APPLICATION OF HYPERSPECTRAL REMOTE SENSING IN STRESS DETECTION AND CROP GROWTH MODELING IN CORN FIELDS

By

Yousef Karimi-Zindashty

Department of Bioresource Engineering
Macdonald Campus of McGill University
Montreal, Canada

March, 2005

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Your file Votre référence
ISBN: 0-494-12867-4
Our file Notre référence
ISBN: 0-494-12867-4

ABSTRACT

This study used hyperspectral data to determine nitrogen, weed, and water stresses in a corn (*Zea mays* L.) field in southwestern Quebec, and incorporated these data in crop growth models for better crop growth simulation under stressful conditions.

In 2000, aerial hyperspectral images (72 wavebands, ranging from 407 to 949 nm) were acquired, and analyzed using a stepwise approach to identify wavebands useful in detecting weed and nitrogen stresses. Discriminant analysis (DA) was used to classify different weed and nitrogen treatments and their combinations. This analysis showed greater classification accuracy (nearly 75%) than those obtained with artificial neural networks (58%) or decision tree algorithms (60%), at the initial growth stages, the time when remedial actions are most needed to alleviate weed and nitrogen stresses.

To explore the possibility of improving nitrogen stress detection in corn in the presence of a confounding water stress, ground-based 2151 narrow-waveband reflectance values (350 to 2500 nm), were collected in 2002. Using DA with the chosen subset of narrow-wavebands, a classification accuracy of greater than 95% was obtained.

For crop growth monitoring, the STICS model was evaluated for yield and biomass estimation in cornfields under different stressful growth conditions using the data collected from 2000 to 2002. Measured yield, biomass, and leaf area index (LAI) were used for both calibration and validation of the model. High correlation coefficients between the measured and estimated grain yield (0.96), biomass (0.98), and LAI (0.93) indicated that the model has good potential in the simulation of corn growth. The model was also linked with LAI values estimated from the hyperspectral observations using the Support Vector Machines technique. Coupling STICS with remote sensing resulted in an overall improvement in the simulation of corn yield (6.3%) and biomass (3.7%).

A new approach was developed to apply crop growth models for yield estimation in weedy areas. The proposed method first corrects the measured/estimated LAI values in weed infested fields for weed effect, and then uses the corrected LAI values as input to the crop growth model. The results showed that the crop yield and biomass predictions were correctly simulated by this method.

RÉSUMÉ

Au cours de cette étude, des données en hyperspace spectral servirent à déterminer les stress en azote, en mauvaises herbes et en eau d'un champ de maïs (Zea mays L.) dans le sud-est du Québec. Ces données furent incorporées dans un modèle de croissance des cultures afin d'améliorer ses simulations sous conditions de stress.

En 2000, des images aériennes en hyperspace spectral (72 gammes d'ondes, variant de 407 à 949 nm) furent acquises et traitées grâce à une analyse de régression par degrés visant à identifier les gammes d'ondes utiles à l'identification des stress en mauvaises herbes et en manque d'azote. Une analyse discriminante (AD) servit à classifier différents niveaux de mauvaises herbes et d'azote et leurs combinaisons. Elle démontra une exactitude de classification plus élevée (près de 75%) comparée à celle de réseaux de neurones formels (58%) ou d'un algorithme d'arbre de décision (60%) lors des stages initiaux de croissance, le moment où une intervention corrective serait la plus efficace à corriger un stress de mauvaises herbes où de manque d'azote.

Afin d'étudier la possibilité d'améliorer la détection de manque d'azote dans le maïs, en présence d'un facteur confondant de stress hydrique, la réflectance de la culture dans 2151 gammes d'ondes étroites (350 à 2500 nm), fut mesurée au sol en 2002. Utilisant une AD avec un sous-ensemble choisi de gammes d'ondes étroites, une exactitude de classification de plus de 95% fut atteinte.

Visant à permettre une suivie de la croissance du maïs, la capacité du modèle STICS d'estimer le rendement et la biomasse au champ sous diverses conditions de croissance imposant un stress, fut évalué avec des données recueillis entre 2000 et 2002. Des mesures au champ de rendement, biomasse et indice foliare (LAI) servirent à la fois à la calibration et à la validation du modèle. Des coéfficients de corrélation élevés entre les valeurs estimées et mesurées du rendement grain (0.96), biomasse (0.98) et LAI (0.93), indiquèrent que le modèle avait un potentiel élevé comme outil de simulation de la croissance du maïs. Le modèle fut également lié à des données de LAI provenant d'observations en hyperespace spectral suivant une technique « support vector machines .» L'accouplement de STICS avec un système de télédétection améliora la simulation du rendement grain du maïs de 6.3%, et celui de la biomasse de 3.7%.

Une nouvelle démarche fut mise en œuvre afin de pouvoir utiliser le modèle de croissance de cultures dans un milieu infesté par de mauvaises herbes. La méthode proposée commence par corriger les valeurs mesurées/estimées de LAI pour les champs infestés de mauvaises herbes pour l'apport en LAI de celles-ci, puis utilise des valeurs de LAI corrigées comme données d'entrée au modèle. Les résultats démontrèrent que par cette méthode le rendement et la biomasse furent simulés avec exactitude

ACKNOWLEDGEMENTS

My sincere thanks to my thesis supervisor, Dr. Shiv O. Prasher, for his continued guidance, understanding, and patience throughout this study - as well as for his assistance in the preparation of this dissertation. I greatly appreciate his deep knowledge in various disciplines, which contributed a lot to this project. Also, I admire his humor, enthusiasm, patience, and kindness during the course of my Ph.D. research. I am highly thankful for the confidence he showed in me for my research work. I would also like to thank my thesis co-adviser Dr. Heather McNairn, for her knowledgeable guidance toward my research project.

I wish to thank the other members of my comprehensive examination committee, Drs. Robert B. Bonnell, Ning Wang, Ian B. Strachan, and Vijaya Raghavan for their constructive criticism and suggestions. Additional thanks are extended to Dr. G.S.V. Raghavan, Professor and former Chairman of the Department of Bioresource Engineering, for his valuable support during the period of my research. I also appreciated the support of Dr. Robert Kok (department chair), Ms. Susan Gregus (administrative assistant), Ms. Trish Singleton (secretary) and Ms. Abida Subhan (secretary).

I acknowledge the help of the Faculty of Graduate Studies and Research, McGill University, and sincerely appreciate receiving a McGill Graduate Studies Fellowship and the Hugh Baily Award from McGill University.

I would like to express my appreciation to my dear friends Dr. Kamran Davary, Mr. Jamshid Jazestani, Dr. Pradeep K. Goel, Mrs. Sara Tafazoli, Mr. Marc-David Andrade, Mr. Yoji Yuno, Dr. M. Rashed, Dr. C.-C. Yang, and Mrs. Sobhalatha Kunjikutty for their assistance and support throughout the course of my research work. I would like to extend my thanks to Mr. P. Kirby, Mr. James, and Mr. Scott for their help in carrying out the fieldwork. I would like to express my appreciation for the late Mr. Peter Alvo, Mr. Vaughn Barnard, Dr. Georges Dodds, and Dr. R.M. Patel for their peer review of the manuscripts presented in this thesis. My special thanks go to Dr. Pierre Dutilleul for his valuable suggestions regarding the statistical analyses presented in this thesis.

I would like to express my deep gratitude to my lovely wife, Tahereh Farajzadeh, and my dear sons, Masoud and Alireza for their understanding, love, and spiritual support. Without them, I could not have carried out my Ph.D. studies.

Finally, I would also like to acknowledge my parents, my brothers and sisters for their help and support to let my dreams come true.

CONTRIBUTION OF AUTHORS

This thesis has been written according to the rules and regulations of the Faculty of Graduate Studies and Research of McGill University Guidelines for a Manuscript Based Thesis, where it is stated that:

"As an alternative to the traditional thesis format, the dissertation can consist of a collection of papers of which the student is an author or co-author. These papers must have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following:

1. Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis.)

2. The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory.

3. The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts.

The thesis must include the following:

(a) a table of contents;

(b) an abstract in English and French;

(c) an introduction which clearly states the rational and objectives of the research;

(d) a comprehensive review of the literature (in addition to that covered in the introduction to each paper);

(e) a final conclusion and summary;

4. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient

detail to allow a clear and precise judgment to be made of the importance and originality of the research reported in the thesis.

5. In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contributions of Authors" as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to clearly specify the responsibilities of all the authors of the co-authored papers".

Manuscripts based on the thesis:

1. Karimi, Y., S.O. Prasher, H. McNairn, R.B. Bonnell, P. Dutilleul, and P.K. Goel. 2004. Classification accuracy of discriminant analysis, artificial neural networks and decision trees for weed and nitrogen stress detection in corn. *Transactions of the ASAE* (In press).
2. Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, P. Dutilleul, and P. K. Goel, 2004. Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Transaction of the ASAE*, (In press)
3. Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, and P. K. Goel , 2004. Incorporating the effects of weeds in crop growth modeling. *Agricultural Systems*, (Submitted for publication).
4. Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, and P. K. Goel, 2004. Coupling STICS crop growth model with remote sensing data for corn growth simulation on eastern Canada. (under preparation).

All the experimental works, analysis of data, and preparation of above manuscripts were completed by the candidate, Y. Karimi-Zindashty, under the supervision of Dr. Shiv O. Prasher and Dr. H. McNairn. Both supervisors gave the guidance and Dr. Prasher provided the necessary funds to carry out this research.

Dr. P. K. Goel, Department of Bioresource Engineering, helped in carrying out the field experiment. The critical suggestion and help of Dr. R. B. Bonnell, Department of Bioresource Engineering, and the contribution of Dr. P. Dutilleul, Department of Plant Science, in the statistical analysis of Chapter three and Chapter four are gratefully acknowledged.

TABLE OF CONTENTS

ABSTRACT i
RÉSUMÉ ii
ACKNOWLEDGEMENTS iv
CONTRIBUTION OF AUTHORS vi
TABLE OF CONTENTS ix
LIST OF FIGURES xii
LIST OF FIGURES cont'd xiii
LIST OF TABLES xiv
LIST OF TABLES cont'd xv
LIST OF SYMBOLS xvi

CHAPTER 1 1
INTRODUCTION 1
1.1. Objectives 4
1.2. Thesis Organization 5
1.3. Scope 6

CHAPTER 2 7
LITERATURE REVIEW 7
2.1 Remote sensing 7
2.1.1 Basics of remote sensing 8
2.1.2 Optical remote sensing 10
2.1.3 Microwave remote sensing 11
2.2 Applications of remote sensing in agriculture 12
2.2.1 Crop growth models 13
2.2.1.1 Model inputs 15
2.2.2 Remote sensing and crop growth modeling 16
2.2.3 Estimation of crop parameters using remote sensing 16
2.2.3.1 Empirical methods 16
2.2.3.2 Semi-empirical models 17
2.2.3.3 Mechanistic models 18
2.2.4 Vegetation indices (VIs) used for the estimation of crop parameters 21
2.2.5 Coupling mechanistic crop models and remote sensing 23
2.2.6 Estimation of yield under weedy conditions 31
2.3 Concluding remarks on current status and future prospects 32

PREFACE TO CHAPTER 3 34
CHAPTER 3 35
WEED AND NITROGEN STRESS DETECTION IN CORN USING DISCRIMINANT ANALYSIS 35
3.1 Abstract 35
3.2 Introduction 36
3.3 Materials and Methods 38
3.3.1 Experimental detail 38

3.3.2 Spectral data collection 38
3.3.3 Weed survey 43
3.3.4 Data analysis 43
3.4 Results and discussion 44
3.4.1 Selection of Wavebands 44
3.4.2 Discriminant analysis 45
3.5 Conclusions 51

PREFACE TO CHAPTER 4 53
CHAPTER 4 54
WATER AND NITROGEN STRESS ASSESSMENT IN CORN BY APPLYING DISCRIMINANT ANALYSIS ON HYPERSPECTRAL DATA 54
4.1 Abstract 54
4.2 Introduction 54
4.3 Materials and methods 57
4.3.1 Study site and experimental details 57
4.3.3 Plant parameters 59
4.3.4 Data analysis 61
4.4 Results and Discussion 63
4.4.1 Selection of wavebands 63
4.4.2 Discriminant analysis 67
4.5 Conclusions 71

PREFACE TO CHAPTER 5 76
CHAPTER 5 77
MODELING CORN GROWTH IN SOUTHWESTERN QUEBEC WITH THE STICS MODEL 77
5.1 Abstract 77
5.2 Introduction 78
5.3 Materials and Methods 80
5.3.1 Study site and experimental details 80
5.3.2 Crop parameters 82
5.3.3 Methodology 82
5.3.3.1 STICS model and its modules 82
5.3.3.1.1 Inputs and outputs 84
5.3.3.1.2 Calibration of the model 85
5.3.3.1.3 Evaluation of the Model Performance 85
5.3.3.2 Estimation of LAI using RS data 89
5.3.3.2.1 Support vector machine 90
5.3.3.2.2 Theory of SVMs 90
5.3.3.3 Coupling STICS with remote sensing 93
5.4 Results and Discussion 93
5.4.1 Calibration and Evaluation of STICS Model 94
5.4.2 Estimation of LAI values 98
5.4.3 Coupling STICS model with RS 99
5.5 Conclusions 106

PREFACE TO CHAPTER 6 107
CHAPTER 6 108
INCORPORATING THE EFFECT OF WEEDS INTO CROP GROWTH MODELING 108
6.1 Abstract 108
6.2 Introduction 108
6.3 Materials and Methods 110
6.3.1 Experimental detail 110
6.3.2 Crop parameters 111
6.3.3 Weed survey 111
6.3.4 Methodology 113
6.4 Results and discussion 115
6.5 Conclusion 122

CHAPTER 7 124
SUMMARY AND GENERAL CONCLUSIONS 124
7.1 Summary 124
7.2 Conclusions 124

CHAPTER 8 127
CONTRIBUTIONS TO KNOWLEDGE AND SUGGESTIONS FOR FUTURE RESEARCH 127
8.1 Contributions to Knowledge 127
8.2 Suggestions for Further Research 128

CHAPTER 9 130
REFERENCES 130

APPENDIX A 152
STICS model and it's modules 152
A.1 Development 152
A.2 Shoot Growth 154
A.3 Yield Components Module 156
A.4 Root Growth 156
A.5 Water Balance 158
A.6 Crop Thermal Environment 159
A.7 Nitrogen Balance 160
APPENDIX B 162

LIST OF FIGURES

Figure 2.1	The schematic of the electromagnetic spectrum	**9**
Figure 2.2	Representation of the forcing strategy: the temporal behavior of one state variable of the model is derived from remote sensing observations and used as the input variable within the model (Source: Moulin *et al.* 1998).	**27**
Figure 2.3	Representation of the recalibration strategy: by comparing modeled LAI profile and 'LAI' derived from remote sensing data, some parameters of the crop model (or some initial conditions) are re-tuned (Source: Moulin *et al.* 1998).	**29**
Figure 2.4	Representation of the assimilation strategy: comparing simulated reflectance profiles (from the coupling of a crop production model and a reflectance model) and remote sensing derived reflectances allows some parameters of the crop model (or some initial conditions) to be re-tuned (Source: Moulin *et al.* 1998).	**30**
Figure 3.1	Experimental layout, 2000	**39**
Figure 3.2	Measured spectral response curves of corn at three different growth stages under different nitrogen application rates and weed control conditions.	**41**
Figure 4.1	Experimental layout, 2002 (N1 - N5 are: 50, 100, 150, 200, 250 kg N ha $^{-1}$, respectively, Irr and N-Irr are: irrigated and non-irrigated treatments)	**58**
Figure 4.2	Measured spectral response curves of corn under irrigated and non-irrigated conditions with different nitrogen application rates, July 31, 2002.	**60**
Figure 4.3	Canonical discriminant analysis of July 31, 2002 data with irrigation and nitrogen as class variables.	**72**
Figure 4.4	Canonical discriminant analysis of August 10, 2002 data with irrigation and nitrogen as class variables.	**73**

LIST OF FIGURES cont'd

Figure 4.5	Canonical discriminant analysis of August 26, 2002 data with irrigation and nitrogen as class variables.	**74**
Figure 5.1	Experimental layout, 2001 (Irr, N-Irr are irrigated and non-irrigated treatments, respectively)	**81**
Figure 5.2	Comparison of measured and estimated yield.	**95**
Figure 5.3	Comparison of measured and estimated biomass.	**95**
Figure 5.4	Comparison of measured and estimated LAI values.	**96**
Figure 5.5	Measured and predicted LAI values based on airborne hyperspectral observations at the initial stage (first flight, June 30, 2000).	**102**
Figure 5.6	Measured and predicted LAI values based on airborne hyperspectral observations at the tasseling stage (second flight, August 5, 2000).	**103**
Figure 5.7	Measured and predicted LAI value based on airborne hyperspectral observation at the full maturity stage (third flight, August 25, 2000).	**104**
Figure 6.1	Measured LAI value in weedy and no-weed plots.	**116**
Figure 6.2	Variation in mean LAI values in weedy treatments (N_{60}, N_{120}, and N_{250} represent low, normal (recommended), and high nitrogen application rates, respectively).	**118**
Figure 6.3	Variation in mean LAI values in no-weed treatments (N_{60}, N_{120}, and N_{250} represent low, normal (recommended), and high nitrogen application rates, respectively).	**119**
Figure 6.4	Comparison of the mean measured crop yield in weedy and no-weed treatments.	**120**
Figure 6.5	Simulated "True" and measured LAI values in weedy plots.	**120**
Figure 6.6	Simulated *vs.* measured yield values in weedy plots using complete data set (validation).	**121**

LIST OF TABLES

Table 3.1	Wavelengths selected by different methods in developing a classification model.	**46**
Table 3.2	Summary of misclassification matrices of DISCRIM procedure (%).	**48**
Table 3.3	Classification matrices for the cross-validation data for the nine weed-nitrogen combinations at the three growth stages.	**49**
Table 3.4	Classification matrices for the cross-validation data for weed controls and for nitrogen application rates, each at three growth stages.	**50**
Table 3.5	Comparison of the risk estimate result obtained from different methods of classification.	**50**
Table 4.1	Plant parameters as a function of different treatments (mean of four replicates).	**62**
Table 4.2a	STEPDISC results for waveband selection for **July 31, 2002, data set**	**64**
Table 4.2b	STEPDISC results for waveband selection for **August 10, 2002, data set**	**65**
Table 4.2c	STEPDISC results for waveband selection for **August 26, 2002, data set**	**66**
Table 4.3	Summary of misclassification matrices obtained from DISCRIM procedure (%).	**69**
Table 4.4a	Classification Matrix obtained from DISCRIM procedure for cross-validation cases in July 31 data sets.	**69**
Table 4.4b	Classification Matrix obtained from DISCRIM procedure for cross-validation cases in August 26 data sets.	**70**
Table 4.5	Canonical discriminant analysis for irrigation with nitrogen as class variable, using the first 20 selected wavebands (July 31, 2002).	**70**

LIST OF TABLES cont'd

Table 4.6	Eigenvalues of the correlation matrix calculated by the PRINCOMP procedure (July 31, 2002).	**75**
Table 5.1	Development stages in STICS.	**83**
Table 5.2	Mean monthly temperature relative humidity, and total precipitation for the growth season during the simulation period and for a normal year.	**86**
Table 5.3	Details on cultural practices.	**87**
Table 5.4	Adjusted model parameters.	**88**
Table 5.5	Statistical criteria used in evaluation of model performance.	**89**
Table 5.6	Basic statistics for estimated and measured corn crop parameters	**96**
Table 5.7	Statistics calculated for comparison of measured and estimated corn crop parameter values.	**97**
Table 5.8	Summary of regression prediction result using SVM software.	**100**
Table 5.9	Summary of regression prediction result using SVM software with leave-one-out method.	**101**
Table 5.10	Statistics calculated for comparison of measured and estimated LAI values using SVM model.	**105**
Table 5.11	Summary of measured and simulated (with and without using remote sensing data) yield and biomass for year 2000 data (Mg ha^{-1}).	**105**
Table 6.1	Details of various crop parameters based on different treatments during entire growth season.	**112**
Table 6.2	Summary of the regression equations and r^2 obtained using different data sets.	**123**
Table 6.3	Statistical parameters calculated for measured and simulated yield on weed treatments.	**123**

LIST OF SYMBOLS

The most commonly used symbols, abbreviations and acronyms are listed below. The specific symbols that are used in a particular equation or section are described at their place of appearance in the text.

μm	micrometer
ANN	artificial neural network
ASAE	American society of agricultural engineers
CASI	compact airborne spectrographic imager
CDA	canonical discriminant analysis
cm	centimeter
CV	coefficient of variation
DT	decision tree
EF	model efficiency
EF1	modified model efficiency
EOS	earth observing satellites
ERS	European remote sensing satellite
GDD	growth degree-day
GI	green index
ha	hectare
HRV	high resolution visible (sensors)
IR	reflectance values in infrared
IRENE	integrated resources for evaluating numerical estimates
JARS	Japanese association of remote sensing
JERS	Japanese earth resources satellite
kg	kilogram
LAI	leaf area index
LANDSAT	land satellite
lat.	latitude

long.	longitude
m	meter
MBE	mean bias error
MD	mean deviation
MSD	mean square deviation
MSAVI	modified soil adjusted vegetation index
MSS	Multispectral Scanner
N	nitrogen
n	number of paired observed-simulated values
NASA	national aeronautics and space administration
NDVI	normalized difference vegetation index
NIR	near-infrared
P	probability
PCA	principal component analysis
PCs	principal components
PCM	precision crop management
PE	processing element
R	reflectance values
r	correlation coefficient
r^2	coefficient of determination
RA	radiation adsorbed
RAR	real aperture radar
RI	ratio index
RMS	root mean square
RMSE	root mean square error
RRMSE	relative RMSE
RS	remote sensing
SAR	synthetic aperture radar
SAVI	soil adjusted vegetation index

SIR	Space Imaging Radar
SPAD	specialty products agricultural division (chlorophyll meter)
SPOT	satellite probatoire d'observation de la Terre
SVM	support vector machine
TM	Thematic Mapper
TSAVI	transformed soil adjusted vegetation index
VI	vegetation index
WDVI	weighted difference vegetation index

CHAPTER 1

INTRODUCTION

Production of food in a cost-effective way is the main goal of every farmer and agricultural manager. Crop growth is affected by different stresses (e.g., water, pest, weed) which can reduce the crop production. The impact of these stresses on crop yield has been widely documented in the literature (Wilcut *et al.*, 1987; McLennan *et* al., 1991; Donald and Khan, 1992; Knezevic *et al.*, 1994; Oerke *et al.*, 1994; Mamolos and Kalburtji, 2001). The impact of insects, weeds and diseases, represents a potential annual loss of 40% of world food production (Oerke *et al.*, 1994). Oerke *et al.* (1994) also indicated that weeds could cause a potential yield loss of 27 % per year in the corn fields of Canada. A recent study indicated that, compared to a weed-free control, mean yield losses in tomato and cabbage crops could rise as high as 80% and 60%, respectively (Tolman *et al.*, 2004). Considering that weeds often exhibit a patchy distribution (Marshall, 1988; Wiles *et al.*, 1992; Vangessel *et al.*, 1995), remote sensing techniques could be useful in identifying weedy areas in infested fields (Deguise *et al.*, 1999; Goel *et al.*, 2003 a and c).

