

**University of Alberta**

Modeling carbon-water-vegetation dynamics  
using remote sensing and climate data

by

Nasreen Jahan

A thesis submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Water Resources Engineering

Department of Civil and Environmental Engineering

©Nasreen Jahan  
Spring 2012  
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

## **Abstract**

Carbon and water fluxes are essential components of biospheric processes which directly or indirectly influence climate, surface energy balance, hydrologic processes and hence influence the vegetation productivity, distribution and characteristics. In this research, promising techniques for simulating carbon (Gross primary production) and water fluxes (soil moisture and evapotranspiration) were developed using remotely sensed data to overcome our dependence on meteorological data which are often not available with sufficient accuracy for regional scale climate studies.

The temporal responses of vegetation to climate were assessed using Artificial Neural Network (ANN) and two remotely sensed vegetation indices (VIs), normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). The results demonstrated a promising potential for monitoring the patterns of terrestrial vegetation productivity from VIs and climate variables in a boreal mixedwood forest of western Canada.

Next, the potential of using the newly available, quad-polarized, RADARSAT-2 synthetic Aperture Radar (SAR) data in retrieving near surface soil moisture in the Canadian Prairies was examined. Ten Radarsat-2 images have been acquired over the Paddle River Basin (PRB), Alberta, Canada and 2250 soil samples have been collected from 9 different sites within the same basin on those 10 days. Soil

moisture was retrieved using the regressions, theoretical Integral Equation model (IEM) and two machine learning techniques: ANN and Support Vector Machine. The results show that combined radar and optical satellite data from the Moderate Resolution Imaging Spectroradiometer (MODIS) can be used to retrieve near surface soil moisture accurately.

Finally, algorithms were developed to model vegetation carbon flux (Gross Primary Production, GPP) and evapotranspiration (ET) for the coniferous and deciduous forests using solely remote sensing data from MODIS. The remotely sensed GPP (R-GPP) and ET model (R-ET) were parameterized and validated using the observed data derived from the eddy covariance towers located in north-eastern USA. The proposed models attempt to exclude the use of ground data or climate data as model input by utilizing MODIS ecosystem and radiation budget variables. Considering the trade-off between sophisticated modeling approach and the uncertainties in obtaining regional scale reliable climate data, it can be concluded that these simple models (R-GPP and R-ET) are practical and promising in providing valuable inputs for regional scale hydrological modeling and water resource planning and managements.

## **Acknowledgements**

I would like to thank my supervisor Dr. Thian Yew Gan for his excellent supervision. I am also thankful to my thesis committee members and anonymous reviewers for their valuable suggestions.

I gratefully acknowledge the financial support from the FS Chia PhD Scholarship, Izaak Walton Killam Memorial Scholarship, Dissertation Fellowship and the National Science and Engineering Research Council of Canada research grant.

The dataset used in this study were obtained from different organizations: the Canadian Centre for Remote Sensing (CCRS), the Scientific Operational Applications Research - Education (SOAR-E) initiative, Alberta Agriculture and Food, the AmeriFlux network, Oak Ridge National Laboratory (ORNL) and NASA Land Processes Distributed Active Archive Center (LP DAAC). I am grateful to those organizations.

I am indebted to my parents for their support and encouragements. Special thanks to my husband Zobeyer and to my son Arish for their love, support and sacrifices.

Above all, I am thankful to the Almighty Allah for his grace.

## Table of Contents

<b>Chapter 1 Introduction and research objectives</b>	<b>1</b>
1.1 Introduction	1
1.2 Statement of the problem and opportunities	4
1.2.1 Vegetation-climate relationships	4
1.2.2 Modeling carbon flux: Gross Primary Production	4
1.2.3 Modeling water flux: evapotranspiration and soil moisture	6
1.3 Specific research objectives	8
1.4 Organization of thesis	9
1.5 References	11
<b>Chapter 2 Modelling the vegetation-climate relationship in a boreal mixedwood forest of Alberta using normalized difference and enhanced vegetation indices</b>	<b>15</b>
2.1 Introduction	15
2.2 Research objective and methodology	19
2.3 Study site in central-eastern Alberta	22
2.4 Data sets	23
2.4.1 Climate data	23
2.4.2 Normalized Difference Vegetation Index (NDVI) data	24
2.4.3 Enhanced Vegetation Index (EVI) data	25
2.5 NDVI-climate relationship	26
2.5.1 Linear (LR) and nonlinear regressions (NLR)	26
2.5.2 Artificial Neural Network (ANN)	28
2.5.3 Discussion of results	31
2.6 EVI-climate relationship	35
2.7 Comparison between NDVI-climate and EVI-climate relationships	37
2.8 Summary and conclusions	39
2.9 References	42
<b>Chapter 3 Soil moisture retrieval from microwave (RADARSAT-2) and optical remote sensing (MODIS) data</b>	<b>65</b>
3.1 Introduction	65
3.2 Soil moisture retrieval models	66
3.3 Research objectives	69
3.4 Study site	71

3.5 Data	72
3.5.1 Soil moisture data	72
3.5.2 Radar data	73
3.5.3 Landsat and MODIS data	75
3.5.4 Digital Elevation data	75
3.5.5 Soil properties data	76
3.6 Soil moisture retrieval algorithms	77
3.6.1 Regressions	77
3.6.2 Integral Equation Model (IEM)	79
3.6.3 Artificial Neural Network (ANN)	82
3.6.4 Support Vector Machine (SVM)	86
3.7 Discussions of results	90
3.7.1 Linear regression (LR)	90
3.7.2 Non-linear regression (NLR)	92
3.7.3 Integral Equation Model (IEM)	93
3.7.4 Radial Basis Function Neural Network (RBFNN)	94
3.7.5 Support Vector Machine (SVM)	96
3.7.6 Results at the watershed scale	99
3.7.7 Comparison with Biftu and Gan (1999)	101
3.8 Summary and conclusions	102
3.9 References	106
<b>Chapter 4 Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables</b>	<b>127</b>
4.1 Introduction	127
4.2 Review of Gross Primary Production (GPP) models	129
4.3 Research objectives	132
4.4 Study sites and data sets	133
4.4.1 Study sites	133
4.4.2 Site specific carbon flux and climate data	134
4.4.3 Remotely sensed data	134
4.5 Research methodology	136
4.5.1 R-GPP model predictors	137
4.5.2 Relationships between model predictors and other environmental variables	142
4.6 R-GPP model development and results	144
4.7 Transferability of R-GPP model	146
4.8 Discussions	148
4.9 Summary and conclusions	151
4.10 References	154

<b>Chapter 5 Modeling gross primary production of coniferous forests from MODIS data</b>	<b>176</b>
5.1 Introduction	176
5.2 Study Sites and data Sets	179
5.2.1 Study sites	179
5.2.2 Carbon flux and remotely sensed data	180
5.3 Data and research methodology	181
5.3.1 Selecting model predictors	181
5.3.2 Model development	186
5.4 Results	190
5.4.1 Calibration results	190
5.4.2 Validation results	192
5.5 Discussion of results	195
5.6 Summary and conclusions	199
5.7 References	202
<b>Chapter 6 An algorithm for estimating evapotranspiration for the deciduous and coniferous forests of north-eastern USA using MODIS data</b>	<b>215</b>
6.1 Introduction	215
6.2 Research objectives	219
6.3 Study sites and data sets	219
6.3.1 Study sites	219
6.3.2 ET and remotely sensed (RS) data	222
6.4 R-ET model predictors	224
6.4.1 Land surface temperature (LST)	226
6.4.2 Short wave albedo ( $\alpha_{sw}$ )	227
6.4.3 Leaf area index (LAI)	230
6.4.4 Gross primary production (GPP)	232
6.5 Remotely sensed-ET (R-ET) model	234
6.6 Discussions of results	236
6.6.1 Calibration of R-ET model	236
6.6.2 Validation and transferability of R-ET model	237
6.6.3 Discussions on modeling ET of forest environments	241
6.7 Summary and conclusions	244
6.8 References	246
<b>Chapter 7 Summary, conclusions and recommendations</b>	<b>268</b>
7.1 Summary and conclusions	268
7.2 Recommendations for future work	271

7.3 References	273
<b>Appendix</b>	<b>275</b>

## List of Tables

Table 2.1. Comparison of MODIS sensor characteristics with that of NOAA-AVHRR.	56
Table 2.2. Climate predictors used to model VI-climate relationships.	56
Table 2.3. Results of modelling NDVI-climate relationship using Back Propagation Neural Network (BPNN).	56
Table 2.4. Correlation between EVI and topographic variables (elevation, slope, aspect).	57
Table 3.1. Description of 7 Radarsat2 images acquired over the Paddle River basin at descending orbit.	118
Table 3.2. Different input cases for the linear and nonlinear regressions.	118
Table 3.3. Results from the linear and nonlinear regressions.	119
Table 3.4. Input cases for Artificial Neural Network and Results.	120
Table 3.5. Input cases for Support Vector Machine and Results.	121
Table 4.1. Correlations between model predictors and GPP at the Harvard forest (for 2000 to 2005).	166
Table 4.2. Correlations between model predictors and different environmental variables (for 2000 - 2005) at the Harvard forest.	166

Table 4.3. Correlation ( $R^2$ ) between observed GPP and either R-GPP model predicted GPP or MODIS GPP. All GPP are mean values of 8 day periods. 167

Table 5.1. Location (lat/long in decimal degrees), annual mean temperature (AMT), annual precipitation (AP), stand age and years of data used for the 4 eddy covariance tower sites used in this study. 210

Table 6.1. Location (lat/long in decimal degrees), vegetation type, annual mean temperature (AMT), annual precipitation (AP), stand age and years of data used for the nine eddy covariance tower sites used in this study. 260

## List of Figures

Figure 2.1. (a) Location of the study site (boreal mixedwood forest) in Alberta, Canada, (b) location and soil type of township grids used for gridded climate dataset (1= Orthic Gray Luvisol, 2= Orthic Humic Gleysol, 3= Orthic Regosol, 4= Eluviated Eutric Burnisol), and (c) location of the climate stations used to produce the gridded climate dataset of 1991-2005. 58

Figure 2.2. Precipitation, temperature, potential evapotranspiration (PET), aridity index (AI) and vegetation indices (NDVI and EVI) time series. 59

Figure 2.3. (a) The ANN system, (b) the network and (c) a sigmoid neuron. 60

Figure 2.4. Performance of linear (LR) and nonlinear (NLR-1 and NLR-2) regressions for different single/ multiple predictors (case1 to case 6) in modelling the NDVI-climate relationship during calibration (left) and validation stages (right), in terms of  $R^2$ , RMSE and  $E_f$ . 61

Figure 2.5. Scatter plots of the observed ( $NDVI_{obs}$ ) and model estimated ( $NDVI_{mod}$ ) NDVI using Back Propagation Neural Network (BPNN) for different input cases; (a) to (f) are the calibration results for case 1 to case 6 respectively, while (g) to (l) are the validation results for case 1 to case 6 respectively. 62

Figure 2.6. Performance of nonlinear regressions (NLR-1 and NLR-2) and Back Propagation Neural Network (BPNN) for different single/ multiple predictors (case1 to case 6) in modelling the EVI-climate relationship during calibration (left) and validation stages (right), in terms of  $R^2$ , RMSE and  $E_f$ . 63

Figure 2.7. Comparing the (a) time series and (b) histogram of NDVI and EVI over the boreal mixedwood forest region of Alberta.	64
Figure 2.8. Comparing summary statistics of NDVI-climate and EVI-climate relationships for six input cases (Time period of 2000 to 2002).	64
Figure 3.1. (a) Location of the study area within the province Alberta, Canada. (b) Location of individual study sites within the Paddle river basin.	122
Figure 3.2. Photographs of 3 landcover types used in this study: (a) Herbaceous land, (b) Agricultural land, (c) Pasture land.	122
Figure 3.3. Architecture of a typical radial basis neural network.	123
Figure 3.4. Details of $\varepsilon$ band with slack variables $\xi_i$ and $\xi_i^*$ and data points (black circles).	123
Figure 3.5. Sensitivity of radar backscatter to volumetric soil moisture with respect to the combined data of 2009 and 2010.	124
Figure 3.6. Regression between observed and IEM simulated volumetric soil moisture during (a) the calibration and (b) the validation stage.	124
Figure 3.7. Observed and IEM simulated volumetric soil moisture at different sites during (a) the calibration and (b) the validation stage.	125

Figure 3.8. Observed soil moisture versus ANN simulated soil moisture where radar backscatter, incidence angle, LST, NDVI, soil properties and terrain properties were used as predictors. 125

Figure 3.9. Observed soil moisture versus SVM simulated soil moisture where radar backscatter, incidence angle, LST, NDVI, soil properties and terrain properties were used as predictors. 126

Figure 3.10. Observed soil moisture versus nonlinear regression (NLR, case 4), IEM, RBFNN and SVM simulated soil moisture for the 10 dates of 2009-2011 at the watershed scale. Each bar represents average soil moisture over the 9 study sites for a particular day. 126

Figure 4.1. (a) Nonlinear/ Linear regression and comparison of seasonal dynamics between GPP and (a) GVMI, (b) Albedo<sub>NIR</sub>, and (c) EVI for 2000 to 2005 at the Harvard forest. All points represent 8-day means. 168

Figure 4.2. Scatter plot and polynomial fit between GPP and (a) LST, (b) scaled LST (LST<sub>s</sub>), and (c) Seasonal dynamics of GPP and LST<sub>s</sub>, for 2000 to 2005 at the Harvard forest. All points represent 8-day means. 169

Figure 4.3. Comparison of seasonal dynamics of photosynthetically active radiation (PAR) with seasonal dynamics of GVMI, Albedo<sub>NIR</sub>, EVI and LST<sub>s</sub> for 2000 to 2005 at the Harvard forest. All points represent 8-day means. 170

Figure 4.4. Comparison of seasonal dynamics of air temperature ( $T_{\text{air}}$ ) with seasonal dynamics of GVMI,  $\text{Albedo}_{\text{NIR}}$ , EVI and  $\text{LST}_s$  for 2000 to 2005 at the Harvard forest. All points represent 8-day means. 171

Figure 4.5. Comparison of seasonal dynamics of vapor pressure deficits (VPD) with seasonal dynamics of GVMI,  $\text{Albedo}_{\text{NIR}}$ , EVI and  $\text{LST}_s$  for 2000 to 2005 at the Harvard forest. All points represent 8-day means. 172

Figure 4.6. Scatter plot of R-GPP model simulated gross primary production (GPP) and eddy covariance tower estimated GPP for the (a) calibration (2000-2003) and (b) validation stages (2004-2005) at the Harvard forest site. All points represent 8-day means. 173

Figure 4.7. Annual cycle of eddy covariance tower estimated gross primary production ( $\text{GPP}_{\text{est}}$ ) and, (a) R-GPP model predicted GPP ( $\text{GPP}_{\text{sim}}$ ) and (b) MODIS GPP ( $\text{GPP}_{\text{MODIS}}$ ) product for the Harvard forest site during the calibration (2000-2003) (left panel) and the validation stage (2004-2005) (right panel), respectively. All points represent 8-day means. 174

Figure 4.8. Annual cycle of eddy covariance tower estimated gross primary production ( $\text{GPP}_{\text{est}}$ ) and, (a) R-GPP model predicted GPP ( $\text{GPP}_{\text{sim}}$ ) and (b) MODIS GPP ( $\text{GPP}_{\text{MODIS}}$ ) product for the Morgan Monroe State Forest (MMSF) during 2000-2005 (left panel) and University of Michigan Biological Station Site (UMBS) during 2000-2003 (right panel). All points represent 8-day means. 175

Figure 5.1. Relationship between EC towers estimated GPP and (a) enhanced vegetation index (EVI), (b) Global vegetation moisture index (GVMI), (c) near infrared albedo, (d) land surface temperature (LST) and (e) scaled LST (LSTs) for all the sites. All points represent 8-day mean. 211

Figure 5.2. Annual cycle of eddy covariance tower estimated gross primary production ( $GPP_{obs}$ ), R-GPP-Coni model predicted GPP ( $GPP_{sim}$ ) and MODIS GPP ( $GPP_{MODIS}$ ) product for the Howland forest site during the calibration period (2000-2004). All points represent 8-day means. 212

Figure 5.3. Scatter plot of eddy covariance tower estimated GPP (Actual GPP) and (a) R-GPP-Coni model simulated GPP; and (b) MODIS GPP for the Howland Forest. All points represent 8-day means. 212

Figure 5.4. Annual cycle of eddy covariance tower estimated gross primary production ( $GPP_{obs}$ ), R-GPP-Coni model predicted GPP ( $GPP_{sim}$ ) and MODIS GPP ( $GPP_{MODIS}$ ) product for the (a) Howland forest west tower site, (b) Duke pine forest and (c) North Carolina Loblolly Pine site. All points represent 8-day means. 213

Figure 5.5. Left panel shows the scatter plot of tower estimated GPP (Actual GPP) and R-GPP-Coni model simulated GPP while right panel shows the scatter plot of eddy covariance tower estimated GPP (Actual GPP) and MODIS GPP. 214

Figure 6.1. Location of the study sites 261

Figure 6.2. Relationships between EC tower estimated ET and (a) enhanced vegetation index (EVI), (b) scaled land surface temperature (LSTs), (c) short wave albedo ( $\alpha_{sw}$ ), (d) leaf area index (LAI), (e) EC tower estimated actual GPP and (f) R-GPP model simulated GPP in the deciduous sites. All data points represent 8-day mean values. 262

Figure 6.3. Relationships between EC towers estimated ET and (a) enhanced vegetation index (EVI), (b) scaled land surface temperature (LSTs), (c) short wave albedo ( $\alpha_{sw}$ ), (d) leaf area index (LAI), (e) EC tower estimated actual GPP, and (f) R-GPP-Coni model simulated GPP for the coniferous sites. All data points represent 8-day mean values. 263

Figure 6.4. Annual cycle of eddy covariance tower estimated evapotranspiration (ET) (actual) and R-ET model predicted ET (simulated) for the deciduous sites. All data points represent 8-day mean values. 264

Figure 6.5. Scatter plot of R-ET model simulated evapotranspiration (ET) and eddy covariance tower estimated ET for the deciduous sites. All data points represent 8-day mean values. 265

Figure 6.6. Annual cycle of eddy covariance tower estimated evapotranspiration (ET) (actual) and R-ET model predicted ET (simulated) for the coniferous sites. All data points represent 8-day mean values. 266

Figure 6.7. Scatter plot of R-ET model simulated evapotranspiration (ET) and eddy covariance tower estimated ET (actual) for the coniferous sites. All data points represent 8-day mean values.

# Chapter 1

## Introduction and research objectives

---

### 1.1 Introduction

Carbon and water fluxes are essential components of biospheric processes which directly or indirectly influence climate, surface energy balance, watershed hydrology, etc., and hence influence the vegetation productivity, distribution and characteristics (Weiss et al., 2004). Vegetation in turn affects those fluxes as well as climate through functioning as a sink for CO<sub>2</sub> (Woodwell et al., 1998) and increasing watershed evapotranspiration (Swank and Douglas, 1974). Therefore, understanding the feedbacks between vegetation and ecosystem fluxes (carbon-water-energy) and modeling the spatial and temporal variations of these fluxes are important in predicting the ecosystem change as well as for the prediction of vegetation productivity under different climatic scenarios.

To quantify the spatial and temporal variation of carbon, water and energy fluxes at ecosystem level, many flux stations have so far been set up throughout the world (Fisher et al., 2008). The FLUXNET network (and its sub-network of USA sites, called AmeriFlux) is a global network of micrometeorological tower sites.

This network was initiated with several primary objectives such as to provide infrastructure for assembling, archiving, and distributing carbon, water, and energy flux measurements over different ecosystem types (e.g., tropical, temperate, and boreal), along with meteorological and other supplementary biological, ecological and disturbance data (e.g., Leaf Area Index (LAI) and phenology, soil carbon, root biomass, and volumetric soil moisture) to the scientific community (Baldocchi et al., 2001). At present more than 500 tower sites from about 30 regional networks are providing flux data spanning diurnal to interannual time scales (<http://daac.ornl.gov/FLUXNET/fluxnet.shtml>). All these flux observations, however, are not adequate enough to capture the spatial and temporal variability of global fluxes (Qinxue et al., 2005).

Beside ground measurements, ecosystem fluxes and vegetation dynamics can be quantified by ecosystem models that are able to describe accurately the physical and biological processes in vegetation cover and soil, as well as physical processes in the atmosphere. Ecosystem models, developed to predict ecosystem processes, range from simple correlation methods to complex ecophysiological models that couple vegetation-atmosphere exchange of energy, mass and momentum. However, most of the existing ecosystem models rely on meteorological data which are not available with sufficient accuracy for regional scale modeling of ecosystem processes (Heinsch et al., 2006)

Given the inadequateness of observation network and problems associated with the complex ecosystem models, it may be highly beneficial to develop robust yet simple models that are capable of simulating carbon, water and vegetation reliably at regional to global scale. Remotely sensing can facilitate these simple models by providing valuable information of land surface and vegetation and can eliminate the necessity of meteorological data partially/ fully. Satellite remote sensing is also a potentially valuable tool for upscaling the carbon and water flux measurements (from large-scale tower networks like FLUXNET) from stand scale to regional or global scale. Vegetation indices (VIs), computed from optical satellites, were first developed in the 1970s to monitor terrestrial vegetation and have been highly successful in assessing vegetation condition, foliage, cover, phenology, and processes such as evapotranspiration (ET) and photosynthesis at regular temporal and spatial resolution (Glenn et al., 2008) . Besides VIs, other radiation and ecosystem variables from the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA's Terra and Aqua satellites are contributing to global climate studies. Several studies have integrated flux data with remote sensing data to quantify carbon and water flux over large areas (Zhang et al., 2007; Yang et al., 2006).

## **1.2 Statement of the problem and opportunities**

### **1.2.1 Vegetation-climate relationships**

Over the past two decades, many studies have been conducted to investigate the relationships between vegetation activity and climatic variables. Many of these studies have used vegetation index (VI) such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), derived from multispectral remote sensing imagery, as a measure of vegetation productivity (Los et al., 2001). However, most of these studies have only examined the relationships between VI and precipitation and temperature; and not much attention has been paid to other climatic variables like evapotranspiration, aridity index, etc. (Deng et al., 2007). But plant growth depends on both energy and water fluxes which will be better represented by precipitation, air temperature, potential evapotranspiration (PET), etc., together rather than by any of these climatic variables individually. Moreover, a comprehensive VI-climate relationship has not yet been obtained due to the short-time scale (e.g. less than 10 years) of the analysis and due to complex relationship between them (Los et al. 2001). Therefore, this research attempts to establish a meaningful relationship between VI and few important climate variables, namely precipitation, temperature, evapotranspiration and aridity index.

### **1.2.2 Modeling carbon flux: Gross Primary Production**

Rising levels of CO<sub>2</sub>, and other greenhouse gases, are of concern to scientists and policy makers because they trap infrared radiation that is emitted by the earth's

surface. Potential consequences of elevated CO<sub>2</sub> concentrations include a warming of the earth's surface (Hansen et al. 1998), melting of polar icecaps and a rising sea level, and an alteration of plant and ecosystem physiological functioning (Amthor 1995; Norby et al. 1999). Forests, occupying about 30% of the earth's land surface, and accounting for about 80–90% of all plant carbon, may act as a valuable sink of CO<sub>2</sub>. However, exact contribution of forests to the global carbon and water budget is still uncertain.

Gross primary production (GPP) represents the gross uptake of CO<sub>2</sub> by vegetation that is used for photosynthesis and is another widely used variable for monitoring vegetation productivity. Estimating GPP of terrestrial ecosystems has been a major challenge in the past decades (Heinsch et al, 2006). A number of satellite-based GPP estimation models (Xiao et al., 2004) showed the potential of remotely sensed (RS) data in GPP modeling. However, most of the currently available GPP estimation models (e.g., Vegetation Photosynthesis model, EC-LUE, C-fix, MODIS GPP algorithm etc.) need meteorological inputs (Sims et al., 2006) in addition to RS data. But for regional scale modeling, measured meteorological data is not always available. Therefore some models rely on coarse resolution (e.g., 1° latitude by 1.25° longitude) climate data obtained from different climate models (e.g., NASA's GEOS-4 climate model) or interpolation of ground station data. However, both approaches may contain error which in turn may introduce significant inaccuracy in GPP estimation (Heinsch et al., 2006). Therefore it will

be very worthwhile if a reliable GPP estimation model can be developed avoiding the meteorological inputs.

### **1.2.3 Modeling water flux: evapotranspiration and soil moisture**

The accurate estimation of water flux (i.e., soil moisture and evapotranspiration (ET)) at regional or larger scale is difficult because it varies highly over space and time (Guyot et al., 2009), and ground measurements are often time consuming and expensive (Biftu and Gan, 1999). Only few effective measurement techniques (such as the eddy correlation technique) are available to monitor ET. Nevertheless, these methods often failed to balance the surface energy components (Amiro, 2009). Moreover, these point measurements sometimes may not be extended to surrounding areas accurately, even for apparently homogeneous vegetation cover (Guyot et al., 2009). Therefore, at watershed scale of 100 to several thousand km<sup>2</sup>, ET are commonly estimated using the standard ET methods such as the Penman-Monteith, the Thornthwaite and the Priestley and Taylor even though the assumptions involved are not always representative of the real situations. Sometimes these ET equations are modified with different theoretical basis to generate global patterns of ET (Nishida et al., 2003). However, many of the traditional energy balance models of ET are too complex and require explicit characterization of numerous physical parameters (e.g. precipitation, soil moisture, soil infiltration capacity, soil texture, etc.) which are sometimes difficult to obtain for regional to global scale modeling (Fisher et al., 2008). Therefore,

development of a robust algorithm to estimate regional evapotranspiration is still a significant challenge (Mu et al., 2007).

Soil moisture is an important variable for many natural resource applications such as hydrological modeling, stream flow forecasting, crop growth modeling, flood forecasting and drought modeling. Over the last three decades many studies have investigated the retrieval of soil moisture values from satellite and airborne synthetic aperture radar (SAR) data by establishing empirical or theoretical relationship between soil moisture and microwave backscattering (Baghdadi and Zribi, 2006). However, most of the existing radar data based soil moisture estimation models (theoretical, empirical and semi-empirical) have used only single polarization data [either HH (horizontal transmit and horizontal receive of microwave) or VV (vertical transmit and vertical receive of microwave)] from the first generation of SAR (such as ERS-1/2, RADARSAT-1).

In December 2007, the Canadian Space Agency (CSA) launched its RADARSAT-2 satellite, Canada's next-generation SAR satellite, as a follow-up to RADARSAT-1. RADARSAT-2 is the first commercial spaceborne SAR satellite to offer fully polarimetric datasets i.e., 4 different polarization (HH, VV, HV (horizontal transmit and vertical receive) and VH (vertical transmit and horizontal receive)) channels are acquired per image. The quad polarization mode of RADARSAT-2 enables it to characterize the scattering field in a comprehensive manner. Although it is expected that polarimetric data set will provide increased

sensitivity to soil, still it is not clear what additional information will be retrievable from these fully polarimetric datasets (McNairn and Brisco, 2004). Only few research studies have so far used fully polarimetric data (HH, VV, HV and VH) available from the Shuttle Imaging Radar (SIR-C) (operation period 1994), Jet Propulsion Laboratory (JPL) AIRSAR (operation period 1988-2004), etc. to retrieve soil moisture (Sokol, 2004). Some of these studies have shown significant improvement in soil moisture retrieval by using multi-polarized data while some other studies showed that use of more than one polarization does not contribute significantly to soil moisture retrieval (Baghdadi and Zribi, 2006, Sokol, 2004). Therefore the potential of using multi-polarization data is not conclusive yet and the relationships between the newly available fully polarimetric datasets of RADARSAT-2 and soil moisture still need to be established (Sokol et al., 2004).

### **1.3 Specific research objectives**

The specific objectives of this research are:

1. To model vegetation-climate relationship and comparison of different vegetation indices in monitoring vegetation activity in response to climate. For this purpose, NDVI and EVI from the NOAA-Advanced Very High Resolution Radiometer (NOAA-AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) will be used to monitor the vegetation activity of a boreal mixed wood forest of central-eastern Alberta, Canada.

2. To assess the feasibility of soil moisture retrieval combining microwave data from the Radarsat-2 satellite and optical remote sensing data from the MODIS using regression, theoretical Integral Equation Model, Artificial Neural Network and Support Vector Machine.
3. To model gross primary production (GPP) of coniferous and deciduous forests using only four remotely sensed variables: two radiation budget variables (land surface temperature (LST) and albedo) and two ecosystem variables (global vegetation moisture index (GVMI) and EVI) from the MODIS.
4. To develop algorithm for estimating evapotranspiration of coniferous and deciduous forests of north-eastern USA using solely remotely sensed variables.

## **1.4 Organization of thesis**

The thesis consists of seven chapters. Chapter 1 provides an overview on carbon, water and vegetation dynamics modeling, limitations of the existing techniques and scope to contribute in modeling carbon-water-vegetation fluxes. Chapter 2 contains the detailed analysis of vegetation climate relationships using NDVI and EVI for the boreal mixed wood forest of central-eastern Alberta, Canada. . In Chapter 3, the potential of multipolarized Radarsat-2 data in soil moisture retrieval has been assessed using regression, theoretical Integral Equation Model, Artificial Neural Network and Support Vector Machine. Chapter 4 and 5 describe in detail the development and validation of a gross primary production model (called remotely sensed GPP model, R-GPP) for modeling carbon flux of deciduous and coniferous ecosystems, respectively, using only satellite data.

Chapter 6 contains the parameterization and validation of a solely remote sensing algorithm for modeling evapotranspiration of deciduous and coniferous ecosystem. Finally, the overall conclusions and recommendations for future work are presented in Chapter 7.

## 1.5 References

- Amthor, J. S. 1995. Terrestrial higher-plant response to increasing atmospheric [CO<sub>2</sub>] in relation to the global carbon cycle. *Global Change Biology*, 1, 243–274.
- Amiro, B. 2009. Measuring Boreal Forest Evapotranspiration using the Energy Balance Residual. *Journal of Hydrology*, 366, no. 1-4, 112-118.
- Baghdadi, N. and Zribi, M. 2006. Evaluation of radar backscatter models IEM, Oh and Dubois using experimental observations. *International Journal of Remote Sensing*, Vol., 27, No. 18, pp. 3831-3852.
- Baldocchi, D.D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, Ch., Davis, K., Fuentes, J., Goldstein et al. 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux densities. *Bull. Am. Meteorol. Soc.* 82, 2415–2435.
- Biftu, G. F., and Gan, T.Y. 1999. Retrieving soil moisture from Radarsat SAR data. *Water Resources Research*, 35(5), p. 1569-1579.
- Deng, F., Su, G. and Liu, C. 2007. Seasonal variation of MODIS vegetation indexes and their statistical relationship with climate over the subtropic evergreen forest of Zhejiang, China. *IEEE Geoscience and Remote Sensing Letters*, 4 (2), pp 236–240.
- Fisher, J. B., Tu, K. P. and Baldocchi D. D. 2008. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data,

validated at 16 fluxnet sites. *Remote Sensing of Environment*, 112, 901-919.

Glenn, E., Huete, A., Nagler, P.L. and Nelson, S.G. 2008. Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape, *Sensors*, vol. 8, pp. 2136-2160.

Guyot, A., Cohard, J-M., Anquetin, S., Galle, S., Lloyd, C.R. 2009. Combined analysis of energy and water balances to estimate latent heat flux of a sudanian small catchment. *Journal of Hydrology*, doi: 10.1016/j.jhydrol.2008.12.027.

Hanson, J. E., Sato, M., Lacias, A., Ruedy, R., Tegen, I. and Matthews, E. 1998. Climate forcings in the industrial era. *Proceedings of National Academy of Science USA*. 95, 12 753–12 758.

Heinsch, F.A., Zhao, M., Running, S.W., Kimball, J.S., Nemani, R.R., et al. 2006. Evaluation of remote sensing based terrestrial production from MODIS using AmeriFlux eddy tower flux network observations. *IEEE Transactions on Geosciences and Remote*, 44, 1908–1925.

Los, S.O., Collatz, G.J., Bounoua, L., Sellers, P.J. and Tucker, C.J. 2001. Global interannual variations in sea surface temperature and land surface vegetation, air temperature, and precipitation. *Journal of Climate*, 14, pp. 1535–1549.

- McNairn, M. and Brisco, B. 2004. The application of C band polarimetric SAR for agriculture: A review. *Canadian Journal of Remote Sensing*, Vol. 30, No. 3, pp. 525-542.
- Mu, Q., Heinsch, F. A., Zhao, M. and Running, S. W. 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, 111, 519-536.
- Nishida, K., Nemani, R. R., Running, S. W., & Glassy, J. M. 2003. An operational remote sensing algorithm of land surface evaporation. *Journal of Geophysical Research-Atmospheres*, 108, 4270.
- Norby, R. J., Wullschleger, S. D., Gunderson, C. A., Johnson, D. W. and Ceulemans, R. 1999. Tree responses to rising CO<sub>2</sub> in field experiments: Implication for the future forest. *Plant, Cell Environment*, 22, 683-714.
- Qinxue, W., Masataka, W. and Zhu, O. 2005. Simulation of water and carbon fluxes using BIOME-BGC model over crops in China. *Agricultural and Forest Meteorology*, 131, 209-224.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, et al. 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment*, 112, pp. 1633-1644.
- Sokol, J., McNairn, H. and Pultz, T. J. 2004. Case studies demonstrating the hydrological applications of C-band multipolarized and polarimetric SAR. *Canadian Journal of Remote Sensing*, Vol. 30, No. 3, 470-483.

- Swank, W. T. and Douglas, J.E. 1974. Streamflow greatly reduced by converting deciduous hardwood stands to pine. *Science*, 185, pp. 857–859
- Weiss, J. L., Gutzler, D. S., Coonrod, J. E. A. and Dahm, C. N. 2004. Seasonal and inter-annual relationships between vegetation and climate in central New Mexico, USA. *Journal of Arid Environment*, 57, 507–534.
- Woodwell, G. M., Mackenzie, F. T., Houghton, R. A., Apps, M. Gorham, E. and Davidson, E. 1998. Biotic feedbacks in the warming of the Earth, *Climate Change*, 40, 495– 518.
- Xiao, X.M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B., Ojima, D. 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, 91, 256–270.

# Chapter 2

## Modelling the vegetation-climate relationship in a boreal mixedwood forest of Alberta using normalized difference and enhanced vegetation indices<sup>\*</sup>

---

### 2.1 Introduction

Vegetation indices (VI) are widely used parameters to monitor seasonal, interannual, and long term variation of terrestrial vegetation productivity. For the last two decades or so, Normalized Difference Vegetation Index (NDVI), has been one of the most popularly used VIs. It is a robust indicator of vegetation productivity (Tucker and Sellers, 1986) and is defined as the reflectance difference between the visible red ( $R$ ) and near-infrared (NIR) bands, over their sum (Equation 1.1).

$$\text{NDVI} = \frac{(\text{NIR}) - R}{(\text{NIR}) + R} \quad (2.1)$$

It indirectly measures the amount of chlorophyll content through the reflected visible red radiation and the spongy mesophyll content through the reflected NIR radiation by the vegetation canopy. In general, if there is much more reflected

---

<sup>\*</sup> A version of this paper has been published. Jahan, N. and Gan, T. Y. 2011. Modeling vegetation-climate relationship in a central mixed wood forest region of Alberta using normalized difference and enhanced vegetation indices. *International Journal of Remote Sensing*, 32, 313-335.

radiation in NIR wavelengths than in visible wavelengths, then the vegetation is likely to be dense (Tucker and Sellers, 1986) and the NDVI value of that pixel will be relatively high. NDVI can be used to predict crop yield (Zhao et al., 2007), to estimate landscape patterns of green biomass (Raynolds et al., 2008; Westra and Wulf, 2007; Prince, 1991) and to monitor droughts, weather impacts and other events important for agriculture and ecology (Bajgiran et al., 2008; Bhuiyan et al., 2006; Kogan, 1990). NDVI also correlates well with other vegetation indices such as leaf area index (LAI), total dry matter accumulation and annual gross primary production (Maselli and Chiesi, 2006; Tucker et. al., 1985; Tucker et. al., 1983). Soudani et al. (2008) studied the onset of green-up in deciduous broadleaf forest using NDVI while Piao et al. (2006) studied the growth of temperate grassland using this index. Researchers have also used NDVI to monitor the temporal response of vegetation to climatic fluctuations in USA (Tan, 2007; Di et al., 1994), in Africa (Martiny et al., 2006; Justice et al., 1986; Townshend and Justice 1986; Malo and Nicholson, 1990), in India (Srivastava et al., 1997), in China (Deng et al., 2007) and at a global scale (Schultz and Halpert, 1993). Many researchers found that the spatial distribution of the vegetation cover is strongly related to mean climatic conditions and therefore NDVI can be used for the study of interannual climate variability (Anyamba and Eastman, 1996; Richard and Pocard, 1998; Chamaille-James et al., 2006).

In recent years, a new generation of advanced optical sensors, the Moderate Resolution Imaging Spectroradiometer (MODIS), on board TERRA (since 2000)

and AQUA satellites (since 2002), came into operation. MODIS has 36 spectral bands specifically designed for land applications with spatial resolution ranging from 250 m to 1 km (Justice et al., 1998). The Enhanced Vegetation Index (EVI) product of MODIS is designed to provide consistent, spatial and temporal comparisons of global vegetation conditions in finer details than the NDVI product of NOAA-AVHRR (Justice et al., 1998). The broad NIR band (0.725 to 1.1  $\mu\text{m}$ ) of the NOAA-AVHRR (Table 2.1) includes several atmospheric water absorption band (such as at 0.82 and 0.905  $\mu\text{m}$ ) and thus the spectral radiance reaching the sensor can be attenuated by the atmospheric water vapor (Huete et al., 1997). On the other hand the NIR band of MODIS (0.84 to 0.87  $\mu\text{m}$ ), shown in Table 2.1, is finer and avoids this water absorbing region. Another disadvantage of NDVI is that it could be saturated easily, which means that a moderately dense vegetation and a very dense rainforests with large amounts of chlorophyll may have the same NDVI value (Huete et al., 2002). NDVI also exhibits very high sensitivity to canopy background (e.g. weathered geologic substrates, leaf litter, water, snow etc) variations and NDVI degradation could be strong with higher canopy background brightness (Huete et al., 2002). EVI, developed to overcome the aforementioned limitations of NDVI, produces vegetation signal with improved sensitivity in high biomass regions and with improved vegetation monitoring through the reduction of soil and atmospheric influences (Waring et al., 2006). EVI is defined as

$$\text{EVI} = G \times \frac{(\text{NIR}) - R}{(\text{NIR}) + C_1 \times R - C_2 \times B + L} \quad (2.2)$$

where NIR,  $R$  and  $B$  are atmospherically corrected surface reflectance in the near-infrared, red and blue bands respectively,  $G$  is the gain factor,  $L$  is the canopy background adjustment factor that addresses nonlinear, differential NIR and red radiant transfer through a canopy; and  $C_1$  and  $C_2$  are the coefficients of the aerosol resistance term, which uses the blue band to correct the aerosol influences in the red band. As Rayleigh scattering is much more pronounced in shorter wavelengths (e.g., blue) than in the longer wavelength (e.g., red), it is possible to estimate and correct the amount of scattering due to the propagation of radiation fluxes by comparing spectral radiances in the blue and red band. The coefficients adopted in the EVI algorithm are,  $L=1$ ,  $C_1=6$ ,  $C_2 = 7.5$ , and  $G$  (gain factor) = 2.5 (Huete et al., 2002). Various studies have utilized MODIS EVI product in determining land use patterns (Vanacker et al., 2005), vegetation phenology (Zhang et al., 2003; Zhang et al., 2004), crop (rice) production (Xiao et al., 2005), crop classification (Wardlow et al., 2007; Nagler et al., 2007) and for studying vegetation-climate relationship (Deng et al., 2007).

Besides climatic factors (such as radiative fluxes and water supply), vegetation cover, being a very sensitive natural system, also depend on factors such as geology, soil characteristics, underground water regime, land use, topography, etc. (Montandon and Small, 2008; Wang et al., 2007; Yang et al., 1997). Conditions of vegetation cover and its dynamics can be explained by the joint effect of these factors. Farrar et al. (1994) found significant differences in the NDVI-rainfall relationships for different soil types. They also found that in

Botswana, vegetation growth is enhanced by exogenous soil water originating as runoff from the surrounding higher terrain. Yang et al. (1997) found that NDVI-climate relations were very strong in areas with low available water capacity and high soil permeability. Narumalani et al. (1999) studied the variation of NDVI with different terrain features in the Island Lake, Nebraska, USA, while Wang et al. (2007) studied the influence of root zone soil moisture on NDVI.

## **2.2 Research objective and methodology**

The primary objective of this study was to determine the possible relationships between VI and climate over a mixedwood forest of central-eastern Alberta, where stagnation moraines with large soil storages dominate its geology (Natural Regions Committee, 2006). This mixedwood forest is a part of the Canadian boreal forest which covers 39% of Canada's land surface and plays an important role in global carbon balance and in the timber production of Canada. Some recent studies of climate models showed that global warming will significantly affect the boreal forest located at higher latitude and small changes in the climate could have large effect on their primary production (Neigh et al., 2006). However only few studies have so far been conducted to study the vegetation-climate relationship over the boreal forest of high latitude (Neigh et al., 2006; De Beurs and Henebry, 2005) and so an improved model on the vegetation-climate relationships of the Canadian boreal forest should be useful for its various stakeholders.

In recent years, although many studies have been conducted to investigate the relationships between vegetation activity and climate factors, however, most of these studies have only examined the relationships between VI and precipitation or VI and temperature (Prasad et al., 2008; Chamaille-James et al., 2006; Prasad et al., 2007); and very little attention has been paid to other climatic variables like evapotranspiration, aridity index, etc. (Suzuki et al., 2006). But plant growth depends on both energy and water fluxes which will be better represented by several key climate variables (precipitation (PPT), air temperature (T), potential evapotranspiration (PET), etc.) together rather than by any of these climate variables individually. Therefore in this study, in addition to single climatic variable, the combined effect of multiple climate variables on VIs was examined.

First, NDVI-climate relationships were attempted using regressions (linear and non-linear) and a more complicated, Back Propagation Neural Network (BPNN). Seasonal, three month moving averaged PPT, T, PET and the Aridity index (AI) over the summer months (May to October) were used as predictors (either individually or combined together) while three-month moving average of the NDVI was used as the predictand. In this study, we used climate variables (predictors) without any time lag because preliminary data analysis showed that for our study site and for seasonal time frame the vegetation index (VI)-climate correlation is generally higher when no lag is considered. Data of 1991 to 1996 have been used for calibrating these models while those of 1997 to 2000 were used for validating the calibrated models. Model performances were assessed in

terms of three classic statistical measures, namely, the coefficient of determination ( $R^2$ ), the root mean square error (RMSE) and the Nash Sutcliffe coefficient ( $E_f$ ) (Nash and Sutcliffe, 1970), given in Equation 2.3.

$$E_f = 1 - \frac{\sum_{i=1}^n (\text{VI}_o - \text{VI}_m)_i^2}{\sum_{i=1}^n (\text{VI}_o - \overline{\text{VI}_o})_i^2} \quad (2.3)$$

Where  $n$  is the sample size,  $\text{VI}_m$  is the model estimated VI,  $\text{VI}_o$  is the observed VI and  $\overline{\text{VI}_o}$  is the mean of observed VI (NDVI or EVI).  $E_f$  is a more stringent statistical measure than the popular coefficient of determination  $R^2$  (Gan et al., 2004). An  $E_f$  of 1 corresponds to a perfect fit while  $E_f$  approaching zero indicates that the model is not better than a model which uses the mean of the observed data as predicted values, and  $E_f$  less than zero indicates that the observed mean is a better predictor than the model prediction.

Similarly, we developed and assessed the corresponding EVI-climate relationship. Given that MODIS data became available since 2000, we chose the EVI data of 2003 to 2005 for calibration while that of 2000 to 2002 for validating the calibrated relationship. After obtaining NDVI-climate and EVI-climate relationships, a comparison between them provides us some insight about which VI better monitors the temporal responses of boreal mixedwood forest of central Alberta to climate.

### **2.3 Study site in central-eastern Alberta**

The study site, located in the central-eastern part of Alberta (Figure 2.1), is part of the western Canadian boreal mixedwood forest and occupies an area of 3525 km<sup>2</sup>. Within Alberta, this ecoregion is the largest natural subregion and occupies more than 40% of the province (Natural Regions Committee, 2006). The site is mainly occupied by deciduous stands like aspen, aspen-white spruce forests with some conifer stands like white spruce and jack pine stands on upland terrain. As the site is occupied by both coniferous and deciduous forest, it is known as 'mixedwood forest' (Natural Regions Committee, 2006). The site is mainly covered by Gray Luvisolic soils with small area composed of Brunisols and Organic soils (Figure 2.1(b)). The climate is characterized by short and warm summer, and long and cold winter. July is generally the month with maximum precipitation and temperature. A detailed description of the soil, topography, vegetation, geology and climate of the study site is given in Natural Regions Committee, 2006 and landuse is given in [http://www.agr.gc.ca/pfra/gis/lcv\\_e.htm](http://www.agr.gc.ca/pfra/gis/lcv_e.htm).

Only a small portion of the area is used for cultivating hay crops and native pasture and domestic livestock grazing, and so such activities should not affect the VI-climate relationship which we attempted to establish in this study for the boreal mixedwood forest. Furthermore, this study site is chosen because its vegetation is subjected to minimal human interference such as forest harvesting, irrigated/ fertilized agriculture, etc ([www.forestwatch.ca](http://www.forestwatch.ca)). So we expect a better

possibility of establishing a meaningful relationship between vegetation and climate for this study sites than other sites of Alberta.

## **2.4 Data sets**

### **2.4.1 Climate data**

The gridded daily precipitation, maximum and minimum air temperature data (1991-2005) for 6900 townships (each township is a 10 km × 10 km square parcel of land for administering legal land title) of Alberta were obtained from Alberta Agriculture, Food and Rural Development (AAFRD) Department (Figure 2.2). AAFRD constructed this gridded dataset from all available and reliable station data using the inverse distance weighting interpolation method for temperature and a hybrid interpolation method for precipitation to preserve the localized and highly spatially variable characteristics of precipitation (Shen et al., 2001). For each grid point, data from the eight nearest stations, located within the specified radius of influence (60 km for precipitation and 200 km for temperature since precipitation is more spatially variable than temperature) of that point, were used in the interpolation (see, Shen et al., 2001 for details). The reliability of the gridded data set was checked by cross validation with observed station data and found that the method used here gives higher accuracy with less error than some other methods like the simple nearest neighbor assignment and the inverse-distance-square weighting for this area (Shen et al., 2001; Griffith, 2002). In this study, the interpolated gridded data is used because they are of sufficient resolution (approximately 10 km × 10 km) which will match better with the VI

data of raster format in establishing the VI-climate-relationship. This database has also been used for other studies, such as Shen et al. (2005) used this dataset to explore the spatial variations of the agroclimatic resources and the potential crop-growing area in Alberta, and Coen et al. (2004) used this data to model wind erosion risk in Alberta soil.

The monthly precipitation and temperature data for each of the 39 towngrids located within the study area were obtained by averaging the daily precipitation and daily mean temperature (average of maximum and minimum) data of that grid, respectively. The potential evapotranspiration (PET) was calculated by the Thornthwaite method (Palmer and Havens, 1958). Then from PPT and PET data, the aridity index, AI ( $=PPT/PET$ ) which is an indicator of the degree of dryness of climate at a given location (Edgell, 2006), was calculated.

#### **2.4.2 Normalized Difference Vegetation Index (NDVI) data**

The NDVI data for 1991 to 2002, derived from the images of NOAA/Advanced Very High Resolution Radiometer (AVHRR), was provided by the Canadian Centre for Remote Sensing (CCRS). The data set, 10-day clear sky composites covering 5700 km × 4800 km area of North America with 1 km spatial resolution, was produced using the Earth Observation Data Manager (EODM) processing system (Latifovic et al. 2005). An algorithm, dubbed SPARC has been used to separate AVHRR pixels into the clear-sky and cloudy scenes (Khlopenkov and Trishchenko, 2006). The data was found to have high georeferencing accuracy

and good radiometric consistency for all sensors from AVHRR NOAA-6 to AVHRR NOAA-17. Other improvements include atmospheric correction and compositing. All necessary corrections were made at CCRS (Latifovic et al., 2005). We combined 3 10-day NDVI images of each month to obtain the monthly NDVI image.

#### **2.4.3 Enhanced Vegetation Index (EVI) data**

The monthly MODIS EVI data for 2000 to 2005, with 1 km spatial resolution, was collected from the NASA Land Processes Distributed Active Archive Center, LP DAAC (<http://lpdaac.usgs.gov/dataproducts.asp>). EVI is computed from MODIS daily surface reflectances which are radiometrically calibrated, cloud-filtered, atmospherically corrected (corrected for molecular scattering, ozone absorption and aerosols), spatially and temporally gridded, and adjusted for view angle influences (see Huete et al. (2002) for details). The quality of EVI data based on quality assurance (QA) flags was checked prior to the analysis of EVI images. MODIS Reprojection Tool, MRT (<http://edcdaac.usgs.gov/datatools.asp>) was used to sub-sample the images to an area covering the study site and subsequently from these new images EVI values at desired towngrid points were extracted.

## 2.5 NDVI-climate relationship

### 2.5.1 Linear (LR) and nonlinear regressions (NLR)

The NDVI-climate relationship is first modeled using simple linear and nonlinear regression models to see whether NDVI is more related to climate in a linear or nonlinear manner. In the case of nonlinear regression (NLR), only two NLR equations (equations 4 and 5) are herein reported although few other NLR equations have been tried but they did not produce appreciable improvement over LR.

$$NDVI = a_0 + a_1(X_1^{b_1}) + a_2(X_2^{b_2}) + \dots + a_n(X_n^{b_n}) \quad (2.4)$$

$$NDVI = K \exp\{a_0 + a_1(X_1^{b_1}) + a_2(X_2^{b_2}) + \dots + a_n(X_n^{b_n})\} \quad (2.5)$$

Where  $a_0, a_1, a_2, \dots, b_1, b_2, \dots, K$ , etc. are the coefficients of the NLR equations,  $X_1, X_2$ , etc. are the predictors (climate variables). The performance of each regression algorithm has been evaluated using six different combinations of predictors (PPT, T, PET, and AI individually and combined) as shown in Table 2.2. For example LR-case1 denotes a linear regression model driven by case -1 predictor (PPT).

When the predictors of a model are strongly interrelated, the regression coefficients will be unstable and suffer from high standard errors (Belsley, 1991). In addition, because of this problem which is known as collinearity or multicollinearity, it will be difficult to estimate the unique effect of an individual predictor on the overall model performance (Roso et al., 2005). A diagnostic tool for multicollinearity is the condition index (CI), which if lies between 15 and 30,

indicates weak multicollinearity among predictors while CI larger than 30 suggests the presence of significant multicollinearity (Roso et al., 2005; Myers and Well, 2003; Belsley et al., 1980). In our study, for all multi-predictor cases, we found the CI to be consistently less than 15, which demonstrates very weak to no multicollinearity.

The calibration of the NLR was done using the Generalized Reduced Gradient (GRG2) optimization technique that solves nonlinear optimization problems (Lasdon, 1978; Lasdon, 1980). The widely used GRG2 optimization algorithm is relatively fast and fairly simple (Spaulding, 1998). This algorithm searches for parameter values that optimize an objective function such as Equation 2.6, which is a simple least square (SLS) that minimizes the sum of square of the difference between the observed and model estimated NDVI. GRG2 has the skill to start from feasible or infeasible starting points and generates a sequence of improved feasible points (Lasdon, 1978) and it has been successfully applied in many fields such as efficient network design, macroeconomic planning, reservoir operation, etc. (Spaulding, 1998).

$$\text{Objective function, } f = \frac{\sqrt{\text{SLS}}}{n} = \frac{\sqrt{\sum_{i=1}^n (\text{NDVI}_{\text{obs}} - \text{NDVI}_{\text{mod}})_i^2}}{n} \quad (2.6)$$

Where  $n$  is the sample size,  $\text{NDVI}_{\text{mod}}$  is the model estimated NDVI and  $\text{NDVI}_{\text{obs}}$  is the observed NDVI.

### **2.5.2 Artificial Neural Network (ANN)**

An ANN is also used to model the NDVI-climate relationship because it can almost fit any type of nonlinear input-output relationships and is considered as a robust/error tolerant method of estimation (Morgan and Bourland, 1990). Because of these established advantages, there are numerous applications of ANN in water resources, such as modelling the rainfall–runoff relationship (Minns and Hall, 1996), river runoff forecasting (Tokar and Johnson, 1999; Cigizoglu, 2003), regional drought analysis (Shin and Salas, 2002), retrieving sea-water turbidity from Landsat TM data (Gan et al., 2004), etc.

In this study, we used the Back Propagation Neural Network (BPNN) that consists of one input layer, one hidden layer and one output layer (Figure 2.3) because BPNN is one of the most popular and proven neural networks (Hagan et al., 1996). The input ( $I$ ) layer introduces the input to the network and the hidden layer processes the input by applying non-linear activation function (also known as transfer function) (Figure 2.3(c)) which allows non-linearity in the network (Zealand, 1997). The output layer represents the response of the network. The layers consist of neurons and connections. Each connection is associated with a weight and each neuron is associated with a bias. Value of each neuron ( $N_j$ ) is computed by using these weights and biases (Figure 2.3(c)). ANN with biases and a single hidden sigmoid layer (a layer which uses sigmoid function as transfer function) is by far the most commonly used network topology, probably because of its capability of approximating any function with a finite number of discontinuities (Cybenko, 1989; Hornik *et al.*, 1989). In training the BPNN, the

predictand is compared to the observed counterpart via the SLS objective function (Equation 2.6). The weights and biases in the network are then iteratively adjusted to optimize the objective function.

The three most common transfer functions in neural network application are the tan-sigmoid, the log-sigmoid, and the linear functions. In this study, we used the tan-sigmoid transfer function in the hidden layer and linear transfer function in the output layer because this combination of transfer functions was found to give better result than other combinations in preliminary investigations. An ANN can basically be trained using an incremental or a batch training approach. In this study the 'batch training' mode has been employed where the weights and biases are only updated after all of the inputs are processed and they have been updated using the Levenberg-Marquardt (LM) training algorithm which is faster and provides better results than many other training algorithms (Hagan and Menhaj, 1994; Tan and van Cauwenberghe, 1999; Islam, 2005). In LM, at first, the derivatives of the network errors with respect to the weights and biases are computed. Then in each iteration, a Jacobian matrix ( $\mathbf{J}$ ) is constructed combining all of the error gradients. The change of weights and biases is calculated using the following expression:

$$\Delta \mathbf{x} = [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e} \quad (2.7)$$

Where  $\mu$  is a scalar,  $\mathbf{I}$  is an identity matrix and  $\mathbf{e}$  is a vector of errors for all neurons in the output layer. The weight and biases for the next iteration ( $t+1$ ) is computed by

$$[\mathbf{x}]^{t+1} = [\mathbf{x}^t + \Delta\mathbf{x}^t] \quad (2.8)$$

To avoid the problem of overfitting an ANN, we applied the Bayesian regularization technique (Foresee and Hagan, 1997) which minimizes a linear combination of squared errors and weights. In this approach the traditional objective function,  $f$  (Equation 2.6) is modified to  $f_{\text{mod}}$  (Equation 2.9) for improving its generalization.

$$f_{\text{mod}} = \beta E_d + \alpha E_w \quad (2.9)$$

Where  $E_d$  is the sum of the squared errors between model estimated and observed VI (NDVI or EVI),  $E_w$  is the sum of the squares of the network weights,  $\alpha$  and  $\beta$  are parameters which are to be optimized in a Bayesian framework.  $\alpha$  and  $\beta$  are optimized such that the network will have smaller weights and biases with smoother responses but less likely to overfit (Xin et al., 1999).

An alternative to regularization is the early stopping (Nelles, 2001). In this technique the available data is divided into three sets: training, validation and test sets. The training set is used for computing the gradient and updating the network weights and biases. During the training process, the model performance on both the training and the validation sets are monitored. The error on the validation set normally decreases during the initial stage of training, followed by an increase as the network begins to overfit the data. Training is stopped when the error on the validation data reaches its minimum (Prechelt, 1998). Then generalization of the model is estimated by evaluating the performance of the trained model on the test

set. The advantage of the Bayesian regularization over early stopping technique is especially noticeable when the size of the data set is small because the later technique needs three different data sets as mentioned before (Skabar, 2005).

