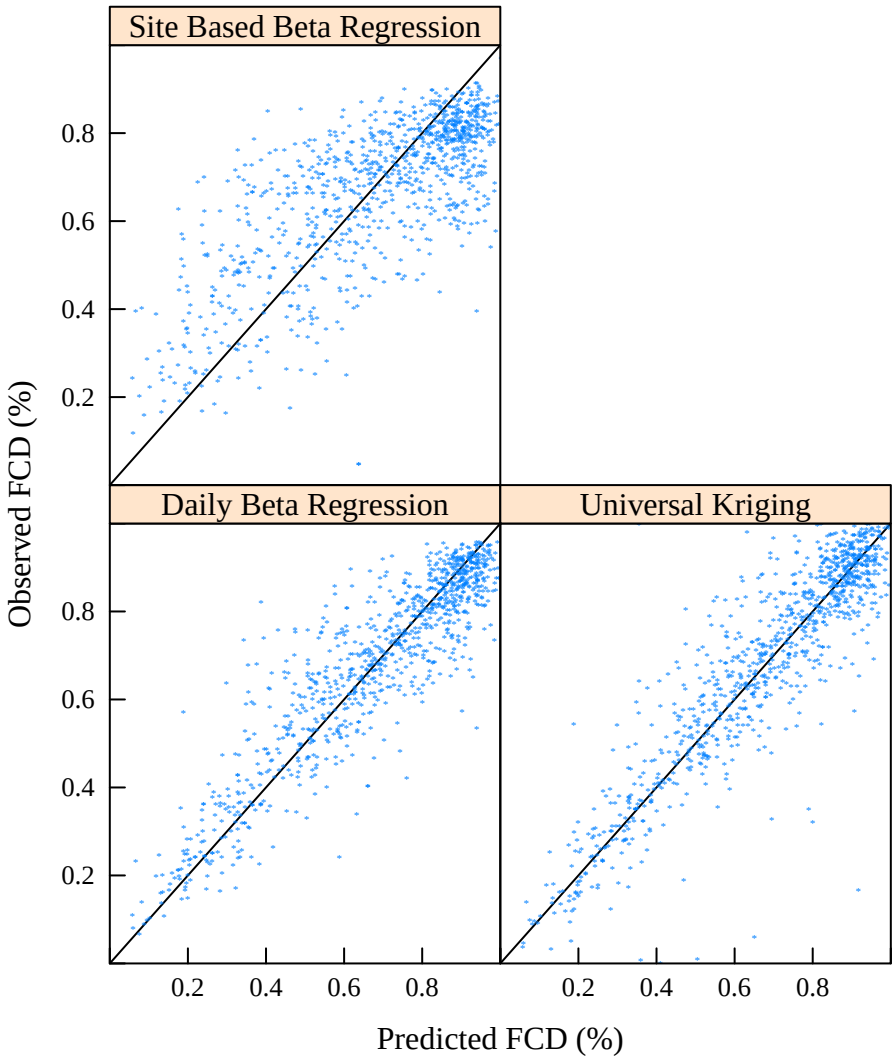


Figure 4.3 Predicted FCD vs. observed FCD for the 1000 randomly chosen day – site combinations that were used for the leave-one-out cross-validation. The daily beta regression model and the universal kriging model both outperformed the site based model. The daily beta regression model tends to overestimate for the lower to middle FCD values and underestimate higher FCD values, but not as badly as the site based model. The universal kriging model does not have any pronounced predictionbias based on FCD value.

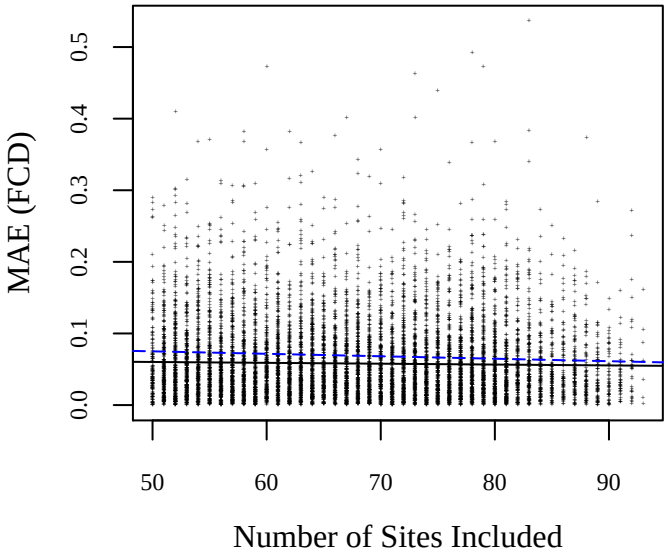


Figure 4.4 MAE for the universal kriging model is plotted with the dashed line. Its slight negative slope indicates some reduction in average MAE as the number of sites is increased. This is to be expected, however, the amount of increase from 50 to 99 sites is not substantial (slope = -0.0005, p-value < 0.00001, $r^2 = 0.005$). MAE for the daily beta regression model is the solid line with a slope of zero. Since the beta regression model does not take advantage of the spatial auto-correlation structure, MAE does not decrease when the number of sites is increased from 50 to 99.

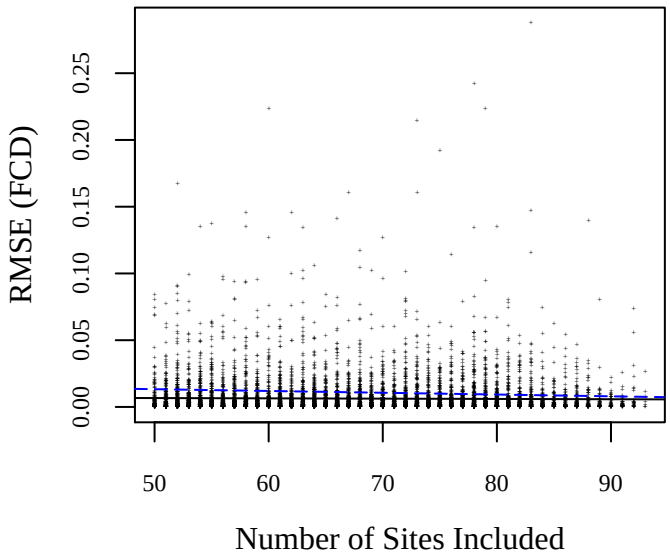


Figure 4.5 RMSE for the universal kriging model is plotted with the dashed line. Its slight negative slope indicates some reduction in RMSE has the number of sites is increased. This is to be expected, however, the amount of increase from 50 to 99 sites is not substantial (slope = -0.0002, p-value < 0.00001, $r^2 = 0.0027$). RMSE for the daily beta regression model is the solid line with a slope of zero. Since the beta regression model does not take advantage of the spatial auto-correlation structure, RMSE does not decrease when the number of sites is increased from 50 to 99.

Literature Cited

Aguilar, C., J. Herrero & M. J. Polo (2010) Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrology & Earth System Sciences Discussions,* 7**,** 2373-2411.

Alsamamra, H., J. A. Ruiz-Arias, D. Pozo-Vázquez & J. Tovar-Pescador (2009) A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain. *Agricultural and Forest Meteorology,* 149**,** 1343-1357.

Bailey, T. C. & A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. New York: Prentice Hall.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Bandyopadhyay, A., A. Bhadra, N. S. Raghuwanshi & R. Singh (2008) Estimation of monthly solar radiation from measured air temperature extremes. *Agricultural and Forest Meteorology,* 148**,** 1707-1718.

Banerjee, S., B. P. Carlin & A. E. Gelfand. 2004. *Preface*. Chapman and Hall/CRC.

Becker, P. & A. P. Smith (1990) Spatial autocorrelation of solar radiation in a tropical moist forest understory. *Agricultural and Forest Meteorology,* 52**,** 373-379.

Branscum, A. J., W. O. Johnson & M. C. Thurmond (2007) Bayesian beta regression: Applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics,* 49**,** 287-301.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Campbell, G. S. & J. M. Norman. 1998. *An Introduction to Environmental Biophysics*. New York: Springer.

Chang, J.-h. 1968. *Climate and Agriculture*. Chicago: Aldine Publishing.

Chien, L.-C. (2010) Diagnostic plots in beta-regression models. *Journal of Applied Statistics*.

Cribari-Neto, F. & A. Zeileis (2010) Beta Regression in R. *Journal of Statistical Software,* 34**,** 1-24.

Donatelli, M. & G. S. Campbell. 1998. A simple model to estimate global solar radiation. In *Proceedings of Fifth European Society of Agronomy Congress, vol 2. The Slovac Agric. univ. Nitra, Slovak Republic*, 133 - 134.

Eskelson, B. N. I., L. Madsen, J. C. Hagar & H. Temesgen (2011) Estimating Riparian Understory Vegetation Cover with Beta Regression and Copula Models. *Forest Science,* 57**,** 212-221.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008a) Influence diagnostics in beta regression. *Computational Statistics & Data Analysis,* 52**,** 4417-4431.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008b) On beta regression residuals. *Journal of Applied Statistics,* 35**,** 407 - 419.

Evrendilek, F. & C. Ertekin (2007) Statistical Modeling of Spatio-Temporal Variability in Monthly Average Daily Solar Radiation over Turkey. *Sensors,* 7**,** 2763-2778.

Ferrari, S. & F. Cribari-Neto (2004) Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics,* 31**,** 799 - 815.

Ferrari, S. L. P. & E. C. Pinheiro (2011) Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation,* 81**,** 431 - 443.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Grant, R. H., S. E. Hollinger, K. G. Hubbard, G. Hoogenboom & R. L. Vanderlip (2004) Ability to predict daily solar radiation values from interpolated climate records for use in crop simulation models. *Agricultural and Forest Meteorology,* 127**,** 65-75.

Grün, B., I. Kosmidis & A. Zeileis. 2011. Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Faculty of Economics and Statistics, University of Innsbruck.

Gueymard, C. A. & D. R. Myers (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy,* 83**,** 171-185.

Haberlandt, U. (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology,* 332**,** 144-157.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Hasenauer, H., K. Merganicova, R. Petritsch, S. A. Pietsch & P. E. Thornton (2003) Validating daily climate interpolations over complex terrain in Austria. *Agricultural and Forest Meteorology,* 119**,** 87-107.

Horel, J. D. & X. Dong (2010) An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology,* 49**,** 1563-1578.

Hunger, M., J. Baumert & R. Holle (2011) Analysis of SF-6D Index Data: Is Beta Regression Appropriate? *Value in Health,* 14**,** 759-767.

Korhonen, L., K. T. Korhonen, P. Stenberg, M. Maltamo & M. Rautiainen (2007) Local models for forest canopy cover with beta regression. *silva Fennica,* 41**,** 671-685.

Lloyd, C. D. & P. M. Atkinson (2004) Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. *International Journal of Applied Earth Observation and Geoinformation,* 5**,** 293-305.

Matsuda, Y., K. Fujita, Y. Ageta & A. Sakai (2006) Estimation of atmospheric transmissivity of solar radiation from precipitation in the Himalaya and the Tibetan Plateau. *Annals of Glaciology,* 43**,** 344-350.

Meek, D. W. (1997) Estimation of maximum possible daily global solar radiation. *Agricultural and Forest Meteorology,* 87**,** 223-241.

Ospina, R., F. Cribari-Neto & K. L. P. Vasconcellos (2006) Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis,* 51**,** 960-981.

Ospina, R. & S. Ferrari (2010) Inflated beta distributions. *Statistical Papers,* 51**,** 111-126.

Ospina, R. & S. L. P. Ferrari (2011) A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. Vienna, Austria}: R Foundation for Statistical Computing.

Rehman, S. & S. G. Ghori (2000) Spatial estimation of global solar radiation using geostatistics. *Renewable Energy,* 21**,** 583-605.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Reusch, D. B. & R. B. Alley (2002) Automatic Weather Stations and Artificial Neural Networks: Improving the Instrumental Record in West Antarctica. *Monthly Weather Review,* 130.

Rocha, A. & F. Cribari-Neto (2009) Beta autoregressive moving average models. *TEST,* 18**,** 529-545.

Rocha, A. & A. Simas (2011) Influence diagnostics in a general class of beta regression models. *TEST,* 20**,** 95-119.

Running, S. W. & P. E. Thornton. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Simas, A. B., W. Barreto-Souza & A. V. Rocha (2010) Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis,* 54**,** 348-366.

Smithson, M. & J. Verkuilen (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods,* 11**,** 54-71.

Thompson, R., M. Ventura & L. Camarero (2009) On the climate and weather of mountain and sub-arctic lakes in Europe and their susceptibility to future climate change. *Freshwater Biology,* 54**,** 2433-2451.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

Thornton, P. E. & S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

White, J. D., S. W. Running, P. E. Thornton, R. E. Keane, K. C. Ryan, D. B. Fagre & C. H. Key (1998) Assessing simulated ecosystem processes for climate variability research at Glacier National Park. *Ecological Applications,* 8**,** 805.

EVALUATING A BETA REGRESSION APPROACH FOR ESTIMATING FRACTION OF CLEAR SKY TRANSMISSIVITY IN MOUNTAINOUS TERRAIN

Contribution of Authors and Co-Authors

Chapter 5: Evaluating a beta regression approach for estimating fraction of clear sky transmissivity in mountainous terrain

First Author: Randall S. Mullen

Contributions: I was responsible for all data acquisition, data filtering and quality control. I wrote the first draft in its entirety, including all graphics, necessary computer code and analysis.

Co-Authors: Brian L. McGlynn, Lucy A. Marshall

Contributions: Lucy Marshall and Brian McGlynn contributed significant critique and ideas for development of intellectual content within the paper. Lucy Marshall and Brian McGlynn edited draft versions of the manuscript. Lucy Marshall and Brian McGlynn were responsible for securing funding.

Manuscript Information

Mullen, R. S., B. L. McGlynn and L. A. Marshall. 2012 Evaluating a beta regression approach for estimating fraction of clear sky transmissivity in mountainous terrain. Manuscript to be submitted to *Hydrology and Earth System Sciences*

- Journal: *Hydrology and Earth System Sciences*
- Status of manuscript (check one)
  - X Prepared for submission to a peer-reviewed journal
  - ____ Officially submitted to a peer-reviewed journal
  - ____ Accepted by a peer-reviewed journal
  - ____ Published in a peer-reviewed journal
- Publisher: European Geosciences Union
- Expected date of submission: May 2012

Abstract

The prediction of global solar radiation (GSR) in complex terrain in the absence of measurements is a difficult task. We tested a beta regression model that predicts the fraction of clear sky transmissivity (and subsequent solar radiation) in a mountainous region using model parameters that were estimated using data from a proximate valley location. We repeated the process using data from a nearby peak within the same mountain range. The model's predictive performance was assessed by considering predictive uncertainty of GSR. The model performed well when predicting fraction of clear day; however, bias was observed when predicting GSR in the absence of site specific clear sky transmissivity data. This finding is in contradiction to earlier studies that found that clear sky transmissivity could be extrapolated across the landscape with little to no bias. However, these studies largely focused on low elevation and agricultural areas, not mountainous regions. Our results suggest that although beta regression is shown to work well in extrapolating fraction of clear day across mountainous terrain, improved estimates of clear sky transmissivity are needed.

Keywords: Global Solar radiation, Beta regression, Fraction of clear day, RAWS, AgriMet, Clear sky transmissivity

Introduction

Predictions of incoming global solar radiation (GSR) in mountainous terrain is fundamental for understanding a broad range of ecological and environmental sciences, including evapotranspiration (Hargreaves and Samani 1982, Allen, Trezza and Tasumi 2006), energy balances of snow cover (Barry et al. 1990, Jin et al. 1999, Marks et al. 1999, Marks and Winstral 2001) vegetation, (Granger and Schulze 1977, Fu and Rich 2002, Pierce, Lookingbill and Urban 2005), NPP (Berterretche et al. 2005, Crabtree et al. 2009) and animal behavior models (Zeng et al. 2010, Keating et al. 2007). The historical dearth of solar radiation sensors (Thornton and Running 1999) has led to the development of multiple methods to estimate monthly GSR (Diodato and Bellocchi 2007) and daily GSR (Glassy and Running 1994, Thornton, Hasenauer and White 2000) leveraging a variety of explanatory climate variables collected at weather stations in complex terrain. These daily models are typically based on the relationship between daily changes in temperature, as well as other climate variables, and daily GSR. This approach has a long history in non-mountainous regions (Bristow and Campbell 1984, Donatelli and Campbell 1998, Fodor and Mika 2011, Hargreaves and Samani 1982, Samani et al. 2011, Ball, Purcell and Carey 2004, Allen 1997, Rivington et al. 2005) and has been referred to as the Bristow and Campbell (B&C) family of models. These models have been validated in mountainous regions using homologous site testing (Glassy and Running 1994, Thornton et al. 2000), that is, explanatory variables used to construct the model were collected at the same site that observations used to test the model. Much has been written on the portability of the fitted parameters in B&C models across agricultural

and relatively low elevation forested lands (Bristow and Campbell 1984, Thornton and Running 1999, Fodor and Mika 2011), however, little work has demonstrated how well fitted parameters from one site can be used to predict GSR at another site in mountainous terrain.

Typically, results from GSR prediction models are compared to results from previously proposed models using standard statistics such as correlation coefficient (r^2), root mean squared error, (RMSE), or mean signed deviance (MSD), without estimates of uncertainty to accompany the predictions. Stochastic methods such as boot strapping, informed simulations, or Monte Carlo analysis could be used to produce uncertainty estimates but such methods are not routinely employed for prediction of GSR. More recently, beta regression has been introduced as a suitable statistical alternative when modeling bounded environmental variables (Cribari-Neto and Zeileis 2010, Smithson and Verkuilen 2006, Eskelson et al. 2011, Korhonen et al. 2007) such as fraction of clear day (FCD), (Mullen, In review) a component of GSR. Beta regression models can provide estimates of variance for all FCD predictions (Ferrari and Cribari-Neto 2004, Smithson and Verkuilen 2006, Simas, Barreto-Souza and Rocha 2010, Ferrari and Pinheiro 2011, Ospina, Cribari-Neto and Vasconcellos 2006) such that estimates of uncertainty can be easily carried through to predictions of GSR. The beta regression method has been used extensively in ecology (Eskelson et al. 2011, Korhonen et al. 2007) and psychology (Smithson and Verkuilen 2006), and was shown to be effective when fitting GSR models in North and South Dakota (Mullen, In review), a region in North America with a relative

lack of complex terrain. The utility of beta regression in modeling solar radiation in mountain terrain has thus far not been tested to our knowledge.

Solar radiation is commonly extrapolated from valley locations to high elevation points by either kriging or inverse distance weighting (IDW) (Jolly et al. 2005), or by fitting models at lower elevations where climate data is available and using that model to predict at higher elevations (Hasenauer et al. 2003, Lo et al. 2011, Huang et al. 2008, Jolly et al. 2005, Diodato and Bellocchi 2007, Running, Nemani and Hungerford 1987). This extrapolation is typically necessary since solar radiation sensors are rarely located in the remote area of interest. There are expanding networks of weather stations in mountainous areas, such as the Remote Automated Weather Stations (RAWS), (Reinbold, Roads and Brown 2005) in the United States. However, nearly all regions of the remain under sampled, especially mountainous regions of the western United States (Reinbold et al. 2005).

There are a variety of complex interactions between air temperature, water vapor and land surface temperature that extend beyond the scope of this paper. One simple relationship though is that higher levels of atmospheric water vapor are associated with higher dew point temperatures. When a cooling air mass approaches dew point, the dispelled latent heat associated with condensation acts as a buffer leading to higher minimum temperatures. Therefore, nightly minimum temperatures can often be associated with atmospheric water vapor content. Bristow and Campbell (1984) demonstrated the usefulness of these relationships with site-based models that predict global solar radiation using ΔT as an indicator of atmospheric transmittance.

There are limitations of B&C models in complex terrain though. The B&C family of models typically relies on the assumption that diurnal amplitude of air temperature (ΔT) is directly related to GSR, which relies on a horizontally stable atmosphere. This assumption though is not always present. Air masses can be heated from below by passing over a warm surface. Additionally, drainages that hold and move cold air downslope, frost pockets, and localized wind effects can influence air temperatures in ways that are not reflected in a daily change of temperature value. Large-scale weather fronts, temperature inversions, and areas where latent heat exchange reduces the change in temperature are examples whereby B&C models may not make appropriate predictions. Despite these exceptions, B&C models have been shown to work reasonably well in complex terrain (Glassy and Running 1994, Thornton et al. 2000).

The original MT-CLIM model (Running et al. 1987) used a modified version of Bristow and Campbell's (1984) original model to predict solar radiation. Two separate studies have validated improved versions of this solar radiation algorithm in complex mountainous terrain (Glassy and Running 1994, Thornton et al. 2000). However, the emphasis in both studies was on testing the diurnal process of these improved algorithms. Consequently, the "base station" sites were identical to the "extrapolated sites". This homologous site testing meant that no corrections for aspect, elevation or slope were needed (Glassy and Running 1994, Thornton et al. 2000). Thornton et al. (2000) did note a small adjustment was needed for snow covered slopes but otherwise the model performed well.

We propose beta regression to predict GSR in mountainous regions where GSR is not measured but other climate variables are available. We test the validity of extrapolation of the GSR model to a remote high elevation location from multiple sites, by first fitting the model in a nearby valley (where data is more likely to be obtained) and then considering a site of similar elevation. Variables including minimum daily temperature, precipitation, relative humidity and the difference between high and low temperature (ΔT) have been effective for predicting GSR for low lying areas using beta regression models (Mullen, In review). Here we evaluate their effectiveness in predicting GSR in mountainous terrain. We further estimate uncertainty in GSR predictions, and propose using those uncertainty estimates as an additional tool for model assessment. We also investigate the effectiveness of extrapolated values of clear sky transmissivity (CST) for GSR predictions in complex mountain terrain.

Methods

Site Description

The remote site, (where predictions are made) for the case study is a small alpine watershed in the United States Forest Service (USFS) Tenderfoot Creek Experimental Forest (TCEF) in Western Montana. TCEF is a small, (3,693 ha) sub-alpine, (1840 – 2421 m) watershed located in the Little Belt Mountains in central Montana. Tenderfoot Creek drains into the Smith River, itself a tributary to the Missouri River. The climate is generally continental, although Pacific maritime influences are possible. Average precipitation is 880 mm, and ranges from 594 mm to 1, 050 mm from the lowest to the

highest elevations. Typically 70% of the annual precipitation falls between the first of November and the last of October. Intense summer thunderstorms are uncommon.

The proposed GSR model was initially parameterized using data from a nearby valley close to the town of White Sulphur Springs, Montana (WSSM). The WSSM site is located approximately 42 kilometers south of TCEF in a valley with extensive agricultural development (Figure 5.1). The nearby Porphyry Peak (Figure 5.1) RAWS station, located 13km to the south east of TCEF at 2510 m elevation was also used to fit a beta regression model, and predict GSR in the TCEF watershed.

Data

The GSR test data for the TCEF model is collected for ongoing hydrological research in the TCEF watershed (Emanuel et al. 2010). The sensors are positioned on a 40 m tower free of topographic shading (46° 55' 51", -110° 52' 42", 2222 m) at approximately 30 m above the ground and approximately 8 m above the lodge pole pine canopy. The radiation sensor is a Campbell Scientific CNR1, a research grade net radiometer that consists of a pyranometer and pyrgeometer pair facing upwards, and another pair facing downward. It is the upward facing pyranometer that is of interest here, as that is recording incoming short wave radiation. The spectral response of the pyranometer is 305 to 2800 nm, with an expected accuracy for daily totals of ± 10%.

The WSSM weather station (46° 33′ 12″, -110° 56′ 48″, 1,515.55 m) is part of the AgriMet network (http://www.usbr.gov/gp/agrimet/) operated by the Great Plains regional office of the Bureau of Reclamation. The AgriMet network is focused on providing weather information for agricultural purposes (Palmer 2011), and thus the sites

are located on valley floors within the western Montana. The solar radiation sensor at these sites is a Li-Cor 200 or similar pyranometer.