Traditionally, in order to obtain greater production, larger quantities of agricultural inputs (e. g., pesticide, herbicides, fertilizer) have been applied, resulting in greater environmental pollution. In this respect, precision farming with the site-specific application of agricultural inputs, leads to an overall reduction in chemical inputs without affecting agricultural production (Tomer *et al.*, 1997; Christensen *et al.*, 1998). However, in implementing this strategy, accessibility to reliable and relevant on time information is imperative. Particularly in the case of management of agricultural market and price-fixing, at both national and regional levels, a beforehand assessment of crop production and available quantities for import/export or storage is crucial.

Currently, there are two main approaches to crop production estimation, namely statistical methods and crop-growth models. Statistical methods (Makowski *et al.*, 2001) are based on historical crop yield data. In these methods, crop yield is related to one or several input parameters (e.g., applied fertilizer). The relationship obtained is used to

predict the yield in an upcoming year. The main advantage of this approach is its ease of use; however, soil-crop parameters and weather data are not directly taken into account. The second approach is the use of crop growth models, which simulate and monitor crop growth by describing relationships between crop growth characteristics (planting date and rate, plant density and variety), environmental factors such as weather parameters (solar irradiation, precipitation, and temperature), soil parameters and nutrient availability (Childs *et al.*, 1977; Weir *et al.*, 1984; Maas, 1988a,b; Spitters *et al.*, 1989; Van Diepen *et al.*, 1989; Eckersten and Jansson, 1991; Hansen *et al.*, 1991; Allen *et al.*, 1996; Brisson *et al.*, 1998; Paz *et al.*, 1998; Irmak *et al.*, 2000; Denisov 2001). The advantage of these dynamic methods is the incorporation of crop-soil parameters and weather data in simulation processes.

Crop growth monitoring and yield estimation with crop models requires both temporal and spatial information about soil and agricultural practices during the entire growing season. However, collection of such information with existing methods is labor-intensive, time-consuming, and, in some cases, impossible. Remote sensing techniques have shown the potential to collect vast amounts of spatial information, frequently and over large areas (Erickson, 1984; Meyer-Roux, 1990; Atkinson and Tatnall, 1997). Past studies have shown that airborne and satellite sensors can observe crop conditions on a much larger scale and can measure parameters related to plant performance during the entire growing period. Such observations can have a significant role to play (Guérif and Duke, 1998; Singh *et al.*, 2002) in precision crop management (PCM). PCM is information and technology- based agricultural management system, which has the capacity to identify, analyze and manage spatial and temporal variability within fields in order to optimize profitability, sustainability and protection of the environment (Robert *et al.*, 1995). One of the most important components of PCM is the variable rate technology, which allows the targeted application of different cropping inputs (e.g., fertilizer, herbicide and seeds) on a specified location. Remote sensing observations can make variable rate technology more easily applicable. These inputs could not be used with traditional agricultural management methods that assume fields to be homogeneous elements.

Furthermore, remote sensing observations could be useful in crop yield estimation using crop models. However, as reported by Moulin *et al.* (1998), the potential of remote sensing in this area is not fully explored. For more quantitative analysis, more complex methodologies are needed to estimate yields on a regional scale, including the coupling of remote sensing data and crop production models, as was suggested by Wiegand *et al.* (1986). The potential of remotely sensed data in supplying required information regarding agricultural fields has been evaluated (Clevers and Leeuwen, 1996; Roerink *et al.*, 1997; Murai and Omatu, 1997; Michaud *et al.*, 1998; Sakthivadivel *et al.*, 1999; Clevers *et al.* 2000; Gillett *et al.*, 2001; Kurosu *et al.*, 2001; Nagendra, 2001; Sannier and Taylor, 2002; Singh *et al.*, 2002). The advantages of this technology (large area coverage and frequency) make it possible for the agribusiness community to become one of the primary users of remotely sensed data.

The collection of information alone is not enough. To fully use the collected and existing information, more dynamic and active crop monitoring and forecasting methods are needed. These methods should be able to use the remotely sensed data as well as the available ground information in order to take advantage of them both in providing information needed for decision-making. In this context, the main challenges for researchers in crop monitoring, crop yield prediction, and soil-crop parameter estimation are (i) to integrate remotely sensed information into crop growth modeling systems (Maas, 1988a,b) and (ii) to develop these models as useful and powerful tools for monitoring crops under different management practices and various meteorological scenarios. In the past, different crop growth models have been linked to remote sensing systems, e.g. SUCROS (Clevers and Leeuwen, 1996; Clevers, 1997; Guérif and Duke, 1998), SWAP (Michaud *et al.*, 1998), CROPGRO-Soybean (Irmak *et al.*, 2000), ROTASK (Clevers *et al.*, 2000) and STICS (Prévot *et al.*, 2000a,b).

Among these models, the STICS model is considered as more generic and can be modified with ease for various crops. The main features associated with the STICS model, which differentiate it from the other crop models, are: (i) it is not fully dependent on the data reflecting current field conditions; (ii) it can operate in various *"agricultural conditions with readily available spatialized inputs"*, and (iii) it is an *"engineering model"*, (i.e., comparatively more adaptable to the needs of farmers and policy makers)

(Brisson *et al.*, 1998). Moreover, STICS can make direct use of observed data, such as Leaf Area Index (LAI), an important parameter reflecting ground biomass cover. Given its adaptability and other characteristics enumerated above, the STICS model seems to have an edge over other models, particularly when attempting to link it to remote sensing observations.

1.1. Objectives

The main goal of this study was to investigate the possibility of using hyperspectral observations for the detection of combined effects of various types of stresses, namely weed and nitrogen, and nitrogen and water. The study also aimed to evaluate the performance of the STICS crop growth simulation model for southwestern Quebec and to link the STICS model with hyperspectral data measured on corn plots under different growth conditions. This investigation was carried out under controlled conditions, reflecting different nitrogen application rates, weed control strategies, and different irrigation regimes. The more specific objectives of the proposed study were:

1- to investigate the use of airborne hyperspectral observations in the classification of weed and nitrogen stresses in a corn field,
2- to examine the ability of very narrow ground-based hyperspectral observations in discrimination of different water and nitrogen stresses,
3- to evaluate the performance of STICS, a crop growth model, in simulating corn growth and yield in southwestern Quebec,
4- to link STICS with hyperspectral data in order to obtain a better estimate of crop biomass and yield in cornfields, and to enhance the performance of the STICS model in estimating crop biomass and yield in corn fields, and
5- to incorporate the effect of weed stresses into crop growth modeling.

To meet these objectives, extensive studies were carried out in structured corn field plots under different cropping management strategies. Experiments were conducted at the Research Farm of Macdonald Campus of McGill University. The studies were carried out over a period of three years. Remote sensing observations were acquired from both a handheld spectroradiometer and an airborne sensor at different critical growth stages. Extensive ground measurements of soil parameters and crop physiological

parameters were also collected. The applied methodology and experimental procedures to meet the objectives of the study are discussed in the following chapters.

1.2. Thesis Organization

This thesis contains nine chapters and one appendix. Chapter 1 gives an introduction to the subject, followed by a listing of the objectives and the scope of this investigation. In Chapter 2, a review of the relevant and applicable literature is provided. Chapter 3 provides information on the first year of the study which focused on monitoring corn growth under different weed control strategies and nitrogen application rates. A hyperspectral airborne sensor was used to acquire spectral observations. The focus of the chapter is on the selection of suitable wavelength regions in order to detect weed and nitrogen stresses, using discriminant analysis. The results of discriminant analysis were further compared with the results obtained using ANN and decision tree methods. A paper based on this chapter has been accepted for publication in the *Transactions of the ASAE* (in press).

Chapter 4 is based on the data collected in the third year of the study, which investigated the effects of irrigation and nitrogen fertilization on the spectral response of corn. A hyperspectral handheld spectroradiometer was used to record spectral observations. Using discriminant analysis, suitable wavelengths or wavelength combinations were selected to detect nitrogen and water stresses in the corn field. A paper based on this chapter has also been accepted for publication in *Transactions of the ASAE* (in press).

Chapter 5 summarizes the results of the work completed on the use of a crop growth model for monitoring crop growth under different agricultural practices and, more precisely, linking of the crop growth model to hyperspectral aerial remote sensing observations. A manuscript based on this chapter is under preparation.

Chapter 6 focuses on the development of a procedure that can incorporate the effect of weeds into crop growth modeling for better estimation of yields. The procedure is based on correcting the LAI values and using these corrected LAI values in the model. A paper based on this work has been submitted for publication in *Agricultural Systems Journal*.

Finally, Chapter 7 summarizes this study and lists the main conclusions derived from this work. Chapter 8 lists the major contributions to knowledge and suggestions for future research. Chapter 9 contains a complete list of references. Detailed explanation on the STICS model is provided in Appendix A.

1.3. Scope

This study was carried out from the year 2000 to 2002 in corn fields in southwestern Quebec. While the first two years of the study were located on clay soil (Bearbrook and Ste. Rosalie from the Dark Gray Gleysolic group), the observation in the last year was made on sandy soil (St. Amable complex). In all three years, the same corn cultivar (*Hybrid DKC42-22*) was used. This study takes into account three types of the crop stresses namely: weed, nitrogen, and water. The results obtained in this work are limited to these conditions only and any extrapolation of the results to other corn cultivars and/or soil types should be done with caution.

CHAPTER 2

LITERATURE REVIEW

In this chapter, the relevant literature on the application of remote sensing in agriculture is reviewed. The subjects of this review work are the application of remote sensing (RS) in stress detection, combining RS data and crop growth simulation models, and prediction of yield under weed infestation, followed by a discussion of current methodological trends. First, the basics of remote sensing are introduced, followed by a definition of the common applications of remotely sensed data in agricultural fields and estimation of the crop growth parameters. The discussion continues with a brief introduction of the various crop growth simulation models, and an overview of relationships between crop state variables and crop growth models. Next, the related research on yield loss estimation resulting from weed infestation effects will be reviewed. Finally, various approaches to coupling crop growth with remote sensing observations are discussed.

2.1 Remote sensing

Remote sensing is the science of collecting information from a target, an object or phenomenon without touching it. This can be accomplished by recording reflected or emitted energy from the target and analyzing and applying the processed information in real conditions. Our eyes are a basic and simple example of remote sensing sensors. While the eyes are reading a page or seeing an object, the amount of reflected energy is being assessed. These data are processed or interpreted by our brain to enable us to identify different aspects. Usually, remote sensing deals with the interaction between incident solar radiation and targets of interest. However, in some cases, it involves the sensing of the emitted energy by the target object.

2.1.1 Basics of remote sensing

The energy source, which is the prerequisite of the remote sensing system, provides electromagnetic radiation for illuminating the target. The next requirement of remote sensing is the interaction of radiation with the targets. Depending on the properties of both the radiation and target, a portion of the energy will be reflected. In this step, the system needs a sensor to record the reflectance. The recorded information needs to be transmitted and processed into an image. The image has to be analyzed to extract the required information about the target of interest. The final element of the system is the application of remote sensing, which is achieved by using extracted data about the target to achieve a better understanding, discover new information or contribute to the solution of specific problems.

The electromagnetic spectrum varies from shorter wavelengths to longer wavelengths (Figure 2.1). The spectrum can be divided into several regions (including visible, infrared, and microwave), which are useful for remote sensing. The first region, which is the most widely used part of the spectrum, is the visible portion and covers the wavelengths of 0.4 to 0.7 μm. The second interesting portion for remote sensing is the infrared (IR), with wavelengths ranging from approximately 0.7 to 100 μm. Based on the radiation properties, this portion can be categorized into three regions: near IR (0.7 – 1.3 μm), mid IR (1.3 – 3.0 μm), and thermal or emitted IR (3.0 – 100 μm). The applications of near and mid-IR in remote sensing are very similar to that of the visible portion. In thermal IR, the radiation is emitted in the form of heat from the surface of the earth. The third region is the microwave region (1 mm – 1 m), which lies at much longer wavelengths of the spectrum and are the longest wavelengths used for remote sensing.

Remote sensing can be classified in different ways. A remote sensing system is called passive if the energy is provided naturally (e.g., the sun). On the other hand, an active remote sensing system provides its own energy source for illumination of the target.

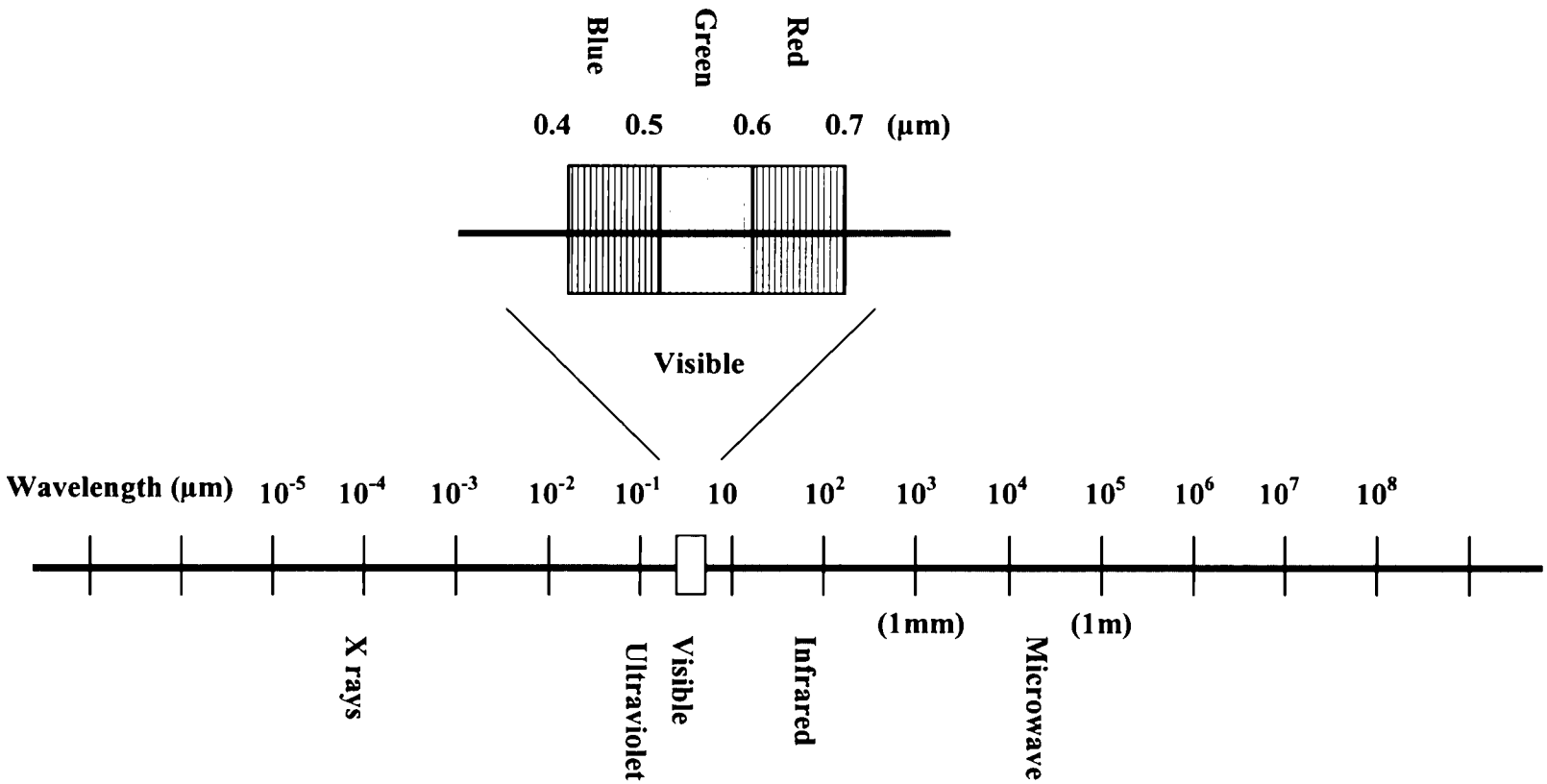


Figure 2.1 The schematic of the electromagnetic spectrum.

The radiation will be transmitted toward the target and reflected radiation from that target is detected and measured by the sensor. The advantage of active remote sensing is the ability to acquire measurements anytime, regardless of the time of day or season (CCRS, 2003). Remote sensing systems can also be classified as optical and radar, based on the region of the spectrum that is used. This will be explained further in the following sections.

2.1.2 Optical remote sensing

Remote sensing systems which use visible, near infrared and short-wave infrared portions of the spectrum are called optical remote sensing systems. Different materials reflect radiation in different ways at various wavelengths, resulting in different colors and brightness levels. Thus, spectral reflectance signatures can be used to differentiate targets within remotely sensed images. Optical imaging systems (commonly in use) can be categorized into three types:

a) **Panchromatic imaging system:** This system has a single channel detector sensor with sensitivity to radiation within a broad wavelength range. When the wavelength ranges within the visible portion, the resulting image looks like a "black-and-white" photograph taken from space. The physical characteristics of the targets could be measured from the brightness of the image. Thus, spectral information or the "color" of the targets is lost. The SPOT HRV-PAN satellite is an example of a panchromatic imaging system (Lillesand and Kiefer, 2000).

b) **Multispectral imaging system:** This system has a multi-channel detector sensor. Each channel can detect the reflectance of the radiation within a wide wavelength band, which provides a multilayer image. LANDSAT MSS, LANDSAT TM, and SPOT HRV-XS (Lillesand and Kiefer, 2000) are some examples of multispectral systems.

c) **Hyperspectral imaging system:** In comparison to multi-spectral remote sensing, which records reflectance from a target in a few broad channels, a hyperspectral imaging system acquires information in more than 100 very narrow, defined continuous spectral bands (Lillesand and Kiefer, 2000). In this system, reflected radiation from any specified target has been obtained continuously, making it

possible to gain detailed information on the materials at any target (Lillesand and Kiefer, 2000). Using these narrow bands, the features that have diagnostic absorption and reflection properties in such narrow wavelength intervals can be differentiated, which was not possible with the wider wavebands used in multispectral sensors. These narrow wavebands make hyperspectral remote sensing systems powerful tools that have the potential to avoid time consuming and labor intensive ground data collection methods. A hyperspectral sensor covers the visible and shortwave infrared regions of the spectrum. The Earth Observing-1 (EO-1) satellite, launched in November, 2000 (NASA), carries onboard hyperspectral sensors (Hyperion) is an example of such a system.

2.1.3 Microwave remote sensing

The microwave portion of the spectrum can provide useful information about land as well as water targets. Both active and passive microwave sensors can be used (JARS, 1999). A passive sensor records the natural microwave emission from the Earth, while an active microwave sensor transmits its own microwave signals and detects the backscattered signals from the target. In radar-based remote sensing, different properties of the target can be obtained from measuring the time delay between the transmitted and received signals and recording the strength of the received signal.

As described previously, the microwave region of the spectrum covers the wavelength range from approximately 1 mm to 1 m. Microwave wavelengths are considerably longer than those in the visible and infrared spectral regions. These longer wavelengths have special properties that are important for remote sensing. Longer wavelength microwave radiation has the capability of penetrating through cloud cover, haze and dust (Lillesand and Kiefer, 2000). Thus radar imagery can be collected under most weather conditions and because radars provide their own source of energy, data collection can take place day or night. Consequently, radar acquisitions are more reliable particularly in regions where persistent cloud cover is a problem. (Lillesand and Kiefer, 2000). Spaceborne radar remote sensing was started by SEASAT in 1978 (Lillesand and Kiefer, 2000). Some of the satellites with radar sensors include SIR, ALMAZ, ERS, JERS, and RADARSAT.

2.2 Applications of remote sensing in agriculture

Remote sensing tools have the potential for providing information about vegetation in various wavebands: using visible wavelengths (Guyot 1996), active or passive microwaves (Prévot *et al.*, 1993) and emitted thermal wavelengths (Seguin *et al.*, 1991, 1994; Moran *et al.*, 1994). Earlier research on applications of remote sensing in agriculture focused on the visible and near-infrared regions of the spectrum (Gates *et al.*, 1965; Woolly, 1971). These studies found an inverse relationship between spectral reflectances and radiances in the red region of the electromagnetic spectrum and *in situ* chlorophyll density. but a direct relationship between spectral reflectances and radiances in the near-infrared region and the green leaf density. Tucker *et al.* (1981) correlated the red (0.65-0.70 μm) and infrared (0.775-0.825 μm) spectral reflectances collected using a hand-held radiometer with total aboveground winter wheat (*Triticum æstivum L.*) biomass accumulation over the growing season. The authors reported a high correlation between the spectral data and the vigor and state of the plant canopy.

Remote sensing techniques have been widely used in different agricultural domains, such as crop identification (Foody *et al.*, 1989; Saha and Jonna, 1994; Cho *et al.*, 2002), detection of different crop stresses (Cibula and Carter, 1992; Carter, 1994; Lelong *et al.*, 1998; Goel *et al.*, 2003a,b), crop diseases (Lorenzen and Jensen, 1989; Peñuelas *et al.*, 1995), weed detection (Curran, 1985; Everitt *et al.*, 1987, 1995, 1996; Brown *et al.*, 1994; Brown and Steckler, 1995; Goel *et al.*, 2002; Vrindts *et al.*, 2002), irrigation performance (Bastiaanssen and Bos, 1999; Sakthivadivel *et al.*, 1999; Bastiaanssen *et al.* 2000), yield estimation (Maas, 1988b; Clevers, 1997; Singh *et al.*, 2002), and in general, crop management and precision farming (Pearson *et al.*, 1994; Wallace, 1994; Moran *et al.*, 1997a; Anderson *et al.*, 1999). Based on these research findings, one can remark that remote sensing has been involved in agricultural projects all over the world.

During the last two decades, several projects have examined the use of remotely sensed observations for crop monitoring and yield estimation. For example, one could mention the Large Area Crop Inventory Experiment (LACIE) in the USA (Erickson, 1984) and the Monitoring Agriculture with Remote Sensing (MARS) program of the European Union (Meyer-Roux, 1990) projects, which are still in progress. Remote

sensing data, specifically satellite data, have successfully been applied in crop acreage estimations (Sharman *et al.*, 1992). Recently, for monitoring crop growth and yield estimation, more attention has been given to linking crop growth models to remote sensing observations (Delécolle *et al.*, 1992; Maas, 1993c; Fischer *et al.*, 1997; Moulin *et al.*, 1998; Bach *et al.*, 2001). In the following section, a brief discussion of crop growth models is provided and the importance of remote sensing for crop growth modeling is summarized.

2.2.1 Crop growth models

The soil–plant–atmosphere continuum, which is the most important part of any agro-ecosystem, is a complicated and multidimensional system. Within this system, nonlinear and complex relationships exist between inputs and plant response. Empirical models are insufficient for the simulation of such systems. Mechanistic crop growth modeling has shown potential for simulating such complex systems. Currently, no model has been able to describe a complex system in a complete and exact manner. Models simulate the system in a given circumstance, and therefore fulfill the purpose of an individual inquiry or application. But a set of different modules are a prerequisite for multiple applications of the model.