### **2.5.3 Discussion of results**

If a set of calibrated parameters, obtained by minimizing the objective function, is realistic or causative then the calibrated model should also perform well in the validation stage (for data not used to calibrate model parameters) and so the drop in RMSE,  $R^2$  and  $E_f$  from the calibration to the validation stage should be fairly modest. Poor validation results may indicate that calibrated parameters were obtained through curve fitting with little physical basis, inappropriate model, non-representative calibration data, overfitting of the model, or possibly erroneous validation data.

#### **2.5.3.1 Results from linear and nonlinear regressions**

The performances of the linear and nonlinear regressions driven by six input cases are presented in Figure 2.4. From Figure 2.4 it is obvious that precipitation alone (case 1) is a poor predictor of NDVI because the net amount of moisture that is available to vegetation growth depends not only on precipitation but also on soil characteristics, temperature, PET loss and others. Case 2 and case 3 based on temperature and PET respectively showed better performance than Case 1. Case 2 (Temperature) was the best among the first three cases based on single predictor partly because the seasonal cycle of temperature and NDVI are highly in phase in

this area. In addition, air temperature is an indication of the amount of solar energy present that is needed for photosynthesis which leads to vegetation growth. Schultz and Halpert (1993) also found higher correlation between temperature and NDVI than that between precipitation and NDVI in cold regions. Given that in this study the PET is estimated from temperature, high correlation is also found between NDVI and PET.

Case 4 to case 6 based on multiple predictors showed better performances (in terms of  $E_f$  and  $R^2$ ) than case 1 to case 3 which was expected because plant growth depends on both energy and water fluxes which will be better represented by precipitation, air temperature, AI and PET together rather than by any of these climate variable individually. Figure 2.4 shows that the calibrated LR and NLR models performed well at the validation stage especially for the multiple predictor cases. The drop in  $R^2$  and  $E_f$  was minimal between the calibration and validation stages for most cases. On the basis of the performance of regression models, it is clear that there is a strong NDVI-climate relationship in the boreal mixedwood forest of central-eastern Alberta and if we can project the climate variables realistically at seasonal lead time, it will be possible to use such models for terrestrial vegetation monitoring.

The results obtained from the LR and two NLR equations show that NLR-1 (Equation 2.4) is generally better than the NLR-2 (Equation 2.5) and LR (Figure 2.4). These comparisons likely imply that the NDVI-climate relationship may be

expressed more accurately by certain nonlinear than linear regressions, which makes sense as we expect vegetation and climate to be non-linearly related. However the relationship may still not be adequately represented using NLR models, partly because of its relatively simplistic structure. Since ANN has been applied to map highly complex processes in a wide range of field of studies, the BPNN (an ANN) was also used herein to model the NDVI-climate relationship.

### **2.5.3.2 Results from Back Propagation Neural Network (BPNN)**

A BPNN can be designed in almost an infinite number of ways by changing the number of layers, number of neurons in each layer, etc. Therefore it is practically difficult to determine what configuration of a BPNN model represents the global optimum (Chan and Nguyen, 2003). Even though some researchers (e.g., Mwale and Gan, 2005; Hsieh and Tang, 1998) found an ANN calibrated by the genetic algorithm (GA) to have better capability of searching the global optimal solution of complex systems, it was not attempted herein because GA could suffer from some serious problems such as: convergence is not guaranteed, convergence may be slow, and a local, instead of global, optimum may be found (Sun et al., 2006; Brackin and Colton, 2002; Chan and Nguyen, 2003).

The same six test cases used in LR and NLR regressions were repeated using BPNN. In this study trials were carried out, for searching an optimum structure, by changing number of neurons in the hidden layer and by using different type of transfer functions (TF) in the hidden and output layers (see Appendix, Algorithm

A1 for the Matlab code). From trial and errors, BPNN consists of one input layer, one hidden layer (consists of 10 to 20 neurons) with a tan-sigmoid TF and one output layer with a linear TF was found to be better than networks with other combination of TF for all input cases. Only the best networks based on different input cases are reported herein.

For case 1, even the BPNN model did not do well, with a  $R^2$  of 0.56 and 0.44 for the calibration and the validation stages, respectively. As expected, a complex model such as BPNN cannot rectify the problem of deficient input data (e.g., precipitation as the only predictor), and therefore its result is similar to that of the simpler regression models. Similar to LR and NLR, case 2 to case 6 results (Table 2.3) and the scatter plots (Figure 2.5) of model estimated versus observed NDVI for BPNN show that better results are obtained in both the calibration and validation stages when multiple input predictors are used instead of a single predictor. Further it has been found that using BPNN,  $R^2$  is the highest for case-6 in both of the calibration ( $R^2 = 0.95$ ) and the validation stage ( $R^2 = 0.92$ ) and drops in  $R^2$  and  $E_f$  were modest from the calibration to the validation stages for all cases.

For all cases, except case 1 as described earlier, BPNN performed much better than LR and NLR-2 (Equation 2.5), in terms of  $R^2$ , RMSE and  $E_f$ , in both stages (Table 2.3 and Figure 2.4). In comparison with NLR-1 (Equation 2.4), it is observed that during calibration stage BPNN performed much better for case 2 and 3 and moderately better for case 4 to 6. During validation, BPNN showed

marginally better performance over Nonlinear1 in terms of RMSE and  $R^2$  but noticeably better in terms of  $E_f$  for case 3 to case 6 (Table 2.3 and Figure 2.4). For case 2, performances of BPNN and NLR-1 are almost the same in the validation stage. In summary, even though the actual NDVI-climate relationship may be more complex than the structure of NLR and BPNN, generally these algorithms can adequately model the relationship if ample and appropriate climate variables are used.

## **2.6 EVI-climate relationship**

Although EVI is developed to overcome some limitations of NDVI, EVI also has some limitations and one of them is its higher sensitivity to topographic conditions than the NDVI, especially in rough terrain (Matsushita et al., 2007). Matsushita et al. (2007) found that the canopy background adjustment factor used for EVI calculation makes it much more sensitive to the direct effect of topography while NDVI can eliminate or weaken the topographic effects because of its band ratio format. The effect of topography on EVI varies with the pixel resolution and with the increase of pixel size this effect may decrease or even disappear (Matsushita et al., 2007). As the pixel size of the EVI images used in this study is 1 km, topography may not have significant effect on the EVI data. However, to verify our assumption, we calculated the slope and aspect from the digital elevation model (DEM) data of the study site and then analyzed the correlations between EVI and those topographic variables (elevation, slope, aspect).

The DEM data of the study site was collected through the GeoBase portal ([www.geobase.ca](http://www.geobase.ca)) of the Canadian Council on Geomatics (CCOG). The resolution of the data set is 0.75 arc seconds. Description of DEM data quality is available in <http://www.geobase.ca/geobase/en/data/cded/description.html>.

Table 2.4 shows the results of the correlation analyses between EVI and topographic variables and indicates that EVI is virtually independent of the terrain features for our case. From the visual comparison of the EVI images of different months with individual raster images of those variables, we also found that EVI images do not show any similar spatial pattern to elevation/slope/aspect images. Therefore, possibly because of the coarse spatial resolution of EVI images (1 km), in this study we can ignore the effect of topography. Some other studies, conducted in areas of rough terrain, have also used coarse resolution EVI data without considering any effect of topography (Takata et al., 2007; Deng et al., 2007).

Since the results obtained for the NDVI-climate relationship shows that vegetation growth is probably more non-linearly than linearly related to climate, we attempted to model EVI-climate relationship using the NLR (equations 4 and 5) and BPNN only. Results from the NLR and BPNN applied to EVI data and driven by the same six cases of climate input are shown in Figure 2.6. As before, a generally better EVI-climate relationship was obtained when multiple climate predictors were used instead of single climate predictors. Among the first three

cases, case 2 and case 3 gave better result than case1 where precipitation was used as predictor.

In terms of  $R^2$ , there was no clear winner between the two NLR models (Figure 2.6). During the calibration stage in some cases NLR-1 performed better while in some cases NLR-2 performed better. Moreover during the validation stage, the two NLR performed almost equally for the first three cases and for the other cases NLR-1 was better than NLR-2. However, BPNN was found to be marginally better than both the NLR during calibration for all cases. During validation, the performance of BPNN was either equivalent (case 1 to case 4) or slightly better (case 5 to case 6) than the NLR. In terms of RMSE, the BPNN model was generally better than the NLR during calibration. On the other hand during validation, its performance was better for case 5 and as good as the NLR for all other cases. In terms of  $E_f$ , it was found that BPNN is better than both NLR in almost all cases of calibration and validation stages. Given that  $E_f$  is a more stringent goodness-of-fit statistics than  $R^2$  (Figure 2.6), it seems that BPNN is a more consistent model than the NLR in modeling the EVI-climate relationship.

## **2.7 Comparison between NDVI-climate and EVI-climate relationships**

The sensitivities of NDVI and EVI to climate are compared with respect to the data of 2000 to 2002. By comparing the seasonal (3-month moving average) time series of both indices (Figure 2.7(a)), it seems that the temporal pattern of EVI is

synchronous with that of NDVI, but EVI values (0.15 to 0.504) are consistently lower than the NDVI values (0.13 to 0.822) in almost all cases. However, Figure 2.7(b) shows that EVI values spread over a broader distribution than NDVI, which tends to be 'spiky' and most of the NDVI values lie between 0.45 and 0.85 with maximum number observations at 0.77.

Given that for both NDVI and EVI, BPNN is a better choice than LR and NLR to model the VI-climate relationship, the effectiveness of these two VIs for vegetation monitoring in response to climate is assessed only in terms of BPNN. Figure 2.8 shows that the EVI-climate relationship has better summary statistics over that of the NDVI-climate relationship in the boreal mixedwood forest of Alberta. This may be partly attributed to the saturation problem of NDVI at higher biomass, as evident in its 'spiky' histogram (Figure 2.7(b)), which causes NDVI to be less sensitive to temporal vegetation variations which in turn may cause it to be less responsive to climate variability. Many other researchers (Deng et al., 2007; Huete et al. 2002) also reported this saturation problem of NDVI in higher biomass. In contrast, EVI values are more evenly spread out on both sides of the distribution (Figure 2.7(b)) with a peak at 0.51, implying that EVI possesses higher dynamic range than NDVI and is less likely to reach saturation as NDVI. Therefore EVI can be regarded as a more sensitive and robust vegetation index than NDVI in depicting the variation in vegetation pattern.

The other major factors behind the differences in the sensitivities of the two VIs can be the atmospheric interference (e.g., aerosol scattering) and canopy background brightness which strongly affect NDVI, making it less sensitive to vegetation changes (Jiang et al., 2008) than EVI. In EVI the aerosol and canopy background effects are minimized by using the blue band and soil adjustment factor (Huete et al., 2002; Jiang et al., 2008), respectively, as discussed earlier. The apparent higher fidelity of the MODIS-EVI may also be partly caused by the increased sensitivity of the MODIS-red band to vegetation (Gitelson & Kaufman, 1998).

## **2.8 Summary and conclusions**

Linear (LR) and nonlinear regression (NLR) models and an artificial neural network called Back Propagation Neural Network (BPNN) have been tested to model the NDVI-climate and EVI-climate relationships for a boreal mixedwood forest in central-eastern Alberta using six different sets of input predictors that consist of precipitation, temperature, potential evapotranspiration (PET) and aridity index either individually or in certain combinations. The results of the study are summarized below:

- (i) For both vegetation indices (VI), results obtained for cases 4 to 6 (multiple predictors) are generally better than that for cases 1 to 3 (single predictor) which probably indicate that a realistic VI-climate relationship are more likely to be obtained from using multiple climate

predictors than using a single climate predictor. This is expected as plant growth depends on both energy and water fluxes which will be better represented by precipitation, air temperature, PET together rather than by any of these climate variables individually;

- (ii) Since the seasonal cycle of NDVI and temperature are closely in phase with each other in this part of Alberta and temperature is closely related to solar energy which plays a significant role in the photosynthesis of vegetation leading to plant growth, temperature (Case 2) showed the highest potential to model VI among the three single predictors. Moreover, as PET is calculated from temperature, high correlation is also achieved for this predictor;
- (iii) In general NLR performed better than LR implying that the NDVI-climate relationship is nonlinear. Between the two NLR tested herein, NLR-1 was found to be better than NLR-2 in modelling NDVI while both NLR models performed comparably in modelling the EVI-climate relationship;
- (iv) BPNN generally modeled the VI better than LR and NLR partly because as an ANN it could better model the structure of a complex system. Among the three models and six cases of input, BPNN achieved the highest skill in both the calibration and the validation stages using the multiple climate predictors; and
- (v) Due to the saturation problem over the dense biomass forests, NDVI derived from NOAA-AVHRR satellite data was found to be less

dynamic and less sensitive to climate than EVI derived from MODIS satellite data. The higher fidelity of the MODIS-EVI data may also be partly attributed to the higher sensitivity of MODIS red band to vegetation than that of NOAA-AVHRR.

Even though this is a site-specific study conducted at the boreal mixedwood forest of Alberta Province, Canada, we believe that the models developed in this study, possibly with minor adjustments, can be applied to the Canadian boreal forest at regional scale, especially to the boreal mixedwood forest of western Canada. For eastern Canada, more modifications to the models will be necessary because of fairly significant differences in the climate between western and eastern Canada. Its potential application can also be expanded to phenology study, yield and drought monitoring. Further studies will be necessary to determine the relationships between vegetation index and climate for different climate regions as well as for different vegetation types, and to explore the possible linkage between a regional and a local regression or neural network model. If the ANN models involve predictors such as precipitation, which are highly variable in space, upscaling the model from local to regional scale would be more complicated than predictors consist mainly of say, temperature. Intuitively, a local model may be extended to other sites with some kind of adjustment factors that reflect the climate relationships between sites, such ideas can be tested in future.

## 2.9 References

- Anyamba, A., and Eastman, J.R. 1996. Interannual variability of NDVI over Africa and its relation to El Nino/Southern Oscillation. *International Journal of Remote Sensing*, 17, pp. 2533- 2548.
- Bajgiran, P. R., Darvishsefat, A. A., Khalili, A. and Makhdoum, M. F. 2008. Using AVHRR-based vegetation indices for drought monitoring in the Northwest of Iran. *Journal of Arid Environments*, 72, pp. 186-196.
- Belsley, D. A., Kuh, E. and Welsch, R. E. 1980. *Regression Diagnostics* (John Wiley and Sons, Inc., New York, NY).
- Bhuiyan, C., Singh, R. P. and Kogan, F.N. 2006. Monitoring drought dynamics in the Arvalli region (India) using different indices based on ground and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 8, pp. 289-302.
- Biftu, G.F. and Gan, T.Y. 2001. Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data. *Journal of Hydrology*, 244, pp. 137–156.
- Boegh, E., Thorsen, M. , Buttsb, M.B. , Hansen, S. , Christiansen, J.S., Abrahamsen, P. , Hasager, C.B. , Jensen, N.O. , Van Der Keur, P. , Refsgaard,J.C. , Schelde, K. , Soegaard, H. and Thomsen, A. 2004. Incorporating remote sensing data in physically based distributed agro-hydrological modeling. *Journal of Hydrology*, 287, pp. 279-299.

- Brackin, P. and Colton, J.S. 2002. Using genetic algorithm to set target values for engineering characteristics in the house of quality. *Transactions of the ASME*, 2, pp. 106-114.
- Chan, C.W. and Nguyen, H.H. 2003. *Intelligent systems: Technology and application*, Leondes C. T. (ed), 5, (CRC press).
- Cigizoglu, H.K. 2003. Estimation, forecasting and extrapolation of acceleration data by artificial neural networks. *Hydrological Science Journal*, 48(3), pp. 349–361.
- Chamaille-James, S., Fritz, H. and Murindagomo, F. 2006. Spatial patterns of the NDVI-rainfall relationship at the seasonal and interannual time scale in an African savanna. *International Journal of Remote Sensing*, 27, pp. 5185-5200.
- Cihlar, J., Tcherednichenko, I., Latifovic, R., Li, Z. and Chen, J. 2001. Impact of variable atmospheric water vapor content on AVHRR data corrections over land. *IEEE Transactions on Geoscience and Remote Sensing*, 39, pp. 173–180.
- Cigizoglu, H.K. and Alp, M. 2006. Generalized regression neural network in modeling river sediment yield. *Advances in Engineering Software*, 37, pp. 63-68.
- Coen, G.M., Tatrko, J., Martin, T.C., Cannon, K.R., Goddard, T.W. and Sweetland, N.J. 2004. A method for using WEPS to map wind erosion risk of Alberta soils. *Environmental Modelling and Software*, 19, pp. 185-190.

- Cybenko, G. 1989. Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2, pp. 303-314.
- De Beurs, K. M. and Henebry, G. M. 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. *Global Change Biology*, 11, pp. 779-790.
- Deng, F., Su, G. and Liu, C. 2007. Seasonal variation of MODIS vegetation indexes and their statistical relationship with climate over the subtropic evergreen forest of Zhejiang, China. *IEEE Geoscience and Remote Sensing Letters*, 4, pp. 236–240.
- Di, L., Rundquist, D.C., and Han, L. 1994. Modeling relationships between NDVI and precipitation during vegetation growth cycles. *International Journal of Remote Sensing*, 15, pp. 2121–2136.
- Edgell, H.S. 2006. *Arabian Deserts Nature Origin and Evolution* (Springer Netherlands).
- Farrar, T.J., Nicholson, S.E. and Lare, A.R. 1994. The influence of soil type on the relationships between NDVI, rainfall and soil moisture in semiarid Botswana. II. NDVI response to soil moisture. *Remote Sensing of Environment*, 50, pp.121-133.
- Foresee, F.D., and Hagan, M.T. 1997. Gauss-Newton approximation to Bayesian regularization. In *Proceedings of the 1997 International Joint Conference on Neural Networks*, pp. 1930-1935.

- Gan, T. Y., Kalinga, O. A., Ohgushi, K. and Araki, H. 2004. Retrieving seawater turbidity from ladsat TM data by regressions and an artificial network. *International Journal of Remote Sensing*, 25, pp. 453-4615.
- Gitelson, A. and Kaufman, Y. 1998. MODIS NDVI optimization to fit the AVHRR data series—spectral consideration. *Remote Sensing of Environment*, 66, 343–350.
- Griffith, D. P., 2002, Processing of daily agroclimatic data. M.S. Thesis, University of Alberta.
- Hagan, M.T. and Menhaj, M.B. 1994. Training feed forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5, pp. 989-993.
- Hagan, M. T., Demuth, H. B. and Beale, M. 1996. *Neural network design* (PWS Publishing Company, Boston).
- Hornik, K., Stinchcombe, M. and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, pp. 359-366.
- Hsieh, W.W. and Tang, B. 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, 79, pp. 1855–1868.
- Huete, A.R., Liu, H.Q., Batchily, K. and Van Leeuwen, W.J.D. 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59, pp 440–451.

- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, pp. 195-213.
- Islam, R. 2005. Improved quantitative estimation of rainfall by radar. M.Sc Thesis, University of Manitoba, Canada.
- Justice, C.O., Holben, B.N. and Gwynne, M.D. 1986. Monitoring East African vegetation using AVHRR data. *International Journal of Remote Sensing*, 7, pp. 1453–1474.
- Jiang, Z., Huete, A.R., Didan, K. and Miura, T. 2008. Development of a two band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, pp 3833 -3845.
- Justice, D.H., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, R., Myneni, R., Knjazihhin, Y., Running, S., Nemani, R., Vermote, E., Townshend, J., Defries, R., Roy, D., Wan, Z., Huete, A., Van Leeuwen, R., Wolfe, R., Giglio, L., Muller, J.P., Lewis, P. and Barnsley, M. 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36, pp. 1228–1249.
- Khlopenkov, K.V. and Trishchenko, A.P. 2006. SPARC: New cloud, clear-sky, snow/ice and cloud shadow detection scheme for historical AVHRR 1-km observations over Canada. *Journal of Atmospheric and Oceanic Technology*, pp. 58.

- Kogan, F.N. 1990. Remote sensing of weather impacts on vegetation on non-homogeneous area. *International Journal of Remote Sensing*, 11, pp. 1405– 1419.
- Lasdon, L.S., Waren, A.D., Jain, A. and Ratner, M. 1978. Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming. *ACM Transactions on Mathematical Software*, 4, pp. 34-50.
- Lasdon, L S., and Waren, A.D. 1980. Survey of nonlinear programming applications. *Operations Research*, 28, No. 5.
- Latifovic, R., Trishchenko, P.A., Chen, J., Park, M.W., Khlopenkov, V.K., Fernandes, A.R., Pouliot, D., Ungureanu, C., Luo, Y., Wang, S., Davidson, A., and Cihlar, J. 2005. Generating historical AVHRR 1-km baseline satellite data records over Canada suitable for climate change studies. *Canadian Journal of Remote Sensing*, 31, pp. 324-346.
- Malo, A.R., and Nicholson S.E. 1990. A study of rainfall and vegetation dynamics in the African Sahel using normalized difference vegetation index. *Journal of Arid Environments*, 19, pp. 1–24.
- Martiny, N., Camberlin, P., Richard, Y. and Philippon, N. 2006. Compared regimes of NDVI and rainfall in semi-arid regions of Africa. *International Journal of Remote Sensing*, 27, pp. 5201-5223.
- Maselli, F. and Chiesi, M., 2006, Integration of multisource NDVI data for the estimation of mediterranean forest productivity. *International Journal of Remote Sensing*, 27, pp. 55-72.

- Matsushita, B. Yang, W., Chen, J., Onda, Y. and Qiu, G., 2007, Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-Density Cypress Forest. *Sensors*, 7, pp. 2636-2651.
- Minns, A.W. and Hall, M.J. 1996. Artificial neural networks as rainfall runoff models. *Hydrological Science Journal*, 41, pp. 399–417.
- Montandon, L. M. and Small, E. E. 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. *Remote Sensing of Environment*, 112, pp. 1835-1845.
- Morgan, N. and Bourland, H. 1990. Generalization and parameter estimation in feedforward nets: Some experiments. In *Advances in Neural Information Processing Systems 2*, Touretzky, D. S. (ed.), pp. 630–637 (Morgan Kaufman Publisher).
- Mwale, D. and Gan, T.Y. 2005. Wavelet analysis of variability, teleconnectivity and predictability of the September-November east African rainfall. *Journal of Applied Meteorology*, 44, pp. 256-269.
- Myers, J. L. and Well, A.D. 2003. *Research design and statistical analysis* (Lawrence Erlbaum Associates USA).
- Nagler, P. L., Glenn, E. P., Emmerich, W., Scott, R. L., Huxman, T. E. and Huete, A. R. 2007. Relationship between evapotranspiration and precipitation pulses in a semiarid rangeland estimated by moisture flux towers and MODIS vegetation indices. *Journal of Arid Environments*, 70, pp. 443-462.

- Narumalani, S, Rundquist, D.C., Meader, J. and Payton, S. 1999. Characterizing patterns and trends of wetland vegetation using the normalized difference vegetation index (NDVI). In Proceedings of 20<sup>th</sup> Asian Conference on Remote Sensing, Hongkong, 1999.
- Nash, J. E and Sutcliffe, J. V. 1970. River flow forecasting through conceptual models Part 1- A discussion of principles. *Journal of Hydrology*, 10(3), pp. 282-290.
- Natural Regions Committee. 2006. Natural Regions and Subregions of Alberta. Compiled by Downing, D.J. and Pettapiece, W.W. Government of Alberta. Pub. No. T/852.
- Neigh, C. S. R., Tucker, C. J. and Townshend, J. R. G. 2007. Synchronous NDVI and surface temperature trends in Newfoundland, 1982 to 2003. *International Journal of Remote Sensing*, 28, no. 11, pp. 2581-2598.
- Nelles, O. 2001. *Nonlinear System Identification* (Springer Publication).
- Palmer, W. C. and Havens, A. V. 1958. A graphical technique for determining evapotranspiration by the Thornthwaite method. *Monthly Weather Review*, 86, pp. 123-128.
- Piao, S., Mohammat, A., Fang, J., Cai, Q. and Feng, J. 2006. NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China. *Global Environmental Change*, Vol. 16, pp. 340-348.
- Prasad, V. K., Badarinath, K. V.S. and Eaturu, A. 2008. Effects of precipitation, temperature and topographic parameters on evergreen vegetation greenery

- in the Western Ghats, India. *International Journal of climatology*, DOI: 10.1002/joc.1662.
- Parasad, A. K., Sarkar, S., Singh, R. P. and Kafatos, M. 2007. Inter-annual variability of vegetation cover and rainfall over India. *Advances in Space Research*, 39, pp. 79-87.
- Prechelt, L. 1998. Early Stopping -- but when? In *Neural Networks: Tricks of the trade*, pp 55-69, *Lecture Notes in Computer Science 1524*, Springer Verlag, Heidelberg.
- Prince, S.D. 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grassland 1981–1988. *International Journal of Remote Sensing*, 12, pp. 1301–1312.
- Raynolds, M. K., Komiso, J. C., Walker, D. A. and Verbyla, D. 2008. Relationship between satellite-derived land surface temperatures, arctic vegetation types, and NDVI. *Remote Sensing of Environment*, 112, pp. 1884-1894.
- Richard, Y. and Pocard, I. 1998. A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing*, 19, pp. 2907 – 2920.
- Roso, V.M., Schenkel, F. S., Miller, S. P. and Schaeffer, L. R. 2005. Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation. *Journal of Animal Science*, 85, pp. 1788-1800.
- Schultz, P.A., and Halpert, M.S. 1993. Global Correlation Of Temperature, NDVI and precipitation. *Advances in Space Research*, 13, pp. 277–280.

- Shen, S.S.P., Dzikowski, P., Li, G. and Griffith, D. 2001. Interpolation of 1961-1997 Daily Temperature and Precipitation Data onto Alberta Polygons of EcoDistrict and Soil Landscapes of Canada. *Journal of Applied Meteorology*, 40, pp. 2162-2177.
- Shen, S.S.P., Yin, H., Canno, K., Howard, A., Chetner, S. and Karl, T.R. 2005. temporal and spatial changes of the agroclimate in Alberta, Canada, from 1901 to 2002. *Journal of Applied Meteorology*, 44, pp. 1090-1105.
- Shin, H.S. and Salas, J.D. 2000. Regional drought analysis based on neural networks. *Journal of Hydrological Engineering* , 5,145-55.
- Skabar, A. 2005. Automatic MLP weight regularization on mineralization prediction tasks. In *Knowledge-based Intelligent Information and Engineering Systems: 9th International Conference, KES 2005, Melbourne, Australia, September 14-16, 2005 : Proceedings*, Khosla et al. (eds).
- Spaulding, K.A. 1998. Neural matamorphic optimization algorithms. Masters Thesis, University of Texas at Austin.
- Srivastava, S.K., Jayaraman, V., Nageswara Rao, P.P., Manikiam, B., and Chandrasekehar, G. 1997. Interlinkages of NOAA/AVHRR derived integrated NDVI to seasonal precipitation and transpiration in dryland tropics. *International Journal of Remote Sensing*, Vol. 18, pp. 2931–2952.
- Soudani, K., Maire, J. L., Dufrene, E., Francois, C., Delpierre, N., Ulrich, E. and Cecchini, S. 2008. Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging

- Spectroradiometer (MODIS) data. *Remote Sensing of Environment*, 112, pp. 2643-2655.
- Sun, X, Liu, D., Chen, A. and Xue, Z. 2006. Neural network with adaptive immune genetic algorithm for eddy current nondestructive testing. In *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006*, pp. 3106-3110.
- Suzuki, R., Masuda, K. and Dye, D. G. 2007 Interannual covariability between actual evapotranspiration and PAL and GIMMS NDVIs of northern Asia. *Remote Sensing of Environment*, 106, pp. 387-389.
- Takata, Y., Funakawa, S., Alshalov, K. Ishida, N. and Kosaki, T. 2007. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. *Soil Science and Plant Nutrition*, 53, pp. 389-399.
- Tan, S. Y. 2007. The influence of temperature and precipitation climate regimes on vegetation dynamics in the US Great Plains: a satellite bioclimatology case study. *International Journal of Remote Sensing*, 28, pp. 4947-4966.
- Tan, Y., and Van Cauwenberghe, A. 1999. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. *Engineering Applications of Artificial Intelligence*, 12, pp. 21-25.
- Tokar, A.S. and Johnson, P.A. 1999. Rainfall-runoff modelling using artificial neural networks. *Journal of Hydrological Engineering*, 4, pp. 232-239.

- Townshend, J.R.G., and Justice, C.O. 1986. Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, 7, pp. 1435–1446.
- Tucker, C.J., Vanpraet, C.L., Boerwinkel, E., and Gaston, A. 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel. *Remote Sensing of Environment*, 13, pp. 461–474.
- Tucker, C.J., Vanpraet, C.L., Sharman, M.J., and Van Ittersum, G. 1985. Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980–1984. *Remote Sensing of Environment*, 17, pp. 233–249.
- Tucker, C.J., and Sellers, P.J. 1986. Satellite remote sensing of primary vegetation. *International Journal of Remote Sensing*, 7, pp. 1395 -1416.
- Vanacker, V., Linderman, M., Lupo, F., Flasse, S. and Lambin, E., 2005, Impact of short-term rainfall fluctuation on interannual land cover change in sub-Saharan Africa. *Global Ecology and Biogeography*, 14, pp.123-135.
- Wang, J., Price, K.P., and Rich, P.M. 2001. Spatial pattern of NDVI in responses precipitation and temperature in Central Great Plains. *International Journal of Remote Sensing*, 22, pp. 3827-3844.
- Wang, X., Xie, H., Guan, H. and Zhou, X. 2007. Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. *Journal of Hydrology*, 340, pp. 12-24.
- Wardlow, B. D., Egbert, S. L. and Kastens, J. H. 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, 108, pp. 290-310.

- Waring, R.H., Coops, N.C., Fan, W. and Nightingale, J.M. 2006. MODIS Enhanced Vegetation Index predicts tree species richness across forested ecoregions in the contiguous U.S.A. *Remote Sensing of Environment*, 103, pp. 218-226.
- Westra, T. and Wulf, D. R.R. 2007. Monitoring Sahelian floodplains using Fourier analysis of MODIS time-series data and artificial neural networks. *International journal of remote sensing*, 28, pp.1595-1610.
- Xiao, X., Boles, S., Liu, J., Zhuang, D., Frohling, S., Li, C., Salas, W. and Moore, B. 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, 95, pp.480–492.
- Xin, J.H., Saho, S. and Chung, K.F. 1999. Color appearance modeling using feedforward networks with bayesian regularization method- Part 1: Forward model. *Color Research and Application*, 25, pp. 424-434.
- Yang, W., Yang, L. and Merchant, J.W. 1997. An assessment of AVHRR/NDVI-ecoclimatological relations in Nebraska, U.S.A. *International Journal of Remote Sensing*, 18, pp. 2161-2180.
- Zealand, C.M. 1997. Short Term Stream Flow Forecasting Using Artificial Neural Networking, M.Sc. thesis, University of Manitoba, Canada.
- Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, C. and Huete, A. 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84, pp. 471–475.

- Zhang, X., Friedl, M. A., Schaff, C. B. and Srahler, A. H. 2004. Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. *Global Change Biology*, 10, pp.1133-1145.
- Zhao, D., Reddy, K.R., Kakani, V.G., Read, J. J., and Koti, S. 2007. Canopy reflectance in cotton for growth assessment and lint yield production. *European Journal of Agronomy*, 26, pp. 335-344.

Table 2.1. Comparison of MODIS sensor characteristics with that of NOAA-AVHRR.

Band	Wavelength ( $\mu\text{m}$ )	
	MODIS	NOAA-AVHRR
Blue	0.459 - 0.479	
Green	0.545 - 0.565	0.55 - 0.68
Red	0.620 - 0.670	
Near-Infrared	0.841 - 0.876	0.725 - 1.1
Middle-infrared	1.230 - 1.250	
Middle-infrared	1.628 - 1.652	
Middle-infrared	2.105 - 2.155	

Table 2.2. Climate predictors used to model VI-climate relationships.

Case	Input
1	PPT
2	T
3	PET
4	PPT, T, PET
5	PPT, T, AI
6	PPT,T,PET, AI

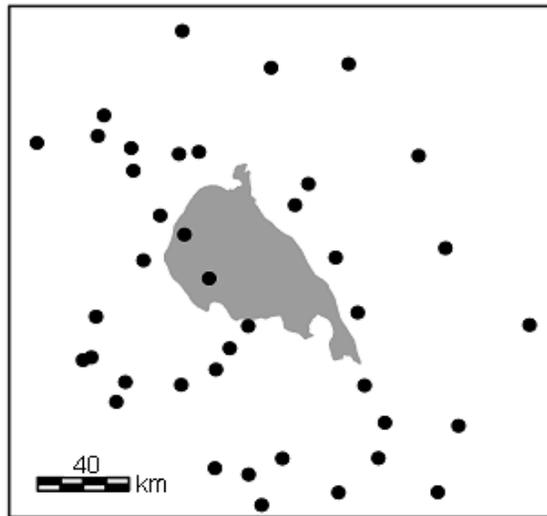
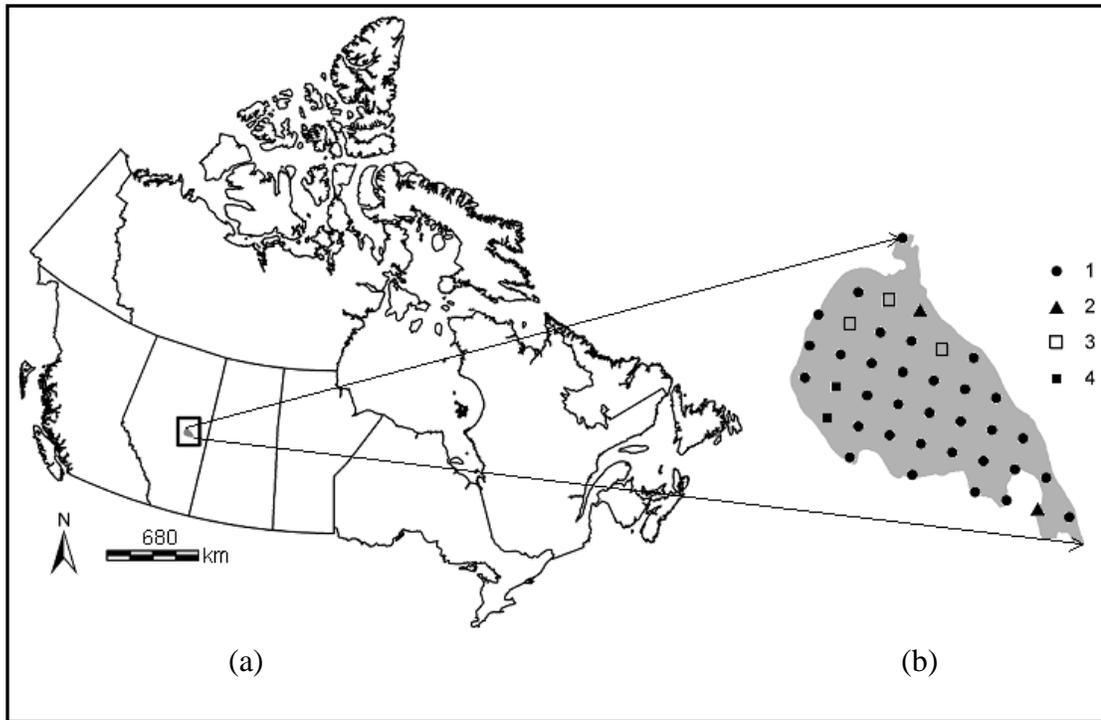
PPT= Precipitation, T= Temperature, PET= Potential evapotranspiration, AI= Aridity index

Table 2.3. Results of modelling NDVI-climate relationship using Back Propagation Neural Network (BPNN).

Case	Calibration			Validation		
	$R^2$	RMSE %	$E_f$	$R^2$	RMSE %	$E_f$
Case 1	0.56	19.05	0.555	0.44	22.40	0.411
Case 2	0.87	10.20	0.872	0.83	12.37	0.829
Case 3	0.87	10.14	0.852	0.83	12.52	0.792
Case 4	0.93	7.69	0.920	0.90	9.58	0.885
Case 5	0.93	7.65	0.917	0.90	9.53	0.876
Case 6	0.95	6.87	0.937	0.92	10.04	0.910

Table 2.4. Correlation between EVI and topographic variables (elevation, slope, aspect).

Variable	Coefficient of determination ( $R^2$ )
Elevation	0.004
Slope	0.02
Aspect	0.01



(c)

Figure 2.1. (a) Location of the study site (boreal mixedwood forest) in Alberta, Canada, (b) location and soil type of township grids used for gridded climate dataset (1= Orthic Gray Luvisol, 2= Orthic Humic Gleysol, 3= Orthic Regosol, 4= Eluviated Eutric Burnisol), and (c) location of the climate stations used to produce the gridded climate dataset of 1991-2005.

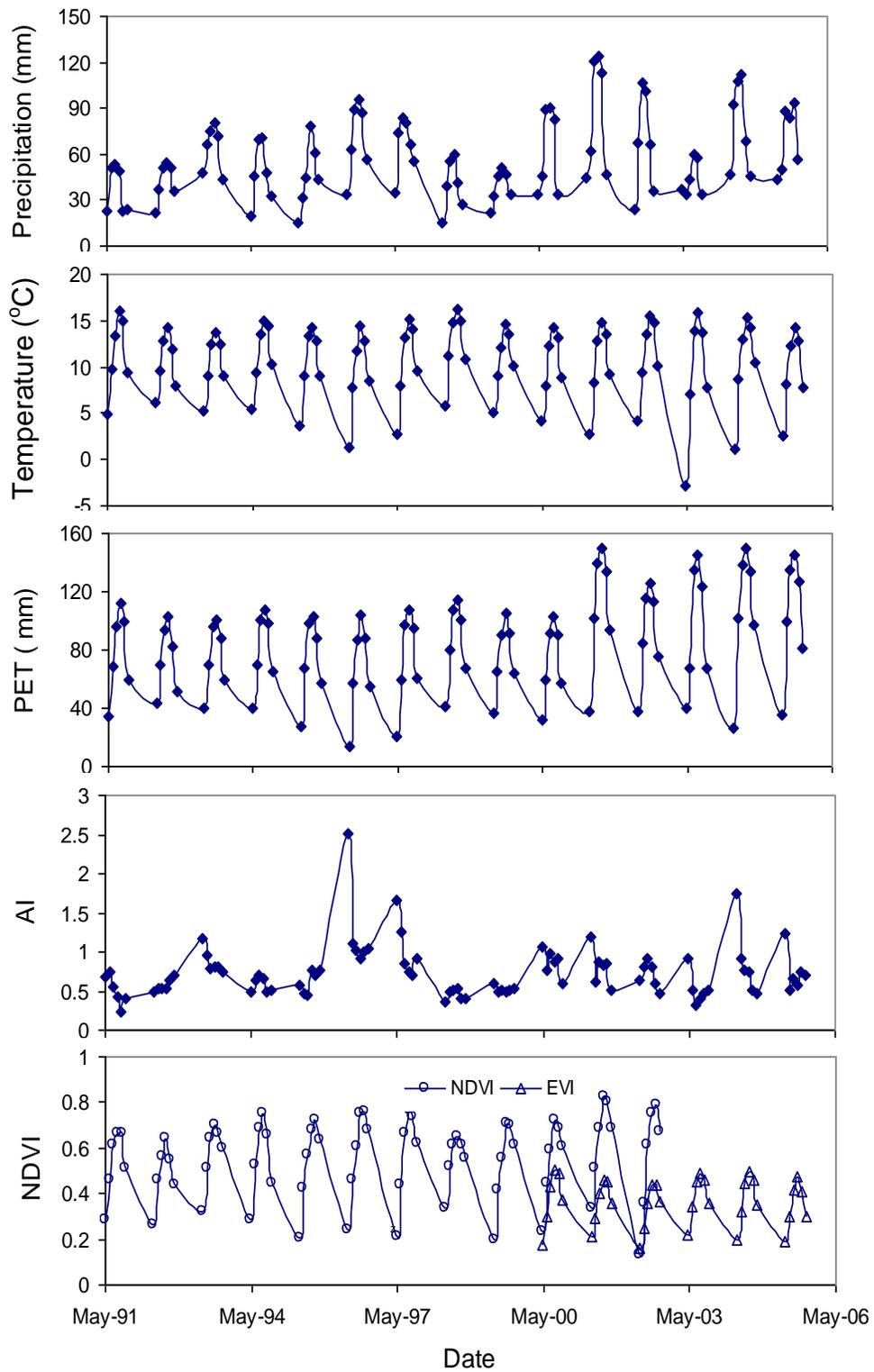


Figure 2.2. Time series of precipitation, temperature, potential evapotranspiration (PET), aridity index (AI) and vegetation indices (NDVI and EVI).

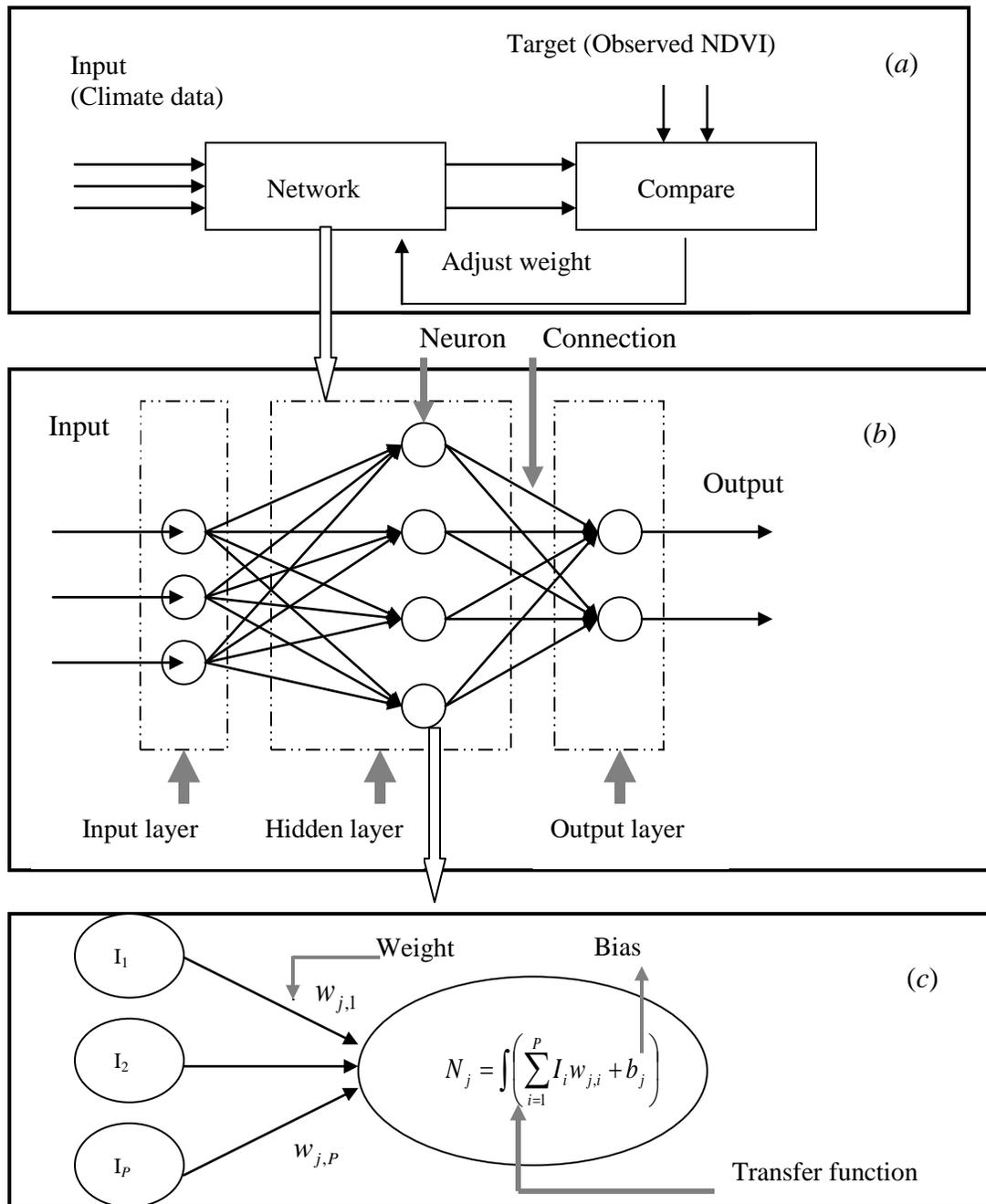


Figure 2.3. (a) The ANN system, (b) the network and (c) a sigmoid neuron.

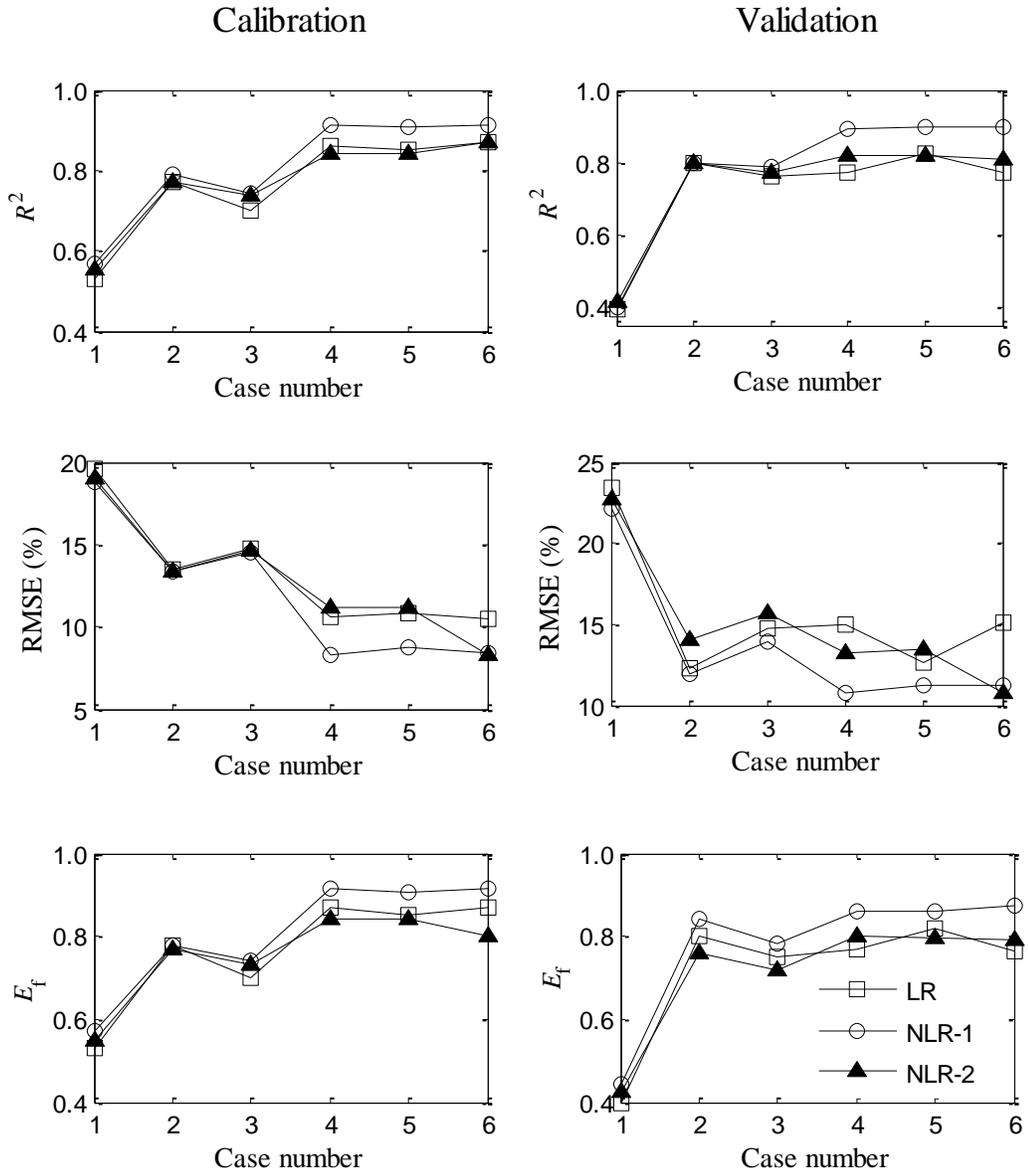


Figure 2.4. Performance of linear (LR) and nonlinear (NLR-1 and NLR-2) regressions for different single/ multiple predictors (case1 to case 6) in modelling the NDVI-climate relationship during calibration (left) and validation stages (right), in terms of  $R^2$ , RMSE and  $E_f$ .

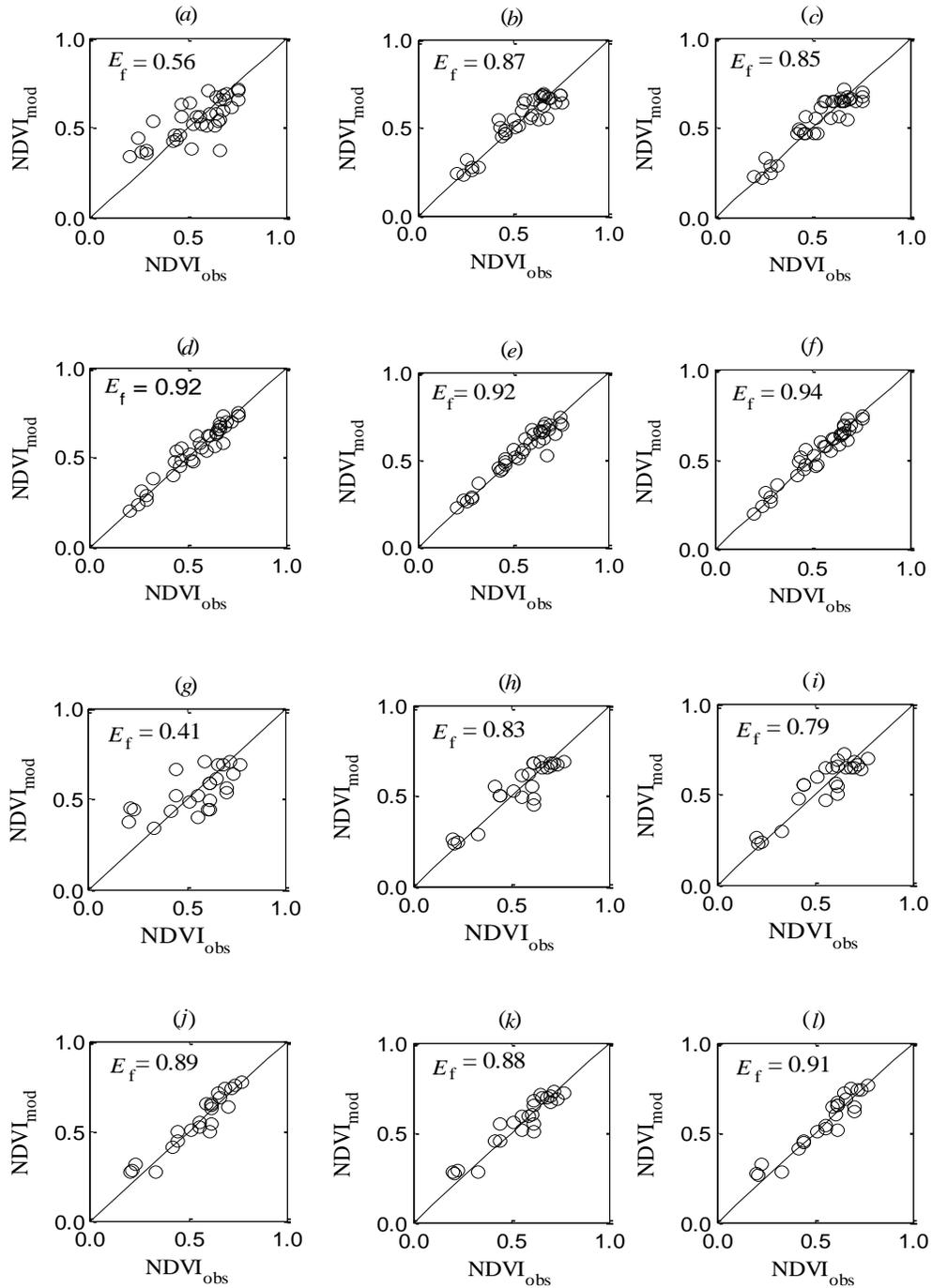


Figure 2.5. Scatter plots of the observed ( $NDVI_{obs}$ ) and model estimated ( $NDVI_{mod}$ ) NDVI using Back Propagation Neural Network (BPNN) for different input cases; (a) to (f) are the calibration results for case 1 to case 6 respectively, while (g) to (l) are the validation results for case 1 to case 6 respectively.

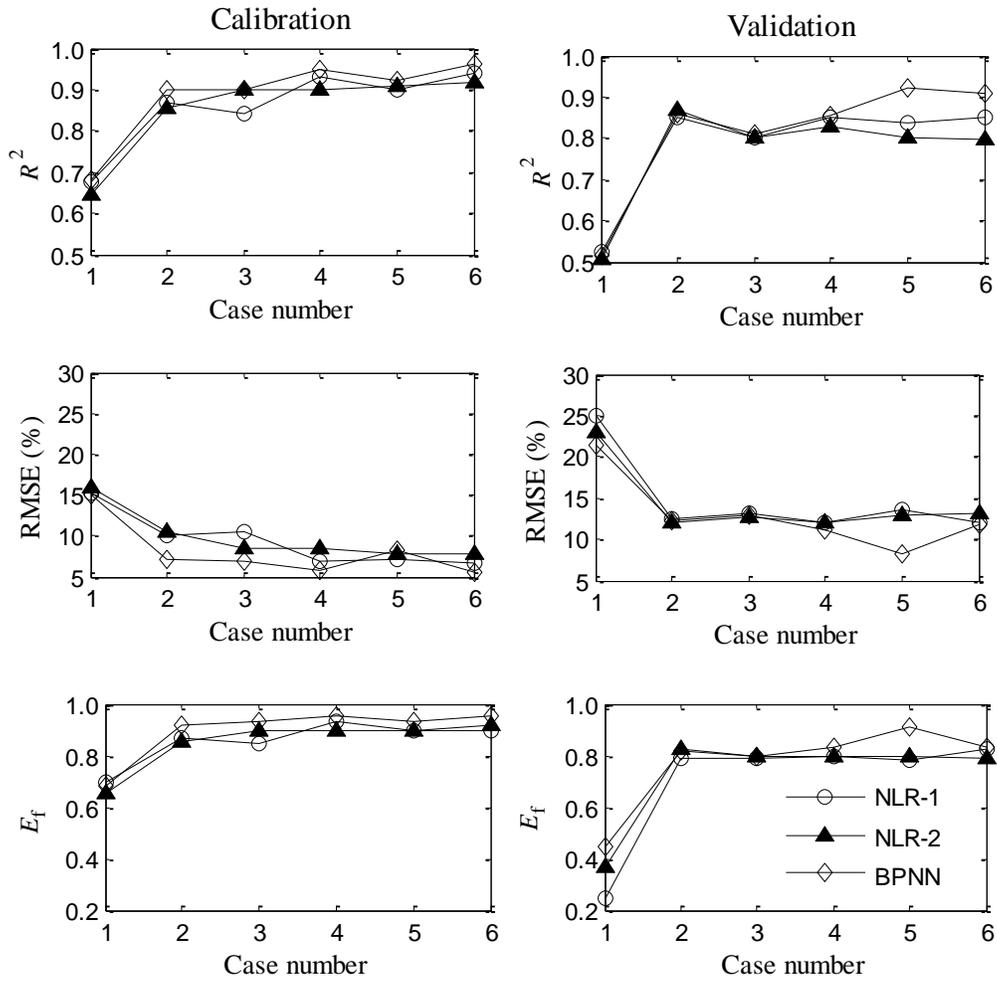


Figure 2.6. Performance of nonlinear regressions (NLR-1 and NLR-2) and Back Propagation Neural Network (BPNN) for different single/ multiple predictors (case1 to case 6) in modelling the EVI-climate relationship during calibration (left) and validation stages (right), in terms of  $R^2$ , RMSE and  $E_f$ .

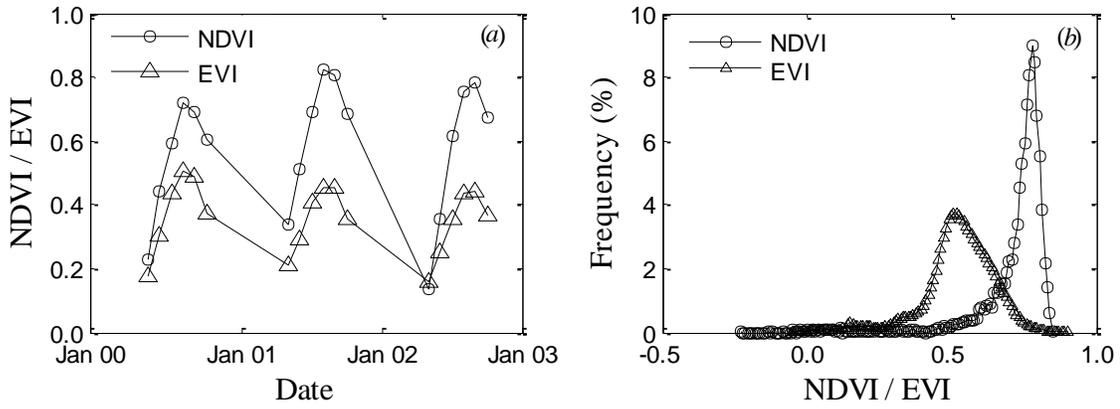


Figure 2.7. Comparing the (a) time series and (b) histogram of NDVI and EVI over the boreal mixedwood forest region of Alberta.