The Porphyry weather station (http://www.raws.dri.edu/) sits atop Porphyry Peak (elevation = 2509 meters) in Southwest Montana (46° 50′ 07″, -110° 43′ 03″) and is part of the RAWS network. This network was initially installed to provide climatological data for fire prediction and weather data for incident commanders of firefighting crews (Horel and Dong 2010, Reinbold et al. 2005). The station is owned and operated by the Unites States Forest Service, but data collected from the site is available through the Western Region Climate Center. The solar radiation sensor is a Li-COR 200 or similar pyranometer.

Components of Global Solar Radiation (GSR)

Daily GSR variability can be high in mountainous regions, even in the valleys, such as the location of the WSSM weather station (Figure 5.2). This variation is less in the winter months when maximum GSR is relatively low (thus a smaller possible range), and highest in the summer months when potential solar radiation is high yet the occurrence of thick clouds is not uncommon. Recorded springtime values of GSR at WSSM from the years 2001 through 2011, ranged from less than 5 MJ $m^{-2}d^{-1}$ to more than 30 MJ $m^{-2}d^{-1}$ (Figure 5.2); at a time of year when estimates of incident solar radiation are critical for applications such as snowmelt and stream flow modeling.

Much of the seasonal variability is attributed to latitudinal gradients caused by the earth's orbit around the sun It is assumed that these effects are well described (Gates 1980) and be easily quantified.

Additional seasonal variation can be explained by the seasonal fluctuation of atmospheric attenuation expressed as clear sky transmissivity (CST). CST is the fraction of ETR that reaches the Earth's surface on a clear day. CST is specific to the site and day of year and deriving it requires either a statistical model fit to an available data set (Fodor and Mica, 2011) or an analytical derivation using a correction from sea level (Running et al. 1987).

Fraction of clear day, (FCD), is the proportion of CST that reaches the earth's surface on any given day and is determined by clouds, dust, water and other aerosols in the atmosphere. Once a prediction of FCD is obtained, GSR can be predicted using the following equation (Fodor and Mika 2011);

$$GSR = ETR \times CST \times FCD \quad (1)$$

This deconstruction parses GSR into global solar geometric effects (ETR), elevation and other site specific effects (CST), and daily fluctuations in atmospheric attenuation (FCD). The B&C family of models estimates GSR by relating explanatory climate variables such as temperature and humidity to daily fluctuations in FCD (Kang and Scott 2008, Bristow and Campbell 1984, Geuder et al. 2004, Thornton et al. 2000, Thornton and Running 1999).

Determining Observed FCD

Extraterrestrial radiation (ETR) was determined for all sites based on their geographical location and time of year (Gates, 1980). CST was derived empirically using historical data and modeled using a Fourier series (Fodor and Mika 2011). Every day of the year was assigned a maximum value based on the highest observed daily sky

transmissivity for that day, or for any day that occurred within a seven day moving window around that day. These maximum values are known to have a yearly sinusoidal pattern (Fodor and Mika 2011), therefore a Fourier series was used to model them. This fitted Fourier series then represents an envelope curve for daily transmissivity values (Figure 5.3), and was used to model the clear sky transmissivity for each day of the year. Observed FCD was then calculated by inverting Equation 1 (Fodor and Mika 2011);

$$FCD = GSR / (ETR \times CST) \tag{2}$$

(where GSR is the measured amount of global solar radiation). These observed FCD values were used in fitting the beta regression model.

Applying the Beta Regression Model

Beta regression is a statistical model developed for analyzing beta distributed dependent variables (Ferrari and Cribari-Neto 2004). The beta distribution is a continuous probability distribution between 0 and 1. Beta regression is therefore appropriate for rates, proportions, and other continuous variables that fall entirely within the standard unit interval (0, 1) and have a mean that can be related to a set of explanatory variables and with a link function and estimable coefficients.. Ferrari and Cribari-Neto (2004) reparameterized the traditional form of the beta distribution by setting $u = p/(p+q)$ and $\phi = p+q$. This yields:

$$f(y;u,\phi) = \frac{\Gamma(\phi)}{\Gamma(u\phi)\Gamma((1-u)\phi)} y^{u\phi-1}(1-y)^{(1-u)\phi-1}, 0 < y < 1 \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function, $0 < u < 1$ and $\phi > 0$. The parameter ϕ is the precision parameter where for fixed u, a larger ϕ gives a smaller variance. A beta distributed variable can

be denoted as $y \sim \beta(u, \phi)$. The expected value of *y* is *u*, or $E(y) = u$. Beta regression is written in matrix notation, as;

$$g(u_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} = \eta_i \tag{4}$$

where $\beta = (\beta_1, ..., \beta_k)^T$ is a *k* x 1 vector of unknown regression parameters, $x_i = (x_{i1}, ..., x_{ik})^T$ is a vector of *k* regressors, or explanatory variables, and η_i is a linear predictor. This is a naturally heteroscedastic function where;

$$Var(y_i) = \frac{u_i(1-u_i)}{1+\phi} \tag{5}$$

When a dependent variable, such as FCD, lies constrained in the standard unit interval, assumptions of normality are usually incorrect. Truncation of the response value makes even an approximate normal distribution unlikely. Almost by definition these variables are highly heteroscedastic with more variation around the mean and less near the boundaries. Like most proportion data, FCD distributions tend to be asymmetric, which results in unreliable confidence intervals and hypothesis testing. Beta regression addresses all of these issues (Ferrari and Cribari-Neto 2004, Ferrari and Pinheiro 2011) and is easy to perform in freely available statistical software programs (Cribrali-Neto, 2010).

The flexibility of beta regression is demonstrated by modeling predictions of FCD using a set of climate variables that are regularly collected at weather stations as regressors. Previously proposed methods for FCD predictions rely on a univariate non-linear relationship between ΔT and FCD (Bristow and Campbell 1984), Some methods require a two-step process (Donatelli and Campbell 1998, Thornton and Running 1999,

Thornton et al. 2000) where dew point or relative is used to parameterize a preliminary model, The parameter is then inserted into the secondary model as a constant in order to predict FCD (Thornton and Running 1999, Donatelli and Marletto 1994, Donatelli and Campbell 1998)

For this study, a beta regression model was constructed using the WSSM data set. The response variable was FCD, with the following climate variables used for explanatory variables; ΔT, (the difference between maximum and minimum temperature), relative humidity, precipitation, low temperature, and the sine and cosine components of the day of year, (Mullen, In review). BIC values were used in order to determine the best model.

This model was then extrapolated to the TCEF site using the explanatory variables collected in TCEF, and the parameter estimates from the WSSM model to predict fraction of clear day. Each prediction of FCD is a beta distributed value, $y_i \sim \beta(u_i, \phi_i)$. If μ and ϕ are assumed to be estimated without error, then the 0.025 and 0.975 quantiles of the distribution can be used as the upper and lower bounds for a 95% estimation interval (Figure 5.4). This assumption distinguishes this interval from a true confidence interval or prediction interval; however, it can be used similarly.

This process was repeated using Porphyry data to parameterize a model. FCD at TCEF was predicted using the parameter estimates from the Porphyry model and the explanatory variables collected at TCEF.

Using FCD to Estimate GSR

The predicted FCD value (FCD_p) can then be used to predict GSR (GSR_p) for the TCEF watershed. To do so, a derivation of CST and of ETR must be made for TCEF. CST could be taken to be 0.6 under standard temperature and pressure at sea level, increasing by 0.00008 for every meter gain in elevation (Glassy and Running 1994). Alternatively, CST at TCEF could be adjusted by adding 0.00008 m^{-1} of elevation difference between measured CST at a site and TCEF (Glassy and Running 1994). Similarly, since TCEF and Porphyry are each unaffected by daytime shading, their proximity to one another would imply similar CST values (Fodor and Mica, 2011). This is true even if daily weather patterns were significantly different. It is assumed that the well-established equations for ETR (Gates 1980) are adequate to estimate ETR. Once the three components are determined, GSR_p can be predicted with,

$$GSR_p = ETR \times CST \times FCD_p \quad (6)$$

Results

Predicting at TCEF Using WSSM Station

Our initial analysis examined the extrapolation of the beta regression parameters from valley to mountain locations. The beta regression was fit using FCD and the suite of explanatory climate variables at the WSSM weather station. This model was then applied to the explanatory climate variables collected in TCEF in order to predict FCD in TCEF. The predicted FCD value was then used to predict GSR at TCEF. The following refers to predicting FCD and GSR at the TCEF tower using data from WSSM.

A simple model based on only calculated transmissivity from the WSSM station was first used to estimate GSR at the TCEF tower. This provided a baseline for comparison against the beta regression models. Transmissivity was calculated at WSSM as the fraction of ETR that reached the Earth's surface. This was simply the measured value of solar radiation divided by ETR. Subsequent analyses partitioned transmissivity into CST and FCD, but this initial analysis did not. No analytical confidence intervals are provided with this model, so there are no capture rates. However, RSME and MSD (observed GSR vs. predicted GSR) were calculated for comparison to the subsequent models (Table 5.2). For the simple transmissivity model, the RMSE was 7.503 MJ $m^{-2}d^{-1}$, and the MSD was -2.371 MJ $m^{-2}d^{-1}$.

Observations at WSSM were used to calculate the daily fraction of clear day (FCD) for fitting the beta regression model. The annual average for observed FCD for the White Sulphur Springs weather station was 70.62%, with highest seasonal value found in summer (77.92%) and lowest found in winter (65.86%) (Table 5.1).

A beta regression model was constructed to predict FCD using explanatory variables of ΔT, (the difference between maximum and minimum temperature), relative humidity, precipitation, low temperature, and the sine and cosine components of the day of year. In models of FCD, ΔT has been considered the most influential independent variable (Bristow and Campbell 1984) and has been shown to have a strong sinusoidal influence on GSR (Bristow and Campbell 1984, Fodor and Mika 2011, Donatelli and Campbell 1998, Thornton and Running 1999). Therefore, a squared and cubed ΔT term

was included in the model to accommodate this, even though the sinusoidal effect was not pronounced in data at WSSM and Porphyry (Figure 5.5).

The initial beta regression model used to fit the White Sulphur Springs and Porphyry data was considered the full model (Mullen, In review);

$$FCD_p = DeltaT \times RelHum \times MinT \times Prec \times \left[\sin(day) + \cos\left(day\right)\right] + DeltaT^2 + DeltaT^3 \quad (7)$$

Variables were subsequently eliminated and the model was rerun. No combination of reduced variables led to a smaller BIC value, so the full model was maintained. FCD_p was used to predict GSR using Eq. 6. Similarly, the 95% upper and lower bounds for FCD were substituted into Eq. 6 to get 95% upper and lower bounds for GSR_p. The 'capture rate' was defined as the percentage of times that the measured value fell in between the upper and lower bounds.

The annual capture rate was 95.38% when using the predicted FCD values at White Sulphur Springs to estimate GSR at White Sulphur Springs, indicating a good overall fit (RMSE = 2.894 MJ $m^{-2}d^{-1}$, MSD = -0.0745 MJ $m^{-2}d^{-1}$). The highest seasonal capture rate was summer (97.87%), with spring second highest (95.65%), and winter and fall the two lowest (94.10% and 94.01% respectively).

The White Sulphur Springs beta regression model was used to predict FCD (Figure 5.6) at TCEF. The annual capture rate of FCD was 82.61%. The highest capture rate was for summer (84.34%). The second and third highest were essentially the same with spring (82.89%) slightly higher than winter (82.76%). The lowest was fall (80.24%) (Table 5.4).

These predictions of FCD were subsequently used to predict GSR at TCEF. CST was assumed to increase by 0.00008 m^{-1}. The rate at which the observed TCEF value of GSR was between the lower and upper confidence bounds of the GSR model was 58.76%, substantially lower than the capture rate for FCD (82.61%). Winter had the highest rate (67.93%), and summer the lowest (52.47%) (Table 5.4). The RMSE (6.655 MJ $m^{-2}d^{-1}$) was lower and the absolute value of MSD (-3.490 MJ $m^{-2}d^{-1}$) was higher than for the simple transmissivity model (Table 5.2).

The model performance discrepancy between predicting FCD and GSR suggests that there were problems in estimating CST. To confirm this, a follow up analysis was performed. CST was modeled using a Fourier series (Fodor and Mika 2011) and daily transmissivity data from the TCEF tower. FCD was predicted using the beta regression model from WSSM as was done in the previous case. In this case, the GSR model performance improved (RMSE = 5.631 MJ $m^{-2}d^{-1}$and MSD = -0.701 MJ $m^{-2}d^{-1}$) (Table 5.2). The overall capture rate for GSR was 82.75% (Table 5.3). The season with the highest capture rate was summer with 84.89 and the lowest was in fall (80.24%) (Table 5.5). These numbers are strikingly close to the predicted FCD values above, confirming that an incorrect derivation of CST can lead to errors in estimating GSR at the target location (Figure 5.7).

An alternative derivation of CST was made by comparing daily CST values at Porphyry and at WSSM (Figure 5.7). A value of 0.00050 m^{-2} was derived by simply averaging the daily differences in CST between the two sites, and determining the average difference per meter of elevation. . When this was used in place of 0.00008 m^{-2},

the GSR estimates were not substantially improved (RMSE = 6.099 $m^{-2}d^{-1}$ and MSD = 1.64 $m^{-2}d^{-1}$) (Table 5.2), but the annual capture rate increased to 89.50% (Table 5.3). The capture rate in the spring was the highest (90.89%) with winter having the lowest capture rate (87.24%) (Table 5.5).

Predicting at TCEF Using Porphyry Station

The follow-up analysis examined the extrapolation of the beta regression parameters nearby mountain locations. This is becoming increasingly possible in places like the Northern Rockies with the steady increase in RAWS or similar meteorology sites (Reinbold et al. 2005). The beta regression was fit using FCD and the suite of explanatory climate variables at the Porphyry weather station, in the same manner it had been done at WSSM. This model was then applied to the explanatory climate variables collected in TCEF to predict FCD in TCEF. The predicted FCD value was then used to predict GSR at TCEF. The following subsections all refer to predicting FCD and GSR at the TCEF tower using data from the Porphyry site.

A simple model using only calculated transmissivity from the Porphyry station was first used to estimate GSR at the TCEF tower, as was done with the WSSM data. This provided a baseline for comparison against the beta regression models. Transmissivity was calculated at Porphyry as the fraction of ETR that reached the Earth's surface. RSME and MSD (observed GSR vs. predicted GSR) were calculated for comparison to later models. The simple transmissivity model had a RMSE of 9.181 MJ $m^{-2}d^{-1}$ and a MSD of -3.699 MJ $m^{-2}d^{-1}$ (Table 5.2), and provided a baseline for comparison for the subsequent analyses

The annual average for calculated FCD for the Porphyry weather station was (55.64%), with highest seasonal value found in summer (63.82%) and lowest seasonal value in winter (48.10%) (Table 1).

The fitted beta regression model was performed well for the homologous site test (RMSE = 4.060 $m^{-2}d^{-1}$, MSD = 0.1898 $m^{-2}d^{-1}$) (Table 5.2) using the Porphyry data set. There was an annual capture rate for the observed value of GSR was 95.41% (Table 5.3). The highest seasonal rates were for summer (97.05%) and spring (97.04%). The capture rate for winter was the lowest (91.40%) (Table 5.5).

The Porphyry beta regression was used to predict FCD, and subsequently GSR at TCEF. The annual capture rate of FCD was 92.98%, significantly higher than the WSSM beta regression when used to predict FCD at TCEF. Spring had the highest capture rate (94.44%) and winter had the lowest rate (90.69%) (Table 5.5).

There was a significant decline in capture rates when estimating GSR at TCEF MSD while using the standard CST elevation adjustment (0.00008 m^{-2}) between the Porphyry site and TCEF (Table 2). The annual capture rate was 62.73% (Table 5.3). Winter and spring were the two highest rates (66.55% and 66.44% respectively) and summer was the lowest (57.59%) (Table 5.5).

The locally derived elevation adjustment value for CST (0.00050 m^{-2}) was then used to estimate CST in TCEF. This adjusted CST was used to predict GSR in TCEF (Figure 5.8) with a reduction in model accuracy. The annual capture rate of GSR was 45%. Clearly, the locally derived CST value biased the results (RMSE = 9.114 MJ $m^{-2}d^{-1}$, MSD = -7.001 MJ $m^{-2}d^{-1}$) (Table 5.2). Overall fit was improved

(RMSE = 5.432 MJ $m^{-2}d^{-1}$, MSD = -2.578 MJ $m^{-2}d^{-1}$) (Table 5.2) when no elevation adjustment of CST was applied between the two mountain sites, but not substantially (annual capture rate = 66%) (Table 5.3).

Finally, CST was derived using historical data at the TCEF tower (Figure 5.8). When this value was used to estimate GSR, overall model fit was improved (RMSE = 5.432 MJ $m^{-2}d^{-1}$, MSD = -2.576 MJ $m^{-2}d^{-1}$) (Table 5.2) and the capture rate was 93.18%. With appropriate derivations of CST, one would expect GSR rates to be nearly identical to FCD capture rates. The discrepancy that occurred when using remotely derived CST rates suggest that use of an inappropriate CST model can lead to significant inaccuracies in GSR estimates.

Discussion

Our study focused on the application of a beta regression model for improved predictions of FCD, and consequently GSR, in complex terrain. To assess the model, we focus on (1) the predictive performance of the beta regression model on FCD (and subsequently GSR); (2) the uncertainty of predictions of FCD and GSR; and (3) the compound error or bias that may be introduced by an inappropriate derivation of clear sky transmissivity (CST).

The beta regression approach produced good predictions of FCD, and consequently GSR in homologous site testing. The 95% upper and lower bounds of GSR captured the measured value 95.34% at WSSM and 95.41% at Porphyry across all available data (Table 5.3). The RMSE for WSSM (7.503 MJ $m^{-2}d^{-1}$) and Porphyry

(9.181 MJ $m^{-2}d^{-1}$) (Table 5.2) compare favorably to previous studies. As comparison, the mean RMSE across six stations in central Oregon for a daily GSR model of the B&C family was 21.86 MJ $m^{-2}d^{-1}$ (Glassy and Running 1994)MSD. Thornton et al (2000) additionally implement a daily B&C style model and report the mean absolute error (MAE) for 24 stations in Austria (high 4.72 and the low was 2.08). The MAE for WSSM (2.22 MJ $m^{-2}d^{-1}$) and Porphyry (3.05 MJ $m^{-2}d^{-1}$) (Table 5.2) were considerably lower. The limited sample size for this case study does not allow a direct comparison of methods, but the case studies here provide evidence that the beta regression method performed well. These results additionally show that that uncertainty intervals produced in the beta regression are valid and useful for GSR prediction in mountainous areas.

When extrapolating the beta regression model to the higher-elevation site at TCEF, we can assess the model in terms of its ability to capture TCEF FCD observations within the model uncertainty limits. The FCD beta regression models derived from WSSM and Porphyry data were slightly lower than when the model was applied to the site at which it was fitted (92.98% and 93.60% respectively) (Table 5.3). This reduction in performance could be partially attributed to the assumption that the mean (μ) and precision parameter (ϕ) of the beta regression models were assumed to be fixed and known. Realistically these fitted parameters have a measure of uncertainty that is not represented in the analysis. It would be difficult to stochastically apply variation to these parameters, as they do not vary independently. Extension of the current analysis to a hierarchical Bayesian approach could be applicable to address this issue.

Another likely reason that the model did not perform as well at TCEF is simply that topographical and climatological factors will affect the ability to extrapolate beta regression parameters to the higher elevation site. The decrease in near-surface maximum temperature with the increase in elevation (the lapse rate) is usually greater than that of minimum temperature, thus reducing ΔT in manner not directly related to GSR (Thornton et al. 2000). Further, surface snow cover has been shown to bias GSR measurements downward such that underestimation of GSR occurs (Thornton et al. 2000) and it is likely that snow persisted at higher elevations long after spring melt in the valley. We note that both of these issues should be less pronounced when moving from Porphyry to TCEF. However, localized inversions, frost pockets, wind patterns, or synoptic front movement through the Little Belt Mountains could reduce the chances of a horizontally stable atmosphere required for any member of the B&C family of models to behave optimally (Glassy and Running 1994).

Most importantly, our case studies highlight the importance of appropriately deriving the components of GSR even with very good predictions of FCD. If we assume that the approaches for ETR are sound, (note ETR estimates were applied equally to all three sites) then the problem of bias lies in the derivation of CST. Previous studies have indicated that spatially interpolated parameters (including values for CST) for GSR models can be effective over large distances, (Fodor and Mika 2011, Thornton and Running 1999, Ball et al. 2004, Winslow, Hunt Jr and Piper 2001). However, these studies focused on low lying areas, and made no attempt to extrapolate the parameters into mountainous areas. When simple adjustments for elevation were applied to obtain a

derivation of CST for TCEF (0.00008 m^{-2}, (Running et al. 1987), there was a clear bias (MSD = -3.49) observable in plots of predicted GSR values versus observed values (Figure 5.7 and Figure 5.8). When applying the WSSM model for prediction in TCEF, this bias was remedied by applying a locally derived elevation adjustment value (MSD = 1.64) (Figure 5.7) or by using CST modeled at TCEF (MSD = -0.701), both of which improved model fit (Fig 5.7 and Fig 5.8). The issue though is that applying the locally derived elevation adjustment value for CST actually led to a worse fit when predicting GSR at TCEF using the Porphyry data (Table 5.2), although the fit was comparable once TCEF CST was derived using the observations available at TCEF (Table 5.2).

CST appears to be less transferable within mountainous regions than within low lying valleys. This presents certain problems for predicting GSR. Clearly, if sufficient GSR data exist to model CST accurately, then this can be used to parameterize a beta regression model for infilling missing data. This is an important contribution when prediction of missing values due to instrumentation failure or power outages or data gaps is required. However, perhaps a greater need for modeling solar radiation is extrapolating to watersheds that do not have solar radiation sensors. To do this without bias, more progress on a reliable derivation of CST for complex terrain is needed. It is important to note that no other member of the B&C family of models is immune to this problem. All of these models rely on partitioning transmissivity into FCD and CST, and consequently need to derive CST at the point on the landscape where a predication of GSR is made in the absence of measured values.

Conclusion

Beta regression has recently presented itself as an attractive alternative to modeling global solar radiation (GSR) in regions where no data exist. Our case studies indicate that beta regression can be appropriate for modeling fraction of clear day (FCD) in mountainous climates even in the absence of observed values of GSR. However, if estimates of GSR are needed (as is often the case) then improvement in the derivation of clear sky transmissivity in the absence of data is needed. When tested at a single site, that is, where the derivation of CST is non-consequential, the beta regression model performed favorably when compared to previously proposed models, is more flexible than previously proposed models and provides intervals of uncertainty.

Uncertainty intervals were used in this study to help assess model fit, however these intervals also serve other purposes. Almost invariably the estimated solar radiation is required as input to another model or for determining requirements for photovoltaic applications. These uncertainty intervals can be propagated through subsequent models, or can be used to support decisions regarding future energy supply or to assess risk of device failure. It would be additionally helpful to reduce the size of uncertainty intervals and develop full confidence intervals (by considering parametric uncertainty).

There are significant remaining challenges for predicting GSR in mountainous terrain, finding unbiased derivations of CST, developing more precise models that have reduced confidence intervals but still perform well in capturing observed values, and doing so with limited, or even no observed climate variables. Despite these challenges, new developments in the field are occurring rapidly. These advancements are concurrent

with an increase of instrumentation around the globe, and we have presented a method that is especially suited to adaptation anywhere a suite of explanatory variables are being measured.

Acknowledgements

The authors would like to thank the National Resource Conservation Service for funding this project, the Western Regional Climate Center for archiving and serving the RAWS data, and the Bureau of Reclamation – Great Plains office for collecting, maintaining, and serving the AgriMet data. In addition, we would like to thank Ryan Emanuel for flux tower data processing support. NSF Grants EAR-0404130, DEB-0807272, EAR-0943640 , EAR-0837937, and EAR-0838193

Tables

Table 5.1 Seasonal trends in calculated FCD values for the two sites used for prediction.