Crop modeling is used as a technique to simulate and monitor crop growth by describing relationships between crop growth and cropping factors (planting date and rate, plant density and variety), environmental factors such as weather parameters (solar irradiation, precipitation, and temperature), soil parameters and nutrient availability (Van Diepen *et al.*, 1989; Williams *et al.*, 1989; Eckersten and Jansson, 1991; Hansen *et al.*, 1991; Maas, 1993a; Brisson *et al.*, 1998). The model's computation is based on the daily rate of crop growth and development. These models can predict dry matter production from emergence until maturity as well as the final yield at harvest time. In these models, biomass is simulated on the basis of the accumulation of growing degree-days (GDDs), which use $0^{\circ}C$ as the base temperature. These models are developed with different techniques such as parameterization, a technique which approximates the response of a physical system by using an empirical function including coefficients over a specified range of environmental conditions (Maas, 1993a). In this context, three variables (live

leaf index, aboveground dry biomass, and accumulated growing degree-days) specifying the state of the simulated crop at any specified time are used. The accuracy of the model depends on achieving the proper parameter values which evaluate the response of the model to environmental situations (Maas, 1993b).

Previously used crop models usually simulated crop growth and yield under homogeneous conditions. With the beginning of precision farming, crop models have been elaborated to be able to simulate plant growth in non-homogeneous areas (Irmak *et al.*, 2000). Crop models have been used for studying spatial problems, including:

a) diagnosing factors that cause observed spatial variability such as variations in soil properties, water stress, nematodes and weeds (Allen *et al.*, 1996; Paz *et al.*, 1998, 1999),
b) evaluating different options for varying management to maximize profit and/or minimize losses of nutrients to the environment (Paz, 2000), and
c) predicting yield of a spatially-variable field as the season progresses, taking into account up-to-date daily weather data (Hoogenboom *et al.*, 1993; Welch *et al.*, 1999; Seidl *et al.*, 2000).

The first generation of mechanistic crop growth models started with simulating the behavior of particular plants. The CORNGRO (Childs *et al.*, 1977), ARCWHEAT (Weir *et al.*, 1984), CERES-wheat (Ritchie and Otter, 1985), SUCROS (Spitters *et al.*, 1989), and CROPGRO-soybean (Irmak *et al.*, 2000) are some examples of such models. These models were formulated to be able to simulate the growth of a specific crop. Their application is limited to that particular crop, unless the model is reconstructed for a new crop type.

In addition to the single crop models, there are also generic models, which were constructed to be able to simulate the growth of a variety of arable crops. These models were built in a manner such that the necessary changes in the specific parameters were provided (e.g., DAISY (Hansen *et al.*, 1991), SOILN (Eckersten and Jansson, 1991), EPIC (Williams *et al.*, 1989), WOFOST (Van Diepen *et al.*, 1989), and STICS (Brisson *et al*, 1998))

According to Sinclair and Seligman (1996), there is no common crop model which could be used in all situations. However, based on Brisson *et al.* (1998), the STICS

crop growth model was developed such that it could be applied to different crop types by modifying only a few parameters and some functions involved in calculating yield. The modular structure of the STICS model makes it possible for new components to be added, modified, and maintained with a minimal effort. Thus, STICS's structure is designed in a manner that makes it possible to work with the existing data as well as new data arising under different agricultural conditions.

2.2.1.1 Model inputs

Most of the crop growth simulation models require several input data. These inputs are meteorological information (daily maximum and minimum temperature, solar radiation and rainfall), management practices (cultivar, plant population, row spacing, fertilizer, and irrigation amount and date), and soil characteristics (soil type and texture, water holding capacity, bulk density), to calculate crop growth as a function of photosynthesis, growth stage, and water and nitrogen stresses. These data must usually be collected on a daily basis. Since the models require such a large number of input parameters, their applicability on a wider scale is limited. In addition, the crop growth modeling is so complex that it is difficult to obtain good reliable estimates with these models without considering the spatial variability of the field.

Khakural *et al.* (1996) and Paz *et al.* (1998) reported variability in soil properties as the main causes of spatial yield variability in soybeans [*Glycine max* (L.) Merr.]. Thus, in site-specific crop management (SSCM), the spatial variabilities of soil properties are defined as important parameters in crop models. These parameters need to be collected or estimated from existing data. A soil parameter measurement for every single grid within a field is time consuming, expensive and impractical. In the past one or more unidentified soil characteristics were derived by correlating estimated and observed yield data over different seasons (Paz *et al.*, 1998, 1999). Alongside the management practices, the leaf area index (LAI) is a major driving variable defining the condition of the crop in many crop growth models for crop monitoring and yield estimation. It reflects the general state and health of the crop and has an important influence on simulating crop growth within the model. However, new and fast techniques with well defined algorithms or rules are

needed to speed up the calibration process involved in incorporating available information into the estimation procedure (Irmak *et al.*, 2000).

2.2.2 Remote sensing and crop growth modeling

Since the very beginning, remote sensing has been used to estimate crop yield at different scales (Moulin *et al.* 1998). However, current research is more directed towards the estimation of various crop conditions, with the ultimate aim of improving yield estimation. Researchers have adopted various approaches for the above purpose. In the following section, work on the direct estimation of crop parameters based on spectral observations and strategies adopted to link crop growth models and remote sensing are discussed.

2.2.3 Estimation of crop parameters using remote sensing

It is possible to relate remotely sensed observations to values of various canopy state variables using completely empirical relationships or physical models. In general, three approaches have been used for estimation of biophysical parameters, such as LAI or yield production. A brief description of these approaches is given in the following sections.

2.2.3.1 Empirical methods

Empirical methods are the most commonly used approaches for estimating crop parameters such as LAI or biomass production. These methods are based on fitting an empirical function between the crop character (e.g., biomass, yield) and vegetation indices [such as Normalized Difference Vegetation Index (NDVI), Ratio Index (RI), etc.]. In these methods, the biological or physical parameters of the crops are not taken into consideration. Tucker *et al.* (1981) and Gilabert *et al.* (1996) have reported strong correlations between production and a cumulated amount of vegetation indices on various crops during the growing season. Similar results were obtained using satellite data (Tucker *et al.*, 1985) based on cumulated indices and proposing an empirical model for biomass production. The successful application of the empirical approaches for

production estimation has been reported in several studies (Hatfield, 1983; Hamar *et al.*, 1996; Hayes and Decker, 1996; Murthy *et al.* 1996).

Using empirical methods, Aase and Siddoway (1981) and Asrar *et al.* (1985a) addressed applicability of multispectral reflectance for predicting dry biomass production in wheat crops. The limiting factor of empirically derived relations for estimation of dry biomass production from spectral reflectance data was the fact that all empirical relationships between cumulated Vegetation Indices (VIs) and dry biomass are only valid on a local scale. Estimation of the production in any condition needs a description of how photosynthetically active absorbed energy is converted into dry biomass and partitioned to harvestable organs (Moulin *et al.*, 1998). This could be achieved using mechanistic or dynamic models to incorporate remotely sensed data and estimate major crops production (wheat, maize, etc.) in exhaustive agricultural systems.

2.2.3.2 Semi-empirical models

In general, the semi-empirical approaches have mostly been used for yield estimation. A semi-empirical model was originally developed by Monteith (1977), which was modified by Kumar and Monteith (1981) to use radiometric observations. According to Monteith's (1977) model's formulation, there is a strong relationship between the daily production of dry matter and the cumulative radiation quantity absorbed by the foliage during the growth period. Final biomass is estimated from the daily production of dry matter, which is incorporated during the growing season. The yield is forecasted from the final biomass.

With Monteith's model, it is possible to estimate dry matter production of crops using remote sensing data in different wavelengths (Asrar *et al.*, 1985; Leblon *et al.*, 1991; Güerif *et al.*, 1993; Loudjani *et al.*, 1995; Sinclair and Muchow, 1999). The main advantage of the semi-empirical models is the simplicity of their application given their limited data requirements, while the disadvantage of the semi-empirical models lies in their lack of an explanation of physiological and biological behavior of the system, which in turn, controls the growth and development of the plant. To achieve the simulation of the effect of these mechanisms in crops, the use of mechanistic models is necessary.

2.2.3.3 Mechanistic models

Mechanistic crop growth models are built and developed to simulate the effect of cultural practices, climate conditions, and soil parameters at the crop-soil-atmosphere interfaces on crop growth and yield. As described in previous sections, there are several available models for the major crops. For some of them, mechanistic relationships describe yield-driving processes and their interactions (e.g., AFRCWHEAT (Weir *et al.*, 1984); CERES-Maize (Jones and Kiniry, 1986); SUCROS (Spitters *et al.*, 1989); STICS (Brisson *et al.*, 1998); ROTASK (Clevers *et al.*, 2000)) with a daily running time step. These models simulate the amount of LAI throughout the growing season, which is of major importance as it drives the absorption of solar radiation and evapotranspiration, and thus carbon assimilation. However, canopy development and allocation of daily assimilates to leaves are described through empirical relationships.

The use of imprecise coefficients within a relationship affecting canopy development may lead to important errors in the estimation of biomass production (Porter 1984). Replacing the uncertain simulations by an estimation of crop state variables, derived from remote sensing throughout the growing season was recommended by some modelers (Maas *et al.*, 1989; Moulin *et al.*, 1998). Currently, researchers have tried to make a link between remote sensing observations and mechanistic crop growth models (Clevers and van Leeuwen, 1996; Guérif and Duke, 1998; Chauki *et al.*, 1999; Prévot *et al.*, 2000a, b; Bach *et al.*, 2001), which will be discussed in detail in coming sections. Moreover, realizing the importance of mechanistic models, some of the most used models are described below. The STICS model, which was selected for the present study, is described in greater detail in the Chapter 5 and Appendix A.

CROPGRO model: CROPGRO is a mechanistic model for grain legumes, which was originally modified from SOYGRO (Wilkerson *et al.*, 1983). It is a process-based model with a modular structure and runs in a daily time-step using weather, crop and soil data. The crop development, crop carbon balance, crop and soil N balance, and soil water balance are the main modules (subroutines) of the model (Boote and Jones, 1998). The crop tissue is derived from carbon balance, which is computed from photosynthesis. Plant tissue growth takes into account the leaf expansion; vegetative tissues; pods; seeds; and

rate calculations for shell growth; seed growth; nodule growth; senescence and carbohydrate mobilization. Crop development stages are juvenile, vegetative, reproductive and senescent phase. Crop simulation accuracy is dependent on how accurate these stages times and duration are predicted. The development stage is described as a function of the temperature, photoperiod, and water deficit. In crop nitrogen balance, a different form of nitrogen exchange takes place. The whole process of nitrogen exchange includes nitrogen from soil uptake, nitrogen fixation, mobilization from vegetative tissues, nitrogen use for new organ growth, and rate of nitrogen losses. The elements of soil water balance are infiltration of rainfall and irrigation, soil evaporation, distribution of water in the soil profile, root water uptake, water drained from the root zone, and crop transpiration.

The CROPGRO-soybean model was calibrated for growth and yield prediction in Galicia, Northwest Spain (Ruíz-Nogueira *et al.*, 2001). The calibration of the model was performed for both fully water supplied and rainfed conditions. The results indicated that an increase in soil water holding capacity would slightly increase root elongation, which in turn would accelerate late phenological development under conditions of water shortage. This, in turn, improved CROPGRO-soybean performance under rainfed conditions.

CERES model: The CERES model is a user-oriented system and initially was developed on the basis of the crop species e.g., CERES-maize (Ritchie *et al.*, 1989), and CERES-wheat (Ritchie and Otter, 1985). It has evolved from a species basis model to a generic model. CERES is a multipurpose simulation model capable of predicting yield over a large area, simulating crop growth within a growing season as well as performing multi-year risk analyses. The CERES model is constructed to simulate the effects of different cropping management, soil water, weather, and nitrogen fertilization rates on crop growth and yield. It runs on a daily time-step using weather data, cropping management and soil information. Like most of the generic models, it has modular structure and its description is comparable with CROPGRO. In crop development modules the growing period is divided into nine stages. Like CROPGRO, the development stage in CERES is described as a function of the temperature.

The CERES model was evaluated for different crops by several researchers. Xie *et al.*, (2001) evaluated the abilities of three crop models (ALMANAC, CERES-Maize and SORKAM) to simulate maize and sorghum [*Sorghum bicolor* (L.) Moench] grain yields in a dry growing season at several sites in Texas. For maize, the root mean square deviation values were 0.56 and 0.83 Mg ha^{-1}, and values for coefficient of determination (r^2) between measured and simulated grain yields were 0.95 and 0.88 with ALMANAC and CERES-Maize, respectively. Comparing the simulation of maize with CERES-Maize under irrigated and dry land conditions, the mean errors were 2.0% and 2.2% respectively; while for ALMANAC the mean error was 6.2% under both conditions. In another work, Hasegawa *et al.* (2000) evaluated the ability of the CERES models to simulate N dynamics during wheat and maize growth following legume cover crop incorporation. Their results showed that the values of soil inorganic nitrogen content and crop nitrogen uptake were generally within 20% of estimated ones. Similarly, Travasso and Magrin (1998) calibrated and validated CERES-barley for different genotypes sown on different dates under optimal growing situations in an Argentine climate. They reported that CERES-barley simulated grain yield reliably under normal sowing dates. Also, Jamieson *et al.* (1998) evaluated the performance of CERES-Wheat in comparison with AFRCWHEAT2, SIRIUS, SUCROS2, and SWHEAT for a winter sowing of wheat under New Zealand conditions. CERES-Wheat simulated yield with reasonable accuracy and performed well compared to AFRCWHEAT2, SIRIUS, and SUCROS2; however, SWHEAT underestimated yield in a fully irrigated treatment and under a varying water supply.

SUCROS (Simple and Universal CROp growth Simulator) is a mechanistic growth model, which was originally developed to simulate potential production (Spitters *et al.*, 1989) and adapted to different crops. Crop yield and dry matter accumulation are simulated using weather information under fully supplied water and nutrient conditions without taking into account the effects of pests, disease and weeds. SUCROS was modified in order to be able to simulate crop growth under water-limited conditions (SUCROS2) (Jamieson *et al.*, 1998). Dry matter accumulation was simulated based on the rate of CO_2 produced by photosynthesis in the crop canopy. The model simulates the

different organs of the crop (leaf, stem, root and storage organ) by the use of radiation, crop characteristics and environmental conditions. During the simulation process, biomass is divided into different organs on the basis of the crop development stage, which is determined by the use of temperature.

The SUCROS model was calibrated and used by Guérif and Duke (1998) to simulate sugar beet (*Beta vulgaris* L.) growth using optical remote sensing data. Their results showed that crop yield estimation was improved. Simane *et al.* (1994) used SUCROS-87 to study the effect of water stresses during the growing season in wheat crops in different regions of Ethiopia. Overall, mean simulated wheat yield was higher than the actual national average. As mentioned earlier, Jamieson *et al.* (1998) also used SUCROS2 along with four other models to study wheat crop growth under drought conditions. SUCROS, along with three of the models, simulated wheat yield with a reasonable accuracy.

2.2.4 Vegetation indices (VIs) used for the estimation of crop parameters

Several VIs based on the combination of two or more wavelengths in different ways have been developed and applied in characterizing plant growth and development (Jackson and Huete, 1991). A large number of studies (Boochs *et al.*, 1990; Miller *et al.*, 1991; Plummer *et al.*, 1991; Bach *et al.*, 1995; Patel *et al.*, 2001) have addressed the usefulness of the red edge of the reflectance spectrum, and especially the wavelength position of the red edge, for vegetation studies on different crops. In most of these studies, a shift of inflection wavelength towards longer wavelengths at different growth stages was reported.

There are more than a dozen vegetation indices (Myneni *et al.*, 1995) which have been correlated with different vegetation characteristics (e.g., vegetation amount, fraction of absorbed photosynthetically active radiation). Some widely used indices are Ratio Index (RI), Normalized Difference Vegetation Index (NDVI), Greenness Index (GI), Soil Adjusted Vegetation Index, (SAVI), Transformed SAVI (TSAVI), and Modified SAVI (MSAVI). These vegetation indices have been used to make quantitative estimates of the LAI, biomass, and percent soil cover (Steven *et al.*, 1983; Asrar *et al.*, 1984; Perry and

Lautenschlager, 1984). Some of the most frequently used vegetation indices are discussed below.

Normalized Difference Vegetation Index (NDVI)

NDVI is calculated as the difference of Near Infrared (NIR) and Red reflectances divided by the sum of those reflectances. NDVI is a widely used index for crop growth monitoring (Asrar *et al.*, 1985b; Moulin and Guérif, 1999; Seidl *et al.*, 2000; Towner *et al.*, 2000). Aase and Siddoway (1981) found a good relationship between NDVI and straw and total dry matter yield from late tillering until the beginning of the flowering stages of crop types. Towner *et al.* (2000) calculated both the RED [(NIR-Red) / (NIR+RED)] and GREEN [(NIR-GREEN) / (NIR+GREEN)] NDVI of corn and soybean plots by using airborne imagery (ADAR 5500) and a handheld (SE-590) hyperspectral scanner, respectively. Towner *et al.* (2000) showed that the Red NDVI saturates at low LAI values, while the Green NDVI saturation happens at higher (>2.5) LAI values (Gitelson *et al.*, 1996). Asrar *et al.* (1985b) used an indirect method, described by the following empirical equation (Asrar *et al.*, 1984) to correlate NDVI values for the effect of soil background and the scattering of the near-infrared radiation by plants in order to get an estimate of radiation absorbed (RA).

$$RA = -0.185 + 1.20 \times NDVI \qquad 2.1$$

where

RA is the estimated fraction of photosynthetic radiation absorbed by plants. This RA value was used to calculate the LAI as follows:

$$LAI = -\overline{\ln(1 - RA)} \,/\, \overline{K} \qquad 2.2$$

Where:

$\overline{\ln(1 - RA)}$ is the arithmetic mean of estimated RA

$\overline{K}$ is a mean leaf angular shape coefficient, which can be calculated as:

$$\overline{K} = 0.5 / \cos\eta \qquad 2.3$$

Where:

η is the solar zenith angle .

In a recent study, Singh *et al.* (2002) used NDVI, derived from satellite data, to estimate crop yield in a small area. Their results showed that NDVI provided more efficient predictions of crop yield.

Ratio Index (RI)

The RI is the simplest index which has appeared in the literature for estimating the LAI values. It is defined as the ratio of near-infrared (NIR) and red reflectance. Maas (1993c) used RI to calculate the LAI by applying the following regression based function.

$$LAI = 0.279 \times RI - 0.448 \qquad 2.4$$

Transformed Soil Adjusted Vegetation Index (**TSAVI**)

The TSAVI is another index which was used by some authors (Baret and Guyot, 1991; Moulin and Guérif, 1999). TSAVI could be calculated with the following function, which was proposed by Baret and Guyot (1991).

$$TSAVI = \alpha(r_2 - \alpha r_1 - \beta)/(\alpha r_2 + r_1 - \alpha\beta + X(1+\alpha^2)) \qquad 2.5$$

Where:

r_1 and r_2 are the canopy reflectance observed in two different wavebands

α and β are the soil line parameters

X is a parameter that is used to minimize the effect of soil background

Extensive efforts are being made to derive vegetation indices to minimize the effects of various other parameters affecting the spectral response of vegetation, and to highlight specific vegetation characteristics.

During the last two decades, as important improvements were achieved regarding the processing and the interpretation of remote sensing observations, the concept of coupling crop production models and remote sensing observations was developed, as will be discussed in the following section.

2.2.5 Coupling mechanistic crop models and remote sensing

Nowadays, for monitoring crop growth and yield prediction using remote sensing, more emphasis is being placed on how effectively remotely sensed data can be linked

with crop growth models. Different methods of combining remotely sensed data with crop growth models have appeared in literature and were reviewed by Maas (1988a). First among these is a system based on using remote sensing as a source for extracting quantitative information about vegetation, soil or meteorological conditions. This information can then be supplied as input values for the model parameters. The basic weakness of this method is that the relationship between the remote sensing data and the parameters are valid only for the periods for which they were measured, and they cannot be extended for other periods or growing seasons.

The second approach, takes advantage of the sequential coverage of the remotely sensed information for within-season calibration of the model (Maas, 1988a; Delécolle *et al.*, 1992). In this approach, which was adopted by Maas (1993c), the values of parameters and initial conditions can be adjusted in a manner such that the simulated LAI matches the estimated LAI (calculated with observed information of crop canopy reflectance) using remote sensing. The result of the Maas' (1993c) study on calibrating the model, using within-season LAI observations obtained from either remote sensing or field sampling, showed that remote sensing could be a useful source of observations for within-season calibration of a crop simulation model.

In the final approach, remote sensing data may be used as direct input to a crop growth model, or it may be used for verifying the results obtained with the model (Maas, 1988a; Delécolle *et al.*, 1992; Maas 1993c). Reviewing the studies in this area, in general, Moulin *et al.* (1998) have identified four separate methods of integrating remote sensing data into crop models as follow:

1. "*The direct use of a driving variable estimated from remote sensing information in the model*
2. *The updating of a state variable of the model derived from remote sensing*
3. *The re-initialization of the model, i.e., the adjustment of an initial condition to obtain a simulation in agreement with the remotely-sensed derived observations*
4. *The re-calibration of the model, i.e. the adjustment of model parameters to obtain a simulation in agreement with LAI derived from the observations.*"

Some of these approaches were first evaluated by Maas *et al.* (1989) using a simple model (GRAMI) which simulates gramineous growth and development. For example, a parameterization of the course of the time evolution of LAI is used to replace the description of the mechanisms driving canopy development and senescence. The parameterization is obtained with the initial LAI value at crop emergence and three parameters for the shape of the LAI seasonal curve. In other works, ground radiometric measurements or airborne/satellite data were used with one of the stated methods. The general strategy in coupling models and observations is to derive variables or parameters from radiometric observations which can be used directly in the modeling procedure. Direct estimation of driving variables using remote sensing data is based on the assumption that remote sensing data are available at a frequent interval (from daily to weekly). However, considering cloud coverage and the properties of sensors and platforms, this is rarely the case, so some interpolation approaches must be used to fill the gaps. Thus, the first approach i.e. the direct use of driving variables in the crop model is generally not practical.

In the second approach, at least one of the model variables is updated from remote sensing observations. Considering the fact that remote sensing observations are not available on a regular basis, model parameters are interpolated for the time between the two observations. Estimated value of the parameters from the model is forced to a value obtained from remote sensing data. The concept behind this approach is presented in Figure 2.2. As described earlier an important link for combining remote sensing data and crop growth models is LAI. LAI information has been most frequently extracted from visible and near-infrared data for different crop types and integrated in crop growth models (Holben and Fan, 1980; Asrar *et al.*, 1985b; Maas, 1993c; Wiegand and Richardson, 1990; Baret and Guyot, 1991; Price and Bausch, 1995). In a similar study, Maas (1993b) and Maas *et al.* (1989) used ground radiometric measurements over maize crops. Optical measurements were used to derive LAI and thermal measurements to derive stress index values. By adjusting the initial LAI value (day of emergence) and three shape coefficients of the LAI seasonal curve, minimization was performed. The correct recovery of these parameters was found to be essential to obtain a reliable prediction of crop production. Using LAI derived from satellite data (SPOT/HRV) over

wheat fields (Delécolle and Guérif, 1988) improved the capacities of ARCWHEAT (Weir *et al.* 1984) in crop monitoring. A simple model of LAI time course was used to interpolate and extend the limited derived LAI values to a daily time step. Using these LAI values as input variables to ARCWHEAT led to a decrease of the mean error in yield prediction. However, early saturation of signals is a limiting factor (Michaud *et al.*, 1998) of LAI estimation.