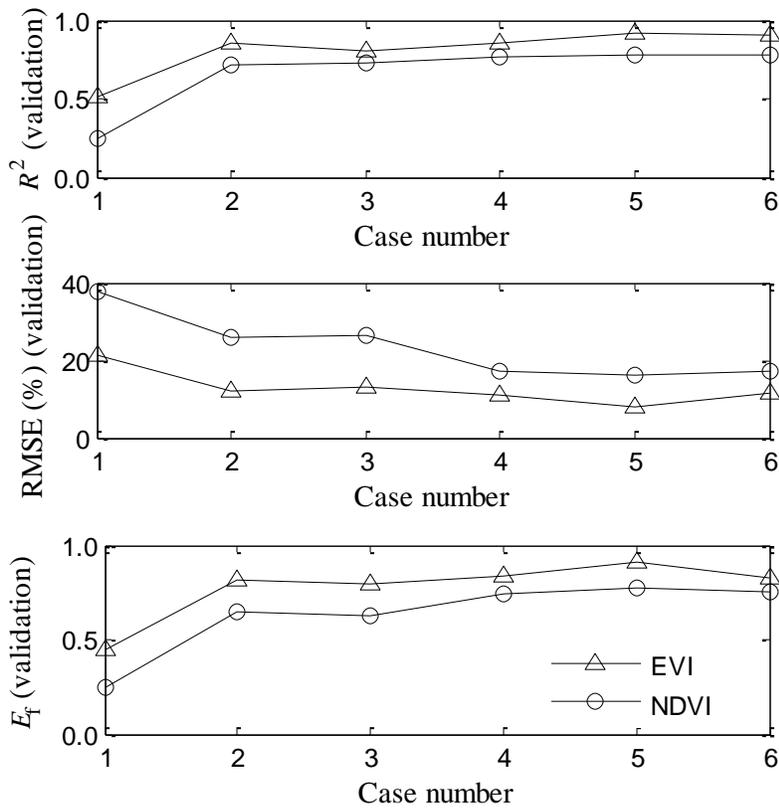


Figure 2.8. Comparing summary statistics of NDVI-climate and EVI-climate relationships for six input cases (Time period of 2000 to 2002).

# Chapter 3

## Soil moisture retrieval from microwave (RADARSAT-2) and optical remote sensing (MODIS) data<sup>\*</sup>

---

### 3.1 Introduction

Even though soil moisture only represents a small fraction of the global water budget, it is nonetheless an important hydrologic variable in many water related applications such as hydrological modeling, crop growth modeling and streamflow forecasting. Accurate and timely measurements of soil moisture are also essential for effective irrigation management, crop selection and plant stress determination. However, the accurate estimation of soil moisture at regional or larger scale is difficult because soil moisture varies highly over space and time, and ground measurements are often time consuming and expensive. Past studies (e.g., Biftu and Gan, 1999; Sokol et al., 2004; Said et al., 2008; Baghdadi et al., 2006 ) have shown that airborne and satellite active microwave sensors can be used to retrieve soil moisture of bare soil or areas with sparse vegetation by developing empirical relationships between soil moisture and microwave

---

<sup>\*</sup> A part of this paper has been submitted for publication. Jahan, N. and Gan, T. Y, 2011. Soil moisture retrieval using microwave (Radarsat-2) and optical remote sensing (MODIS) data. *Canadian Journal of Remote Sensing*.

backscattering. Moran et al. (2004) reported that a soil moisture retrieval model primarily based on microwave remote sensing such as Synthetic Aperture Radar (SAR) sensors is an efficient approach for obtaining spatially distributed soil moisture. The advantages of radar (Radio Detection and Ranging) remote sensing are that its active microwave radiation can penetrate cloud cover and also it can operate both day and night.

### **3.2 Soil moisture retrieval models**

An estimation of soil moisture from radar data is generally done by either theoretical (Fung et al., 1992) or empirical (Oh et al., 1992; Dubais et al., 1995) or semi empirical models (Chen et al., 1995; Oh, 2004). The Integral Equation Model (IEM) (Fung et al., 1992) is one of the most widely used theoretical models for retrieving soil moisture. Unlike some other models, such as the small perturbation model (SPM), geometrical optic model (GOM), physical optic model (POM), etc., which are used specifically for either smooth or rough surfaces, the IEM is applicable to a wide range of surface roughness conditions (Baghdadi and Zribi, 2006). The application of IEM requires three roughness parameters: the standard deviation of surface height,  $\sigma$ ; the surface correlation length,  $L$ ; and the shape of surface autocorrelation function,  $ACF$ . Unfortunately, these parameters are difficult to measure in the field (Baghdadi et al., 2008; Oh and Kay, 1998), and as a result, they are sometimes optimized on the basis of measured soil moisture and radar backscatter (Biftu and Gan, 1999). Besides IEM, several empirical and semi-empirical models with an emphasis on using multi-polarized

and multi-angular data have been developed in recent years (Oh et al., 1992, Oh 2004; Dubois et al., 1995). Oh et al. (1992) used the HH (horizontal transmit and receive), HV (horizontal transmit and vertical receive) and VV (vertical transmit and receive) data as ratios [e.g., co-polarization ratio (HH/VV) and the cross-polarization ratio (HV/VV)] to retrieve surface roughness and soil moisture from two empirical equations. By combining these two equations, the surface roughness can be factored out and these ratios can then be solved for soil moisture. Dubois et al. (1995) also developed a semi-empirical model for estimating soil moisture and surface roughness using copolarized (HH and VV) images. Although extensive studies have evaluated various models, inconsistent results have been obtained (Baghdadi and Zribi., 2006 ; Thoma et al., 2006; Alvarez-Mozos et al., 2007). Some studies have reported good agreement between measured and modeled soil moisture (Sokol et al., 2004; Oh, 2004), while some others have found large discrepancies between them (Baghdadi et al. 2004; Alvarez-Mozos et al., 2007).

In recent years, machine learning techniques such as artificial neural network (ANN) and Support vector machine have been used to retrieve soil moisture. An ANN possesses an ability to learn from past examples without the requirement of explicit physics (Kartalopoulos, 2000). ANNs are capable of approximating any function with a finite number of discontinuities (Phil,2000; Cybenko 1989) as long as sufficient data are available for learning or training of the network. Among the existing methods, the back-propagation neural network or the radial

basis function networks are more common and have widely been used to simulate soil moisture from radar data (Baghdadi et al., 2002; Said et al., 2008; Jana et al., 2008). Besides radar backscatter past studies have used additional predictors such as vegetation variables, topographic variables, soil texture, etc. as input to their ANN models. As for example, Del Frate et al. (2002) used several vegetation variables such as the moisture of stems and leaves, density of stems, length width and thickness of leaves, radius and height of leaves and stems as inputs while Said et al. (2008) used the incidence angle of radar beam, land cover, terrain height, leaf area index, and field measured surface roughness height and plant water content as input variables. Ramirez and Beltran (2008) have used soil temperature, soil texture, terrain elevation, precipitation, field measured soil moisture at 20 cm depth, etc. to determine soil moisture at different depth ranging from 5 to 100 cm using ANN. But estimating parameters such as moisture of stems and leaves, density of stems, plant water content, surface roughness or soil moisture at different height through field measurements may be expensive, time consuming and difficult. Therefore more research has been recommended by the researchers to determine a robust and feasible method for estimating soil moisture avoiding relatively unavailable data (Said et al., 2008; Baghdadi and Zribi, 2006; Alvarez-Mozos, 2007). In this study we will explore the capability of radial basis neural network in estimating soil moisture from radar and optical satellite data.

The support vector machine (SVM) is based on statistical learning theory and can be used to predict a variable through the use of a trained model that utilizes past

data. This learning strategy was developed by Vapnik and coworkers in early 1990s for classification problem and later Vapnik (1995) extended SVM for regression problem. Although, SVMs have successfully been used for pattern recognition and regression in bioinformatics and artificial intelligence, there are also a few applications of SVM in hydrology. Lin et al. (2009) applied SVM to forecast hourly typhoon rainfall in Fei-Tsui Reservoir Watershed in northern Taiwan and compared the results with ANN model while Yang et al. (2006 b) have applied SVM for modeling continental scale evapotranspiration. In this study, we will investigate the potential of SVM in soil moisture retrieval.

### **3.3 Research objectives**

The first generation of SAR sensors, such as ERS-1/2 (European remote sensing) with VV polarization and RADARSAT-1 with HH polarization, provided valuable data based on a single polarization. These two sensors operating in the C-band have been widely used for retrieving both soil moisture and surface roughness (e.g. Baghdadi *et al.*, 2002; Leconte et al., 2004; Baghdadi and Zribi, 2006; Said et al., 2008). However, the unavailability of dual/ quad polarization data from the earlier sensor has lagged this research because it is sometimes difficult to capture the dynamic nature of soil moisture and to assess the influence of surface roughness and vegetation cover on soil moisture from only single polarization data (McNairn and Brisco, 2004). Only few studies have used fully polarimetric data (HH, VV, HV and VH) available from the Shuttle Imaging

Radar (SIR-C) (operation period 1994), Jet Propulsion Laboratory (JPL) AIRSAR (operation period 1988-2004), etc. to retrieve soil moisture (Sokol et al., 2004).

In December 2007, the Canadian Space Agency (CSA) launched RADARSAT-2 (5.405 GHz), the follow-on satellite to RADARSAT-1. RADARSAT-2 is the first commercial spaceborne SAR satellite that produces fully polarimetric datasets (HH, HV, VV and VH). It acquires images at incidence angles ranging from 20° to 60°, swath width ranging from 20- 500 km and a resolution varying from 3 to 100 m. The sensor's quad polarization mode enables it to capture the comprehensive characteristics of the scattering field of a surface. The horizontally polarized wave can penetrate more than the vertically polarized wave and hence provide significant information about the soil moisture. On the other hand the VV polarized backscatter is useful in determining vegetation growth stage, height, type and health while HV and VH polarized backscatters provide complementary information about vegetation structure (McNairn and Brisco, 2004). As the radar backscatter is affected by both vegetation and soil moisture therefore data of HH, VV and HV or VH polarizations should provide an advantage over the data of single HH or VV polarization to retrieve soil moisture from coupled vegetation-soil backscatter. Even though quad polarization data of Radarsat 2 should be better than single polarization data in monitoring soil moisture and roughness, past studies have not adequately investigated the usefulness of quad polarization datasets in this regard (McNairn and Brisco, 2004). Therefore it is time to vigorously investigate the usefulness of the newly available, fully polarimetric

datasets of RADARSAT-2 in characterizing soil surface properties (Sokol et al., 2004).

The objectives of this study are:

- i)** To examine the sensitivity of HH, VV, HV and VH radar backscatter from Radarsat-2 to soil moisture
- ii)** To assess the accuracy of soil moisture retrieved from the quad-polarization data using IEM and empirical regression algorithms;
- iii)** To test the usefulness of optical satellite data (surface temperature and vegetation index) as supplementary predictors in the regression model;
- iv)** To explore the potential of statistical machine learning techniques, ANN and SVM, in soil moisture retrieval;
- v)** To compare the soil moisture retrieval results from the Radarsat-2 data with the results from the single (HH) polarized RADARSAT-1 data, applied to the same Paddle River basin of central Alberta by Biftu and Gan (1999).

### **3.4 Study site**

The study site is located in the Paddle River Basin (53°52'North, 115°32'West) of Alberta, Canada. The basin consists of about 50% mixed forest, 21% coniferous forest, 15% agricultural land, 11% pasture land with short grass, 2% water body, and 1% impervious lands (Biftu and Gan, 1999). The predominant vegetation of this basin is deciduous and aspen forest (Alberta Energy and Natural Resources,

1977). The slope of the basin is 3-5% and the annual runoff coefficient is of about 0.28. The major soil type of this basin is characterized as orthic grey luvisol belonging to Hubalta series, non saline, and moderately fine-textured glacial till. The typical soil texture is clay loam at a depth of 0-10 cm (Biftu and Gan, 1999). The mean temperature of the basin is -15.5°C in January to 15.6°C in July and the mean annual precipitation is about 500 mm. The field sites chosen for collecting soil moisture data are comprised of agricultural land, pasture land (with short grass) and herbaceous land (with grass, weed, short bush) (Figure 3.1 and 3.2). In each year of 2009 to 2011 soil samples were collected on selected days of the early growing period (Table 3.1).

## **3.5 Data**

### **3.5.1 Soil moisture data**

Soil moisture data from 9 sites (4 agricultural, 3 pasture land, 2 herbaceous land) were collected in selected 10 days of 2009-2011 (Table 3.1) when the Radarsat-2 satellite passed over the study site. These dates were chosen because on these dates SAR images were acquired with low incidence angle which is desirable for soil moisture retrieval since at low incidence angle the effect of surface roughness and vegetation on radar backscatter is minimal. The sites were chosen at inter-site distance of 1 to 10 km from each other. On each day, soil moisture data were collected within 2 to 3 hours of Radarsat's image acquisition. At each site, data was collected from 25 sampling points and at a depth ranging from 0 to 5 cm. The sampling points were located at approximately 50 m apart and covered a plot of

about 200m × 200m in area. On those 10 days a total of 2250 soil samples were collected from the 9 sites. The soil samples were collected in sealed plastic bags and after finishing the sample collection on each day, the samples were kept in a moisture room at the University of Alberta to avoid moisture loss. Later these soil samples were analyzed in the soil laboratory at the University of Alberta to obtain gravimetric soil moisture and then converted to volumetric soil moisture.

### 3.5.2 Radar data

10 RADARSAT-2 images were provided by the Canadian Centre for Remote Sensing (CCRS) under the Scientific Operational Applications Research - Education (SOAR-E) initiative which is a joint program between the MacDonald Dettwiler and Associates Ltd. - Geospatial Services Inc (MDA-GSI) and the Government of Canada, represented by the Canadian Space Agency (CSA) and CCRS. The images were obtained in the form of SAR Georeferenced Extra Fine Resolution product (Path Image Plus, SGX) format. The images were obtained in descending mode and low incidence angles. Details of the images are given in Table 1. Each scene was subjected to radiometric calibration, speckle filtering and georeferencing. Radiometric calibration was done following the approach of Shepherd (2000) using the lookup table provided with the data.

$$\sigma_{k,j}^o = \beta_{k,j} + 10 \times \log_{10}(\sin I_j) \quad (3.1)$$

$$\beta_{k,j} = 10 \times \log_{10}((DN^2 + A_0) / A_j) \quad (3.2)$$

Where,  $\sigma_{k,j}^0$  and  $\beta_{k,j}$  are the radar back scatter and brightness of scanline  $k$  and column  $j$ , respectively,  $I_j$  is the incidence angle of column  $j$ ,  $DN$  is the digital number,  $A_0$  is the gain offset and  $A_j$  is the expanded gain scaling value of column  $j$ .

Speckle is a noise that naturally exists in an image due to the coherent interference of scattered wave from the surface elements. Constructive interference causes an increase while destructive interference causes a decrease from the mean intensity and thus randomly modulates the intensity of each pixel. Reducing these effects increases the radiometric resolution which is necessary for the proper interpretation of the image. In this study speckles were filtered using the 7 by 7 Gamma map filter (Lopes et al., 1993) which minimizes the loss of texture information important for mapping forested and agricultural area. Next the images were geometrically corrected by coregistering to a Landsat TM image with respect to many ground control points uniformly distributed across the whole image and based on a first-order polynomial transformation. The original image with a pixel spacing of  $3.125 \times 3.125$  m was then resampled to image of  $25 \times 25$  m resolution using the nearest neighbor method. Lastly, the backscatter values of the study sites were extracted from the processed images. All these operations were performed using the PCI Geomatica software.

### **3.5.3 Landsat and MODIS data**

The Landsat images of the study site dates were acquired from the US Geological Survey (USGS) (<http://landsat.usgs.gov/>). However, no Landsat image of the study area was available for the same dates when Radarsat images were acquired because Landsat images are available at 16-day intervals. Moreover, all these available images suffer from partial to full cloud covers over the study area which hindered the use of these Landsat images. For example we could not get any cloud free data Landsat image for May, 2010. Therefore instead of Landsat images, we used images of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor which views the entire earth surface in every one to two days. For each of the selected dates, MODIS images were available. However in some days, due to the cloud cover effect, data of some portions of the study area were not available and for those effected portions, we used the spectral data of images acquired in the next day. MODIS surface reflectance products and Land surface temperature (LST) data were obtained from the USGS' Land Processes Distributed Active Archive Centre. Reflectance products are atmospherically corrected and available at 250 m resolution. On the other hand, MODIS LST, is available at 1 km resolution and retrieved using the Split Window algorithm and the thermal infra-red bands of MODIS (Wan and Dozier, 1996)

### **3.5.4 Digital Elevation data**

The digital elevation model (DEM) data of the Paddle River basin was collected through the GeoBase portal ([www.geobase.ca.](http://www.geobase.ca/)) of the Canadian Council on

Geomatics (CCOG). The resolution of the data set is 0.75 arc seconds. From the elevation data, topographic variables such as slope and aspect, were calculated using ArcGIS. A detail description of the DEM data quality is available in <http://www.geobase.ca/geobase/en/data/cded/description.html>.

### **3.5.5 Soil properties data**

The soil properties of the study sites were obtained from the Agricultural Region of Alberta Soil Inventory Database (AGRASID) which is a database describing the soil properties and landscapes within the agricultural region of Alberta. Since the beginning of soil mapping in Alberta in 1920, more than 70 reconnaissance soil surveys have been conducted by different organizations, such as the Alberta Research Council, Agriculture and Agri-Food Canada and the University of Alberta. The soils maps and reports were compiled by different mappers, at different map scales (ranging from 1:30000 to 1:750000) and AGRASID is a compilation of these databases. Details of the soil properties are available in the website of Alberta Agriculture (<http://www1.agric.gov.ab.ca/>). For each site we extracted the information of % of sand, silt and clay and water holding capacity of the soil.

## **3.6 Soil moisture retrieval algorithms**

### **3.6.1 Regressions**

Many previous studies have used regression based models for retrieving soil moisture directly from radar backscatter data (Biftu and Gan, 1999; Shoshany et al., 2000; Kelly et al., 2003). However, results obtained in these studies vary from poor (e.g., Sano et al., 1998;  $R^2=0.09$ ), to moderate (Biftu and Gan, 1999;  $R^2=0.45$  to  $0.69$ ), and to excellent (e.g., Shoshany et al., 2000;  $R^2=0.92$ ). Moreover, most of the regression based soil moisture retrieval models have been developed from data of single polarization (either HH or VV). But radar backscatter is influenced by both soil moisture and soil surface characteristics such as surface roughness, vegetation type, vegetation condition in terms of health and growth stage (Said et al., 2008) and single polarization alone may not always be sufficient to capture these variations. In this study we used the quadpolarization backscatter data (HH, VV, HV or VH) of Radarsat-2 either alone or together with vegetation data of MODIS as predictors in the regression model. Because HH backscatter can penetrate the vegetation and provide useful information about soil moisture while the VV backscatter is useful for distinguishing crop type and vertical structure (McNairn and Brisco, 2004; Wooding et al., 1995). On the other hand HV (or VH) backscatter can provide information such as crop conditions such as productivity (McNairn et al. 2004) and biomass of crop (Ferrazzoli et al., 1997). Studies also showed that VV backscatter is related to the leaf water content (Martin et al., 1989) which should be related to the status of available soil moisture.

Besides radar backscatter, some recent studies have also used the normalized difference vegetation index (NDVI) to incorporate the effect of vegetation (Makkeasorn et al., 2006; Yang et al., 2006) on soil moisture retrieval from radar data. NDVI is the reflectance difference between the visible red (R) and near-infrared (NIR) bands, over their sum (Equation 3.1).

$$\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}} \quad (3.3)$$

NDVI is an index of vegetation growth and density. Past studies have used NDVI to study landscape patterns of green biomass, primary production and monitor droughts (Reynolds et al., 2008; Bajgirani et al., 2008). Studies also showed that NDVI reflects the mean climatic condition and therefore can be used to study the climatic response of vegetation (Jahan and Gan, 2011). Furthermore NDVI is related to vegetation type and different vegetation possesses different surface roughness which affect the radar backscatter differently (Gupta et al., 2002; Makkeasorn et al., 2006). Therefore including NDVI as a predictor in addition to the radar backscatter, may provide supplementary information about soil the moisture and surface characteristics of a study site. Some studies (Sandholt et al., 2002; Carlson, 2007; Patel et al., 2009) have reported that the land surface temperature (LST) and NDVI can together model the stress and moisture conditions of vegetation effectively. From the perspective of past findings, we will use radar and optical sensor's data separately, and in several combinations as predictors of soil moisture retrieval by regression algorithms (Table 3.2).

### 3.6.2 Integral Equation Model (IEM)

In this study we used the IEM to retrieve soil moisture. It is a theoretical backscatter model developed for randomly rough dielectric surface and applicable for wide range of surface roughness. Besides IEM, the other theoretical algorithms for modeling the backscattering from rough surfaces are the small perturbation model (valid for slightly rough surface), the standard Kirchhoff model (valid for small surface slope and large surface height) and the geometric optic solution (appropriate for large surface slope with multiple scattering). The complete version of IEM is supposed to be valid for all scales of roughness and a wide range of wavelength. However, due to the complexity of a complete version, an approximate version of IEM is more practical (Altese et al., 1996, Biftu and Gan, 1999, Baghdadi and Zribi., 2006).

The IEM calculates the radar backscatter( $\sigma^o$ ) on the basis of the radar frequency, polarization, angle of incidence, soil's dielectric constant, the root mean squared (rms) surface height ( $\sigma$ ), surface correlation length ( $L$ ) and autocorrelation function ( $ACF$ ). This model can be inverted to estimate the dielectric constant and so the soil moisture, once the other parameters are known. However, an accurate estimation of the surface roughness is still challenging, especially at large basins (Baghdadi et al., 2006). Recently Baghdadi et al. (2008) reported that the roughness variables estimated from field measurements are very sensitive to the profile length. Oh and Kay (1998) found that the accuracy of the rms surface height and correlation length estimated for a surface depend on the length and the

horizontal resolution of the roughness profiles. To overcome the problems associated with the roughness variables, some of the recent studies have proposed notable improvements to the initial version of IEM model. Baghdadi et al. (2004, 2006) proposed an empirical approach to replace the measured correlation length by a model parameter which considers both the true correlation length and the imperfections of the IEM. This calibration parameter determined through calibration is dependent on the roughness, incident angle, polarization and wavelength. Apparently their model simulated backscatters agreed closely with the observed counterparts. Chen et al. (2000) and Wu and Chen (2004) proposed some other improvements to the IEM model and compared the model results with both numerical simulations and laboratory measurements. Since these new models have not been tested extensively in other areas, we have decided to test the original version of IEM (Fung et al., 1992).

In the IEM model backscatter is expressed as follows

$$\sigma^o = \frac{k^2}{2} \exp(-2kz^2 \sigma^2) \sum_{n=1}^{\infty} \sigma^{2n} |I^n|^2 \frac{W^n(-2kx)}{n!} \quad (3.4)$$

where

$$I^n = (2kz)^n f_{hh} \exp(\sigma^2 kz^2) + kz^n [F_{hh}(-kx, 0) + F_{hh}(kx, 0)] / 2$$

$$f_{hh} = -(2R / \cos \vartheta)$$

$$F_{hh}(-kx, 0) + F_{hh}(kx, 0) = \frac{\sin^2 \vartheta (1 + R)^2}{\cos \vartheta} \left[ \left(1 - \frac{1}{\mu}\right) + \frac{\mu \varepsilon - \sin^2 \vartheta - \mu \cos^2 \vartheta}{\mu \cos^2 \vartheta} \right]$$

$$R = \frac{\cos \vartheta - \sqrt{\varepsilon - \sin^2 \vartheta}}{\cos \vartheta + \sqrt{\varepsilon - \sin^2 \vartheta}}$$

$\varepsilon$  = dielectric constant

$\mu$  = magnetic permeability

$k$  = wave number

$\vartheta$  = incidence angle

$k_x = k \sin \vartheta$ ,  $k_z = k \cos \vartheta$

$W^n(-2k_x) =$  roughness spectrum of the surface related to the  $n^{\text{th}}$

power of the surface correlation function  $\rho(\xi)$  by the Fourier transform:

$$W^n(u,0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho^n(\xi) \exp(-ju\xi) d\xi \quad (3.5)$$

To apply IEM some simplifying assumptions are made: the surface correlation function is isotropic and can be represented either by a Gaussian,  $\rho^n(\xi) = \exp[-(\xi^2 / L^2)]$  or the exponential function,  $\rho^n(\xi) = \exp[-(|\xi| / L)]$  and only the real part of the relative dielectric constant ( $\varepsilon$ ) is considered. This assumption is valid in the frequency range of 1 to 6 GHz where the increase of the real part of  $\varepsilon$  is much higher than the increase of the imaginary part of  $\varepsilon$  and therefore the later can be ignored. In this study we have used the exponential correlation function because previous studies (Biftu and Gan, 1999; Chen et al., 1995) showed that the exponential function is less sensitive to the roughness variables than the Gaussian function. Therefore by using this correlation function we can minimize the effect of roughness on the backscatter and so theoretically the soil moisture retrieved from the backscatter should be more accurate.

In summary, according to IEM,  $\sigma^o = f(\text{frequency, polarization, } \theta, \varepsilon, \sigma, L, \text{ correlation function})$ . We obtain the backscatter coefficient and the radar configuration (*frequency, polarization,  $\theta$* ) from the radar image and we assume the correlation function as exponential. So the remaining unknowns in the IEM are roughness variables ( $\sigma$  and  $L$ ) and  $\varepsilon$ . Now from the measured soil moisture ( $\theta$ ) of 2010,  $\varepsilon$  can be computed by inverting the empirical equation (Equation 3.6) of Topp et al. (1985).

$$\theta = (-530 + 292\varepsilon - 5.5\varepsilon^2 + 0.043\varepsilon^3) \times 10^{-4} \quad (3.6)$$

Then using the  $\varepsilon$  obtained from Equation 3.6 and radar backscatter from Radarsat-2 images, the surface parameters ( $\sigma$  and  $L$ ) of IEM will be optimized using the global optimization algorithm, the Shuffled-Complex Evolution (Duan et al, 1993; 1994). This optimization algorithm is a combination of probabilistic and deterministic approaches and uses the concept of competitive evolution and complex shuffling. Once the  $\sigma$  and  $L$  are optimized from the combined data of 2009 and 2010, then  $\varepsilon$  will be computed for each image of 2011 using these optimized surface variables ( $\sigma, L$ ) and the radar data (radar configuration, polarization, incidence angle and backscatter) through IEM inversion. Then soil moisture will be calculated from  $\varepsilon$  using Equation 3.6

### **3.6.3 Artificial Neural Network (ANN)**

ANN is developed on the basis of the neuron cell structure of biological nervous system and can fit almost any type of nonlinear input-output relationships (Hsu et al., 1990; Phil, 2000). ANN has been found to be a robust tool for modeling many

nonlinear hydrological processes because they are able to capture the nonlinear relationships of such processes without requiring the users to understand the physics of the process (Nor et al., 2007). In that regards, using an ANN to retrieve soil moisture is an appealing idea. In this study, the radial basis function network (RBF) is used to retrieve soil moisture from radar data. RBF networks are widely used for nonlinear modeling, pattern recognition and modeling of complex and chaotic dynamical systems (Nor et al., 2007; Islam et al., 2005; Mann and McLaughlin 2000; Lucks and Oki 1999).

RBF neural networks (RBFNN), shown in Figure 3.3, consist of one input, one hidden and one output layers (Broomhead and Lowe, 1988). The input layer takes the input to the network while the hidden layer processes the input by certain non-linear activation functions and sends the computed value to the output layer. The activation function (also known as transfer function) allows non-linearity in the network (Zealand, 1997). The output layer computes the final response of the network. Each layer of a RBF consists of certain number of neurons and connections. The main controlling parameters of any ANN are the strengths of the connections between neurons represented by weights, and biases. RBF is a supervised, feed forward neural network that uses a linear transfer function for the output layer and a nonlinear transfer function, normally Gaussian, for the hidden layer. The value of a neuron in the hidden layer is computed as;

$$N_j = \exp \left( - \frac{\sum_{i=1}^I (N_i - w_{j,i})^2}{1.44\sigma_1^2} \right) \quad (3.7)$$

Where  $N_i$  is the value of the  $i^{\text{th}}$  (where,  $i = 1, 2, \dots, I$ ) unit in the input layer,  $N_j$  is the value of the  $j^{\text{th}}$  (where,  $j = 1, 2, \dots, J$ ) unit in the hidden layer,  $w_{j,i}$  are weights connecting  $j^{\text{th}}$  unit of the hidden layer to the  $i^{\text{th}}$  unit of input layer,  $\sigma_1$  is an adjustable parameter called width or spread. The optimum value of  $\sigma_1$  is chosen by trial and error approach (Phil, 2000). The number of neurons in the hidden layer is kept the same as the number of input patterns.

The output of the hidden layer ( $N_j$ ) is passed on as input to the output layer. The neurons of the output layer do not have any activation function but a bias ' $b_k$ ' is added. The value of an output neuron is computed as

$$O_k = \sum_{j=1}^J (N_j w_{k,j}) + b_k \quad (3.8)$$

where  $O_k$  is the value of the  $k^{\text{th}}$  ( $k = 1, 2, \dots, K$ ) neuron of the output layer,  $w_{k,j}$  is the weight connecting the  $j^{\text{th}}$  neuron in the hidden layer to the  $k^{\text{th}}$  neuron of the output layer.

In a RBFNN, the value of the weights and biases for the hidden layer and the output layer are computed as follows:

### ***Computing hidden layer's weights and $\sigma_1$***

The hidden layer weight (connecting input neurons to the  $j^{\text{th}}$  neuron in the hidden layer) is set equal to the  $j^{\text{th}}$  input pattern presented in the network. The value of  $\sigma_1$  is chosen by trial and error approach (Phil, 2000).

### ***Computing output layer's weights and biases***

The weights and biases in the output layer are calculated from the target (output data of the training data set) data for different input patterns of the training data set. For a given training data set, if the total number of neurons in the in the input, hidden and output layers are  $I$ ,  $J$  and  $K$ , respectively, then the weights and biases for the output layer can be determined by solving the following system of equations,

$$\begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,J} & b_1 \\ w_{2,1} & w_{2,2} & \dots & w_{2,J} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ w_{K,1} & w_{K,2} & \dots & w_{K,J} & b_K \end{bmatrix} * \begin{bmatrix} N_{1,1} & N_{1,2} & \dots & N_{1,I} \\ N_{2,1} & N_{2,2} & \dots & N_{2,I} \\ \dots & \dots & \dots & \dots \\ N_{J,1} & N_{J,2} & \dots & N_{J,I} \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} T_{1,1} & T_{1,2} & \dots & T_{1,I} \\ T_{2,1} & T_{2,2} & \dots & T_{2,I} \\ \dots & \dots & \dots & \dots \\ T_{k,1} & T_{k,2} & \dots & T_{k,I} \end{bmatrix} \quad (3.9)$$

where  $w_{k,j}$  are weights in the output layer,  $b_k$  are biases in the output layer,  $N_{j,i}$  are neurons in the hidden layer and  $T_{k,i}$  are the targets. The value of  $I$  and  $J$  should be equal. In Equation 3.9, the values of  $N_{j,i}$  are already computed from the weights and biases for the hidden layer following Equation 3.7 and target  $T_{k,i}$  is known for a given training data set. Then by solving Equation 3.9, the weights and biases of the output layer are determined. A more detailed description of the radial basis function network is available in Phil, 2000. In this study RBFNN modeling was done using Matlab (see Appendix, Algorithm A2).

In this study, in addition to radar backscatter, NDVI and LST, we have used soil properties (% of sand, silt and clay, water holding capacity), terrain information (slope and aspect) and incidence angles as input to the RBFNN. However the use of incidence angle, soil and terrain properties did not provide any improvement in

the regression algorithms but caused deterioration in their performances, and so the results are not reported here. Soil properties determine the soil infiltration process and the water holding capacity of soil after a rainfall event (Makkeasorn et al., 2006). On the other hand, both slope and aspect may affect the radar backscatter (van Zyl et al., 1993; Baghdadi et al., 2002, Makkeasorn et al., 2006). Therefore these variables were used to train and validate the RBF model. Different network were tested by changing the combination of input variables (Table 3.4). For each case of input, we tested different  $\sigma$  for the RBFNN and the best network is presented here.

#### **3.6.4 Support Vector Machine (SVM)**

SVM is a statistical machine learning technique that transforms nonlinear regression into linear regression by mapping the original low-dimensional input space to a higher dimensional feature space by a nonlinear mapping function  $\phi$  and then performs linear regression in the feature space. SVM is based upon the structural risk minimization (SRM) theory where both the empirical error (e.g., mean squared error) and the model complexity should be minimized simultaneously. The use of SRM principle enhances the general capability of SVMs.

The main objective of SVM regression is to determine the functional dependency between independent variables  $\{x_1, x_2, \dots, x_n\}$  and dependent variables  $\{y_1, y_2, \dots, y_n\}$ . In other words, SVM finds a function  $f(x)$  that yields the output  $\hat{y}$ ,

which is the best estimate of the actual output  $y$  with a small error tolerance of  $\varepsilon$ . In the other words, we do not care about errors as long as they are less than  $\varepsilon$ , but any deviation larger than this will not be accepted (Vapnik, 1995). This is done through the use of an  $\varepsilon$ -insensitive loss function (described later). First, the input vector  $x$  is mapped onto a higher dimensional feature space by a using a nonlinear function  $\phi(x)$ . Then the linear regression is performed in the feature space and can be expressed as

$$\hat{y} = f(x) = \langle w \cdot \phi(x) \rangle + b \quad (3.10)$$

where,  $w$  is the weight vector,  $b$  is the bias, and  $\langle w \cdot \phi(x) \rangle$  is the dot product between  $w$  and  $\phi(x)$ . On the basis of the structural risk minimization principle,  $w$  and  $b$  are estimated by minimizing the following structural risk function:

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n L_{\varepsilon}(y, x, f(x)) \quad (3.11)$$

where,  $L_{\varepsilon}(y, x, f(x))$  is the Vapnik's  $\varepsilon$ -insensitive loss function (Vapnik, 1995), defined as

$$L_{\varepsilon}(y, x, f(x)) = \begin{cases} 0 & \text{for } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \quad (3.12)$$

where  $\|w\|$  is the Euclidean norm of the weight vector. Minimizing  $\|w\|^2$  corresponds to minimizing the model complexity. The parameter  $C$  determines the tradeoff between the model complexity and the training error (Smola and Scholkopf, 1998). A large value of  $C$  means that the model complexity part will be negligible during optimization while small value of  $C$  means that the training error (also called empirical error) has less influence in the optimization

formulation. If  $C = 1$  then it represents that both the model complexity is as important as the empirical error. The loss function of Equation 3.11 and 3.12 are represented by two variables (called slack variables) when data can't be estimated by the function under the precise  $\varepsilon$ . By introducing two slack variables, the optimization problem of Equation 3.11 can be expressed as

$$\begin{aligned}
 & \text{Minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
 & \text{Subject to } y_i - (\langle \mathbf{w} \cdot \phi(x_i) \rangle + b) \leq \varepsilon + \xi_i \\
 & \qquad \qquad \qquad (\langle \mathbf{w} \cdot \phi(x_i) \rangle + b) - y_i \leq \varepsilon + \xi_i^* \tag{3.13} \\
 & \qquad \qquad \qquad \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, 3, \dots, n
 \end{aligned}$$

where  $\xi_i$  and  $\xi_i^*$  are slack variables, represents the lower and upper training errors, respectively and determine the degree to which sample points are penalized if the error is larger than  $\varepsilon$  (Figure 3.4).

The optimization problem of Equation 3.13 can be solved by dual formulation. The key idea of dual formulation is to construct a Lagrange function using both the objective function and corresponding constraints by introducing a dual set of Lagrange multipliers,  $\alpha$  and  $\alpha^*$  (Smola and Scholkopf, 1998; Mangasarian, 1969). The partial derivatives of this Lagrange function with respect to  $w$ ,  $b$ ,  $\xi_i$  and  $\xi_i^*$  should be zero to satisfy the optimizing condition. Substituting the partial derivative output into Equation 3.13 transform this optimization Equation into the following form

Minimize

$$\varepsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) - \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \langle \phi(x_i), \phi(x_j) \rangle$$

Subject to

$$\sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, n \quad (3.14)$$

Then solving Equation 3.14 with constraints determines the Lagrange multipliers

( $\alpha$  and  $\alpha^*$ ) and the regression Equation 3.10 can be rewritten as

$$\begin{aligned} y = f(x) &= \sum_{i=1}^n (\alpha_i^* - \alpha_i) \langle \phi(x_i), \phi(x) \rangle + b \\ &= \sum_{i=1}^n (\alpha_i^* - \alpha_i) K(x_i, x) + b \end{aligned} \quad (3.15)$$

where  $K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle$  is the kernel function that determines the nonlinear dependence between the two input variables  $x$  and  $x_i$ . It should be mentioned that it is not necessary to know the analytical form of the nonlinear transformation function which is difficult to find. In Equation 3.14, only the dot product  $K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle$  is necessary for the optimization and we can generalize the dot product to other functions. The kernel function used in this study is the radial basis function (RBF), which has the following form

$$K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x\|^2\right) \quad (3.16)$$

There are different other kernel functions such as linear, polynomial and sigmoid.

In this study RBF kernel is used because it requires only one parameter which makes the computation process easier (Yang et al., 2006 b).

The steps involved in SVM modeling are as follows: (1) selecting a suitable kernel function and kernel parameter (kernel width,  $\sigma$ ); (2) specifying the ‘ $\varepsilon$ ’ for the width of an insensitive error band; and (3) specifying the parameter, ‘C’ for the cost of error. More details about SVM are available in Vapnik (1995) and Vapnik (1998). The SVM modeling was done using the SVM software WEKA (Hall et al., 2009), developed at the University of Waikato, New Zealand.

### **3.7 Discussions of results**

#### **3.7.1 Linear regression (LR)**

In this study soil moisture is modeled using both linear (single or multiple variable), and nonlinear regression. We first employed linear regression (LR) because of its simplicity. The mean Radarsat backscatter of nine study sites, each averaged over an area of 200m  $\times$  200m (averaged over 64 pixels of 25 m resolution) were regressed against their corresponding mean soil moisture (averaged from 25 samples per site collected over an area of 200m  $\times$  200m each day). Data of 2009-2010 were used for calibration and data of 2011 were used for validating the calibrated models. The HH, VV and HV backscatter data were used as predictor either alone or together.

Table 3.3 and Figure 3.5 show the results obtained for the regression models for the calibration stage. In all 3 simple linear regression algorithms (Equation I to III, Table 3.3) with single radar backscatter as the only predictor, the obtained root mean square error (RMSE) and the correlation coefficients (R) are similar

(ranging from 0.67 to 0.65). Results for the validation stage are similar to the calibration stage. The best result in the calibration stage was obtained using the HH backscatter (Equation I, Table 3.3) as the predictor. Our result is consistent with results reported in previous studies using multipolarized data (Sokol et al., 2004; Baghdadi and Zribi, 2006). Sokol et al. (2004) reported a correlation (R) range of 0.857 to 0.712 between soil moisture and HH, VV and HV radar backscatter of the Spaceborne Imaging Radar (SIR-C) for bare fields.

In the next stage we used all the quad polarization backscatteras predictors in a multi-linear regression (MLR) (Equation IV, Table 3.3). In using MLR, we also checked the multicollinearity between two or more predictor variables to see if they are highly correlated. In case of strong multicollinearity between predictors, the regression coefficients become unstable and suffer from high standard error (Belsely, 1991). A diagnostic tool for multicollinearity is the variable inflation factor (VIF).

$$\text{VIF}_i = \frac{1}{(1 - R_i^2)} \quad (3.17)$$

Where  $\text{VIF}_i$  is the VIF of the predictor variable  $X_i$  and  $R_i^2$  is the coefficient of determination resulting regressing  $X_i$  on all the remaining predictors. Usually, VIF values greater than 10 suggest the presence of multicollinearity (Chatterjee et al., 2000). In our study, the computed VIF for each of the variables has a value that is less than 10. The MLR algorithm showed an improvement over the simple LR with a single backscatter in both the calibration and the validation stages. This is expected because HH, VV and HV backscatter together provide more

information of the surface characteristics and the effects of vegetation on soil moisture, than individual backscatter.

### **3.7.2 Non-linear regression (NLR)**

In addition to LR, the nonlinear regression technique was also applied (Table 3.3). Although different NLR equations were investigated, only the equations that gave good results are reported in Table 3.3. When all three backscatters were used as predictors in the NLR (Equation V, Table 3.3), the result obtained was comparable to the MLR case of same predictors in the calibration stage but in the validation stage, the results of NLR in terms of RMSE was better than that of MLR (Equation IV, Table 3.3).

Next, optical data, NDVI and LST were used either individually (Case 2 and 3, Table 3.2) or together (Case 4, Table 3.2) in addition to radar backscatter in the NLR algorithm. In this study we found that soil moisture is correlated with NDVI and LST ( $R= 0.42$  and  $0.50$ ). Table 3.3 (case 2 to 4) shows that when all radar backscatter (HH, VV and HV backscatter), LST and NDVI were used as predictors, best result was obtained. Although  $R$  in the validation stage was little poorer than the other cases (Equation IV to VII) but a reduction in RMSE during both calibration and validation stages was noticed. It seems that additional information about the vegetation (by NDVI) and the soil surface temperature (by LST), have the potential to improve the accuracy of soil moisture retrieval under a vegetated landscape. This is likely because NDVI provides valuable information

about the vegetation growth stage and condition which affect the surface roughness and hence the backscatter. Moreover, NDVI is an indicator of the vegetation density which controls the depletion of soil moisture via transpiration. On the other hand, LST has been established as a well known indicator of the water stress of vegetation which is related to available soil moisture (Jackson et al., 1981; Patel et al., 2009).

### **3.7.3 Integral Equation Model (IEM)**

Figure 3.6 and 3.7 show the IEM results in which the correlation coefficients obtained are 0.70 and 0.67 and RMSE are 9.74 and 7.25 (% of soil moisture) during the calibration and validation stages, respectively. Even though IEM is a theoretical model, the results of IEM are poorer than that of NLR because of possible difficulties and limitations in applying IEM. First, IEM was originally developed for bare soil and so the effect of vegetation was not explicitly incorporated in this model (Bindlish and Barros, 2001). Therefore applying IEM to a vegetated landscape may incur some errors or uncertainties to the results (Thoma et al., 2006). Secondly, in this study, instead of using measured surface roughness, we have calibrated the surface roughness parameters of IEM using the known soil moisture of 2010 and corresponding radar backscatter and configuration to avoid the limitations of field measurement of surface roughness reported in past studies (Baghdadi et al., 2008; Thoma et al., 2006; Verhoest et al., 2000; Oh and Kay, 1998). A range of surface roughness values, taken from the literature for agricultural, pasture and herbaceous landuses were used in the

optimization algorithm (Biftu and Gan, 1999; Jackson et al., 1997). Later those calibrated surface roughness values were used to retrieve soil moisture for the validation stage of 2009, assuming these calibrated values as time invariant. Although examples (Biftu and Gan, 1999; Thoma et al.; 2006) of using calibrated roughness parameter as time invariant is available in literature, it could possibly lead to some errors. We tried to minimize this problem by choosing spring and early summer of each year for our study when the vegetation was at the early growth stage. In both years, field measurements and corresponding image acquisitions were done during the end of May to the middle of July (approximately 6 weeks). Over this short period it may be reasonable to assume similar roughness for the pasture and herbaceous lands as reported by Leconte et al. (2004) for these vegetation types. On the hand, for agricultural fields, variation of surface roughness may be small during the early growing season because of the combined effect of roughness increase due to vegetation growth and roughness decrease due to weathering and rainfall erosion (Leconte et al., 2004). Thirdly, the presence of residual speckle, uncertainty in the field measured soil moistures, row effect in agricultural field also may produce some errors in the simulated soil moisture.

#### **3.7.4 Radial Basis Function Neural Network (RBFNN)**

To determine the best network structure, we started to train the network with minimum number of input variables, then gradually increased the number of variables and trained the corresponding network. This procedure of adding and

changing the combination predictor variables helped us to determine which predictors are useful in soil moisture retrieval. Data of 2009 and 2010 were used to train the RBFNN network while data of 2011 was used to validate the trained model.

Results from the RBFNN are presented in Table 3.4. For each input case we tested different network by varying the width parameter  $\sigma$  and the best results are shown in Table 3.4 where seven combinations of input variables have been presented. The performances of the networks are presented in terms of R and RMSE. Results show that radar backscatter alone is a weak predictor of soil moisture which means that a complex model such as RNFNN cannot perform well if the input data is deficient (e.g., only backscatter was used as the predictor) and therefore the result obtained is similar to that of the regression models. The worst result was obtained when only LST, NDVI, incident angles, soil and terrain variables were used without the radar backscatter (input Case A-6, Table 3.4). This indicates that radar backscatters are essential to estimate near surface soil moisture. Results from Case A-3 and A-4 were comparable where radar backscatter, LST, NDVI, angles along with either soil or terrain properties were used as predictors. A comparison between the NLR and RBFNN results shows that for the same input variables (backscatter, LST and NDVI), RBFNN (Case A-2) produce slightly better result than the NLR (Case 4) which is expected as in general RBFNN possesses a stronger ability to model nonlinear relationships (Phil, 2000). It is noteworthy that the results of Case A-5 (all variables except

LST and NDVI were used as predictors) were inferior to the results of the input Case A-7 where all variables (including LST and NDVI) were used. Better results obtained for the Case A-7 probably indicates the importance of LST and NDVI in soil moisture retrieval. The best result (Figure 3.8) was obtained for the Case A-7 when all predictor variables were used. The RMSE (RMSE=3.56 and 3.52% during the calibration and the validation stages, respectively) of the Case A-7 are significantly less than the other cases, because these input variables together probably provide valuable information about soil and vegetation surface conditions which control the soil moisture status of a vegetated surface.

### **3.7.5 Support Vector Machine (SVM)**

We used the data of 2009-2010 for calibrating the SVM model which was then validated with the data of 2011. All the input variables were scaled to the range of  $-1$  to  $1$  following standard SVM techniques to eliminate the influence of variables with different magnitudes. The following procedure was followed during model calibration: First, we selected the widely used radial basis function (RBF) kernel which requires only one parameter,  $\sigma$  (Yang et al., 2006). Second, optimal values of  $C$  (cost of errors),  $\varepsilon$  (width of insensitive error band), and  $\sigma$  (kernel parameter) were searched using a grid search method (Chang and Lin, 2005). The SVM parameters that produced the lowest cross-validation errors were selected. At the beginning, we conducted a coarse grid search for  $C$  ( $2^{-2}$ ,  $2^{-1}$ ,  $2^0$ , . . . ,  $2^4$ ),  $\varepsilon$  ( $2^{-5}$ ,  $2^{-4}$ ,  $2^{-3}$ , . . . ,  $2^{-1}$ ) and  $\sigma$  ( $2^{-5}$ ,  $2^{-4.5}$ ,  $2^{-4}$ , . . . ,  $2^4$ ) to identify the values of  $C$ ,  $\varepsilon$ , and  $\sigma$  that produce the lowest mean RMSE during the cross validation.

We then used a gradually finer grid search until the variance of the RMSE was smaller than 0.01.

We explored seven cases of input data, the same as RBFNN, to determine the usefulness of different input variables in soil moisture retrieval. For each input case, we followed the above procedure to train the SVM model. The trained model was later used to simulate the soil moisture of 2011. Results of different input cases are given in Table 3.5. The performance of SVM was compared with respective LR, NLR and RBFNN input cases in terms of RMSE and R. The results (Table 3.5) show that SVM performs much better than the LR, NLR and RBFNN in all input cases, except for Case A-4 and A-5 where the results of SVM were comparable to RBFNN. Among the 7 input cases, the worst result was obtained when radar backscatter was the only input and the results were comparable to that of RBFNN or regression. Including other variables such as LST, NDVI, soil and terrain properties significantly improved the results. Poor result was also obtained when backscatter was omitted from the input data set (RMSE = 4.16 and 4.40% during the calibration and validation stages, respectively). The SVM model driven by all the predictors produced the best overall result (RMSE = 3.23 and 3.11% for the calibration and validation stage, respectively) among all of the retrieval algorithms and the input cases tested (Figure 3.9).

Recent studies have also reported that the superiority of SVM over ANN in different fields of hydrology (Kalra and Ahmad, 2009; Asefa et al., 2005; Yang et al., 2006b, Lin et al., 2009). According to the statistical learning theory, SVMs possess better generalization ability than ANNs and the optimization algorithm for SVM is more robust than that of ANN (Lin et al., 2009). Therefore in general a more reliable model can be obtained by using SVM rather than ANN. A drawback of RBFNN is that it produces a network with as many hidden neurons as there are input vectors. Because of this, RBFNN may sometimes produce less accurate results when many hidden neurons are needed to accurately define a network. Probably because of these advantages SVM gave better results than the RBFNN. However, there are also some drawbacks of SVM. The time taken to tune a SVM model could be too long for some applications of large database (Oyang et al., 2005).

Although we attempted to incorporate different information of soil and vegetation in the RBFNN or SVM, however there are numerous other factors, such as difference in tillage condition and soil compaction, which control soil moisture but were not included as predictors since these information are seldom available. Therefore, some differences between the actual and the simulated soil moisture were noticed. The presence of residual speckle, uncertainty in the field measured soil moistures and uncertainties in the other input data may also cause the discrepancies between actual and simulated soil moisture. Given the data driven nature of machine learning techniques (i.e, ANN or SVM), it is generally

beneficial to have long data series to train the network. But in this study, we have only 10 days of data (90 observations) which may not be sufficient enough to adequately train the RBFNN/ SVM. However examples of using short data set for training machine learning techniques are also available in literature. As for example, Alcazar et al. (2008) used 46 cases to train a neural network to estimate environmental flows. Kagoda et al. (2007) suggest that the amount of data necessary to train an ANN or SVM depends largely on the complexity of the relationship between the input and output data series which is difficult to know a priori. Therefore, it is difficult to determine the optimum amount of input for successful training of any machine learning technique.

### **3.7.6 Results at the watershed scale**

At field scale, the observed soil moisture of each field was computed by averaging all 25 samples (at approximately 50 m spacing over a plot size of 200m × 200 m) collected, while the soil moisture retrieved from RADARSAT2-SAR data was based on the average backscatters of all pixels covering that same area. These 25 point samples may not always be sufficient to capture the actual field average soil moisture, because soil moisture is highly variable spatially, especially in the presence of tile drain and small scale surface heterogeneities. To verify this concern, we checked the observed data of each site and found that in many cases the observed soil moisture of 25 sampling points, collected from the same field on the same day varied by few percentages from each other. So the average soil moisture obtained from 25 samples may not be truly representative of the actual

soil moisture over that  $200 \times 200$  m plot. Another possible error was that the row direction effect of agricultural fields which can significantly affect radar backscatters was ignored. Two fields with the same crop type and conditions but of different row directions may produce different backscatters. In case of HH or VV polarizations, the radar backscatters should be notably higher for fields with row directions perpendicular to the radar look direction, than for backscatters obtained from fields with row directions parallel to the radar look direction, if other conditions are the same (McNairn and Brisco, 2004, Beaudoin et al., 1990; Leconte et al., 2004). The increase of radar backscatters due to the row direction effect is generally within few several decibels but can be as high as 10 decibels (McNairn and Brisco, 2004). Therefore, ignoring the row effect might induce some error in retrieving soil moisture from agricultural fields.

To assess the performance of the models at the watershed scale, we averaged the soil moisture data of all 9 sites, collected on a particular day. Figure 3.10 shows the results obtained from the IEM model and the best NLR algorithm (Case-4), ANN (Case A-7) and SVM (Case A-7) for retrieving soil moisture from radar backscatter data at the watershed scale. The results indicate an improvement at the watershed scale (R of 0.90, 0.85, 0.94 and 0.96 with RMSE of 1.96, 5.56, 1.18, 0.98 % of soil moisture, for the IEM, NLR, ANN and SVM, respectively) over the field scale (R of 0.69, 0.75, 0.81 and 0.85 with RMSE of 9.06, 4.06, 3.55 and 3.19%, for the IEM, NLR, ANN and SVM, respectively) for the combined data of 2009, 2010 and 2011. Results for the watershed scale were better than the field

scale because at a larger spatial scale, the errors caused by heterogeneities at field scale partially cancelled out each other so the overall errors were reduced. A number of studies have also reported more accurate soil moisture retrieval results obtained at the watershed than at the field scales (Leconte et al., 2004; Thoma et al., 2006; Rahman et al., 2008; Kelly et al., 2003). Since the grid resolutions used to model basin hydrology in a distributed manner are mostly larger than the field size (200 m × 200 m) used in this study, we believe that retrieval algorithms developed for the RADARSAT-2 SAR data have the potential to provide valuable soil moisture information at higher spatial resolution than soil moisture obtained from passive microwave sensors such as the Advanced Microwave Sounding Radiometer (AMSR)-E (Njoku et al. 2003), Scanning Multichannel Microwave radiometer (SMMR) (Njoku et al., 1999) or Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001) which can only provide soil moisture information at low spatial resolutions (25 km for AMSER-E and SMMR and 40 km for SMOS).

### **3.7.7 Comparison with Biftu and Gan (1999)**

Biftu and Gan (1999) applied LR and IEM to retrieve soil moisture over the Paddle river basin using the HH backscatter from Radarsat-1. We compared their results with our results of Equation I (Table 3.3) and IEM, where we used HH backscatter as only predictor, same as them. Biftu and Gan (1999) found an R of 0.67 and 0.83 for 1996 and 1997, respectively, while we found R=0.64, 0.73 and 0.68 for 2009 and 2010, respectively, for LR using HH backscatter. Biftu and

Gan (1999) also found better result with the IEM ( $R=0.92$  to  $0.79$  for different dataset) than the current study ( $R= 0.70$  and  $0.67$  for the calibration and the validation stage) using HH backscatter. The possible reason behind this poorer result in our study can be that we conducted our study in the early growing season and so radar backscatter was subjected to more pronounced effect of vegetation. On the other hand Biftu and Gan (1999) conducted their study during the post harvest period when the crops of agricultural fields have already been harvested and vegetation in the grass and pasture land has reached to mature stage, i.e., change of vegetation in their study period was negligible.

### **3.8 Summary and conclusions**

This study investigated the potential of retrieving soil moisture from the newly available quad polarization backscatter dataset from RADARSAT-2. Ten Radarsat-2 images were obtained in 2009-2011 along with simultaneous field measurement of soil moisture at 9 selected sites of the Paddle River Basin, Canada. Data of 2009-2010 was used to calibrate the regression, IEM, RBFNN and SVM algorithms and data of 2011 were used to validate the calibrated algorithms. In the IEM algorithm, calibrated surface roughness parameters were used instead of measured surface roughness variables to avoid the uncertainty usually induced from the profilometer used to measure surface roughness at fields. In this study roughness variables were calibrated from the radar images and soil moisture data of 2009-2010 and later these calibrated values were applied to

retrieve the soil moisture from 2011's data. Results from this study can be summarized as follows:

- i) All HH, VV and HV radar backscatter were significantly correlated with soil moisture (R ranging from 0.64 to 0.68).
- ii) The combined use of all 3 radar backscatters as predictors produced better soil moisture retrieval results than the use of a single backscatter, which is expected partly because VV and HV backscatters provide information about the vegetation dynamics while the HH backscatter provide the information about soil moisture by penetrating through vegetation.
- iii) Using NDVI and LST, from the MODIS sensor, as additional predictors to the radar backscatter data further improved the soil moisture retrieval results with  $R=0.76$  and  $0.74$  during the calibration and validation stage, respectively for the non-linear regression (NLR) case.
- iv) The IEM performed poorer than the non-linear regressions (NLR) where both radar and optical data were used as predictors. The assumption of calibrated surface roughness as time invariant variable could be one of the limitations of applying IEM in this study. Further, the presence of residual speckles even after speckle filtering, uncertainty in field measured soil moisture, row effect in the agricultural field may also introduce some errors and uncertainties in the IEM and other algorithms.