	WSSM	Porphyry
	Units are fraction of clear day (%)	
Yearly	70.62	55.64
Winter	65.86	48.1
Spring	68.83	52.57
Summer	77.92	63.82
Fall	69.84	57.18

Table 5.2 RMSE and MSD for predicted GSR at TCEF using two different prediction data sets (WSSM and Porphyry) and 5 different models. The simple transmissivity model just uses transmissivity at either WSSM or Porphyry. The other 4 models differ only in how CST is estimated.

	WSSM		Porphyry	
	RMSE	MAE	RMSE	MAE
Transmissivity model	7.503	-2.371	9.181	-3.699
Pred GSR – homologous site testing	2.893	-0.074	4.06	0.189
Pred GSR at TCEF with CST adjusted using 0.00008 m^{-1}	6.655	-3.49	7.38	-5.184
Pred GSR at TCEF with CST derived from TCEF data	5.631	-0.701	5.428	-2.559
Pred GSR at TCEF with CST derived using locally derived adjustment value	6.099	1.64	9.114	-7.001

Table 5.3 Capture rates for each model constructed from each base site for the year.

	WSSM	Porphyry
	Yearly	
Pred GSR – homologous site testing	95.38	95.41
Pred FCD at TCEF using site listed above	82.61	92.98
Pred GSR at TCEF with CST adjusted using 0.00008 m^{-1}	58.76	62.73
Pred GSR at TCEF with CST derived from TCEF data	82.89	93.6
Pred GSR at TCEF with CST derived using locally derived adjustment value	89.5	93.18

Table 5.4 Seasonal values for capture rates for each model constructed from the WSSM site data.

	WSSM			
	Winter	Spring	Summer	Fall
Pred GSR at WSSM	94.1	95.65	97.87	94.01
Pred FCD at TCEF	82.76	82.89	84.34	80.24
Pred GSR at TCEF with CST adjusted using 0.00008 m^{-1}	67.93	61.11	52.47	54.5
Pred GSR at TCEF with CST derived from TCEF data	82.41	83.11	84.89	80.24
Pred GSR at TCEF with CST derived using locally derived adjustment value	87.24	90.89	89.29	89.82

Table 5.5 Seasonal values for capture rates for each model constructed from the Porphyry site data.

	Porphyry			
	Winter	Spring	Summer	Fall
Pred GSR at Porphyry	91.4	97.04	97.05	95.52
Pred FCD at TCEF using site listed above	90.69	94.44	92.86	93.11
Pred GSR at TCEF with CST adjusted using 0.00008 m^{-1}	66.55	66.44	57.59	59.88
Pred GSR at TCEF with CST derived from TCEF data	91.38	95.11	92.86	94.31
Pred GSR at TCEF with CST derived using locally derived adjustment value	90.69	95.11	92.86	93.11

Figures

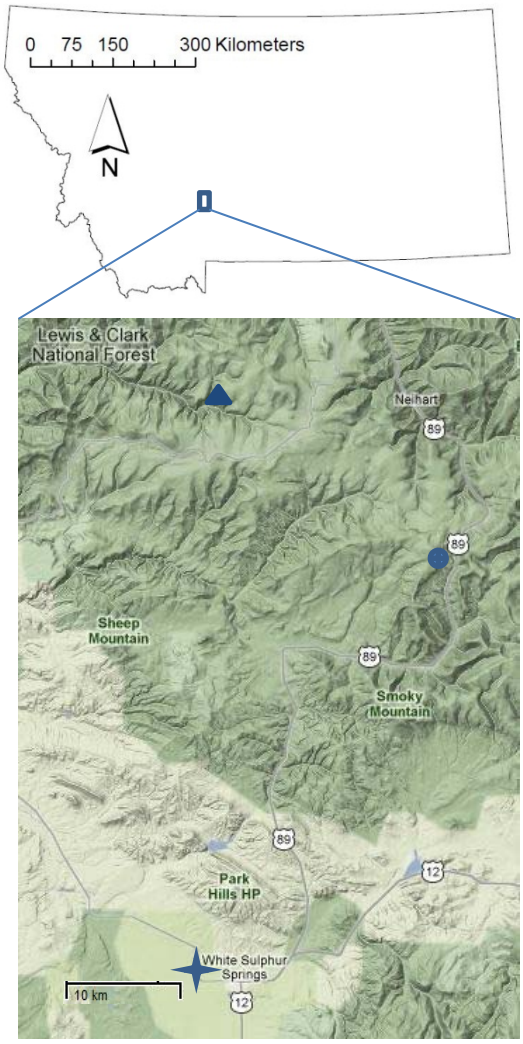


Figure 5.1 Map of the three study sites with reference to their location within the state of Montana. Note the change of terrain complexity between the WSSM site (blue star) and the TCEF watershed (Blue square). The Porphyry site (blue circle) is also in the mountains.

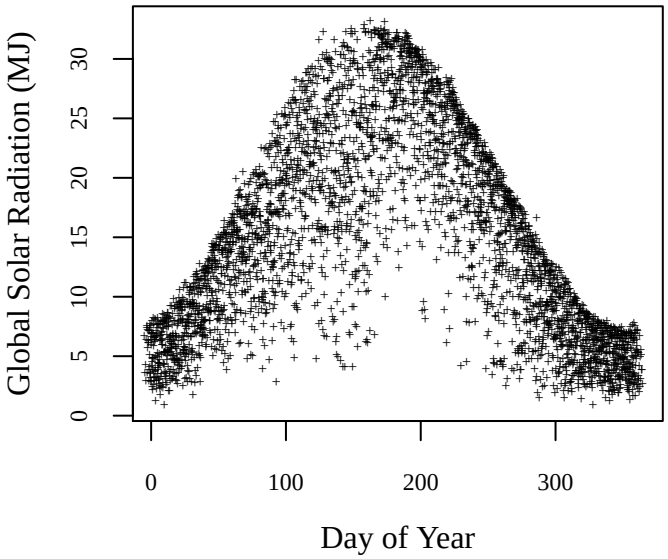


Figure 5.2 A scatter plot showing the measured GSR in mega joules by day of year at the WSSM weather station.

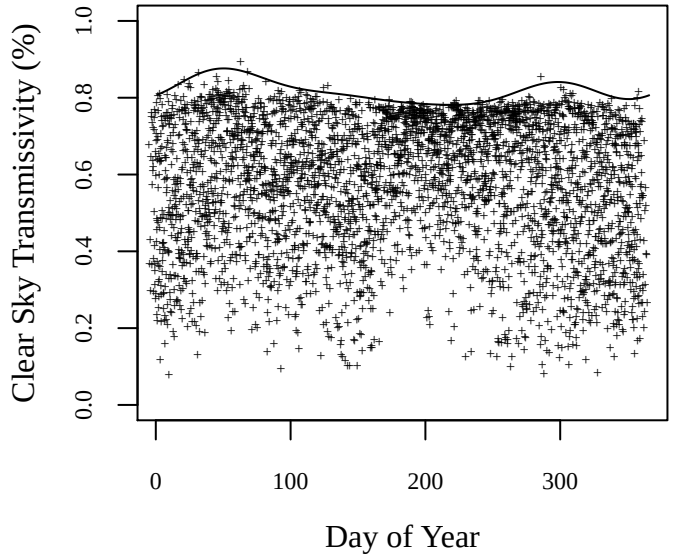


Figure 5.3 A visual representation of an envelope curve for modeling CST. The scatter plot points are daily sky transmissivity values, while the curve is a fitted Fourier series used to model CST.

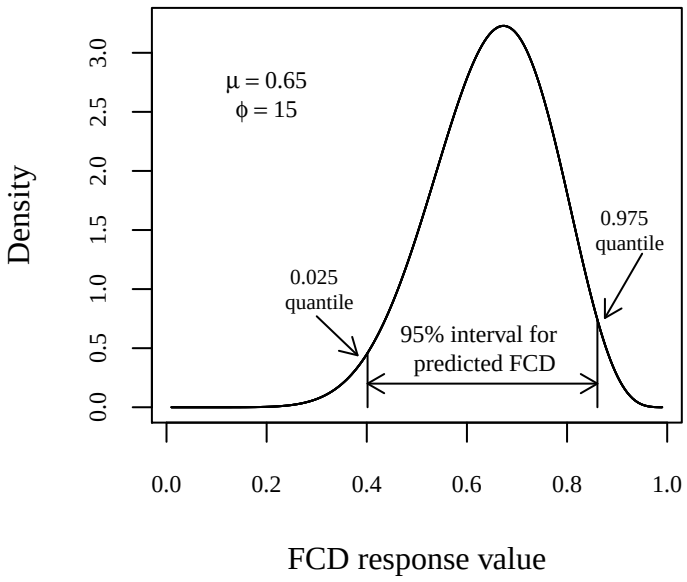


Figure 5.4 An example of a beta distribution. In this case, μ equals 0.65, and ϕ equals 15. The 0.025 and 0.975 quantile represent the lower and upper bounds for the middle 95% of the distribution. The value along the x –axis at these quantiles can be considered an upper and lower bound for an uncertainty interval. Since μ and ϕ are assumed to be fixed, these are not true confidence intervals.

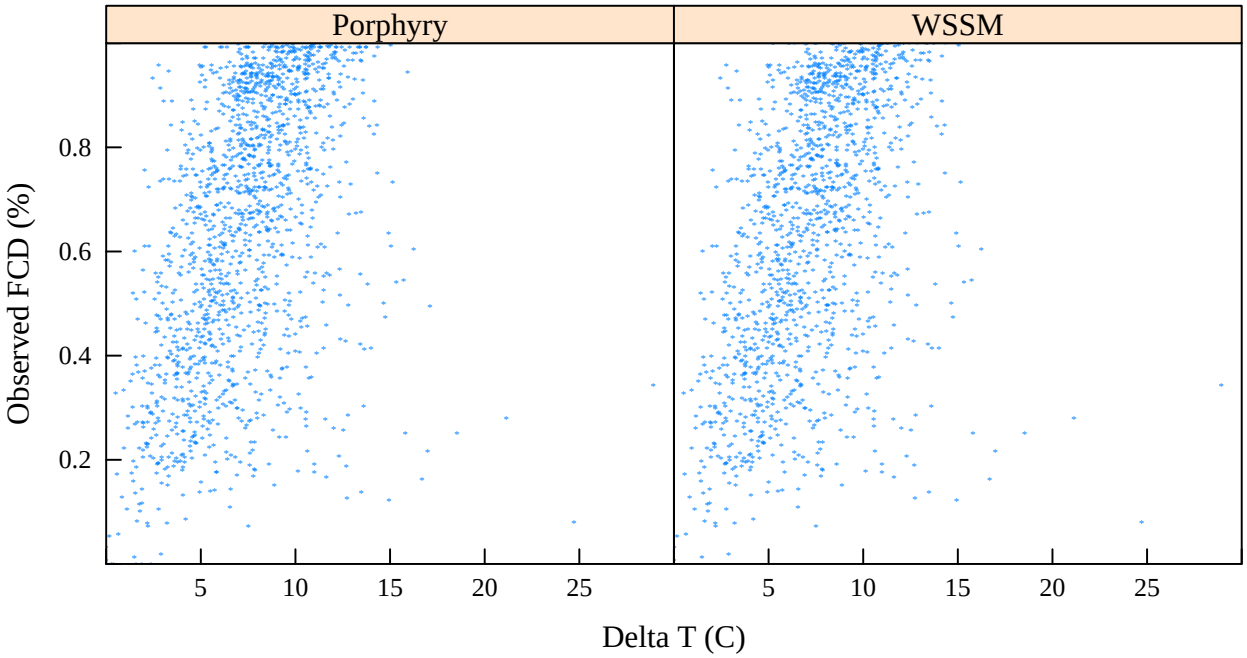


Figure 5.5 A scatter plot of ΔT vs. observed FCD at the two sites used for prediction. This plot indicates that the sinusoidal relationship is not as pronounced as earlier studies have found in non-mountainous areas.

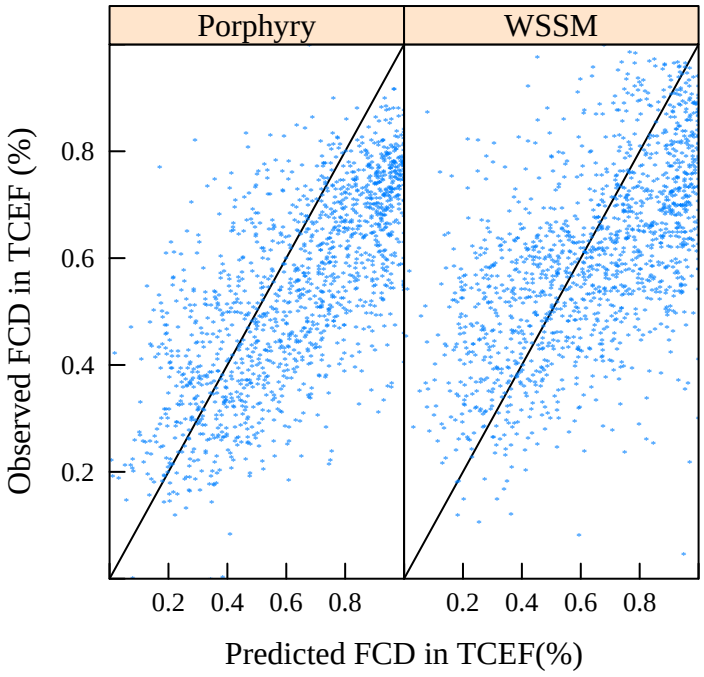


Figure 5.6 A scatter plot of predicted FCD vs. observed FCD at TCEF based on the two beta regression models fit at WSSM and Porphyry.

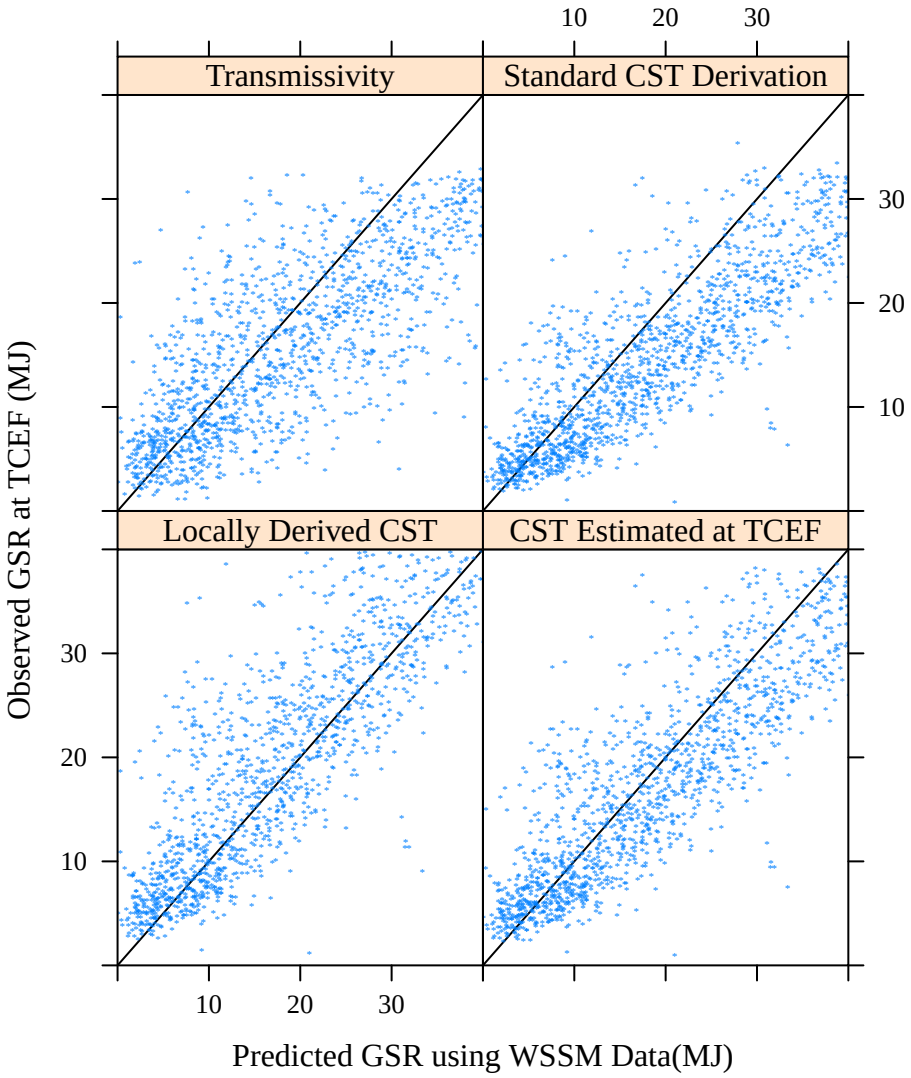


Figure 5.7 Predicted versus observed GSR at TCEF for four different models when using the beta regression FCD model fitted at WSSM. This upper left is a simple transmissivity model, the upper right uses a standard adjustment of CST for elevation, and this value was applied to WSSM CST to estimate CST at TCEF. The lower left show the results when a locally derived elevation adjustment value for CST was used, and the lower right shows the result for when the TCEF data was used to model CST using a Fourier series. Note the improved model fit for the lower two models, however, these models would not be possible in the absence of data in the target location.

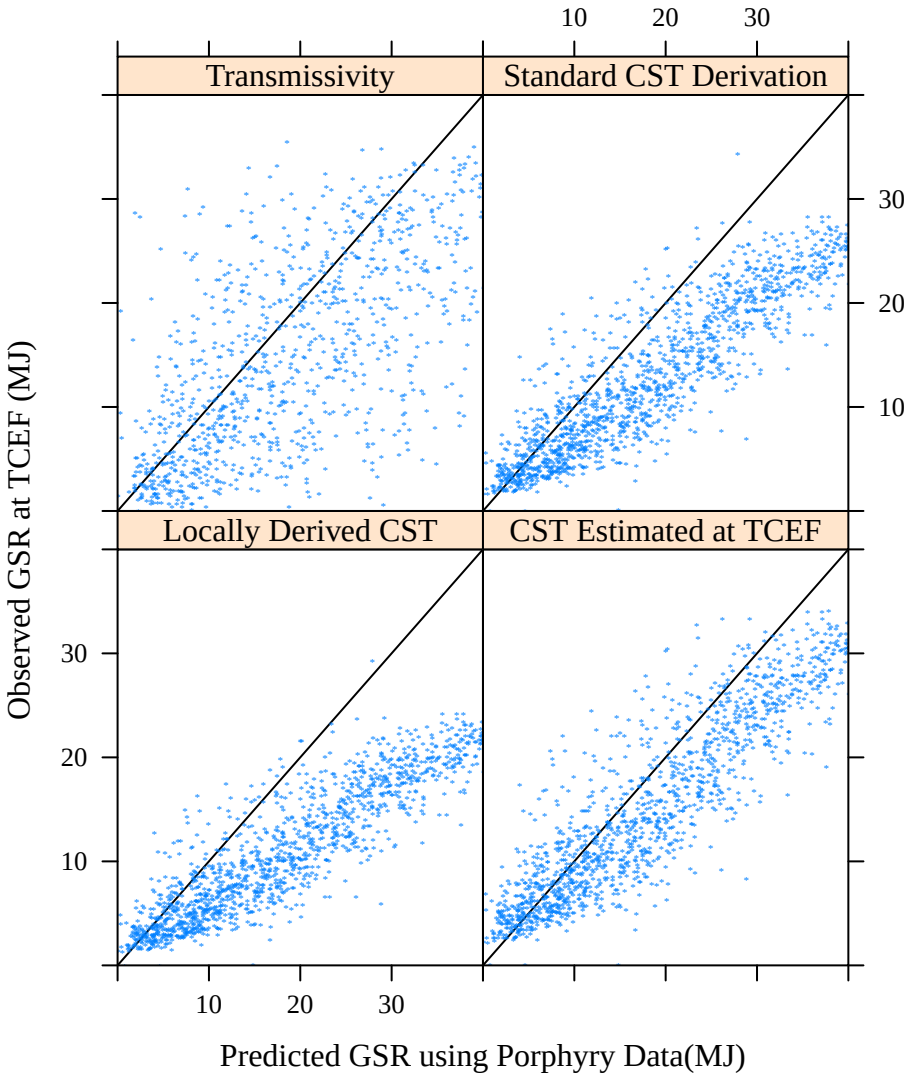


Figure 5.8 Predicted versus observed GSR at TCEF for four different models when using the beta regression model from Porphyry. This upper left is a simple transmissivity model, the upper right uses a standard adjustment of CST for elevation, and this value was applied to Porphyry CST to estimate CST at TCEF. The lower left show the results when a locally derived elevation adjustment value for CST was used, and the lower right shows the result for when the TCEF data was used to model CST using a Fourier series.

Literature Cited

Allen, R. G. (1997) Self-Calibrating Method for Estimating Solar Radiation from Air Temperature. *Journal of Hydrologic Engineering,* 2**,** 56-67.

Allen, R. G., R. Trezza & M. Tasumi (2006) Analytical integrated functions for daily solar radiation on slopes. *Agricultural and Forest Meteorology,* 139**,** 55-73.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Barry, R., M. Prévost, J. Stein & A. P. Plamondon (1990) Application of a snow cover energy and mass balance model in a balsam fir forest. *Water Resour. Res.,* 26**,** 1079-1092.

Berterretche, M., A. T. Hudak, W. B. Cohen, T. K. Maiersperger, S. T. Gower & J. Dungan (2005) Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment,* 96**,** 49-61.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Crabtree, R., C. Potter, R. Mullen, J. Sheldon, S. Huang, J. Harmsen, A. Rodman & C. Jean (2009) A modeling and spatio-temporal analysis framework for monitoring environmental change using NPP as an ecosystem indicator. *Remote Sensing of Environment,* 113**,** 1486-1496.

Cribari-Neto, F. & A. Zeileis (2010) Beta Regression in R. *Journal of Statistical Software,* 34**,** 1-24.

Diodato, N. & G. Bellocchi (2007) Modelling solar radiation over complex terrains using monthly climatological data. *Agricultural and Forest Meteorology,* 144**,** 111-126.

Donatelli, M. & G. S. Campbell. 1998. A simple model to estimate global solar radiation. In *Proceedings of Fifth European Society of Agronomy Congress, vol 2. The Slovac Agric. univ. Nitra, Slovak Republic*, 133 - 134.

Donatelli, M. & V. Marletto. 1994. Estimating surface solar radiation by means of air temperature. In *Proceedings of third annual ESA Congress, September 18-22. Abano, Italy*, 352-353.

Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth & P. D'Odorico (2010) Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains. *Water Resour. Res.,* 46**,** W11553.

Eskelson, B. N. I., L. Madsen, J. C. Hagar & H. Temesgen (2011) Estimating Riparian Understory Vegetation Cover with Beta Regression and Copula Models. *Forest Science,* 57**,** 212-221.

Ferrari, S. & F. Cribari-Neto (2004) Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics,* 31**,** 799 - 815.

Ferrari, S. L. P. & E. C. Pinheiro (2011) Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation,* 81**,** 431 - 443.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Fu, P. & P. M. Rich (2002) A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture,* 37**,** 25-35.

Gates, D. M. 1980. *Biophysical Ecology*. New York: Springer-Verlag.

Geuder, N., V. Quaschning, P. Viebahn, F. Steinsiek, J. Spies & C. Hendriks. 2004. Comparison of Terrestrial and Space Power Generation for Europe. In *4th International Conference on Solar Power from Space - SPS*. Granada, Spain.

Glassy, J. M. & S. W. Running (1994) Validating Diurnal Climatology Logic of the MT-CLIM Model Across a Climatic Gradient in Oregon. *Ecological Applications,* 4**,** 248-257.

Granger, J. E. & R. E. Schulze (1977) Incoming Solar Radiation Patterns and Vegetation Response: Examples from the Natal Drakensberg. *Vegetatio,* 35**,** 47-54.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Hasenauer, H., K. Merganicova, R. Petritsch, S. A. Pietsch & P. E. Thornton (2003) Validating daily climate interpolations over complex terrain in Austria. *Agricultural and Forest Meteorology,* 119**,** 87-107.

Horel, J. D. & X. Dong (2010) An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology,* 49**,** 1563-1578.

Huang, S., P. Rich, R. Crabtree, C. Potter & P. Fu (2008) Modeling Monthly Near-Surface Air Temperature from Solar Radiation and Lapse Rate: Application over Complex Terrain in Yellowstone National Park. *Physical Geography,* 29**,** 158-178.