In the third approach, re-parameterization and/or re-initialization of the crop production model, the goal is to minimize the difference between simulated and derived state variables or radiometric signals (Maas, 1988a; Moulin *et al.*, 1998). This approach is further subdivided into two groups, first using remotely-sensed derived parameters in the model, and the second directly using spectral information in the model. Details of the concepts behind these approaches are presented in Figures 2.3 and 2.4. Bouman (1992) used ground radiometric measurements for re-initializing/re-parameterizing of the SUCROS model, the measurements acquired in solar spectrum and active microwave ranges were simulated by linking a reflectance model and a radar reflectivity model with a crop model. Using satellite data (Landsat/MSS), Maas derived a green LAI time course for sorghum (1988b) and winter wheat crops (1991) by re-initializing/re-parameterizing the GRAMI model. For different fertilization and irrigation treatments, the technique was helpful for estimation of yields. In this context, radiometric information can be used directly for re-parameterizing and/or re-initializing a crop model. Coupling a radiative transfer model to the crop production model can produce the temporal behavior of canopy surface reflectance (Bouman, 1992; Major *et al.*, 1992; Moulin *et al.*, 1995; Fischer *et al.*, 1996, 1997).

Clevers (1988, 1989) described a simplified semi-empirical reflectance (CLAIR) model for LAI estimation of a green canopy. The model first calculated the weighted difference vegetation index (WDVI) as a weighted difference between the measured near-infrared and red reflectance. The calculated WDVI was corrected for the influence of soil background. Subsequently, LAIs were estimated from this WDVI using an inverse exponential function.

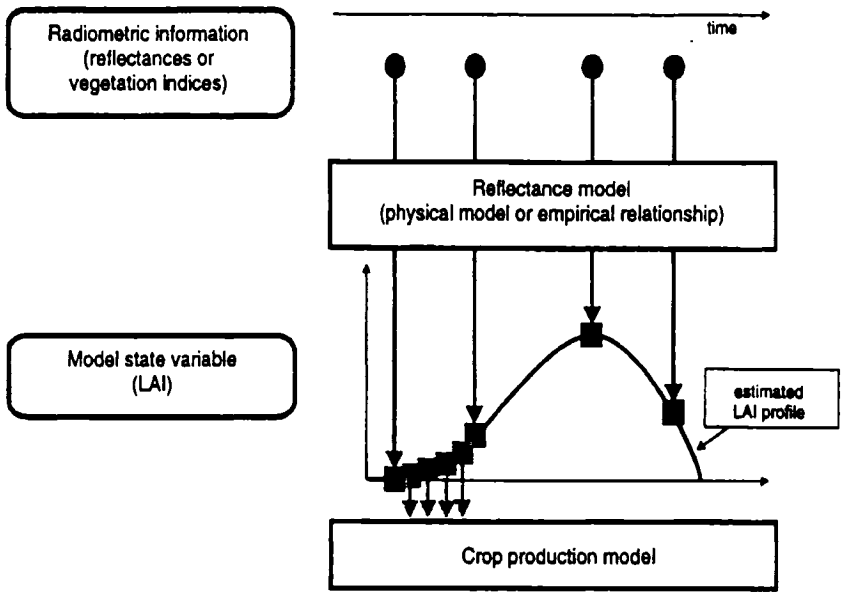


Figure 2.2 Representation of the forcing strategy: the temporal behavior of one state variable of the model is derived from remote sensing observations and used as the input variable within the model (Source: Moulin *et al.* 1998).

The CLOUD model (Attema and Ulaby, 1978) is a simplified semi-empirical radar model that was used to estimate LAI (Clevers and Van Leeuwen, 1996). This model is very useful in describing the radar backscatter of the crops. The model is valid for a short period in the beginning of the growing season because after that the signal will be saturated. For the calibration of the model, it is required to collect high temporal resolution data, which is a limiting factor of the model.

The CLOUD radar model was coupled to the STICS (Brisson *et al.* 1998) crop growth model by Chauki *et al.* (1999) using satellite data (ERS and RDARSAT) over wheat fields. In this approach, the amount of canopy water and soil moisture were simulated within the STICS model and used as input parameters for calculating radar backscatter in the CLOUD model. The combined STICS+CLOUD model allowed the estimation of new values for some key parameters of the STICS model. In another study, Prévot *et al.* (2000a, b) linked the STICS model with a radiative transfer (RT) model in the solar (SAIL) and microwave (water-cloud) domains. The accuracy of STICS predictions was improved.

Bach *et al.* (2001) coupled the PROMET-V crop model to the GeoSAIL canopy reflectance model using LANDSAT images. Their study was based on minimizing the differences between observed reflectance spectra derived from images and the modeled surface reflectance spectra. As a result, the total LAI fraction of brown leaves and surface soil moisture were estimated. The regained and simulated LAI were matched by re-initializing the PROMET-V model, which improved the biomass and yield estimation.

The accuracy of predicted LAI was evaluated by comparing the predicted values with the ground measured LAI in different studies (Asrar *et al.*, 1984; Hatfield *et al.*, 1985; Xu and Jaggard, 1996; Rastogi *et al.*, 2000), and a significant correlation between measured and predicted LAI was achieved.

In a study over sugar beet fields, using ground and airborne radiometric measurements, Clevers and van Leeuwen (1996) calibrated the SUCROS model. They derived LAI from measurements in optical and microwave wavebands. Sowing date, growth rate, light use efficiency and maximum leaf area were then adjusted as were some initial conditions. The ability of optical data to improve yield estimation was pointed out by the authors. In another study, Guérif and Duke (1998) combined the SUCROS (crop)

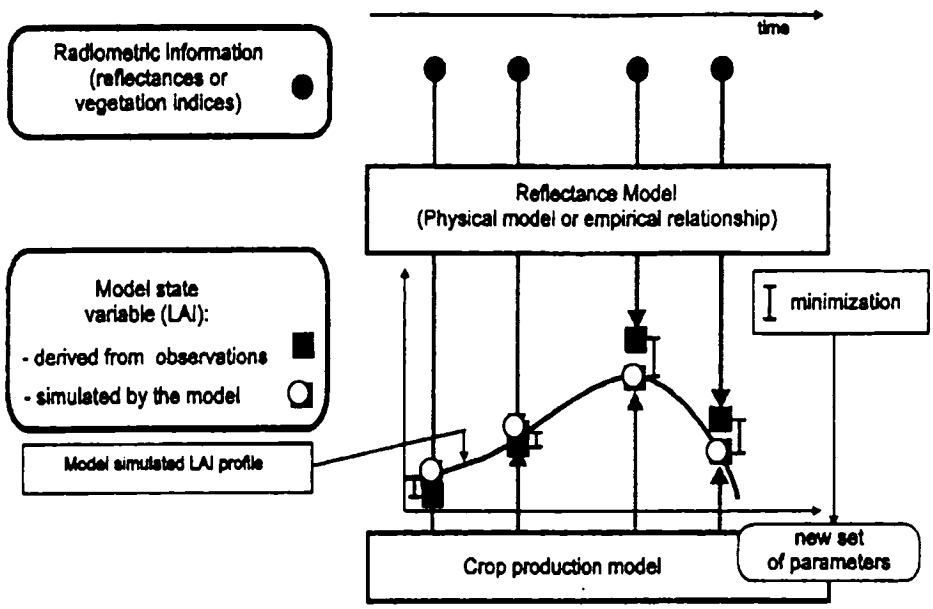


Figure 2.3 Representation of the recalibration strategy: by comparing modeled LAI profile and 'LAI' derived from remote sensing data, some parameters of the crop model (or some initial conditions) are re-tuned (Source: Moulin *et al.* 1998).

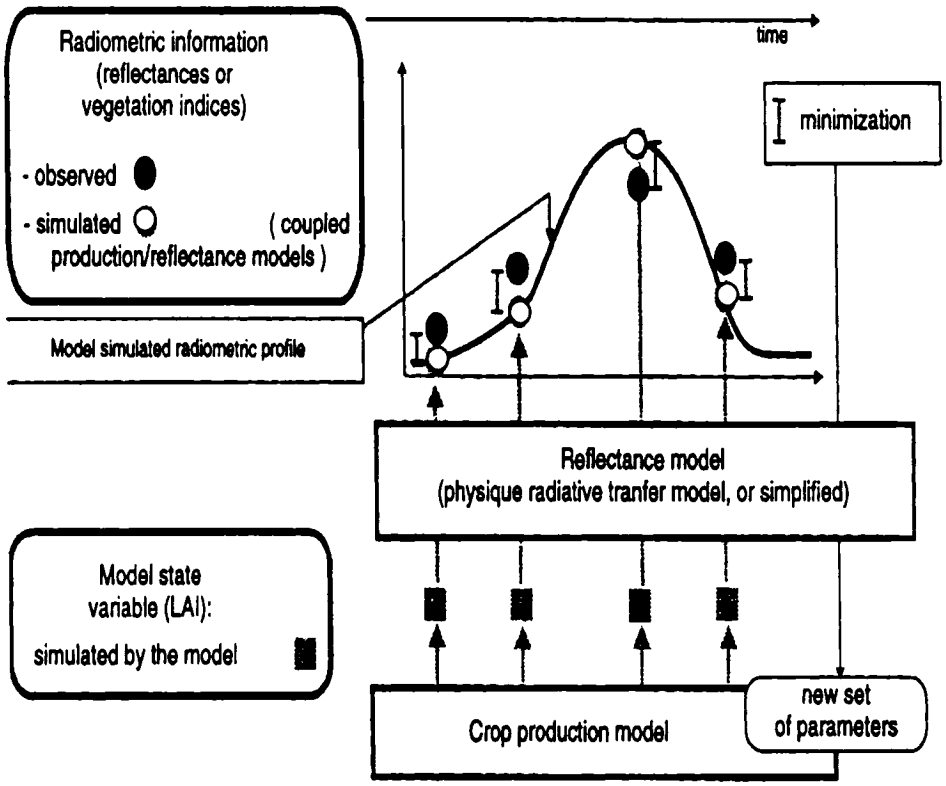


Figure 2.4 Representation of the assimilation strategy: comparing simulated reflectance profiles (from the coupling of a crop production model and a reflectance model) and remote sensing reflectance's allows some parameters of the crop model (or some initial conditions) to be re-tuned (Source: Moulin *et al.* 1998).

and SAIL (reflectance) models. The calibration and evaluation of the resulting model was done by adjusting the emergence and early growth parameters of the SUCROS model under standard and test conditions, respectively. This led to the use of new parameters that allowed accurately prediction of crop yields.

2.2.6 Estimation of yield under weedy conditions

The effect of weeds on crop yield has been widely documented by various researchers (Wilcut *et al.*, 1987; McLennan *et al.*, 1991; Donald and Khan, 1992; Knezevic *et al.*, 1994; Mamolos and Kalburtji, 2001); however, reliable methods for quantitative estimation of weed effects on crop growth are still lacking.

Based on the literature; many researchers have established empirical regression models to estimate the crop yield loss by relating crop yield to one or more weed characteristics (Dew, 1972; Cousens, 1985a,b; Cousens *et al.*, 1987; Kropff and Lotz, 1992; Knezevic *et al.*, 1997; Ngouajio *et al.*, 1999a; Canner *et al.*, 2002, Lemieux *et al.*, 2003). The potential extrapolations of these models results are limited due to the variation among the parameters values obtained between sites and years. To achieve general values for the models' parameters, numerous experiments are needed with different crop and weed combinations worldwide.

Cousens (1985a) developed a simple two-parameter hyperbola model by describing yield loss as a function of weed density. In a similar study considering the crop density, Cousens (1985b) explained a three-parameter model to relate crop yield loss with weed stresses. In comparison to previously published models, he reported that both methods were able to provide better yield loss estimates. Ngouajio *et al.*, (1999b) and Lemieux *et al.*, (2003) used relative leaf area based models to develop a decision support procedure for post-emergence herbicide applications. Accurate estimation of crop losses resulting from weed infestations is the most complicated component of a decision procedure because of the complex relationship between weeds and crop. Using path analysis, Mamolos and Kaburtji (2001) studied the effect of Canada thistle (*Cirsium arvensis*) density on wheat yield, relating the wheat grain to density and nitrogen

concentration of Canada thistle. They found that both Canada thistle density and aboveground biomass had an effect on wheat yield.

These models were generally developed for a single crop and a single weed species in a specified location, which becomes the major limiting factor for the application of such models. Furthermore, collecting input data (e.g. weed density, relative leaf cover) for such models is time consuming and labor intensive. Further, a primarily knowledge of crop and weed conditions are necessary to the implementation of these models.

Recent research in precision agriculture has demonstrated the ability of optical remote sensing in determining the crop growth conditions (Foody *et al.*, 1989; Cibula and Carter, 1992; Pearson *et al.*, 1994; Saha and Jonna, 1994; Clevers, 1997; Moran *et al.*, 1997; Lelong-Camille *et al.*, 1998; Anderson *et al.*, 1999; Singh *et al.*, 2002). It is possible to not only identify areas in the field that are under water or nitrogen stress but also those under weeds (Borregaard *et al.* 2000, Cho *et al.*, 2002; Geol *et al.* 2002, Karimi *et al.*, 2004). Moreover, studies have also shown the value of optical remote sensing in the estimation of various crop biophysical parameters, such as LAI, leaf nitrogen content, plant height, etc. (Gamon *et al.*, 1992; Peñuelas and Filella, 1998; McNairn *et al.*, 2001; Pacheco *et al.*, 2001; Haboudane *et al.*, 2002; Strachan *et al.*, 2002).

2.3 Concluding remarks on current status and future prospects

From the above review it may be concluded that multispectral and hyperspectral imagery can be used to determine crop growth, crop vigor conditions and crop production. However, remote sensing applications in the field of agriculture were limited due to the lack of effective and efficient methodology in processing, delivery and application of data (Moran *et al.*, 1997a, b). The main difficulty in utilizing the full capabilities of remote sensing lies in the conversion of the data into meaningful information, such as crop growth or crop vigor parameters.

High temporal reliability in delivered data is the most important requirement for remote sensing to be more marketable in the field of agriculture (Moulin *et al.*, 1995; Guérif and Duke, 1998; Anderson *et al.*, 1999). Anderson *et al.* (1999) mentioned that in most conditions, the time lag between acquisition of raw (digital) data and delivery of the

final product to the user should not be more than 24 hours, and a greater spatial resolution (about 5 m) is required. Until recently, it was difficult to obtain this resolution from satellite platforms. However, with the launch of new high-resolution satellites and the use of airborne sensors, such resolutions could be obtained. Current research is directed towards evaluating these systems for various purposes. Higher spectral resolution was addressed as a third requirement. Proposed hyperspectral satellite sensors should meet this spectral requirement.

Combining crop growth models with remote sensing techniques shows promise in monitoring crop growth and yield prediction. Most of these studies were based on satellite images, which have a low resolution and less frequency in data acquisition. In this context, extending the available methodology using hyperspectral data was suggested (Bach *et al.*, 2001). It is expected that with the use of hyperspectral data, much more accurate information could be obtained for various crop growth parameters. Moreover, recent advances in crop growth modeling should translate into better prediction results with more accurate crop parameters inputs. Thus, by coupling an advance crop growth model with hyperspectral remote sensing systems, better predictions could be made with respect to crop growth.

PREFACE TO CHAPTER 3

According to the literature on the application of remote sensing, further research is needed to highlight the capability of airborne hyperspectral sensors in weed and nitrogen stress detection. A pilot field experiment was initiated in the summer of 2000 to study the spectral response of corn, under different weedy conditions and nitrogen application rates. Spectral observations were acquired from a Compact Airborne Spectrographic Imager (CASI) in 72-wavebands (from 407 nm to 949 nm). The aim of the study was to investigate the usefulness of hyperspectral observations from airborne platforms in monitoring corn growth under specific weed conditions and different nitrogen application rates. More specifically, the usefulness of hyperspectral airborne observations to discriminate weed and nitrogen stresses was investigated using discriminant analysis and stepwise approaches.

Research papers based on the chapter:

Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, P. Dutilleul, and P. K. Goel. 2004. Classification accuracy of discriminant analysis, artificial neural networks and decision trees for weed and nitrogen stress detection in corn. *Transactions of the ASAE* (in press).

CHAPTER 3

WEED AND NITROGEN STRESS DETECTION IN CORN USING DISCRIMINANT ANALYSIS

3.1 Abstract

Hyperspectral images of experimental plots, cropped with corn and to which the twelve combinations of three nitrogen application rates and four weed management strategies were applied, were obtained with a 72-waveband Compact Airborne Spectrographic Imager. The images were taken at three times during the year 2000 growing season: early growth, tasseling, and full maturity. Nitrogen application rates were 60, 120, and 250 kg N/ha. Weed controls were: none, control of grasses, control of broadleaf weeds, and full weed control. The objective of this study was to evaluate discriminant analysis as a tool for classifying images with respect to the nitrogen and weed management practices applied to the experimental plots, and to compare the classification accuracy of this technique with those obtained by artificial neural network (ANN) and decision tree (DT) algorithms on the same data. Significant wavebands were selected, among the 72 available, using the stepwise option and DISCRIM procedure (SAS software). Classification accuracy was determined for the full set of selected wavebands and for subsets thereof, for three problems: distinguishing between the 12 combinations of factor levels, differentiating between nitrogen levels only, and separating weed controls only. Misclassification rates of images, taken at the initial growth stage, were substantially lower for each of these tasks (25, 17 and 13%, respectively) when discriminant analysis was used. The ANN approach was best for images taken at the tasseling and full maturity stages. However, from the precision-farming point of view, it is easier to apply site-specific remedies to weed and nitrogen stresses early in the season than when the corn crop has reached the tasseling stage so the results obtained with the discriminant analysis are noteworthy.

3.2 Introduction

One aspect of overcoming the problem of minimizing the impact of agriculture on environmental quality, in the face of the higher demand for food by the world's human population, is the development of more efficient approaches to crop production. Precision-farming, which is based on site-specific application of agricultural inputs (e.g., fertilizer, herbicide, pesticide) is very promising in this respect since it has been shown to lead to a reduction of the overall quantities applied (Tomer *et al.*, 1997; Christensen *et al.*, 1998).

The success of precision farming depends on accurate and fast collection and analysis of field data and development of variable rate application techniques. The new generation of multispectral and hyperspectral sensors can provide digital images at much finer spatial and spectral resolutions at a reasonable cost (Lamb, 1998). The large amount of information contained in these images could ease the automation processes; however better methods of data transfer, storage, and analysis must be developed (Thenkabail *et al.*, 2003). Moreover, automated methods of data analysis, requiring a minimum of manual activities for testing, are essential if optimal use of the data is to be achieved (Soh, 1999). The fundamental problem of analysis of hyperspectral images is to determine how variations in the wavebands are related to differences in the features of images to be classified, according to one or more target parameters.

Various artificial intelligence and statistical approaches have been used for this purpose. Yang *et al.* (1999) used Artificial Neural Networks (ANNs) to recognize weeds in a corn crop based on color photographs. Best success was obtained in differentiating between corn and any weed but results could not be considered of practical significance for finer differentiation of species of weed. In hyperspectral remote sensing studies, where observations are collected in a large number of wavebands, it is computationally not efficient to processes all the data. The main purpose behind these methods is to determine the possibility of using fewer data (subset or transformed data) without losing major information (Rao, 1964; Kenkel *et al.*, 2002). Approaches based on ANNs have successfully been used in agricultural remote sensing (Deck *et al.*, 1995; Ghazanfari *et al.*, 1996; Nakano, 1997; Wilkinson, 1997; Yang *et al.*, 1999; Goel *et al.*, 2003a). More

recently, analyses based on decision trees technology have also been used (Goel *et al.*, 2003a; Friedl *et al.*, 1999; Soh, 1999; Friedl and Brodley, 1997; Hansen *et al.*, 1996).

Furthermore, Cho *et al.* (2002) used both ANN and discriminant analysis to distinguish between radish and weeds. Goel *et al.* (2003c) used decision trees (DTs), and ANNs to differentiate between plots cropped with corn, according to the nitrogen application rate and weed control they had been subjected to, based on hyperspectral images in 72 wavebands from the visible to the near-infrared, at three stages of growth of corn. The misclassification rates obtained by Goel *et al.* (2003c) were dependent on growth stage, the complexity of the classification problem, and the number of pixels used from each image. Reasonable accuracy was obtained when the problem involved distinguishing between nitrogen levels or between weed controls at a given growth stage and the authors noted that ANNs performed somewhat better than DTs. Goel *et al.* (2003a) also compared these and five other classifiers on this data set, but found that there was no consistently better classifier for all problems presented. Although ANN models can do an acceptable job, they cannot identify important wavebands in a data set (Deck *et al.*, 1995; Ghazanfari *et al.*, 1996; Nakano, 1997; Wilkinson, 1997; Yang *et al.*, 1999). On the other hand, both decision tree and multivariate data analysis methods can discover important hidden relationships in data. Like the stepwise method in multivariate data analysis, a decision tree approach can also identify important wavebands. However, while constructing rules in a decision tree, once an input data is selected to become part of a tree, it cannot be removed. On the other hand, the significance of the selected variables is tested continuously in a stepwise method, and a previously selected input may be removed, if warranted by later analyses.

Kenkel *et al.* (2002) reported that multivariate statistical approaches were being used in agricultural domains only recently. In their paper, they illustrated the usefulness of different multivariate methods, for both descriptive and predictive modeling, in weed research. Meyer *et al.* (1998) applied discriminant analysis methods to recognize different weed types from soil based on textural image processing. Using a color machine-vision system, Vrindts *et al.* (2002) used ANN and discriminant functions to detect weeds, based on three-band ratios of canopy reflectance. They reported a better accuracy for discriminant analysis over ANN. Misclassification for differentiating corn

from weeds was less than 1% in the laboratory but it was 85% for corn and 3% for weeds when tested on field data. The application of discriminant analysis in agricultural remote sensing has been reported in many recent studies (Slaughter *et al.*, 2003; Burks *et al.*, 2002; Cho *et al.*, 2002; Terawaki *et al.*, 2002).

The overall objective of the study was to analyze the data of Goel *et al.* (2003c) using discriminant analysis and a stepwise approach. A more specific goal was to identify the most significant wavebands that can be used in classification of different treatments. The results of discriminate analysis will be compared to results obtained with ANN and DT techniques.