- v) The RBFNN generally performed better than the regression algorithms probably because of the better capability of neural network to simulate complex nonlinear relationships.
- vi) In general, SVM was better than RBFNN because of the better generalization capability of SVM over RBFNN. Among all the methods and input cases tested, the best result was obtained with SVM (RMSE of 3.23 and 3.11% for the calibration and validation stage, respectively) when radar backscatter, LST, NDVI, incidence angles, soil and terrain properties were used as input variables.
- vii) The models showed better performances at the watershed scale (R of 0.90, 0.85, 0.94 and 0.96 with RMSE of 1.96, 5.56, 1.18, 0.98 % of soil moisture, for the IEM, NLR, ANN and SVM, respectively) over the field scale (R of 0.69, 0.75, 0.81 and 0.85 with RMSE of 9.06, 4.06, 3.55 and 3.19%, for the IEM, NLR, ANN and SVM, respectively) for the combined data of 2009, 2010 and 2011 which demonstrates the potential of accurate soil moisture retrieval at larger spatial scale.
- viii) Our results, in terms of HH backscatter as the only predictor, are slightly inferior to that of Biftu and Gan (1999) who also retrieved soil moisture for the Paddle River basin using single (HH) polarized RADARSAT-1 SAR data probably because we conducted our study in the early growing season and so soil moisture retrieval was subjected

to more pronounced effect of vegetation than that of Biftu and Gan (1999) who conducted their study during the post-harvest season.

### 3.9 References

- Alberta Energy and Natural Resources. 1977. A study of watershed management options for controlling flood runoff in the headwaters of the Paddle River Basin. Report. 44, pp. 1-39, Edmonton, Alberta, Canada.
- Altese, E., Bolognani, O. and Mancini, M. 1996. Retrieving soil moisture over bare soil from ERS1 synthetic aperture radar data: sensitivity analysis based on a theoretical surface scattering model and field data. *Water Resources Research*, 32, pp. 653–661.
- Alcazar, J., Palau, A., Vega-Garcia, C. 2008. A neural net model for environmental flow estimation at the Ebro River Basin, Spain. *Journal of Hydrology*, 349 (1–2), 44–55.
- Alvarez-Mozos, J., Gonzalez-Audicana, M. and Casali, J. 2007. Evaluation of empirical and semi-empirical backscattering models for surface soil moisture estimation. *Canadian Journal of Remote Sensing*, Vol. 33 (3), 176-188.
- Asefa, T., M. Kemblowski, Lall, U. and Urroz, G. 2005. Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series, *Water Resources Research*, 41, W12422, doi:10.1029/2004WR003785.
- Lopes, A., Nezry, E., Touzi, R. and Laur, H. 1993. Structure detection and statistical adaptive speckle filtering in SAR images. *International Journal of Remote Sensing*, vol. 14 , pp.1735 - 1758 .

- Baghdadi, N., Gaultier, S. and King, C. 2002. Retrieving surface roughness and soil moisture from SAR data using neural network. *Canadian Journal of Remote Sensing*, 28, pp. 701–711.
- Baghdadi, N., Gherboudj, I., Zribi, M., Sahebi, M., Bonn, F. and King, C. 2004, Semiempirical calibration of the IEM backscattering model using radar images and moisture and roughness field measurements. *International Journal of Remote Sensing*, 25, pp. 3593–3623.
- Baghdadi, N. and Zribi, M. 2006. Evaluation of radar backscatter models IEM, Oh and Dubois using experimental observations. *International Journal of Remote Sensing*, Vol., 27, No. 18, pp. 3831-3852.
- Baghdadi N, Holah N. and Zribi M. 2006. Calibration of the Integral Equation Model for SAR data in C-band and HH and VV polarizations. *International Journal of Remote Sensing* 27(3–4), pp. 805–816.
- Baghdadi, N., Cerdan, O., Zribi, M., Auzet, V., Darboux, F., El Hajj, M., et al. 2008. Operational performance of current synthetic aperture radar sensors in mapping soil surface characteristics: application to hydrological and erosion modelling. *Hydrological Processes*, vol. 22, no. 1, pp. 9–20.
- Beaudoin, A., Le Toan, T., and Gwyn, Q. H. J. 1990. SAR observations and modeling of the C-band backscatter variability due to multiscale geometry and soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, pp. 886–895.

- Bajgiran, P. R., Darvishsefat, A. A., Khalili, A. and Makhdoum, M.F. 2008, Using AVHRR-based vegetation indices for drought monitoring in the Northwest of Iran. *Journal of Arid Environments*, 72, pp. 186–196.
- Belsley, D.A. 1991. *Conditioning Diagnostics—Collinearity and Weak Data in Regression* (New York: John Wiley and Sons).
- Biftu, G. F. and Gan, T.Y. 1999. Retrieving near-surface soil moisture from Radarsat SAR data. *Water Resources Research*, 35(5), pp. 1569-1579.
- Bindlish, R. and Barros, A.P. 2001. Parameterization of vegetation backscatter in radar-based soil moisture estimation. *Remote Sensing of Environment*, 76, pp. 130-137.
- Broomhead, D. S., and Lowe, D. 1988. Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2, 302–355.
- Carlson, T. N. 2007. Review: an overview of the “triangle method” for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors*, Vol. 7, pp. 1612–1629.
- Chang, C. C. and Lin, C. J. 2005. LIBSVM—A Library for Support Vector Machines, Version 2.8. Software Available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- Chatterjee, S., Hadi A. S., and Price, B. 2000. *Regression analysis by example*. 3<sup>rd</sup> edition. John Wiley and Sons, Inc, New York, N.Y.
- Chen, K. S., Yen, S. K. and Huang, W. P. 1995. A simple model for retrieving bare soil moisture from radar scattering coefficients. *Remote Sensing of Environment*, vol. 54, pp. 121-126.

- Chen, K.S., Wu, T.D., Tsay, M.K. and Fung, A.K. 2000. A note on the multiple scattering in an IEM model. *IEEE Transactions on Geoscience and Remote Sensing*, 38, pp. 249–256.
- Cybenko, G. 1989. Approximation by superposition of a sigmoidal function. *Mathematics of Control. Signals and Systems*, 2, 303–314.
- Del Frate, F., et al. 2002. Wheat cycle monitoring using radar data and a neural network trained by a model. *IEEE Transactions on Geoscience and Remote Sensing*, 5772–5781.
- Duan, Q., Gupta, V. K. and Sorooshian, S. 1993. A Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and its Applications*, Vol. 76(3), pp 501-521.
- Duan, Q., Sorooshian, S and Gupta, V. K. 1994. Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models. *Journal of Hydrology*, Vol.158, 265-284.
- Dubois, P., Van Zyl, J. and Engman, T. 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, 33, pp. 915–926.
- Ferrazzoli, P., Paloscia, S., Pampaloni, P., Schiavon, G., Sigismondi, S., and Solimini, D. 1997. The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 35 (1), pp. 5-17.

- Fung, A.K., Li, Z. and Chen, K.S. 1992. Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing*, 30, pp. 356–369.
- Gupta, R. K., Prasad, T. S. and Vijayan, D. 2002. Estimation of roughness length and sensible heat flux from WiFS and NOAA AVHRR data. *Advances in Space Researches*, 29(1), 33– 38.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). *The WEKA Data Mining Software: An Update*. *SIGKDD Explorations*, Volume 11, Issue 1.
- Hsu, K., Gupta, H. V. and Sorooshian, S. 1995. Artificial neural networking modeling of the rainfall-runoff Process. *Water Resources Research*, vol. 3, no 10, pp 2517- 2530.
- Islam, R. 2005. Improved quantitative estimation of rainfall by radar. M.SC Thesis, University of Manitoba, Canada.
- Jackson, R.D., Idso, S.B., Beginato, R.J. and Pinter, P.J. 1981. Canopy temperature as a crop water stress indicator. *Water Resources Research*, 17, pp. 1133–1138.
- Jackson, T. J., McNairn, H., Wertz, M. A., Brisco, B. and Brown, R. J. 1997. First order surface roughness of active microwave observations for estimating soil moisture, *IEEE Transactions on Geoscience and Remote Sensing*, 35(4), 1065– 1069.
- Jahan, J., and Gan, T. Y. 2009. Modeling gross primary production of deciduous forest using remotely sensed radiation budget and ecosystem variables.

Journal of Geophysical Research, Vol. 114, G04026,  
doi:10.1029/2008JG000919.

Jahan, J., and Gan, T. Y. 2011. Modeling vegetation-climate relationship in a central mixed wood forest of Alberta using normalized difference and enhanced vegetation indices... *International Journal of Remote Sensing*, vol. 32(2), pp. 313-335.

Jana, B. R., Mohanty, B. P. and Springer, E. P. 2008. Multiscale Bayesian neural networks for soil water content estimation. *Water Resources Research*, vol. 44, W08408, doi:10.1029/2008WR006879.

Kagoda, P. A., Ndiritu, J., Ntuli, C. and Mwaka, B. 2010. Application of radial basis function neural network to short term stream flow forecasting. *Physics and Chemistry of the Earth*, 35, pp. 571-581.

Kalra, A. and Ahmad, S. 2009. Using oceanic-atmospheric oscillations for long lead time streamflow forecasting. *Water Resources Research*, 45:W03413. doi:10,1029/ 2008WR006855.

Kartalopoulos, S. T. 2000. Understanding neural network sand fuzzy logic: Basic concepts and applications. Prentice-Hall of India, New Delhi, India.

Kelly, R. E. J., David, T. J. A. and Atkinson, P. M. 2003. Explaining temporal and spatial variation in soil moisture in a bare field using SAR imagery. *International Journal of Remote Sensing*, 24(15), 3059– 3074.

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J.M., Font, J., and Berger, M. 2001. Soil moisture retrieval from space: the soil moisture and

- ocean salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1729–1735.
- Leconte, R., Brissette, F., Galarneau, M. and Rousselle, J. 2004. Mapping near surface soil moisture with Radarsat-1 aperture radar data. *Water Resources Research*, VOL. 40, W01515, doi:10.1029/2003WR002312.
- Lin, G. F, Chen, G. R., Wu, M. C., Chou, Y. C. 2009. Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resources Researches*, 45, W08440.
- Lucks, M. B., and Oki, N. \_1999\_. A radial basis function for function approximation. *IEEE Transactions on Neural Networks*, 5, 1099–1101.
- Makkeasorn, A., Chang, B., Beaman, S., Wyatt, C. and Slater, C. 2006. Soil moisture estimation in a semi-arid watershed using RADARSAT-1 satellite imagery and genetic programming. *Water resources Research*, Vol. 42, W09401.
- Mangasarian, O. L. 1969. *Nonlinear Programming*. McGraw-Hill, New York, 1969.
- Mann, I., and McLaughlin, S. 2000. Dynamical system modelling using radial basis function. *IEEE Transactions on Neural Networks*, 7, 461–465.
- Martin, R. D., Asrar, G. and Kanemasu, E. T. 1989. C-band Scatterometer measurements of a tall grass prairie. *Remote Sensing of Environment*, Vol., 29, pp. 281-292.

- McNairn, H. and Brisco, B. 2004. The application of C band polarimetric SAR for agriculture: A review. *Canadian Journal of Remote Sensing*, Vol. 30, No. 3, pp. 525-542.
- McNairn, H., Hochheim, K. and Rabe, N. 2004. Applying polarimetric radar imagery for mapping the productivity of wheat crops. *Canadian Journal of Remote Sensing*, Vol. 30, No. 3, pp 517-524.
- Moran, M. S., Peters-Lidard, C. D., Watts, J. M., & McElroy, S. 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian Journal of Remote Sensing*, 30, 1–22.
- Njoku, E. G., and Li, L. 1999. Retrieval of land surface parameters using passive microwave instruments at 6 – 18 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1), 79– 93.
- Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K. and Nghiem, S. V. 2003. Soil moisture retrieval from AMSR-E, *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 215– 229.
- Nor, N. I. A., Harun, S., Kassim, A. H. M. 2007. Radial basis function modeling of hourly streamflow hydrograph. *Journal of Hydrologic Engineering*, 112, pp. 113-123.
- Oh, Y., and Kay, Y. 1998. Condition for precise measurement of soil surface roughness. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 691–695.

- Oh, Y., Sarabandi, K., and Ulaby, F.T. 1992. An empirical model and inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, pp. 370–381.
- Oh, Y. 2004. Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 42, pp. 596–601.
- Oyang, Y. J., Hwang, S. C., Ou, Y. Y., Chen C. Y. and Chen, Z. W. 2005. Data classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 225-236.
- Patel, N. R., Anapashsha, R., Kumar, S., Saha, S. K. and Dadhwal, V. K. 2009. Assessing potential of MODIS derived temperature/vegetation condition index (TVDI) to infer soil moisture status. *International Journal of Remote Sensing*, Vol.30, No.1, pp-23-39.
- Phil, P. 2000. *Neural Networks*, 2nd Edition, Grass root Series, MacMillan Publisher, England.
- Rahman, M. M., Moran, M.S., Thoma, D. P., Bryant, R., Collins, C. D. H., Jackson, T., Orr, B. J. and Tischler, M. 2008. Mapping surface roughness and soil moisture using multi-angle radar imagery without ancillary data. *Remote Sensing of Environment*, Vol. 112, pp. 391-402.
- Ramrez-Beltran, N. D., Castro, J. M., Harmsen, E. and Vaquez, R. 2008. Stochastic transfer function model and neural networks to estimate soil

- moisture. *Journal of the American Water Resources Association*, vol., 44(4), pp. 847-865.
- Raynolds, M.K., Komiso, J.C., Walker, D.A. and Verbyla, D. 2008. Relationship between satellite-derived land surface temperatures, arctic vegetation types, and NDVI. *Remote Sensing of Environment*, 112, pp. 1884–1894.
- Said, S., Kothyari, U. C. and Arora, M. K. 2008. ANN based soil moisture retrieval over bare and vegetated surface areas using ERS-2 SAR data. *Journal of Hydrologic Engineering*, Vol. 13, No. 6, pp. 461-471.
- Sandholt, I., Rasmussen, K. and Andersen, J. 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79, 213-224.
- Sano, E. E., Huete, A. R., Troufleau, D., Moran, M. S. and Vidal, A. 1998. Relation between ERS-1 synthetic aperture radar data and measurements of surface roughness and moisture content of rocky soils in a semiarid rangeland. *Water Resources Research*, 34(6), 1491– 1498.
- Shepherd N. 2000. Extraction of Beta-Nought and Sigma-Nought from RADARSAT CDPF Products. Canadian Space Agency Document AS97-5001.
- Shoshany, M., Svoray, T., Curran, P. J., Foody, G. M. and Perevolotsky, A. 2000. The relationship between ERS-2 SAR backscatter and soil moisture: Generalization from a humid to semi-arid transec. *Internatinal Journal of Remote Sensing*, 21(11), 2337– 2343.

- Smola, A. J. and Scholkopf, B. 1998. A tutorial on Support Vector Regression. NeuroCOLT2 Technical Report Series (NC2-TR-1998-030), University of London, UK.
- Sokol, J., Mcnairn, H. and Pultz, T. J. 2004. Case studies demonstrating the hydrological applications of C band multipolarized and polarimetric SAR. Canadian Journal of Remote Sensing, vol. 30, no. 3, pp. 470-483.
- Thoma, D. P., Moran, M. S., Bryant, R., Rahman, M. and Holifield-Collins, C. D., Skirvin, S., Sano, E. E., and Slocum, K. 2006. Comparison of four models to determine surface soil moisture from C-band radar imagery in a sparsely vegetated semiarid landscape. Water Resources Research, 42, W01418, doi:10.1029/2004WR003905.
- Topp, G. C., and J. L. Davis. 1985. Measurements of soil water content using time domain reflectometry: A field evaluation. Soil Science Society of American Journal, 49, 19-24, 1985.
- van Zyl, J. J., Chapman, B. D., Dubois, P. and Shi, J. 1993. The effect of topography on SAR calibration. IEEE Transaction Geoscience and Remote Sensing, 31(5), 1036– 1043.
- Vapnik, V. N. 1995. The nature of statistical learning theory. Springer, New York.
- Vapnik, V. N. 1998. Statistical Learning Theory. Wiley, New York.
- Verhoest, N. E. C., Hoeben, R., De Troch, F. P. and Troch, P. A. 2000. Soil moisture inversion from ERS and SIR-C imagery at the Zwalm catchment,

- Belgium. *IEEE Transactions on Geoscience and Remote Sensing*, 5, 2041–2043.
- Wan, Z., and Dozier, J. 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34, 892–905, doi:10.1109/36.508406.
- Wooding, M.G., Attema, E., Aschbacher, J., Borgeaud, M., Cordey, R.A., de Groof, H., Harms, J., Lichtenegger, J., Nieuwenhuis, G., Schmullius, C., and Zmuda, A.D. 1995. *Satellite radar in agriculture: experience with ERS-1*. ESA Special Publication SP-1185. 72 pp.
- Wu, T.D. and Chen, K.S. 2004. A reappraisal of the validity of the IEM model for backscattering from rough surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 42, pp. 743–753.
- Yang, H., Shi, J., Li, Z. and Guo, H. 2006. Temporal and spatial soil moisture change pattern detection in an agricultural area using multi-temporal Radarsat Scan SAR data. *International Journal of Remote Sensing*, Vol. 27, No. 19. pp. 4199-4212.
- Yang, F., White, M. A., Michaelis, A. R. Ichii, K., Hashimoto, H., Votava, P., Zhuu and Nemani, R. R. 2006. Prediction of continental-scale evapotranspiration by combining MODIS and Ameriflux data through support vector machine. *IEEE Transaction on Geoscience and Remote Sensing*, 44 (11), pp. 3452 -3461.
- Zealand, C. M. 1997. *Short Term Stream Flow Forecasting Using Artificial Neural Networking*, M.Sc. thesis, University of Manitoba, Canada.

Table 3.1. Description of 7 Radarsat2 images acquired over the Paddle River basin at descending orbit.

Acquisition date	Scene Centre		Beam mode	Angle of incidence
	Longitude	Latitude		
29 May, 2009	115°34'04" W	53°52'33" N	FQ8	27.83
22 June, 2009	115°34'04" W	53°52'28" N	FQ8	27.83
16 July, 2009	115°33'40" W	53°52'30" N	FQ8	27.83
17 May, 2010	115°32'38" W	53°53'33" N	FQ4	23.20
24 May, 2010	115°33'32" W	53°53'14" N	FQ8	27.83
June 10, 2010	115°33'13" W	53°53'29" N	FQ4	23.20
July 4, 2010	115°32'42" W	53°53'31" N	FQ4	23.20
May 19, 2011	115°33'39" W	53°53'47" N	FQ8	27.83
June 5, 2011	115°33'12" W	53°53'54" N	FQ4	23.20
June 12, 2011	115°33'52" W	53°53'40" N	FQ8	27.83

Table 3.2. Different input cases for the linear and nonlinear regressions.

Case Number	Variables
Case 1	Polarimetric radar data ( $\sigma_{HH}^{\circ}$ , $\sigma_{VV}^{\circ}$ , $\sigma_{HV}^{\circ}$ )
Case 2	Polarimetric radar data and NDVI
Case 3	Polarimetric radar data and LST
Case 4	Polarimetric radar data, NDVI and LST

Table 3.3. Results from the linear and nonlinear regressions.

Case Number	Regression Equations (Equation number)	Calibration		Validation	
		R	RMSE (%)	R	RMSE (%)
	$\theta = 1.64 \sigma_{HH}^o + 42.81$ (I)	0.67	4.73	0.68	4.65
	$\theta = 1.55 \sigma_{VV}^o + 42.36$ (II)	0.65	4.81	0.64	4.85
	$\theta = 1.81 \sigma_{HV}^o + 58.76$ (III)	0.66	4.81	0.64	4.80
Case 1	$\theta = 0.99 \times (-\sigma_{HH}^o) + 0.32 \times (-\sigma_{VV}^o) + 0.69 \times (-\sigma_{HV}^o) + 51.06$ (IV)	0.70	4.63	0.75	4.847
	$\theta = 67.05 \times (-\sigma_{HH}^o)^{-0.18} + 194 \times (-\sigma_{VV}^o)^{-2.37} + 133.77 \times (-\sigma_{HV}^o)^{-0.22} - 89.55$ (V)	0.70	4.61	0.76	4.35
	$\theta = 65.15 \times (-\sigma_{HH}^o)^{-0.19} + 100.62 \times (-\sigma_{VV}^o)^{-1.82} + 120.83 \times (-\sigma_{HV}^o)^{-0.18} + 4.14 \times \text{NDVI}^{1.59} - 91.42$ (VI)	0.71	4.54	0.77	4.21
Case 2	$\theta = 31.82 \times (-\sigma_{HH}^o)^{-0.12} + 30.83 \times (-\sigma_{VV}^o)^{-0.07} + 143.61 \times (-\sigma_{HV}^o)^{-0.24} + 129.63 \times \text{LST}^{-0.88} - 104.34$ (VII)	0.74	4.41	0.75	4.15
Case 3	$\theta = 24.78 \times (-\sigma_{HH}^o)^{-0.28} + 45.69 \times (-\sigma_{VV}^o)^{-1.08} + 137.24 \times (-\sigma_{HV}^o)^{-0.52} + 115.25 \times \text{LST}^{-0.07} - 6.31 \times \text{NDVI}^{20.53} - 113.56$ (VIII)	0.76	4.30	0.74	3.97
Case 4					

Table 3.4. Input cases for Artificial Neural Network and Results.

Input Case	Input variables	Calibration results		Validation results	
		R	RMSE (%)	R	RMSE (%)
A-1	Radar backscatter ( $\sigma_{HH}^{\circ}$ , $\sigma_{VV}^{\circ}$ , $\sigma_{HV}^{\circ}$ )	0.70	4.51	0.73	4.21
A-2	Radar backscatter, LST, NDVI	0.76	4.22	0.73	3.90
A-3	Radar backscatter, LST, NDVI, angle, soil properties(% of sand, silt, clay, water holding capacity)	0.80	3.84	0.77	3.55
A-4	Radar backscatter, LST, NDVI, angle, terrain properties (slope and aspect)	0.82	3.61	0.76	3.53
A-5	Radar backscatter, soil properties, angle, terrain properties	0.75	3.77	0.70	4.32
A-6	LST, NDVI, angle, soil properties and terrain properties	0.68	4.65	0.41	5.85
A-7	Radar backscatter, LST, NDVI, angle, soil properties and terrain properties	0.83	3.56	0.76	3.52

Table 3.5. Input cases for Support Vector Machine and Results.

Input	Input variables	Calibration results		Validation results	
Case		R	RMSE (%)	R	RMSE (%)
A-1	Radar backscatter ( $\sigma^{\circ}_{HH}$ , $\sigma^{\circ}_{VV}$ , $\sigma^{\circ}_{HV}$ )	0.70	4.49	0.77	4.10
A-2	Radar backscatter, LST and NDVI	0.77	4.13	0.76	3.55
A-3	Radar backscatter, LST, NDVI, angle, soil properties(% of sand, silt, clay, water holding capacity)	0.85	3.35	0.78	3.51
A-4	Radar backscatter, LST, NDVI, angle, terrain properties (slope and aspect)	0.82	3.57	0.76	3.63
A-5	Radar backscatter, soil properties, angle, terrain properties	0.79	3.77	0.70	4.28
A-6	LST, NDVI, angle, soil properties and terrain properties	0.76	4.16	0.49	4.40
A-7	Radar backscatter, LST, NDVI, angle, soil properties and terrain properties	0.88	3.23	0.84	3.11

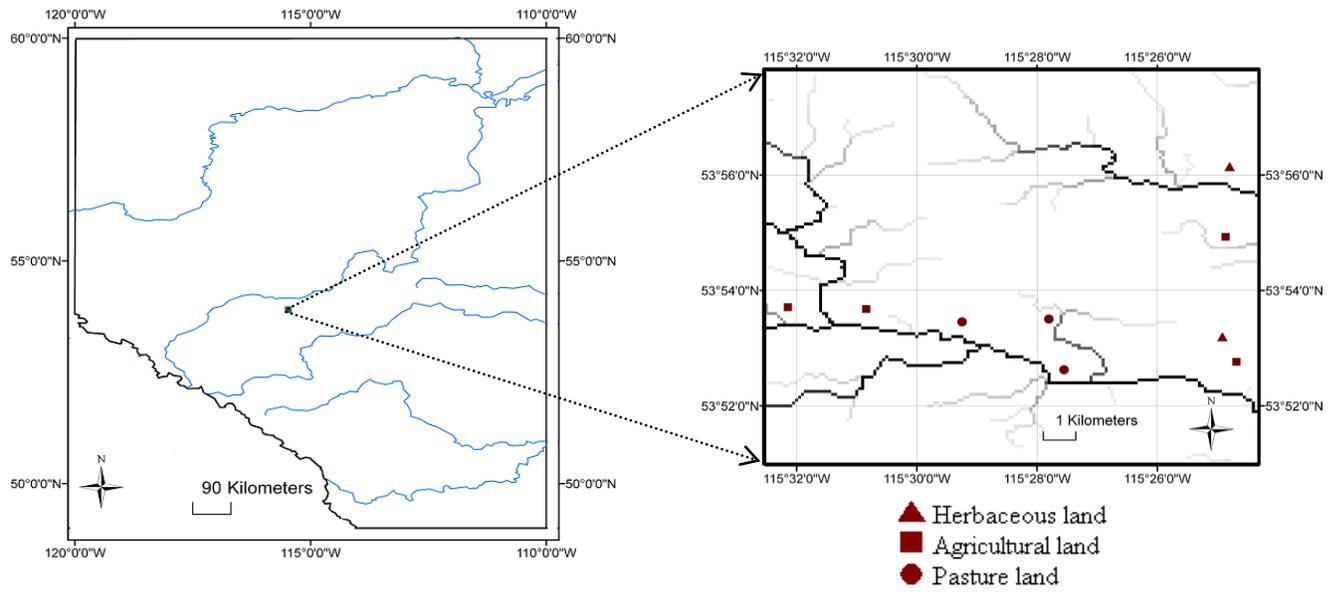


Figure 3.1. (a) Location of the study area within the province Alberta, Canada. (b) Location of individual study sites within the Paddle river basin.

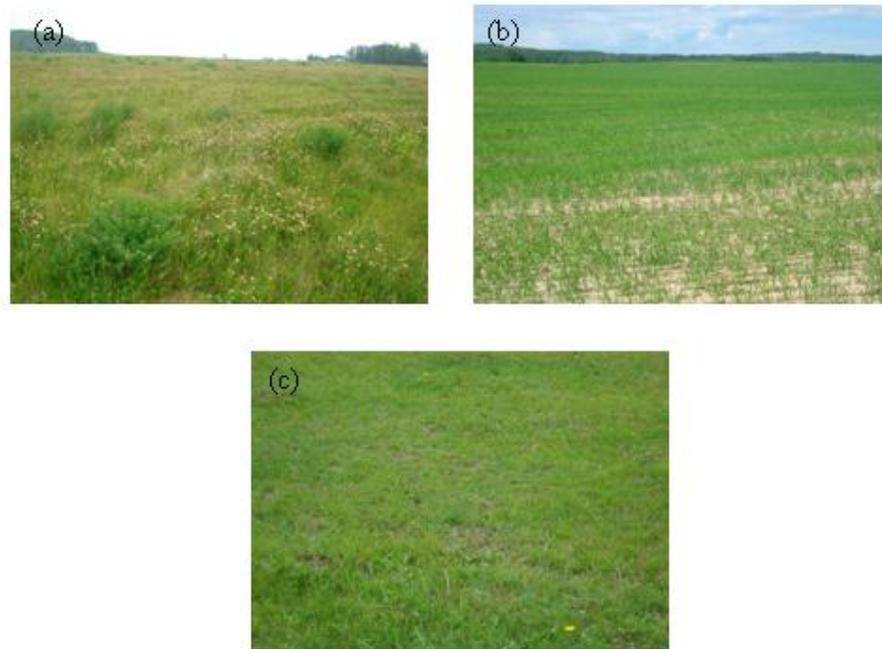


Figure 3.2. Photographs of 3 landcover types used in this study: (a) Herbaceous land, (b) Agricultural land, (c) Pasture land.

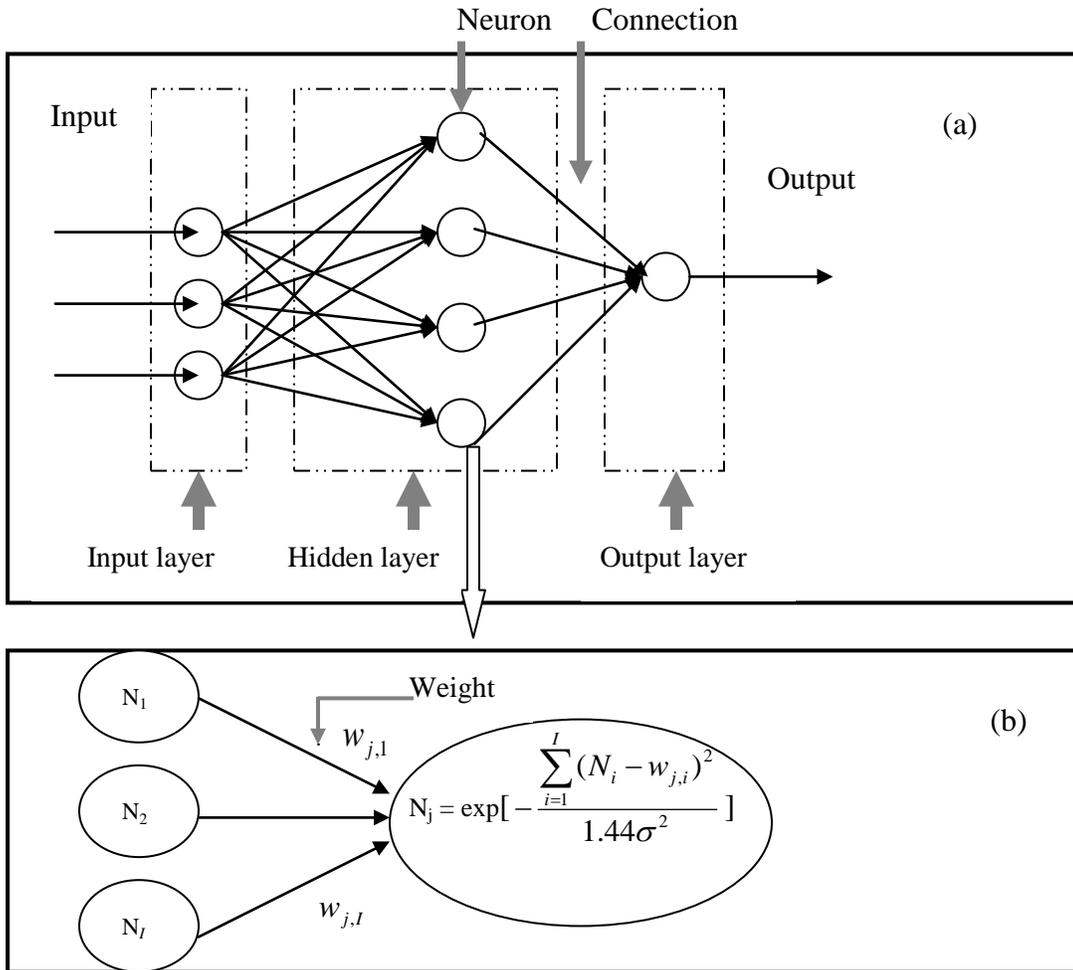


Figure 3.3. Architecture of a typical radial basis neural network.

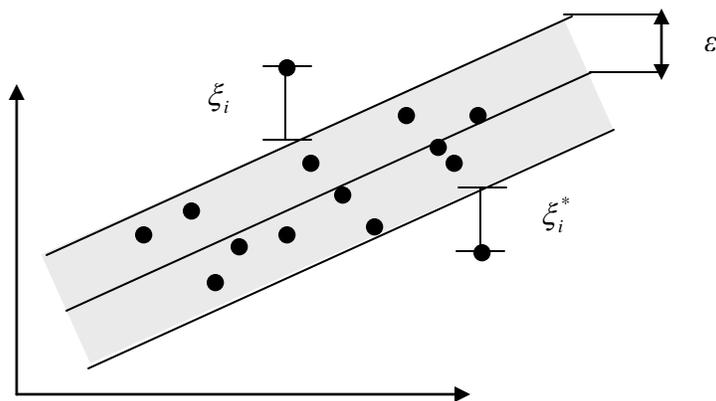


Figure 3.4. Details of  $\epsilon$  band with slack variables  $\xi_i$  and  $\xi_i^*$  and data points (black circles).

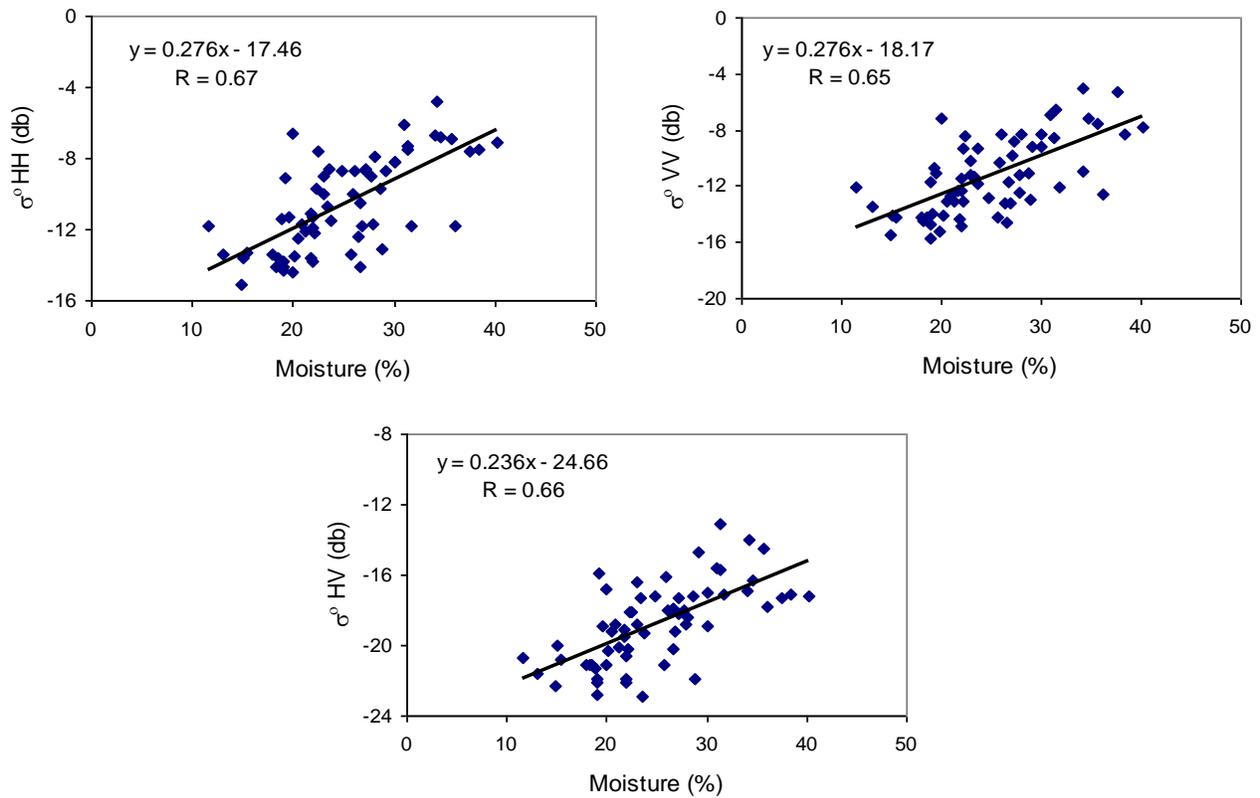


Figure 3.5. Sensitivity of radar backscatter to volumetric soil moisture with respect to the combined data of 2009 and 2010.

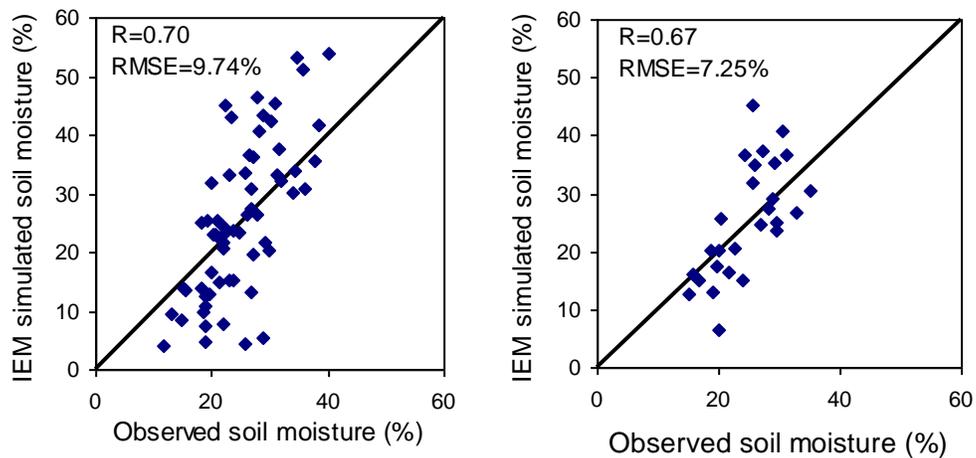


Figure 3.6. Regression between observed and IEM simulated volumetric soil moisture during (a) the calibration and (b) the validation stage.

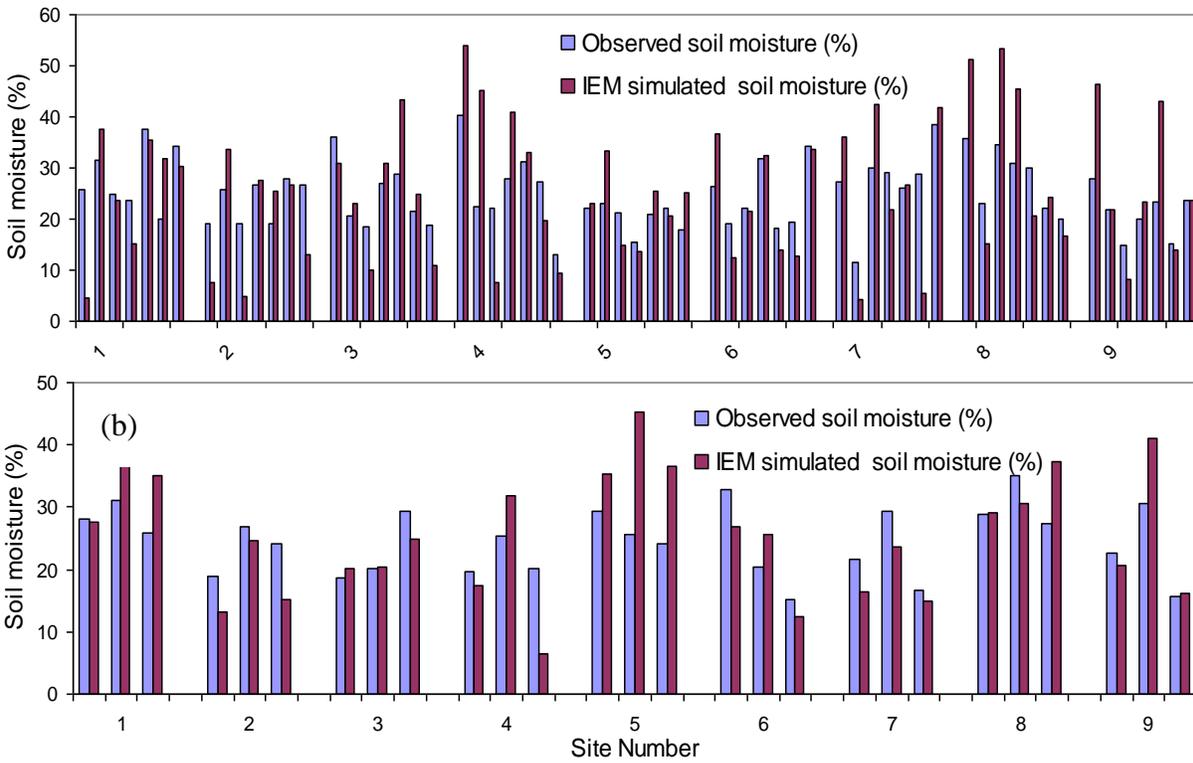


Figure 3.7. Observed and IEM simulated volumetric soil moisture at different sites during (a) the calibration and (b) the validation stage.

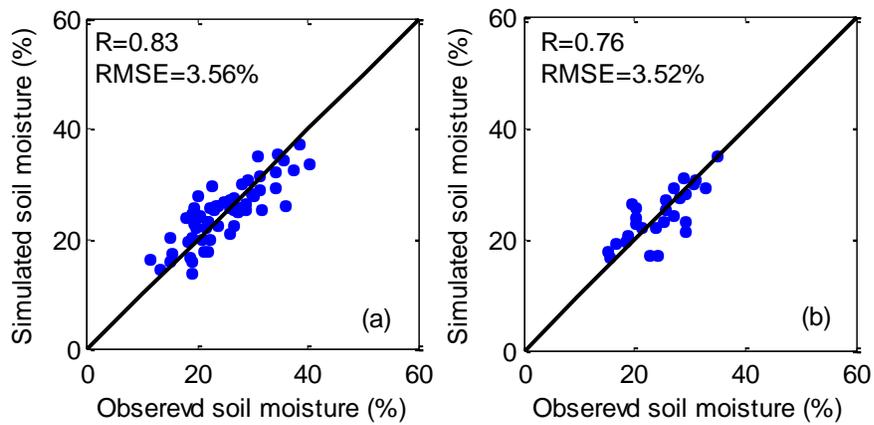


Figure 3.8. Observed soil moisture versus ANN simulated soil moisture where radar backscatter, incidence angle, LST, NDVI, soil properties and terrain properties were used as predictors.

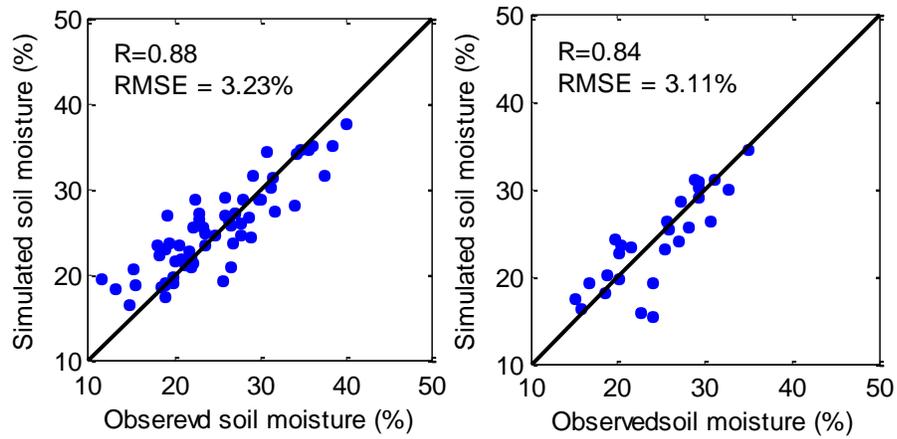


Figure 3.9. Observed soil moisture versus SVM simulated soil moisture where radar backscatter, incidence angle, LST, NDVI, soil properties and terrain properties were used as predictors.

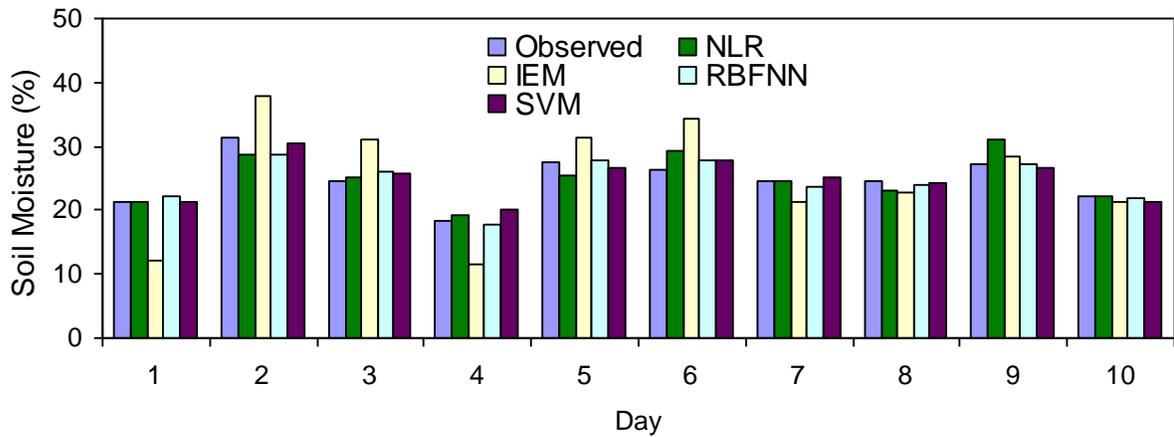


Figure 3.10. Observed soil moisture versus nonlinear regression (NLR, case 4), IEM, RBFNN and SVM simulated soil moisture for the 10 dates of 2009-2011 at the watershed scale. Each bar represents average soil moisture over the 9 study sites for a particular day.

# Chapter 4

## Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables\*

---

### 4.1 Introduction

The gross primary production (GPP) of an ecosystem represents the gross uptake of carbon dioxide (CO<sub>2</sub>) by vegetation for photosynthesis. It is the primary conduit of carbon flux from atmosphere to land and a key source of energy that fuels economies. On the other hand, CO<sub>2</sub> from fossil fuel burning and ecosystem respiration is a major contributor to global warming or greenhouse effect. Fossil fuel burning has perturbed the carbon cycle, and affected the global climate, leading to worldwide research on climate change and the carbon cycle (IPCC, 2007; Heinsch et al, 2006; Urbanski, et al., 2007). However, considerable uncertainties still remain regarding the dynamics of carbon fluxes over both short and long timescales and effective strategies are necessary to acquire relevant information about carbon flux processes, to locate and to quantify terrestrial

---

\* A version of this paper has been published. Jahan, N. and Gan, T. Y. 2009. Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables. *Journal of Geophysical Research*, Vol.114, No. G04026.

sources and sinks of carbon (Rahman et al, 2005). As GPP is a measure of carbon uptake by vegetation, an improved knowledge about GPP can provide us a useful measure of the health of ecosystem and the global carbon cycle.

Estimating GPP of terrestrial ecosystems has been challenging because of its dependence on a variety of environmental factors (Makela et al., 2008). Among the existing methods, the light use efficiency (LUE) model proposed by Monteith (1972) had been widely used (e.g., Potter et al., 1993; Landsberg and Waring, 1997; Coops et al., 2005; Running et al., 2000; Xiao et al., 2004; Yuan et al., 2007) to simulate the spatial and temporal dynamics of GPP because of its theoretical basis and practicality (Running et al., 2000). LUE is defined as the amount of carbon uptake per unit of absorbed photosynthetically active radiation (APAR) by photosynthetic biomass. In LUE, it is assumed that (1) the ecosystem GPP is directly related to amount of APAR and (2) the actual LUE may be lesser than its theoretical value because of environmental stresses such as low temperatures or water deficits (Yuan et al., 2007). The general form of LUE is:

$$\text{GPP} = \varepsilon \times \text{fPAR} \times \text{PAR} \quad (4.1)$$

$$\varepsilon = \varepsilon_{\max} \times f \quad (4.2)$$

Where PAR is the incident photosynthetically active radiation ( $\text{MJ m}^{-2}$ ) per unit time, fPAR is the fraction of incident PAR absorbed by the canopy,  $\varepsilon_{\max}$  is the potential LUE ( $\text{g C m}^{-2} \text{MJ}^{-1} \text{APAR}$ ) without environment stress,  $f$  is a scalar ranging from 0 to 1 representing the reduction of potential LUE under

environmental stresses,  $fPAR \times PAR$  gives the APAR and  $\varepsilon_{\max} \times f$  gives the realized LUE ( $\varepsilon$ ).

In recent years, carbon fluxes measured by the eddy covariance (EC) tower sites set up over forest, grasslands, savannas, etc., have provided useful field measurements for us to parameterize and to validate GPP models. Furthermore, it has been shown that combining these EC tower measurements with remotely sensed (RS) data has the potential to enhance modeling of GPP based on LUE. The MODIS-GPP Algorithm (Running et al., 2004), Vegetation Photosynthesis Model (Xiao et al., 2004), EC-LUE (Yuan et al., 2007), etc., are some examples of successful application of RS data in GPP modeling. The objective of this study is to investigate the applicability of several RS variables in GPP modeling, and to develop a solely RS data based GPP prediction model which does not depend on any supplementary meteorological data.

## **4.2 Review of Gross Primary Production (GPP) models**

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the Terra and Aqua satellites, provides GPP product (MOD 17) using the LUE method and inputs from the MODIS LAI/*fPAR* (MOD15A2) product, land cover, and biome-specific climatologic data from NASA's Data Assimilation Office (DAO). In this model, the light use efficiency ( $\varepsilon$ ) is calculated as

$$\varepsilon = \varepsilon_{\max} \times m(T_{\min}) \times m(VPD) \quad (4.3)$$

Where  $m(T_{min})$  and  $m(VPD)$  are multipliers that reduce  $\epsilon_{max}$  when cold temperatures and high vapor pressure deficit (VPD), respectively, limit photosynthesis. These factors range linearly from 0 to 1 where 1 denotes no inhibition and 0 denotes total inhibition. Values of  $\epsilon_{max}$ ,  $m(T_{min})$ ,  $m(VPD)$  are listed in the Biome Properties Look-Up Table (BPLUT). By comparing the MODIS GPP product with EC tower estimated GPP across a range of biomes, Heinsch et al. (2006) identified three potential sources of errors: 1) errors in meteorological input data derived from NASA's Goddard Earth Observing System (GEOS-4) climate model, 2) errors in MODIS LAI/fPAR product, and 3) errors in the land cover classification.

The Vegetation Photosynthesis Model (VPM) is another LUE model developed by Xiao et al. (2004) to estimate GPP using PAR, Enhance Vegetation Index (EVI), Land surface water index (LSWI) and coarse resolution temperature data according to Equation 4.1. In VPM  $\epsilon$  is computed as:

$$\epsilon = \epsilon_{max} \times T_{scalar} \times W_{scalar} \times P_{scalar} \quad (4.4)$$

Where  $T_{scalar}$ ,  $W_{scalar}$ ,  $P_{scalar}$  are scalars to account for the effects of temperature, water and leaf age on  $\epsilon_{max}$ , respectively. For this GPP model, it is critical to measure PAR accurately at large spatial scale because PAR is highly variable spatially.

Sims et al. (2006) developed a simple model with EVI as the only predictor, but it estimated GPP that were as good or even better than the MODIS GPP product for

some sites during periods of active photosynthesis. However, this model gave poor GPP estimates for sites subjected to summer drought or sites dominated by evergreen vegetation. Sims et al. (2008) improved this model (Temperature and Greenness (TG) Model) by adding an additional predictor, Land Surface Temperature (LST). This model computes GPP (Equation 4.5) for a 16-day period.

$$\text{GPP} = \text{scaledLST} \times \text{scaledEVI} \times m \quad (4.5)$$

Where,  $m$  is a scalar,

$$\text{scaledLST} = \min \left[ \left( \frac{\text{LST}}{30} \right); (2.5 - (0.05 \times \text{LST})) \right] \quad (4.6)$$

$$\text{scaledEVI} = \text{EVI} - 0.1 \quad (4.7)$$

The Global Production Efficiency Model proposed by Prince and Goward (1995) also uses APAR to calculate the global GPP using Equation 4.1 but its  $\varepsilon$  is based on

$$\varepsilon = \varepsilon_{\max} \times T_s \times SM \times \text{VPD} \quad (4.8)$$

Where  $T_s$  is the soil temperature and  $SM$  the soil moisture index.

The C-Fix model of Veroustraete et al. (2002), driven by temperature, radiation and fPAR, assumes that  $\varepsilon = \varepsilon_{\max}$  which is a fixed value ( $1.1 \text{ g C m}^{-2} \text{ MJ}^{-1} \text{ APAR}$ ) for calculating GPP (Equation 4.9) while others (e.g., Yuan et al., 2007) suggest reducing  $\varepsilon_{\max}$  to  $\varepsilon$  under limiting environmental conditions.

$$\text{GPP} = p(T_{\text{am}}) \times CO_2 \text{ fert} \times \varepsilon \times \text{fPAR} \times c \times S_{g,d} \quad (4.9)$$

Where  $p(T_{am})$  is the normalized temperature dependency factor (value =0 to 1),  $CO_2fert$  the normalized CO<sub>2</sub> fertilization factor (value =1, for no fertilization and value >1, for fertilization),  $c$  the climate efficiency factor (= 0.48) and  $S_{g,d}$  the daily incoming global solar radiation (MJ/m<sup>2</sup>/day).

### 4.3 Research objectives

Many of the aforementioned GPP models use BPLUT for LUEs and coarse resolution (e.g., 1° latitude by 1.25° longitude) meteorological inputs which may contain errors and lead to erroneous GPP estimates (Turner et al., 2005; Zhao et al., 2005; Heinsch et al., 2006). Therefore the objectives of this study are:

(1) To develop a GPP model called the “Remotely Sensed-GPP” model (the R-GPP model) without relying on coarse resolution meteorological variables, but only on four RS variables - two radiation budget variables (Albedo<sub>NIR</sub> and LST) and two ecosystem variables (Global Vegetation Moisture Index (GVMI) and EVI); and

(2) To assess the transferability of the proposed R-GPP model and its potential to map carbon fluxes of other deciduous forests.

If a dependable GPP model solely relying on RS data can be developed, it may be possible to estimate GPP accurately at global scale with a spatial resolution the same as that of the satellite data, which for MODIS is 1 km.

## 4.4 Study sites and data sets

### 4.4.1 Study sites

In this study, GPP estimated from three EC towers located in three deciduous forests were used:

(1) The Harvard Forest EC tower (42.54 °N, 72.17 °W) within the Harvard Forest, Massachusetts, USA, is part of the Ameriflux network and is one of the longest running tower sites in the world since 1991 (Goulden et al., 1996; Urbanski et al., 2007). The site primarily consists of 60- to-80-year-old deciduous broadleaf forest dominated by red oak, red maple, black birch, white pine and hemlock (Goulden et al., 1996). The climate of this forest is temperate, with warm humid summers and annual mean temperature of about 7.9 °C, annual precipitation of about 1066 mm, and an average annual plant growing season of about 161 days (Waring et al., 1995);

(2) The Morgan Monroe State forest (MMSF) EC tower (39.32 °N, 86.41°W) of Indiana, USA consists of 60 to 90 years old mixed hard wood forest and is dominated by sugar maple, tulip poplar, white oak and black oak. Its mean annual temperature is 11.1°C, and mean annual precipitation is 1012 mm (Curtis et al., 2002); and

(3) The University of Michigan Biological Station (UMBS) tower (45.56 °N, 84.7°W) of Michigan, USA is dominated by 90 years old deciduous forest. Other species are mid-aged conifer, northern hardwood, pine understay, aspen and hemlock. Its mean annual temperature is 6.2°C, and its mean annual precipitation is 750 mm (Curtis et al., 2002).

#### 4.4.2 Site specific carbon flux and climate data

All the carbon flux data used in this study are mean of 8-day period. EC towers do not measure GPP directly but they measure CO<sub>2</sub> exchange between vegetation and the atmosphere in terms of net ecosystem exchange (NEE) using the eddy covariance technique (Goulden et al., 1996). Then GPP is calculated from the daytime NEE ( $NEE_d$ ) and daytime ecosystem respiration ( $R_d$ ) by:

$$GPP = R_d - NEE_d \quad (4.10)$$

$R_d$  is usually estimated from daytime temperature and a temperature-respiration relationship usually developed from nighttime NEE measurements that represent nighttime respiration (autotrophic and heterotrophic) because plants do not photosynthesize at night.

#### 4.4.3 Remotely sensed data

Among the 36 spectral bands of MODIS, with spatial resolution ranging from 250 m to 1 km (Justice et al., 1998), seven spectral bands are primarily designed for the study of vegetation and land surface: blue (459–479 nm), green (545–565 nm), red (620–670 nm), near infrared (NIR) (841–875 nm, 1230–1250 nm) and shortwave infrared (1628–1652 nm, 2105–2155 nm). MODIS daily surface reflectances are radiometrically calibrated, cloud-filtered, atmospherically corrected for molecular scattering, ozone absorption and aerosols, spatially and temporally gridded and adjusted for view angle influences. For the three study sites, the 8-day surface reflectance data (MOD09A1, Collection 5) of the four spectral bands, blue, red, NIR (841–875 nm), and shortwave infrared (1628–1652

nm) were collected for 2000 to 2005 and then used to calculate vegetation indices, EVI and GVMl.