Jin, J., X. Gao, S. Sorooshian, Z.-L. Yang, R. Bales, R. E. Dickinson, S.-F. Sun & G.-X. Wu (1999) One-dimensional snow water and energy balance model for vegetated surfaces. *Hydrological Processes,* 13**,** 2467-2482.

Jolly, W. M., J. M. Graham, A. Michaelis, R. Nemani & S. W. Running (2005) A flexible, integrated system for generating meteorological surfaces derived from point sources across multiple geographic scales. *Environmental Modelling & Software,* 20**,** 873-882.

Kang, H. & D. Scott (2008) An integrated spatio-temporal GIS toolkit for exploring intra-household interactions. *Transportation,* 35**,** 253-268.

Keating, K. A., P. J. P. Gogan, J. M. Vore & L. R. Irby (2007) A Simple Solar Radiation Index for Wildlife Habitat Studies. *The Journal of Wildlife Management,* 71**,** 1344-1348.

Korhonen, L., K. T. Korhonen, P. Stenberg, M. Maltamo & M. Rautiainen (2007) Local models for forest canopy cover with beta regression. *silva Fennica,* 41**,** 671-685.

Lo, Y.-H., J. A. Blanco, B. Seely, C. Welham & J. P. Kimmins (2011) Generating reliable meteorological data in mountainous areas with scarce presence of weather records: The performance of MTCLIM in interior British Columbia, Canada. *Environmental Modelling & Software,* 26**,** 644-657.

Marks, D., J. Domingo, D. Susong, T. Link & D. Garen (1999) A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes,* 13**,** 1935-1959.

Marks, D. & A. Winstral (2001) Comparison of Snow Deposition, the Snow Cover Energy Balance, and Snowmelt at Two Sites in a Semiarid Mountain Basin. *Journal of Hydrometeorology,* 2**,** 213-227.

Ospina, R., F. Cribari-Neto & K. L. P. Vasconcellos (2006) Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis,* 51**,** 960-981.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

Pierce, K. B., T. Lookingbill & D. Urban (2005) A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology,* 20**,** 137-147.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Rivington, M., G. Bellocchi, K. B. Matthews & K. Buchan (2005) Evaluation of three model estimations of solar radiation at 24 UK stations. *Agricultural and Forest Meteorology,* 132**,** 228-243.

Running, S. W., R. R. Nemani & R. D. Hungerford (1987) Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Canadian Journal of Forest Research,* 17**,** 472-483.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Simas, A. B., W. Barreto-Souza & A. V. Rocha (2010) Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis,* 54**,** 348-366.

Smithson, M. & J. Verkuilen (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods,* 11**,** 54-71.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

Thornton, P. E. & S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Winslow, J. C., E. R. Hunt Jr & S. C. Piper (2001) A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling,* 143**,** 227-243.

Zeng, Z. G., P. S. A. Beck, T. Wang, A. K. Skidmore, Y. L. Song, H. S. Gong & H. H. T. Prins (2010) Effects of plant phenology and solar radiation on seasonal movement of golden takin in the Qinling Mountains, China. *Journal of Mammalogy,* 91**,** 92-100.

CONCLUSIONS

Efficient implementation of off-grid power generation using photovoltaic technology is enhanced with accurate site-specific predictions of solar radiation characteristics such as frequency and duration of periods of low solar radiation. While simple monthly and yearly averages are useful, they do not provide the characterization of solar radiation at a variety of temporal scales, and they do not convey the day to day variability of solar radiation. This is especially important when implementing any type of directly coupled photovoltaic system that does not use a battery bank. A particular example of this is the directly couple photovoltaic pumping system, (DC-PVPS), which among many other uses, is employed to provide affordable access to clean drinking water for many third world communities (Lynn 2010, Posorski 1996) and provide necessary water for livestock (Boutelhig, A.Hadjarab and Bakelli 2011, Boutelhig et al. 2012). Like any natural system, this temporal solar radiation variability is accompanied by spatial variability.

To address the spatial variability of solar radiation, high density networks of solar monitoring sites are expanding (Reinbold, Roads and Brown 2005, Palmer 2011). As these currently large data sets are increasing, there is a need to accurately summarize and present this data to photovoltaic practitioners. To address this issue, we adapted intensity-duration-frequency (IDF) curves (Sherman 1931, Bernard 1932) for summarizing daily cumulative values of global solar radiation and demonstrated the usefulness (Chapter 2). The solar intensity-duration-frequency (SIDF) curves and the short-term solar intensity-duration-frequency (SSIDF) curves presented provide a quick and easy way to

summarize years of time series data into simple graphical output that is easily interpreted. Any end user implementing any photovoltaic device can predict roughly how often periods of inadequate solar radiation will be realized and to what extent should adaptations be implemented (i.e. larger array, larger storage tank, etc).

The third, fourth and fifth chapters of this dissertation all deal with predicting solar radiation in the absence of measured solar radiation data. Reliable predictions of solar radiation are a requisite to models of soil moisture (Spokas and Forcella 2006), carbon flux and plant growth (van Dijk, Dolman and Schulze 2005), wildlife behavior (Keating et al. 2007), evapotranspiration (Hargreaves and Samani 1982), weed management (Spokas and Forcella 2006) , hydrology (Zhou and Wang 2010), net primary productivity (Crabtree et al. 2009) and others. Numerous models have been proposed to predict solar radiation at ungauged locations because of the frequent lack of instrumentation to directly measure it (Thornton and Running 1999), but one particular family of models was the subject of these three chapters. This family of models, known as the B&C (Bristow and Campbell 1984) family, takes advantage of one simple relationship between higher levels of atmospheric water vapor and higher dewpoint temperatures. When a cooling air mass approaches dewpoint, the dispelled latent heat associated with condensation acts as a buffer leading to higher minimum temperatures. Therefore, night time minimum temperatures can often be associated with atmospheric water vapor content. Bristow and Campbell (1984) demonstrated the usefulness of these relationships with site-based models that predict global solar radiation using ΔT as an indicator of atmospheric transmittance.

Although all belonging to one family of models, a wide variety of models have been implemented that estimate solar radiation based on observations of the difference between the daily maximum and the daily minimum temperature (ΔT). One of the earliest was proposed by Hargreaves and Samani (1982) where the square root of extraterrestrial radiation is multiplied by ΔT and a coefficient, initially fixed at 0.75, but later adjusted by relative humidity. Bristow and Campbell (1984) proposed a model where transmittivity is a function of smoothed ΔT and three fitted parameters that are estimated for an individual site using historical data. Richardson (1985) proposed a simple model where ΔT is a function of two site specific empirical parameters and extraterrestrial radiation. Liu and Scott (2001) compare nine models that estimate solar radiation, three of which use only ΔT, two that use only precipitation and four that use both. Samani et al (2011) propose a modified version of Allen (1997), a model self-calibrated by season and location.

Thornton and Running (1999), proposed a ΔT method enhanced with precipitation and dew point data. Their motivation was to better estimate solar radiation for locations where no previously collected data is available. Fodor and Mika (2011) revisited ΔT models, and compared an 'S-shaped' function borrowed from soil science with Donatelli and Campbell's (1998) function for estimating the fraction of solar radiation that hits the Earth's surface.

Each of these models rely on global solar radiation being decomposed into three components in order to relate to ΔT to daily fluctuations in atmospheric attenuation. Extraterrestrial radiation (ETR) is the amount of solar radiation that hits the outside of the atmosphere. Clear sky transmittivity (CST), is the amount of ETR that will reach the

Earth's surface on a clear day. Fraction of clear day, (FCD), is the fraction of CST that hits the Earth's surface on any given day. By decomposing GSR in this manner, daily fluctuation in FCD can be related to daily fluctuations in climate variables.

ΔT is not the only climate variable that can be used to predict FCD. At a given temperature, and independent of barometric pressure, dew point is a function of absolute humidity. Night time low temperatures do tend towards dew point, however, for particularly dry nights the amount of heat released when the water vapor undergoes a phase change to frost is minimal, and thus temperatures can drop well below dew point. For this reason, low temperature can be included in any model that predicts FCD. Humidity and precipitation have also been found to be useful when modeling FCD (Thornton and Running 1999, Thornton et al. 2000). As more developments exploited the relationship between various climate variables and FCD, multi-step approaches were proposed. (Samani et al. 2011, Thornton and Running 1999, Thornton et al. 2000, Donatelli and Campbell 1998).

In an attempt to simplify the FCD modeling process, this dissertation presented the use of beta regression to relate FCD to a suite of explanatory climate variables. Like any multiple regression framework, beta regression can incorporate numerous variables, is flexible depending on which variables are available, provides measurements of uncertainty regarding predictions, and is robust to non-normal distributions and small data sets (Cribari-Neto and Zeileis 2010, Smithson and Verkuilen 2006).

In order to demonstrate the advantages of beta regression for estimating FCD, it was compared to a recently proposed model that had performed well when compared to

earlier models (Fodor and Mika 2011). The beta regression model outperformed the previous model (Fodor and Mika 2011) with a lower root mean squared error (RMSE) and mean absolute error (MAE) (Table 3.1 and 3.2). Overall, the RMSE was reduced an average of 17% and the MAE by 24%. The mean signed deviance (MSD) was generally higher in the beta regression model but in every case by less than 0.25 MJ m^{-2} d^{-1}. This slight increase in bias should not be a problem for most analyses.

Another advantage to using the beta regression model is the ability to combine strata through the addition of variables. When inspecting data output from networks of solar monitoring sites, it is not unusual to have low sample sizes for numerous strata (Fig. 3.3). This problem can be alleviated by combining strata. A single beta regression model was used to analyze the Redfield, SD data to determine if seasonal (spring, summer, etc.) and climate (wet vs. dry) stratification is necessary. The y*earday* variable was transformed to radians, (as it is circular data) and the sine and cosine components were entered into the model as covariates. Precipitation was left in the model as a continuous variable. The resulting RMSE was 19.735, which is lower than the RMSE from each of the individual models run on separate strata (19.989). This indicates that indeed one model per site can outperform eight separate models for the same site.

A natural extension of the regression framework is the inclusion of auto-regression, either temporal or spatial. In the case of chapter 3, the advantages of incorporating spatial auto-correlation were investigated. Whereas traditional models relied on data collected at one site over time (Fodor and Mika 2011, Bristow and Campbell 1984, Richardson 1985, Hargreaves and Samani 1982, Samani et al. 2011,

Ball, Purcell and Carey 2004, Running and Thornton 1999, Thornton, Hasenauer and White 2000), we presented a daily model that investigate the relationship between explanatory climate variables and FCD across many sites on one day. Two different daily models, one a beta regression, the other a universal kriging model, outperformed a beta regression site-based model with a lower RMSE and MAE (Table 4.1). The beta regression model displayed less bias throughout the response range of FCD, with the universal kriging less still (Fig 3.3).

Finally, the beta regression model was validated for use in mountainous regions (Chapter 5). The original MT-CLIM model (Running, Nemani and Hungerford 1987) used a modified version of Bristow and Campbell's (1984) original model to predict solar radiation. Two separate studies have validated improved versions of this solar radiation algorithm in complex mountainous terrain (Glassy and Running 1994, Thornton et al. 2000). However, both studies used homologous site (Glassy and Running 1994, Thornton et al. 2000). Thornton et al. (2000) did note a small adjustment was needed for snow covered slopes but otherwise the model performed well.

The beta regression approach produced good predictions of FCD in homologous site testing. The 95% upper and lower bounds of GSR captured the measured value 95.34% at WSSM and 95.41% at Porphyry across all available data (Table 5.3). The RMSE for WSSM (7.503 MJ $m^{-2}d^{-1}$) and Porphyry (9.181 MJ $m^{-2}d^{-1}$) (Table 5.2) compare favorably to previous studies. As comparison, the mean RMSE across six stations in central Oregon for a daily GSR model of the B&C family was 21.86 MJ $m^{-2}d^{-1}$ (Glassy and Running 1994)MSD. Thornton et al (2000) additionally implement a daily

B&C style model and report the mean absolute error (MAE) for 24 stations in Austria (high 4.72 and the low was 2.08). The MAE for WSSM (2.22 MJ $m^{-2}d^{-1}$) and Porphyry (3.05 MJ $m^{-2}d^{-1}$) (Table 5.2) were considerably lower. The limited sample size for this case study does not allow a direct comparison of methods, but the case studies here provide evidence that the beta regression method performed well. These results additionally show that that uncertainty intervals produced in the beta regression are valid and useful for GSR prediction in mountainous areas.

For predicting FCD and GSR, beta regression models consistently outperformed earlier models. The added advantages of accompanying estimates of uncertainty for predictions, flexibility of model construction, robustness to non-normal data and small data sets, ease of implementation in common statistical software, and the strong theoretical foundation that support it imply that the beta regression model is the best choice for modeling FCD when using a B&C model approach to predicting global solar radiation.

The beta regression models introduced here improve model fit, and offer numerous benefits over previously described models for predicting FCD. However, their overall effectiveness for predicting GSR in mountainous regions is limited by the reconstruction of GSR, and the method in which CST is derived. That is, after partitioning GSR into ETR, CST and FCD, one has to reconstruct GSR using predicted values of FCD and derived values of CST. As shown in Chapters 3, the uncertainty contributed to the final GSR predictions from deriving CST from historical data is very small, such that it was left out of the final predictions. However, when attempting to

derive CST from nearby locations in mountainous areas, clearly there is a bias (Figure 5.7 and Figure 5.8). Future studies should investigate this bias, and determine if it is possible to use either locally derived CST values or independently derived CST values to overcome this issue. Additional issues on which future studies should focus are increasing the precision of FCD predictions, incorporating the estimates of uncertainty into subsequent models, and refining automation of fitting spatial auto-correlation structure such that high volumes of daily data can be analyzed incorporating this information.

Despite these additional challenges to overcome, we introduced a new method to summarize and communicate solar radiation reliability and short-term variability. We applied a beta regression model that fully exploits explanatory meteorological variables, and outperformed previously proposed models when predicting global solar radiation. We refined the beta regression model to take advantage of implicit spatial auto-correlation, and investigated the use of universal kriging to explicitly model spatial auto-correlation for improved predictions. We investigated the use of the beta regression model in complex terrain, and identified weaknesses for further research. Our results illustrate substantial advancements in modeling the temporal and spatial variability of solar radiation.

Literature Cited

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Bernard, M. M. (1932) Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers,* 96**,** 592-624.

Boutelhig, A., A.Hadjarab & Y. Bakelli (2011) Comparison Study to Select an Optimum Photovoltaic Pumping System (PVPS) Configuration upon Experimental Performances Data of Two Different DC Pumps Tested at Ghardaïa Site. *Energy Procedia,* 6**,** 769-776.

Boutelhig, A., Y. Bakelli, I. Hadj Mahammed & A. Hadj Arab (2012) Performances study of different PV powered DC pump configurations for an optimum energy rating at different heads under the outdoor conditions of a desert area. *Energy,* 39**,** 33-39.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Crabtree, R., C. Potter, R. Mullen, J. Sheldon, S. Huang, J. Harmsen, A. Rodman & C. Jean (2009) A modeling and spatio-temporal analysis framework for monitoring environmental change using NPP as an ecosystem indicator. *Remote Sensing of Environment,* 113**,** 1486-1496.

Cribari-Neto, F. & A. Zeileis (2010) Beta Regression in R. *Journal of Statistical Software,* 34**,** 1-24.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Glassy, J. M. & S. W. Running (1994) Validating Diurnal Climatology Logic of the MT-CLIM Model Across a Climatic Gradient in Oregon. *Ecological Applications,* 4**,** 248-257.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Lynn, P. A. 2010. *Electricity from sunlight: An introduction to photovoltaics*. John Wiley & Sons, Ltd.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

Posorski, R. (1996) Photovoltaic water pumps, an attractive tool for rural drinking water supply. *Solar Energy,* 58**,** 155-163.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Richardson, C. W. (1985) Weather simulation for crop management models. *Transactions of the ASAE,* 28**,** 1602.

Running, S. W., R. R. Nemani & R. D. Hungerford (1987) Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Canadian Journal of Forest Research,* 17**,** 472-483.

Running, S. W. & P. E. Thornton. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Sherman, C. W. (1931) Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers,* 95**,** 951-960.

Smithson, M. & J. Verkuilen (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods,* 11**,** 54-71.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

LITERATURE CITED

Aguilar, C., J. Herrero & M. J. Polo (2010) Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrology & Earth System Sciences Discussions,* 7**,** 2373-2411.

Allen, R. G. (1997) Self-Calibrating Method for Estimating Solar Radiation from Air Temperature. *Journal of Hydrologic Engineering,* 2**,** 56-67.

Allen, R. G., R. Trezza & M. Tasumi (2006) Analytical integrated functions for daily solar radiation on slopes. *Agricultural and Forest Meteorology,* 139**,** 55-73.

Alsamamra, H., J. A. Ruiz-Arias, D. Pozo-Vázquez & J. Tovar-Pescador (2009) A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain. *Agricultural and Forest Meteorology,* 149**,** 1343-1357.

Badescu, V., C. A. Gueymard, S. Cheval, C. Oprea, M. Baciu, A. Dumitrescu, F. Iacobescu, I. Milos & C. Rada (2012) Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models. *Renewable and Sustainable Energy Reviews,* 16**,** 1636-1656.

Bailey, T. C. & A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. New York: Prentice Hall.

Ball, R. A., L. C. Purcell & S. K. Carey (2004) Evaluation of Solar Radiation Prediction Models in North America. *Agron J,* 96**,** 391-397.

Bandyopadhyay, A., A. Bhadra, N. S. Raghuwanshi & R. Singh (2008) Estimation of monthly solar radiation from measured air temperature extremes. *Agricultural and Forest Meteorology,* 148**,** 1707-1718.

Banerjee, S., B. P. Carlin & A. E. Gelfand. 2004. *Preface*. Chapman and Hall/CRC.

Barry, R., M. Prévost, J. Stein & A. P. Plamondon (1990) Application of a snow cover energy and mass balance model in a balsam fir forest. *Water Resour. Res.,* 26**,** 1079-1092.

Becker, P. & A. P. Smith (1990) Spatial autocorrelation of solar radiation in a tropical moist forest understory. *Agricultural and Forest Meteorology,* 52**,** 373-379.

Belcher, B. N. & A. T. DeGaetano (2007) A revised empirical model to estimate solar radiation using automated surface weather observations. *Solar Energy,* 81**,** 329-345.

Bernard, M. M. (1932) Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers,* 96**,** 592-624.

Berterretche, M., A. T. Hudak, W. B. Cohen, T. K. Maiersperger, S. T. Gower & J. Dungan (2005) Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment,* 96**,** 49-61.

Boutelhig, A., A.Hadjarab & Y. Bakelli (2011) Comparison Study to Select an Optimum Photovoltaic Pumping System (PVPS) Configuration upon Experimental Performances Data of Two Different DC Pumps Tested at Ghardaïa Site. *Energy Procedia,* 6**,** 769-776.

Boutelhig, A., Y. Bakelli, I. Hadj Mahammed & A. Hadj Arab (2012) Performances study of different PV powered DC pump configurations for an optimum energy rating at different heads under the outdoor conditions of a desert area. *Energy,* 39**,** 33-39.

Branscum, A. J., W. O. Johnson & M. C. Thurmond (2007) Bayesian beta regression: Applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics,* 49**,** 287-301.

Bristow, K. L. & G. S. Campbell (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology,* 31**,** 159-166.

Campbell, G. S. & J. M. Norman. 1998. *An Introduction to Environmental Biophysics*. New York: Springer.

Chang, J.-h. 1968. *Climate and Agriculture*. Chicago: Aldine Publishing.

Chien, L.-C. (2010) Diagnostic plots in beta-regression models. *Journal of Applied Statistics*.

Crabtree, R., C. Potter, R. Mullen, J. Sheldon, S. Huang, J. Harmsen, A. Rodman & C. Jean (2009) A modeling and spatio-temporal analysis framework for monitoring environmental change using NPP as an ecosystem indicator. *Remote Sensing of Environment,* 113**,** 1486-1496.

Cribari-Neto, F. & A. Zeileis (2010) Beta Regression in R. *Journal of Statistical Software,* 34**,** 1-24.

Dingman, S. L. 2002. *Physical Hydrology*. Upper Saddle River: Prentice Hall, Inc.

Diodato, N. & G. Bellocchi (2007) Modelling solar radiation over complex terrains using monthly climatological data. *Agricultural and Forest Meteorology,* 144**,** 111-126.

Donatelli, M. & G. S. Campbell. 1998. A simple model to estimate global solar radiation. In *Proceedings of Fifth European Society of Agronomy Congress, vol 2. The Slovac Agric. univ. Nitra, Slovak Republic*, 133 - 134.

Donatelli, M. & V. Marletto. 1994. Estimating surface solar radiation by means of air temperature. In *Proceedings of third annual ESA Congress, September 18-22. Abano, Italy*, 352-353.

Emanuel, R. E., H. E. Epstein, B. L. McGlynn, D. L. Welsch, D. J. Muth & P. D'Odorico (2010) Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains. *Water Resour. Res.,* 46**,** W11553.

Eskelson, B. N. I., L. Madsen, J. C. Hagar & H. Temesgen (2011) Estimating Riparian Understory Vegetation Cover with Beta Regression and Copula Models. *Forest Science,* 57**,** 212-221.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008a) Influence diagnostics in beta regression. *Computational Statistics & Data Analysis,* 52**,** 4417-4431.

Espinheira, P. L., S. L. P. Ferrari & F. Cribari-Neto (2008b) On beta regression residuals. *Journal of Applied Statistics,* 35**,** 407 - 419.

Evrendilek, F. & C. Ertekin (2007) Statistical Modeling of Spatio-Temporal Variability in Monthly Average Daily Solar Radiation over Turkey. *Sensors,* 7**,** 2763-2778.

Evrendilek, F. & C. Ertekin (2008) Assessing solar radiation models using multiple variables over Turkey. *Climate Dynamics,* 31**,** 131-149.

Federer, C. A. & C. B. Tanner (1966) Sensors for Measuring Light Available for Photosynthesis. *Ecology,* 47**,** 654-657.

Ferrari, S. & F. Cribari-Neto (2004) Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics,* 31**,** 799 - 815.

Ferrari, S. L. P. & E. C. Pinheiro (2011) Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation,* 81**,** 431 - 443.

Firatoglu, Z. A. & B. Yesilata (2004) New approaches on the optimization of directly coupled PV pumping systems. *Solar Energy,* 77**,** 81-93.

Fodor, N. & J. Mika (2011) Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology,* 151**,** 78-86.

Fu, P. & P. M. Rich (2002) A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture,* 37**,** 25-35.

Gates, D. M. 1980. *Biophysical Ecology*. New York: Springer-Verlag.

Geuder, N. & V. Quaschning (2006) Soiling of irradiation sensors and methods for soiling correction. *Solar Energy,* 80**,** 1402-1409.

Geuder, N., V. Quaschning, P. Viebahn, F. Steinsiek, J. Spies & C. Hendriks. 2004. Comparison of Terrestrial and Space Power Generation for Europe. In *4th International Conference on Solar Power from Space - SPS*. Granada, Spain.

Glassy, J. M. & S. W. Running (1994) Validating Diurnal Climatology Logic of the MT-CLIM Model Across a Climatic Gradient in Oregon. *Ecological Applications,* 4**,** 248-257.

Goh, K. L. & A. H. Pooi (1997) Adjustment of Prediction Intervals in Nonlinear Regression. *Biometrical Journal,* 39**,** 719-731.

Gopinathan, K. K. & A. Soler (1996) Effect of sunshine and solar declination on the computation of monthly mean daily diffuse solar radiation. *Renewable Energy,* 7**,** 89-93.

Granger, J. E. & R. E. Schulze (1977) Incoming Solar Radiation Patterns and Vegetation Response: Examples from the Natal Drakensberg. *Vegetatio,* 35**,** 47-54.

Grant, R. H., S. E. Hollinger, K. G. Hubbard, G. Hoogenboom & R. L. Vanderlip (2004) Ability to predict daily solar radiation values from interpolated climate records for use in crop simulation models. *Agricultural and Forest Meteorology,* 127**,** 65-75.