3.3 Materials and Methods

3.3.1 Experimental detail

A field experimentation was laid out in the growing season of the year 2000 at the Lods Agronomy Research Center of Macdonald Campus, McGill University, Ste-Anne-de-Bellevue, Québec, Canada (45°25′45"N lat., 73°56′00"W lon.). Figure 3.1 shows a schematic diagram of the experimental layout. The soils at the study site are classified as Bearbrook clay and Ste. Rosalie clay. Both soils belong to the Dark Gray Gleysolic group. Corn was planted and subjected to four weed controls and three nitrogen application rates levels in a four-replicate split plot design with weed controls assigned as the main treatment units. The plot size was 20 m × 20 m with a 3 m buffer strip between plots. The weed treatments consisted of no weed control (W1), control of grass species (W2), control of broadleaf species (W3), and full weed control (W4). The three nitrogen application rates were: low nitrogen (60 kg N ha^{-1}, N_{60}), normal nitrogen (120 kg N ha^{-1}, N_{120}), and high nitrogen (250 kg N ha^{-1}, N_{250}).

3.3.2 Spectral data collection

Hyperspectral aerial images were obtained from a Compact Airborne Spectrographic Imager (CASI). The CASI sensor collects information in 72 narrow

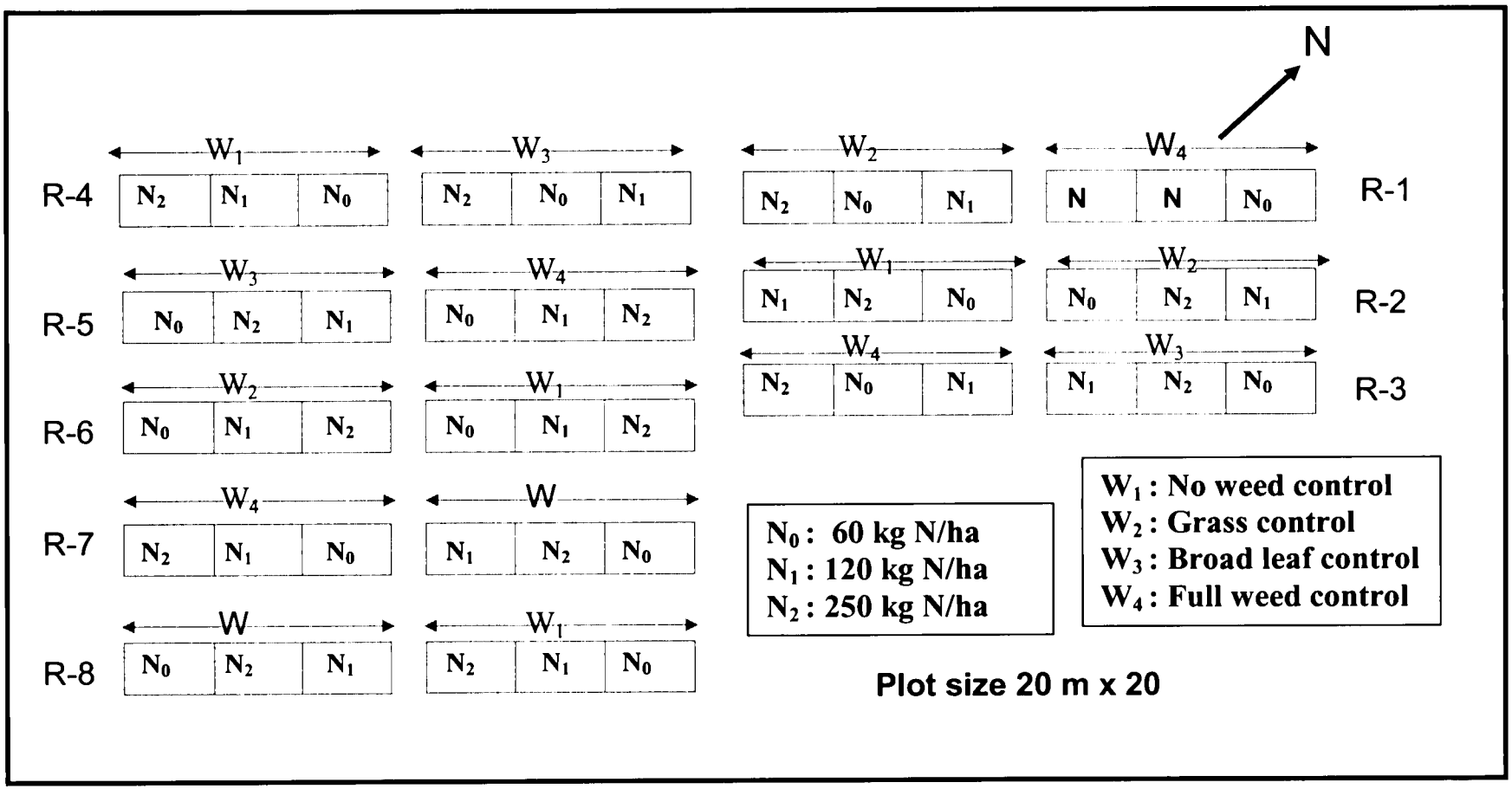


Figure 3.1 Experimental layout, 2000.

bands from 407 to 949 nm (visible to near-infrared). The bandwidths varied from 4.27 to 4.41 nm, and the distance between two band centers ranged from 7.4 to 7.74 nm. More details on the wavebands properties can be found in Goel *et al.* (2003a). Spatial resolution was 2 m at the flight altitude. Three images, representing the major growth stages of corn were acquired during the growing season: early growth (June 30, 2000), tasseling (August 05, 2000), and fully mature (August 25, 2000).

Pre-processing of the images was completed by the data supplier. Pre-processing included application of radiometric and atmospheric corrections, using calibration coefficients established by CRESTech (Center for Research in Earth and Space Technology). The CASI hyperspectral data were processed to at-sensor radiance. Transformation of at-sensor radiance to ground-reflectance data was accomplished by using the CAM5S atmospheric correction model (O'Neill *et al.*, 1997). Further, to improve reflectance image data cubes, spectrally-flat uniform areas in each image (asphalt, bare soil and concrete surfaces) were used to apply flat field adjustments. Using GPS data onboard the aircraft, geometric corrections and geo-referencing of the images were performed for the aircraft movements (yaw, pitch, and roll). The corrected images were geocoded to UTM geographic coordinates. White boards at the corners of the field assisted in the geometric and error assessment. These processes resulted in an estimated RMSE (root mean square error) of about 0.5 of a pixel.

Reflectance values from the corrected images were extracted using ENVI software (ENVI 3.1, Research System, Inc., Boulder, Colorado, USA). A total of 20 random points per plot were selected arbitrarily to get representative reflectance values for the plot for each flight. This resulted in a total of 720 values per flight. The average spectral response of corn under different nitrogen levels and weed control treatments is illustrated in Figures 3.2a to 3.2c for the three flight data sets. The figures clearly demonstrate the difference in reflectance values for various treatments, with comparatively greater differences among treatments in the visible region in early growth stages where the field had less crop coverage than at the other stages. They also show that certain wavebands in the visible region and many more in the near infrared region can be used to classify the treatments.

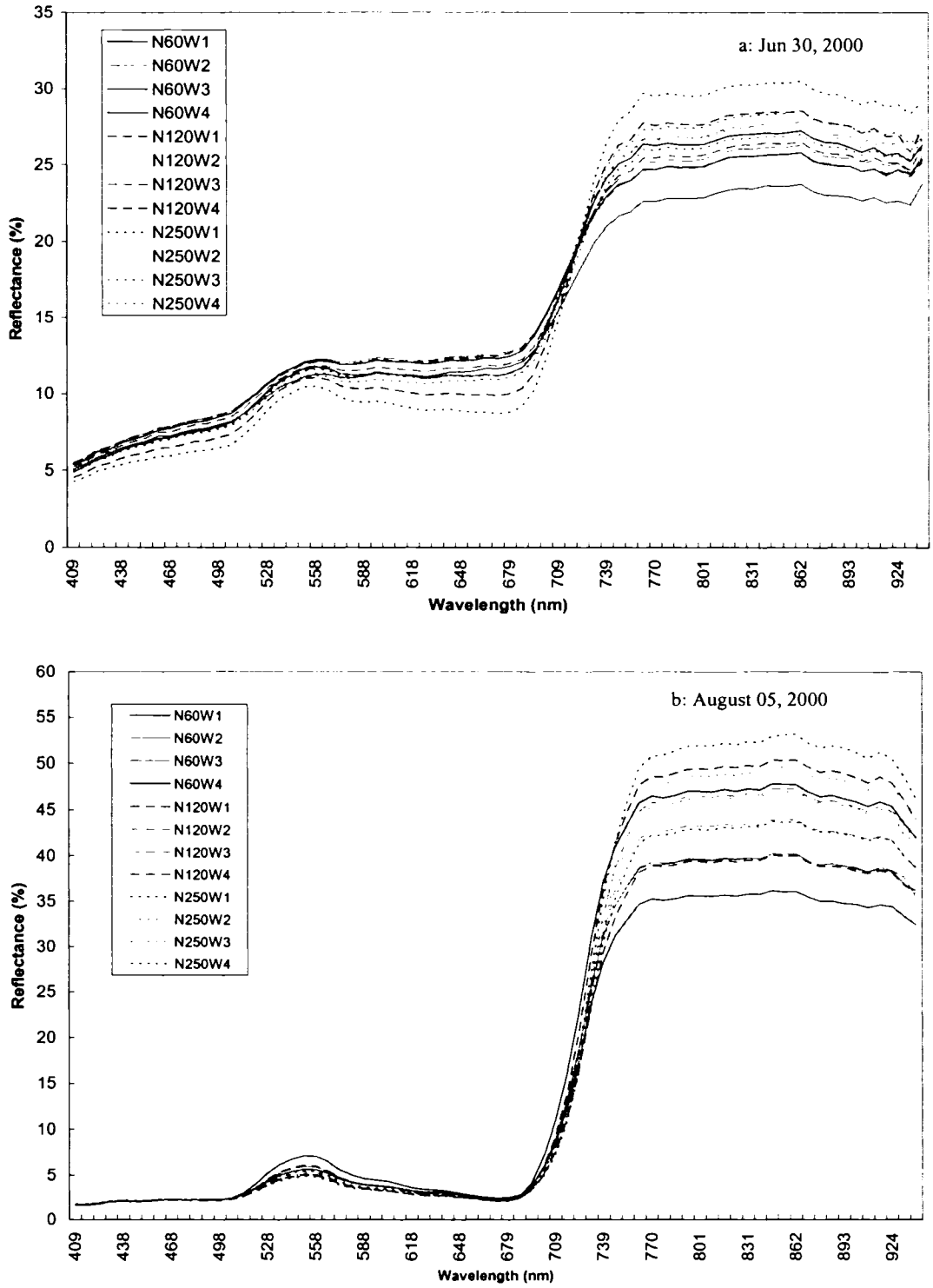


Figure 3.2 Measured spectral response curves of corn at three different growth stages under different nitrogen application rates and weed control conditions. (N60, N120, and N250 are nitrogen treatments of 60, 120, and 250 kg N ha^{-1} and W1, W3, and W4 are weed treatments with no weed control, broadleaf control, and full weed control, respectively).

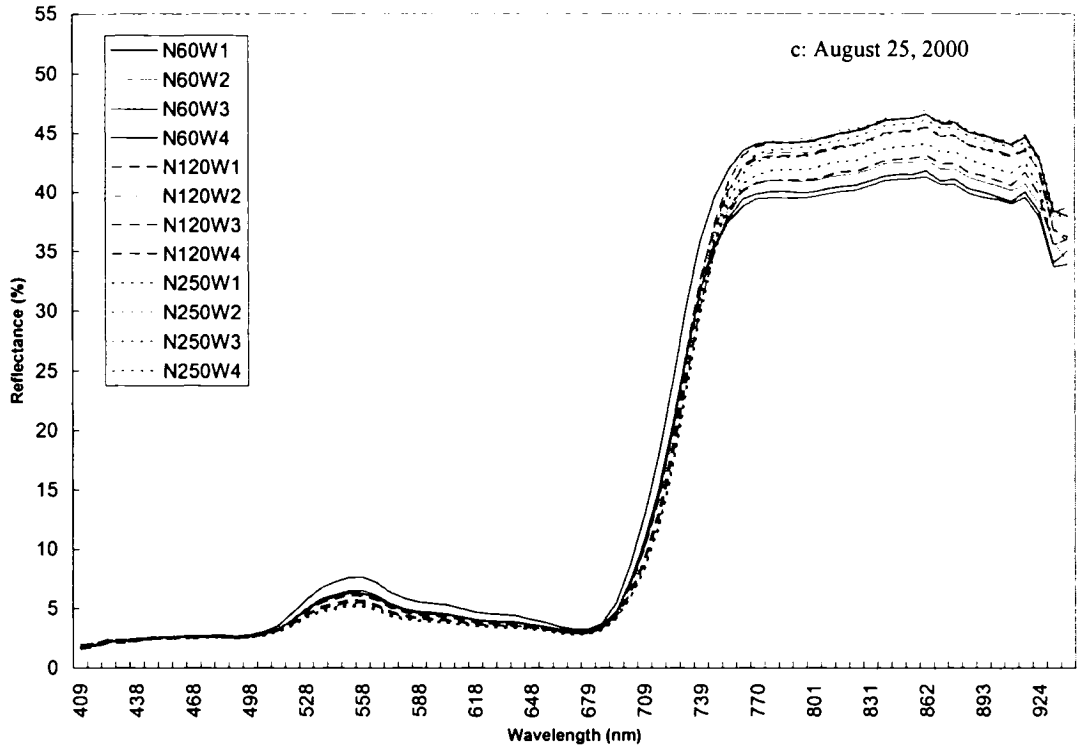


Figure 3.2 (cont.) Measured spectral response curves of corn at three different growth stages under different nitrogen application rates and weed control conditions. (N60, N120, and N250 are nitrogen treatments of 60, 120, and 250 kg N ha^{-1} and W1, W3, and W4 are weed treatments with no weed control, broadleaf control, and full weed control, respectively).

3.3.3 Weed survey

Information on weeds in each plot was collected on July 14 and August 15, 2000 using three and four randomly chosen 50 x 50 cm quadrates on the two dates, respectively. Observations included weed types, density, and percent ground coverage. The weed cover was determined visually in each quadrate. The most dominant grassy weeds were Yellow Nutsedge (*Cyperus esculenthus)*, Barnyard grass (*Echinochloa crusgalli)*, and Crab grass (*Digitaria ischaemum*), whereas the dominant broad-leaves were Canada thistle (*Cirsium arvensis*), Sow thistle (*Sonchus oleraceus)*, Redroot pigweed (*Amaranthus retroflexus*), and lamb's quarter (*Chenopodium album*). More details are given in Goel *et al.* (2003c).

3.3.4 Data analysis

Analyses of the spectral reflectance were performed for the three images individually. As stated earlier, the objectives of the research were to find the most significant wavebands for categorizing data into different classes. The stepwise option of SAS procedure STEPDISC was used to select the wavebands most likely to be associated with differences in the pixels associated with the different plots. The criterion used for selection was the significance of the F value calculated from an analysis of covariance. Although it is assumed that for discrimination of the treatments, the most suitable wavebands will be selected using stepwise procedure, there are no guarantees that the selected variables are the best set of variables, particularly when there is high collinearity (Murray, 1977; Vrindts *et al.*, 2002; Johnson, 1998), which is often the case with very-narrow waveband hyperspectral data sets. "Contrary to what might be expected, a subset of well-chosen variables will often do a better job of discriminating between groups than you can do using all possible variables" (Johnson, 1998).

Next, the discrimination capability of the selected wavebands was evaluated using the DISCRIM procedure of SAS (Version 8.20, North Carolina, USA). Based on the quantitative variables, the DISCRIM procedure calculates different discriminant functions for categorizing observations into groups. In this procedure, the discriminant

functions can be developed using a classification variable, describing groups of observations and quantitative variables. Discriminant criteria include the individual within-group covariance matrices, the pooled covariance matrix, and the prior probabilities of the groups. The suitability of the discriminant functions for a given classification problem can be compared using a cross-validation method, available in the DISCRIM procedure, which involves calculation of misclassification matrices.

3.4 Results and discussion

Verification of the weed populations in different treatments indicated that there were no broadleaf weeds in the second treatment (W2, grass control). As a result the plots with weed-nitrogen combinations involving W2 were not included in the analysis. Inspection of the data from the 72nd waveband showed that the signal was too noisy to be used. Thus, only 9 weed-nitrogen groups were to be differentiated on the basis of data from 71 wavebands.

3.4.1 Selection of Wavebands

Johnson (1998) recommended that the significance level for the inclusion of the variables in the stepwise approach be set somewhere between 0.25 and 0.5 for entering a variable, and 0.15 for removing a variable. In this study, the significance levels for the stepwise STEPDISC procedure were set at 0.10 and 0.15 for inclusion and removal, respectively. Although it was noted in the literature that better results are often obtained by using a well-chosen subset of the available variables, there is not as yet, a method, other than an iterative one, to determine the best subset. The STEPDISC approach was applied for three classification problems: a) the 9 combinations of weed and nitrogen treatments, b) the 3 weed treatments alone, and c) the three nitrogen application rates alone.

For the first classification problem (a), a total of 34, 42, and 42 wavebands were respectively found to be significant in explaining the variability for the three growth stages. For the second problem (b), a total of 26, 32, and 31 wavebands were significant, while for the classification according to nitrogen, a total of 28, 19 and 19 wavebands were considered to be explanatory. The selection of a greater number of

significant wavebands for the first classification problem, i.e. both weeds and nitrogen together, underscores the complexity of classification in such problems.

3.4.2 Discriminant analysis

Different DISCRIM models were trained and used to test the performance of the selected wavebands and/or subsets from selected wavebands for their ability to discriminate between various groups. DISCRIM determines the classification accuracy using a one data-out approach for cross-validation. The risk estimate, for resubstitution and cross-validation, represents the percentage of the wrongly classified in the training and validation data sets, respectively, although there is no guarantee that the selected wavebands are the best possible subset. In this study, different DISCRIM models were applied using different wavebands chosen on the basis of their order of entry in the stepwise approach (Table 3.1). The wavelengths of the selected subset for all three classification problems are summarized in Table 3.1. For the first problem, although, the selected wavebands for the first flight are generally different from those for the other two flights, over 75% of the selected wavebands for the second and third flights were the same. For the second classification problem, while 40% of the selected wavebands for the first flight were also selected for the other two flights data sets, the wavebands selected for the second and third flights were identical. For the third problem, the selected wavebands for the first flight are generally different from the other two flights. The wavebands for the third flight are included in the wavebands selected for the second flight. The wavebands selected by the discriminant analysis are compared to those identified by the DT approach in Table 3.1. Generally speaking, a lower number of wavebands were selected by the DT method for all three classification problems.

A summary of calibration and cross-validation misclassifications for all three data sets and different subsets of wavebands is presented in Table 3.2 In the case of discriminating combined weed and nitrogen treatments, the best results were obtained using subsets of 19, 20, and 19 wavebands with an accuracy of 75, 69, and 71 % for early growth, tasseling and full growth stages, respectively. Generally speaking, using all the selected wavebands resulted in far greater accuracy in the calibration data sets than in the

Table 3.1 Wavelengths selected by different methods in developing classification model.

Classification method	Treatments	First Flight	Second Flight	Third flight
		Selected wavebands	Selected wavebands	Selected wavebands
Stepwise procedure*	Combined effect of weed and Nitrogen	893, 732, 641, 717, 535, 588, 679, 446, 701, 476, 801, 747, 461, 816, 610, 565, 724, 663, 453	717, 885, 732, 663, 755, 778, 558, 739, 901, 939, 520, 461, 490, 505, 762, 709, 694, 610, 453, 446	717, 641, 885, 739, 770, 558, 755, 732, 901, 762, 939, 461, 490, 679, 505, 520, 847, 453, 446
	Weeds only	739, 663, 409, 505, 932, 701, 535, 893, 717, 610, 528, 453, 438, 939, 755	901, 490, 543, 679, 755, 461, 801, 762, 505, 453, 550, 732, 739, 409, 770, 785, 939, 446, 513, 431, 916, 908, 528, 595, 633, 694	901, 490, 543, 679, 755, 461, 801, 762, 505, 453, 550, 732, 739, 409, 939, 446, 770, 785, 513, 431, 916, 908, 528, 595, 633, 694
	Nitrogen only	588, 724, 490, 847, 701, 679, 641, 461, 565, 739, 755, 831, 610, 424, 686, 656, 633, 839	717, 893, 732, 770, 739, 939, 520, 535, 694, 679, 641, 558, 528, 816	717, 732, 770, 739, 939, 816, 520, 558, 595, 679, 641, 528
Decision tree **	Combined effect of weed and Nitrogen	409, 446, 588, 641, 717, 878, 932, 939	483, 490, 513, 535, 550, 558, 724, 755, 770	409, 416, 431, 461, 543, 565, 701, 739
	Weeds only	409, 513, 701, 724, 732, 939	424, 446, 732, 739	416, 453, 490, 762
	Nitrogen only	409, 476, 550, 641, 724, 732, 939	490, 535, 747, 770	409, 424, 558, 701
ANN **	Combined effect of weed and Nitrogen	All 71 wavebands	All 71 wavebands	All 71 wavebands
	Weeds only			
	Nitrogen only			

* Wavelengths are given in the order they entered into the model

** These data are taken from Geol *et al.* (2003c)

cross-validations. Using subsets, fewer wavebands decreased classification accuracy of the calibration data sets. For the cross-validations, accuracy usually increased up to a certain numbers of wavebands and decreased thereafter. Moreover, reflectances from the early growth stage for different subsets resulted in better accuracy (69 to 75 %). With respect to weed control, this could help identify the weediest patches as early as possible. However, more accurate results were obtained by considering one factor at a time (weed or nitrogen). For cross-validation, the DISCRIM procedure could differentiate between weed treatments with 87 % accuracy in early growth and 78 % accuracy in tasseling and full growth stages in the optimal conditions. For nitrogen, classification accuracies of 83, 81, and 82 % were obtained for the three stages, respectively. The best misclassification matrices for cross-validation are presented in Table 3.3 (weed-nitrogen combinations) and Table 3.4 (weeds only and nitrogen only). The values in the tables provide both the number of correctly classified and misclassified cases for each flight. Close inspection of the misclassification matrices shows that in all cases, most misclassified cases were classified into the next nearest category. In other words, it was very rare for N0 (low nitrogen application rate) to be misclassified as N2 (high nitrogen application rate).