The other 8-day composite MODIS data sets used in this study include the 1 km LST (MOD11A2, collection 5) and 1 km GPP product (MOD17A2, Collection 5) described in Section 4.2. MOD11A2 is retrieved using the Split-Window algorithm and the thermal infra-red bands of MODIS (Wan and Dozier, 1996). We also collected MODIS albedo product which is produced every 8 days with 16 days of acquisition. The Bidirectional Reflectance Distribution Function (BRDF) coefficients from MCD43A1 were used to calculate the actual albedo for the visible (VIS), NIR and shortwave bands (0.3– 0.7, 0.7–5.0, and 0.3–5.0  $\mu\text{m}$ , respectively) as a function of optical depth, solar zenith angle, band (Schaaf et al., 2002; Lucht et al., 2000; <http://daac.ornl.gov/MODIS/MODIS-menu/MCD43.html>).

We estimated GPP with a 1 km resolution which is the same as that of MODIS GPP. In other words, we used a 1 km  $\times$  1 km area within which the EC tower is located to calibrate the R-GPP model instead of the more common approach of using RS data averaged over of  $N \times N$  km area ( $N=3$  or 5, or even larger number) because the footprint of the EC tower, which depends on the height of flux tower, wind speed, topography, etc., is usually few hundred meters to 1 km (Schmid, 2002; Xiao et al., 2004). Therefore using predictors averaged over areas of  $N \times N$  km in size may be too coarse to represent a tower footprint and may cause a scale

mismatch problem between simulated and tower GPP. From this perspective, using RS data at  $1\text{ km} \times 1\text{ km}$  resolution is a better strategy as long as MODIS data have been properly corrected geometrically. Since the RS data are of 1 km (LST, MODIS GPP) and 500 m (reflectance and Albedo) resolutions, for LST and MODIS GPP, we extracted digital values of a 1-km pixel within which the EC tower is located; while for reflectance and albedo, we used the average value of  $2 \times 2$  pixels which represents the same  $1\text{ km} \times 1\text{ km}$  area.

#### **4.5 Research methodology**

The research approach undertaken in this study can be summarized as follows:

- (1) Selecting EVI, GVMI,  $\text{Albedo}_{\text{NIR}}$  and LST as the model predictors and investigating the relationships between these model predictors and few environmental variables such as air temperature, PAR, VPD, which have been widely used to account for the environmental stresses on GPP;
- (2) Calibrating and validating the R-GPP model using the Harvard forest tower estimated GPP data of year 2000-2003 and 2004-2005, respectively, and comparing its results with the MODIS GPP product (MODIS-17).
- (3) Testing the transferability of the R-GPP model calibrated for the Harvard forest to two other deciduous forests sites - Morgan Monroe State forest and University of Michigan Biological station of USA.

## 4.5.1 R-GPP model predictors

### 4.5.1.1 Global Vegetation Moisture Index (GVMI)

Previous studies have demonstrated the possibility of using NIR and short wave infra-red bands to retrieve leaf and canopy water content ( $\text{g/m}^2$ ) using Landsat-TM data (Hunt and Rock, 1989), hyperspectral data (Gao, 1996; Serrano et al., 2000), and VEGETATION (VGT) sensor data (Ceccato et al., 2001). Recently, Ceccato et al. (2002 a, b) proposed to retrieve equivalent water thickness (EWT) at the canopy level using GVMI from the VGT sensor:

$$\text{GVMI} = \frac{(\text{NIR} + 0.1) - (\text{SWIR} + 0.02)}{(\text{NIR} + 0.1) + (\text{SWIR} + 0.02)} \quad (4.11)$$

Where NIR and SWIR are reflectance of the rectified NIR band and short wave infrared bands, respectively. Ceccato et al. (2002b) tested GVMI in retrieving EWT from four different ecosystems and found that water content retrieved from GVMI was consistent with field measured water content. Other studies also demonstrated the applicability of GVMI in retrieving EWT (e.g., Danson and Bowyer, 2004; Du et al., 2005).

To incorporate the effect of water stress in the R-GPP model, we used GVMI computed from MODIS reflectance products. Although GVMI is not correlated with the GPP of Harvard forest throughout the year (Table 4.1), they are significantly correlated ( $R^2=0.60$ ) during the growing season (mid April to 27 October (Urbanski et al., 2007)) of this deciduous study site. However, in early spring, late fall and winter, high values of GVMI could still be observed because of snow cover above or below the canopy (Figure 4.1a). Therefore during these

cold periods, GPP is not related to GVMI but is probably controlled by the LST which will be explained later (Section 4.5.1.4).

Figure 4.1a (right panel) shows that GPP increases with GVMI (related to soil moisture) during the growing season. However, Figure 4.1a (left panel) also shows that when GVMI is around 0.4 to 0.5, GPP fluctuates widely from 5 to 14 showing little relation to GVMI because when there is sufficient soil moisture (water is not a limiting factor), photosynthesis will probably depend more on temperature which is related to the incoming solar radiation. Yuan et al. (2007) also reported that GPP is controlled either by air temperature or by soil moisture, whichever is the most limiting.

#### 4.5.1.2 Near-infrared Albedo (Albedo<sub>NIR</sub>)

Albedo ( $\alpha$ ), the fraction of incident solar radiation reflected by a surface (Equation 4.12), plays a key role in partitioning the total radiative flux into absorbed, sensible, latent, and reflected fluxes (Bounoua et al., 2000). The net radiation  $R_n$  is given as

$$\begin{aligned} R_n &= S_{in} - S_{out} + L_{in} - L_{out} \\ &= S_{in}(1 - \alpha) + L_{in} - L_{out} \end{aligned} \quad (4.12)$$

Where  $S_{in}$  and  $S_{out}$  are the incoming and outgoing solar (shortwave) radiation;  $L_{in}$  and  $L_{out}$  are the downwelling and upwelling longwave radiation at the surface, respectively.

Albedo influences the radiation absorbed by plant canopies and thereby affects physical and bio-geochemical processes such as photosynthesis, energy balance, evapotranspiration, and respiration (Wang et al., 2001, 2002a, b). Furthermore, surface albedo also affects rainfall, vegetation growth (e.g., Bounoua et al., 2000; Laval and Picon 1986; Wang and Davidson, 2007) and even droughts that could lead to desertification (Dirmeyer and Shukla, 1996; Knorr et al., 2001). The albedo of vegetation, unlike that of bare soil, shows temporal variability due to the seasonal behavior of plant phenology such as green-up, peak greenness, dry-down, and senescence. For example, Song (1998) found that the albedo of a wheat field decreased from the peak green to senescence stage. Although some previous studies on GPP (e.g., Ichii et al., 2003; Landsburg and Waring, 1997; Gebremichael and Barros, 2006; Kimball et al. 1997) used albedo to calculate radiative fluxes, as far we know, none of them reported a direct relationship between NIR albedo and GPP, and most of these models used a constant albedo without considering its temporal variability.

In this study, albedo at the NIR band,  $\text{Albedo}_{\text{NIR}}$  (0.7 to 5  $\mu\text{m}$ ) has been used because the reflectance of vegetation is very strong at NIR band and likely because of this reason, it is the most commonly used albedo in ecosystem modeling (Wang and Davidson, 2007; Ghulam et al., 2007; Ollinger et al., 2008). Since only 16-day resolution albedo data is available from MODIS, we have used that 16-day albedo product produced every 8 days (e.g., albedo of Date 1 corresponds to average albedo of day 1 to 16 while albedo of date 9 corresponds

to average albedo of Date 9 to 24). To estimate the GPP of any 8 day period we have used  $\text{Albedo}_{\text{NIR}}$  averaged over that particular 8-day and the previous 8-day while the other predictors were averaged over that particular 8-day. For example, to calculate the average GPP of day 9 to 16 (17 to 24), we have used the average albedo of day 1 to 16 (9 to 24) while the other predictors were averages of day 9 to 16 (17 to 24). Therefore the R-GPP remains as an 8-day GPP model.

Table 4.1 and Figure 4.1 (b) show that the seasonal dynamics of  $\text{Albedo}_{\text{NIR}}$  and GPP are strongly correlated with each other for the Harvard Forest site with a  $R^2=0.82$  for the 2000-2005 data which indicates that using only  $\text{Albedo}_{\text{NIR}}$ , GPP may be modeled with comparable or better accuracy than the GPP estimates from MODIS ( $R^2=0.78$  for 2000-2005) for this site.  $\text{Albedo}_{\text{NIR}}$  gradually increases with the green-up of deciduous forest because of the high reflectance of canopy leaves in the NIR band and continues until the peak green stage and then gradually decreases with the senescence of leaves (Figure 4.1 b, right panel), as was also observed by Wang (2005) for a boreal deciduous forest of Saskatchewan, Canada.

#### **4.5.1.3 Enhanced Vegetation Index (EVI)**

EVI produces vegetation signal with improved vegetation monitoring through canopy background and atmospheric corrections (Waring et al., 2006). It is more sensitive than the popular normalized difference vegetation index (NDVI) in high biomass regions.

$$\text{EVI} = G \frac{\text{NIR} - \text{R}}{\text{NIR} + C_1 \text{R} - C_2 \text{B} + L} \quad (4.13)$$

where NIR, R and B are atmospherically corrected surface reflectance in the near-infrared, red and blue bands respectively,  $G$  is the gain factor,  $L$  is the canopy background adjustment factor that addresses nonlinear, differential NIR and red radiant transfer through a canopy; and  $C_1$  and  $C_2$  are the coefficients of the aerosol resistance term, which uses the blue band to correct the aerosol influences in the red band. In the EVI algorithm,  $L=1$ ,  $C_1=6$ ,  $C_2 = 7.5$ , and  $G = 2.5$ . EVI has been shown to be a good predictor of growing season GPP for many sites and it was used as a predictor in some previous models (Xiao et al., 2004). In this study we found that the seasonal dynamics of GPP agrees reasonably well with EVI ( $R^2 = 0.84$ ) for the Harvard forest (Figure 4.1 c) and so EVI was selected as a predictor.

#### 4.5.1.4 Land Surface Temperature (LST)

LST is a potential predictor for GPP estimation because it can incorporate the effect of temperature and VPD on vegetation (Hashimoto et al., 2008). It is highly correlated with vegetation dynamics (Sun and Kafatos, 2007) and is positively correlated with NDVI in high latitudes (Karnieli et al., 2006). Boegh et al. (1998) found the slope of LST/NDVI to be related to the evapotranspiration of Sahel.

The scatter plot of GPP with LST (Figure 4.2 a) shows that below 0 °C, there is no photosynthesis while above 0 °C, GPP slowly increases with LST, which implies that 0 °C can be used as a temperature threshold for this deciduous forest

to define periods of active photosynthesis. Some studies (Sims et al., 2008; Yuan et al., 2007) reported that photosynthesis is predominantly controlled by temperature only at the beginning and the end of a growing season, but by moisture conditions throughout the growing season. Therefore we used a scaled LST ( $LST_s$ ) (Equation 4.14) to set GPP to zero when LST is below 0 °C.

$$LST_s = \begin{cases} \frac{LST}{LST_{max}}; & \text{when } LST > 0^\circ \text{ C} \\ 0; & \text{when } LST \leq 0^\circ \text{ C} \end{cases} \quad (4.14)$$

Where LST is the observed LST and  $LST_{max}$  is the maximum LST. In this study  $LST_{max}$  is set to 30°C partly because it has been used as the optimum LST in some other studies (Sims et al., 2008). Figure 4.2 (b) shows that GPP is strongly correlated with  $LST_s$  ( $R^2=0.71$ ). Throughout the summer, GPP increases with increasing  $LST_s$ ; but as the season enters into fall,  $LST_s$  decreases and GPP drops because the deciduous forest slowly drops its leaves (Figure 4.2 c). However, it is also found that GPP does not respond instantaneously with temperature rise during the early growing season (at low  $LST_s$ ) which is attributed to the lag in the leaf development of deciduous forest in the spring (Figure 4.2 b, c). Furthermore, low  $LST_s$  during the start and the end of each growing season restricts water and nutrient uptake and hence it affects photosynthesis (Sims, et al., 2008).

#### **4.5.2 Relationships between model predictors and other environmental variables**

The relationships between the R-GPP model predictors (EVI,  $LST_s$ , GVMI and  $Albedo_{NIR}$ ) and some environmental variables (PAR, VPD, air temperature)

measured at the EC tower site of Harvard Forest were examined. These environmental variables have been popular predictors of carbon flux but they vary substantially over space and usually they are only available as limited ground measurements or coarse resolution, gridded data. Therefore, if we can establish meaningful relationships between PAR, air temperature, VPD and aforementioned RS predictors, it will be possible to get a continuous estimation of carbon fluxes on the basis of RS predictors as they are acquired on a continuous basis.

PAR was found to be reasonably correlated with the R-GPP model predictors for 2000 to 2005 (Table 4.2, Figure 4.3). As GVMI influences photosynthesis only in the active photosynthesis period, and during non-active period (winter) it is affected by snow cover above or below the canopy, the seasonal cycle of GVMI and PAR did not match during winter and the overall correlation between PAR and GVMI was relatively poor (Table 4.2). However,  $LST_s$ , EVI, and  $Albedo_{NIR}$  followed the seasonal variation of PAR quite systematically, which imply that these variables can possibly replace PAR which is one of the most critical predictors in the estimation of GPP (Xiao et al., 2004).

EVI,  $Albedo_{NIR}$  and LST were also found to be consistently correlated with air temperature ( $T_{air}$ ) (Figure 4.4) and VPD (Figure 4.5) which have been key predictors in many GPP models (Yuan et al., 2007). Since the correlation between  $T_{air}$  and  $LST_s$  is very high ( $R^2 = 0.92$ ), it likely means that  $LST_s$  can replace  $T_{air}$ . Moreover, using  $LST_s$  instead of  $T_{air}$  will allow us to use data of fine

spatial resolution instead of limited ground measured data from climate stations or coarse resolution data simulated from global climate models (e.g., NASA's Data Assimilation Office GEOS-4 global climate model). The high values of GVMI because of snowcover effects during winter weakened the overall correlation (Table 4.2) of GVMI with  $T_{\text{air}}$  and VPD (Figure 4.5 a). Given that GPP is controlled by LST during winter, the poor correlation of GVMI with GPP and other environmental variables during winter may only have marginal effect on the winter GPP simulated by the R-GPP model.

#### 4.6 R-GPP model development and results

Given that GVMI, EVI,  $\text{Albedo}_{\text{NIR}}$  and LSTs are correlated to GPP and to PAR, VPD and air temperature which are key elements of many GPP models, we propose a Remotely Sensed GPP (R-GPP) model (Equation 4.15) based on these four RS predictors only,

$$\text{GPP} = k \times \text{GVMI}^a \times \text{LST}_s^b \times \text{Albedo}_{\text{NIR}}^c \times \text{EVI}^d \quad (4.15)$$

Where  $k$  is a scalar, and  $a$ ,  $b$ ,  $c$ , and  $d$  are exponents. These model parameters were estimated using the estimated GPP of 2000 to 2003 from the EC tower located at the Harvard forest site and a nonlinear optimization scheme, the Generalized Reduced Gradient (GRG2) (Lasdon, 1978; Spaulding, 1998). By GRG2, the optimized values of  $k$ ,  $a$ ,  $b$ ,  $c$ ,  $d$  have been found to be 114, 0.885, 1.05, 0.695 and 0.933, respectively.

The performance of R-GPP was evaluated in terms of the coefficient of determination ( $R^2$ ) and root mean squared error (RMSE). The calibration results show that R-GPP model could capture the seasonal dynamics of the observed GPP accurately (Figure 4.6, 4.7 a (left panel), Table 4.3). With respect to the EC tower estimated GPP, it is clear that R-GPP ( $R^2=0.95$ , RMSE= 1.02  $\mu\text{mol C/ m}^2/\text{sec}$ ) was more efficient than the MODIS GPP algorithm ( $R^2 = 0.80$ , RMSE=2.78  $\mu\text{mol C/ m}^2/\text{sec}$ ) for the Harvard Forest site, especially during the peak growing season. In almost all years tested, the MODIS algorithm showed a marginal over-estimation in the early part of the plant growing season and an under-estimation in the peak growing season (June to September) (Figure 4.7 b, left panel), as was also reported by Xiao et al. (2004) for the Harvard Forest site. The poor estimate of MODIS-GPP arises likely from uncertainties related to meteorological inputs, and erroneous land cover classification and LAI/fPAR product (Heinsch et al., 2006) used in the MODIS algorithm (as explained in Section 4.2). Moreover,  $\varepsilon_{\max}$  used in the MODIS GPP algorithm, which is biome specific, has been found to be smaller than the  $\varepsilon_{\max}$  value observed at the Harvard Forest (Turner et al., 2003). Using an underestimated  $\varepsilon_{\max}$  may cause an underestimated GPP.

The calibrated R-GPP model ( $R^2=0.92$ , RMSE=1.62  $\mu\text{mol C/ m}^2/\text{sec}$ ) also outperformed the MODIS GPP algorithm ( $R^2=0.77$ , RMSE=3.35  $\mu\text{mol C/ m}^2/\text{sec}$ ) in the validation stage (2004-2005) of the Harvard forest. The MODIS GPP algorithm under predicted the EC tower GPP quite substantially during the growing season (Figure 4.7 b, right panel) of 2004-2005 whereas the R-GPP

model's prediction was relatively close to the EC tower GPP (Figure 4.7 a, right panel).

Even though the overall simulated GPP of the R-GPP model matched closely to the observed, occasionally there were large discrepancies between them (especially in the validation stage) partly because of the limitations of R-GPP and possibly because of the error in the observed GPP estimated from  $NEE_d$  and  $R_d$  which are subjected to uncertainties (Xiao et al., 2004). Moreover there are gaps in both NEE and  $R_d$  data and gap-filling steps of these data are still subjected to debates (Falge et al., 2001). Therefore, even though the gap-filled 'estimated GPP' data can be used to assess the performance of GPP models reasonably accurately, some uncertainties are expected.

#### **4.7 Transferability of R-GPP model**

Given that EC towers are established only in limited sites, it will be useful to examine the transferability of the proposed R-GPP model calibrated for the Harvard forest site to other deciduous forests located in northeastern USA to estimate their carbon fluxes. Intuitively, the degree of transferability will depend on the degree of similarity in terms of vegetation types, climate regimes and how accurately the four predictors ( $GVMI$ ,  $Albedo_{NIR}$ ,  $EVI$ ,  $LST_s$ ) measure the basic environmental properties such as moisture condition, reflectivity and surface temperature. Gilmanov et al. (2005) argued that models based on vegetation indices such as NDVI (and presumably EVI) are transferable as long as vegetation

types and age are comparable between the sites since these are two important factors to be considered in estimating GPP (Desai et al., 2008). To test its transferability, the R-GPP model developed out of the Harvard forest site was applied to UMBS and MMSF forest sites which are also mature deciduous forests with stand age 60 to 90 years.

Figure 4.8 (a, left panel) shows that the R-GPP model developed for the Harvard forest simulated the observed GPP of MMSF more accurately ( $R^2=0.93$ , RMSE=1.47  $\mu\text{mol C/ m}^2/\text{ sec}$ ) than the MODIS GPP algorithm ( $R^2=0.74$ , RMSE=3.54  $\mu\text{mol C/ m}^2/\text{ sec}$ ). The MODIS GPP algorithm consistently showed an underestimation in the peak growing season for most of the years (2000-2005) and an overestimation in the early growing seasons for some years (Figure 4.8 b, left panel). In contrast, the R-GPP model only showed minor over-estimation during the peak growing season of 2000-2002 and marginal underestimation in 2004.

For the UMBS site, the R-GPP model's predicted GPP also followed the seasonal dynamics of the observed GPP very well (Figure 4.8 a, right panel) and the agreement is marginally better than the GPP estimated by the MODIS algorithm, e.g.,  $R^2=0.94$  versus  $R^2=0.91$  (Table 4.3). On a whole, for UMBS, the MODIS GPP product showed better estimation during the early growing season and peak season (Figure 4.8 b, right panel) than it did for the other two sites.

Given that both UMBS and MMSF test sites are located about more than 1000 km away from the Harvard forest, it seems that the proposed R-GPP model using the four selected RS predictors can generally estimate the GPP of deciduous forests located at the north-eastern USA. Even though simple in nature and built on the basis of the RS data only, the R-GPP model possesses the necessary physical basis to capture the basic ecological and environmental functioning of deciduous ecosystems, which is probably why R-GPP turns out to be more effective than the MODIS GPP in characterizing the seasonal variability of GPP of three deciduous ecosystems of north-eastern USA.

#### **4.8 Discussions**

In recent years RS data based models have demonstrated strong potential in GPP modeling, e.g, MODIS GPP algorithm (Running et al., 2004), TG model (Sims et al., 2008), VPM (Xiao et al., 2004), MOD-SIM-Cycle (Hazarika et al., 2005), EC-LUE (Yuan et al., 2007), etc. Our proposed R-GPP model is different from other GPP models because it is dependent solely on RS data whereas majority of the RS based models (e.g. MODIS GPP) requires supplementary meteorological inputs which are often available with spatial resolutions poorer than the RS variables, and as a result may produce significant errors in regional scale GPP estimation (Heinsch et al., 2006). Our proposed R-GPP model is likely closest to the TG model of Sims et al. (2008) which is also solely RS data based. However, the R-GPP model estimates 8-day means of GPP while the TG model computes 16-day means of GPP using EVI and LST, as described before.

Although the overall performance of the R-GPP model was encouraging, there are still discrepancies when compared with field observations, especially for the Harvard forest during the growing season of the validation period. The surface reflectance products from 8-day composite images are likely a key factor that may affect the accuracy of GPP predicted by the R-GPP model. The compositing method (e.g., currently MODIS reflectance data is composited on the basis of a minimum-blue criterion that selects the clearest conditions over the period) could result in some bias so that EVI and GVMi computed from the reflectance products may not reflect the average condition of that 8-day period (Xiao et al., 2004). Therefore GPP estimated by the R-GPP model may differ from the observed 8-day mean GPP. This problem can be partly resolved by using daily images as input to the R-GPP model, although this would incur large increases in computer processing.

Another factor that may affect the results of the R-GPP model is the 16-day albedo of MODIS. To estimate the GPP of any 8-day period, the R-GPP model uses  $\text{Albedo}_{\text{NIR}}$  averaged over that particular 8-day and the previous 8-day (Section 4.5.1.2). This averaging may introduce some discrepancies. Moreover, the nutrition limit is not explicitly considered in the R-GPP model, which partly contributed to the discrepancies between the R-GPP model output and that estimated from the EC tower.

This study demonstrated that combining indices such as GVMI, EVI, Albedo<sub>NIR</sub>, etc., in a meaningful manner can capture the temporal dynamics of photosynthetic activities of deciduous ecosystems in north-eastern USA. GVMI and EVI enabled us to account for the soil moisture state and the overall status of vegetation, while albedo and LST provided crucial information about the surface energy necessary for plant growth (Huete, 2005). This study has demonstrated the applicability of these predictors and their quantitative relationships with GPP. It may be useful to examine other vegetation indices (e.g., the normalized difference water index) to more comprehensively model the seasonal dynamics of GPP across different ecosystems.

At present about 400 EC tower sites are operating worldwide, under the FLUXNET network, on a continuous and long term basis to collect information on carbon, moisture and energy fluxes (<http://daac.ornl.gov/FLUXNET/>). These towers are located in different part of the world and belong to different climatic regime. Multi-year GPP data from EC towers located in various deciduous forests can be used to validate the R-GPP model. However, some of these flux data are not yet publicly available, because the analysis and publications of flux data are time consuming (Xiao et al., 2004). When data from many EC tower sites become publicly available, we will be in a better position to more comprehensively validate this R-GPP model, to better identify various sources of error and to fine tune the model.

The time scale of the R-GPP model is dictated by the temporal resolution of the MODIS data. Among the four model predictors, albedo is not available on a daily basis and so it is not possible to compute daily GPP using the R-GPP model which for now can only operate at an 8-day period. Some LUE based models (e.g., Yuan et al., 2007; Makela et al., 2008 etc.) can be used to model daily GPP variations, which, however, depend on meteorological data whose limitations have already been discussed in Section 4.2.

#### **4.9 Summary and conclusions**

We have developed a GPP estimation model, called the R-GPP model solely based on four remotely sensed (RS) variables, namely the EVI, near infrared albedo, GVMi and LST as model predictors. The model was calibrated (2000-2003) and validated (2004-2005) on the basis of GPP estimated from fluxes of an eddy covariance tower located in the Harvard Forest, USA. The summary of the results are listed below:

- (1) The proposed model predicted the GPP of the Harvard Forest accurately, with  $R^2=0.95$  and  $R^2=0.92$  in the calibration and the validation periods, respectively, which is much better than the MODIS-GPP algorithm ( $R^2=0.80$  and  $0.77$  in the calibration and validation stages, respectively) even though the latter is relatively complex and requires meteorological inputs which are mostly available in coarse resolution only.

(2) The model predictors individually showed strong correlation to the GPP of the Harvard forest for 2000-2005 ( $R^2=0.60, 0.84, 0.82, 0.71$  for GVMi (during growing season), EVI, Albedo<sub>NIR</sub> and LST<sub>s</sub>, respectively). Furthermore, for the Harvard forest, the Albedo<sub>NIR</sub> or the EVI itself could predict GPP marginally better than the MODIS GPP ( $R^2=0.78$  for 2000-2005). Therefore, the R-GPP model out-performed the MODIS-GPP algorithm since it is designed to take the advantage of the combined contributions of all these four RS predictors.

(3) The R-GPP model predictors, such as EVI, Albedo<sub>NIR</sub>, LST, have been shown to be correlated with few other environmental variables such as air temperature, PAR and VPD which have been widely used as predictors in modeling GPP. The relationships between them imply that the predictors of R-GPP model, which are available in relatively fine spatial resolutions, can replace meteorological predictors of coarse spatial resolutions in GPP modeling.

(4) The transferability of the R-GPP model, calibrated for the Harvard forest, was tested by applying it to two other deciduous forest sites, MMSF and UMBS. The R-GPP model captured the seasonal dynamics of the observed GPP of MMSF ( $R^2 = 0.93$ ) and UMBS ( $R^2 = 0.94$ ) more accurately than the MODIS GPP algorithm ( $R^2 = 0.74$  and  $0.91$  for MMSF and UMBS, respectively). Apparently the R-GPP model is transferable and can estimate the GPP of other similar deciduous forests, especially those that are located in north-eastern USA.

Although the proposed R-GPP model has shown promising results in estimating GPP of several deciduous forests of north-eastern USA, further validation is

needed to test the robustness of the R-GPP model and its applicability in different climatic and biophysical conditions. The model parameters may need to be refined for other climatic regimes and biomes. Further study is also needed to determine whether net primary production can be estimated from the RS variables used in this study.

## 4.10 References

- Berry, J. and Bjorkman, O. 1980. Photosynthetic response and adaptation to temperature in higher plants. *Annual Review of Plant Physiology*, 31, 491-543.
- Betts, A.K. and Ball, J.H. 1997. Albedo over the boreal forest. *Journal of Geophysical Research* 102, 28901- 28909.
- Boegh, E., Soegaard, H., Hanan, H., Kabat, P. and Lesch, L. 1998. A remote sensing study of the NDVI-Ts relationship and the transpiration from sparse vegetation in the Sahel based on high resolution data, *Remote Sensing of Environment*, 69, 224– 240.
- Bounoua, L., Collatz, G. J., Los, S. O., Sellers, P. J., Dazilich, D. A., Tucker C. J., and Randall, D. A. 2000. Sensitivity of climate to changes in NDVI, *J. of Climate*, 13, 2277-2292.
- Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud S. and Gregoire, J. M. 2001. Detecting vegetation leaf water content using reflectance in the optical domain, *Remote Sensing of Environment*, 77, 22– 33.
- Ceccato, P., Flasse, S. and Gregoire, J. M. 2002a. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications, *Remote Sensing of Environment*, 82, 198– 207.
- Ceccato, P., Gobron, N., Flasse, S., Pinty B. and Tarantola, S. 2002b. Designing a spectral index to estimate vegetation water content from remote sensing

- data: Part 1. Theoretical approach, *Remote Sensing of Environment*, 82, 188–197.
- Coops, N. C., Waring, R. H. and Law, B. E. 2005. Assessing the past and future distribution and productivity of ponderosa pine in the pacific northwest using a process model, 3-PG. *Ecological Modeling*, 183, 107–124.
- Curtis, P. S., Hanson, P.J., Bolstad, P., Barford, C., Randolph, J. C., Schmid H. P. and Wilson, K. B. 2002. Biometric and eddy-covariance based estimates of annual carbon storage in five eastern North American deciduous forests, *Agric. For. Meteorol.*, 113, 3-19.
- Danson, F. M. and Bowyer, P. 2004. Estimating live fuel moisture content from remotely sensed reflectance, *Remote Sensing of Environment*, 92, 30-321.
- Desai, A. R., Noormets, A., Bolstad, P. V., Chen, J., Cook, B. D., Davis, K. J., Euskirchen, E. S., Gough, C., Martin, J. G., Ricciuto, D. M., Schmid, H. P., Tang J. and Wang, W. 2008. Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: implications for regional scaling, *Agric. For. Meteorol.*, 148, 288-308.
- Dirmeyer, P. A. and Shukla, J. 1996. The effect on regional and global climate of expansion of the world's desert, *Quarterly Journal of the Royal Meteorological Society*, 122, 451–482.
- Du, X., Wang, S. and Zhou, Y. 2005. Monitoring and spatio-temporal evolution researching on vegetation leaf water in China, *IEEE Trans. Geosci. Remote Sens.*, 4, 2365-2368.

- Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, G., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K. and Wofsy, S. 2001. Gap filling strategies for long term energy flux data sets. *Agric. For. Meteorol.*, 107, 71– 77.
- Gao, B. C. 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, 58, 257– 266.
- Gebremichael, M. and Barros, E. P. 2006. Evaluation of MODIS gross primary productivity (GPP) in tropical monsoon region, *Remote Sensing of Environment*, 100, pp. 150-166.
- Ghulam, A., Qin, Q. and Zhan, Z. 2007. Designing of the perpendicular drought index. *Environmental Geology*, 52, 1045-1052.
- Gilmanov, T. G., Tieszen, L. L., Wylie, B. K., Flanagan, L. B., Frank, A. B., Haferkamp, M. R., Meyers, T. P. and Morgan, J. A. 2005. Integration of CO<sub>2</sub> and remotely sensed data for primary production and ecosystem respiration analyses in the northern Great Plains: Potential for quantitative spatial extrapolation, *Global Ecology and Biogeography*, 14, 271-292.

- Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C. and Woofsy, S. C. 1996. Measurements of carbon storage by long-term eddy correlation: Methods and a critical evaluation of accuracy, *Global Change Biol.*, 2, 169-182.
- Hashimoto, H., Dungan, J. L., White, M. A. Yang, F., Michaelis, A. R., Running, S. W. and Nemani, R. R. 2008. Satellite-based estimation of surface vapor pressure deficits using MODIS land surface temperature data, *Remote Sens. of Environ.*, 112, 142-155.
- Hazarika, M. K., Yasuoka, Y., Ito, A. and Dye, D. 2005. Estimation of net primary productivity by integrating remote sensing data with an ecosystem model, *Remote Sens. of Environ.*, 94, 298-310.
- Heinsch, F.A., Zhao, M., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J. Bolstad, P. V., Cook, B. D., Desai, A. R., Ricciuto, D. M., Law, B. E., Oechel, W. C., Kwon, H., Luo, H., Wofsy, S. C., Dunn, A. L., Munger, J. W., Baldocchi, D. D., Xu, L., Hollinger, D. Y., Richardson, A. D., Stoy, P. C., Siqueira, M.B.S., Monson, R. K. Burns, S.P. and Flanagan, L.B. 2006. Evaluation of remote sensing based terrestrial production from MODIS using AmeriFlux eddy tower flux network observations, *IEEE Trans. Geosci. Remote Sens.*, 44, 1908–1925.
- Huete, A. R. 2005 Global variability of terrestrial surface derived from MODIS visible to thermal-infrared channels, *Proceeding of Geoscience and Remote Sensing Symposium, 2005. IGARSS. 2005*, 7, 4938-4941.

- Hunt, E. R. and Rock, B. N. 1989. Detection of changes in leaf water content using near-infrared and middle-infrared reflectances, *Remote Sensing of Environment*, 30, 43– 54.
- Ichii, K., Matsui, Y., Murakami, K., Mukai, T., Yamguchi, Y. and Ogawa, K. 2003. A simple global carbon and energy coupled cycled model for global warming simulation: sensitivity to the light saturation effect. *Tellus*, 55, pp. 676-691.
- Intergovernmental Panel on Climate Change, 2007. *Climate Change 2007 - The Physical Science Basis, Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge.
- Karnieli, A., Bayasgalan, M., Bayasgalan, Y., Agam, N., Khudulmur, S. and Tucker, C. J. 2006. Comments on the use of the vegetation health index over Mongolia, *Int. J. Remote Sens.*, 27, 2017– 2024.
- Kimball, J. S., White, M. A. and Running, S. W. 1997. Biome BGC simulations of stand hydrologic processes for BOREAS, *J. Geophys. Res.*, 102, 2043-29051.
- Knorr, W., Schnitzler, K.G. and Govaerts, Y. 2001. The role of bright desert regions in shaping North African climate, *Geophys. Res. Lett.*, 28, 3489– 3492.
- Landsberg, J. J. and Waring, R. H. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning, *Forest Ecology Management*, 95, 209–228.

- Lasdon, L. S., Waren, A. D., Jain, A. and Ratner, M. 1978. Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming, *ACM Transactions on Mathematical Software*, 4, 34-50.
- Laval, K. and Picon, L. 1986. Effect of change of the surface albedo on the Sahel climate, *Journal of Atmospheric Sciences.*, 43, 2418–2429.
- Law, B. E., Waring, R. H., Anthoni, P. M. and Aber, J. D. 2000. Measurements of gross and net ecosystem productivity and water vapour exchange of a *Pinus ponderosa* ecosystem, and an evaluation of two generalized models, *Global Change Biol.*, 6, 155–168.
- Lucht, W., Schaaf, C.B. and Strahler, A. H. 2000. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Trans. Geosci. Remote Sens.*, 38, 977–998.
- Makela, A., Pulkkinen, M., Kolari, P., Lagergren, F., Lindroth, P. A. Loustau, D., Nikinmaa, E., Vesala, T. and Hari, P. 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting coniferous sites in Europe, *Global Change Biol.*, 14, pp. 92-108.
- Monteith, J. 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, 9, 747-766.
- Ollinger, S. V., Richardson, A. D., Martin, M. E., Hollinger, D. Y., Frolking, S. E., Reich, P. B., Plourde, L. C., Katul, G. G., Munger, J. W., Oren, R., Smith, M.-L., Paw, K. T., Bolstad, P. V., Cook, B. D., Day, M. C., Martin, T. A., Monson, R. K. and Schmid, H. P. 2008. Canopy nitrogen, carbon

- assimilation, and albedo in temperate and boreal forest: Functional relations and potential climate feedbacks, *PNAS*, 105, 19366-19341.
- Potter, C. B., Randerson, J. T., Field, C. B., Matson, P.A., Vitousek, P. M., Mooney, H. A. and Klooster, S. A. 1993. Terrestrial ecosystem production: a process model based on global satellite and surface data, *Global Biogeochem. Cycle*, 7, 811–841.
- Potter, C., Klooster, S., Huete, A. and Genovese, V. 2007. Terrestrial carbon sinks for the United States predicted from the MODIS satellite data and ecosystem modeling, *Earth Interactions*, 11, 1-21.
- Prince, S. D. and Goward, S. N. 1995. Global primary production: a remote sensing approach. *Journal of Biogeography*, 22, 815–835.
- Rahman, A. F., Sims, D. A., Cordova, V. D. and El-Masri, B. Z. 2005. Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes, *Geophys. Res. Lett.*, 32, L19404, doi:10.1029/2005GL024127.
- Running, S. W., Thornton, P. E., Nemani, R. and Glassy, J. M. 2000. Global terrestrial gross and net primary productivity from the earth observing system, in *Methods in Ecosystem Science*, edited by O.E. Sala, R. B. Jackson and H. A. Mooney, pp. 44–57 Springer-Verlag, New York.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves M. and Hashimoto, H. 2004. A continuous satellite-derived measure of global terrestrial primary production, *Bioscience*, 54, 547–560.

- Schaaf, C. B., Gao, F. Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J. P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entermont, R., Hu, B., Liang, S., Privette, J. and Roy, D. 2002. First operational BRDF, albedo nadir reflectance products from MODIS, *Remote Sensing of Environment*, 83, 135–148.
- Schmid, H. P., Su, H. B., Vogel, C. S. and Curtis, P. S. 2002. Footprint modeling for vegetation atmosphere exchange studies: a review and perspective, *Agric. For. Meteorol.*, 113, 159– 183.
- Serrano, L., Gamon, J. A. and Penueles, J. 2000. Estimation of canopy photosynthetic and nonphotosynthetic components from spectral transmittance, *Ecology*, 81, 3149– 3162.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Flanagan, L. B., Goldstein, A.H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schimd, H. P., Wofsy, S. C. and Xu, L. 2006. On the use of MODIS EVI to assess gross primary productivity of North American ecosystems, *Journal of Geophysical Research* 111, G04015.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Flanagan, L. B., Goldstein, A.H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schimd, H. P., Wofsy, S. C. and Xu, L. 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface

- temperature from MODIS, *Remote Sensing of Environment*, 112, 1633-1644.
- Song, J. 1998. Diurnal asymmetry in surface albedo, *Agricultural and Forest Meteorology*, 92, 181–189.
- Spaulding, K.A. 1998. Neural matamorphic optimization algorithms, Masters Thesis, University of Texas at Austin.
- Sun, D. and Kafatos, M. 2007. Note on the NDVI-LST relationship and the use of temperature related drought indices over north America, *Geophys. Res. Lett.*, 34, L24406.
- Toomey, M., and Vierling, L. A. 2005. Multispectral remote sensing of landscape level foliar moisture: techniques and applications for forest ecosystem monitoring, *Canadian Journal of Forest Research*, 35, pp 1087–1097.
- Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Zhao, M., Running, S. W., Wofsy, S. C., Urbanski, S., Dunn, A. L. and Munger, J. W. 2003. Scaling gross primary production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation, *Remote Sensing of Environment*, 88, 256-270.
- Turner, D. P., Ritts, W. D., Cohen, W. B., Maeirsperger, T. K., Gower, S. T., Kirschbaum, A. A., Running, S. W., Zhao, M., Wofsy, S. C., Dunn, A. L., Law, B. E., Campbell, J. L., Oechel, W. C., Kwon, H. J., Meyers, T. P., Small, E. E., Kurc, S. A. and Gamon, J. A. 2005. Site-level evaluation of satellite-based global terrestrial GPP and NPP monitoring, *Global Change Biol.*, 11(4), 666–684.

- Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M. and Munger, J. W. 2007. Factors controlling CO<sub>2</sub> Exchange on time scales from hourly to decadal at Harvard Forest, *J. Geophys. Res.*, 112, G02020, doi:10.1029/2006JG000293.
- Veroustraete, F., Sabbe, H. and Eerens, H. 2002. Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data, *Remote Sensing of Environment*, 83, 376–399.
- Wan, Z., and Dozier, J. 1996. A generalized split-window algorithm for retrieving landsurface temperature from space. *IEEE Trans. Geosci. Remote Sens.*, 34, 892–905.
- Wang, S., Grant, R. F., Versegny, D. L. and Black, T. A. 2001. Modelling plant carbon and nitrogen dynamics of a boreal aspen forest in CLASS—the Canadian Land Surface Scheme, *Ecological Modeling*, 142, 135–154.
- Wang, S., Grant, R. F., Versegny, D. L. and Black, T.A. 2002a. Modelling carbon dynamics of boreal forest ecosystems using the Canadian Land Surface Scheme, *Climatic Change*, 55, 451–477.
- Wang, S., Chen, W. and Cihlar, J. 2002b. New calculation methods of diurnal distributions of solar radiation and its interception by canopy over complex terrain, *Ecological Modeling*, 155, 191–204.
- Wang, S. 2005. Dynamics of surface albedo of a boreal forest and its simulation, *Ecological Modeling*, 183, 477-494.

- Wang, S. and Davidson, A. 2007. Impact of climate variations on surface albedo of a temperate grassland, *Agric. For. Meteorol.*, 142, pp 133-142.
- Waring, R. H., Law, B. E., Goulden, M. L., Bassow, S. L., McCreight, R. W., Wofsy, S. C. and Bazzaz, F. A. 1995. Scaling gross ecosystem production at Harvard Forest with remote-sensing—a comparison of estimates from a constrained quantum-use efficiency model and Eddy-correlation, *Plant, Cell & Environment*, 18, 1201– 1213.
- Waring, R. H., Coops, N. C., Fan, W. and Nightingale, J. M. 2006. MODIS Enhanced Vegetation Index predicts tree species richness across forested ecoregions in the contiguous U.S.A., *Remote Sensing of Environment*, 103, 218-226.
- Xiao, X. M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B. and Ojima, D. 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data, *Remote Sensing of Environment*, 91, 256–270.
- Yuan, W., Liu, S. Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A. H., Goulden, M. L., Hollinger, D. Y., Hu, Y., Law, B. E., Stoy, P. C., Vesala, T. and Wofsy, S. C. 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes, *Agric. For. Meteorol.*, 143, 189-207.

Zhao, M., Heinsch, F. A., Nemani, R. R and Running, S. W. 2005. Improvement of the MODIS terrestrial gross and net primary production global data set, *Remote Sensing of Environment*, 95, 164–176.

Table 4.1. Correlations between model predictors and GPP at the Harvard forest (for 2000 to 2005).

Predictor	Coefficient of Determination ( $R^2$ ) <sup>a</sup>
GVMI	0.11
GVMI (growing season)	<b>0.60</b>
Albedo <sub>NIR</sub>	<b>0.82</b>
EVI	<b>0.84</b>
LST <sub>s</sub>	<b>0.71</b>

<sup>a</sup> Correlations significant at 1% significant level are shown in bold text.

Table 4.2. Correlations between model predictors and different environmental variables (for 2000 - 2005) at the Harvard forest.

Variables \ Predictors	Coefficient of determination ( $R^2$ ) <sup>a</sup>			
	GVMI	Albedo <sub>NIR</sub>	EVI	LST <sub>s</sub>
PAR	0.02	<b>0.47</b>	<b>0.46</b>	<b>0.64</b>
T <sub>air</sub>	0.01	<b>0.67</b>	<b>0.64</b>	<b>0.92</b>
VPD	0.01	<b>0.30</b>	<b>0.29</b>	<b>0.60</b>

<sup>a</sup> Correlations significant at 1% significant level are shown in bold text.

Table 4.3. Correlation ( $R^2$ ) between observed GPP and either R-GPP model predicted GPP or MODIS GPP. All GPP are mean values of 8 day periods.

Site	Study Period	Observed GPP vs simulated GPP (R-GPP model)		Observed GPP vs MODIS GPP	
		$R^2$	RMSE <sup>a</sup>	$R^2$	RMSE <sup>a</sup>
Harvard Forest	2000-2003 (Calibration)	0.95	1.02	0.80	2.78
	2004-2005 (Validation)	0.92	1.62	0.77	3.35
MMSF	2000-2005	0.93	1.47	0.74	3.54
UMBS	2000-2003	0.94	0.95	0.91	1.48

<sup>a</sup> Unit of RMSE is  $\mu\text{mol C/ m}^2/\text{sec}$

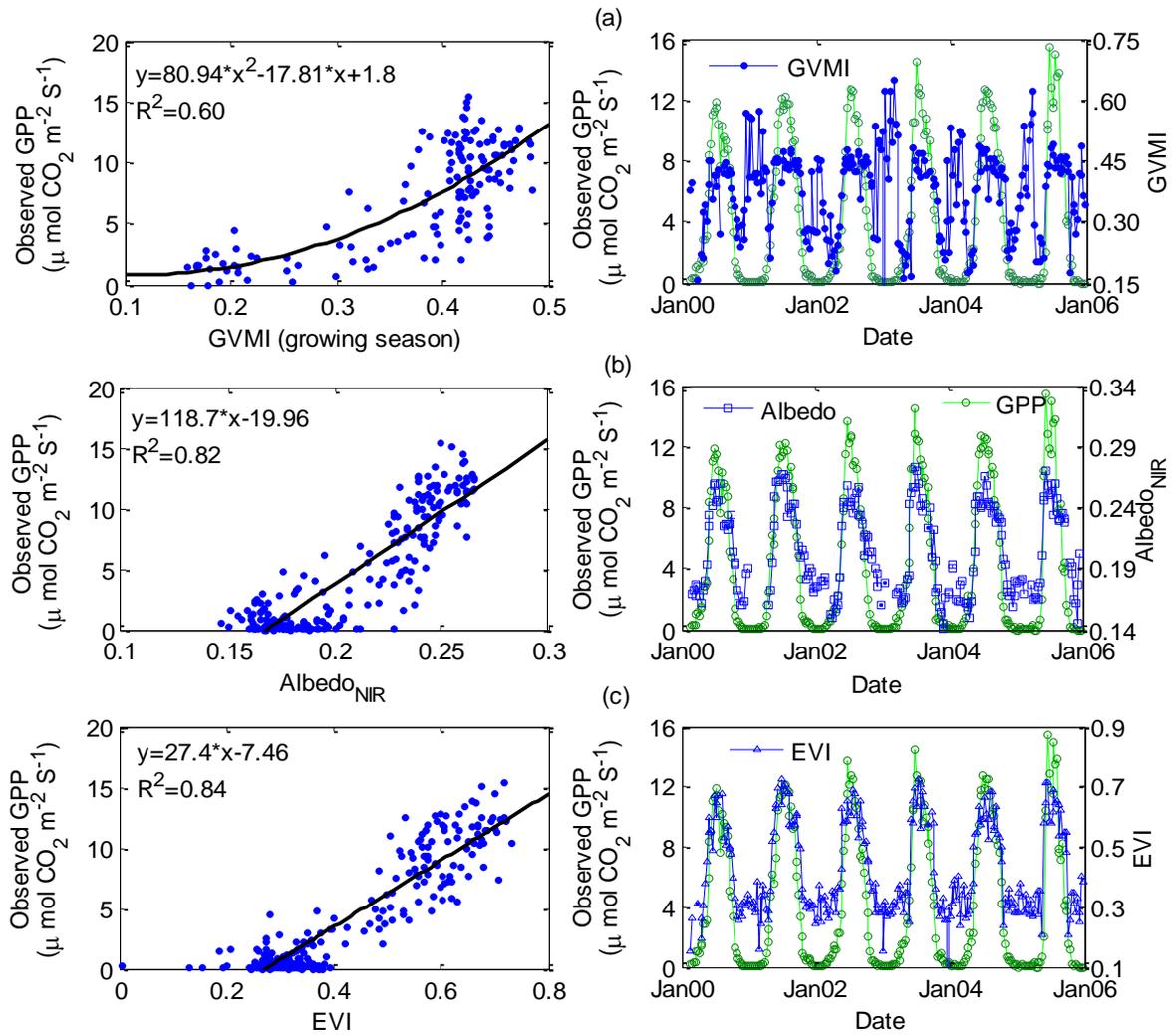
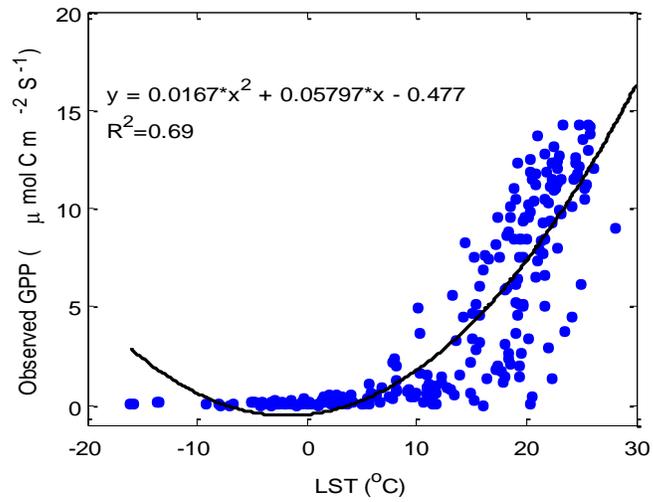
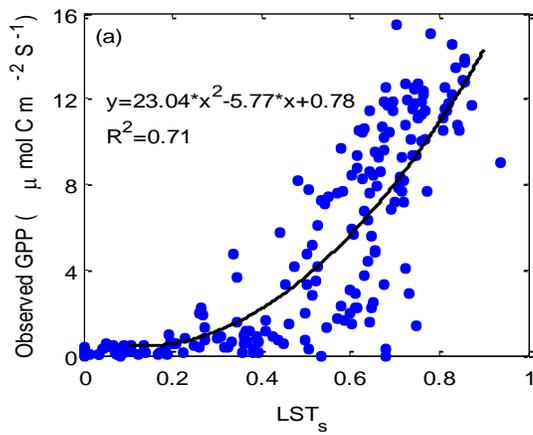


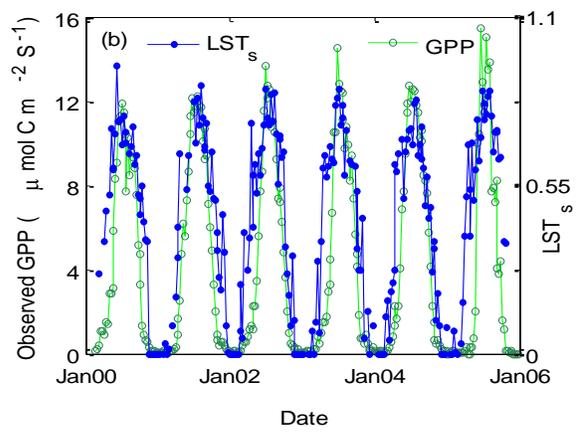
Figure 4.1. (a) Nonlinear/ Linear regression and comparison of seasonal dynamics between GPP and (a) GVMI, (b)  $\text{Albedo}_{\text{NIR}}$ , and (c) EVI for 2000 to 2005 at the Harvard forest. All points represent 8-day means.



(a)



(b)



(c)

Figure 4.2. Scatter plot and polynomial fit between GPP and (a) LST, (b) scaled LST ( $\text{LST}_s$ ), and (c) Seasonal dynamics of GPP and  $\text{LST}_s$ , for 2000 to 2005 at the Harvard forest. All points represent 8-day means.

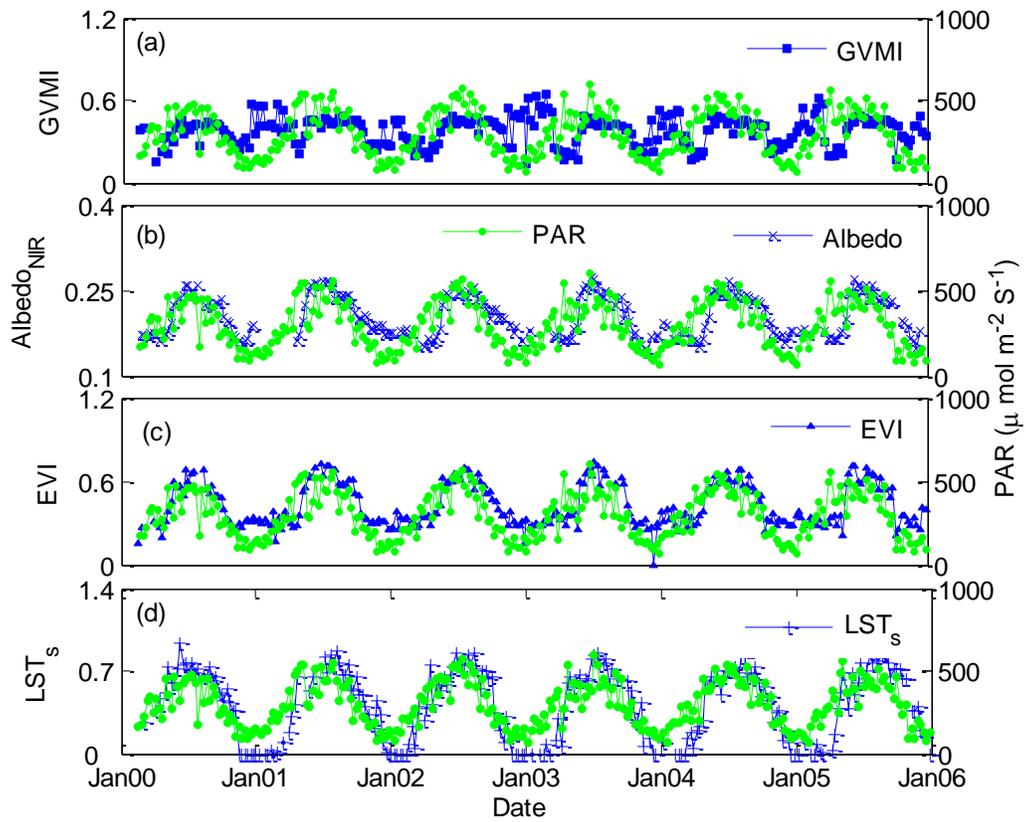


Figure 4.3. Comparison of seasonal dynamics of photosynthetically active radiation (PAR) with seasonal dynamics of GVMi, Albedo<sub>NIR</sub>, EVI and LST<sub>s</sub> for 2000 to 2005 at the Harvard forest. All points represent 8-day means.

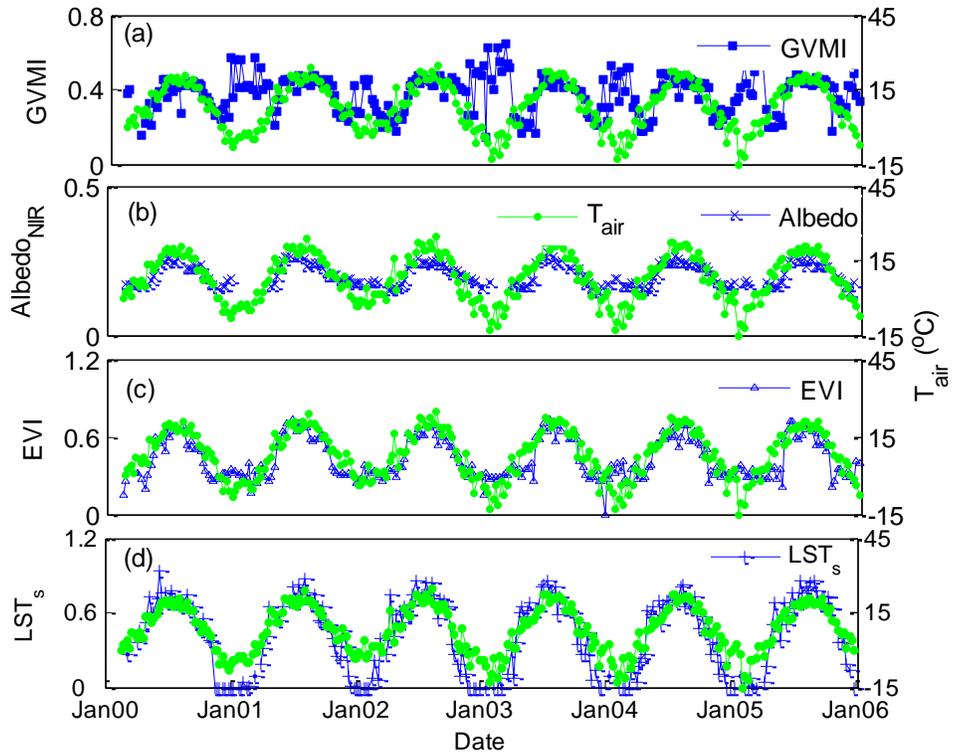


Figure 4.4. Comparison of seasonal dynamics of air temperature ( $T_{\text{air}}$ ) with seasonal dynamics of GVMI,  $\text{Albedo}_{\text{NIR}}$ , EVI and  $\text{LST}_s$  for 2000 to 2005 at the Harvard forest. All points represent 8-day means.