Grün, B., I. Kosmidis & A. Zeileis. 2011. Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Faculty of Economics and Statistics, University of Innsbruck.

Gueymard, C. A. & D. R. Myers. 2008. Solar Radiation Measurement: Progress in Radiometry for Improved Modeling. In *Modeling Solar Radiation at the Earth's Surface*, 1-27.

Gueymard, C. A. & D. R. Myers (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy,* 83**,** 171-185.

Haberlandt, U. (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology,* 332**,** 144-157.

Hargreaves, G. H. & Z. A. Samani (1982) Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers,* 108**,** 225-230.

Hasenauer, H., K. Merganicova, R. Petritsch, S. A. Pietsch & P. E. Thornton (2003) Validating daily climate interpolations over complex terrain in Austria. *Agricultural and Forest Meteorology,* 119**,** 87-107.

Horel, J. D. & X. Dong (2010) An Evaluation of the Distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology,* 49**,** 1563-1578.

Huang, S., P. Rich, R. Crabtree, C. Potter & P. Fu (2008) Modeling Monthly Near-Surface Air Temperature from Solar Radiation and Lapse Rate: Application over Complex Terrain in Yellowstone National Park. *Physical Geography,* 29**,** 158-178.

Hunger, M., J. Baumert & R. Holle (2011) Analysis of SF-6D Index Data: Is Beta Regression Appropriate? *Value in Health,* 14**,** 759-767.

Jin, J., X. Gao, S. Sorooshian, Z.-L. Yang, R. Bales, R. E. Dickinson, S.-F. Sun & G.-X. Wu (1999) One-dimensional snow water and energy balance model for vegetated surfaces. *Hydrological Processes,* 13**,** 2467-2482.

Jolly, W. M., J. M. Graham, A. Michaelis, R. Nemani & S. W. Running (2005) A flexible, integrated system for generating meteorological surfaces derived from point sources across multiple geographic scales. *Environmental Modelling & Software,* 20**,** 873-882.

Kang, H. & D. Scott (2008) An integrated spatio-temporal GIS toolkit for exploring intra-household interactions. *Transportation,* 35**,** 253-268.

Keating, K. A., P. J. P. Gogan, J. M. Vore & L. R. Irby (2007) A Simple Solar Radiation Index for Wildlife Habitat Studies. *The Journal of Wildlife Management,* 71**,** 1344-1348.

Korhonen, L., K. T. Korhonen, P. Stenberg, M. Maltamo & M. Rautiainen (2007) Local models for forest canopy cover with beta regression. *silva Fennica,* 41**,** 671-685.

Koutsoyiannis, D., D. Kozonis & A. Manetas (1998) A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology,* 206**,** 118-135.

Liu, D. L. & B. J. Scott (2001) Estimation of solar radiation in Australia from rainfall and temperature observations. *Agricultural and Forest Meteorology,* 106**,** 41-59.

Lloyd, C. D. & P. M. Atkinson (2004) Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. *International Journal of Applied Earth Observation and Geoinformation,* 5**,** 293-305.

Lo, Y.-H., J. A. Blanco, B. Seely, C. Welham & J. P. Kimmins (2011) Generating reliable meteorological data in mountainous areas with scarce presence of weather records: The performance of MTCLIM in interior British Columbia, Canada. *Environmental Modelling & Software,* 26**,** 644-657.

Lynn, P. A. 2010. *Electricity from sunlight: An introduction to photovoltaics*. John Wiley & Sons, Ltd.

Marks, D., J. Domingo, D. Susong, T. Link & D. Garen (1999) A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes,* 13**,** 1935-1959.

Marks, D. & A. Winstral (2001) Comparison of Snow Deposition, the Snow Cover Energy Balance, and Snowmelt at Two Sites in a Semiarid Mountain Basin. *Journal of Hydrometeorology,* 2**,** 213-227.

Matsuda, Y., K. Fujita, Y. Ageta & A. Sakai (2006) Estimation of atmospheric transmissivity of solar radiation from precipitation in the Himalaya and the Tibetan Plateau. *Annals of Glaciology,* 43**,** 344-350.

Meek, D. W. (1997) Estimation of maximum possible daily global solar radiation. *Agricultural and Forest Meteorology,* 87**,** 223-241.

Myers, D. R. (2005) Solar radiation modeling and measurements for renewable energy applications: data and model quality. *Energy,* 30**,** 1517-1531.

Ospina, R., F. Cribari-Neto & K. L. P. Vasconcellos (2006) Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis,* 51**,** 960-981.

Ospina, R. & S. Ferrari (2010) Inflated beta distributions. *Statistical Papers,* 51**,** 111-126.

Ospina, R. & S. L. P. Ferrari (2011) A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*.

Palmer, P. L. 2011. AgriMet: A Reclamation Tool for Irrigation Water Management. eds. R. E. Beighley Ii & M. W. Kilgore, 279-279. Palm Springs, CA: ASCE.

Pierce, K. B., T. Lookingbill & D. Urban (2005) A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology,* 20**,** 137-147.

Posorski, R. (1996) Photovoltaic water pumps, an attractive tool for rural drinking water supply. *Solar Energy,* 58**,** 155-163.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. Vienna, Austria}: R Foundation for Statistical Computing.

Rehman, S. & S. G. Ghori (2000) Spatial estimation of global solar radiation using geostatistics. *Renewable Energy,* 21**,** 583-605.

Reinbold, H. J., J. O. Roads & T. J. Brown (2005) Evaluation of the Experimental Climate Prediction Center's fire danger forecasts with remote automated weather station observations. *International Journal of Wildland Fire,* 14**,** 19-36.

Reusch, D. B. & R. B. Alley (2002) Automatic Weather Stations and Artificial Neural Networks: Improving the Instrumental Record in West Antarctica. *Monthly Weather Review,* 130.

Richardson, C. W. (1985) Weather simulation for crop management models. *Transactions of the ASAE,* 28**,** 1602.

Rivington, M., G. Bellocchi, K. B. Matthews & K. Buchan (2005) Evaluation of three model estimations of solar radiation at 24 UK stations. *Agricultural and Forest Meteorology,* 132**,** 228-243.

Rocha, A. & F. Cribari-Neto (2009) Beta autoregressive moving average models. *TEST,* 18**,** 529-545.

Rocha, A. & A. Simas (2011) Influence diagnostics in a general class of beta regression models. *TEST,* 20**,** 95-119.

Running, S. W., R. R. Nemani & R. D. Hungerford (1987) Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Canadian Journal of Forest Research,* 17**,** 472-483.

Running, S. W. & P. E. Thornton. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Samani, Z. A., G. H. Hargreaves, V. D. Tran & A. S. Bawazir (2011) Estimating solar radiation from temperature with spatial and temporal calibration. *Journal Irrigation and Drainage Engineering*.

Sherman, C. W. (1931) Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers,* 95**,** 951-960.

Simas, A. B., W. Barreto-Souza & A. V. Rocha (2010) Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis,* 54**,** 348-366.

Smithson, M. & J. Verkuilen (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods,* 11**,** 54-71.

Spiess, A.-N. & N. Neumeyer (2010) An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology,* 10**,** 6.

Spokas, K. & F. Forcella (2006) Estimating Hourly Incoming Solar Radiation from Limited Meteorological Data. *Weed Science,* 54**,** 182-189.

Srivastava, S. K., O. P. Singh & G. N. Pandey (1995) Correlations for the estimation of hourly global solar radiation. *Applied Energy,* 52**,** 55-64.

Tham, Y., T. Muneer & B. Davison (2010) Estimation of hourly averaged solar irradiation: evaluation of models. *Building Services Engineering Research and Technology,* 31**,** 9-25.

Thompson, R., M. Ventura & L. Camarero (2009) On the climate and weather of mountain and sub-arctic lakes in Europe and their susceptibility to future climate change. *Freshwater Biology,* 54**,** 2433-2451.

Thornton, P. E., H. Hasenauer & M. A. White (2000) Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology,* 104**,** 255-271.

Thornton, P. E. & S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. In *Agricultural & Forest Meteorology*, 211.

Thornton, P. E., S. W. Running & M. A. White (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology,* 190**,** 214-251.

van Dijk, A. I. J. M., A. J. Dolman & E.-D. Schulze (2005) Radiation, temperature, and leaf area explain ecosystem carbon fluxes in boreal and temperate European forests. *Global Biogeochem. Cycles,* 19**,** GB2029.

Veneziano, D., C. Lepore, A. Langousis & P. Furcolo (2007) Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall. *Water Resour. Res.,* 43.

White, J. D., S. W. Running, P. E. Thornton, R. E. Keane, K. C. Ryan, D. B. Fagre & C. H. Key (1998) Assessing simulated ecosystem processes for climate variability research at Glacier National Park. *Ecological Applications,* 8**,** 805.

Winslow, J. C., E. R. Hunt Jr & S. C. Piper (2001) A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling,* 143**,** 227-243.

Wu, H., K. G. Hubbard & J. You (2005) Some Concerns when Using Data from the Cooperative Weather Station Networks: A Nebraska Case Study. *Journal of Atmospheric and Oceanic Technology,* 22**,** 592-602.

Yesilata, B. & Z. A. Firatoglu (2008) Effect of solar radiation correlations on system sizing: PV pumping case. *Renewable Energy,* 33**,** 155-161.

Zeng, Z. G., P. S. A. Beck, T. Wang, A. K. Skidmore, Y. L. Song, H. S. Gong & H. H. T. Prins (2010) Effects of plant phenology and solar radiation on seasonal movement of golden takin in the Qinling Mountains, China. *Journal of Mammalogy,* 91**,** 92-100.

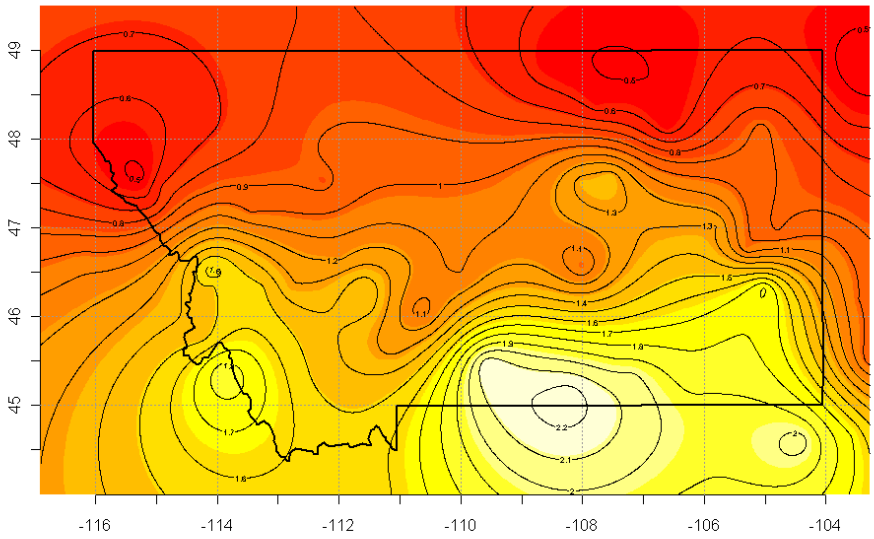
Zhou, Z.-m. & X.-c. Wang. 2010. Distributed Hydrological Model Based on Topography Trend. In *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on*, 1-4.

APPENDIX A

IDF CURVES, ED CURVES AND STATEWIDE MAPS FOR SOLAR RADIATION IN THE STATE OF MONTANA

Included is an excerpt from the full manual. All of the front matter is present, and one example of an SSIDF curve, ED curve, and Monthly statewide ED curve for each threshold is included. The manual has 366 pages total.

IDF Curves, ED Curves and Statewide Maps for Solar Radiation in the State of Montana.



Randall Mullen

Brian McGlynn Lucy Marshall

Department of Land Resources and Environmental Science
College of Agriculture
Montana State University
Bozeman, MT USA

September, 2011

MONTANA
STATE UNIVERSITY

An electronic version of this manual is available online.

http://watershed.montana.edu/analysis

The electronic version of this manual has navigation bookmarks.

If these are not currently seen, activate in Adobe Reader by clicking on

View → Show/Hide → Navigation Panes → Bookmarks

For all questions and correspondence, please contact

Randall Mullen …… randall.mullen@gmail.com

Dr. Lucy Marshall ……. lucyamarshall@gmail.com

Dr. Brian McGlynn …… blmcglynn@gmail.com

TABLE OF CONTENTS

SECTION 1 FRONT MATTER 6

Section 1.1 Introduction 6
Section 1.2 How to use this manual 8
Section 1.3 Intensity-Duration-Frequency curves 8
Section 1.4 How to read an IDF curve 9
Section 1.5 Exceedence-Duration curves 12
Section 1.6 How to read an ED curve 12
Section 1.7 Interpolated ED curve maps 14
Section 1.8 Determining reference station 15
Section 1.9 Acknowledgments 15
Section 1.10 Cited literature 15

SECTION 2 TABLES AND MAPS 17

Table of weather stations 18
Map of weather stations in Montana 20
Map with Thiessen polygons 21

SECTION 3 IDF CURVES (ALPHABETICAL BY FOUR LETTER CODE) 22

Ashton, ID (ahti) 23
Antelope Range, SD (anra) 26
Beach, ND (bech) 29
Blackfeet, MT (bfam) 32
Big Flat, MT (bftm) 35
Bowman, ND (bomn) 38
Broken-O Ranch, MT (bomt) 41
Bozeman, MT (bozm) 44
Buffalo Rapids – Glendive, MT (brgm) 47
Brorson, MT (brsn) 50
Buffalo Rapids – Terry, MT (brtm) 53
Corvalis, MT (covm) 56
Creston, MT (crsm) 59
Crosby, ND (csby) 62
Dworshak – Dent Acres, ID (deni) 65
Dillon, MT (dlnm) 68
Deer Lodge, MT (drlm) 71
Greenfields, MT (gfmt) 74
Glasgow, MT (glgm) 77
Harlem, MT (hrlm) 80
Helena Valley, MT (hvmt) 83
Kriley Creek, ID (ikri) 86
Jefferson River Valley, MT (jvwm) 89
Lower Musselshell, MT (mbra) 92
Malta, MT (matm) 95
Bradshaw Creek, MT (mbra) 98
Sheep Mountain, MT (mbsm) 101
Chain Buttes, MT (mcha) 104
Fishtail, MT (mfis) 107
Fort Howes, MT (mfoh) 110

Ginger, MT (mgin) 113
Knowlton, MT (mkno) 116
Little Bighorn, MT (mlih) 119
Nine Mile, MT (mnin) 122
Philipsburg, MT (mphl) 125
Pistol Creek, MT (mpis) 128
Plains, MT (mpln) 131
Poplar, MT (mpop) 134
Ronan, MT (mron) 137
Seeley Lake, MT (msee) 140
Sawmill Creek, MT (mssc) 143
Stevi, MT (mste) 146
St. Regis, MT (mstr) 149
Thompson Falls AP, MT (mtho) 152
West Fork, MT (mwef) 155
Wolf Mountain, MT (mwol) 158
Moccasin, MT (mwsm) 161
Nisland, SD (nsld) 164
Ruby River Valley, MT (rbym) 167
Roundbutte, MT (rdbm) 170
Rathdrum Praire, ID (rthi) 173
Rexburg, ID (rxgi) 176
Sidney, MT (sdny) 179
St. Ignatius, MT (sigm) 182
Shields Valley, MT (svwm) 185
Toston, MT (tosm) 188
Teton River, MT (trfm) 191
Upper Musselshell, MT (umhm) 194
Bear Lodge, WY (wbea) 197
Echeta, WY (wech) 200
Hillsboro, MT (whil) 203
Williston, ND (wlsn) 206
Rochelle Hills, WY (wrch) 209
White Sulphur Springs, MT (wssm) 212

SECTION 4 ED CUVES (ALPHABETICAL BY FOUR LETTER CODE) 215

Ashton, ID (ahti) 216
Antelope Range, SD (anra) 217
Beach, ND (bech) 218
Blackfeet, MT (bfam) 219
Big Flat, MT (bftm) 220
Bowman, ND (bomn) 221
Broken-O Ranch, MT (bomt) 222
Bozeman, MT (bozm) 223
Buffalo Rapids – Glendive, MT (brgm) 224
Brorson, MT (brsn) 225
Buffalo Rapids – Terry, MT (brtm) 226
Corvalis, MT (covm) 227
Creston, MT (crsm) 228
Crosby, ND (csby) 229
Dworshak – Dent Acres, ID (deni) 230
Dillon, MT (dlnm) 231
Deer Lodge, MT (drlm) 232
Greenfields, MT (gfmt) 233
Glasgow, MT (glgm) 234

Harlem, MT (hrlm) 235
Helena Valley, MT (hvmt) 236
Kriley Creek, ID (ikri) 237
Jefferson River Valley, MT (jvwm) 238
Lower Musselshell, MT (mbra) 239
Malta, MT (matm) 240
Bradshaw Creek, MT (mbra) 241
Sheep Mountain, MT (mbsm) 242
Chain Buttes, MT (mcha) 243
Fishtail, MT (mfis) 244
Fort Howes, MT (mfoh) 245
Ginger, MT (mgin) 246
Knowlton, MT (mkno) 247
Little Bighorn, MT (mlih) 248
Nine Mile, MT (mnin) 249
Philipsburg, MT (mphl) 250
Pistol Creek, MT (mpis) 251
Plains, MT (mpln) 252
Poplar, MT (mpop) 253
Ronan, MT (mron) 254
Seeley Lake, MT (msee) 255
Sawmill Creek, MT (mssc) 256
Stevi, MT (mste) 257
St. Regis, MT (mstr) 258
Thompson Falls AP, MT (mtho) 259
West Fork, MT (mwef) 260
Wolf Mountain, MT (mwol) 261
Moccasin, MT (mwsm) 262
Nisland, SD (nsld) 263
Ruby River Valley, MT (rbym) 264
Roundbutte, MT (rdbm) 265
Rathdrum Praire, ID (rthi) 266
Rexburg, ID (rxgi) 267
Sidney, MT (sdny) 268
St. Ignatius, MT (sigm) 269
Shields Valley, MT (svwm) 270
Toston, MT (tosm) 271
Teton River, MT (trfm) 272
Upper Musselshell, MT (umhm) 273
Bear Lodge, WY (wbea) 274
Echeta, WY (wech) 275
Hillsboro, MT (whil) 276
Williston, ND (wlsn) 277
Rochelle Hills, WY (wrch) 278
White Sulphur Springs, MT (wssm) 279

SECTION 5 STATEWIDE MAPS 280

January ....
400 WATTS/M^2 ....*281*
300 WATTS/M^2 ....*281*
200 WATTS/M^2 ....*282*
100 WATTS/M^2 ....*282*
February ....
500 WATTS/M^2 ....*283*
400 WATTS/M^2 ....*283*
300 WATTS/M^2 ....*284*

200 WATTS/M² *284*
100 WATTS/M² *285*
March .....
700 WATTS/M² *285*
600 WATTS/M² *286*
500 WATTS/M² *286*
400 WATTS/M² *287*
300 WATTS/M² *287*
200 WATTS/M² *288*
100 WATTS/M² *288*
April .....
800 WATTS/M² *289*
700 WATTS/M² *289*
600 WATTS/M² *290*
500 WATTS/M² *290*
400 WATTS/M² *291*
300 WATTS/M² *291*
200 WATTS/M² *292*
100 WATTS/M² *292*
May .....
900 WATTS/M² *293*
800 WATTS/M² *293*
700 WATTS/M² *294*
600 WATTS/M² *294*
500 WATTS/M² *295*
400 WATTS/M² *295*
300 WATTS/M² *296*
200 WATTS/M² *296*
100 WATTS/M² *297*
June .....
900 WATTS/M² *297*
800 WATTS/M² *298*
700 WATTS/M² *298*
600 WATTS/M² *299*
500 WATTS/M² *299*
400 WATTS/M² *300*
300 WATTS/M² *300*
200 WATTS/M² *301*
100 WATTS/M² *301*
July .....
900 WATTS/M² *302*
800 WATTS/M² *302*
700 WATTS/M² *303*
600 WATTS/M² *303*
500 WATTS/M² *304*
400 WATTS/M² *304*
300 WATTS/M² *305*
200 WATTS/M² *305*
100 WATTS/M² *306*
August .....
800 WATTS/M² *306*
700 WATTS/M² *307*
600 WATTS/M² *307*
500 WATTS/M² *308*
400 WATTS/M² *308*
300 WATTS/M² *309*

200 watts/m² *309*
100 watts/m² *310*
September .....
700 watts/m² *310*
600 watts/m² *311*
500 watts/m² *311*
400 watts/m² *312*
300 watts/m² *312*
200 watts/m² *313*
100 watts/m² *313*
October .....
600 watts/m² *314*
500 watts/m² *314*
400 watts/m² *315*
300 watts/m² *315*
200 watts/m² *316*
100 watts/m² *316*
November .....
400 watts/m² *317*
300 watts/m² *317*
200 watts/m² *318*
100 watts/m² *318*
December .....
300 watts/m² *319*
200 watts/m² *319*
100 watts/m² *320*

1.1 Introduction

As high density *solar monitoring site* networks expand and data sets grow, there is increasing need to accurately summarize and present this data to photovoltaic practitioners. Graphic summaries should be succinct, easy to read and communicate daily fluctuations (diurnal fluctuations), weather variations (the impact of cloud cover) and seasonal variations (related to the daily path of the sun) of both daily cumulative solar radiation values and threshold exceedence values. To meet this demand for the state of Montana, we adapt intensity-duration-frequency (IDF) curves (Sherman 1931, Bernard 1932) for summarizing daily cumulative values of global solar radiation, and introduce exceedence-duration (ED) curves for summarizing the average number of hours per day over a certain solar radiation threshold. Both of these techniques directly utilize solar radiation data collected at the earth's surface from multiple sites. Thiessen Polygons were created so that end users can determine which site (or combination of sites) to reference. State-wide interpolated maps were created for the ED curves.

There are a variety of ways to estimate solar radiation and users may notice that results presented here differ from other mapped output. Some maps produced by the photovoltaic industry use models that rely heavily on solar geometry and make broad corrections for atmospheric attenuation (the amount of solar radiation filtered out by the atmosphere). Other products incorporate maps from the National Renewable Energy Lab (NREL) which rely on the 1961-1990 data from the National Solar Radiation Database (NSRD). These maps of yearly and monthly averages use actual data but only from 9 sites for the entire state of Montana. The products presented here rely on 64 sites for up to 23 years of data. This method captures subtle trends in atmospheric attenuation that are difficult to model while providing good spatial and temporal resolution for estimating representative values of available solar radiation.

The 64 sites used for this analysis are from several sources. The Bureau of Reclamation operates 26 weather stations referred to as AgriMet (Agricultural Meteorology) sites in Montana. Twenty-one of these are east of the continental divide and are operated by the Great Plains regional office[1] (Figure 1.1). The Pacific Northwest regional office operates 5 stations west of the continental divide[2] (Figure 1.1). The High Plains Regional Climate Center (HPRCC) operates the Automated weather Data Network (AWDN) through North and South Dakota, Wyoming, and other Great Plains states[3]. They have two sites in Eastern Montana, and several in North and South Dakota that were used for this study (Figure 1.1). The Western Regional Climate Center stores and serves data collected as part of the Remote Automated Weather Station (RAWS) system[4]. The intent of this data is to aid in the monitoring of air quality, provide information for research applications and rate fire danger. These stations are not all in agricultural settings, and some are located at elevations higher than then the range of inference for this project (Figure 1.1). Each station in and near Montana was assessed for usability. This assessment

[1] http://www.usbr.gov/gp/agrimet/index.cfm

[2] http://www.usbr.gov/pn/agrimet/index.cfm

[3] http://www.hprcc.unl.edu/index.php

[4] http://raws.fam.nwcg.gov/

determined if there was significant topographical or local shading that could render the data unusable, and if the altitude was appropriate. Twenty-six RAWS sites were eventually chosen for inclusion.