The best classification results obtained with the discriminant analysis are given in Table 3.5 along with the best results, obtained for the same data set, with the DT and ANN approaches by Goel *et al.* (2003c). The risk estimate values were calculated by dividing the number of incorrectly classified cases by the total number of cases used in the classification. Thus, a lower risk estimate would indicate better classification accuracy. For the early growth stage, higher classification accuracies were obtained with the discriminant analysis for all three classification problems (Table 3.5). For the concurrent weed and nitrogen classification problem, a classification accuracy of 75 % was achieved using the discriminant analysis, while the DT and ANN methods were 60% and 58% accurate, respectively. The combined weed and nitrogen problem also happens to be the most complex classification problem in our study, and it appears that the discriminant analysis has performed well. When only one factor was considered (weed or nitrogen), the accuracy of the discriminant analysis was 87% for weeds and 83% for nitrogen. The DT and ANN accuracies for these cases were 76% and 81%, respectively, for weeds and 68% and 69%, respectively, for nitrogen. At the tasseling stage, the DT

Table 3.2 Summary of misclassification matrices of DISCRIM procedure (%)

a) Weed and nitrogen treatments

First flight			Second flight			Third flight		
# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation
34	0.1	30.8	42	0.0	40.6	42	0.0	39.2
30	1.1	31.8	38	0.3	39.2	38	0.3	38.3
25	3.8	29.2	34	0.7	36.5	35	0.8	37.4
22	5.6	28.2	30	2.2	35.0	30	2.1	34.6
19	6.7	25.1	25	4.2	33.2	25	4.2	34.0
17	7.9	25.8	20	10.1	31.2	20	9.0	30.4
15	8.9	27.4	18	11.5	32.5	19	9.6	28.9
14	9.7	26.7	16	13.1	32.5	15	15.2	30.3

b) Weed treatments

First flight			Second flight			Third flight		
# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation
26	5.7	13.6	32	6.4	21.9	31	8.1	22.6
23	5.7	13.2	31	6.5	21.5	28	9.2	23.1
20	7.9	13.6	29	8.3	22.4	26	10.7	22.1
18	7.9	13.6	26	10.4	21.7	22	12.8	23.2
16	8.9	14.1	24	12.4	22.5	18	15.8	25.1
15	9.7	13.1	20	14.3	22.4	15	19.9	26.9
12	11.2	14.4	15	19.0	27.1	10	25.1	28.5
8	15.6	16.8	10	25.4	28.5	5	34.4	35.6
4	22.1	23.5	5	34.3	35.8	3	40.4	41.2

c) Nitrogen treatments

First flight			Second flight			Third flight		
# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation
28	9.2	20.3	19	14.3	19.7	19	14.3	19.9
25	9.4	20.0	16	15.3	20.3	17	16.2	20.1
20	11.0	19.6	14	15.3	19.0	15	17.1	20.4
18	13.2	17.2	12	17.1	19.3	12	16.5	18.5
15	14.2	18.3	10	19.0	20.1	10	18.2	19.7
10	17.1	20.1	7	17.8	19.4	7	18.3	19.0
7	20.7	22.2	5	20.8	21.5	5	20.8	21.5
4	23.8	24.2	3	22.4	22.8	2	22.8	22.9

Table 3.3 Classification matrices for the cross-validation data for the nine weed-nitrogen combinations at the three growth stages.

a) Early growth stage

		Predicted									
		N_{60} W1	N_{60} W3	N_{60} W4	N_{120} W1	N_{120} W3	N_{120} W4	N_{250} W1	N_{250} W3	N_{250} W4	Total
Actual	N_{60}W1	65	1	1	12	0	0	1	0	0	80
	N_{60}W3	4	65	6	1	4	0	0	0	0	80
	N_{60}W4	0	8	63	0	1	7	0	0	1	80
	N_{120}W1	11	1	0	51	5	0	12	0	0	80
	N_{120}W3	0	1	0	5	60	6	3	5	0	80
	N_{120}W4	0	2	14	0	9	49	0	0	6	80
	N_{250}W1	0	0	0	13	2	0	62	3	0	80
	N_{250}W3	0	0	0	1	3	0	8	64	4	80
	N_{250}W4	0	0	1	0	4	11	0	4	60	80
	Total	80	78	85	83	88	73	86	76	71	720

b) Tasseling stage

		Predicted									
		N_{60} W1	N_{60} W3	N_{60} W4	N_{120} W1	N_{120} W3	N_{120} W4	N_{250} W1	N_{250} W3	N_{250} W4	Total
Actual	N_{60}W1	62	8	0	9	1	0	0	0	0	80
	N_{60}W3	7	60	3	4	3	3	0	0	0	80
	N_{60}W4	1	8	55	8	0	7	1	0	0	80
	N_{120}W1	11	9	1	35	12	0	8	4	0	80
	N_{120}W3	0	1	1	10	56	6	3	2	1	80
	N_{120}W4	1	2	8	3	1	54	4	3	4	80
	N_{250}W1	0	0	0	8	7	2	54	7	2	80
	N_{250}W3	0	0	0	1	2	1	10	58	8	80
	N_{250}W4	0	0	0	0	1	6	6	6	61	80
	Total	82	88	68	78	83	79	86	80	76	720

c) Full growth stage

		Predicted									
		N_{60} W1	N_{60} W3	N_{60} W4	N_{120} W1	N_{120} W3	N_{120} W4	N_{250} W1	N_{250} W3	N_{250} W4	Total
Actual	N_{60}W1	67	5	0	8	0	0	0	0	0	80
	N_{60}W3	6	60	7	3	2	1	1	0	0	80
	N_{60}W4	3	4	61	4	0	8	0	0	0	80
	N_{120}W1	9	6	1	47	6	0	7	4	0	80
	N_{120}W3	0	1	0	9	53	11	4	1	1	80
	N_{120}W4	0	1	10	2	2	55	5	4	1	80
	N_{250}W1	0	0	0	10	4	2	55	6	3	80
	N_{250}W3	0	0	0	2	4	2	16	50	6	80
	N_{250}W4	0	0	0	0	0	6	5	5	64	80
	Total	85	77	79	85	71	85	93	70	75	720

N_{60}, Low N; N_{120}, Normal N; N_{250}, High N; W1, No weed control; W3, Broadleaf control; W4, Full weed control.

Table 3.4 Classification matrices for the cross-validation data for weed controls and for nitrogen application rates, each at three growth stages.

a) Early growth stage

Weed treatments		Predicted			
		W1	W3	W4	Total
Actual	W1	220	19	1	240
	W3	27	192	21	240
	W4	0	27	213	240
	Total	247	238	235	720

Nitrogen treatments	Predicted			
	N_{60}	N_{120}	N_{250}	Total
N_{60}	203	37	0	240
N_{120}	29	169	42	240
N_{250}	3	35	202	240
Total	235	241	244	720

b) Tasseling stage

Weed treatments		Predicted			
		W1	W3	W4	Total
Actual	W1	196	33	11	240
	W3	45	163	32	240
	W4	17	20	203	240
	Total	258	216	246	720

Nitrogen treatments	Predicted			
	N_{60}	N_{120}	N_{250}	Total
N_{60}	196	43	1	240
N_{120}	26	172	42	240
N_{250}	1	24	215	240
Total	223	239	258	720

c) Full growth stage

Weed treatments		Predicted			
		W1	W3	W4	Total
Actual	W1	188	42	10	240
	W3	38	167	35	240
	W4	10	24	206	240
	Total	236	233	251	720

Nitrogen treatments	Predicted			
	N_{60}	N_{120}	N_{250}	Total
N_{60}	196	43	1	240
N_{120}	23	173	44	240
N_{250}	1	21	218	240
Total	220	237	263	720

N_{60}, Low N; N_{120}, Normal N; N_{250}, High N; W1, No weed control; W3, Broadleaf control; W4, Full weed control.

Table 3.5 Comparison of the risk estimate result obtained from different methods of classification.

	Method used	Early growth stage	Tasseling stage	Full growth stage
Combined effect of weed and nitrogen	DISCRIM	25.1	31.2	28.9
	Decision Tree	40.4	28.5	40.8
	ANN	41.7	29.9	36.8
Effect of weed alone	DISCRIM	13.1	21.5	22.1
	Decision Tree	23.5	16.2	30.8
	ANN	18.8	11.8	14.6
Effect of nitrogen alone	DISCRIM	17.2	19.0	18.5
	Decision Tree	32.1	18.9	22.8
	ANN	30.6	11.8	11.8

approach gave the highest classification accuracy for the combined weed and nitrogen case (71%). For the other two classification problems, the ANN method gave the most accurate results (88% for both weed and nitrogen classification problems). At the full growth stage, once again the discriminant analysis gave the best results for the combined case (79%). Also, ANN results were the best for the other two cases (weeds alone– 85% and nitrogen alone– 88%).

It is noteworthy that, in all three classification problems, the discriminant analysis resulted in the best classification results for the early growth stage, whereas the ANNs performed better at tasseling and full maturity. The higher classification accuracy obtained by the discriminant analysis in the early growth stage emphasizes the usefulness of this approach over the DT and ANN methods. It is difficult to say why one method is performed better than the other as they are based on different principles. However, it is clear that different classification methods should be tried for a given situation to determine the most accurate method. As stated earlier, compared to the other stages, in the early growth stage, more reflectance differences were evident between treatments in the visible portion, where the most significant wavebands are selected by the stepwise approach. This may explain why discriminant analysis provided the best results.

In precision agriculture, critical decisions need to be made at the early growth stage of the crop to determine site-specific herbicide and fertilizer application rates. Our study shows that the discriminant analysis may be a better classification method at this stage.

3.5 Conclusions

This study compared three different methods of analyzing hyperspectral data, namely ANNs, DTs, and discriminant analysis for identifying weed and nitrogen stresses in corn. The discriminant analysis was found to provide the best classification accuracy at the early growth stage (more than 75%), whereas, better accuracy was obtained with ANN models, at the tasseling and full maturity stages. Misclassification rates for two target variables at the same time were too high (25% or greater) to be considered of practical use, regardless of the method of classification used. However, when one target

at a time was considered (nitrogen application rate, or weed treatment) misclassification rates were below 20%, and as low as 11% for cross-validation data.

This study has demonstrated that correctly chosen narrow waveband hyperspectral aerial observations can be used as an important source of information in identifying weed and nitrogen stresses. The choice of image classification method is important, but dependent on the stage of maturity of the crop. Discriminant functions based on a restricted set of wavebands, as determined by a stepwise selection technique, can be used for this purpose.

PREFACE TO CHAPTER 4

The results of Chapter 3 indicated that airborne imagery could be used for weed and nitrogen stress determination by the use of discriminant analysis method, especially during the early growth stages of the crop. Thereafter, the study was focused on using very narrow-waveband, ground-based spectral observations for discriminating nitrogen and water stress. In the summer of 2002, an experiment was laid out to study the spectral response of corn, under different irrigation regimes and nitrogen application rates. Spectral observations were acquired from a handheld spectroradiometer in 2150 narrow wavebands from 350-2500 nm. The overall goal of the study was to examine the applicability of ground-based narrow-waveband hyperspectral observations in monitoring corn growth under different water availability conditions and different nitrogen application rates. Discriminant analysis and stepwise approaches were used for this investigation.

Research papers based on the chapter:

Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, P. Dutilleul, and P. K. Goel. 2004. Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Transaction of the ASAE* (in press).

CHAPTER 4

WATER AND NITROGEN STRESS ASSESSMENT IN CORN BY APPLYING DISCRIMINANT ANALYSIS TO HYPERSPECTRAL DATA

4.1 Abstract

The development and implementation of both economically and environmentally sustainable precision crop management systems can be greatly enhanced through the use of remote sensing. In this study, the potential of narrow-waveband hyperspectral observations in the discrimination of nitrogen and water stresses in corn was investigated. A field experiment was conducted in the summer of 2002 at the Macdonald Research Farm, McGill University, Ste-Anne-de-Bellevue, Quebec, Canada. Corn was grown in forty, 9.0 m x 10.0 m, test plots laid out in a split plot design with irrigation (non-irrigated, irrigated) as the main treatment and nitrogen fertilizer application rate (50, 100, 150, 200, and 250 kg ha^{-1}) as the sub-treatment. Hyperspectral measurements in 2151 wavebands (350 to 2500 nm) were made with a field spectroradiometer during the entire growing season. Using a stepwise procedure, the most effective wavebands capable of discriminating treatment effects were selected. By applying a discrimination procedure with a well-chosen subset of the selected wavebands, treatments were correctly classified with more than 95% accuracy. Specific narrow wavebands, from different portions of the spectrum, allowed the discrimination of plots differing in their irrigation and nitrogen treatments. This study supports past work suggesting that greater spectral resolution should lead to more consistent relationships between the spectral data and different crop status indicators.

4.2 Introduction

The applications of remote sensing (RS) in agricultural field include: crop identification (Saha and Jonna, 1994; Foody *et al.*, 1989), crop stress detection (Cibula

and Carter, 1992; Carter, 1994; Lelong-Camille *et al.*, 1998), crop disease detection (Lorenzen and Jensen, 1989; Peñuelas *et al.*, 1995), weed detection (Everitt *et al.*, 1995, 1996; Goel *et al.*, 2002), irrigation performance (Bastiaanssen and Bos, 1999; Sakthivadivel *et al.*, 1999; Bastiaanssen *et al.*, 2000), yield estimation (Maas, 1988; Clevers, 1997; Singh *et al.*, 2002), and, in general, crop management and precision farming (Moran *et al.*, 1997; Pearson *et al.*, 1994; Anderson *et al.*, 1999). Besides the large area that can be studied, the increased application of RS systems in agricultural management is mainly due to the improvements in spatial, spectral and temporal resolution of remotely sensed observations (Yang *et. al.*, 2002). Multispectral remote sensing systems offered information in limited bands, whereas hyperspectral sensors collect information in a large number of very narrow wavebands (Lillesand and Kiefer, 2000). With the launch of new hyperspectral RS satellites, the focus of current research has shifted to investigating the possibilities of hyperspectral RS.

Hyperspectral RS could offer great opportunities in recognizing, modeling, and categorizing terrestrial ecological characteristics (Thenkabail *et al.*, 2003). Specific narrow-band RS observations within certain portions of the spectral regions can significantly improve the classification accuracies and discrimination capabilities of RS in vegetation and agricultural crops (Bork *et al.*, 1999; Elvidge and Chen, 1995; Thenkabail *et al.*, 2002). As Gao (1999) demonstrated in the mapping of mangrove forests in a temperate zone, greater spectral resolution can now play a relatively greater role than spatial resolution.

Thenkabail *et al.* (2002) showed that the narrowband vegetation indices provided a more accurate estimation of crop parameters than did equivalent broadband-based indices. Similarly, Blackburn (1999) reported that for estimation of chlorophyll a, chlorophyll b and carotenoids, wavebands of 680 nm, 635 nm and 470 nm, respectively, were optimal. The literature supports the conclusion that hyperspectral data could provide considerable additional information in the estimation of various crop characteristics compared to similar information obtained from broadband sensors (Blackburn, 1998; Carter, 1998; Elvidge and Chen, 1995; Thenkabail *et al.*, 2002 and 2000; Asner *et al.*, 2000).

However, unlike multispectral data analysis, hyperspectral data are more complicated, given the large volume of information obtained over a short period of time, it requires the development and testing of new methods and algorithms of data analysis (Thenkabail *et al.*, 2002). These issues are becoming more crucial with the advent of new generations of satellites using hyperspectral sensors. For example, the data storage requirements for Hyperion imagery (220 narrow wavebands with 30 m spatial resolution) is about 70-fold that of Landsat imagery for a similar aerial coverage (Thenkabail *et al.*, 2002).

For the same electromagnetic spectrum, hyperspectral data will typically have hundreds, if not thousands, of wavebands, thus reducing the bandwidth of individual wavebands. Therefore, it will be important to pre-determine a sub-set of narrow-wavebands that could be used to provide detailed information on targets of interest, such as vegetation, soil, etc. However, more research needs to be conducted in this area, including establishing acceptable spatial and spectral accuracy levels.

Currently, efforts are being made to recognize important wavebands that can discriminate among different targets (e.g. crop vs. weeds). In a recent study, by comparing hyperspectral and multispectral-based vegetation indices, Thenkabail *et al.* (2002) found that narrow wavebands strongly correlated with a number of crop characteristics. In another study, Vrindts *et al.* (2002) used discriminant analysis, with both laboratory and field observations, to discriminate between weeds and crops (sugar beet and maize). The laboratory study allowed crop-weed classification with <1% misclassification errors. However, the misclassification rate was 85% for corn and 3% for weeds in the field study. The value of shortwave infrared (SWIR) spectral reflectance for determining equivalent water thickness (EWT) at leaf level was investigated by Ceccato *et al.* (2001). They concluded that although the SWIR is sensitive to EWT, the SWIR reflectance alone is not sufficient for assessing EWT at leaf level. Estimating LAI using data acquired by the Hyperion satellite and applying vegetation indices derived from NIR and SWIR wavebands was investigated by Gong *et al.* (2003). They found that the wavelengths of 820, 1040, 1200, 1250, 1650, 2100, and 2260 nm were the most valuable bands for estimation of LAI.

This study was undertaken to determine the best hyperspectral narrow wavebands to discriminate between different water and nitrogen stresses and their interactions in a corn field. In this study we used different multivariate discriminant analysis approaches. Canopy reflectance observations, collected in the range from 350 to 1800 nm, were analyzed to determine which wavebands could be used to classify types of water and nitrogen stresses.

4.3 Materials and methods

4.3.1 Study site and experimental details

During the 2002 growing season, a split-plot experiment was conducted on sandy soil at the Research farm of Macdonald Campus, McGill University in Ste-Anne-de-Bellevue, Quebec, Canada. A schematic diagram of the experimental design is presented in Figure 4.1.

This soil belongs to the St. Amable complex, which has deep sandy deposits subtended by a clay layer (Lajoie, 1960). Corn (*Zea mays* L. cv., *Hybrid DKC42-22*) was grown in forty, 9.0 m x 10.0 m, test plots laid out in a split-plot design with 4 blocks. In each block, irrigation (non-irrigated, irrigated from 19 July 2002 onward) was the main treatment and nitrogen fertilizer application rate (50, 100, 150, 200, and 250 kg ha^{-1}) was the sub-treatment. Corn was sown on 22 May 2002. Corn plots were planted under two irrigation schemes and five different nitrogen treatments. Each sub-plot consisted of 12 rows of corn with rows running east-west, with 75 cm row spacing and population of 76,000 plants per ha.

4.3.2 Collection of Spectral Data

Hyperspectral data were acquired throughout the season, using a handheld spectroradiometer (FieldSpec Pro model, Analytical Spectral Device, Boulder, CO, USA) with 2151 wavebands from 350 to 2500 nm (1.0 nm bandwidth) and a 15° field of view. This spectroradiometer contains three different sensors to cover the spectral range of 350 to 2500nm. The raw spectra acquired with the field spectroradiometer was converted to

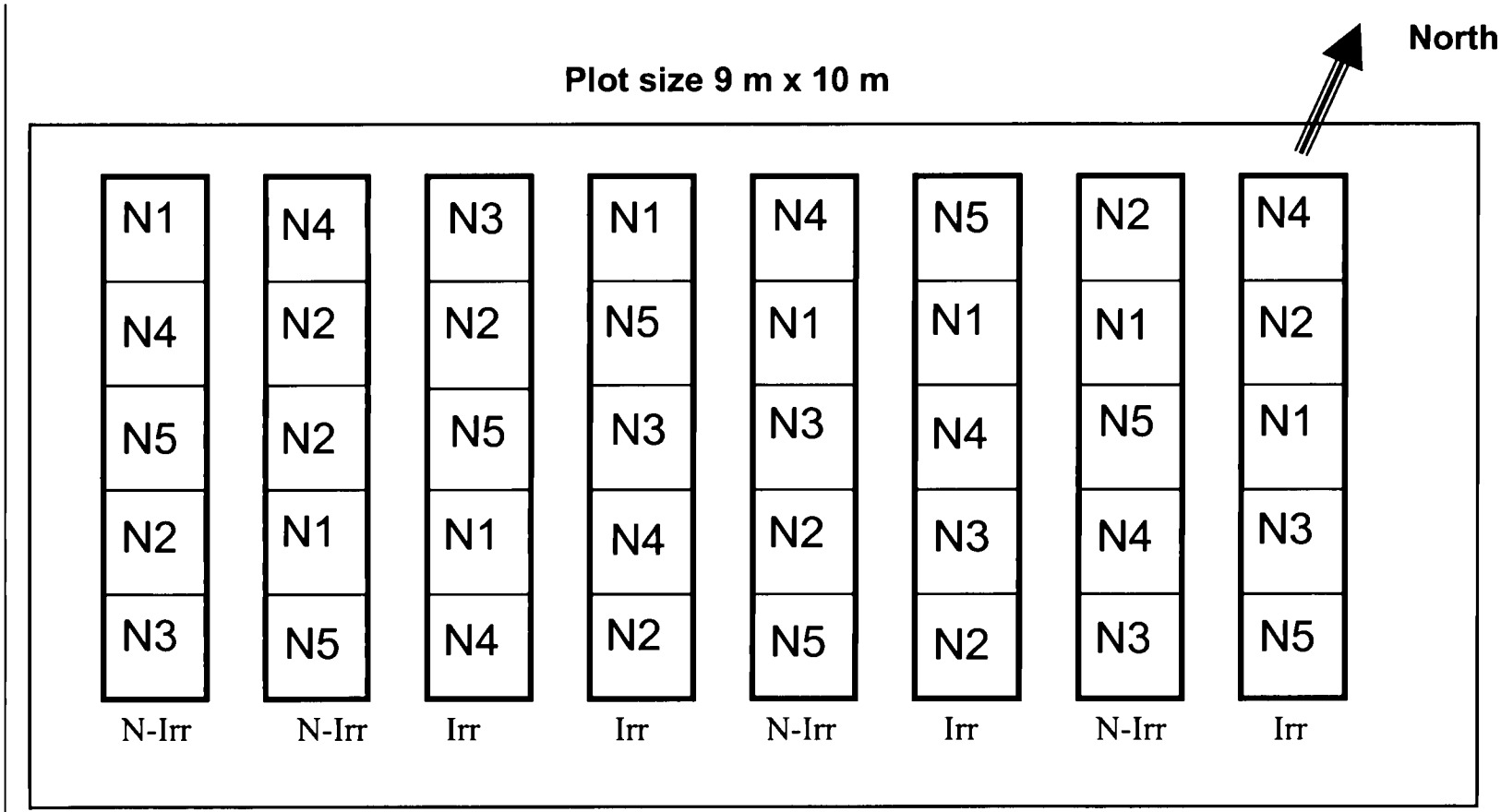


Figure 4.1 Experimental layout, 2002 (N1 - N5 are: 50, 100, 150, 200, 250 kg N ha $^{-1}$, respectively, Irr and N-Irr are: irrigated and non-irrigated treatments).

reflectance using a white reference panel. A white reference was taken to measure all the reflected light from a white reference panel, which is considered to be equal to the radiation reaching the target. This was checked regularly at a 10 to 15 min interval. Six measurements were collected per plot and the mean value was used in the data analysis. For each plot, the first three measurements were made with the spectroradiometer sensor placed directly over corn plants and the other three were made with the instrument located between corn rows. Due to excessive noise in the water absorption region and some technical problems in the calibration of the third sensor, the measurements ranging from 1345 to 1430 nm and 1800 to 2500 nm, respectively, were removed from the spectral data set, thereby leaving 1364 wavebands for analysis. Since irrigation only began on 19 July, attempts to discriminate the irrigation treatment were limited to sampling dates after this date: 31 July, 10 August, and 26 August. The mean spectral response of corn under different nitrogen application rates and water treatments is illustrated in Figure 4.2 a and b for the data sets collected on 31 July from irrigated and non-irrigated treatments, respectively. In non-irrigated plots, the treatments that received 200 and 250 kg N ha^{-1} had comparatively greater reflectance responses. Whereas, in irrigated plots the treatments that received 50 and 100 kg N ha^{-1} provided a greater response in the visible and near infrared. Although, it is hard to assess why such response occurred, it may be attributable to different chemical components of the plants tissues (e.g., chlorophyll). Similarly, the lower response from the treatments with higher nitrogen application rates could be explained by increased osmotic potential due to excessive nitrogen in the irrigated treatments. Furthermore, at the spectral range of 1450-1750 nm the reflectance values were higher in the non-irrigated plots. In addition, there was greater variation between treatments in that region. The figure clearly demonstrates the difference in reflectance values for the various treatments, and it also shows that certain wavebands in the visible region and many more in the near infrared region can be useful in classifying different treatments.