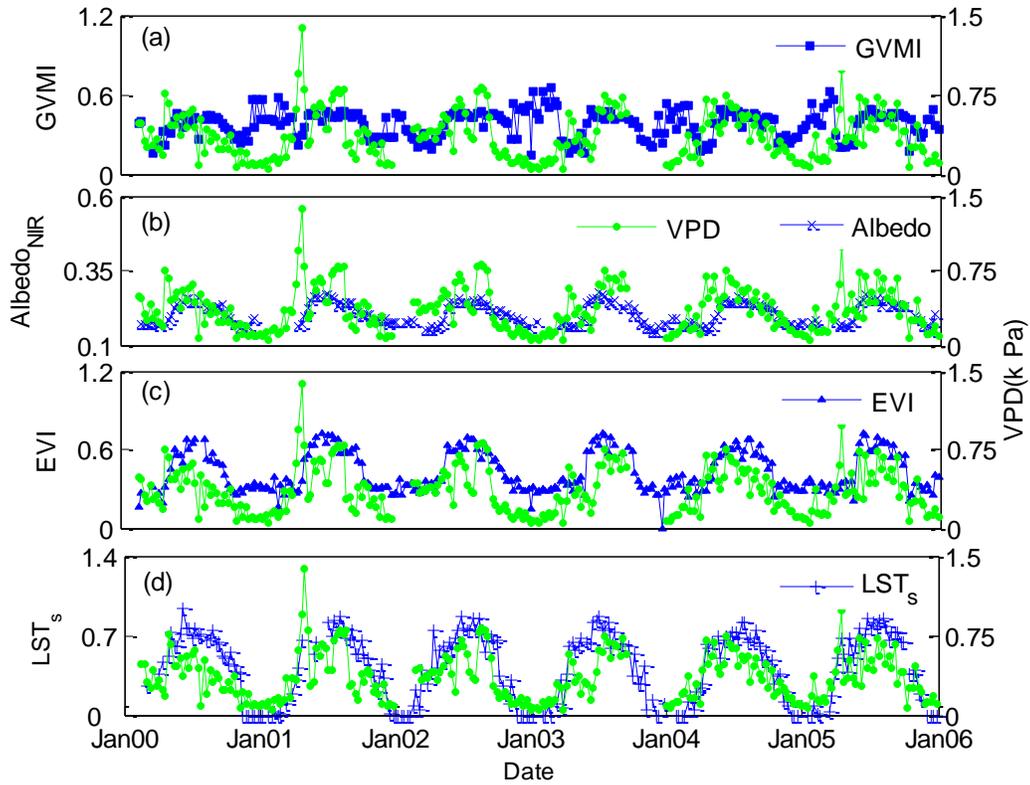


Figure 4.5. Comparison of seasonal dynamics of vapor pressure deficits (VPD) with seasonal dynamics of GVM,  $\text{Albedo}_{\text{NIR}}$ , EVI and  $\text{LST}_s$  for 2000 to 2005 at the Harvard forest. All points represent 8-day means.

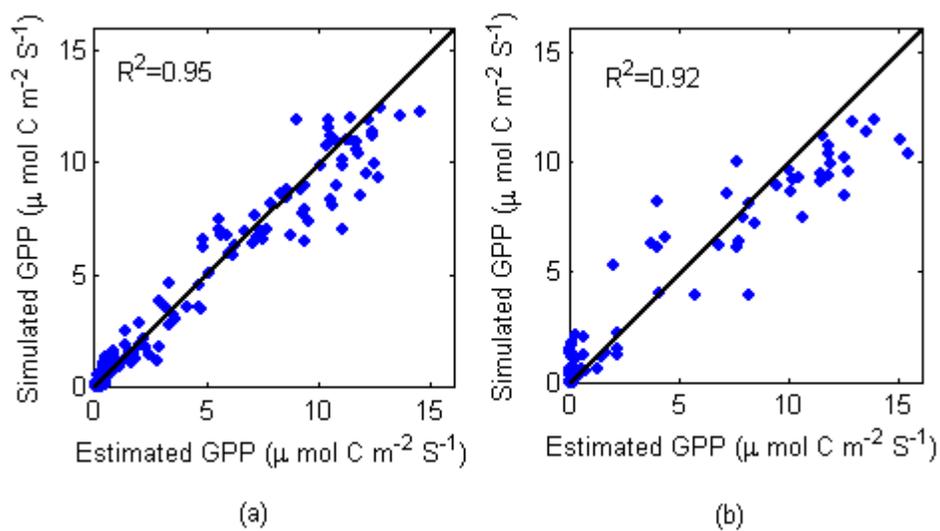


Figure 4.6. Scatter plot of R-GPP model simulated gross primary production (GPP) and eddy covariance tower estimated GPP for the (a) calibration (2000-2003) and (b) validation stages (2004-2005) at the Harvard forest site. All points represent 8-day means.

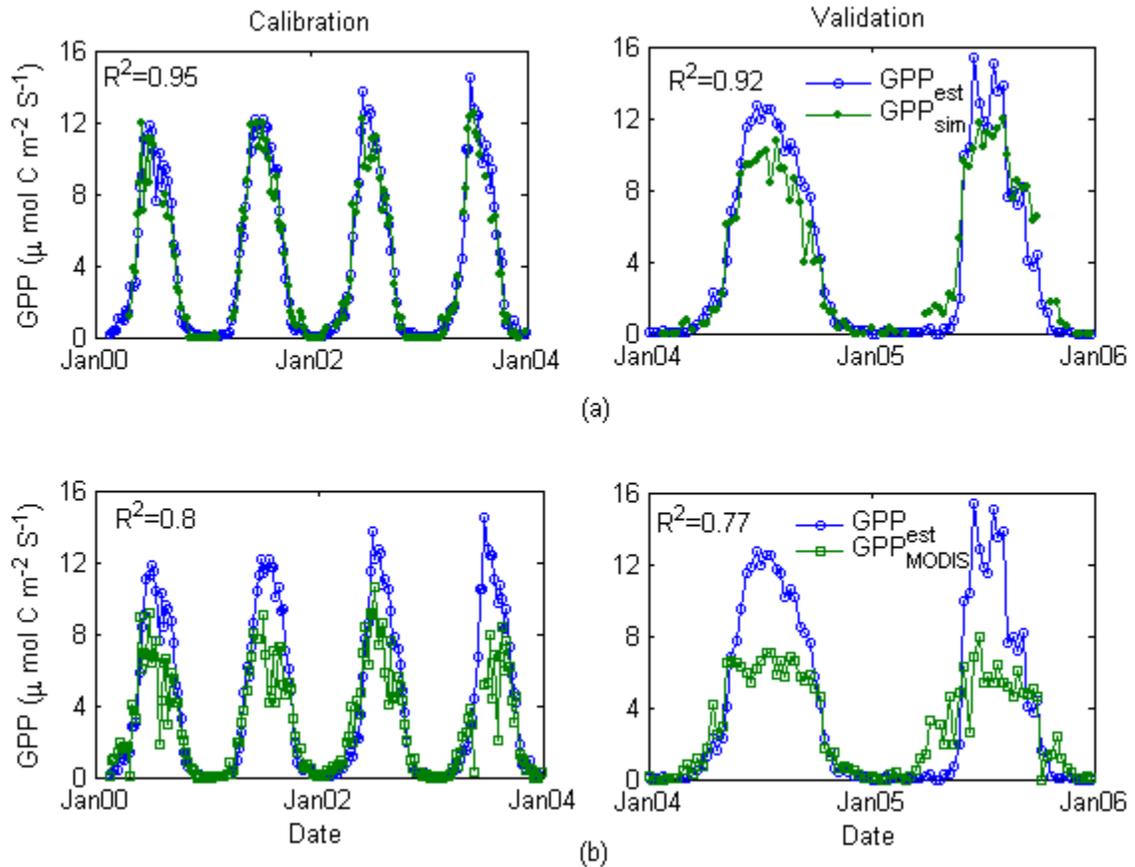


Figure 4.7. Annual cycle of eddy covariance tower estimated gross primary production ( $GPP_{est}$ ) and, (a) R-GPP model predicted GPP ( $GPP_{sim}$ ) and (b) MODIS GPP ( $GPP_{MODIS}$ ) product for the Harvard forest site during the calibration (2000-2003) (left panel) and the validation stage (2004-2005) (right panel), respectively. All points represent 8-day means.

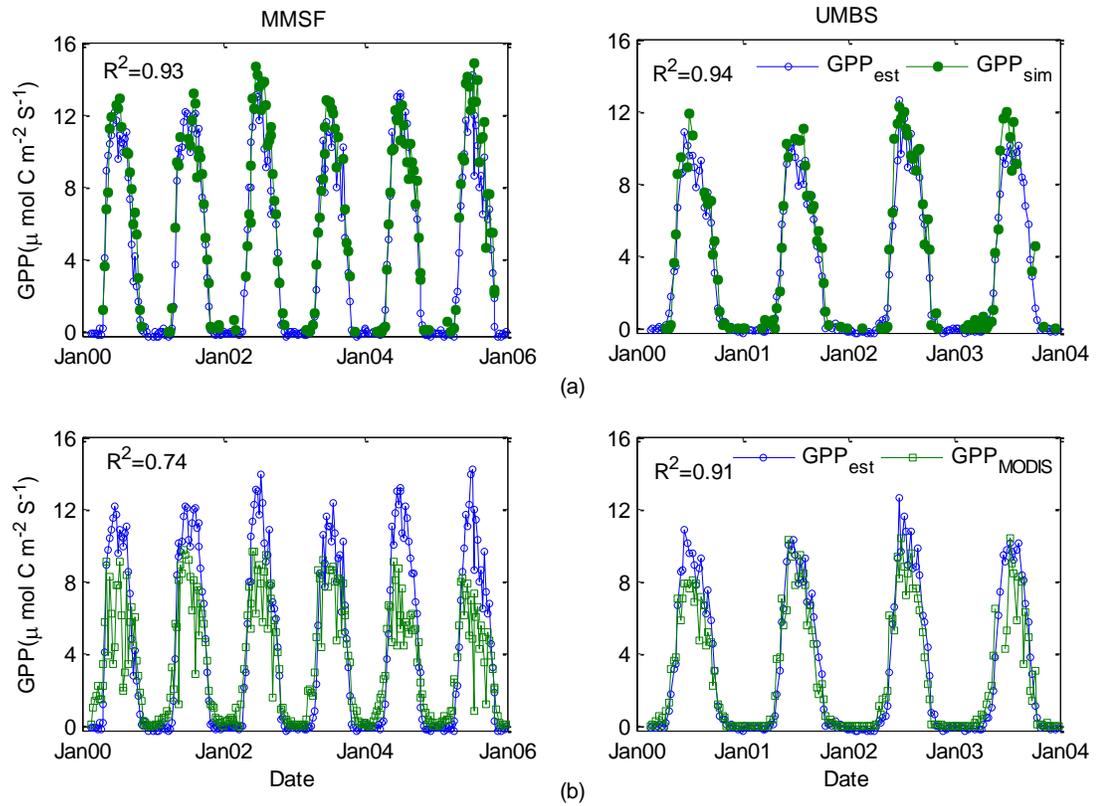


Figure 4.8. Annual cycle of eddy covariance tower estimated gross primary production ( $\text{GPP}_{\text{est}}$ ) and, (a) R-GPP model predicted GPP ( $\text{GPP}_{\text{sim}}$ ) and (b) MODIS GPP ( $\text{GPP}_{\text{MODIS}}$ ) product for the Morgan Monroe State Forest (MMSF) during 2000-2005 (left panel) and University of Michigan Biological Station Site (UMBS) during 2000-2003 (right panel). All points represent 8-day means.

# Chapter 5

## Modeling gross primary production of coniferous forests from MODIS data

---

### 5.1 Introduction

Accurate predictions of ecosystem carbon fluxes, which still suffer from considerable uncertainties, are important for understanding the carbon cycle and for decision makings related to climate change concerns. The gross primary production (GPP) of an ecosystem is a measure of the gross uptake of carbon dioxide (CO<sub>2</sub>) by plants for photosynthesis. It is the main source of carbon flux from atmosphere to land and plays a key role on the balance of carbon fluxes between the biosphere and the atmosphere. GPP is influenced by a number of environmental variables which complicate an accurate estimate of terrestrial GPP (Xiao et al., 2010). The eddy covariance (EC) technique is one of the most appropriate micrometeorological techniques used for measuring carbon, water and energy exchange fluxes at ecosystem level. Over 500 EC towers, set up over diverse landcover and climate regimes, are currently being operated throughout the world. However, EC tower estimated fluxes are representative only over the tower foot print which varies from few hundred meters to a kilometer (Xiao et al.,

2004). Satellite remote sensing is a useful technique to upscale these tower measurements from tree stand scale to regional scale.

Besides ground measuring techniques, GPP can also be estimated from ecosystem models. The light use efficiency (LUE) model (Equation 5.1) proposed by Monteith (1972) is one of the most widely used algorithms (e.g., Running et al., 2004; Xiao et al., 2004; Yuan et al., 2007) which has shown strong potential to simulate the dynamics of GPP because of its strong theoretical basis (Yuan et al., 2007).

$$\text{GPP} = \text{fPAR} \times \text{PAR} \times \varepsilon_{\max} \times f$$

Where, PAR is the incident photosynthetically active radiation ( $\text{MJ. d}^{-1}$ ), fPAR is the fraction of photosynthetically active radiation attendant on the canopy,  $\varepsilon_{\max}$  is the maximum LUE ( $\text{g C MJ}^{-1} \text{ APAR}$ ) and  $f$  is a scaling factor ranging from 0 to 1 used to reduce  $\varepsilon_{\max}$  because of environmental factors such as cold temperature, high vapor pressure deficit, etc., that limit plant function, and  $\varepsilon_{\max} \times f$  produces realized LUE. Many past remote sensing (RS) based studies used this LUE approach to estimate GPP independently or as part of an integrated ecosystem model (Running et al., 2004; Yuan et al., 2007; Landsberg and Waring, 1997; Xiao et al., 2004). MODIS-GPP Algorithm (Running et al., 2004), Vegetation Photosynthesis Model (VPM) (Xiao et al., 2004), and EC-LUE (Yuan et al., 2007), etc., are examples of LUE based models that successfully use RS data to accurately predict carbon fluxes at regional to continental scales with fine spatial resolutions and regular temporal sampling.

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, mounted on both the Terra and Aqua satellites, provides an 8-day GPP product (MOD 17) at 1 km resolution over global, vegetated earth surface based on the LUE method. It uses the MODIS LAI/fPAR (MOD15A2) product, land cover, and biome-specific climatologic data from NASA's Data Assimilation Office (DAO) as inputs. Heinsch et al. (2006) compared the MODIS GPP product with EC tower estimated GPP across diverse landcover types and climate regimes and identified different probable sources of errors related to the inputs of MODIS GPP algorithm: i) inaccuracies in meteorological data simulated from the NASA's Goddard Earth Observing System (GEOS-4) climate model, ii) inaccuracies in MODIS leaf area index (LAI) and fPAR product, and iii) inaccuracies in land cover classifications and errors from the biome specific, radiation use efficiencies.

The Remotely Sensed-GPP model (R-GPP model) by Jahan and Gan (2009) successfully computed GPP for three deciduous forests of northeastern USA from only four RS variables: two radiation budget variables (near infrared albedo ( $\text{Albedo}_{\text{NIR}}$ ) and land surface temperature (LST)) and two ecosystem variables (Global Vegetation Moisture Index (GVMI) and Enhanced vegetation index (EVI)). The R-GPP model explained 93 to 95% of the observed GPP variation for the study sites of Harvard forest, Morgan Monroe State Forest and University of Michigan Biological Station of USA. The objective of this study is to investigate the applicability of these variables in estimating GPP of coniferous forests, and to develop a GPP prediction model solely based on RS data (R-GPP-Coni) for this

forest type. The results of the R-GPP-Coni model will be compared with the GPP estimates from the primary GPP estimation model, the MODIS GPP. The validity and transferability of the R-GPP-Coni model will be further tested at three independent validation sites. All the predictor variables will be derived from the MODIS satellite. As most of the existing GPP estimation models (e.g, MODIS GPP) still rely on ground based meteorological data which are not available with sufficient accuracy for regional scale studies (Heinsch et al., 2006; Turner et al., 2005), therefore we believe that the proposed R-GPP-Coni model, based solely on RS data, will provide an practical and innovative approach to model GPP at regional scale. If the results of this study are encouraging, it will be useful to extensively test the proposed model to simulate the GPP of coniferous forests of other geographical regions at an 8-day temporal and a 1 km, spatial resolutions.

## **5.2 Study Sites and data Sets**

### **5.2.1 Study sites**

In this study we used the carbon flux data from 4 coniferous forests located in different parts of USA (Table 5.1). The 4 sites represent variation in regions, climate, age and species compositions. The Howland forest site, a dense evergreen forest with closed canopy and little understory, is dominated by red spruce, eastern hemlock, balsam fir, white pine, northern white cedar, red maple, paper birch, etc. The average stand height of this boreal-northern hardwood transition forest is about 20 m. The soils of this forest are generally glacial tills with low fertility and high organic composition and the climate is mainly cold and

humid. The forest was logged selectively around 1910 but has been minimally disturbed since that time. The Howland forest west tower is located at a sufficient (775 m) distance from the Howland forest main tower so that the flux source regions, called footprints, of these two towers do not overlap (Hollinger et al., 2004). Within the 1 km west tower footprint, the vegetation is 90% evergreen needle leaf and 10% broadleaf deciduous and the dominant species are hemlock and spruce. The Duke Forest Loblolly Pine Site is an early successional planted pine forest. The site was established in 1985 after a clear cut and burn. Then the loblolly pine seedlings were planted at a regular spacing. This forest area has warm, humid summers and mild winters. The North Carolina loblolly pine site is located in a pine plantation within the mixed forests of the North Carolina lower coastal plain. The average stand age is 18 years and the average canopy height is 14.1 m (Noormets et al., 2010).

### **5.2.2 Carbon flux and remotely sensed data**

All the carbon flux and RS data used in this study are averaged over 8 days and have been collected from the Ameriflux website maintained by the Oak Ridge National Laboratory (<http://public.ornl.gov/ameriflux/dataproducts.shtml>). In this study we have used EVI, GVMi, LST and Albedo<sub>NIR</sub> as input to the proposed R-GPP-Coni model while MODIS derived GPP (MOD17A2) was used to compare the performance of the proposed model with respect to MODIS GPP. The spatial and temporal resolutions of the R-GPP-Coni model were chosen to be 1 km and 8-day, respectively, which are the same as those of MODIS GPP. In this study we

used the 8-day surface reflectance (MOD09A1, collection 5) for four spectral bands: blue (459-479 nm), red (620-670 nm), near infrared (841-875 nm) and shortwave infrared (1628-1652 nm) to calculate the EVI and GVM. The MODIS LST product (MOD11A2, collection 5) is calculated by the generalized Split Window algorithm. It is an 8-day average of cloud-free daily LST. The MODIS albedo is computed from the Bidirectional Reflectance Distribution Function (BRDF) coefficients as a function of optical depth, solar zenith angle and band (Schaff et al., 2002). In this study we used RS data of a 1 km $\times$ 1 km area that covers the EC tower to simulate the GPP (Jahan and Gan, 2009; Xiao et al., 2004) instead of using RS data averaged over N $\times$ N km (N= 3 or 5 or large number) area as used by some previous studies (Sims et al., 2008). In general, the EC tower footprint that varies with the flux tower height, wind speed, topography, etc., is few hundred meters to 1 km in size (Xiao et al., 2004). Therefore using predictors averaged over areas of 3 $\times$ 3 or 5 $\times$ 5 MODIS pixels of 1-km resolution may be too large to represent a tower footprint and may be too coarse in case of complex topography.

## **5.3 Data and research methodology**

### **5.3.1 Selecting model predictors**

The photosynthesis of plants is controlled by a number of physical and physiological variables. At the leaf level, photosynthesis is controlled by incoming solar radiation, air and soil temperature, vapor pressure deficit, available moisture, nitrogen availability, and others while at the ecosystem level,

photosynthesis is also influenced by ecosystem disturbance, leaf area index (LAI), and canopy phenology (Richardson et al., 2009, Ruimy et al., 1995). RS offers the opportunity to monitor some of these variables from a space platform. Vegetation indices (VIs) computed from surface reflectances serve as indicators of plant growth and vegetation greenness and are diagnostic of the canopy cover and architecture (Jiang et al., 2008). VIs are used as proxies to assess many biophysical and biochemical properties such as leaf area (Boegh et al., 2002), canopy chlorophyll content (Gitelson et al., 2005), fractional vegetation cover (e.g., Xiao et al., 2010), and fPAR absorbed by vegetation (Di Bella et al., 2004). EVI (Equation 5.1) is one of the most commonly used vegetation indices which is highly responsive to LAI and canopy structural variations such as canopy type, plant physiognomy, and canopy architecture (Gao et. al., 2000). In addition, EVI is less prone to saturation than some other vegetation indices such as the normalized difference vegetation index (NDVI) in high biomass region, relatively insensitive to residual aerosol contaminations and provides improved vegetation monitoring through a decoupling of the canopy background signal (Huete et al., 2002). EVI corrects the aerosol influences in the red band reflectance as a function of the blue band reflectance. EVI is given as

$$EVI = G \frac{NIR - R}{NIR + C_1 R - C_2 B + L} \quad (5.1)$$

where NIR, R and B are atmospherically corrected surface reflectance in the near-infrared, red and blue bands respectively, G the gain factor, L the canopy background adjustment factor; and  $C_1$  and  $C_2$  are the coefficients of the aerosol resistance term. In this algorithm,  $L=1$ ,  $C_1=6$ ,  $C_2 = 7.5$ , and  $G = 2.5$ . EVI has been

successfully used in the study of landcover change (Wardlow et al., 2007), monitoring of vegetation phenology (Xiao et al., 2006), and modeling of evapotranspiration (Nagler et al., 2005) and GPP (Rahman et al., 2005).

EVI is also capable of capturing the changes in leaf optical properties associated with changes in biochemical, and biophysical properties at different leaf ages (Xiao et al., 2005). Evergreen needle leaf trees consist of needles of various ages. The leaf thickness, dry weight, chlorophyll content and nitrogen content of a needle leaf change with age (Rock et al., 1994). In general, old leaves have higher chlorophyll and water content in comparison to young leaves, which influences the absorbance, transmittance, and reflectance of leaves. Rock et al. (1994) compared the reflectance properties of red spruce first- and second-year needles. They found that the reflectances in the blue band increases with age due to a build up of surface waxes which give needles a whitish appearance. They also found that absorbance in NIR was higher in second-year leaves due to the increase in pigmentation concentration. Moreover, the thickness of leaves increases with age and contain more water which is effective in absorbing NIR radiation. As the computation of EVI involves both the blue and the NIR bands, EVI is able to capture age related changes in canopies (Xiao et al., 2005).

Past studies have used NIR and short wave infra-red (SWIR) data from different sensors to estimate leaf and canopy water content ( $\text{g/m}^2$ ). As for example Hunt and Rock (1989) used NIR and middle infrared reflectances from the Landsat-TM

while Ceccato et al. (2002) used NIR and SWIR from the VEGETATION (VGT) sensor to model vegetation water content. Global vegetation moisture index (GVMI), one of the recently developed RS based indices useful in retrieving equivalent water thickness (EWT) at the canopy level (Ceccato et al., 2002), is computed as

$$GVMI = \frac{(NIR + 0.1) - (SWIR + 0.02)}{(NIR + 0.1) + (SWIR + 0.02)} \quad (5.2)$$

Where NIR and SWIR are reflectance of the rectified NIR band and short wave infrared bands, respectively. Ceccato et al. (2002) examined the capability of GVMI in retrieving EWT for four different ecosystems and reported that the water content derived from GVMI was consistent with field measured water content. Successful applications of GVMI had also been demonstrated in other studies (e.g., Danson and Bowyer, 2004; Du et al., 2005).

Albedo represents the fraction of incident solar radiation reflected by a surface. It is a biophysical property by which canopy interacts with climate (Ollinger et al., 2008). Various studies have been conducted to demonstrate the influence of albedo on the climate. Albedo determines the amount of radiation absorbed at the surface and thereby affects different physical and bio-geochemical processes including energy balance, evapotranspiration, primary production of vegetation, etc. Any change in the vegetation density generally changes albedo because of the strong absorption in the PAR region of the solar spectrum by vegetation. In general, when the canopy is brighter than the soil background, albedo increases with an increase in the density of vegetation (Bounoua et al., 2000). Albedo also

varies with the canopy chemistry and structure. Ollinger et al. (2008) found that there exists a strong and positive relationship between albedo and canopy CO<sub>2</sub> uptake capacity. They also showed that albedo is correlated with the nitrogen concentration of canopy which is a key constraint on the carbon assimilation of canopy. For both deciduous and coniferous forests they found that canopies with higher albedo contain higher concentrations of nitrogen in foliage and possess a greater capacity to remove CO<sub>2</sub> from the atmosphere than the canopies with lower nitrogen.

Land surface temperature (LST) acquired by satellite sensors is another prospective predictor for estimating GPP. It provides vital information on the land surface state and is widely used for studying ecosystem's energy and water budgets. It can incorporate the effect of temperature on vegetation (Sims et al., 2008) and is highly correlated with vegetation dynamics (Sun and Kafatos, 2007). Schwarz et al (1997) found that soil temperature is a better indicator of photosynthesis than air temperature. Studies also showed that LST is positively correlated with vapor pressure deficits which control the uptake of CO<sub>2</sub> and water loss through plant stomata. LST served as a proxy variable in many studies to assess evapotranspiration (Anderson et al., 2007), vegetation water stress (Moran et al., 1994), soil moisture (Ottle et al., 1994) and thermal inertia (Jang et al., 2006).

In this study we will first examine the correlation of these variables with the GPP of coniferous forests (Section 5.3.2) and then test the accuracy of GPP estimated from these predictors. We expect that key physical and physiological factors that control GPP can be assessed in terms of these RS variables. For example, vegetation condition and available soil moisture status will be assessed through EVI and GVMI; the radiation budget will be incorporated through albedo while temperature and vapor pressure deficits status will be assessed through LST.

### **5.3.2 Model development**

Figure 5.1 shows a scatter plot of GPP with predictors EVI, GVMI, LST and albedo for all the study sites. As both GPP and the predictors were averaged over 8-day periods, this analysis will only represent the longer term temporal variation and not the short term (hours to days) variability. It should be noted that 8-day tower GPP values generally contain all days (both sunny and cloudy) over each 8-day period, while RS data comprised of only cloud free images over that period. Furthermore, some predictors such as albedo and GVMI can be significantly affected by snow and high values can be observed because of snow cover above or below the canopy. These snow-affected albedo and GVMI values may not be representative of the actual radiation or moisture status of canopies. Therefore snow-affected RS data were excluded in the scatterplots of GPP versus predictors shown in Figure 5.1.

Figure 5.1 (a), the scatter plot of GPP with EVI, shows a fairly consistent relationship between them ( $R^2 = 0.65$ ) which is expected because green and healthy vegetation will generally photosynthesize more than unhealthy, non-green vegetation. EVI is also related to the vegetation condition. For example, in December 2002, there was an ice storm in the Duke forest which damaged needles of pines in the forest (Stoy et al., 2006). The damage to pine needles was reflected in the EVI of 2003 which was lower than that of 2004 (results not shown) when the plants had recovered from the damage. As a result of the damage, both LAI and GPP of 2003's growing season were also lower than those of 2004's growing season. It seems that EVI is reflective of the vegetation condition and some important biophysical variables such as LAI which controls photosynthesis.

Figure 5.1 (b) shows that GPP generally increases with GVMI (related to soil moisture). However, when GVMI is around 0.4 to 0.5, Figure 5.1 (b) shows that GPP tends to show little relation to GVMI. This is because other than soil moisture, plant growth is also dependant on other resources (such as temperature or solar radiation). According to the 'Law of the minimum', originally enunciated by Liebig in 1855, biological or plant growth is controlled by the resource available in the smallest amount (Barnes et. al., 1998). Several studies on vegetation photosynthesis also support this hypothesis. For example, for different terrestrial biomes, Yuan et al. (2007) reported that photosynthesis of vegetation during the growing season is controlled either by air temperature or by soil

moisture, whichever is the most limiting. On the other hand, Goulden et al. (1997) reported that photosynthesis is predominantly controlled by the incident light during the growing season for a spruce forest of Manitoba.

In this study, albedo at the NIR band,  $\text{Albedo}_{\text{NIR}}$  (0.7 to 5  $\mu\text{m}$ ) has been used partly because it is the most commonly used albedo in modeling ecosystem processes (Ollinger et al., 2008) and partly because vegetation has high reflectance in the NIR. Figure 5.1(c) shows the seasonal dynamics of  $\text{Albedo}_{\text{NIR}}$  and GPP for the study sites.  $\text{Albedo}_{\text{NIR}}$  generally increases with an increase in the greenness of forest because of high canopy reflectance in the NIR band. Moreover, as described earlier, a higher albedo is an indication of a higher concentration of nitrogen (N) in the canopy foliage and canopies with more N possess greater capacity to absorb  $\text{CO}_2$  for photosynthesis than canopies with lower albedo (Ollinger et al., 2008). Albedo at all forest sites followed a seasonal pattern with higher values during growing season and lower values during spring and fall. The effect of ice storm was also evident in the Duke forest's albedo. A perusal of the albedo time series of Duke Forest showed that albedo was lower in 2003, after the ice storm in December 2002, than that of 2004 when the vegetation recovered from the ice damage (data not shown).

Figure 5.1(d), the scatter plot of GPP with LST (Figure 5.1 (d)), shows a positive and strong relationship between LST and GPP. It also shows that below  $0^\circ\text{C}$ , there is no photosynthesis while above  $0^\circ\text{C}$ , GPP gradually increases with LST. Similar

phenomena were also reported in other studies on coniferous species (Goulden et al., 1997; Hollinger et al., 1999; Sims et al., 2008). The possible reason behind this pattern is that plant metabolic activity and growth depend on the availability of water in liquid form, and during winter low soil temperatures limit this water availability by reducing the water uptake through an increase in the viscosity of water and by affecting the growth of new fine roots (Mellander et al., 2004). Moreover during winter, plant's ability to absorb nutrients and to transport those from roots to shoots are also reduced. Many studies have reported this inhibiting effect of low soil temperature on photosynthesis, including that in conifers (Sims et al., 2008; Schwarz et al., 1997; Mellander et al., 2004). For example, Schwarz et al. (1997) reported this winter inhibition for the red Spruce while Ellsworth reported that for the loblolly pine. Therefore in this study, 0°C was chosen as a temperature threshold to define periods of active photosynthesis by a scaled LST ( $LST_s$ ) (Equation 5.3) such that GPP is set to zero when LST is below 0 °C.

$$LST_s = \begin{cases} \frac{LST}{LST_{max}}; & \text{when } LST > 0^\circ \text{ C} \\ 0; & \text{when } LST \leq 0^\circ \text{ C} \end{cases} \quad (5.3)$$

where LST is the observed LST and  $LST_{max}$  is the maximum LST. In this study  $LST_{max}$  is set to 30°C because it has been used as the optimum LST by some other researchers (Sims et al., 2008; Jahan and Gan, 2009). Figure 5.1 (e) shows that GPP is strongly correlated with  $LST_s$  throughout the growing season with a  $R^2$  of 0.86.

Given that GVMI, EVI, Albedo<sub>NIR</sub> and LSTs are correlated to GPP, we propose a Remotely Sensed GPP model (Equation 5.4) for coniferous forest (R-GPP-Coni) based on these four RS predictors only,

$$\text{GPP} = k \times \text{GVMI}^a \times \text{LST}_s^b \times \text{Albedo}_{\text{NIR}}^c \times \text{EVI}^d \quad (5.4)$$

where k is a scalar, and a, b, c, and d are exponents. These model parameters were optimized using the estimated GPP of 2000 to 2004 from the EC tower located at the Howland forest site and a global optimization scheme, the Shuffled-Complex Evolution (Duan et al, 1993; 1994). By Shuffled-Complex Evolution, the optimized values of k, a, b, c, d have been found to be 38.309, 0.396, 0.673, 0.332 and 0.604, respectively.

## 5.4 Results

### 5.4.1 Calibration results

The GPP predicted by the R-GPP-Coni model was compared against the GPP measured at the EC towers. We did not attempt to gap fill the missing RS data because our objective was to assess the performance of the R-GPP-Coni model when reliable RS data were available. The calibration results show that the model could capture the seasonal dynamics of the observed GPP quite accurately (Figure 5.2). With respect to the EC tower estimated GPP, it is clear that R-GPP-Coni ( $R^2=0.94$ ; root mean squared error, RMSE = 1.10 gm C/ m<sup>2</sup>/ day) could predict more accurate GPP (Figure 5.3) than the MODIS GPP algorithm ( $R^2 = 0.87$ , RMSE=2.02 gm C/ m<sup>2</sup>/ day) for the Howland Forest site, especially during the peak growing season. The MODIS algorithm consistently showed an under-

estimation of GPP in the peak growing season (June to September) (Figure 5.2). Besides underestimation, the MODIS-GPP time series also seem to exhibit more variability than that of the R-GPP-Coni model (Figure 5.2 and 5.3). Uncertainties related to meteorological inputs, erroneous land cover classifications and LAI/fPAR product used in the MODIS algorithm could contribute to the relatively large discrepancies found between the observed and the MODIS simulated GPP (Heinsch et al., 2006).

Richardson et al. (2007) reported that environmental variations (i.e. variation in solar radiation, precipitation, temperature) is directly related to short term (daily to monthly) variations in CO<sub>2</sub> uptake for the Howland forest. Studies showed that at the Howland forest, carbon uptake was higher than normal when spring (April) and fall temperature (November December) were above-average (Hollinger et al., 2004; Richardson et al., 2007). They also found that the CO<sub>2</sub> uptake was lower than normal when soil moisture levels were either too higher or too lower than when soil moisture levels were intermediate during the growing season. Such variations of GPP in response to climatic variations are noticed in this study. For example, in 2001 the spring temperature at the Howland forest was above normal and summer precipitation was little lower than normal (Richardson et al., 2007) which were favorable for photosynthesis. On the other hand, in 2002-2004 its spring temperature were below normal (Richardson et al., 2007) and summer precipitation were either much lower or higher than normal which were unfavorable for carbon uptake which consequently caused the GPP of 2002-2004

to be lower than that of 2001 (Figure 5.2). These variations of GPP in response to climatic variations were accurately modeled by the proposed R-GPP-Coni model.

#### **5.4.2 Validation results**

Although EC towers generally provide accurate estimates of GPP at stand scales, only limited numbers of EC towers have been established throughout the world. Therefore, if the proposed model, calibrated for a coniferous forest, can also be applied to estimate the GPP of other coniferous forests located in north-eastern USA, it will be very useful to estimate GPP at sites where EC tower data are not available. Moreover EC tower measured fluxes are representative of actual GPP only at a scale of tower footprint which varies from few hundred meters to 1 km depending on vegetation and fetch, as discussed earlier. In other words, the propose model will also be valuable to upscale the EC tower measurements from stand scale to regional or continental scales. In general, the efficiency in up-scaling or the transferability of a GPP model to another region will depend on how accurately the predictor variables (EVI, LST<sub>s</sub>, albedo, GVMI) represent the physical (e.g, moisture condition, surface temperature, VPD, etc) and biological drivers (leaf area, vegetation condition, etc.) of GPP. The transferability of the proposed, R-GPP-Coni model was assessed with respect to EC tower data observed at 3 other coniferous forest sites: Howland forest west tower site, North Carolina loblolly pine site and Duke Pine site.

The primary production of many ecosystems is generally limited by the amount of available nitrogen (Gaige et al., 2007). The application of nitrogen helps to overcome this limitation and increases the sequestration of CO<sub>2</sub> in plant biomass. The canopy processes following nitrogen applications had been examined in many studies. The Howland forest west tower site is one of these experimental sites which receive nitrogen treatment. Every year, 18 kg of dissolved nitrogen fertilizer is applied by helicopter over a 21 ha area around the tower. Figures 4(a) and 5(a) show the results from the R-GPP-Coni model and MODIS GPP algorithm for this experimental site. Results shows that the proposed model simulated the observed GPP of this site more accurately ( $R^2=0.89$ , RMSE= 1.25 gm C/ m<sup>2</sup>/ day) than the MODIS GPP algorithm ( $R^2=0.82$ , RMSE=1.54 gm C/ m<sup>2</sup>/ day) that suffered from an underestimation and large fluctuations of the GPP in the peak growing season for most years (2000-2004). However, estimations from the MODIS algorithm were quite accurate in early and late growing seasons (Figure 5.4 (a)). In contrast, the R-GPP-Coni model only showed slight over-estimations during the peak growing season of 2001-2003 and marginal underestimation in 2000 and 2004. We suspect that the underestimation problem might be caused by the saturation problem of RS data acquired by optical sensors (discussed in Section 5.5). However, the overall accuracy of GPP simulated by the R-GPP-Coni model shows that the model is capable of simulating GPP for sites subjected to nitrogen treatment which should be confirmed through more similar studies. The R-GPP-Coni model is better than the MODIS model partly

because it has albedo as one of its predictors which reflects the availability of canopy nitrogen, as discussed earlier (Ollinger et al., 2008).

For the Duke pine site, the R-GPP-Coni model's predicted GPP also followed the seasonal dynamics of the observed GPP reasonably well (Figure 5.4 (b)) and again the agreement is noticeably better than the GPP estimated by the MODIS algorithm (Figures 4 (b) and 5(b)), e.g.,  $R^2= 0.84$ ,  $RMSE= 1.64 \text{ gm C/ m}^2/ \text{ sec}$  versus  $R^2= 0.64$ ,  $RMSE=2.73 \text{ gm C/ m}^2/ \text{ sec}$ . Among the four sites, the MODIS's performance was worst for this site with large discrepancies in the summer of 2004 when some errors were as high as 90%. The R-GPP-Coni model could also accurately estimate the GPP at the beginning of 2003 following an ice storm in December 2002 which damaged leaves and caused a large reduction in the LAI (Stoy et al., 2006) and GPP. On the other hand, the MODIS algorithm underestimated the observed GPP for the same period of 2003. Because by 2004 the canopies had generally recovered from the ice damage, the growing season GPP of 2004 was expectedly higher than the growing season GPP of 2003 even though precipitation (i.e. available soil moisture) was lower in 2004 and these variations of GPP in response to ice damage were well captured by the R-GPP-Coni model. The better performance of the R-GPP-Coni model than the MODIS GPP in both growing seasons of 2003 and 2004 indicates that the proposed model is capable of modeling reasonably accurate GPP under variable biological (e.g., LAI) and climatic (e.g., precipitation) conditions.

Figure 5.4 (c) shows the seasonal variation of the observed, R-GPP-Coni and MODIS estimated GPP for the North Carolina Loblolly Pine site. In general, R-GPP-Coni captured the seasonal variations well with  $R^2 = 0.88$  (RMSE=1.19 gm C/ m<sup>2</sup>/ sec). On the other hand during summer, the MODIS-GPP was consistently lower than the observed GPP ( $R^2 = 0.76$ ; RMSE=1.78 gm C/ m<sup>2</sup>/ sec) and in few instances, especially in July 2006, the differences between them were quite large.

## 5.5 Discussion of results

Even though the R-GPP-Coni model was successful in modeling the GPP of 4 coniferous forests, occasionally there were large differences (especially in the validation stage) between the observed and simulated GPP. These discrepancies can partly be attributed to the limitations of the R-GPP-Coni model given that it is a simple model based on only 4 predictors. Since the proposed model has been parameterized for a particular site, we generally expect using this model to estimate the GPP of other sites to incur some discrepancies. Further, modeling the interannual GPP variation of forest ecosystems is still challenging, even for a relatively complex model such as the MODIS GPP algorithm (Heinsch et al, 2006; Richardson et al., 2007). Moreover, some important factors that also influence GPP, such as solar radiation, soil fertility and soil moisture availability, have not explicitly incorporated into this model. Past studies (White et al., 2000; Esser, 1987) argued that incorporation of soil fertility or nutrient availability increase the accuracy in the estimation of primary product. But the extent to which, and how, soil fertility or nutrient limitations affect the rates of

photosynthesis in vegetation is still poorly known and thus limits their applicability in the GPP model (Raich et al., 1991). Moreover, in this study 0°C has been used as the temperature threshold to set the GPP to zero because low temperature limits photosynthesis. But at leaf and stand levels, the minimum temperature threshold can vary widely between species, even within the same biome type (Heinsch et al., 2003; Mu et al., 2007; Raich et al., 1991). Heinsch et al. (2003) have used -8°C as the minimum temperature ( $T_{\min}$ ) threshold for the evergreen forest when such a temperature stress will cause the stomata to close almost completely, stopping photosynthesis. On the other hand, Raich et al. (1991) used 2.5°C as the  $T_{\min}$  in their terrestrial ecosystem process based model while Sims et al. (2008) used 0°C as the  $T_{\min}$  in their GPP model, both for coniferous forests. Therefore we think that a single minimum temperature may not be appropriate for all the four sites and may cause some errors in modeling the cold season GPP.

Stand age is an important structural determinant of the canopy carbon gain. A reduction of photosynthesis with age has been found in conifer trees due to a lower stomatal conductance, reduced light interception per unit leaf area (Kostner et al., 2002). In addition, a reduction in hydraulic conductivity with age could cause a reduction in the water use efficiency (Kostner et al., 2002). Rautiainen et al. (2011) conducted a study on the coniferous stand in Finland and found that albedo decrease with age due to increase in stand volume. Therefore, even though stand age was not directly incorporated in the R-GPP-Coni model, albedo might

have indirectly incorporated this aging effect on photosynthesis, since the four conifer sites have trees that vary over a wide range of age (17 to 140 years).

As expected with any model, the performance of R-GPP-Coni is dependent upon the quality of the data used to calibrate and to drive the model. In this study, we have used the EC tower estimated GPP to calibrate and to validate the proposed R-GPP-Coni model. These GPP data were estimated from the daytime net ecosystem exchange ( $NEE_d$ ) and daytime respiration ( $R_d$ ) data which are subjected to uncertainties (Xiao et al., 2004). Moreover gap-filling techniques of these  $NEE$  and  $R_d$  data are still subjected to a great deal of discussions (Xiao et al., 2004). Even though GPP estimated from gap-filled data are reasonably reliable and being widely used to calibrate and validate the performance of GPP models, some uncertainties are expected. Moreover in this study we compared the observed GPP with the estimated GPP derived from RS data of  $1 \text{ km}^2$  resolution. In contrast, the size of a flux tower footprint depends on the tower height and environmental conditions such as the wind speed and thus it varies from tower to tower and even for the data of a given tower, it varies temporally, depending on the environmental condition. Therefore the R-GPP-Coni model simulated GPP, estimated over a  $1 \text{ km}^2$ , area may not always be representative of the EC tower measured GPP.

The growing-season GPP estimated by the R-GPP-Coni model can be affected by the saturation problem of RS data acquired from the optical sensors. In an

environment with dense canopies, the reflectance from leaves of lower layers can be obscured by that of upper layers and so the reflectance of canopies becomes non-representative. Due to this saturation problem, the reflectance data for a medium dense and highly dense vegetation may be similar to each other. In other words, using “saturated” RS data in estimating GPP is expected to produce underestimated GPP (Myneni et al., 2002). The GPP predicted by the R-GPP-Coni is also dependent on the compositing procedure of the MODIS 8-day composite images. The MODIS reflectance data is composited on the basis of a minimum-blue criterion through which only data for the clearest atmospheric conditions, not the average conditions, is selected. Therefore, EVI and GVMi (input for R-GPP-Coni) computed from the reflectance products may not be representative of the average condition of each 8-day period (Xiao et al., 2004). Exclusion of days with high and low reflectance could cause either an underestimation or an overestimation of GPP, respectively. This problem can be eliminated by using daily RS data to drive the R-GPP-Coni model. However, because of cloud cover effect, daily data are not always available. Furthermore, the computational efforts of using daily data must be justified against the increased efficiency of the model. The errors may also result from the use of 16-day albedo from the MODIS sensor. Because until now only 16 day albedo product is available from MODIS, for any 8-day period, the R-GPP-Coni model uses NIR albedo averaged over that particular 8-day and the previous 8-day as input. This 16-day product may introduce some uncertainties to the estimated GPP.

Although a number of satellite data driven GPP models are currently available for estimating GPP of different ecosystems, there are only a few GPP models which are based on solely RS data. Sims et al. (2006) proposed a simple model for estimating GPP using EVI as the only predictor. They demonstrated that EVI alone can estimate GPP that were as good as or better than the MODIS GPP for many sites. But the main drawback of this model is that it could not define the photosynthetically active period for the strongly evergreen vegetation. Moreover it showed poor performance for the sites suffered from summer drought. Later, Sims et al (2008) modified this model by incorporating another model predictor and the new model (called the Temperature and Greenness (TG) model) remains to be solely RS data based. But this model provides GPP estimates only at 16 day scale, as compared to the R-GPP-Coni model designed to estimate GPP at 8-day scale using four RS predictors. The proposed R-GPP-Coni model, in contrast to T-G model, uses two additional RS-based predictors, moisture index (GVMI) and radiation variable (NIR albedo), to incorporate the soil moisture and the radiation information in the estimation of GPP.

## **5.6 Summary and conclusions**

In this study, we have modified the R-GPP model developed by Jahan and Gan (2009) for deciduous forests to the R-GPP-Coni model for predicting the 8-day average GPP for coniferous forests of northern USA. The R-GPP-Coni model captured the seasonal dynamics of the observed 8-day GPP successfully by explaining 88 to 94% of the observed variations of GPP with a RMSE ranging

from 1.06 to 1.64 gm C/ m<sup>2</sup>/ day over the 4 study sites. On the other hand, the primary RS based GPP estimation algorithm of MODIS only explained 64% to 87% of the GPP variability of these sites with a RMSE ranging from 1.06 to 2.73 gm C/ m<sup>2</sup>/ day. The successful validation of the model both temporally and spatially reveals that this simple model is capable of modeling GPP of coniferous forest of north eastern USA, even though these sites have tree stands that are diversified in species and age, and subjected to different management practices such as the nitrogen application. This study also demonstrated that this simple and empirical approach has the potential for up-scaling EC tower GPP data to regional scale.

GPP is the photosynthetic response to climate, nutrients, and disturbances, and can vary considerably within a small latitudinal range. This study will improve our ability to model the dynamics of the terrestrial carbon cycle, increasing our understanding of the interactions between climate, ecosystem process and RS variables. This study demonstrated that RS variables such as GVMI, EVI, albedo and LST are useful predictors capable of representing certain important climatic, physiological factors and the nitrogen availability that influence GPP. Therefore, the R-GPP-Coni model based on these predictors could capture the temporal dynamics of GPP of coniferous ecosystems in the north-eastern USA. However, the robustness of the R-GPP-Coni model should be further validated to other coniferous forests of different geographical and climatic regimes such as the boreal forests of Canada. In addition, it may be necessary to modify the model

parameter to effectively apply the proposed R-GPP-Coni model to other biomes, such as grassland and savannas.

## 5.7 References

- Anderson, M. C., Norman, J. M., Mecikalski, J. R., Otkin, J. A. and Kustas, W. P. 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research*, 112, D10117,
- Barnes, B.V., Zak, D. R., Denton, S. R. and Spurr, S. H. 1998. *Forest Ecology*, 4<sup>th</sup> Edition, John Wiley & Sons.
- Boegh, E., Soegaard, H., Broge, N., Hasager, C. B., Jensen, N. O., Schelde, K., et al. 2002. Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment*, 81, 179–193.
- Ceccato, P., S. Flasse, S. and Gregoire, J. M. 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote Sensing of Environment*, 82, 198– 207.
- Danson, F. M. and Bowyer, P. 2004. Estimating live fuel moisture content from remotely sensed reflectance. *Remote Sensing of Environment*, 92, 30-321.
- Di Bella, C. M., Paruelo, J. M., Becerra, J. E., Bacour, C. and Baret, F. 2004. Effect of senescent leaves on NDVI-based estimates of f APAR: experimental and modelling evidences. *International Journal of Remote Sensing*, 25, 5415–5427.
- Du, X., Wang, S. and Zhou, Y. 2005, Monitoring and spatio-temporal evolution researching on vegetation leaf water in China. *IEEE Transactions Geoscience and Remote Sensing*, 4, 2365–2368.

- Duan, Q., Gupta, V. K. and Sorooshian, S. 1993. A Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and its Applications*, Vol 76(3), pp 501-521.
- Duan, Q., Sorooshian, S and Gupta, V. K. 1994. Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models. *Journal of Hydrology*, Vol.158, 265-284, 1994,
- Esser, G. 1987. Sensitivity of global carbon pools and fluxes to human and potential climatic impacts. *Tellus*, 39B, 245- 260.
- Gaige, E., Dail, D.B., Hollinger, D.Y., Davidson, E.A., Fernandez, I.J., Sievering, H., White, A., Halteman, W. 2007. Changes in canopy processes following whole-forest canopy nitrogen fertilization of a mature spruce-hemlock forest. *Ecosystems*, 10:7, 1133-1147.
- Gao, B. C. 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, 58, 257– 266.
- Gitelson, A. A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, 161, 165–173.
- Heinsch, F. A, Reeves, M. C., Votava, P., Kang, S., Milesi, C., Zhao, M., et al. 2003. User's guide: GPP and NPP (MOD17A2/A3) products, NASAMODIS land algorithm (pp. 1–57). Missoula, MT: The University of Montana.
- Heinsch, F.A., Zhao, M., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., Bolstad, P. V., Cook, B. D., Desai, A. R. et al. 2006, Evaluation of

- remote sensing based terrestrial production from MODIS using AmeriFlux eddy tower flux network observations. *IEEE Transactions Geoscience and Remote Sensing*, 44, 1908–1925.
- Hollinger, D. Y., Aber, J., Dail, B., Davidson, E. A., Goltz, S. M., Hughes, H., Leclerc, M. Y. Lee, J. T., Richardson, A. D., Rodrigues, C., Scott, N. A., Achuatavariar, D. and Walsh, J. 2004. Spatial and temporal variability in forest-atmosphere CO<sub>2</sub> exchange. *Global Change Biology*, 10, 1689-1706.
- Hollinger, D. Y., Goltz, S. M., Davidson, E. A., Lee, J. T., Tu, K. and Valentine, H. T. 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. *Global Change Biology* 5:8. 891-902.
- Huete, A., K. didan, T. Miura, E. P., Rodriguez, X., GAO, X. and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, pp. 195–213.
- Hunt, E. R. and Rock, B. N. 1989. Detection of changes in leaf water content using near-infrared and middle-infrared reflectances, *Remote Sensing of Environment*, 30, 43– 54.
- Jahan, N. and Gan, T. Y. 2009. Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables. *Journal of Geophysical Research*, Vol.114, No. G04026.
- Jahan, N. and Gan, T. Y. 2011. Modeling vegetation-climate relationship in a central mixed wood forest region of Alberta using normalized difference and

- enhanced vegetation indices. *International Journal of Remote Sensing*, 32, 313-335.
- Jang, J. D., Viau, A. A. and Anctil, F. 2006. Thermal-water stress index from satellite images. *International Journal of Remote Sensing*, 27, 1619– 639.
- Jiang, Z., Huete, A. R., Didan, K. and Miura, T. 2008. Development of a two-band enhanced vegetation index without blue band. *Remote Sensing of Environment*, 112, pp. 3833-3845.
- Kostner, B., Falge, E., Tenhunen, J. D. 2002. Age-related effects on leaf area/sapwood area relationships, canopy transpiration and carbon gain of Norway spruce stands (*Picea abies*) in the Fichtelgebirge, Germany. *Tree Physiology*, 22, 567-574.
- Monteith, J. 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, 9, 747-766.
- Moran, M. S., Clarke, T. R., Inoue, Y. and Vidal, A. 1994. Estimating crop water deficit using the relation between surfaceair temperature and spectral vegetation index. *Remote Sensing of Environment*, 49, 246–263.
- Mellander P-E, Bishop, K., Lundmark, T. 2004. The influence of soil temperature on transpiration: a plot scale manipulation. *Forest Ecology and Management* 195: 15–28.
- Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., et al. 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83, 214–231.

- Nagler, P. L., Cleverly, J., Glenn, E., Lampkin, D., Huete, A. R., & Wan, Z. 2005. Predicting riparian evapotranspiration from MODIS vegetation indices and meteorological data. *Remote Sensing of Environment*, 94, 17–30.
- Noormets, A., Gavazzi, M. J., McNulty, S.G., Domec, J-C., Sun, G., King, J.S., Chen, J. 2010. Response of carbon fluxes to drought in a coastal plain loblolly pine forest. *Global Change Biology*, vol. 16, pp 272-287.
- Ollinger, S. V., A. D. Richardson, M. E. Martin, D. Y. Hollinger, et al. 2008. Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forest: Functional relations and potential climate feedbacks. *PNAS*, 105, 19366-19341.
- Ottle, C., and Vidalmadjar, D. 1994. Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-Mobilhy region. *Journal of Hydrology*, 158, 241–264.
- Rahman, A. F., Sims, D. A., Cordova, V. D., & El-Masri, B. Z. 2005. Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes. *Geophysical Research Letters*, 32, L19404. doi:10.1029/2005GL024127.
- Raich, J. W., Rastetter, E. B., Melillo, J. M., Kicklighter, D. W., Steudler, P. A., Peterson, B. J., Grace, A. L., Moore, B., & Vorosmarty, C. J. 1991. Potential net primary productivity in South-America—application of a global-model. *Ecological Applications*, 1, 399–429.

- Rautiainen, M., Stenberg, P., Motts, M., Manninen, T. 2011. Radiative transfer simulation link boreal forest structure and shortwave albedo. *Boreal Environment Research*, 16, 91-100.
- Richardson, A. D., Hollinger, D.Y., Aber, J. D., Ollinger, S. V., Braswell, B.H. 2007. Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange. *Global Change Biology* 13:4, 788-803.
- Richardson, A. D., Hollinger, D. Y., Dail, D. B., Lee, J. T., Munger, J. W., & O'Keefe, J. 2009. Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests. *Global Change Biology*, 29, 321–331.
- Rock, B. N., Williams, D. L., Moss, D. M., Lauten, G. N., & Kim, M. 1994. High-spectral-resolution field and laboratory optical reflectance measurements of red spruce and eastern hemlock needles and branches. *Remote Sensing of Environment*, 47, 176–189.
- Ruimy, A., Jarvis, P. G., Baldocchi, D. D., & Saugier, B. 1995. CO<sub>2</sub> fluxes over plant canopies and solar radiation: A review. *Advances in Ecological Research*, 26, 1–68.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M. and Hashimoto, H. 2004. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, 54, 547–560.

- Schwarz, P.A., Fahey, T. J., Dawson, T.E. 1997. Seasonal air and soil temperature effects on photosynthesis in red spruce (*Picea rubens*) saplings. *Tree Physiology*, 17, 187–194.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D. et al. 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS, *Remote Sensing of Environment*, 112, 1633-1644.
- Stoy, P. C., Gabriel, Katul, G., Siqueira, M. B. S., Juang, J., Novick, K. A., McCarthy, H. R., Oishi, A. C., Uebelherr, J. M., Kim, H. and Oren, R. 2006. Separating the effects of climate and vegetation on evapotranspiration along a successional chronosequence in the southeastern US. *Global Change Biology*, 12, 2115-2135
- Sun, D. and Kafatos, M. 2007. Note on the NDVI-LST relationship and the use of temperature related drought indices over north America. *Geophysical Research Letters*, 34, L24406.
- Wardlow, B. D., Egbert, S. L., & Kastens, J. H. 2007. Analysis of time-series MODIS 250m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, 108, 290–310.
- White, M. A., Thornton, P. E., Running, S. W. and Nemani R. R. 2000. Parameterization and sensitivity analysis of the Biome-BGC terrestrial ecosystem model: Net primary production control. *Earth Interaction*, Vol. 4, pp. 1-84.

- Xiao, X. M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B. and Ojima, D. 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, 91, 256–270.
- Xiao, X., Zhang, Q., Saleska, S., Hutyyra, L., De Camargo, P., Wofsy, S., Frohking, S. et al. 2005. Satellite based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment*, 94, 105-120.
- Xiao, X., Hagen, S., Zhang, Q., Keller, M., & Moore, B. 2006. Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images. *Remote Sensing of Environment*, 103, 465–473.
- Xiao, J., Zhuang, Q., Law, B. E., Chen, J., Baldocchi, D.D. et al. 2010. A continuous measure of gross primary production for the conterminous United States from MODIS and AmeriFux data. *Remote Sensing of Environment*, 114, 576-591.
- Yuan, W., S. Liu, G. Zhou, G. Zhou, L. L. Tieszen, D. Baldocchi, C. Bernhofer, H. Gholz, A. H. Goldstein, M. L. Goulden, D. Y. Hollinger, et al. 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agriculture and Forest Meteorology*, 143, 189-207.

Table 5.1. Location (lat/long in decimal degrees), annual mean temperature (AMT), annual precipitation (AP), stand age and years of data used for the 4 eddy covariance tower sites used in this study.