All stations used in this study have been placed in locations generally free of topographical and local shading. While we propose state-wide coverage, the understanding is that interpolations will not be applicable to areas of the state above 5400 feet (1646 meters). Furthermore, uncertainty from extrapolations into valleys, sites with local shading, and regions that do not have a weather station will have to be considered by the end user.

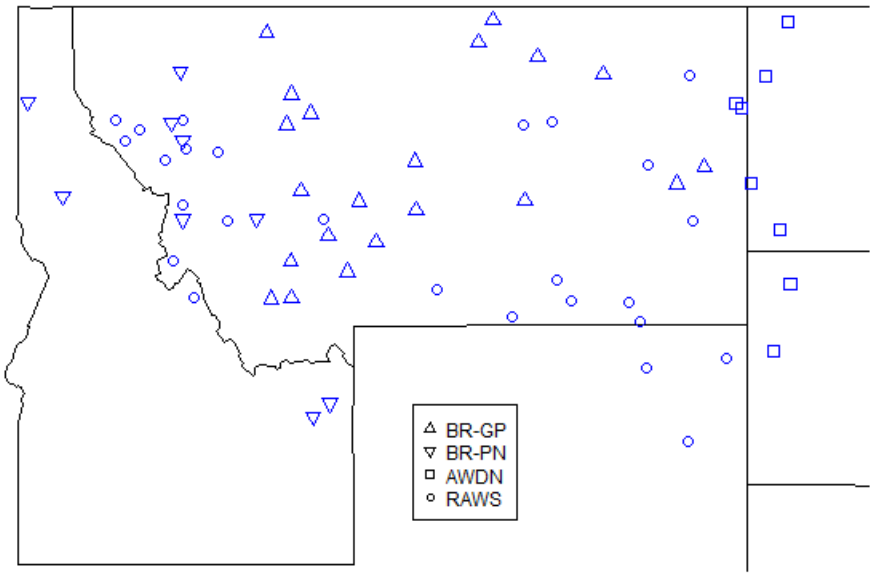


Figure 1.1 Map of station locations. Bureau of Reclamation sites are shown with triangles. Those pointed down indicate stations operated by the Pacific Northwest office while those pointed up indicate those operated by the Great Plains office. Squares shown stations operated by the high Plains Climate regional Climate Center. Circles indicate stations that are part of the Remote Automated Weather Station network that is served by the Western Regional Climate Center.

Each agency uses a Licor LI-200 (or similar) pyranometer designed primarily for field measurement of global solar radiation in agricultural, meteorological and solar energy studies. This sensor uses a silicon photovoltaic detector[5]. This sensor has been shown to have less than 5% error under natural daylight conditions (Federer and Tanner 1966) or as high as 25% error under adverse conditions (Geuder and Quaschning 2006) Using actual solar radiation data has definitive advantages over using modeled data. However, is not without problems. Gueymard (2009) outlines potential biases with common pyranometers used in many of today's weather stations. Different agencies have varying policies regarding maintenance and calibration for the pyranometers. If instruments are not cleaned regularly, they can record less solar radiation than

[5] http://www.licor.com/env/Products/Sensors/200/li200_description.jsp

what is present. Data is often corrupted by failures in data loggers, remote power sources, and temporary events that take days to weeks to fix (i.e. dust collecting on sensors, temporary shading due to obstruction, and broken weather stations). Agencies collecting the data provide varying amounts of quality control for hourly and cumulative data. The data presented herein has undergone additional quality control measures, but no guarantees can be made regarding the quality of the data. One distinct advantage of using IDF curves is that they are robust to missing data, so filtering can involve simple removal of corrupt data and does not necessarily need to be imputed. ED curves are less robust to missing data.

1.2 How to use this manual

We present here two types of solar radiation summaries. Intensity-duration-frequency (IDF) curves characterize trends in daily cumulative solar radiation values, and threshold-exceedence (ED) curves describe the expected average number of hours per day over various thresholds. IDF curves convey return intervals for runs of low (and high) radiation, while ED curves report the average number of hours per day over a given threshold. The derivation of these products is described in detail in sections 1.3 to 1.6. Monthly IDF (Section 3) and ED (Section 4) curves are presented for 64 sites in and near Montana. Maps are also provided that help users determine which weather station should be used as a reference station (described in detail in section 1.8 and shown in section 2). Table 2.1 lists the full name and corresponding 4-letter code of each weather station.

Summary information for the ED curves has also been interpolated state-wide and is provided in a series of maps (section 5). These maps are designed to supplement individual site ED values and to show general spatial trends. The derivation of these maps is explained in detail in section 1.7.

1.3 Intensity-Duration-Frequency Curves

Intensity-Duration-Frequency (IDF) curves were introduced in the 1930's to characterize return intervals for significant rainfall events (Bernard 1932; Dingman 2002; Sherman 1931). Traditionally, data might be yearly maximums for 1-hr, 6-hr and 24-hour rainfall of varying return periods (Dingman 2002). Herein, IDF curves have been adapted to characterize short-term variability of solar radiation intensities by estimating return intervals for 1 to 10 day spans of high and low solar radiation.

The process involves analysis of the full run of available solar radiation time series for a given site. For instance, if 20 years of daily cumulative values are available for a given location, and there is no missing data, that would yield 7305 data points, (365 days x 16 years + 366 days x 4 leap years = 7305 days), or $n = 7305$. The steps for creating the IDF curves in this publication are as follows.

1) Spans of 1, 3, 7 and 10 days were selected as time periods of interest for runs of low or high solar radiation values (note that time spans between 1 and 10 days can be estimated from the resulting curves).
2) Moving averages are calculated for each time span over the entire time series for a specific site.

3) All of the moving averages for each month are extracted and analyzed by month. For instance, in order to create IDF curves for July, data from July for all available years are extracted and then combined in this step.
4) Moving averages are ranked and the 0.04, 0.1, 0.2, 0.5, 0.8, 0.9 and 0.96 quantiles are calculated to determine the 25-day, 10-day, 5-day, and 2-day return intervals for the periods of high radiation and the periods of low radiation.
5) These solar radiation quantiles are graphed for each time span, resulting in a separate curve for each return interval. Each curve represents a specific return interval (e.g 10-day) and connects the points on the graph representing that quantile for each span in days.
6) Steps 2-4 are repeated for each site.

1.4 How to Read an IDF Curve

Once constructed, using an IDF curve to determine return intervals for periods of *low solar radiation* can be done as follows;

1) Determine the span in days that is relevant for the system. For example, if a three day period of insufficient incoming solar radiation is unacceptable, then locate 3 days on the x-axis and draw a vertical line up the chart.
2) Determine the daily cumulative solar radiation value of interest. This is the amount of incoming solar radiation that is necessary to operate the system. Draw a horizontal line across the chart.
3) The intersection of these lines indicates the return interval. If the point falls between two return interval curves then the return interval can be approximated by its proximity to both curves.
4) If the point lies above the lower 5-day return interval line and below the upper 5 day return interval line, then it falls into the expected solar radiation level (normal range) and interpreting return intervals does not make as much sense.

To determine return intervals for periods of *high solar radiation*, follow the instructions above except use the top 3 curves for return intervals. Similarly, if the point falls between two return interval curves then the return interval can be approximated by its proximity to both curves. As described in #4 above, any point lying between the 5-day return interval lines is falling into the expected solar radiation zone and does not lend itself to interpreting return intervals.

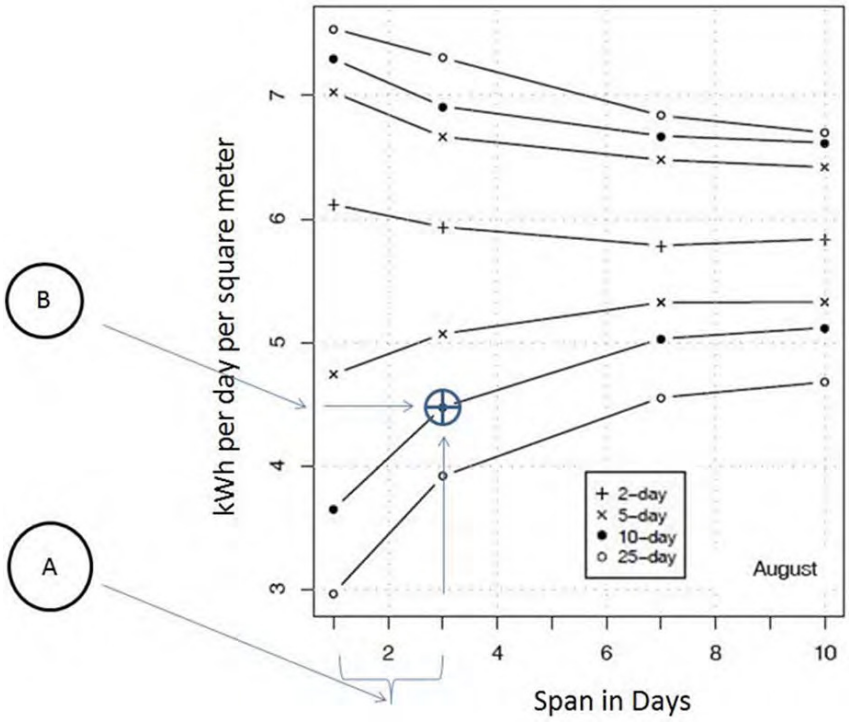


Figure 1.2 To interpret an IDF chart, A) Find the time span in days on the x-axis, in this case 3 days. Draw a vertical line up. B) Since the system requires 4.5 kWh (per m^2) of solar input per day, draw a horizontal line right from 4.5 kWh. 3) Interpret this case as, "about every 10 days, expect to start a run of 3 days of 4.5 kWh or less per day".

As an example, assume an off grid pumping system and tank that can store 3 days of water and requires 4.5 kWh/m^2 of solar radiation each day to run at full capacity. This system will run in August. Once the reference weather station has been determined (see section 1.8), consult the IDF curve corresponding to the month of August for that station (using the table of contents on page 1 to locate). Find the time span of interest, in this case 3 days, on the x-axis. Draw a line up from that point. Find the required kWh needed, in this case 4.5 kWh, on the y-axis and draw a horizontal line out from that point. These two lines cross at a point on the 10-day return interval line. This can be interpreted as, "a 3 day span of 4.5 kWh or less is expected to start about every 10 days during the month of August" (Figure 1.2). Figure 1.3 shows how to interpret different locations on an IDF graph.

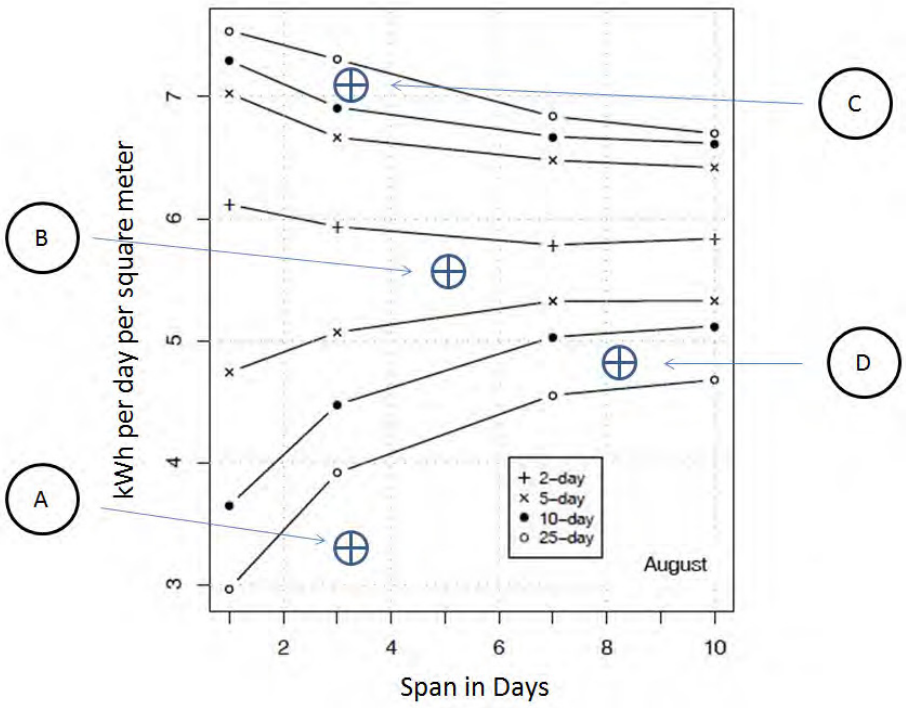


Figure 1.3 Refer to these examples of how to interpret the results. A) Three consecutive days with average intensity of 3.5 kWh or less will have a return interval greater than 25 days in the month of August. B) The 5.5 kWh average for 4 consecutive days falls between the upper and lower 5-day return interval lines, therefore, this is considered 'normal' solar radiation. C) Expect 3 day spans of average intensity *greater* than 7 kWh per day to have a return interval greater than 10 days but less than 25 days. D) Expect 8 day spans of 4.75 kWh or less per day to have a return interval between 10 days and 25 days.

For a second example, assume a home solar radiation system requires a minimum 5 kWh/m^2 of solar radiation for operation and to maintain a battery bank at 95% during the month of June. In order to properly maintain battery life, auxiliary power is utilized to recharge the system if the system drops below 95% for a period longer than five days. About how many times will auxiliary power be needed during a typical month of June? Using figure 1.4 as an example IDF curve, an estimate can be obtained as such. Find the 5 day span tick mark on the x-axis. Go up from there and find the intersection with the 5 kWh/m^2 tick mark. In this case, that intersection falls on the 10-day return interval line. This implies that about every 10 days, the end user should expect a span of 5 days below 5 kWh/m^2 to start. This would mean that auxiliary power would be needed about 3 times in the month of June. The end user can use this information to increase the number of solar panels if they desire to decrease the number of times auxiliary power is needed.

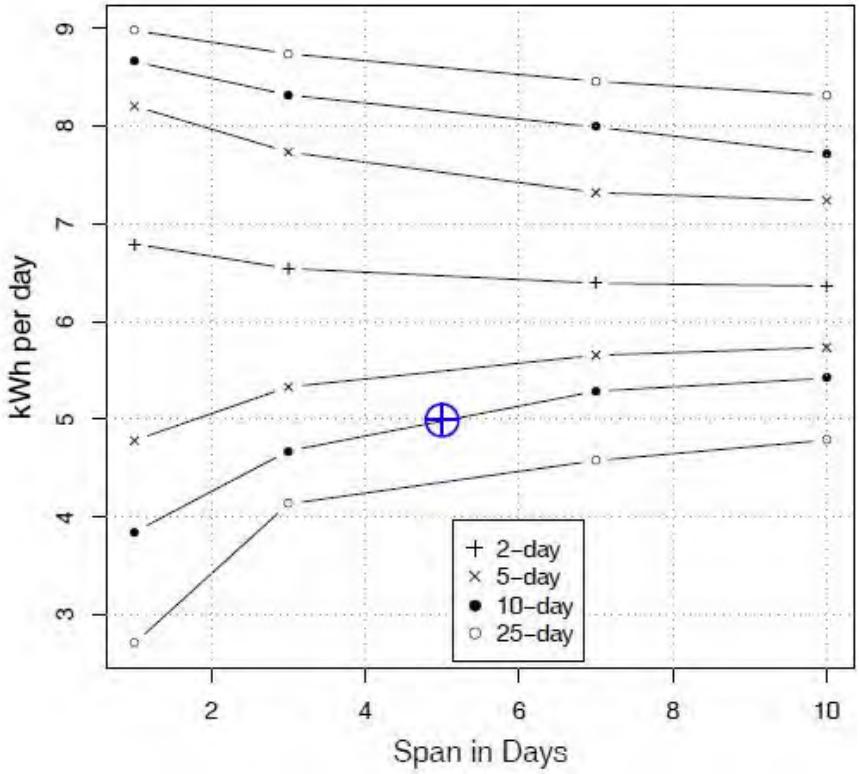


Figure 1.4 See text for complete detail. Assume end user wishes to estimate how many instances of a 5 day span of less than or equal to 5 kWh/m^2 are to be expected in a typical June. Assume the IDF curve above is from the appropriate station (see section 1.8). Expect a five day period to start about every 10 days. This would lead to about 3 instances during the month of June.

1.5 Exceedance-Duration Curve

Exceedance-duration (ED) curves are used for estimating the length of time in hours per day exceeding a given threshold (e.g. 600 watts/m^2). To construct each ED curve, the number of hours above a specified threshold is calculated for each day throughout the entire historical data set. This is repeated for various thresholds in increments of 100 watts/m2 (1000, 900, 800…100 watts/m^2). Then, the average numbers of hours per day that exceed this threshold are calculated for each month. These values are graphed for easy interpretation and interpolation (Figure 1.5).

1.6 How to read an ED Curve

Locate the threshold necessary for operation on the x-axis; the y-axis is the average number of observed hours over the selected threshold for the time period of interest. See figure 1.5 for an example on how to read an ED curve. Using this example ED curve, if the threshold of interest is 400 watts/m^2, then the number of hours per day expected in the month of September would be about 5 hours a day (Figure 1.5).

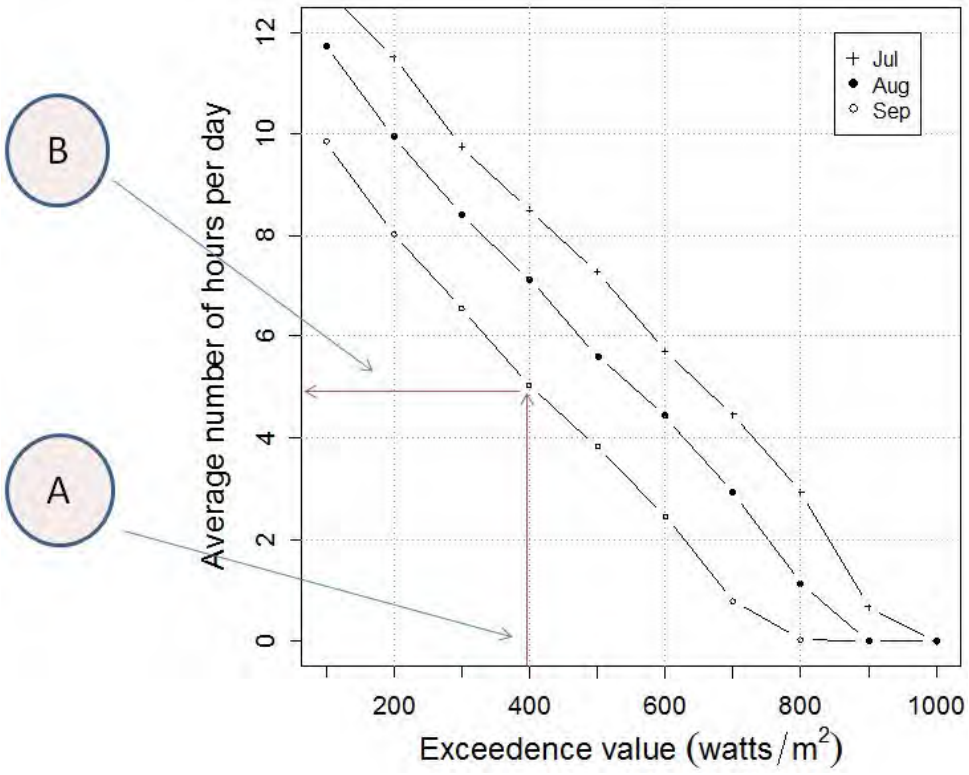


Figure 1.5 How to read an ED curve. Determine the critical threshold for the system. Here, that is 400 watts/m^2. A) From 400 watts/m^2, go up until hitting the month of interest. September is shown here. B) From that point, follow a horizontal line to the y-axis, in this case, 5. Interpret as, "an average of 5 hours per day greater than 400 watts/m^2 has been observed at this site for the month of September".

Consider a location in north eastern Montana, near Malta, Montana, (latitude 48° 37′, longitude -107° 78′, elevation 2270 feet) for a second example. Assume an end user wishes to pump 500 gallons of water per day from a water well to a storage tank. A typical fixed position 1 m^2 panel can produce the energy needed to run a typical pump when incoming solar radiation exceeds 600 watts per meter square. The AgriMet site near Malta, Montana has an Excedence Duration Curve shown in figure 1.6. Reading up from the 600 watts/m^2 value on the x-axis it can be seen that an average of about 4 hours per day exceeding 600 watts per meter square have been observed in the past during the month of May. Since the pump can operate for 4 hours and pump 2.14 gallons per minute, (according to manufacturer) then the total amount of water pumped per day is about 513 gallons per day in May. This assumes that past observations are indicative of future solar radiation trends.

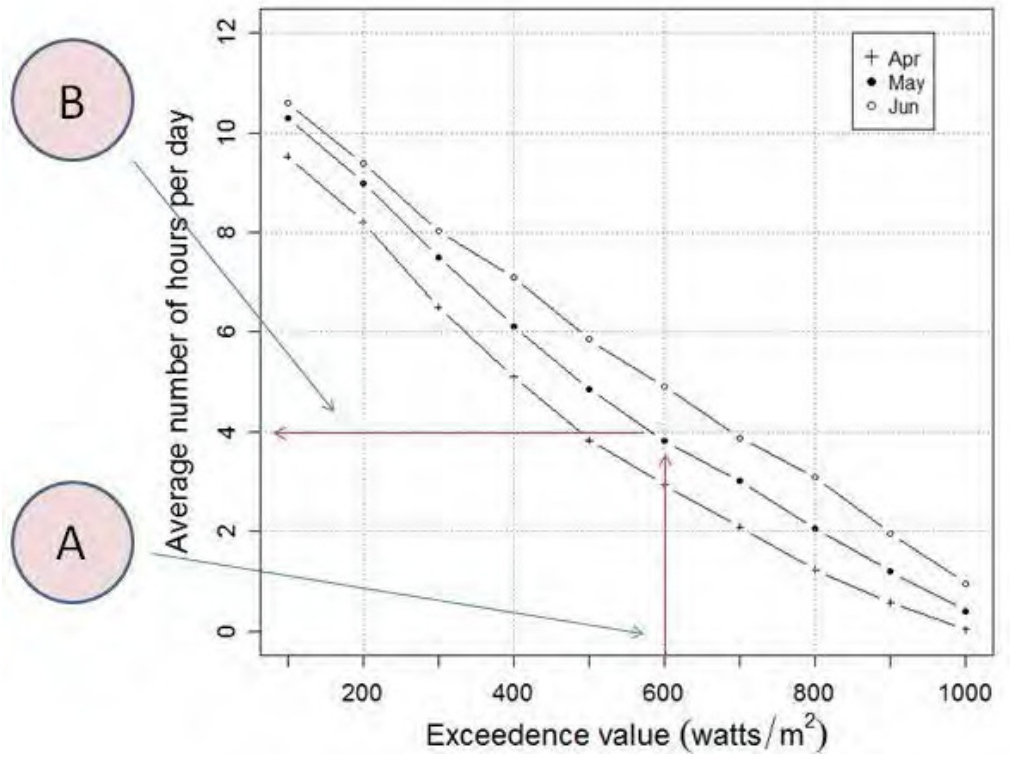


Figure 1.6 How to read an ED curve. Determine the critical threshold for the system. Here, that is 600 watts/m^2. A) From 600 watts/m^2, go up until hitting the month of interest. September is shown here. B) From that point, follow a horizontal line to the y-axis, in this case, 4. Interpret as, "an average of 4 hours per day greater than 600 watts/m^2 has been observed at this site for the month of September".

1.7 Interpolated ED Curve maps

The information contained in ED curves has been interpolated state-wide in the form of contour maps (Section 5). The method used for this was principally ordinary kriging. However, when not enough spatial auto-correlation was present for the kriging procedure, inverse distance weighting with a distance limit of 300 miles and 10 stations was used. A small 'idw' in the lower right hand corner of a map indicates that inverse distance weighting was used.

NOTE: All weather stations used for this analysis were below 5400 feet and were free of topographic and local shading. State-wide maps must be read with this in mind, and corrections must be made for mountainous regions.

1.8 Determining reference station

Three factors should be considered while choosing a reference weather station; distance, weather patterns, and length of data set.