4.3.3 Plant parameters

Various crop parameters, such as plant height, leaf greenness, leaf area index (LAI), biomass, and soil moisture, were measured during the entire growing season. Crop

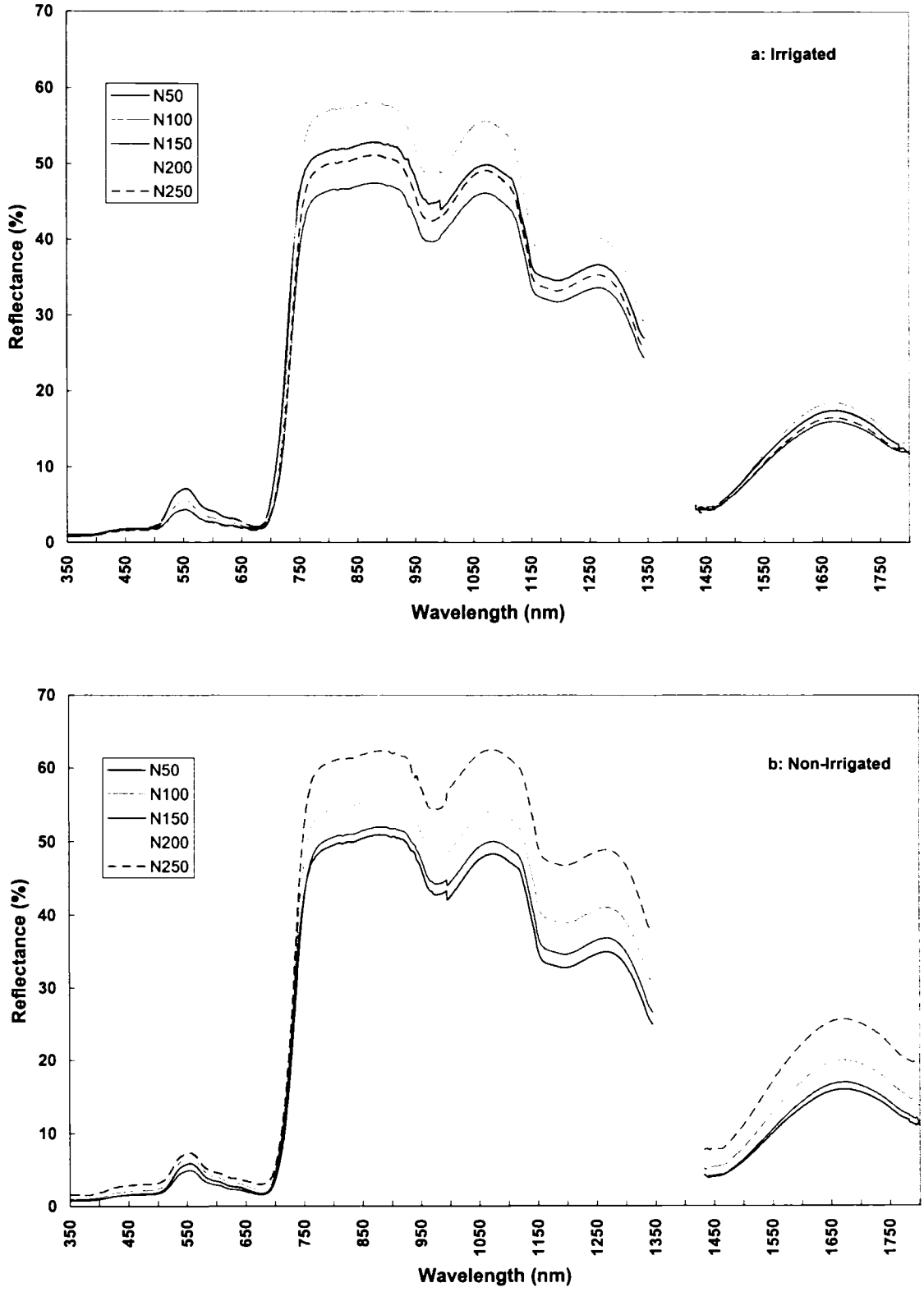


Figure 4.2 Measured spectral response curves of corn under irrigated and non-irrigated conditions with different nitrogen application rate at July 31, 2002 (N_{50}-N_{250} = 50, 100, 150, 200, 250 kg ha^{-1} nitrogen treatments).

yield and final biomass were also recorded at the harvesting time. In each plot, the height of ten different plants was measured. LAI was measured using a LAI-2000 Plant Canopy Analyzer (Li-Cor, Inc., Lincoln, Nebraska, USA). LAI has no dimension; however, it can be expressed as m^2 foliage area per m^2 ground area. SPAD Chlorophyll meter (Minolta Camera Ltd., Osaka, Japan) was used to measure the greenness or the amount of chlorophyll in plants. Ten representative plants from each plot were harvested and weighed for the estimation of biomass. Crop yield was calculated by harvesting ten representative plants, from each plot. Table 4.1 shows the variations in plant parameters as a function of different water and nitrogen treatments.

4.3.4 Data analysis

The ground-based canopy hyperspectral data were analyzed with the statistical software package SAS (Version 8.20, Cary, NC, USA). Firstly, to discriminate between different treatments, a set of suitable wavelengths were selected using the STEPDISC procedure of SAS. Next, in order to evaluate the discrimination capability of the selected wavelengths, two SAS procedures, namely DISCRIM and CANDISC, were used. In addition, Principal Component Analysis (PCA) was applied to not only reduce the dimensionality of the datasets, but also to generate an uncorrelated data set from correlated data.

The STEPDISC procedure can perform discriminant analysis using one of three procedures: forward selection, backward elimination, or stepwise. The stepwise procedure is a combination of forward selection and backward elimination methods, where forward selection is used for the inclusion of a variable and backward elimination for the removal of variables no longer significant in the model. It is recommended that the entry significance level for variables be set somewhere between 0.25 and 0.5, and 0.15 for variable removal (Johnson, 1998). Although the STEPDISC procedure is supposed to select the most significant variables suitable for discrimination of the treatments, there are no guarantees that the finally selected variables are the best set of variables, particularly when there is high collinearity in data (Murray, 1977; Johnson, 1998; Vrindts *et al.*, 2002). Johnson (1998) noted that while it was counterintuitive, a small subset of well-chosen variables often allows a better discrimination between treatments than with the entire set of variables.

Table 4.1 Plant parameters as a function of different treatments (average of four replicates).

		Irrigated					Non-Irrigated				
Nitrogen		**N_{50}**	**N_{100}**	**N_{150}**	**N_{200}**	**N_{250}**	**N_{50}**	**N_{100}**	**N_{150}**	**N_{200}**	**N_{250}**
Yield	**kg ha^{-1}**	**10101**	**10318**	**12003**	**12467**	**12406**	**6638**	**7048**	**7890**	**7914**	**8820**
Biomass (kg ha^{-1})	**25-Jul**	**5063**	**5360**	**5109**	**6375**	**5256**	**4966**	**4837**	**4238**	**5395**	**5411**
	27-Aug	**12319**	**14838**	**13962**	**15204**	**13920**	**9335**	**9889**	**11837**	**11778**	**11696**
	Final	**18275**	**20569**	**21859**	**23032**	**23689**	**13700**	**13953**	**15480**	**15698**	**17738**
LAI	**31-Jul**	**2.37**	**2.45**	**2.70**	**2.88**	**2.54**	**1.91**	**2.44**	**2.60**	**2.59**	**2.61**
	7-Aug	**3.64**	**4.02**	**4.20**	**4.23**	**4.08**	**3.08**	**3.35**	**3.70**	**3.88**	**3.67**
	16-Aug	**3.22**	**3.29**	**3.59**	**3.64**	**3.69**	**2.21**	**2.34**	**2.68**	**2.84**	**2.77**
SPAD	**31-Jul**	**37.5**	**40.5**	**47.1**	**46.8**	**48.8**	**37.6**	**37.5**	**44.3**	**44.5**	**44.3**
	9-Aug	**32.8**	**44.7**	**48.1**	**52.4**	**47.2**	**38.3**	**45.9**	**44.8**	**54.8**	**42.7**
	21-Aug	**36.3**	**44.8**	**49.2**	**51.9**	**50.2**	**32.2**	**35.2**	**36.9**	**36.4**	**36.9**
Plant Height (cm)	**31-Jul**	**142.6**	**153.3**	**159.8**	**174.8**	**166.2**	**141.2**	**140.7**	**149.5**	**148.2**	**144.2**
	11-Aug	**204.3**	**222.5**	**223.2**	**233.2**	**222.2**	**202.1**	**181.7**	**200.9**	**201.3**	**199.3**
	21-Aug	**215.2**	**229.7**	**230.0**	**242.0**	**232.7**	**210.4**	**191.0**	**196.8**	**201.3**	**210.9**

N_{50}-N_{250}= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments

Using the DISCRIM and CANDISC procedures of SAS, the suitability of the selected wavelengths and/or a subset of selected wavebands were examined with discriminant models. Based on the quantitative variables, the DISCRIM procedure calculates different discriminant functions for categorizing observations into groups.

Different parameters that can be used for the development of discriminant functions are the individual within-group covariance matrices, or the pooled covariance matrix, and prior probabilities of the groups. The misclassification matrices were calculated by determining the number of wrongly classified groups in any single classes. In CANDISC, all the data set is used for developing the model and no subset is used for validation, while the performance of the discriminant models is evaluated using a cross-validation method in the DISCRIM procedure.

Also PCA, a procedure which can be employed with correlated data, can generate a set of uncorrelated variables from a set of correlated variables. These newly generated variables are called principal components (PCs). Overall, the goal of using PCA is to reduce the dimensionality of the dataset to an optimal level and, thus, introducing a new set of meaningful variables (Johnson, 1998).

4.4 Results and Discussion

4.4.1 Selection of wavebands

Considering the combination of two irrigation and five nitrogen fertilizer levels, there were ten groups (or populations), which were used for discrimination purposes with the STEPDISC procedure. The summary of the results of the STEPDISC procedure are presented in Table 4.2 a to c for reflectance data collected on 31 July, and 10 and 26 August 2002, respectively. The STEPDISC procedure was able to select the most effective of the 1364 wavebands for discriminating between different treatments. The numbers of selected wavebands were 39 for 31 July and 10 August, and 38 for 26 August. Also, among all three datasets, some ultraviolet wavebands (350-400 nm) were selected. For the 31 July data set, the selected wavebands chosen were primarily in the visible portion of the spectrum, rather than in the infrared region and shortwave infrared (SWIR), as has been found in other two data sets.

Table 4.2 a STEPDISC results for waveband selection for **July 31, 2002, data set**

Number		Wavelength		Partial		
Step	In	Entered	Removed	R-Square	F Value	Pr > F
1	1	420		0.5135	3.52	0.0044
2	2	453		0.5678	4.23	0.0014
3	3	448		0.5244	3.43	0.0058
4	4	407		0.5857	4.24	0.0017
5	5	488		0.5255	3.20	0.0097
6	6	471		0.5142	2.94	0.0159
7	7	446		0.5001	2.67	0.0265
8	8	678		0.5004	2.56	0.0333
9	9	426		0.4862	2.31	0.0525
10	10	498		0.5916	3.38	0.0103
11	11	677		0.5448	2.66	0.0329
12	12	365		0.5140	2.23	0.0673
13	13	360		0.6614	3.91	0.0067
14	14	389		0.6556	3.60	0.0112
15	15	428		0.6292	3.02	0.0261
16	16	445		0.6373	2.93	0.032
17	17	398		0.6939	3.53	0.0173
18	18	414		0.7210	3.73	0.0159
19	19	392		0.7352	3.70	0.0191
20	20	489		0.7825	4.40	0.0121
21	21	364		0.7445	3.24	0.0406
22	22	647		0.8675	6.55	0.005
23	23	662		0.8157	3.94	0.0333
24	22		446	0.6982	2.06	0.1616
25	23	643		0.8408	4.69	0.0201
26	24	456		0.8303	3.81	0.0459
27	25	642		0.9277	8.55	0.0084
28	26	513		0.9393	8.59	0.0145
29	27	493		0.9646	17.01	0.0032
30	28	583		0.9947	92.98	0.0003
31	29	691		0.9929	52.08	0.0039
32	30	413		0.9999	1940.24	0.0005
33	31	594		0.999	279.72	0.0036
34	32	499		1	77901.90	0.0028
35	33	436		1	103678.00	0.0024
36	34	354		0.9999	2864.65	0.0142
37	35	588		1	Infty	<.0001
38	36	388		1	Infty	<.0001
39	37	654		1	Infty	<.0001
40	38	496		1	Infty	<.0001
41	39	613		1	Infty	<.0001

Table 4.2 b STEPDISC results for waveband selection for **August 10, 2002, data set**

Number		Wavelength		Partial		
Step	In	Entered	Removed	R-Square	F Value	Pr > F
1	1	496		0.4046	2.26	0.0451
2	2	370		0.7309	8.75	<.0001
3	3	510		0.5662	4.06	0.002
4	4	508		0.5167	3.21	0.009
5	5	1500		0.4913	2.79	0.0196
6	6	1789		0.4733	2.50	0.0343
7	7	432		0.5091	2.77	0.0225
8	8	375		0.462	2.19	0.062
9	9	1504		0.541	2.88	0.0207
10	10	511		0.5385	2.72	0.0281
11	11	492		0.4991	2.21	0.0666
12	12	1503		0.4715	1.88	0.1175
13	13	394		0.5003	2.00	0.1004
14	14	414		0.7249	4.98	0.0022
15	15	506		0.5384	2.07	0.0974
16	16	1502		0.5809	2.31	0.073
17	17	380		0.6692	3.15	0.0269
18	18	373		0.6449	2.62	0.0559
19	19	493		0.654	2.52	0.0689
20	20	454		0.7701	4.09	0.0157
21	21	457		0.8233	5.18	0.0085
22	22	443		0.7789	3.52	0.0373
23	23	452		0.7749	3.06	0.065
24	24	499		0.8618	4.85	0.0246
25	25	703		0.9587	15.46	0.0017
26	26	413		0.9596	13.20	0.0055
27	27	445		0.9627	11.47	0.0158
28	28	1129		0.9925	43.84	0.005
29	29	420		0.9968	68.16	0.0145
30	30	367		0.9968	76.69	0.0129
31	31	1514		1	94594.6	0.0025
32	32	1471		1	2879414.00	0.0005
33	33	794		1	5951096.00	0.0003
34	34	1145		1	9092832.00	0.0003
35	35	598		1	252066.00	0.0015
36	36	467		0.9999	4220.56	0.0113
37	37	699		1	Infty	<.0001
38	38	1142		1	Infty	<.0001
39	39	770		1	Infty	<.0001

Table 4.2 c STEPDISC results for waveband selection for **August 26, 2002, data set**

Number		Wavelength		Partial		
Step	In	Entered	Removed	R-Square	F Value	Pr > F
1	1	461		0.6469	6.11	<.0001
2	2	388		0.4994	3.21	0.008
3	3	425		0.4949	3.05	0.0113
4	4	724		0.4931	2.92	0.015
5	5	1798		0.4670	2.53	0.0311
6	6	396		0.4919	2.69	0.0244
7	7	1231		0.5319	3.03	0.0145
8	6		461	0.3646	1.53	0.1938
9	7	1227		0.546	3.21	0.0108
10	8	426		0.5291	2.87	0.0198
11	9	379		0.5168	2.61	0.0318
12	10	373		0.5696	3.09	0.0159
13	11	386		0.5274	2.48	0.0435
14	12	1561		0.5671	2.77	0.0296
15	13	435		0.5603	2.55	0.0436
16	12		373	0.3972	1.32	0.2944
17	13	456		0.5822	2.79	0.0307
18	14	1167		0.6439	3.42	0.0141
19	15	444		0.6448	3.23	0.0198
20	16	1232		0.7379	4.69	0.0043
21	17	1210		0.6937	3.52	0.0174
22	16		1231	0.521	1.69	0.1822
23	17	1295		0.7594	4.91	0.0042
24	18	420		0.6987	3.35	0.024
25	19	1177		0.7656	4.36	0.0103
26	20	458		0.7722	4.14	0.0151
27	21	384		0.7907	4.2	0.0176
28	22	1173		0.8773	7.15	0.0036
29	23	460		0.8451	4.85	0.0183
30	24	1240		0.9095	7.82	0.0064
31	25	1184		0.9405	10.53	0.0048
32	26	375		0.9168	8.26	0.0095
33	27	1196		0.974	23.4	0.0015
34	28	1297		0.9953	104.93	0.0002
35	29	370		0.9887	32.94	0.0077
36	30	451		0.9988	204.9	0.0049
37	31	1239		0.9989	250.38	0.004
38	32	1168		1	49415.8	0.0035
39	33	1149		1	32459.6	<.0001
40	34	432		1	7091.84	0.009
41	35	1159		1	6.78E+08	<.0001
42	36	440		1	Infty	<.0001
43	37	1200		1	Infty	<.0001
44	38	1312		1	Infty	<.0001

As mentioned earlier, Johnson (1998) had reported that, although it was difficult to explain, a smaller subset of well-chosen variables may often permit better discrimination between treatments than can be obtained with the entire set of variables, selected by the STEPDISC procedure, particularly when inter-collinearity exists in the data. However, there is no systematic method for selecting this subset of variables. Below, we propose a systematic method of variable selection based on the DISCRIM procedure of SAS.

4.4.2 Discriminant analysis

The DISCRIM procedure was used to select the best subset of data which could discriminate between different categories within the data. For this purpose, different DISCRIM models were applied using various numbers of wavebands, chosen on the basis of their order of entry in the STEPDISC procedure selection. In the development of DISCRIM models, the data can be divided into two parts. The first part is used for the construction and development of the model (calibration), while the second "unseen" part is used to validate the classification accuracy of the model. Since our data sets are small, we repeatedly used "take one data-out approach" for cross-validation. A summary of the results for all three data sets is presented in Table 4.3. Generally speaking, using all wavebands selected with the STEPDISC procedure provided higher accuracy in the calibration phase than during cross-validation. However, using subsets of wavebands, based on their order of entry into the STEPDISC procedure, decreased classification accuracy for the calibration data sets. For cross-validation, the accuracy first increased up to a certain numbers of wavebands and then decreased thereafter. The classification matrices for the 31 July and 10 August data sets showed that the DISCRIM procedure was fully able to classify all different treatments. For the 26 August data set, the misidentification rates were 2.5% and 5% respectively for calibration and validation data sets (Table 4.3).

The classification results for the cross validation data sets are provided in Table 4.4 a and b for the 31 July and 26 August datasets. The results for the August 10 data set are similar to those for the July 31 data set. The wavelengths of the selected wavebands

are also listed in the table. For the 31 July data set, all selected wavebands are within 350 – 700 nm (about 50% are in the green region). In the case of 26 August data set, nearly 44% of the selected wavebands are located in near-infrared region. Although 43% of the selected wavebands are located in the green region for the August 10 data set, some wavebands from the near-infrared and SWIR (20 %) region were also chosen. This supports previous work that showed the SWIR spectral data along with that from other wavelengths can be useful for water stress detection. These wavebands are mainly concentrated around a wavelength of 1500 nm.

Comparing results for different chosen subsets, one notes that, whenever the F value approaches infinity, the DISCRIM procedure is no longer able to discriminate treatments in a significant way. For example, on 31 July, the discrimination error risk of the DISCRIM procedure was estimated to be 57.5 percent when applied to all selected wavebands from STEPDISC (39 wavebands), whereas for the first 34 wavebands (for which F was not very large), the DISCRIM procedure was able to fully and correctly discriminate the different treatments. Similarly, zero discrimination error was obtained with the first 30 selected wavebands on 10 August. For 26 August, in the best case (first 25 wavebands), the DISCRIM procedure exhibited a 95% accuracy. In the case of the 31 July data, the DISCRIM procedure was able to fully and correctly classify variables, using anywhere from 26 to 34 wavebands.

In addition to the DISCRIM procedure, a canonical discriminant analysis (CDA) was performed by applying CANDISC procedure. The main challenge was how to choose an effective subset from the wavebands, selected from the STEPDISC procedure. For this purpose the CANDISC procedure was performed with different subsets. In the selection of subsets, different wavebands were again selected based on their entrance order in the STEPDISC procedure. The best results for the three sampling dates, with classes being the irrigation and nitrogen treatments, were obtained using a subset of 20 wavebands (results of CANDISC for 31 July are given in Table 4.5). In all three cases, a squared canonical correlation of approximately 0.99 was obtained. The total variance explained by only the first two canonical functions were 94.6, 83.7, and 93.7 % for the 31 July, and 10 and 26 August data sets, respectively. Using the first two canonical variables,

Table 4.3 Summary of misclassification matrices obtained from DISCRIM procedure (%)

July 31, 2002			August 10, 2002			August 26, 2002		
# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation	# of wavebands	Calibration	Cross validation
39	0.0	57.5	39	0.0	70.0	38	2.5	42.5
34	0.0	0.0	36	0.0	27.5	35	2.5	10.0
30	0.0	0.0	30	0.0	0.0	31	2.5	7.5
25	0.0	5.0	25	0.0	10.0	27	2.5	5.0
20	0.0	25.0	20	0.0	27.5	25	2.5	5.0
15	0.0	40.0	15	0.0	42.5	20	2.5	15.0
10	10.0	45.0	10	10.0	50.0	15	5.0	37.5
5	15.0	77.5	5	0.0	95.0	10	7.5	50.0
3	47.5	70.0	3	20.0	95.0	5	2.5	92.5

Table 4.4a Classification Matrix obtained from DISCRIM procedure for cross validation case in July 31 data sets

		Predicted										
		Irr-N_{50}	Irr-N_{100}	Irr-N_{150}	Irr-N_{200}	Irr-N_{250}	NIrr-N_{50}	NIrr-N_{100}	NIrr-N_{150}	NIrr-N_{200}	NIrr-N_{250}	Total
Actual	Irr-N_{50}	4	0	0	0	0	0	0	0	0	0	4
	Irr-N_{100}	0	4	0	0	0	0	0	0	0	0	4
	Irr-N_{150}	0	0	4	0	0	0	0	0	0	0	4
	Irr- N_{200}	0	0	0	4	0	0	0	0	0	0	4
	Irr- N_{250}	0	0	0	0	4	0	0	0	0	0	4
	NIrr- N_{50}	0	0	0	0	0	4	0	0	0	0	4
	NIrr- N_{100}	0	0	0	0	0	0	4	0	0	0	4
	NIrr- N_{150}	0	0	0	0	0	0	0	4	0	0	4
	NIrr- N_{200}	0	0	0	0	0	0	0	0	4	0	4
	NIrr- N_{250}	0	0	0	0	0	0	0	0	0	4	4
	Total	4	4	4	4	4	4	4	4	4	4	40
Selected wavebands		**420, 453, 448, 407, 488, 471, 678, 426, 498, 677, 365, 360, 389, 428, 445, 398, 414, 392, 489, 364, 647, 662, 643, 456, 642, 513, 493, 583, 691, 413**										

Irr = irrigated, NIrr = non-irrigated, N_{50}-N_{250}= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments.