<b>Site name</b>		<b>Latitude</b>	<b>Longitude</b>	<b>AMT</b>	<b>AP</b>	<b>Stand age</b>	<b>Years</b>
				<b>(°C)</b>	<b>(mm)</b>	<b>(years)</b>	
Howland	Forest	45.20 °N	68.74°W	6.6	523-	95-140	2000-2004
	(main tower)				1032		
Duke Pine Forest		35.97°N	79.10°W	15.5	1145	17	2003-2005
Howland	Forest	45.21°N	68.75°W	6.5	523-	109	2000-2004
	West Tower				948		
North	Carolina	35.8°N	76.67°W	15.5	1320	18	2005-2006
	Loblolly Pine Forest						

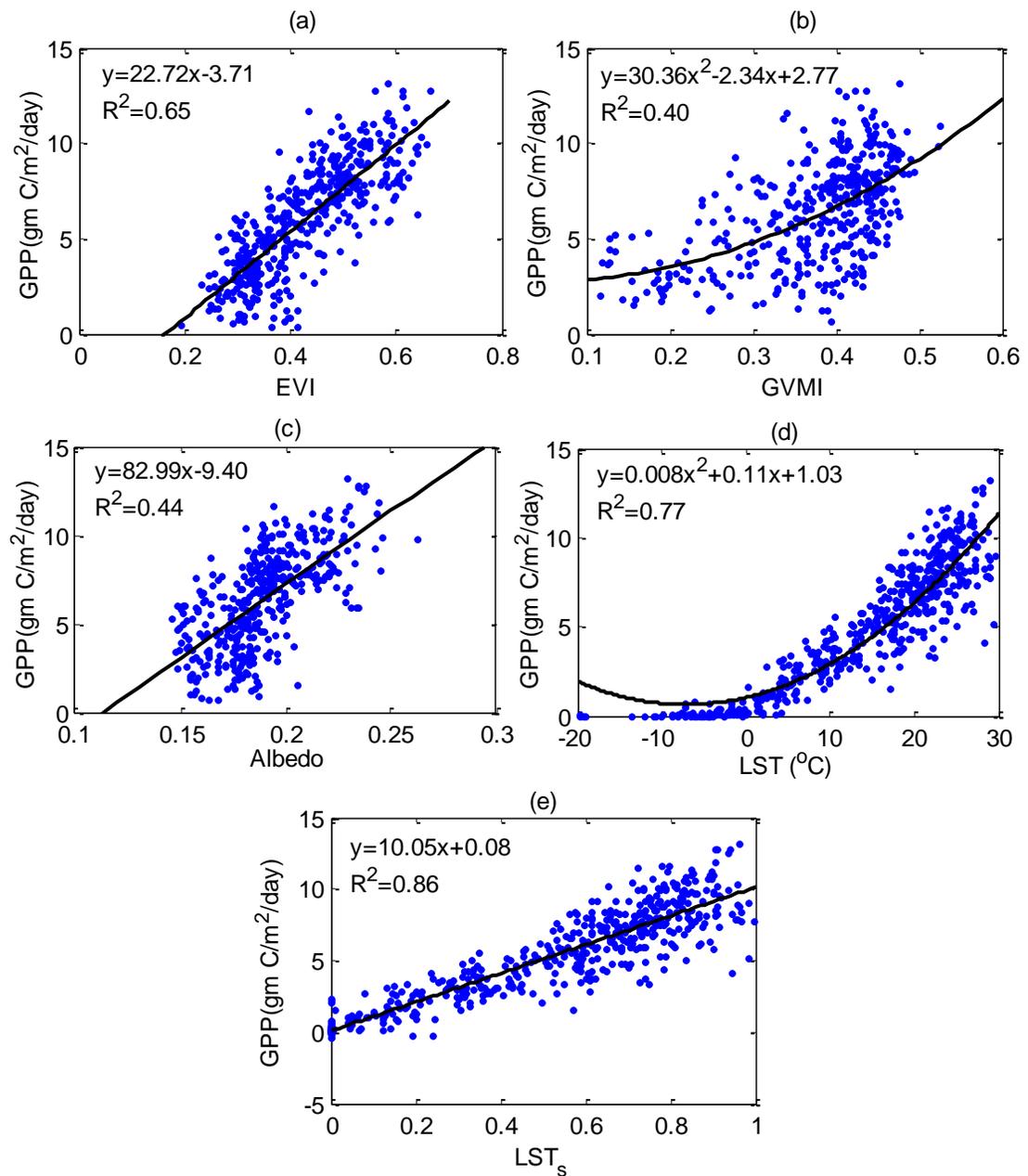


Figure 5.1. Relationship between EC towers estimated GPP and (a) enhanced vegetation index (EVI), (b) Global vegetation moisture index (GVMI), (c) near infrared albedo, (d) land surface temperature (LST) and (e) scaled LST ( $\text{LST}_s$ ) for all the sites. All point represent 8-day mean.

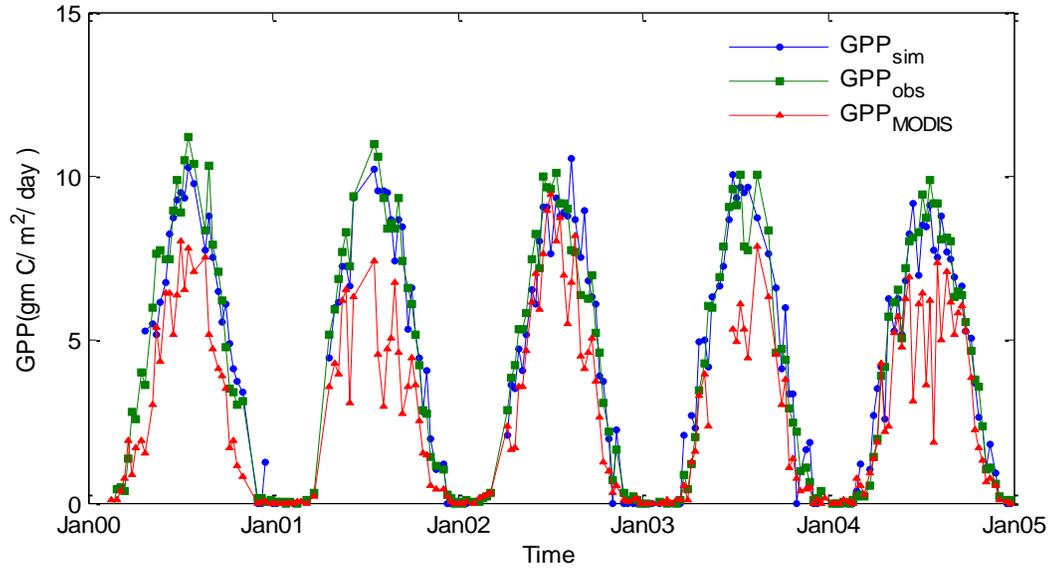


Figure 5.2. Annual cycle of eddy covariance tower estimated gross primary production ( $GPP_{obs}$ ), R-GPP-Coni model predicted GPP ( $GPP_{sim}$ ) and MODIS GPP ( $GPP_{MODIS}$ ) product for the Howland forest site during the calibration period (2000-2004). All points represent 8-day means.

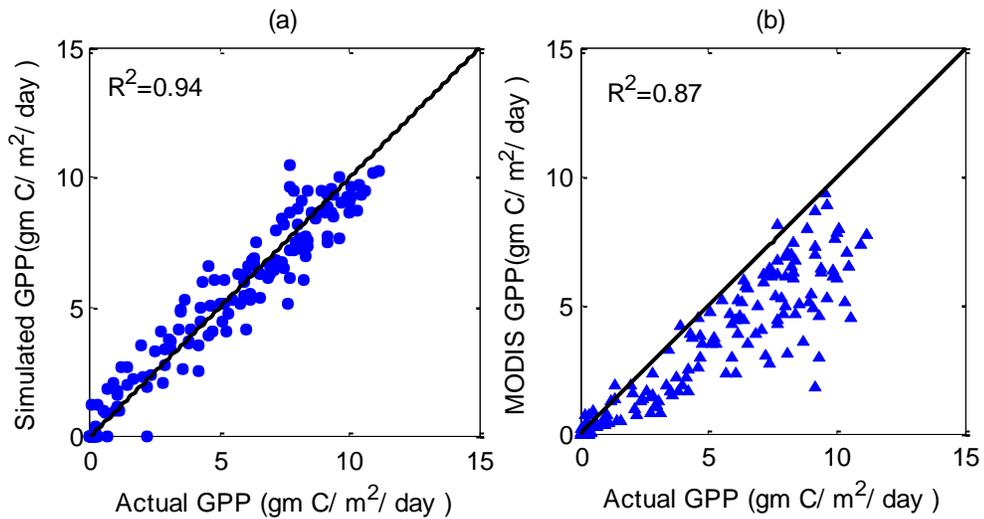


Figure 5.3. Scatter plot of eddy covariance tower estimated GPP (Actual GPP) and (a) R-GPP-Coni model simulated GPP; and (b) MODIS GPP for the Howland Forest. All points represent 8-day means.

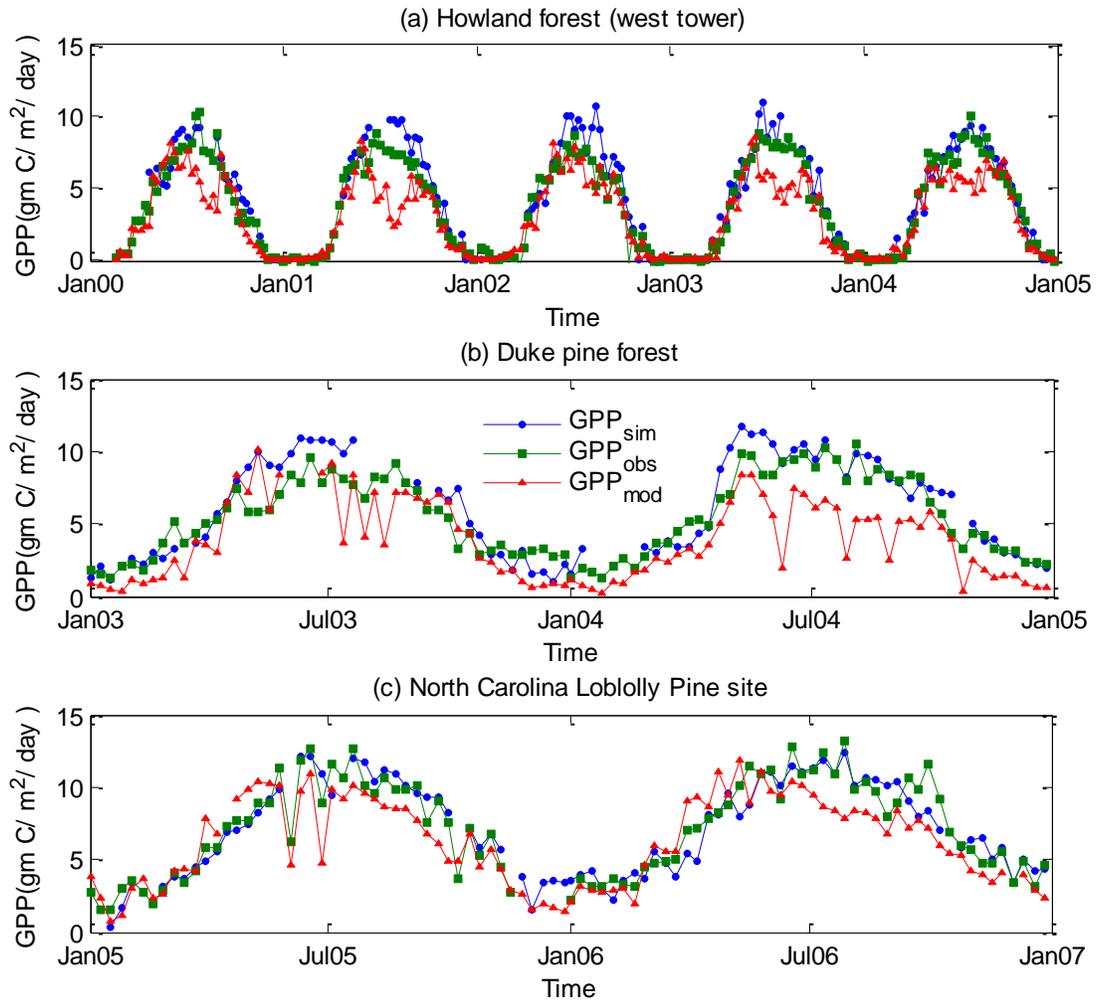


Figure 5.4. Annual cycle of eddy covariance tower estimated gross primary production ( $GPP_{obs}$ ), R-GPP-Coni model predicted GPP ( $GPP_{sim}$ ) and MODIS GPP ( $GPP_{MODIS}$ ) product for the (a) Howland forest west tower site, (b) Duke pine forest and (c) North Carolina Loblolly Pine site. All points represent 8-day means.

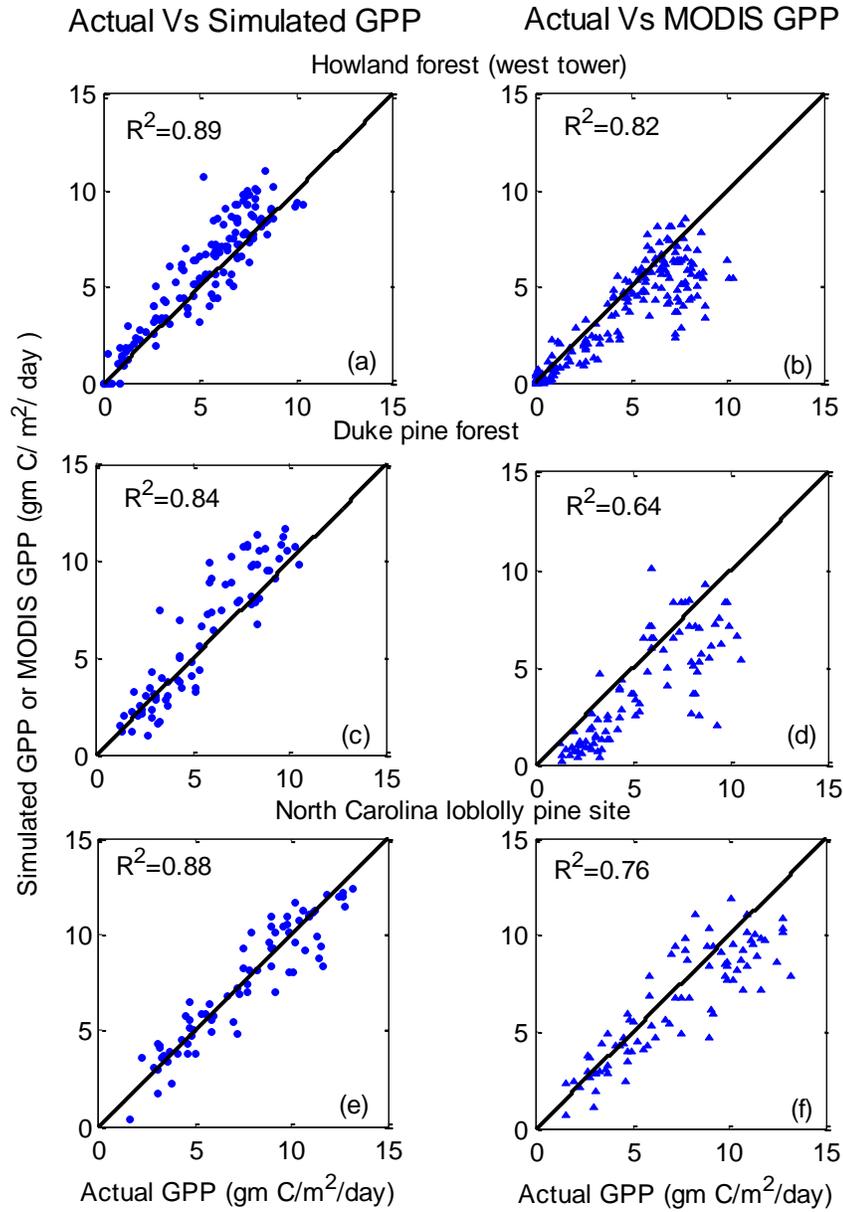


Figure 5.5. Left panel shows the scatter plot of tower estimated GPP (Actual GPP) and R-GPP-Coni model simulated GPP while right panel shows the scatter plot of eddy covariance tower estimated GPP (Actual GPP) and MODIS GPP.

# Chapter 6

## **An algorithm for estimating evapotranspiration for the deciduous and coniferous forests of north-eastern USA using MODIS data**

---

### **6.1 Introduction**

Evapotranspiration (ET) is the water lost to the atmosphere through soil evaporation and plant transpiration. A reliable estimate of the temporal and spatial variability of ET is important for the water resource management of regions where ET constitutes a major component of the water budget at regional scale, and where there are strong interactions and feedbacks between ET and other hydrological variables, such as soil moisture which is important for agriculture (Betts et al., 1997). Accurate ET estimation is also crucial for improving drought detection and assessment and for climate model simulations (McCabe and Wood, 2006). However, in general ET is highly variable spatially, further complicated by the heterogeneity of landscape and its dependency on many controlling factors such as climate, plant biophysics, soil properties, and topography (Mu et al., 2007; Friedl, 1996). Generally, because of the high spatial variability of ET, the spatial interpolations of point ET values are subjected to errors, even for homogeneous vegetation covers (Guyot et al., 2009). Therefore, such ET

estimates may not be reliable input for spatially distributed hydrological modeling and for decision makings in managing water resources (Jiang et al., 2009). Typically, for watersheds of scales ranging from say, a 100 to several thousand km<sup>2</sup> in area, ET are mostly estimated using standard ET models of Penmann-Monteith, Thornthwaite and Priestley and Taylor, even though assumptions associated with these models may not be representative of the real situations. Sometimes, these ET equations are modified with different theoretical basis to generate regional scale ET patterns (Cleugh et al., 2007; Nishida et al., 2003; Fisher et al., 2008).

During the last decade, due to the development of newer computation methods, the accuracy of regional scale ET estimation has improved, but there are still errors and uncertainties associated with each method (Drexler et al., 2004; Nagler et al., 2005). The eddy covariance (EC) method is generally regarded as one of the most accurate small scale, 0.1 to 1 km, methods of estimating ET (Nagler et al., 2005). In recent years, flux data collected from EC tower sites set up over major ecosystems, have provided valuable field measurements to parameterize and to validate ET models and to gain understanding about the factors controlling the seasonal dynamics of water fluxes. However, ET estimated from EC towers sometimes suffers from the lack of energy balance closure (Nagler et al., 2005; Zhang et al., 2009b), i.e., the sum of latent heat flux (LH) and sensible heat flux (H) measured by the EC method is sometimes different from the difference between net radiation ( $R_n$ ) and ground heat flux (G) (Wilson et al., 2002). The

reasons behind such discrepancies are still not fully understood (e.g., Brotzge and Crawford, 2003; Twine et al., 2000; Nagler et al., 2005).

The advent of satellite technology has inspired many researchers throughout the world to combine remotely sensed (RS) and meteorological data to estimate regional scale to global scale ET (Gillies and Carlson, 1995; Nishida et al., 2003; Mu et al., 2007). The most common approaches to model ET using RS data are empirical model that relate ET to vegetation index (VI), land surface temperature (LST), etc., (Zhang et al., 2009a), or surface energy balance models (Bastiaanssen et al., 1998, Cleugh et al., 2007). However, operational applications of such RS data driven surface energy balance models are limited by the requirement of meteorological forcing data, aerodynamic and surface resistances parameters that are mostly not readily available (Cleugh et al., 2007; Zhang et al., 2009b). Some ET models require so much climate and soil data that they are either not available or not representative over large areas (Jiang et al., 2009; Fisher et al., 2008). For example, the triangle method for ET estimation (Gillies and Carlson, 1995; Nishida et al., 2003) uses the slope of surface temperature versus the Normalized Difference Vegetation Index (NDVI) to estimate the surface resistance to ET, following the idea of Nemani and Running (1989). However, this method requires a wide range of soil moisture data for bare soil (from dry to saturated) or the vegetation status for fully vegetated soil (from water stress to well-watered condition) to provide the necessary range of surface conditions (Zhang et al., 2009a).

Several researchers tried to extrapolate EC tower estimated, local scale ET to regional scale ET using RS data based statistical models. For example, Zhang et al. (2009a) related the enhanced vegetation index (EVI) to water use efficiency ( $WUE=ET/\text{gross primary production}$ ) and then estimated ET from WUE and gross primary production (GPP) for an old-growth temperate mixed wood forest. Nagler et al. (2005) developed a regression based model combining RS data with EC and Bowen ratio energy balance (BREB) flux site measurements to predict ET over the reaches of western U.S. rivers. These studies demonstrated the possibility of using statistical techniques to extrapolate EC tower estimated, local scale ET to regional scale ET.

In recent years, there have been attempts to develop solely RS based model for estimating the evaporative fraction, ratio of ET and available energy (Batra et al., 2006, Nishida et al., 2003). Batra et al. (2006) combined the RS based radiation estimate and evaporative fraction to compute ET. They also compared ET estimated from different satellites (Moderate Resolution Imaging Spectroradiometer (MODIS), NOAA16 and NOAA14) to the ground measurements over the Southern Great Plains with an overall root mean square error (RMSE) of 53, 51 and 56.24  $W/m^2$ , respectively. Sobrino et al. (2007) developed a simple algorithm for retrieving ET from NOAA-AVHRR data but their method seems to work well only if the RS data contain information about extreme surface temperature. Venturini et al. (2008) proposed a RS based ET

retrieval algorithm which estimated ET with a RMSE and bias of 33.89 and  $-10.96 \text{ W/m}^2$ , respectively, for the Southern Great Plains of USA.

## **6.2 Research objectives**

Given that RS data are more readily available than ground measurements and various uncertainties are associated with meteorological data, this study proposes a simple model to estimate ET solely from RS data. The proposed model offers a practical approach to estimate ET of fine spatial resolution and frequent temporal sampling at regional scale without requiring aerodynamic or surface resistances parameters. The objectives of our study are:

- i) To select RS variables that explains the variability of ET by adequately characterizing the physical and plant physiological environments,
- ii) To develop an ET model only based on RS data of the MODIS satellite (R-ET);
- iii) To validate and to check the transferability of the proposed R-ET model to other similar ecoregions.

## **6.3 Study sites and data sets**

### **6.3.1 Study sites**

In this study, ET estimated from nine EC towers (Table 6.1, Figure 6.1) located in different forest sites were used to calibrate and validate the model. Among the 9

sites, 6 are deciduous and 3 are coniferous forests. The 6 deciduous forest sites represent a variety of species composition and climate. Precipitation among the sites varies from 750 to 1300 mm. The Harvard Forest EC tower within the Harvard Forest, Massachusetts, USA, is one of the sites with the longest running towers in the world since 1991 (Goulden et al., 1996; Urbanski et al., 2007). It is a part of the Ameriflux network. The site primarily consists of red oak, red maple, black birch, white pine and hemlock (Goulden et al., 1996). The Morgan Monroe State Forest (MMSF) is an extensive managed forest, primarily covered by secondary successional broadleaf forest. Dominant species are sugar maple, tulip poplar, white oak and black oak (Schmid et al., 2000). The University of Michigan Biological Station (UMBS) tower, located in Michigan, USA, is in the transition zone between mixed hardwood and boreal forests. Soils are mostly excessively well drained (Schmid et al., 2003). The forest was logged around 1880, and the area was disturbed repeatedly by fire until 1923 which essentially burned the entire region. Mean stand age is 90 years. The Willow Creek Study Site is located in Wisconsin, USA and is dominated by sugar maple, American basswood, etc. Climate of this forest is northern continental with short and moist summer and cold winter. The forests around the tower were clear cut once in the late 1800s and later in 1933. The Bartlett Experimental Forest is a part of North American Carbon Program and dominated by red maple, American beech, paper birch, eastern hemlock. Average air temperature is 9.8°C in January and 19.8°C in July. The Duke Hardwood Forest comprises of mixed hardwood species with a mean canopy height of 25m. The stand is dominated by hickories, yellow poplar

and oaks. The regional climate of this deciduous site is characterized by warm and humid summers, and mild winters. The three evergreen sites also represent a variety of temperature regimes, species composition and age (Table 6.1). The Howland forest is located within the boreal-northern hardwood transition zone and the average stand height forest is about 20 m. The site is dominated by conifers such as red spruce, eastern hemlock, balsam fir, white pine, and northern white cedar, and hardwoods such as red maple and paper birch. The soils of this forest are generally glacial tills and climate is mainly cold and humid. The forest was selectively logged around 1910 but has been minimally disturbed since that time. The Howland forest west tower is located at 775 m distant from the Howland forest main tower. This distance was chosen to ensure that the flux source regions, called footprints, of these two towers do not overlap. Within the 1km west tower footprint, the vegetation is 90% evergreen needleleaf (hemlock and spruce) and 10% broadleaf deciduous. The Duke Forest Loblolly Pine Site is located in a planted pine forest which was established in 1985 after a clear cut and burn. The mean temperature is 9°C in winter and 21.7°C in summer. Besides variation in climate and species, the selected sites also include a wide range in the time since disturbance, from recently planted plantation stands (Duke pine forest) to mature forests (UMBS and Howland forest sites). Moreover, one site, the Howland forest west tower site, has received treatment (nitrogen fertilization).

### **6.3.2 ET and remotely sensed (RS) data**

All the ET and MODIS RS data, except albedo (details given in Section 6.4.2), used in this study are values averaged over an 8-day period and were downloaded from the Ameriflux website (<http://public.ornl.gov/ameriflux/>). Among 36 spectral bands of MODIS data, seven spectral bands are mainly designed for monitoring terrestrial vegetation and land surfaces: blue (459–479 nm), green (545–565 nm), red (620–670 nm), near infrared (NIR) (841–875 nm, 1230–1250 nm) and shortwave infrared (SWIR) (1628–1652 nm, 2105–2155 nm). MODIS daily surface reflectance products are subjected to radiometric corrections, cloud-filter, atmospheric corrections for molecular scattering, ozone absorption and aerosols. In this study, the 8-day surface reflectances (MOD09A1, Collection 5) of the four spectral bands, blue, red, NIR (841–875 nm), and SWIR (1628–1652 nm) were collected to calculate the enhanced vegetation index (EVI) and global vegetation moisture index (GVMI) for the study period (Table 6.1) of each site.

The other MODIS data used in this study include the 1 km land surface temperature (LST) (MOD11A2, collection 5) and leaf area index (LAI) (MOD15A2, collection 5). LST is obtained using the Split-Window algorithm and the thermal IR bands of MODIS (Wan and Dozier, 1996). Using the spectral information of MODIS surface reflectances, the MODIS LAI algorithm estimates the LAI, the number of equivalent layers of leaves relative to a unit of ground area. We also collected the Bidirectional Reflectance Distribution Function (BRDF) coefficients (MCD43A1) of MODIS which is produced every 8 days with 16 days of acquisition (Details given in Section 6.4.2). Actual albedo for the

NIR and SWIR bands (0.7– 5 and 0.3–5.0  $\mu\text{m}$ , respectively) were then calculated from the BRDF as a function of optical depth, solar zenith angle, and spectral bands (Schaaf et al., 2002; <http://daac.ornl.gov/MODIS/MODIS-menu/MCD43.html>).

The GPP of a deciduous forest was estimated from EVI, GVMI, LST and NIR albedo using the R-GPP model (Jahan and Gan, 2009) which gave very promising results in modeling the 8-day GPP of the Harvard forest, MMSF and UMBS forest sites. The R-GPP model (details given in Section 6.4.4) predicted GPP with  $R^2$  ranging from 0.92 to 0.95 and its performance was better than that of the MODIS GPP ( $R^2$  ranging from 0.74 to 0.91). The GPP of coniferous forests were estimated using the R-GPP-Coni model (Jahan and Gan, 2011, submitted) which was also calibrated and validated for different coniferous sites of USA (details given in Section 6.4.4). This model explained 88% to 94% of the observed GPP variability for 4 coniferous sites located in different parts of USA and also outperformed the primary RS-based GPP algorithm of MODIS.

In this study, our goal was to compute ET at 1 km resolution using RS data averaged over a 1 km $\times$ 1 km area within which each EC tower is located instead of using RS data averaged over of a N $\times$ N km area (N=3 or 5, or even larger number) which is more commonly used. The rationale for choosing the 1 km $\times$ 1 km area for the input variable is that the footprint of the EC tower, which depends on the flux tower height, wind speed, topography, etc., is usually of few hundred

meters to 1 km in size (Schmid, 2002; Xiao et al., 2004). Therefore RS data extracted from  $N \times N$  MODIS pixels may not be appropriate to represent a flux tower footprint and may lead to error. To prepare 1 km resolution RS data (LST, LAI), we either used the digital value of a single pixel within which the EC tower is located (in case of 1 km resolution RS data) or the average of  $2 \times 2$  pixels (in case 500 m resolution RS data) which represents the same area.

#### **6.4 R-ET model predictors**

ET is controlled by both climatic (e.g., temperature, vapor pressure deficits (VPD), soil moisture, wind, etc.) and plant physiological factors (e.g., vegetation type, leaf area, canopy conductance, etc.). In forested ecosystems, transpiration is the more dominant component of ET than soil evaporation (Daley, 2007; Oishi et al., 2008) because bare soil occupies relatively small areas in many forested ecosystems and so soil evaporation from a forest floor contributes minor energy fluxes since turbulent eddies and solar radiation cannot fully penetrate the dense and tall vegetation (Nishida et al., 2003). Stoy et al. (2006) found that during the growing season, ET was linearly correlated with the product of climatic (precipitation and VPD) and the biological drivers (LAI and stomatal conductance) for three adjacent ecosystems of Southeastern USA. Therefore it is necessary to carefully select RS predictors which can account for climatic factors and plant physiological stresses that control ET.

EVI is an advanced vegetation index which produces more sensitive vegetation signal than the popular vegetation index NDVI in high biomass regions (Zhang et al., 2004).

$$EVI = G \frac{NIR - R}{NIR + C_1 R - C_2 B + L} \quad (6.1)$$

where NIR, R and B are atmospherically corrected surface reflectances in the near-infrared, red and blue bands respectively, G is the gain factor, L is the canopy background adjustment factor and  $C_1$  and  $C_2$  are the coefficients of the aerosol resistance term, which uses the blue band to correct the aerosol influences in the red band. In the EVI algorithm (Equation 6.1),  $L=1$ ,  $C_1=6$ ,  $C_2 = 7.5$ , and  $G = 2.5$ . Structural changes of canopy including change in leaf area, stomatal conductance, canopy capacity, surface roughness, canopy albedo, and canopy heat capacity, etc. influence transpiration and hence the ET. These canopy properties vary with the phenological cycle of vegetation (Lawrence and Slingo, 2004) and EVI has been shown to be a good predictor of this vegetation phenology and vegetation condition (Zhang et al., 2004). Past studies (Jahan and Gan, 2011) also showed that EVI possesses good correlation with potential ET and some other climate variables (such as precipitation, temperature, aridity index) which control ET. Therefore EVI was used as a predictor in some previous ET estimation models. As for example, Mu et al. (2007) used MODIS EVI and meteorological data to predict ET at the continental scale, and achieved a  $R^2 = 0.76$  based on ET measured at 19 Ameriflux towers. Yang et al. (2006) also combined EVI and meteorological data to predict ET using a machine learning program at 19 towers sites of USA with a  $R^2 = 0.75$ . Wang et al. (2007) used EVI,

air temperature and surface net radiation to predict ET in the Southern Great Plains of USA with  $R^2 = 0.83$ . In this study we also examined the correlation between EVI and ET. Results (Figure 6.2 (a) and 6.3 (a)) show that the dynamics of ET agree reasonably well with the EVI variation for both deciduous ( $R^2 = 0.64$ ) and evergreen sites ( $R^2 = 0.52$ ). However, in this study instead of using EVI, we have used GPP as model predictor which contains the information of EVI but possesses some additional advantages as demonstrated by the higher correlation between GPP and ET (Details Section 6.4.4).

#### **6.4.1 Land surface temperature (LST)**

LST is a potential predictor for ET because it is strongly correlated with VPD, soil and canopy temperature and thus can incorporate the effects of these variables on ET (Hashimoto et al., 2008). High VPD causes plants to decrease their stomatal apertures to prevent excessive water loss through transpiration and thus it controls ET. Boegh et al. (1998) found the slope of the LST/NDVI curve to be related to the ET of Sahel. LST is also related to the stomatal conductance (Nemani and Running 1989) and the surface bowen ratio (Goward and Hope, 1989), etc., which are also key factors that affect ET. Many studies have employed LST to compute soil moisture (Kimura, 2007; Goetz, 1997, Goward et al., 2002) which controls both the soil evaporation and the transpiration. Low soil temperatures increase the viscosity of water, hence decrease soil and plant hydraulic conductance and thus the amount of water uptake (e.g. Kramer and Boyer, 1995; Mellander et al., 2004). These effects in turn hamper the

production of new roots and consequently reduce the amount of transpiration. So it is expected that using LST as a predictor will help us to incorporate the effects of temperature, VPD, soil moisture and stomatal conductance on ET.

In this study, we have scaled the LST ( $LST_s$ ) with respect to the maximum ( $LST_{max}$ ) and the minimum LST ( $LST_{min}$ ):

$$LST_s = \left\{ \frac{LST - LST_{min}}{LST_{max} - LST_{min}} \right. \quad (6.2)$$

where LST is the observed LST. From preliminary data analysis,  $LST_{min}$  is set to 243 Kelvin (-30° C) while  $LST_{max}$  to 303 Kelvin (30°C). The scatterplots of ET as a function of LST for all the study sites, shown in Figure 6.2 (b) and 6.3 (b), demonstrate that ET increases linearly with LST. The relationship between them is strong for both the coniferous ( $R^2=0.76$ ) and the deciduous forests ( $R^2=0.64$ ). Wang et al. (2007) also found similar strong relationships between ET and LST for the Southern Great Plains area of USA.

#### **6.4.2 Short wave albedo ( $\alpha_{sw}$ )**

Albedo ( $\alpha$ ) is the fraction of incident solar radiation that is reflected from the surface. It influences the radiation absorbed by plant canopies and thereby it affects photosynthesis, energy balance, evapotranspiration, respiration (Wang et al., 2001), rainfall, drought and vegetation growth (e.g., Bounoua et al., 2000; Wang and Davidson, 2007). The albedo of a typical vegetated surface shows

temporal variation because of the seasonality of plant phenology (Song, 1998), and due to seasonal variations of surface soil wetness (Lawrence and Slingo, 2004). Various studies reported the important role of albedo in relation to gross productivity (Jahan and Gan, 2009), surface temperature and precipitation (Dickinson, 1996).

Net radiation ( $R_n$ ) has been used for modeling ET in many studies (Mu et al., 2007).  $R_n$  is given as

$$\begin{aligned}
 R_n &= S_{in} - S_{out} + L_{in} - L_{out} \\
 &= S_{in} (1 - \alpha) + L_{in} - L_{out}
 \end{aligned}
 \tag{6.3}$$

Where  $S_{in}$  and  $S_{out}$  are the incoming and outgoing shortwave solar radiation;  $L_{in}$  and  $L_{out}$  are the downward and upward longwave radiation at the surface, respectively. Many studies on regional scale land surface radiation balances have relied on the data of geostationary satellites for  $R_n$ , which provide data with fine temporal resolutions (less than 3 h) but of coarse spatial resolutions of 20–250 km scale (Harries et al., 2005; Jacobs et al., 2002). Coarse resolution radiation data are available from models such as the GMAO (Global Modeling and Assimilation Office) and the ISLSCP-II (the International Satellite Land Surface Climatology Project Initiative II). The GMAO  $R_n$  data, available at  $1.00^\circ \times 1.25^\circ$  resolution, combines all globally available surface weather observations to produce an estimate of climatic conditions for the world. On the other hand ISLSCP is a compilation of RS data, complementary and supplementary ground measurements (Los et al., 2000). The problem with such reanalysis data is that their accuracy

depends on the availability of the meteorological input data from the observation networks which may not be sufficient and so extrapolation from a limited number of ground stations may affect the reliability of these data (Mu et al., 2007; Heinsch et al., 2006). In addition, the grid scale of the reanalysis data is usually too coarse to be combined with finer resolution RS data. At scale of 1 km resolution, some studies have relied on either the field measurements of solar radiation (Boegh et al., 2002) or on meteorological data (Su et al., 2005). The problem of using measured  $R_n$  is that the availability of reliable  $R_n$  data is usually limited (Alados et al., 2003). Moreover, extrapolation of point measurements over large watersheds may incur unpredictable errors (Guyot et al., 2009). On the other hand, reliable  $R_n$  data derived entirely from RS data at moderate resolutions (e.g., 1 km) are also rare (Ryu et al., 2008). Therefore, instead of using measured or climate model simulated  $R_n$ , we have used the short wave (SW) albedo from MODIS satellite which is an important part of the SW surface energy exchange (Ollinger et al., 2008).

Indices based on shortwave radiation data are generally reliable for assessing the vegetation condition and dynamics when the vegetation cover is moderate to dense (e. g., forest) (Huete et al., 1985). Therefore, short wave albedo,  $\alpha_{sw}$  (0.3 to 5  $\mu m$ ) has been used to assess the radiation status in our proposed model. Moreover  $\alpha_{sw}$  is the most commonly used albedo in ecosystem modeling (Wang and Davidson, 2007; Ollinger et al., 2008). In this study, we have used MODIS  $\alpha_{sw}$  which is produced every 8 days but averaged over a 16-day cycle (e.g.,

albedo of Date 1 represents the average albedo of day 1 to 16 while albedo of Date 9 represents the average albedo of Date 9 to 24). To estimate the ET of any 8 day period we used  $\alpha_{sw}$  averaged over that particular 8-day and the previous 8-day while the other RS predictors were averaged over that particular 8-day only.

Figures 6.2 (c) and 6.3 (c) show that  $\alpha_{sw}$  and ET are reasonably correlated to each other for both coniferous and deciduous sites ( $R^2=0.51$  and 0.49) which indicates that  $\alpha_{sw}$  can be a potential predictor of ET. During winter, because of snow cover, albedo becomes very high and therefore those albedos have been excluded in Figure 6.2 (c) and 6.3 (c). However, there exists a different relationship between winter time albedo and ET as winter ET is mainly governed by snow sublimation due to radiation (Hadley et al., 2008). Therefore in this study we have used separate parameterization schemes for computing the winter ET (Details given in Section 6.5).

#### **6.4.3 Leaf area index (LAI)**

LAI represents the surface area of the plants' canopies for transpiration and is an indication of a plant's biophysical capacity for absorbing solar energy. This biophysical capacity is related to the plant's physiological capacity to transpire. Under limited water conditions, plants reduce their LAI by leaf rolling/shedding and could even undergo chronic stomatal closure under prolonged droughts (high VPD) (Mailhol et al., 1997). Stoy et al. (2006) found that during growing seasons ET is controlled by LAI in three adjacent ecosystems: an abandoned agricultural

field, an early successional planted pine forest, and a late-successional hardwood forest of Southeastern USA. Mu et al. (2007) used LAI to compute the canopy conductance which plays a dominant role in transpiration. Mailhol et al. (1997) and Fisher et al. (2008) computed the net radiation available to soil ( $R_{ns}$ ) for evaporation as function of LAI and then determined the radiation available to canopy by deducting  $R_{ns}$  from the total  $R_n$ . Since transpiration is more dominant than soil evaporation when LAI is high (Mu et al., 2007), LAI is often one of the governing variables to simulate ET for vegetated surfaces in physically-based hydrological models (Arora, 2002).

In this study, we also examined the relationship between LAI and ET. Results (Figures 2 (d) and 3 (d)) show that the dynamics of ET agree reasonably well with the LAI variation for both forest types. However, the overall relationship between LAI and tower ET was stronger for the deciduous forests ( $R^2 = 0.70$ ) than for the evergreen forests ( $R^2 = 0.58$ ). It should be mentioned that the results in terms of  $R^2$  also varied among sites, even within the same forest type. In winter, transpiration from conifers, even with their needles intact, is generally very small due to the dormancy of vegetation and significant low-temperature constraints which cause the overall ET to decrease. Therefore, during winter leaves of conifers (i.e. LAI) do not contribute much to transpiration or ET and the overall  $R^2$  of conifer LAI-ET relationship drops when both the summer and winter seasons are considered together.

#### **6.4.4 Gross primary production (GPP)**

For forested ecosystems, transpiration and photosynthesis are closely related (Zhang et al., 2009a; Freitag et al., 2008) because, the release of water due to transpiration is proportional to the carbon fixation due to photosynthesis (Freitag et al., 2008). Various studies have reported the strong relationship between GPP and ET. Law et al. (2002) investigated the seasonal variation of GPP and water vapor exchange across many forests, grasslands, crops, and tundra sites and found that GPP generally increases with ET, with  $R^2$  of 0.78 for deciduous broadleaf and  $R^2$  of 0.58 for evergreen forest. McCaughey et al. (2006) also found a strong relationship ( $R^2 = 0.72$ ) between GPP and ET for a boreal mixed wood forest of Ontario, Canada.

In this study we found that the  $R^2$  between the EC tower estimated GPP and ET for the deciduous (Figure 6.2 (e)) and coniferous forests (Figure 6.3 (e)) were 0.82 and 0.78, respectively. The strong relationship is partly due to the similarity in the seasonal cycles of ET and GPP. During the growing season ET gradually increases with GPP because of a general increase in radiation, LST, and growing vegetation. On the other hand, during most autumn, winter and spring, both GPP and ET are low because of low air temperatures, low incident solar radiation which affect the stomatal conductance, and, consequently, the photosynthesis and transpiration of both coniferous and deciduous forests (Sims et al., 2008). Similar observations were also reported for a boreal mixed wood forest of Canada (McCaughey et al., 2006) and for a red oak dominated deciduous and a hemlock

dominated coniferous forests, located in the Harvard forest of USA (Hadley et al., 2008).

As the objective of this study was to develop a solely RS based ET estimation model, therefore instead of using EC tower estimated GPP, we estimated GPP from the R-GPP (for deciduous forest) and R-GPP-Coni models (for coniferous forest). In the R-GPP model, GPP was estimated (Equation 6.4) from the Global Vegetation Moisture Index (GVMI) (Ceccato et al., 2002), scaled LST (scaled  $LST = LST / LST_{max}$  when  $LST > 0^{\circ} C$  or scaled  $LST = 0$  when  $LST \leq 0^{\circ} C$ ), EVI and NIR albedo ( $Albedo_{NIR}$ ).

$$GPP = k_c \times GVMI^w \times (\text{scaled LST})^x \times Albedo_{NIR}^y \times EVI^z \quad (6.4)$$

Where  $k_c$  is a scalar and  $w$ ,  $x$ ,  $y$ , and  $z$  are the exponents. The optimized values of  $k_c$ ,  $w$ ,  $x$ ,  $y$ ,  $z$  are 114, 0.885, 1.05, 0.695 and 0.933, respectively. The R-GPP-Coni (Jahan and Gan, 2011) model predicted the GPP of coniferous forest using the same four predictors, however with different coefficient values ( $k_c$ ,  $w$ ,  $x$ ,  $y$ ,  $z$  are 38.31, 0.396, 0.673, 0.332 and 0.603, respectively). The GPP simulated by the R-GPP and R-GPP-Coni models are strongly related to the 8-day ET for both coniferous and deciduous forests with a  $R^2=0.77$  (Figure 6.2 (e) and 6.3 (e)) which is almost the same as the  $R^2$  between EC tower measured GPP and ET. Therefore the R-GPP model simulated GPP can potentially replace ground measured GPP as a predictor of ET.

## 6.5 Remotely sensed-ET (R-ET) model

Given that LST and albedo characterize relevant physical factors (e.g., energy, temperature, VPD, soil moisture) while LAI and GPP characterize key plant physiological factors (vegetation condition and phenology) that control ET, we therefore propose a R-ET model (Equation 6.5) based on four RS predictors:

$$ET = k \times \alpha_{sw}^a \times LST_s^b \times GPP^c \times \exp(d * LAI) + m \quad (6.5)$$

Where  $k$  and  $m$  are scalars, and  $a$ ,  $b$ ,  $c$  and  $d$  are exponents. These model parameters were estimated for coniferous and deciduous forests separately. The estimated ET from the EC tower located at the Duke hardwood forest (2001-2005) and MMSF (2000-2005) sites were used to calibrate coefficients for the deciduous forest (called deciduous R-ET model) while the ET data of Howland forest (2000-2004) was used to calibrate the coefficients for the coniferous forest (called Coniferous R-ET model). A nonlinear optimization scheme, the Generalized Reduced Gradient (GRG2) (Lasdon, 1978; Spaulding, 1998) was used to calibrate the model (Equation 6.5) and the optimized values for  $k$ ,  $a$ ,  $b$ ,  $c$ ,  $d$  and  $m$  are 47.566, 0.211, 2.474, 0.333, 0.078 and 3.475, respectively, for the deciduous R-ET model. For the coniferous R-ET model, the optimized values are 1183.708, 1.497, 2.094, 0.259, 0.022 and -0.5.866, respectively.

During winter, low soil temperatures increase the viscosity of water which prevents plants from replenishing their water supply, causing them to drastically decrease the transpiration and the photosynthesis. Weaver and Mogensen (1919) investigated 3 deciduous and 6 coniferous species and concluded that the

transpiration losses from conifers in winter are scarcely greater than those from defoliated stems of broad-leaved trees under uniform soil types, texture, soil temperature, and identical aerial environment. They also concluded that the decrease in transpiration from broad-leaved trees resulting from defoliation in autumn is gradual, and is similar to the decrease in transpiration of coniferous forests. Hadley et al. (2008) also reported a similar transpiration behavior for a red-oak dominated deciduous forest and a hemlock dominated coniferous forest. They concluded that during winter, soil evaporation and snow sublimation mainly contribute to a forest's ET and so both a leafless deciduous forest and a leafy coniferous forest located in similar climatic regime are expected to produce similar amount of ET during winters. As transpiration from vegetation is minimal to almost zero during winter and ET is mainly controlled by radiative variables in that period, the proposed R-ET model for computing winter ET excluded predictor variables which are mainly responsible for transpiration:

$$ET = k_w \times \alpha_{SW}^{a_1} \times LST_s^{b_1} + m_1 \quad (6.6)$$

where,  $k_w$  and  $m_1$  are the scalars and  $a_1$  and  $b_1$  are the exponents. Winter ET data of Duke hardwood and MMSF forests have been used to parameterize the model (Equation 6.6) for deciduous species. By GRG-2, the calibrated values of  $k_w$ ,  $m_1$ ,  $a_1$  and  $b_1$  are 24.36, -0.3, 2.98 and 2.89, respectively. For the coniferous forest, the values obtained are 233.54, 0.10, 6.26 and 0.38, respectively. Previous studies (Sims et al., 2008; Jahan and Gan, 2009) reported that vegetation activity declines when LST is less than 0° C which therefore was used to define the inactive winter

period in their studies, and also in this study for both coniferous and deciduous forest.

## **6.6 Discussions of results**

The model predicted ET was compared against the ET measured at EC towers. We did not attempt to gap fill the RS data because our objective was to be able to predict ET when reliable RS data were available. The calibrated model was validated against four other deciduous forest and two other coniferous forest sites.

### **6.6.1 Calibration of R-ET model**

The calibration results show that the R-ET model could capture the seasonal dynamics of the observed ET reasonably accurately for both deciduous calibration sites (Figure 6.4). The  $R^2$  were 0.90 and 0.85 (RMSE were 12.40 and 14.95  $W/m^2$ ) for the Duke hardwood forest and MMSF, respectively. The model estimated the winter season ET very accurately for both sites although some discrepancies were noticed during the growing season, especially for the MMSF forest site. The model did not simulate the peak ET accurately probably because of the saturation problem of RS data (Yang et al., 2006, Myneni et al., 2002). In case of densely vegetated forests, the reflectance from lower leaf layers can be obscured by that of upper leaf layers and so the reflectance data for medium dense and highly dense vegetation may be very similar to each other. This problem of reflectance saturation in very dense vegetation might also affect other vegetation

indices (e.g., EVI and GVMI) which were computed from reflectance and might cause the performance of the R-ET model to drop.

For the Duke forest, the model simulated summer ET agreed closely with the EC tower estimated ET. However, the simulated peak ET was marginally lower than the actual peak ET during the summer 2003 which was a very wet year. The growing season precipitation (P) was 790mm in 2003 which was about 1 standard deviation ( $\sigma$ ) above the long-term (111 year) mean growing season P of  $632 \pm 132$  mm (Stoy et al., 2006). This above normal P of 2003 caused both evaporation and transpiration to increase during that growing season (Stoy et al., 2006) which was not captured by the R-ET model. However, the model was successful in capturing the ET dynamics during the mild and severe drought years of 2001 and 2002 when the precipitation was  $1 \sigma$  and  $2 \sigma$  below the long term average, respectively.

At the calibration stage, the R-ET model was also able to estimate ET of the coniferous Howland forest with a  $R^2$  of 0.85 and a RMSE of  $10.89 \text{ W/m}^2$  for 2000-2004 (Figure 6.6 and 6.7). With respect to the EC tower estimated ET, it is clear that the results obtained from the R-ET model is promising for this coniferous site.

### **6.6.2 Validation and transferability of R-ET model**

We validated and tested the transferability of the calibrated R-ET model by applying it to other sites of similar ecosystems located at northeastern USA. As

only limited numbers of EC towers have so far been established throughout the world and their measurements are representative only over the tower footprint, therefore it will be beneficial if the proposed R-ET model can be used to upscale the EC tower measurements from small scale to regional scale i.e. to estimate the ET of other similar types of forest. In general, for forests located in similar climatic regimes, the degree of transferability will depend on different factors such as similarity in vegetation types and capability of the explanatory variables ( $\alpha_{sw}$ , EVI, LST<sub>s</sub>, LAI, GPP) to represent the physical (e.g, moisture condition, surface temperature, VPD, etc.) and the biological (leaf area, vegetation condition, etc.) drivers of ET.

The deciduous R-ET model was validated against 4 other deciduous sites while the coniferous R-ET model validated against 2 other coniferous sites. The model simulated ET were fairly accurate for all the 4 deciduous validation sites with R<sup>2</sup> of 0.85, 0.82, 0.92 and 0.81, and RMSE of 13.24, 14.44, 12.13 and 14.13 W/m<sup>2</sup> for the Bartlett experimental site (2004-2006), Willow creek (2004-2005), UMBS (2000-2001) and Harvard forest site (2000-2003), respectively (Figure 6.4 and 6.5). In few occasions, discrepancies were noticed between the simulated and the observed peaks, which could be partly due to the limitations and simplicity of the R-ET model and partly because of possible errors in the observed ET.

The R-ET model calibrated against the coniferous Howland forest EC tower data was validated against the Duke Pine forest and the Howland forest west tower

data. At both sites, the multiple annual time series plots (Figure 6.6) of the measured and simulated ET showed good agreements in phase and in changes in magnitude. The  $R^2$  between the simulated and observed ET were 0.84 and 0.83 while the RMSE were 14.38 and 11.60  $W/m^2$  for the Duke pine and Howland forest west tower sites, respectively (Figure 6.6 and 6.7). The Howland forest west tower site is a managed site which receives nitrogen treatment. Each year 18 kg of dissolved nitrogen fertilizer is applied over a 21 ha area around the tower by helicopter (Gaike et al., 2007). On the other hand Duke pine forest is an 'early successional' forest which was established after a clear cut and burning in 1983 which affected the atmospheric  $CO_2$  and nitrogen deposition of that forest (Stoy et al., 2006). The R-ET model was found to be successful for both of these forests: one subjected to treatment and another subjected to disturbance. Another thing should be noted that the above-average precipitation in 2003 increased the ET of Duke hardwood forest in that year in comparison with the ET of other years when precipitation was below-normal (2001 and 2002) to normal (2004). But at the Duke Pine forest, the effect of 2003's above-average precipitation was not noticed in its corresponding ET (Figure 6.6). This is probably because of the ice storm that occurred in December 2002. This storm damaged the pine needles causing a large reduction in LAI and consequently reduced the transpiration in 2003. The increase in ET in 2003 due to high precipitation was nearly equal to that reduction in transpiration due to reduced LAI after the ice storm and thus the overall ET was not noticeably different from the other years (Stoy et al., 2006). It is worth

noting that the R-ET model was able to capture this overall variation of ET subjected to the ice-storm damage and excessive precipitation.

Comparison with previous studies shows that the results of this study in terms of  $R^2$  and RMSE are comparable or even better in some cases. Jiang and Islam (2001) proposed a method for ET estimation using primarily NOAA-AVHRR data with minimum ground observation for the southern great plain (SGP) of USA. They used the linear NDVI-LST spatial variation method and reported a RMSE of  $85.3 \text{ W/ m}^2$  (29% of the mean ET) and  $R^2$  of 0.64. Batra et al. (2006) estimated ET from data of MODIS, NOAA16 and NOAA14 with a RMSE of 53, 51 and  $56.24 \text{ W/ m}^2$ , and a correlation coefficient of 0.84, 0.79 and 0.77, respectively. Nishida et al. (2003) developed an ET model driven by NOAA-AVHRR data and found a bias, RMSE and correlation coefficient (R) of  $5.59 \text{ W/m}^2$ ,  $45.06 \text{ W/m}^2$  and 0.86, respectively, for several Ameriflux sites spread over the continental United States. Norman et al. (2003) estimated ET by combining low (GOES) and high (aircraft) resolution ( $\sim 24 \text{ m}$ ) remotely sensed data and reported a RMSE of about  $40 \text{ W/m}^2$ . According to Seguin et al. (1999), the recommended ET retrieval accuracy varies according to application, but typically it is about  $50 \text{ W/m}^2$ . Jiang et al. (2004) showed that a reasonable upper limit to the accuracy of ET obtained from remote sensing methods should be about 20%. The accuracy of our simulated results meets these recommendations.

Given that the proposed R-ET model is validated for sites located far away (up to 1400 km) from the calibration sites (Figure 6.1), it seems that this proposed

model, using the four selected RS predictors, can generally estimate the ET of deciduous and coniferous forests located at north-eastern USA with sufficient accuracy for water resources management or hydrologic modeling purposes. The predictors of the R-ET model possess the capabilities to represent the basic climatic and ecological status of forest ecosystems. This is probably why the proposed R-ET model turns out to be effective in simulating the seasonal dynamics of ET of deciduous and coniferous ecosystems of north-eastern USA, given the variations in climate, species, age structures, management practice and time since disturbance.

### **6.6.3 Discussions on modeling ET of forest environments**

In recent years RS data based ET models have demonstrated strong potential in modeling the ET of various forest environments (Nishida et al., 2003; Jiang and Islam, 2001; Mu et al., 2007). The new aspect of the proposed model is that it is dependent solely on RS data whereas majority of the RS-based ET models (e.g., Fisher et al., 2008; Mu et al., 2007) require supplementary meteorological inputs which are often not available with sufficient spatial resolutions as that of the RS data and as a result may affect the reliability of those models in the estimation of regional ET (Heinsch et al., 2006). However, the robustness of the R-ET model should be further tested for different climatic regimes and surface conditions. We expect the need to refine the model parameters for other biomes, such as grassland and savannas.

Even though the overall performance of the proposed model was promising, occasionally there were large differences between the actual and modeled ET partly because of the simplicity and the limitations of the R-ET model and partly because of the possible errors in the observed actual ET. In this study we have used the gap-filled data and gap-filling of these data are still subjected to debates (Falge et al., 2001). Moreover most of the gap-filling methods have been evaluated against the measured net ecosystem exchange of CO<sub>2</sub> but not against the measured ET which further limit our ability to obtain reliable gap-filled estimates of actual ET (Mu et al., 2007). Another thing is that the methods used to gap-fill EC measured ET (Falge et al., 2001; Stoy et al., 2006) may not provide accurate estimation for potentially high evaporation from wet canopy and forest floor after rain events if relationships are derived from dry surface conditions. In this case we would expect the model to underestimate the ET following rain events when surface conductance is high (Oishi et al., 2008).

Estimating the growing season ET by the R-ET model can also be affected by the problem of saturation in RS data acquired by optical sensors. For dense canopies, the reflectance from the lower leaf layers can be masked by the upper layers and so the measured reflectances (as well as EVI and GVMi computed from them) become non-representative of actual vegetation. Use of non-representative EVI and GVMi can lead to erroneous ET estimation. Moreover, without considering soil moisture as one of the predictors in the R-ET model may incur some additional error in the simulated ET even though some studies argued that when

the soil moisture is above the wilting point the influence of soil moisture on ET is relatively small (Wang et al., 2007; Jaksic et al., 2006). This is especially true for the vegetation-covered areas. However in the long run, soil moisture does influence ET, especially during prolonged drought periods. Therefore, discrepancies between the actual and simulated ET is expected to generally increase under extreme soil moisture conditions.

Some studies found that the EC towers have surface energy balance closure problem due to complexities in wind variations, footprint representation and sampling variability (Wilson et al., 2002; Twine et al., 2000). Recently Glenn et al. (2008) reported that EC tower estimated ET might have an error of about 10–30% based on comparison of ET from multiple towers located at the same site, or by comparison with separate measurements of ET by other methods. Uncertainties associated with the non-closure of energy balance in the EC method will introduce uncertainties in ET estimation when the algorithm is based on the relationships between these tower measured ET fluxes and RS variables, which is our case. However, researchers are trying to employ effective correction procedures for the lack of energy balance closure (e.g. Griffis et al., 2004).

Another factor that may affect the accuracy of ET predicted by the R-ET model is the surface reflectance products derived from the 8-day composite images of MODIS. Currently MODIS 8-day reflectance data, used for EVI and GVMi computation, is composited on the basis of the clearest conditions over the 8-day

period but not the average condition (Xiao et al., 2004). Elimination of days with high and low reflectance values can cause an underestimation or overestimation of ET values, respectively. This problem can be resolved by using daily RS data as input to the R-ET model. But in this case the increase in computational effort must be justified against the improvement in estimation. Errors in ET may also result from the use of 16-day albedo products of MODIS.