The Thiessen polygon map (Figure 2.2) should be consulted to determine which weather station is the shortest distance to the point of interest. Thiessen polygons are constructed so that each point inside of the polygon is closest to the weather station with which it shares a polygon. Polygons at the edge of the state with no station shown reference an out-of-state station. The 4 letter code is shown and is valid for all uses in this manual. As a cautionary note, the closest weather station may not be the station that shares the same weather pattern as the location of interest. An end user may want to pick a station further away if it shares a more similar weather pattern with the location of interest depending on their knowledge of the site.

Each station has been collecting solar radiation data for a different length of time. This can vary from 3 to 23 years. The time frame that the data set covers at each station is listed on each page with IDF and ED graphs. In some cases, it may be beneficial to reference a station that is slightly further away, but has a substantially longer data set.

1.9 Acknowledgements

The authors would like to thank the entities that made these products possible. The Bureau of Reclamation operates the AgriMet weather stations. A special thanks to Tim Groove in the Great Plains office and Peter Palmer in the Pacific Northwest office. The High Plains Regional Climate center operates and maintains the Automated Weather Data Network (AWDN) sites in Montana, North Dakota and South Dakota that were used for this study. We would also like to thank the Western Region Climate Center for serving and distributing the data that is collected by various cooperative agencies for the Remote Automated Weather Station (RAWS) network.

Several vendors have provided input into this product, and we'd like to thank them as well. Dwight Patterson of GenPro Energy Solutions provided some excellent insight into what would help vendors of photovoltaic technologies. Sarah Ray and Sara Biddle of Independent Power Systems provided excellent feedback and motivation for this project.

The authors would like to extend a special thank you to the National Resources Conservation Service (NRCS) for funding this project and providing excellent motivation for the analysis. Specifically, we'd like to thank Steve Becker who helped us to focus on the needs of the end users.

1.10 Literature Cited

Bernard, M. M. (1932) Formulas for rainfall intensities of long duration. *Transactions of the American Society of Civil Engineers,* 96, 592-624.

Dingman, S. L. 2002. *Physical Hydrology*. Upper Saddle River: Prentice Hall, Inc.

Federer, C. A. & C. B. Tanner (1966) Sensors for Measuring Light Available for Photosynthesis. *Ecology,* 47, 654-657.

Geuder, N. & V. Quaschning (2006) Soiling of irradiation sensors and methods for soiling correction. *Solar Energy,* 80, 1402-1409.

Sherman, C. W. (1931) Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the American Society of Civil Engineers,* 95, 951-960.

Section 2
Tables and Maps

Code	Name / location	Elev (ft)	Lat	Lon	State	Source
AHTI	Ashton	5300	44.03	-111.47	ID	Bureau of Reclamation - Pacific Northwest
ANRA	Antelope Range	2890	45.52	-103.28	SD	High Plains Regional Climate Center
BECH	Beach	2899	46.78	-103.97	ND	High Plains Regional Climate Center
BFAM	Blackfeet	3905	48.68	-112.59	MT	Bureau of Reclamation - Great Plains
BFTM	Big Flat	3103	48.84	-108.56	MT	Bureau of Reclamation - Great Plains
BOMN	Bowman	2994	46.20	-103.47	ND	High Plains Regional Climate Center
BOMT	Broken-O Ranch	3890	47.52	-112.25	MT	Bureau of Reclamation - Great Plains
BOZM	Bozeman	4775	45.67	-111.15	MT	Bureau of Reclamation - Great Plains
BRGM	Buffalo Rapids-Glendive	2140	46.99	-104.80	MT	Bureau of Reclamation - Great Plains
BRSN	Brorson	2266	47.78	-104.25	MT	High Plains Regional Climate Center
BRTM	Buffalo Rapids-Terry	2270	46.78	-105.30	MT	Bureau of Reclamation - Great Plains
COVM	Corvallis	3597	46.33	-114.08	MT	Bureau of Reclamation - Pacific Northwest
CRSM	Creston	2950	48.19	-114.13	MT	Bureau of Reclamation - Pacific Northwest
CSBY	Crosby	2086	48.80	-103.32	ND	High Plains Regional Climate Center
DENI	Dworshak-Dent Acres	1660	46.62	-116.22	ID	Bureau of Reclamation - Pacific Northwest
DLNM	Dillon	5000	45.33	-112.51	MT	Bureau of Reclamation - Great Plains
DRLM	Deer Lodge	4680	46.34	-112.77	MT	Bureau of Reclamation - Pacific Northwest
GFMT	Greenfields	3820	47.66	-111.81	MT	Bureau of Reclamation - Great Plains
GLGM	Glasgow	2084	48.14	-106.61	MT	Bureau of Reclamation - Great Plains
HRLM	Harlem	2358	48.54	-108.83	MT	Bureau of Reclamation - Great Plains
HVMT	Helena Valley	3673	46.68	-111.98	MT	Bureau of Reclamation - Great Plains
IKRI	Kriley Creek	5200	45.36	-113.89	ID	Western Regional Climate Center
JVWM	Jefferson River Valley	4415	45.80	-112.17	MT	Bureau of Reclamation - Great Plains
LMMM	Lower Musselshell	2951	46.56	-108.01	MT	Bureau of Reclamation - Great Plains
MATM	Malta	2270	48.37	-107.78	MT	Bureau of Reclamation - Great Plains
MBRA	Bradshaw Creek	3930	45.06	-105.95	MT	Western Regional Climate Center
MBSM	Big Sheep Mountain	3200	47.02	-105.82	MT	Western Regional Climate Center
MCHA	Chain Buttes	2928	47.52	-108.03	MT	Western Regional Climate Center
MFIS	Fishtail	4550	45.46	-109.57	MT	Western Regional Climate Center
MFOH	Fort Howes	3380	45.30	-106.16	MT	Western Regional Climate Center
MGIN	Ginger	4370	46.33	-111.59	MT	Western Regional Climate Center
MKNO	Knowlton	3320	46.31	-105.02	MT	Western Regional Climate Center
MLIH	Little Bighorn	3400	45.57	-107.44	MT	Western Regional Climate Center
MNIN	Nine Mile	3300	47.07	-114.40	MT	Western Regional Climate Center
MPHL	Philipsburg	5280	46.32	-113.30	MT	Western Regional Climate Center
MPIS	Pistol Creek	5000	47.22	-114.02	MT	Western Regional Climate Center

Code	Name / location	Elev (ft)	Lat	Lon	State	Source
MPLN	Plains	2400	47.45	-114.87	MT	Western Regional Climate Center
MPOP	Poplar	2423	48.13	-105.07	MT	Western Regional Climate Center
MRON	Ronan	3060	47.57	-114.08	MT	Western Regional Climate Center
MSEE	Seeley Lake	4235	47.18	-113.45	MT	Western Regional Climate Center
MSSC	South Sawmill Creek	3290	47.56	-107.53	MT	Western Regional Climate Center
MSTE	Stevi	3365	46.51	-114.09	MT	Western Regional Climate Center
MSTR	St. Regis	2680	47.31	-115.11	MT	Western Regional Climate Center
MTHO	Thompson Falls AP	2460	47.58	-115.29	MT	Western Regional Climate Center
MWEF	West Fork	5200	45.82	-114.26	MT	Western Regional Climate Center
MWOL	Wolf Mountain	5217	45.31	-107.17	MT	Western Regional Climate Center
MWSM	Moccasin	4243	47.06	-109.95	MT	Bureau of Reclamation - Great Plains
NSLD	Nisland	2912	44.68	-103.57	SD	High Plains Regional Climate Center
RBYM	Ruby River Valley	5250	45.35	-112.15	MT	Bureau of Reclamation - Great Plains
RDBM	Roundbutte	3040	47.54	-114.28	MT	Bureau of Reclamation - Pacific Northwest
RTHI	Rathdrum Prairie	2210	47.80	-116.83	ID	Bureau of Reclamation - Pacific Northwest
RXGI	Rexburg	4875	43.85	-111.77	ID	Bureau of Reclamation - Pacific Northwest
SDNY	Sidney	1918	47.73	-104.15	MT	High Plains Regional Climate Center
SIGM	St. Ignatius	2980	47.33	-114.08	MT	Bureau of Reclamation - Pacific Northwest
SVWM	Shields Valley	5310	46.05	-110.65	MT	Bureau of Reclamation - Great Plains
TOSM	Toston	3920	46.17	-111.48	MT	Bureau of Reclamation - Great Plains
TOSM	Toston	4058	46.12	-111.49	MT	Bureau of Reclamation - Great Plains
TRFM	Teton River	3854	47.90	-112.16	MT	Bureau of Reclamation - Great Plains
UMHM	Upper Musselshell	4360	46.45	-109.94	MT	Bureau of Reclamation - Great Plains
WBEA	Bear Lodge	5280	44.60	-104.43	WY	Western Regional Climate Center
WECH	Echeta	4320	44.47	-105.85	WY	Western Regional Climate Center
WHIL	Hillsboro	3986	45.10	-108.22	MT	Western Regional Climate Center
WLSN	Williston	2099	48.13	-103.73	ND	High Plains Regional Climate Center
WRCH	Rochelle Hills	5199	43.55	-105.09	WY	Western Regional Climate Center
WSSM	White Sulphur Springs	4969	46.55	-110.95	MT	Bureau of Reclamation - Great Plains

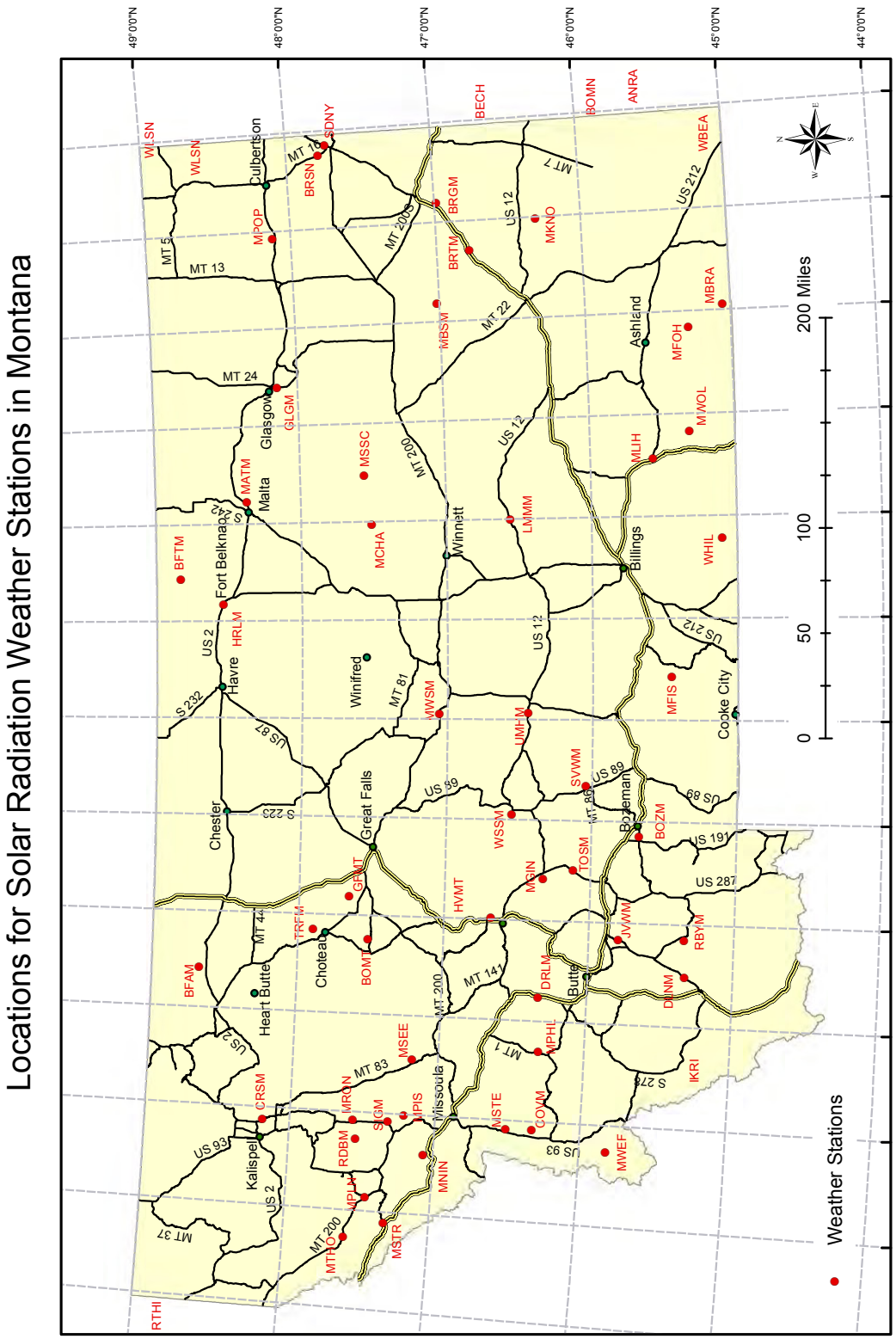


Figure 2.1 Weather stations that were used for solar radiation analysis are shown with red circles. Corresponding name codes are shown in red. Codes can be matched with full names, exact locations, and other information in appendix A.

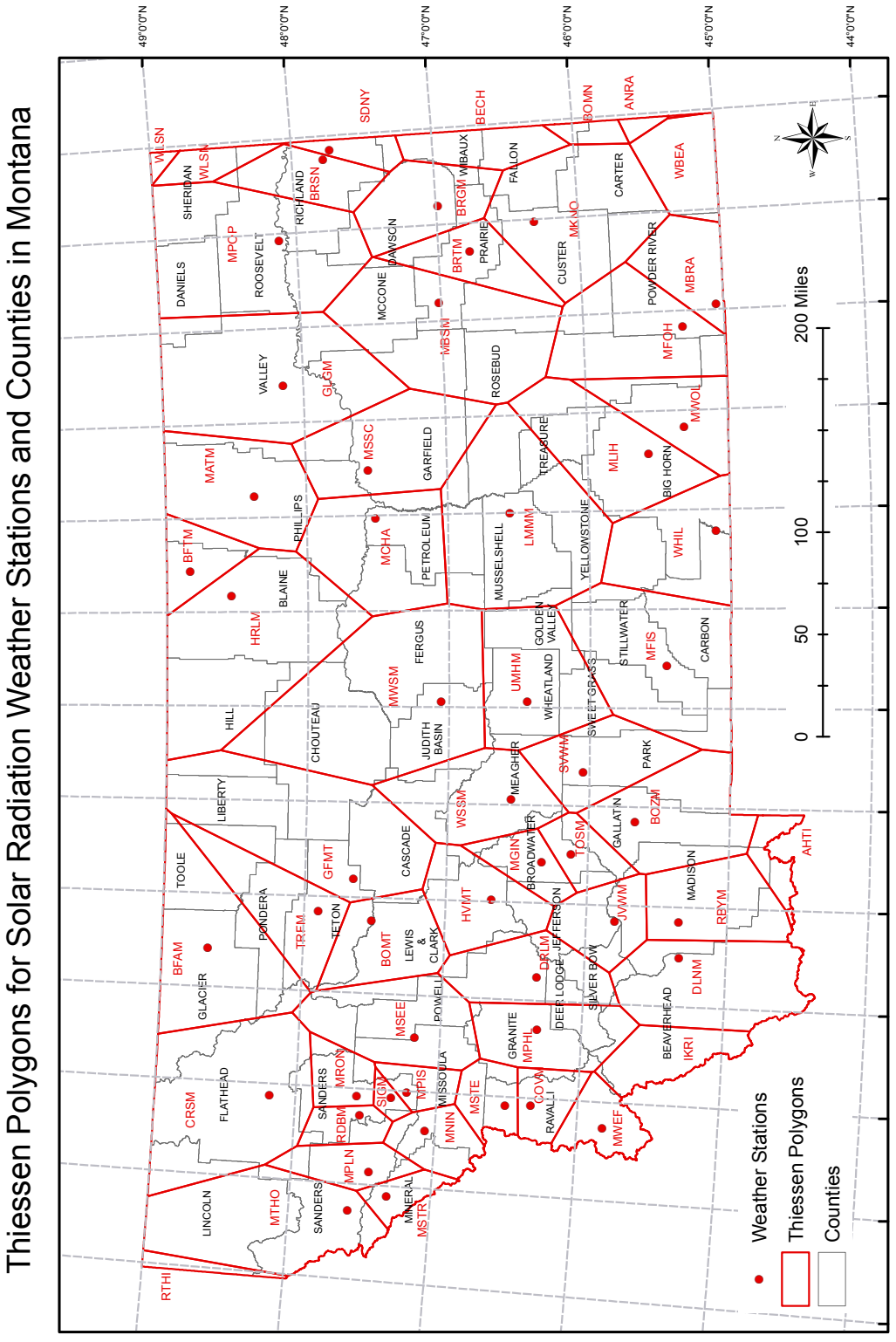
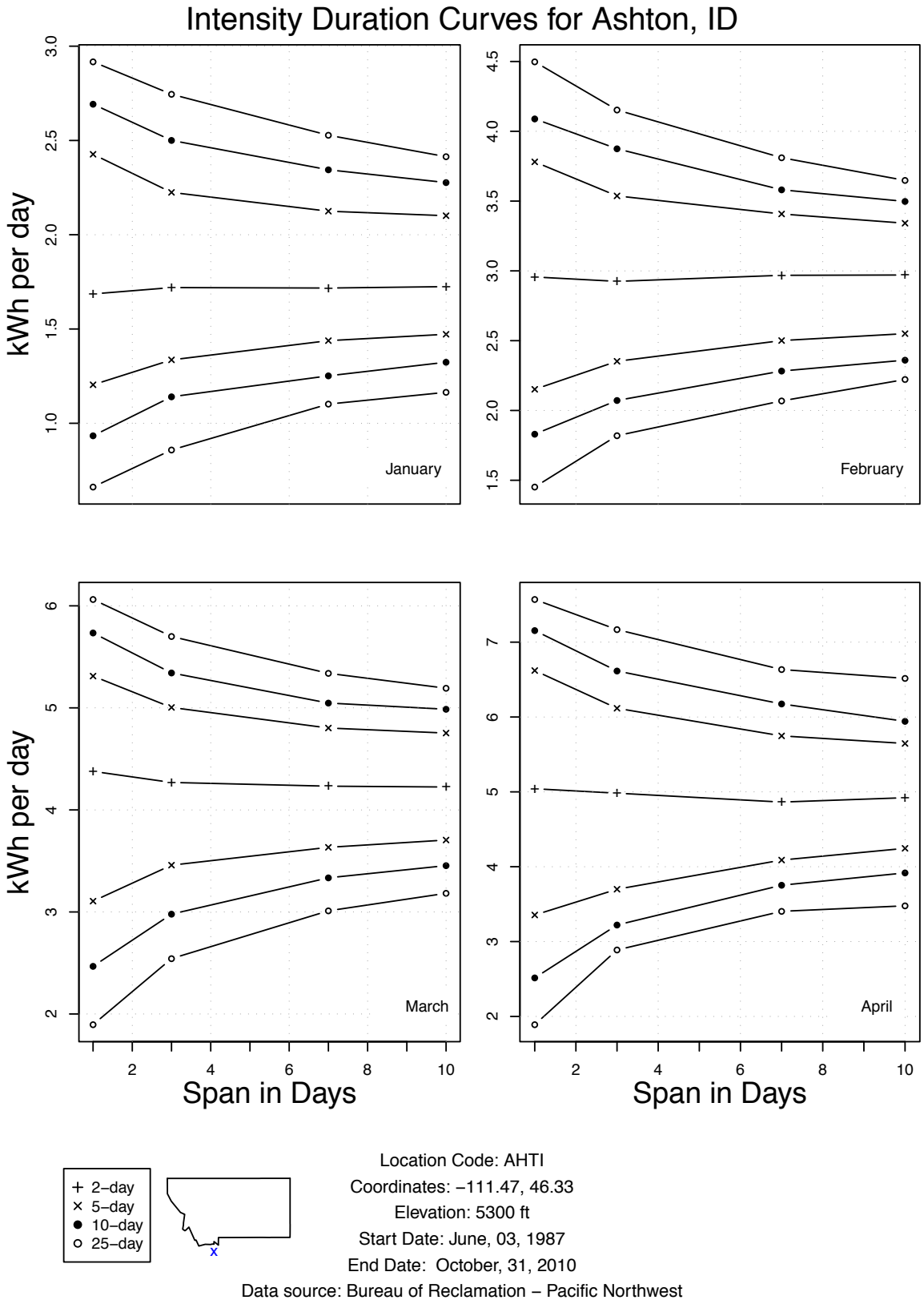
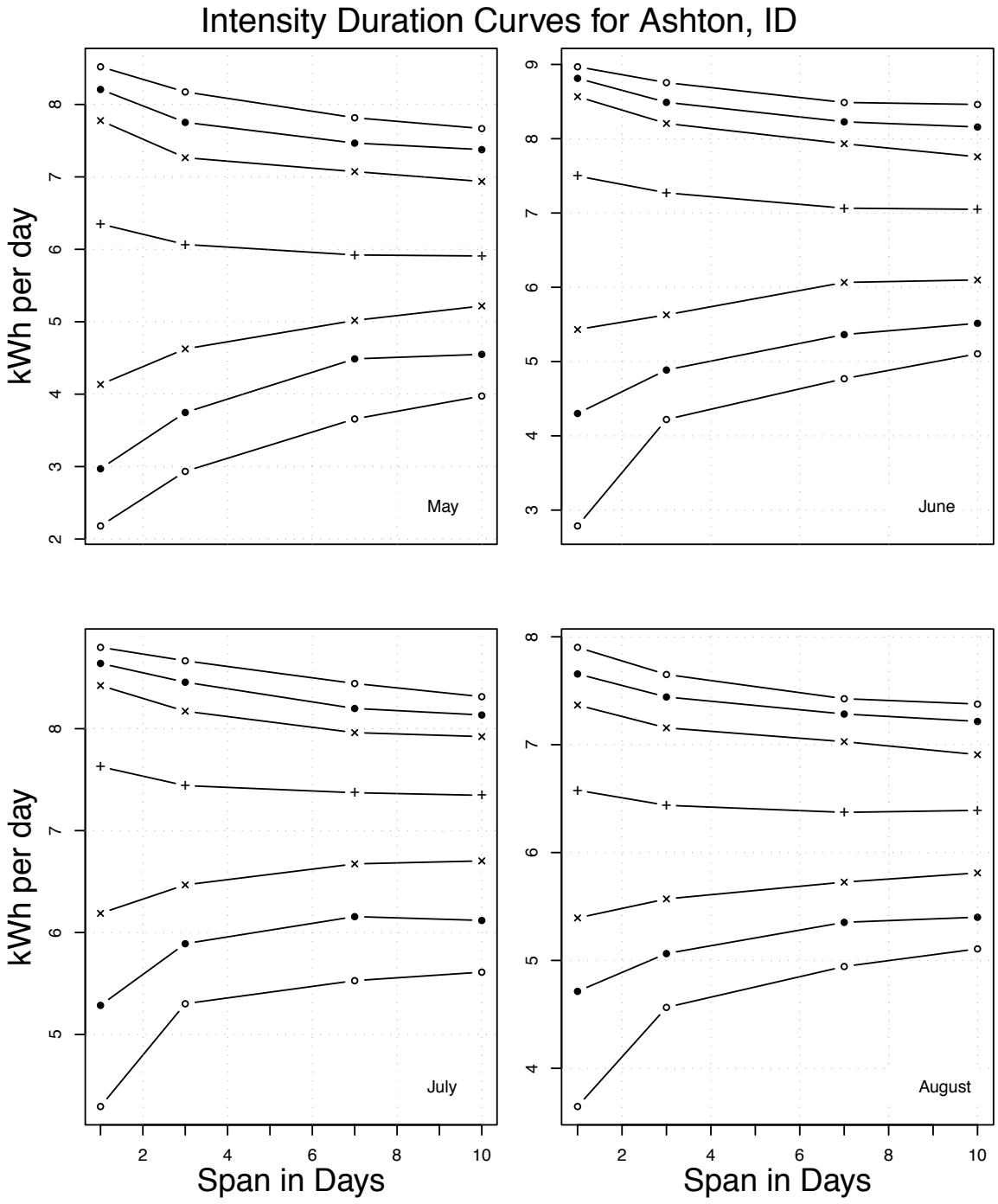


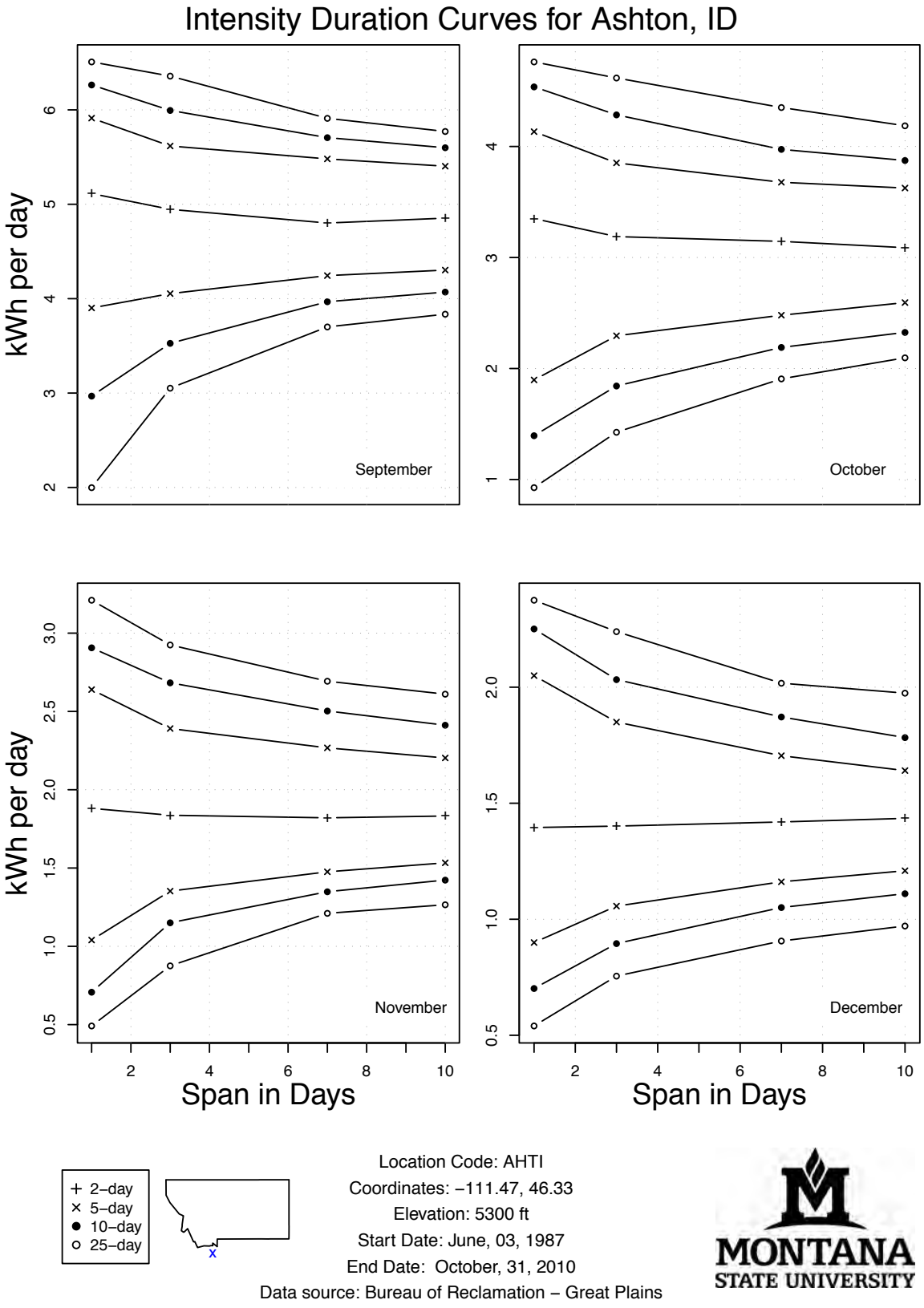
Figure 2.2 Weather stations that were used for solar radiation analysis are shown with red circles. Corresponding name codes and Thiessen polygons are also shown in red. By definition, all points inside of a polygon are closest to the weather station in that polygon. Codes can be matched with full names, exact locations, and other information in appendix A.

Intensity – Duration – Frequency (IDF) Curves for the State of Montana and nearby sites.

Please see Section 1.4 in the introduction for directions on how to interpret an IDF curve.


Intensity Duration Curves for Ashton, ID
kWh per day
Span in Days
January
February
March
April
2–day
5–day
10–day
25–day
Location Code: AHTI
Coordinates: –111.47, 46.33
Elevation: 5300 ft
Start Date: June, 03, 1987
End Date: October, 31, 2010
Data source: Bureau of Reclamation – Pacific Northwest

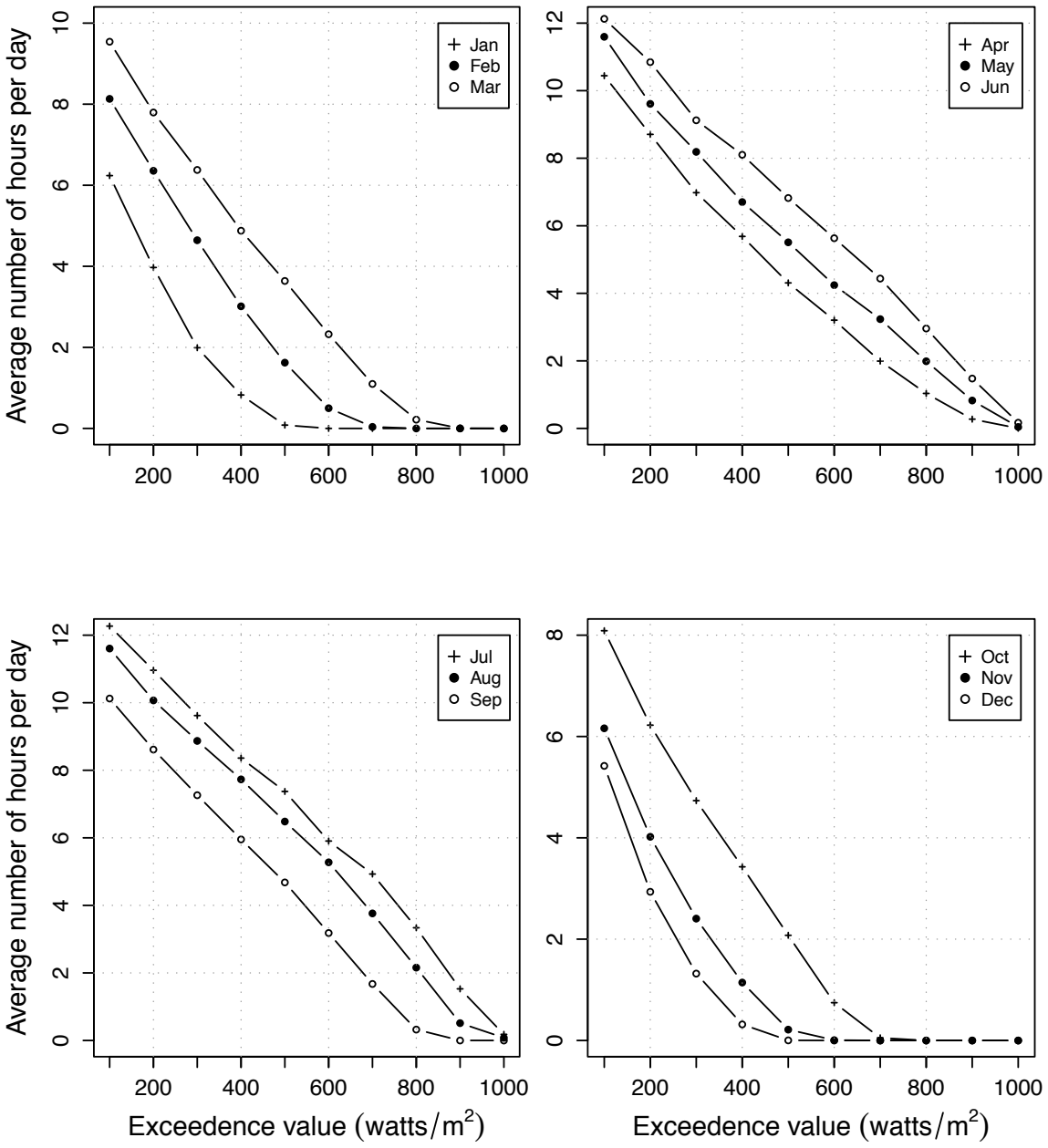

Intensity Duration Curves for Ashton, ID
kWh per day
May
June
July
August
Span in Days
Span in Days
+ 2–day
× 5–day
● 10–day
○ 25–day
Location Code: AHTI
Coordinates: –111.47, 46.33
Elevation: 5300 ft
Start Date: June, 03, 1987
End Date: October, 31, 2010
Data source: Bureau of Reclamation – Great Plains

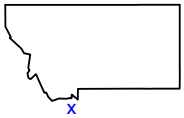

Intensity Duration Curves for Ashton, ID
kWh per day
September
October
November
December
Span in Days
+ 2–day
× 5–day
● 10–day
○ 25–day
Location Code: AHTI
Coordinates: –111.47, 46.33
Elevation: 5300 ft
Start Date: June, 03, 1987
End Date: October, 31, 2010
Data source: Bureau of Reclamation – Great Plains
MONTANA
STATE UNIVERSITY

Exceedence – Duration (ED) Curves for the State of Montana and nearby sites.

Please see Section 1.6 in the introduction for directions on how to interpret an ED curve.

Exceedence Duration Curves for Ashton, ID





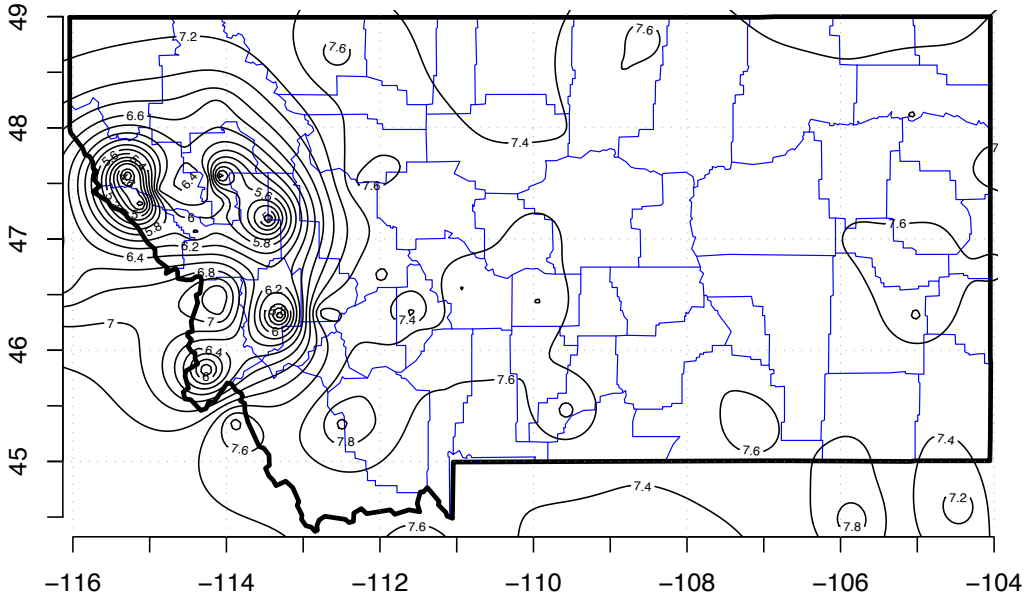
Location Code: AHTI
Coordinates: –111.47, 46.33
Elevation: 5300 ft
Start Date: June, 02, 1987
End Date: December, 31, 2010
Data source: Bureau of Reclamation – Pacific Northwest

MONTANA STATE UNIVERSITY

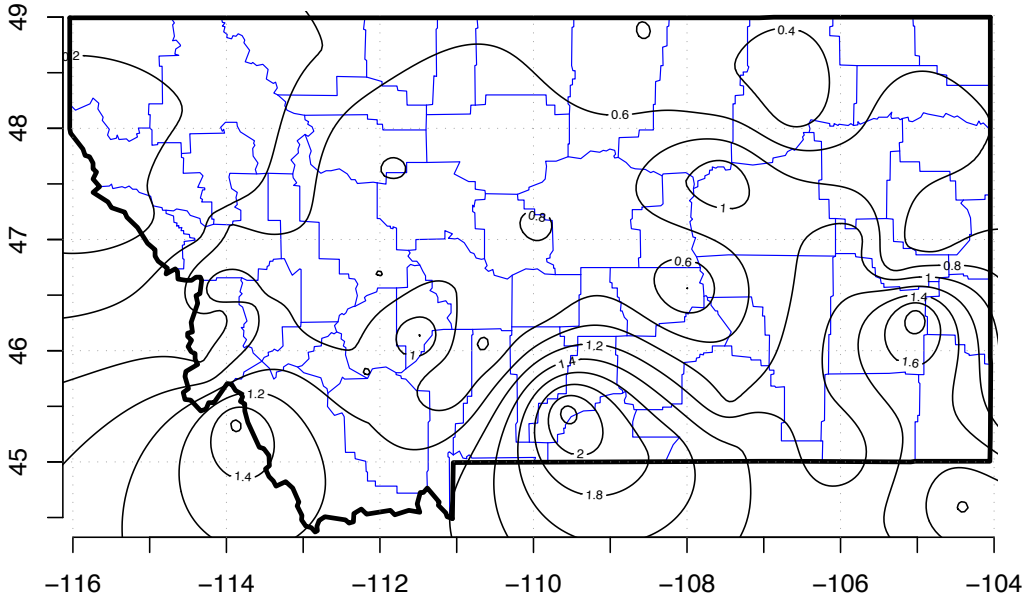
Interpolated ED curve maps.

Please see section 1.7 for instructions on the use of these statewide maps.

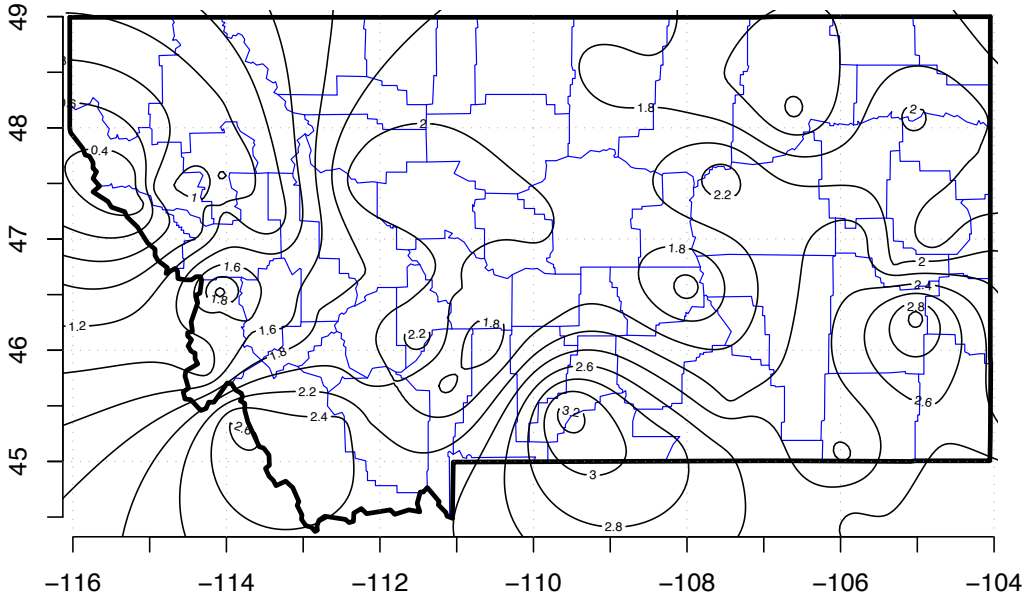
Number of Hours per day exceeding 100 Watts/m^2, Feb



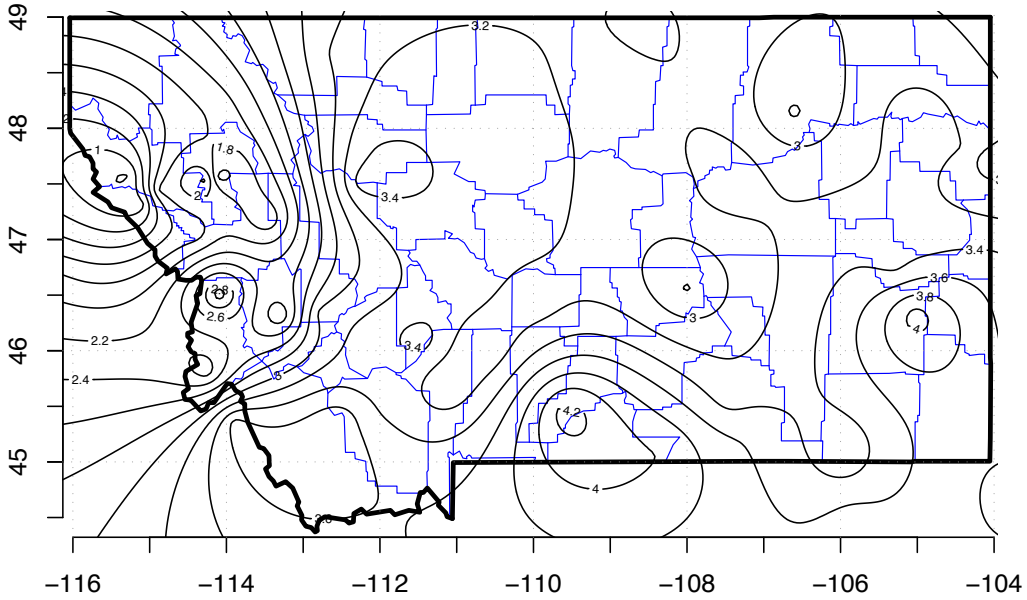
Number of Hours per day exceeding 700 Watts/m^2, Mar



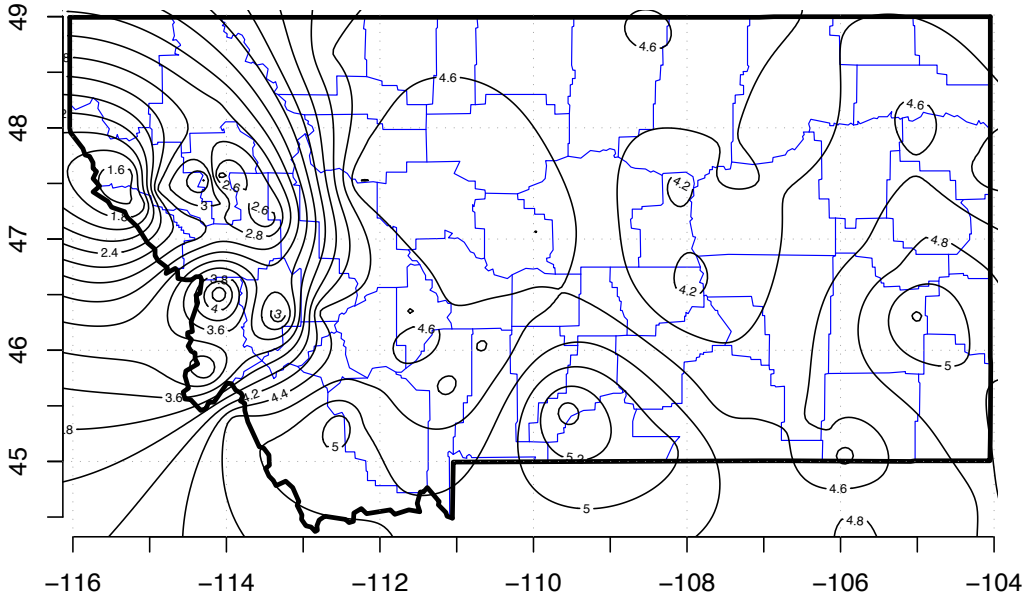
Number of Hours per day exceeding 600 Watts/m^2, Mar



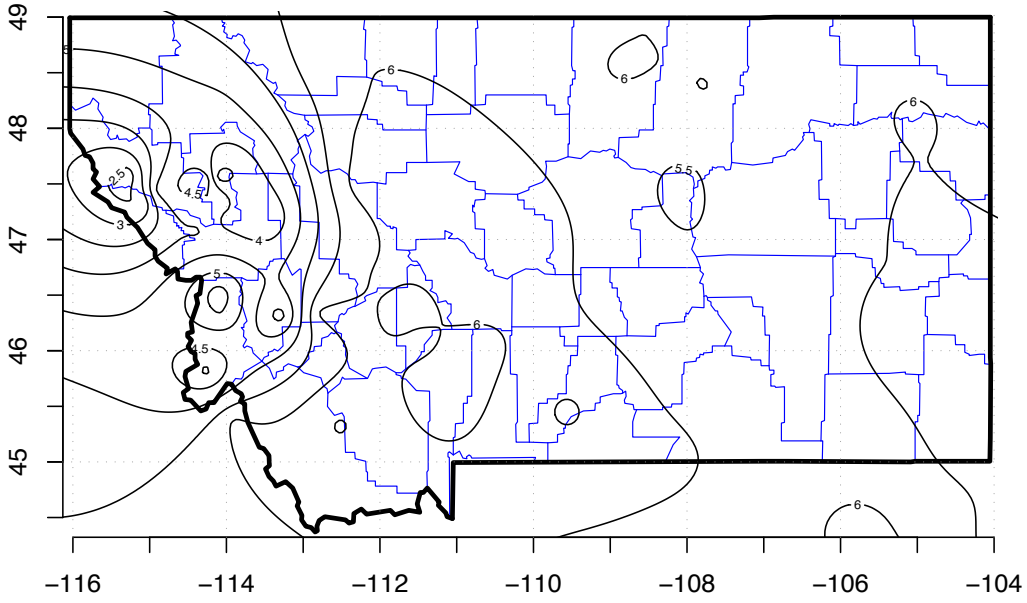
Number of Hours per day exceeding 500 Watts/m^2, Mar



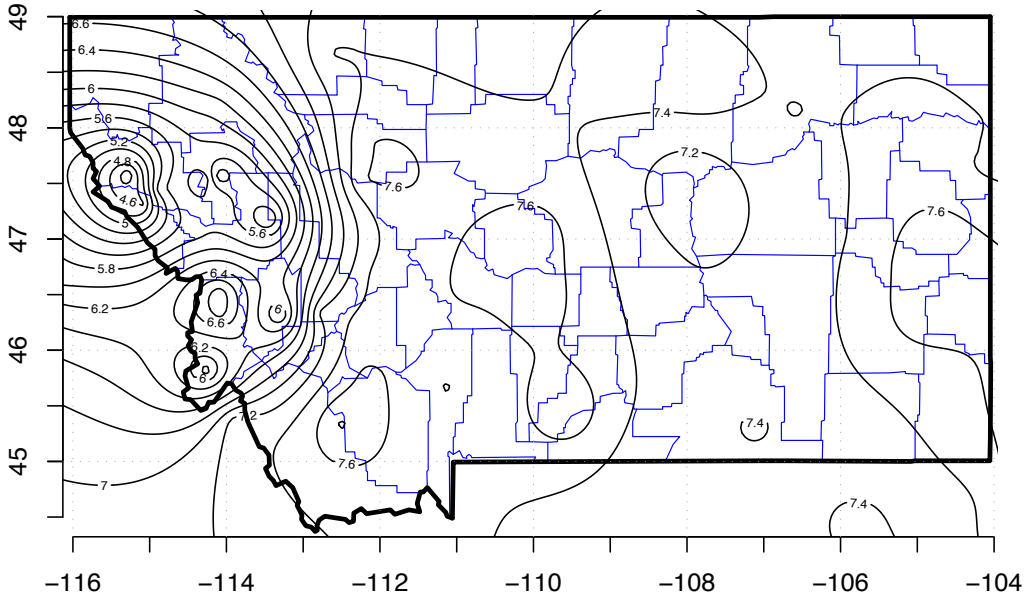
Number of Hours per day exceeding 400 Watts/m^2, Mar



Number of Hours per day exceeding 300 Watts/m^2, Mar



Number of Hours per day exceeding 200 Watts/m^2, Mar



Number of Hours per day exceeding 100 Watts/m^2, Mar