## 6.7 Summary and conclusions

In this study, we have developed a remote sensing algorithm for modeling ET at 8-day intervals for coniferous and deciduous forests of northeastern USA. The summary of the study is listed below:

1. A comparison between the observed ET and model predictors demonstrated that the selected RS variables are capable of capturing the seasonal dynamics of 8-day ET. For the deciduous sites,  $R^2$  between ET and GPP (simulated either from the R-GPP or R-GPP-coni model), LST,  $\alpha_{sw}$  and LAI were 0.77, 0.64, 0.49 and 0.70, respectively while for the coniferous sites,  $R^2$  between them were 0.77, 0.76, 0.51 and 0.58, respectively.
2. During the calibration stage, the model captured the seasonal dynamics of the observed ET successfully with  $R^2 = 0.90$  and  $0.85$  (RMSE= 12.40 and  $14.95 \text{ W/m}^2$ ) for the deciduous sites of Duke hardwood forest and MMSF, respectively. The model was also successful for the coniferous calibration

site of Howland forest, with  $R^2 = 0.85$  (RMSE = 10.89 W/m<sup>2</sup>) between the actual and simulated ET.

3. The transferability of the proposed model was tested by applying it to independent validation sites. The model captured the variations of ET satisfactorily with  $R^2$  varying from 0.81 to 0.92 (RMSE varying from 12.13 to 14.44 W/m<sup>2</sup>) over the 4 deciduous validation sites of Harvard forest, Willow Creek, UMBS and Bartlett experimental site and with  $R^2$  of 0.84 and 0.83 (RMSE =14.38 and 11.60 W/m<sup>2</sup>) for the 2 coniferous validation sites of Duke pine and Howland forest west tower sites, respectively.

The agreement between the R-ET model simulated ET with EC tower estimated ET demonstrate the potential applicability of RS data in mapping the ET process of forest ecosystems with sufficient accuracy. This proposed model attempts to overcome the necessity of ground data or climate data as model input by utilizing MODIS ecosystem and radiation budget variables. Considering the trade-off between sophisticated modeling approach and the uncertainties in obtaining regional scale reliable climate data, we conclude that this simple R-ET model is a practical and promising approach for regional scale hydrological modeling and water resource planning and managements.

## 6.8 References

- Alados, I., Foyo-Moreno, I., Olmo, F.J. and Alados-Arboledas, L. 2003. Relationship between net radiation and solar radiation for semi-arid shrubland, *Agricultural and Forest Meteorology* **116** (3–4) (2003). pp. 221–227.
- Arora, V. 2002. Modeling vegetation as a dynamic component in soil vegetation-atmosphere transfer schemes and hydrological models, *Rev. Geophys.*, 40(2), 1006, doi:10.1029/2001RG000103.
- Bastiaanssen, W., Menenti, G. M., Feddes, R. A. and Holtslag, A. A. M. 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation, *Journal of Hydrology*, 212–213, 198–212.
- Batra, N., Islam, S., Venturini, V., Bisht, G. and Jiang, L. 2006. Estimation and comparison of evapotranspiration from MODIS and AVHRR sensors for clear sky days over the Southern Great Plains, *Remote sensing of Environment*, 103, 1-15.
- Betts, A. K., Chen, F., Mitchell, K. E. and Janjic, Z. I. 1997. Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta Model using FIFE data, *Monthly Weather Review*, 125(11), 2896–2916.
- Boegh, E., Soegaard, H., Hanan, N., Kabat, P. and Etsch, L. 1998. A remote sensing study of the NDVI–Ts relationship and the transpiration from sparse vegetation in the Sahel based on high resolution satellite data. *Remote Sensing of Environment*, 69, pp. 224–240.

- Boegh, E., Soegaard, H. and Thomsen, A. 2002. Evaluating evapotranspiration rates and surface conditions using Landsat TM to estimate atmospheric resistance and surface resistance. *Remote Sensing of Environment*, 79, 329–343.
- Bounoua, L., Collatz, G. J., Los, S. O., Sellers, P. J., Dazilich, D. A., Tucker, C. J. and Randall, D. A. 2000. Sensitivity of climate to changes in NDVI. *Journal of Climate*, 13, 2277– 2292,
- Brotzge, J. A. and Crawford, K. C. 2003. Examination of the surface energy budget: A comparison of eddy correlation and Bowen ratio measurement systems. *Journal of Hydrometeorology*, 4, 160–178.
- Ceccato, P., Flasse, S. and Gregoire, J. M. 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote Sensing of Environment*, 82, 198–207.
- Cleugh, H., Leuning, R., Mu, Q., Running, S. (2007) Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment*, 106, 285-304.
- Daley, M. J. 2007. Changes in ecohydrological function due to the loss and replacement of eastern hemlock in a New England forest. PhD thesis, Boston University.
- Dickinson, R. 1996. In: *Amazonian Deforestation and Climate*. edited by J. Gash, C. Nobre, J. Roberts, R. Victoria, pp. 399–411.

- Drexler, J., Snyder, R., Spano, D., Ta, K. and Paw, U. 2004. A review of models and micrometeorological methods used to estimate wetland evapotranspiration. *Hydrological Processes*, 18, 2071– 2101.
- Falge, E., Baldocchi, D., Olson, R. J., Anthoni, P., Aubinet, M. and Bernhofer, C. 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. *J. Agricultural and Forest Meteorology* 107, 43–69.
- Fisher, J. B., Tu, K. P., Baldocchi, D. 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, 112, 901-919.
- Friedl, M. A. 1996. Relationships among remotely sensed data, surface energy balance, and area-averaged fluxes over partially vegetated land surfaces, *Journal of Applied Meteorology*, 35, 2091–2103.
- Freitag, H., Ferguson, P. R., Dubois, K., Hayford, E. K., von Vordzobec, V., Veizer, J. 2008. Water and carbon fluxes from savanna ecosystems of the Volta River watershed, West Africa, *Global and Planetary Change*, 61, 3-14.
- Gaige, E., Dail, D.B., Hollinger, D.Y., Davidson, E. A., Fernandez, I. J., Sievering, H., White, A., Halteman, W. 2007. Changes in canopy processes following whole-forest canopy nitrogen fertilization of a mature spruce-hemlock forest. *Ecosystems* 10:7 1133-1147.

- Gillies, R.R. and Carlson, T. N. 1995. Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *Journal of Applied Meteorology*, 34, pp. 745–756.
- Glenn, E. P., Morino, K., Didan, K., Jordan, F., Carroll, K.C., Nagler, P.L. et al. 2008. Scaling sap flux measurements of grazed and ungrazed shrub communities with fine and coarse-resolution remote sensing. *Ecohydrology*, 1(4), 316–329.
- Goetz, S.Z., 1997, Multisensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site, *International Journal of Remote Sensing*, 18, pp. 71–94.
- Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C., Wofsy, S. C. 1996) Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global Change Biology*, 2, 169-182.
- Goward, S.N. and Hope, A. S. 1989, Evapotranspiration from combined reflected solar and emitted terrestrial radiation: preliminary FIFE results from AVHRR data. *Advances in Space Research*, 9, pp. 239–249.
- Goward, S. N., Xue, Y. and Czajkowski, K.P. 2002. Evaluating land surface moisture conditions from the remotely sensed temperature/vegetation index measurements: an exploration with the simplified simple biosphere model. *Remote Sensing of Environment*, 79, pp. 225–242.
- Griffis, J., Black, T. A., Gaumont-Guay, D., Drewitt, G. B., Nestic, Z., Barr, A.G., Morgenstem, K. and Kljun, N. 2004. Seasonal variation and partitioning

of ecosystem respiration in a southern Boreal aspen forest. *Agricultural and Forest meteorology*, 125 (3–4), 207–223.

Guyot, A., Cohard, J-M., Anquetin, S., Galle, S., Lloyd, C. R. 2009. Combined analysis of energy and water balances to estimate latent heat flux of a sudanian small catchment. *Journal of Hydrology*, doi: 10.1016/j.jhydrol.2008.12.027.

Hadley, J. L., Khuzja, P. S., Daley, M. J., Phillips, N. G., Malcahy, T., Singh, S. 2008. Water use and carbon exchange of red oak- and eastern hemlock dominated forests in the northeastern USA: implications for ecosystem-level effects of hemlock woolly adelgid. *Tree Physiology*, 28, 615-627.

Harries, J.E., Russell, J.E., Hanafin, J.A., Brindley, H., Futyan, J., Rufus, J., Kellock, S., Matthews, G., Wrigley, R., Last, A., Mueller, J., Mossavati, R., Ashmall, J., Sawyer, E. et al. 2005. The geostationary earth radiation budget project. *Bulletin of American Meteorological Society*, 86 (7), 945–960.

Hashimoto, H., Dungan, J. L., White, M. A., Yang, F., Michaelis, A. R., Running, S. W. and Nemani, R. R. 2008. Satellite-based estimation of surface vapor pressure deficits using MODIS land surface temperature data. *Remote Sensing of Environment*, 112, 142–155, doi:10.1016/j.rse.2007.04.016.

Heinsch, F. A., Zhao, M. S., Running, S.W., Kimball, J.S., Nemani, R. R. et al. 2006. Evaluation of remote sensing based terrestrial production from

- MODIS using AmeriFlux eddy tower flux network observations. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1908 – 1925.
- Huete, A. R., Jackson, R. D. and Post, D.F. 1985. Spectral response of a plant canopy with different soil backgrounds. *Remote sensing of Environment*, 17, 37-53.
- Jacobs, J.M., Myers, D.A., Anderson, M.C., Diak, G.R. 2002. GOES surface insolation to estimate wetlands evapotranspiration. *Journal of Hydrology*, 266 (1–2), 53–65.
- Jahan, N. and Gan, T. Y. 2009. Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables. *Journal of Geophysical Research*, Vol.114, No. G04026.
- Jahan, N. and Gan, T. Y. 2011. Modeling vegetation-climate relationship in a central mixed wood forest region of Alberta using normalized difference and enhanced vegetation indices. *International Journal of Remote Sensing*, 32, 313-335.
- Jahan, N. and Gan, T. Y. 2011 (submitted). A continuous estimate of gross primary production of coniferous forest using MODIS data, *Remote Sensing of Environment*.
- Jaksic, V., Kiely, G., Albertson, J., Oren, R., Katul, G., Leahy, P. and Byrne, K. A. 2006. Net ecosystem exchange of grassland in contrasting wet and dry years. *Agricultural and Forest Meteorology*, 139, 323– 334.

- Jiang, L. and Islam, S. 2001. Estimation of surface evaporation map over southern Great Plains using remote sensing data. *Water Resources Research*, 37(2), 329–340.
- Jiang, H., Liu, S., Sun, P., An, S., Zhou, G., Li, C. et al. 2004. The influence of vegetation type on the hydrological process at the landscape scale. *Canadian Journal of Remote Sensing*, 30(5), 743– 763.
- Jiang, L., Islam, S., Guo, W. et al. 2009. Satellite-Based Daily Actual Evapotranspiration Estimation Algorithm over South Florida. *Global and Planetary Change*, doi: 10.1016/j.gloplacha.2008.12.008.
- Kimura, R. 2007. Estimation of moisture availability over the Liudaogou river basin of the Loess Plateau using new indices with surface temperature, *Journal of Arid Environments*, vol. 70, pp. 237-252.
- Kramer, P.J. and Boyer, J.S. 1995. *Water Relations of Plants and Soil*. Academic Press, San Diego. ISBN 0-12-425060-2.
- Lasdon, L. S., Waren, A. D., Jain, A. and Ratner, M. 1978. Design and testing of a generalized reduced gradient code for nonlinear programming. *Transaction on Mathematical Software*, 4, 34– 50, doi:10.1145/355769.355773.
- Law, B.E., Falge, E., Gu, L., Baldocchi, D. D., Bakwin, P. et al. 2002. Environmental controls over carbon dioxide and water vapour exchange of terrestrial vegetation. *Agricultural and Forest Meteorology*, 113, 97–120.

- Los, S. O., Collatz, G. J., Malmstrom, C. M., Pollack, N. H., DeFries, R. S., Bounoua, L. et al. 2000. A global 9-year biophysical land surface dataset from NOAA AVHRR data. *Journal of Hydrometeorology*, 1, 183–199.
- Lawrence, D. M. and Slingo, J. M. 2004. An annual cycle of vegetation in a GCM. Part I: implementation and impact on evaporation. *Climate Dynamics*, 22, 87-105.
- Mailhol, J. C., Olufayo, A. A., Ruelle, P. (1997). Sorghum and sunflower evapotranspiration and yield from simulated leaf area index. *Agricultural Water Management*, 35,167-182.
- McCabe, M.F. and Wood, E. F. 2006. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sensing of environment*, Vol., 105, pp. 271-285.
- McCaughey, J. H., Pejam, M.R., Arain, M. A. and Cameron, D. A. 2006. Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada. *Agricultural and Forest Meteorology*, 140, 79-96.
- Mellander, P., Bishop, K., Lundmark, T. 2004. The influence of soil temperature on transpiration: a plot scale manipulation in a young Scots pine stand. *Forest ecology and management*, 195, 15-28.
- Mu, Q., Heinsch, F., Zhao, M., Running, S. 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, 111, 519-536.
- Myneni, R. B., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Hoffman, S., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M.,

- Morisette, J. T., Votava, P., Nemani, R. R. and Running, S. W. 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83, 214– 231.
- Nagler, P. L., Scott, R. L., Westenburg, C., Cleverly, J. R., Glenn, E. P. and Huete, A. R. 2005. Evapotranspiration on western U.S. rivers estimated using the Enhanced Vegetation Index from MODIS and data from eddy covariance and Bowen ratio flux towers. *Remote Sensing of Environments*, 97, 337-351.
- Nemani, R. R. and Running, S. W. 1989. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal infrared AVHRR data. *Journal of Applied Meteorology*, 28, 276–284.
- Nishida, K. and Nemani, R. R. et al. 2003. Development of an evapotranspiration index from Aqua/MODIS for monitoring surface moisture status. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2).
- Norman, J. M., Anderson, M. C., Kustas, W. P., French, A. N., Mecikalski, J., Torn, R. et al. 2003. Remote sensing of surface energy fluxes at 101-m pixel resolutions. *Water Resources Research*, 39(8), 1221.
- Oishi, A., Oren, R. and Stoy, P. C. 2008. Estimating components of forest evapotranspiration: A footprint approach for scaling sap flux measurements. *Agricultural and Forest Meteorology*, 148, 1719-1732.
- Ollinger, S.V, Richardson, A. D., Martin, M. E., Hollinger, D. Y. et al. 2008. Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal

forests: Functional relations and potential climate feedbacks. *The National Academy of Sciences of the USA*, vol 105, No 49 pp. 19336–19341

Ryu, Y., Kang, S., Moon, S. and Kim, J. 2008. Evaluation of land surface radiation balance derived from moderate resolution imaging spectroradiometer (MODIS) over complex terrain and heterogeneous landscape on clear sky days, *Agricultural and Forest meteorology*, vol. 148, pp. 1538-1552.

Schaaf, C. B., et al. 2002. First operational BRDF, albedo nadir reflectance products from MODIS, *Remote Sensing of Environment*, 83, 135 – 148.

Schmid, H.P., Grimmond, C. S. B., Offerle, B., Cropley, F.D. and Su, H. B. 2000. Measurements of CO<sub>2</sub> and energy fluxes over a mixed hardwood forest in the Midwestern United States. *Agricultural and Forest Meteorology*, 103, 355–373.

Schmid, H. P., Su, H. B., Vogel, C. S. and Curtis, P. S. 2002. Footprint modeling for vegetation atmosphere exchange studies: A review and perspective. *Agricultural and Forest Meteorology*, 113, 159– 183, doi:10.1016/S0168-1923(02)00107-7.

Schmid, H.P., Su, H.-B., Vogel, C. S., Curtis, P.S. 2003. Ecosystem atmosphere exchange of carbon dioxide over a mixed deciduous forest in northern lower Michigan, *Journal of Geophysical Research*, 108 (D14), 4417 doi:10.1029/2002JD003011.

- Seguin, B., Becker, F., Phulpin, T., Gu, X. F., Guyot, G., Kerr, Y. 1999. IRSUTE: a minisatellite project for land surface heat flux estimation from field to regional scale. *Remote Sensing of Environment*, 68, 357–369.
- Sims, D. A., et al. 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment*, 112, 1633–1644.
- Sobrino, J. A., Gomez, M., Jiménez-Muñoz, J. C., Oliso, A. 2007. Application of a simple algorithm to estimate daily evapotranspiration from NOAA–AVHRR images for the Iberian Peninsula. *Remote Sensing of Environment*, 110, 139-148.
- Song, J. 1998. Diurnal asymmetry in surface albedo, *Agric. For. Meteorol.*, 92, 181–189.
- Spaulding, K. A. 1998. Neural metamorphic optimization algorithms, Masters thesis, Univ. of Texas at Austin, Austin.
- Stoy, P.C., Katul, G. G., Siqueira, M.B.S. et al. 2006. Separating the effects of climate and vegetation on evapotranspiration along a successional chronosequence in the southeastern US. *Global Change Biology*, 14:6, 1409-1427.
- Su, H.B., McCabe, M. F., Wood, E. F., Su, Z., Prueger, J. H. 2005. Modeling evapotranspiration during SMACEX: comparing two approaches for local- and regional-scale prediction. *Journal of Hydrometeorology*, 6 (6), 910–922.

- Twine, T., Kustas, W., Norman, J., Cook, D., Houser, P., Meyers, T. et al. 2000. Correcting eddy-covariance flux underestimates over a grassland. *Agricultural and Forest Meteorology*, 103, 279– 300.
- Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D. et al. 2007. Factors controlling CO<sub>2</sub> exchange on time scales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research*, G 112:G02020.
- Venturini, V., Bisht, G., Islam, S., and Jiang, L. 2004. Comparison of evaporative fractions estimated from AVHRR and MODIS sensors over South Florida. *Remote Sensing of Environment*, 93:77-86.
- Wan, Z., and Dozier, J. 1996. A generalized split-window algorithm for retrieving land-surface temperature from space, *IEEE Trans. Geosci. Remote Sens.*, 34, 892– 905,
- Wang, S., and Davidson, A. 2007. Impact of climate variations on surface albedo of a temperate grassland. *Agricultural Forest Meteorology*, 142, 133–142.
- Wang, S., Grant, R. F., Versegny, D. L. and Black, T. A. 2001. Modeling plant carbon and nitrogen dynamics of a boreal aspen forest in CLASS—The Canadian Land Surface Scheme. *Ecological Modelling*, 142, 135 – 154.
- Wang, K., Wang, P., Li, Z., Cribb, M., Sparrow, M. 2007. A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *Journal of Geophysical Research – Atmospheres*, 12, Art. No. D15107.

- Weaver, J. E. and Mogensen, A. 1919. Relative Transpiration of Coniferous and Broad- Leaved Trees in Autumn and Winter. *Botanical Gazette*, 68, 393-424.
- Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P. et al. 2002. Energy balance closure at FLUXNET sites. *Agricultural and Forest Meteorology*, 113, 223–243.
- Xiao, X. M., Zhang, Q. Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B. and Ojima, D. 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, 91, 256 – 270
- Yang, F., White, M., Michaelis, A., Ichii, K., Hashimoto, H., Votava, P., Zhu, H. and Nemani, R. R. 2006. Prediction of continental-scale evapotranspiration by combining MODIS and Ameriflux data through support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 3452-3461.
- Zhang, X., Friedl, M. A., Schaff, C. B. and Srahetr, A. H. 2004. Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. *Global Change Biology*, 10, pp.1133-1145.
- Zhang, J., Hua, Y., Xiao, X., Chen, P., Han, S., Song, G., Yu, G. 2009a. Satellite-based estimation of evapotranspiration of an old-growth temperate mixed forest. *Agricultural and forest meteorology*, 149, 976-984
- Zhang, KE, Kimball, J. S., Mu, Q., Jones, L. A., Goetz, S. J., Running, S. W. 2009b. Satellite based analysis of northern ET trends and associated

changes in the regional water balance from 1983 to 2005. *Journal of Hydrology*, 379, 92-110.

Table 6.1. Location (lat/long in decimal degrees), vegetation type, annual mean temperature (AMT), annual precipitation (AP), stand age and years of data used for the nine eddy covariance tower sites used in this study.

Site name	Latitude	Longitude	Vegetation	AMT (°C)	AP (mm)	Stand age (years)	Years
Harvard Forest	42.54°N	72.17°W	Deciduous Broadleaf	7.9	1066	60-80	2000-2003
Morgan Monroe State Forest (MMSF)	39.32°N	86.41°W	Deciduous Broadleaf	11.1	1012	70	2000-2005
University of Michigan Biological station (UMBS)	45.56°N	84.71°W	Deciduous Broadleaf	6.2	750	90	2000-2001
Willow Creek	45.81 °N	90.08°W	Deciduous Broadleaf	4.1	815	55-90	2004-2005
Bartlett Experimental Forest	44.06 °N	71.29°W	Deciduous Broadleaf	6.0	1300	99	2004-2006
Duke Hardwood Forest	35.97 °N	79.10°W	Deciduous Broadleaf	14.3	1154	80-100	2001-2005
Howland Forest	45.20 °N	68.74°W	Evergreen Needleleaf	6.6	523-1032	95-140	2000-2004
Duke pine Forest	35.97°N	79.10°W	Evergreen Needleleaf	15.5	1145	17	2003-2005
Howland Forest West Tower	45.21°N	68.75°W	Evergreen Needleleaf	6.5	523-948	109	2000-2004

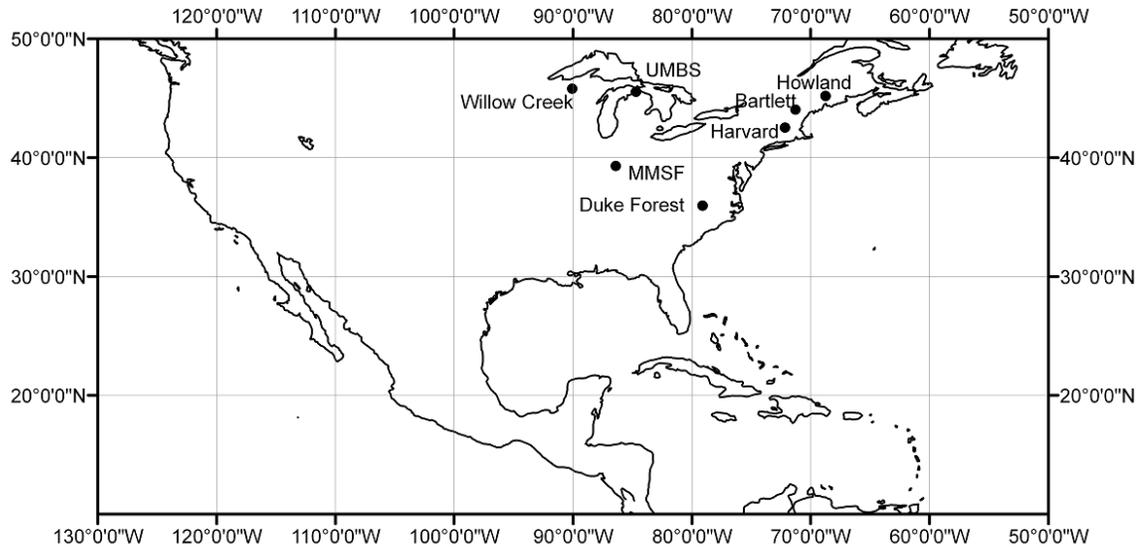


Figure 6.1. Location of the study sites

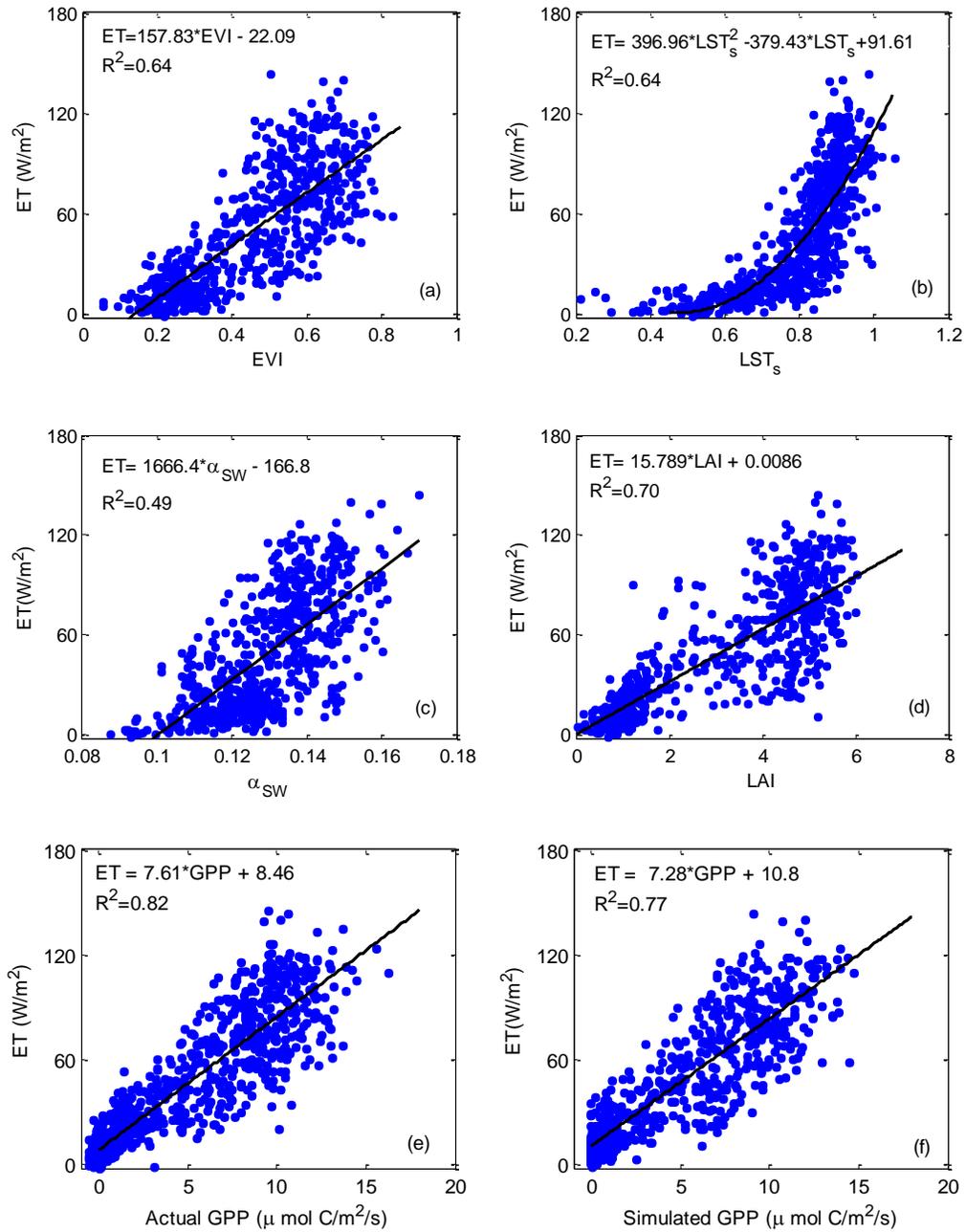


Figure 6.2. Relationships between EC tower estimated ET and (a) enhanced vegetation index (EVI), (b) scaled land surface temperature (LSTs), (c) short wave albedo ( $\alpha_{sw}$ ), (d) leaf area index (LAI), (e) EC tower estimated actual GPP and (f) R-GPP model simulated GPP in the deciduous sites. All data points represent 8-day mean values.

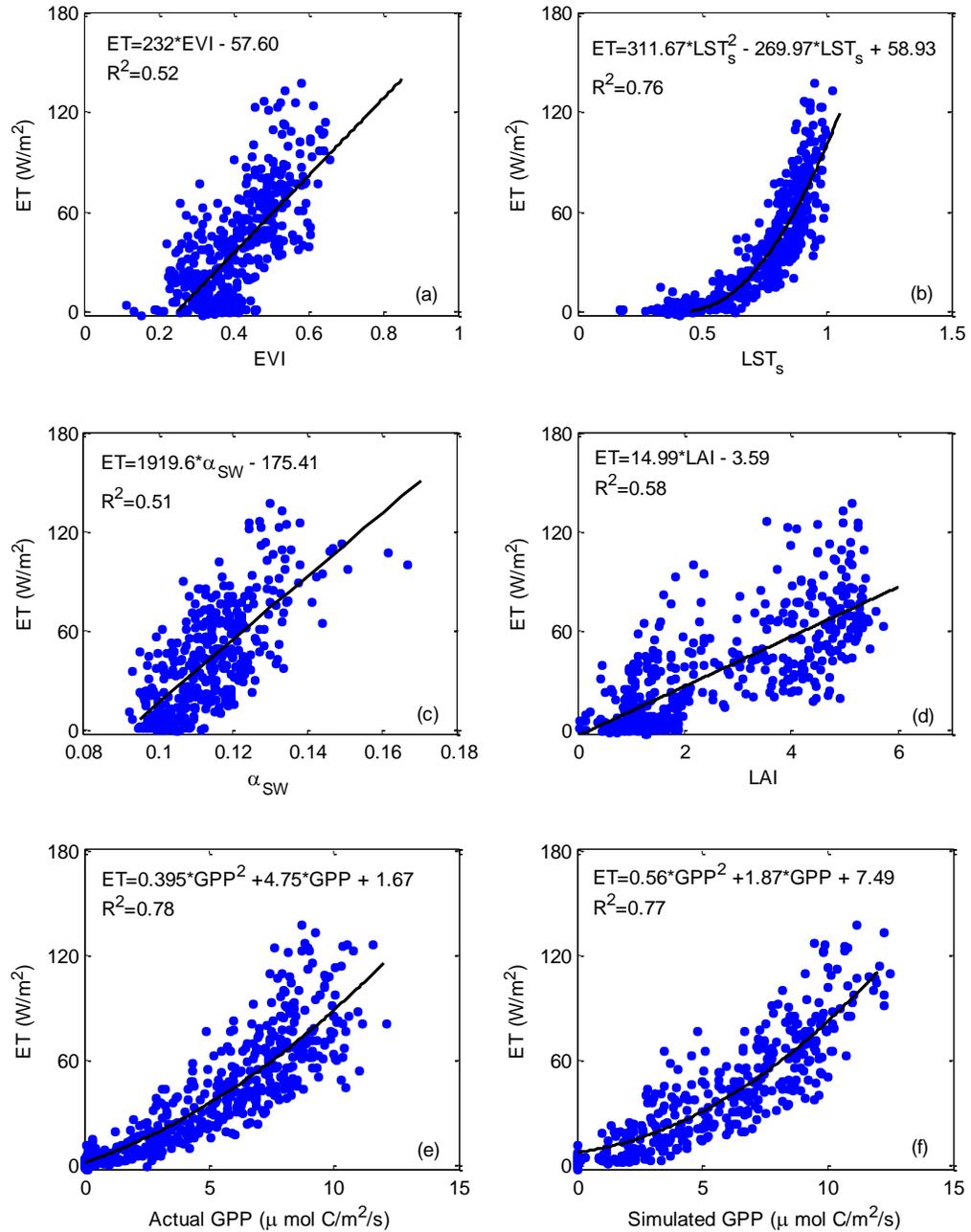


Figure 6.3. Relationships between EC towers estimated ET and (a) enhanced vegetation index (EVI), (b) scaled land surface temperature (LSTs), (c) short wave albedo ( $\alpha_{sw}$ ), (d) leaf area index (LAI), (e) EC tower estimated actual GPP, and (f) R-GPP-Coni model simulated GPP for the coniferous sites. All data points represent 8-day mean values.

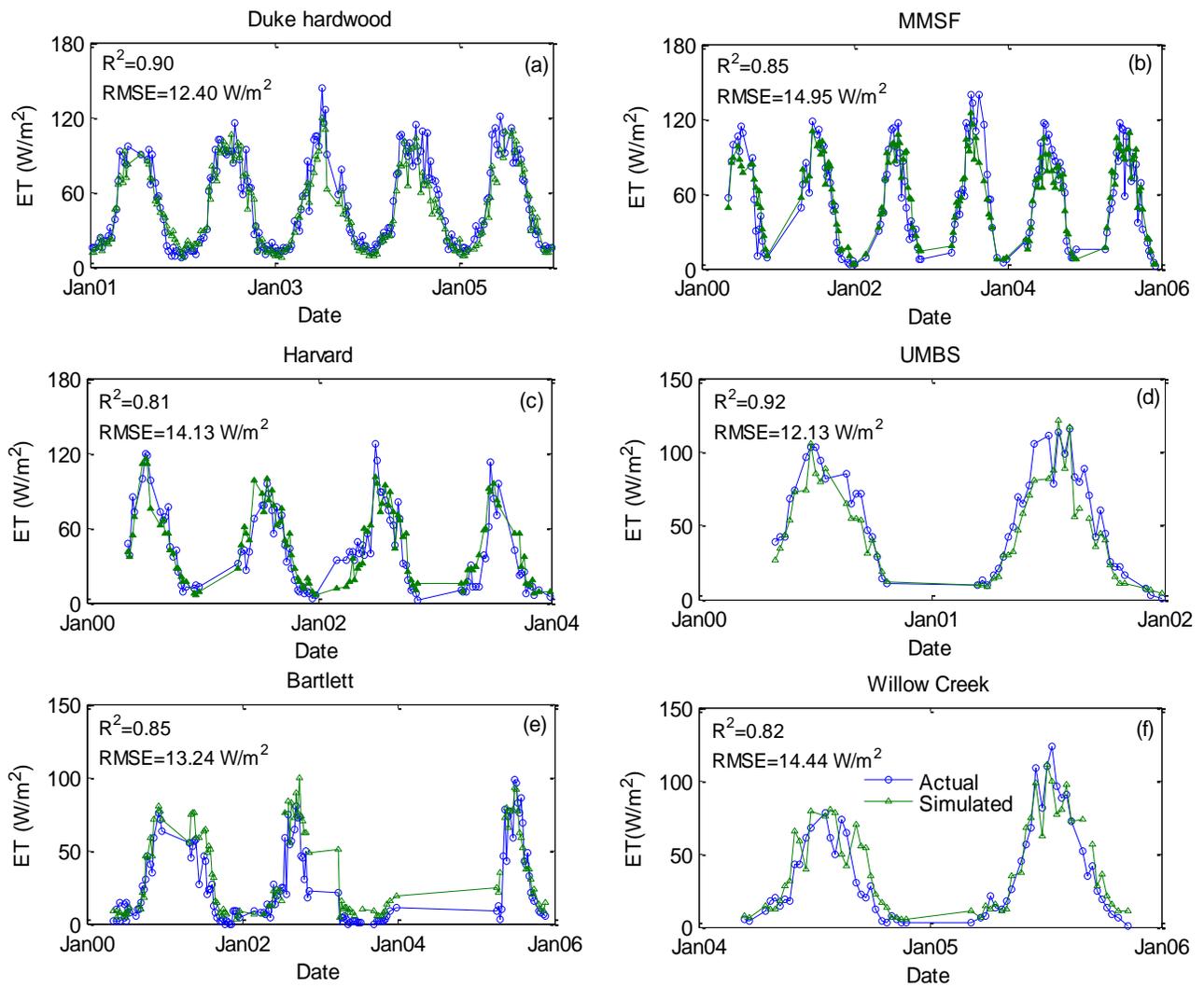


Figure 6.4. Annual cycle of eddy covariance tower estimated evapotranspiration (ET) (actual) and R-ET model predicted ET (simulated) for the deciduous sites. All data points represent 8-day mean values.

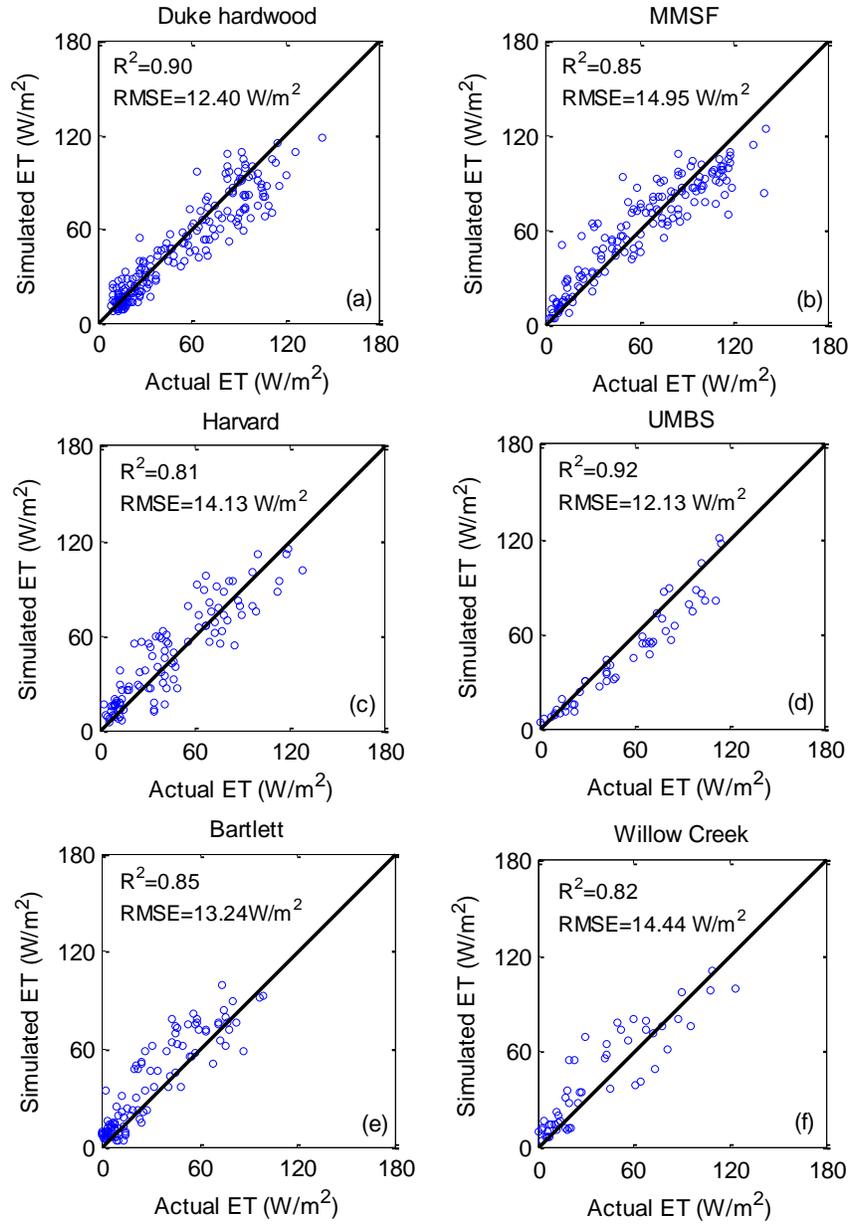


Figure 6.5. Scatter plot of R-ET model simulated evapotranspiration (ET) and eddy covariance tower estimated ET for the deciduous sites. All data points represent 8-day mean values.

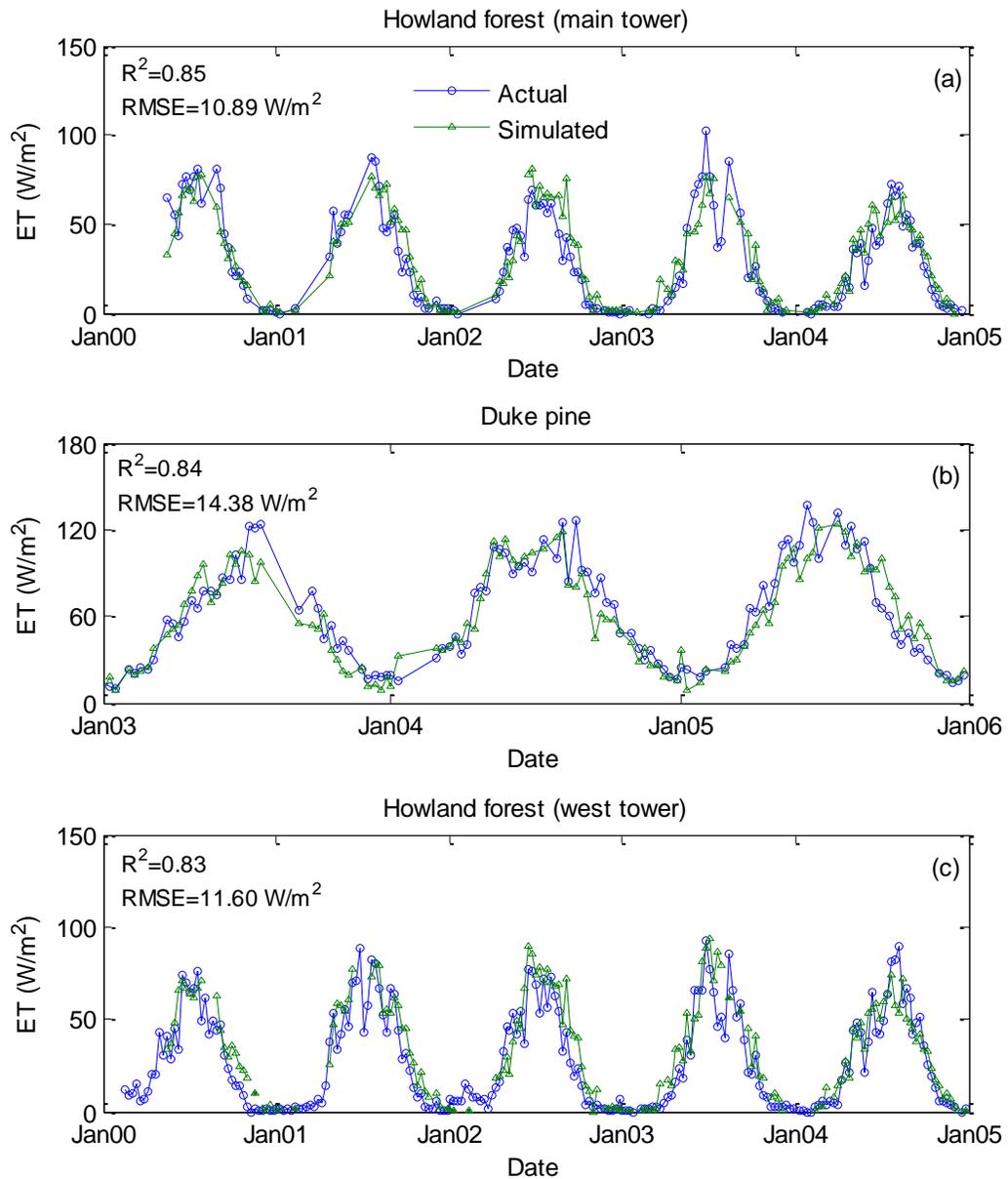


Figure 6.6. Annual cycle of eddy covariance tower estimated evapotranspiration (ET) (actual) and R-ET model predicted ET (simulated) for the coniferous sites. All data points represent 8-day mean values.

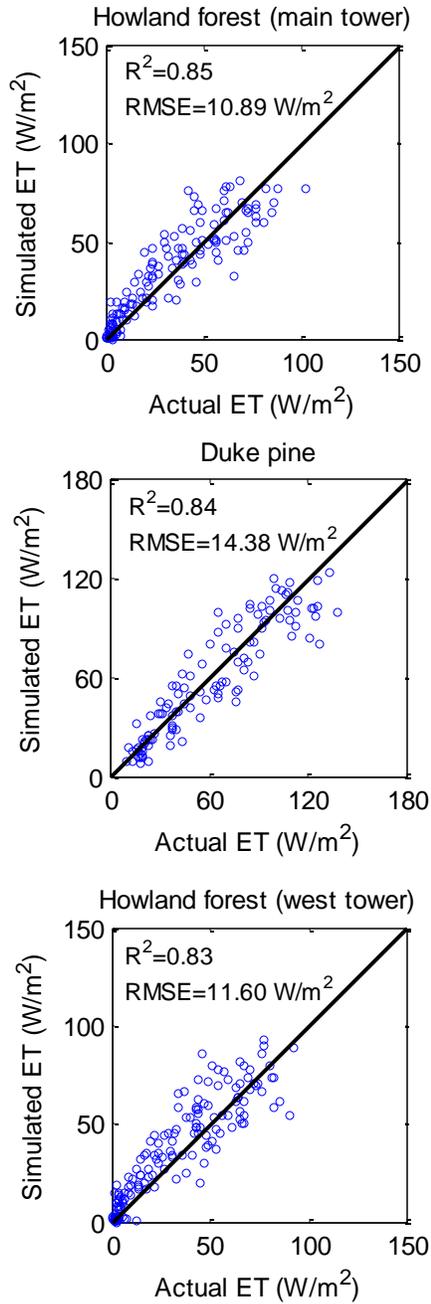


Figure 6.7. Scatter plot of R-ET model simulated evapotranspiration (ET) and eddy covariance tower estimated ET (actual) for the coniferous sites. All data points represent 8-day mean values.

# Chapter 7

## Summary, conclusions and recommendations

---

### 7.1 Summary and conclusions

On a vegetated surface, carbon and water cycles are closely linked with vegetation. Physiological processes of vegetation have strong influences on the thermal and chemical composition of the atmosphere and thus on climate. Therefore understanding and quantifying the land–atmosphere exchange of turbulent fluxes of carbon, energy and water is fundamental to the understanding of the functioning of the ecosystem.

This study developed algorithms for estimating carbon and water fluxes using solely remotely sensed (RS) data. As these fluxes are primarily controlled by the vegetation and the climate, therefore the first objective of this research was to investigate whether RS data, such as vegetation indices (VIs), are capable of relating the responses of vegetation to climate variables. In Chapter 2, the relationships between vegetation and climate were investigated using an Artificial Neural Network (ANN) and two satellite derived VIs, the normalized difference vegetation index (NDVI) from the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and the enhanced vegetation index (EVI) from the Moderate Resolution Imaging

Spectroradiometer (MODIS). This study reveals that EVI is more sensitive to climate variables than NDVI probably because of the saturation problem of NDVI over dense vegetation, such as forests. In general, this study demonstrates the promising potential of monitoring terrestrial vegetation productivity from climate variables over the boreal mixedwood forest of central-eastern Alberta, Canada.

This study explored the potential of multipolarized radar data from the newly launched Radarsat-2 synthetic aperture radar (SAR) in retrieving near surface soil moisture. Ten Radarsat-2 images were acquired over the Paddle River basin (PRB) of Alberta, Canada and about 2250 soil samples were collected from the 9 sites of different landuses within the PRB on those 10 days. Then soil moisture was retrieved using regressions, the theoretical Integral Equation Model, ANN and Support Vector Machine (Chapter 3). Besides radar data, various combinations of optical satellite data (NDVI and LST) from the MODIS, soil properties and topographic variables (slope and aspect) were used in different combination as supplementary input to the regression, ANN and SVM algorithms. The results demonstrated promising approaches of retrieving near surface soil moisture, with root mean squared error of 3.11%, from multi-sensor data.

This study also explored the potential application of two remotely sensed (RS) variables, the Global Vegetation Moisture Index and the near-infrared albedo, in estimating the gross primary production (GPP) of deciduous (Chapter 4) and coniferous (Chapter 5) forests. A statistical model, called the Remotely Sensed

GPP (R-GPP) model, was developed to estimate GPP using only 4 RS variables: two radiation (land surface temperature (LST),  $\text{Albedo}_{\text{NIR}}$ ) and two ecosystem variables (enhanced vegetation index (EVI) and GVM). The R-GPP model was calibrated and validated against the GPP estimates derived from the eddy covariance flux tower located at different coniferous and deciduous sites in the north-eastern USA. It outperformed the primary RS based GPP algorithm of the MODIS sensor which suffered from an underestimation of the GPP during the peak growing seasons in most of the study sites.

Evapotranspiration (ET) is a key component of the water cycle and it plays an important role in the climate dynamics, ecosystem productivity and watershed hydrology. Due to the scarcity of ground-based meteorological stations, the spatial coverage of ground-based measured ET is rarely sufficient to accurately estimate the evaporation budget or for modeling hydrologic processes at regional scale. This study developed an algorithm based on remote sensing data (Chapter 6) for estimating ET (R-ET) without relying on ground measurements or climate data but solely on four remotely sensed (RS) variables, namely, land Surface temperature, short-wave Albedo, leaf area index and gross primary production (GPP). The model was calibrated against the eddy covariance tower estimated ET data using a nonlinear optimization algorithm (Generalized Reduced Gradient, GRG2) for computing ET for the coniferous and the deciduous forests. Then the calibrated ET model was validated for 4 deciduous and 2 coniferous sites independent of the calibration experience. Considering the trade-off between

complex modeling approach and the uncertainties in obtaining reliable regional scale climate data, it seems that this simple, yet robust R-ET model is a practical approach for modeling regional ET processes.

## **7.2 Recommendations for future work**

Although the proposed R-GPP and R-ET model has shown promising results in estimating GPP and ET, respectively, for several deciduous and coniferous forests of north-eastern USA, further validation is needed to test the robustness of these models and their applicability in different climatic and biophysical conditions. The model parameters may need to be refined for other climatic regimes and biomes such as grassland, cropland and savannas. Further study is also needed to determine whether the net primary production can be estimated from the RS variables used in this study.

Water use efficiency ( $WUE = GPP/ET$ ) is a measure of the amount of dry matter produced by plant during photosynthesis relative to the water lost by the ecosystem (i.e. ET) and has been used to determine the adaptation and productivity of plants in water-limited areas (Xu and Hsiao, 2004). There have been only few attempts to compute WUE solely on the basis of satellite data (Zhang et al., 2009). Future studies can examine the potential of WUE estimation based on a combined version of the R-GPP and R-ET models developed in this study and compare the model results with the observed data.

Many studies on the modeling of regional scale GPP and ET fluxes have utilized reanalyzed net radiation ( $R_n$ ) data available from different models which use supplementary or complementary ground based climate data as input. Due to scarcity of ground based stations, these models often use extrapolated climate data and extrapolation from a limited number of ground stations may affect the reliability of these data (*Mu et al.*, 2007; *Heinsch et al.*, 2006). On the other hand, reliable  $R_n$  data derived entirely from RS data at moderate resolutions (e.g, 1 km) are also rare (*Ryu et al.*, 2008). Therefore, instead of using measured or climate model simulated  $R_n$ , in this research, albedo from the MODIS satellite was used as it is an important component of the surface energy exchange (*Ollinger et al.*, 2008). Further research is necessary to model the net radiation reliably from solely satellite data and examine whether better estimation of GPP or ET can be obtained by replacing albedo with  $R_n$ .

In this study site specific relationships have been developed between climate variables and VIs using regression and neural network. Further study is necessary to explore the possible linkage between a regional and a local regression (or neural network) model. In general, a local model may be extended to other sites with some kind of adjustment factors that reflect the climate relationships between sites, such ideas can be tested in future.

### 7.3 References

- Gebremichael, M., and Barros, E. P. 2006. Evaluation of MODIS gross primary productivity (GPP) in tropical monsoon region, *Remote Sensing of Environment*, 100, 150–166.
- Heinsch, F.A., Zhao, M., Running, S.W., Kimball, J.S., Nemani, R.R., et al. 2006. Evaluation of remote sensing based terrestrial production from MODIS using AmeriFlux eddy tower flux network observations. *IEEE Transactions on Geosciences and Remote*, 44, 1908–1925.
- Ichii, K., Matsui, Y., Murakami, K., Mukai, T., Yamguchi, Y. and Ogawa, K. 2003. A simple global carbon and energy coupled cycled model for global warming simulation: Sensitivity to the light saturation effect, *Tellus, Series B*, 55, 676– 691.
- Mu, Q., Heinsch, F. A., Zhao, M. and Running, S. W. 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, 111, 519-536.
- Ollinger, S.V, A. D. Richardson, M. E. Martin, D. Y. Hollinger, et al. (2008) Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks, *The National Academy of Sciences of the USA*, vol 105, No 49 pp. 19336–19341
- Ryu, Y., S. Kang, S. Moon, J. Kim (2008). Evaluation of land surface radiation balance derived from moderate resolution imaging spectroradiometer (MODIS) over complex terrain and heterogeneous landscape on clear sky days, *Agricultural and Forest meteorology*, vol. 148, pp. 1538-1552.

- Xu, L. and Hsiao, T. C. 2004. Predicted vs measured photosynthetic water-use efficiency of crop stands under dynamically changing field environments. *Journal of experimental Botany*, Water saving Agriculture Special issue, 1-17.
- Zhang, J., Hu, Y., Xiao, X., Chen, P., Han, S., Song, G., Yu, G. 2009. Satellite-based estimation of evapotranspiration of an old growth temperate mixed forest. *Agricultural and Forest Meteorology*, 149, 976-984.

## Appendix

### Algorithm: A1

```
load Pareas; %Precipitaion input
load PETareas; %Potential evapotranspiration input
load Tareas; %Temperature input
moisareas=Pareas./PETareas; %aridity index input
load NDVareas; %NDVI input

[row, col1]=size(Pareas);

Input=[ (Pareas)'; (Tareas)'; PETareas';(moisareas)']; %input variables
[Inputn,meanin,stdin,outn,meanout,stdout] = prestd(Input,NDVareas');
%standardization of input variables so that its mean is 0 and the standard
deviation is 1
[rown,coln]=size(Inputn);
% val.P = Inputn(:,coln*6/10+1:coln); val.T = outn(:,coln*6/10+1:coln);

ptr = Inputn(:,1:coln*6/10); %training input
ttr = outn(:,1:coln*6/10); %training output

net=newff(minmax(ptr),[20,1],{'tansig','tansig'},'trainbr');

net.trainParam.show = 500;
net.trainParam.lr = 0.05;
net.trainParam.epochs = 5000;
net.trainParam.goal = 1e-5;

net.trainParam.mu=1;
net.trainParam.mu_dec=0.9;
net.trainParam.mu_inc=1.5;
[net,tr]=train(net,ptr,ttr);%,[],[],val);
plot(tr.epoch,tr.perf);% ,tr.epoch,tr.vperf)
legend('Training');%,'Validation',-1);
ylabel('Squared Error'); xlabel('Epoch')

an = sim(net,Inputn);
a = poststd(an,meanout,stdout); % simulated NDVI, postprocessed by poststd

reg1=0:.1:1;

mm=mean(NDVareas(1:row*6/10,1));
[Rc,P1]=corrcoef(a(:,1:coln*6/10),(NDVareas(1:row*6/10,1)));

[Rv,P2]=corrcoef(a(:,coln*6/10+1:coln),(NDVareas(row*6/10+1:row,1)));
```

```

b=a';
z1=(NDVareas(1:row*6/10,1)-mean(NDVareas(1:row*6/10,1))).^2;
z2=(b(1:row*6/10,1)-NDVareas(1:row*6/10,1)).^2;
Efc=(mean(z1)-mean(z2))/mean(z1); %efficiency coefficient_calibration stage
z11=(NDVareas(row*6/10+1:row,1)-mean(NDVareas(row*6/10+1:row,1))).^2;
z22=(b(row*6/10+1:row,1)-NDVareas(row*6/10+1:row,1)).^2;
Efv=(mean(z11)-mean(z22))/mean(z11); %efficiency coefficient_validation stage

```

## Algorithm: A2

```

clear all;
load backscatter; %Radar backscatter
load MC;%Actual soil moisture
load Variables; % other predictor variable

[row,col1]=size(MC);

Input=[backscatter'; Variables'];
Inputn= (Input+35)/35;
outn=MC';
outn=(outn)/50;
[rown,coln]=size(Inputn);

ptr = Inputn(:,1:coln*7/10); ttr = outn(:,1:coln*7/10);
net = newrbe(ptr,ttr,30);

an= sim(net,Inputn);
a=an*50;

reg1=0:10:60;

[R1,P1]=corrcoef(a(:,1:coln*7/10),(MC(1:row*7/10,1)))
Error1 =(a(:,1:coln*7/10)-(MC(1:row*7/10,1))).^2;
RMSE1=sqrt(mean(Error1))

[R2,P2]=corrcoef(a(:,coln*7/10+1:coln),(MC(row*7/10+1:row,1)))
Error2 =(a(:,coln*7/10+1:coln)-(MC(row*7/10+1:row,1))).^2;
RMSE2=sqrt(mean(Error2))

subplot(1,2,1)
plot((MC(1:row*7/10,1))',a(:,1:coln*7/10),reg1,reg1);xlabel('Actual soil moisture (%)');ylabel('Simulated soil moisture (%)');axis square;
subplot(1,2,2)

```

```
plot((MC(row*7/10+1:row,1))',a(:,coln*7/10+1:coln),reg1,reg1);  
xlabel('Actual soil moisture (%)');ylabel('Simulated soil moisture (%)');axis  
square;
```