Table 4.4b Classification Matrix obtained from DISCRIM procedure for cross validation case in August 26, data sets

		Predicted										
		Irr-N_{50}	Irr-N_{100}	Irr-N_{150}	Irr-N_{200}	Irr-N_{250}	NIrr-N_{50}	NIrr-N_{100}	NIrr-N_{150}	NIrr-N_{200}	NIrr-N_{250}	Total
Actual	Irr-N_{50}	4	0	0	0	0	0	0	0	0	0	4
	Irr-N_{100}	0	4	0	0	0	0	0	0	0	0	4
	Irr-N_{150}	0	0	4	0	0	0	0	0	0	0	4
	Irr- N_{200}	0	0	0	3	0	0	0	1	0	0	4
	Irr- N_{250}	0	0	0	0	4	0	0	0	0	0	4
	NIrr- N_{50}	0	0	0	0	0	4	0	0	0	0	4
	NIrr- N_{100}	0	0	0	0	0	0	4	0	0	0	4
	NIrr- N_{150}	0	0	0	1	0	0	0	3	0	0	4
	NIrr- N_{200}	0	0	0	0	0	0	0	0	4	0	4
	NIrr- N_{250}	0	0	0	0	0	0	0	0	0	4	4
	Total	4	4	4	4	4	4	4	4	4	4	40
Selected wavebands		**388, 425, 724, 1798, 396, 1227, 426, 379, 386, 1561, 435, 456, 1167, 444, 1232, 1210, 1295, 420, 1177, 458, 384, 1173, 460, 1240, 1184**										

Irr = irrigated, NIrr = non-irrigated, N_{50}-N_{250}= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments.

Table 4.5 Canonical discriminant analysis for Irrigation with Nitrogen as class variable, using the first 20 selected wavebands (July 31, 2002)

	Eigenvalues	Canonical Correlations	Squared Canonical Correlations	Cumulative Variability	Approx F Value	Pr > F
1	368.650	0.999	0.997	0.833	5.15	<.0001
2	50.295	0.990	0.981	0.946	3.06	<.0001
3	8.480	0.946	0.895	0.966	2.11	<.0001
4	6.270	0.929	0.862	0.980	1.8	0.0026
5	4.684	0.908	0.824	0.990	1.48	0.0436
6	1.927	0.811	0.658	0.995	1.11	0.3439
7	1.183	0.736	0.542	0.997	0.93	0.5916
8	0.842	0.676	0.457	0.999	0.79	0.7321
9	0.339	0.503	0.253	1	0.54	0.8642

the irrigation and nitrogen treatments were distinguished into different classes for each of the sampling dates (Figure 4.3-4.5). These Figures show that, using only two canonical variables, CDA was able to differentiate most of the classes.

The PCA was carried out next using the PRINCOMP procedure of the SAS software. PRINCOMP was applied both to wavebands, selected through the STEPDISC procedure, and to the complete reflectance data sets for the three days. For the 31 July data set, the first 10 eigenvalues of the covariance matrix, the difference between successive eigenvalues, the proportion of the covariance explained by each eigenvalue, and the cumulative proportion of the explained variance is presented in Table 4.6. With pre-selected wavebands, 93.5% of the variance was explained by the first PC, whereas, using all wavebands, only 89.2 percent of the variance was explained by the first PC. Similarly, using all wavebands, the variance explained by the first PC was 77.5% and 99.2% for August 10 and 26, respectively. It was 78.7% and 61.7% for August 10 and 26, respectively, using the wavebands selected with the STEPDISC procedure. In all cases, more than 99 percent of variance was explained by the first five PCs. Moreover, the ability of PCs for classifying the different treatments with DISCRIM and CANDISC procedures was also explored. For all three days, although the DISCRIM procedure was able to identify the different treatments during the calibration step, no satisfactory results were obtained during the cross-validation phase. This may have resulted from the limited number of samples for each treatment (four in our case). Similarly, CANDISC models were also developed using different number of PCs. The results showed that CANDISC may produce better classification accuracies with higher numbers of PCs.

4.5 Conclusions

Appropriately selected narrow waveband hyperspectral observations can be used as a valuable source of information regarding agricultural crop parameters. A stepwise approach was highly effective in selecting the significant wavebands for further analysis. Using all wavebands identified by the stepwise procedure and various discriminant analysis methods, differentiation between the various irrigation and nitrogen treatments

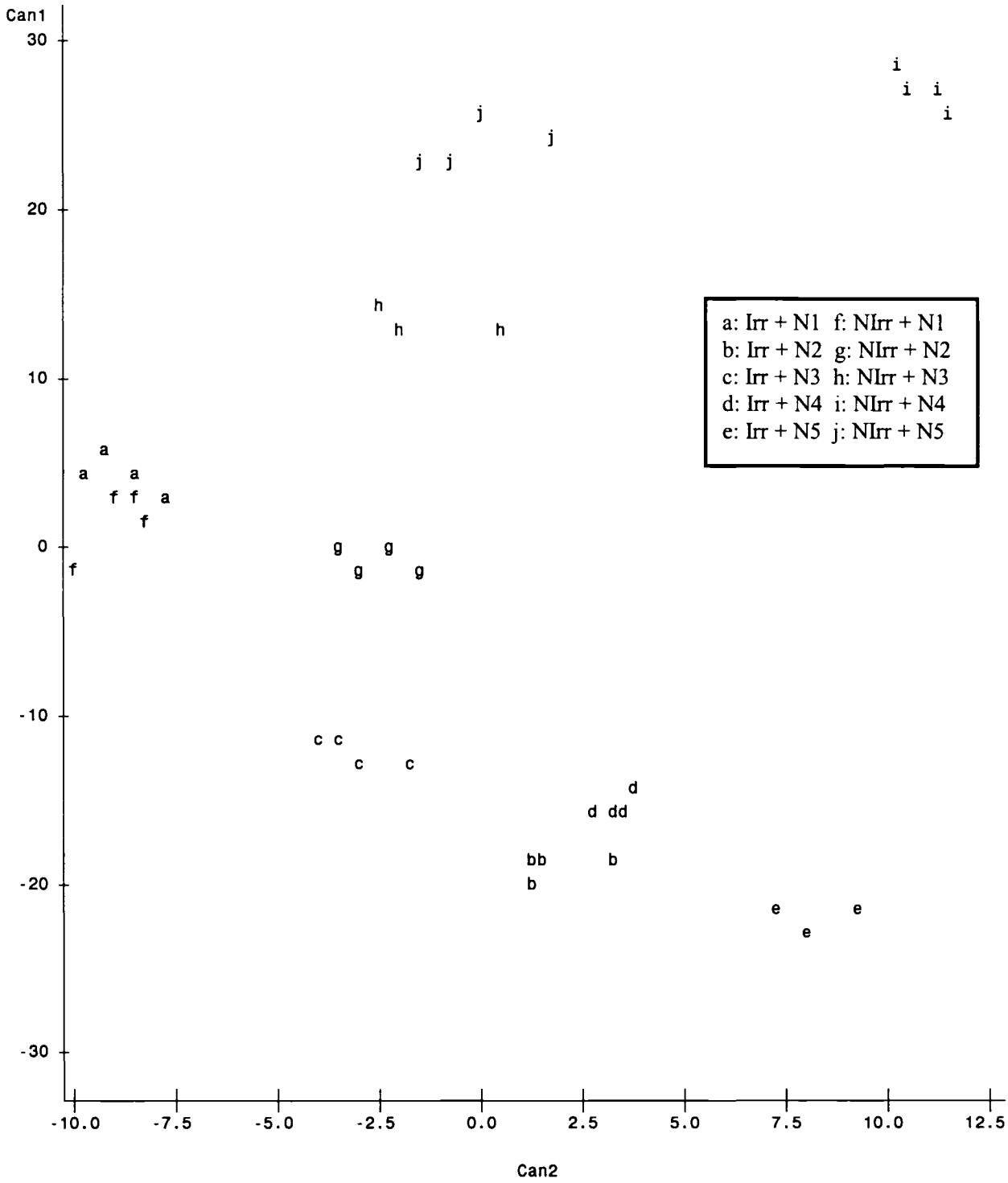


Figure 4.3 Canonical discriminant analysis of July 31 data set with irrigation and nitrogen as class variables (**Irr = irrigated, NIrr = non-irrigated, N1-N5= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments).**

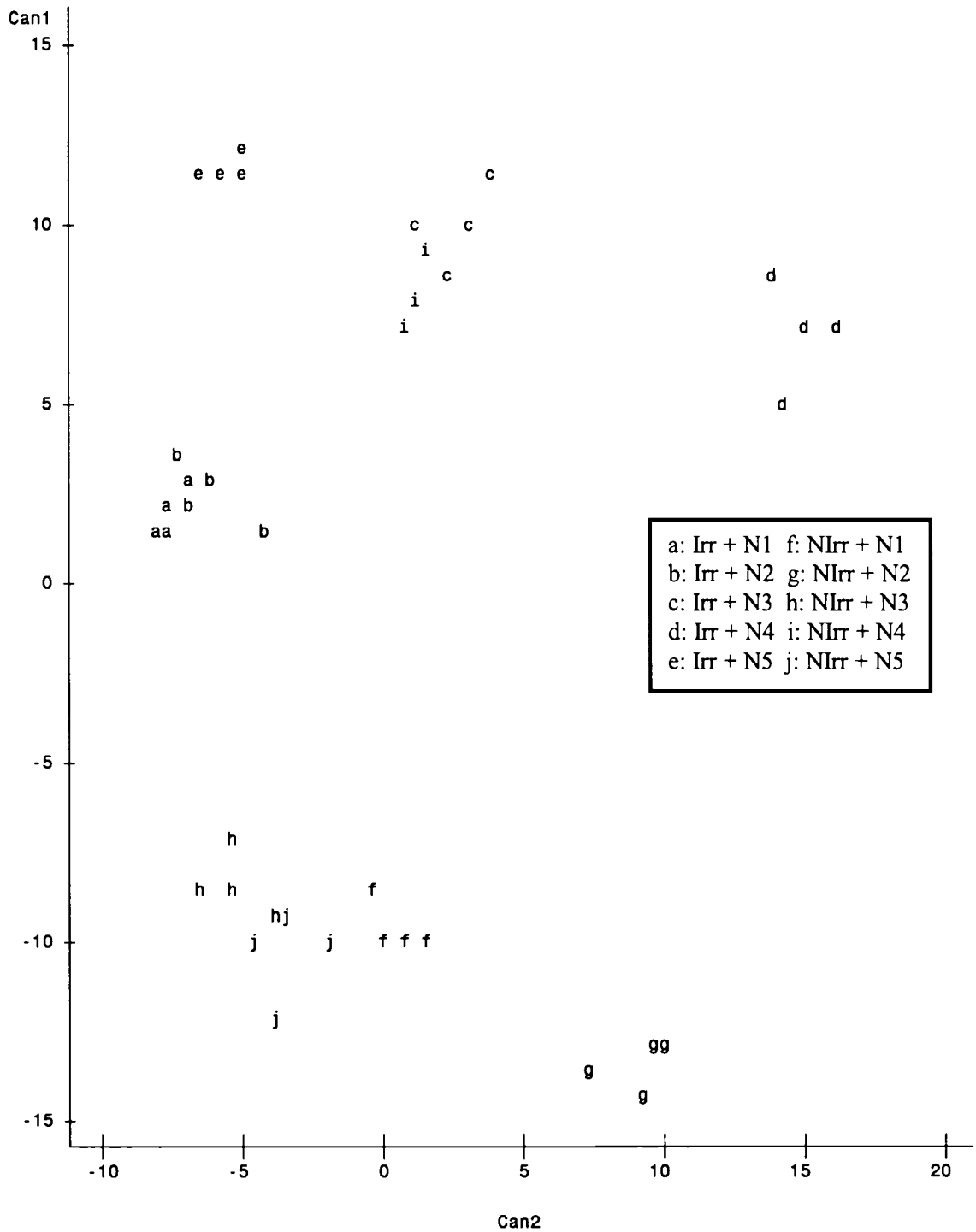


Figure 4.4 Canonical discriminant analysis of August 10 data set with irrigation and nitrogen as class variables (**Irr = irrigated, NIrr = non-irrigated, N1-N5= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments).**

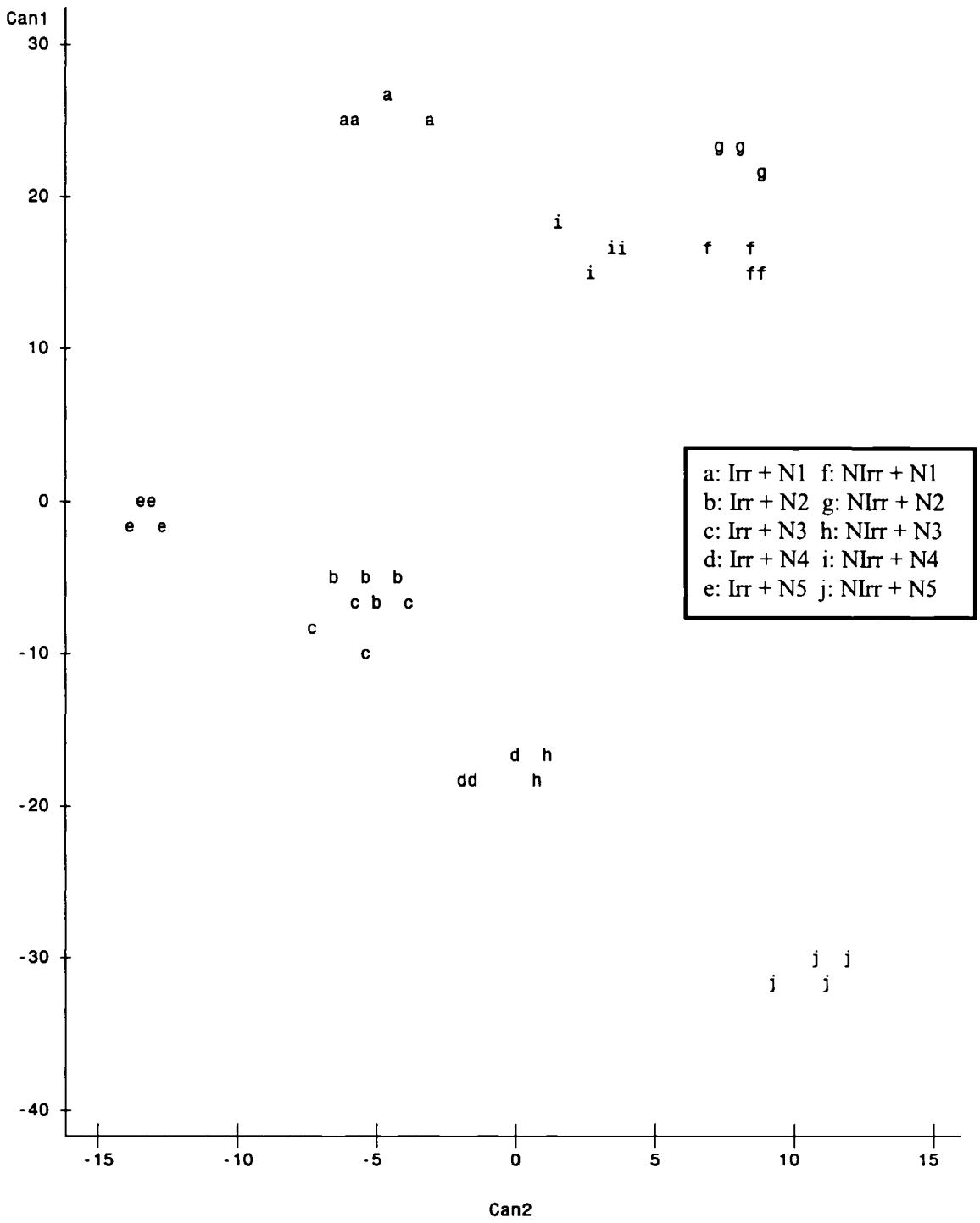


Figure 4.5 Canonical discriminant analysis of August 26 data set with irrigation and nitrogen as class variables (**Irr = irrigated, NIrr = non-irrigated, N1-N5= 50, 100, 150, 200, and 250 kg ha^{-1} nitrogen treatments).**

Table 4.6 Eigenvalues of the correlation matrix calculated from PRINCOMP procedure (July 31, 2002)

Principal Components	Using all wavebands				Using selected wavebands from STEPDISC			
	Eigenvalues	Difference	Proportion	Cumulative	Eigenvalues	Difference	Proportion	Cumulative
Prin1	1216.791	1139.710	0.892	0.892	35.530	33.347	0.935	0.935
Prin2	77.081	28.699	0.057	0.949	2.184	2.051	0.058	0.993
Prin3	48.382	33.047	0.036	0.984	0.133	0.072	0.004	0.996
Prin4	15.335	12.954	0.011	0.995	0.061	0.037	0.002	0.998
Prin5	2.380	1.028	0.002	0.997	0.025	0.015	0.001	0.998
Prin6	1.353	0.629	0.001	0.998	0.010	0.001	0.000	0.999
Prin7	0.723	0.333	0.001	0.999	0.009	0.002	0.000	0.999
Prin8	0.390	0.092	0.000	0.999	0.007	0.001	0.000	0.999
Prin9	0.298	0.079	0.000	0.999	0.005	0.001	0.000	0.999
Prin10	0.219	0.057	0.000	0.999	0.005	0.001	0.000	0.999

was not always successful. However, when a well-chosen subset of selected wavebands was used, the results from the discriminant analysis methods were greatly improved.

The discriminant analysis method was found to be very useful in identifying different treatments, showing only 5% misclassification rate for the 26 August observations, and 0% misclassification rates for the 31 July and 10 August observations. However, the number of significant wavebands varied with time. The CDA showed good capability in identifying the different treatments. The variability that could be explained by the first two canonical functions was more than 94% for the 31 July and 26 August data sets, and about 84% for the 10 August data. Principal component analysis was also found to be useful in highlighting and reducing data dimensionality.

PREFACE TO CHAPTER 5

We have so far determined that hyperspectral information, collected from both airborne and handheld sensors, could be used for differentiating between weed, nitrogen and water stresses in a corn field. Thereafter, the study was focused on using a crop growth model, namely STICS, for the prediction of crop yield under such stresses. First, the applicability of model was evaluated for our climatic conditions, and then efforts were made to link the model with hyperspectral data. This was accomplished by estimating the LAI values of the crop canopy from remote sensing observation, and using these LAI values in the model. Chapter 5 is focused on nitrogen and water treatments. The effect of weed treatments on crop yield estimation, using the STICS model, is discussed in Chapter 6.

Research papers based on the chapter:

Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, and P. K. Goel. 2004. Coupling STICS crop growth model with remote sensing data for corn growth simulation in Eastern Canada. (under preparation).

CHAPTER 5

MODELING CORN GROWTH IN SOUTHWESTERN QUEBEC WITH THE STICS MODEL

5.1 Abstract

Crop growth models have shown promise in simulating crop biomass and in yield prediction on a field scale. However, their applications have remained restricted due to their huge data requirements. Considering the capability of remote sensing in data collection, the main objectives of the study were to: (i) evaluate the performance of STICS, a crop growth model, for Eastern Canada, (ii) to estimate LAI values from hyperspectral observations, and (iii) to incorporate the LAI values thus determined into the STICS model.

The data for the model evaluation were obtained from 2000 to 2002 at the Macdonald Campus Farm of McGill University, Ste-Anne-de-Bellevue, Quebec, Canada. Corn was grown under different agricultural practices, including different rates of irrigation and nitrogen application. The model was calibrated using the data collected from the treatments that received the recommended nitrogen fertilizer rate for the region. The remaining data were used for the validation of the model. Both calibration and validation were performed on the basis of measured yield, biomass, and LAI. The calibration and evaluation of the STICS model illustrated that the model has the potential capacity to simulate corn growth in Eastern Canada.

The year 2000 data was used for LAI estimation from airborne hyperspectral observations. Hyperspectral observations were collected using a Compact Airborne Spectrographic Imager (CASI) sensor, in 72 narrow wavebands from 407-949 nm. Observations were made three times over the growing season: 30, 66 and 86 days after sowing, corresponding to early growth, tasseling and fully-mature stages, respectively. Crop growth parameters such as LAI, leaf greenness (SPAD), and plant height were also measured at the same time. A support vector machine algorithm (SVM) was used to develop regression models for LAI estimation from hyperspectral data. Using a leave-

one-out method for cross-validation the capacity of the SVM-based model to estimate LAI was investigated.

Crop yield and biomass were better simulated than under the standard application of the model, which confirmed previous studies.

5.2 Introduction

The growing demand for greater agricultural production makes agriculture a highly strategic domain. In this regard, availability of reliable and accessible information on crop production estimates, early in the growing season, is an important issue. Two main approaches to generating such information are statistical methods and crop growth models. With statistical methods (Makowski *et al.*, 2001), measured crop yield is related to one or more parameters. The relationship obtained is used to predict the yield in upcoming years. The simplicity of these methods is their main advantage. However, the disadvantage of these methods is that they do not consider variations in growth parameters over the entire season. Crop growth models, on the other hand, simulate crop growth by defining relationships between various crop growth characteristics, weather parameters, soil properties, and nutrient availability.

Crop growth models simulate crop growth and development in terms of daily dry matter production, thereby allowing them to make simulations at any time during the growth season, as well as the final yield at harvest. Earlier versions of crop models usually simulated crop growth and yield under homogeneous field conditions. However, recently developed models are able to simulate plant growth in non-homogeneous areas (Irmak *et al.*, 2000).

The use of these models in crop growth monitoring is limited due to enormous data requirements of models in terms of soil conditions and agricultural practices over the entire growing season. With the existing methods, data collection is labor-intensive and time-consuming. In this regard, with remote sensing techniques large quantities of such information can be easily and more frequently collected for large areas (Erickson, 1984; Atkinson and Tatnall, 1997). However, the potential of remote sensing in crop production estimation is not fully investigated (Moulin *et al.*, 1998). For a more

quantitative analysis, more advanced approaches are required to estimate yields, including the incorporation of RS observations into crop growth models (Wiegand *et al.*, 1986). Many researchers have tried to link remote sensing systems with various crop growth models. These include SUCROS (Clevers and Leeuwen, 1996; Clevers, 1997; Guérif and Duke, 1998), SWAP (Michaud *et al.*, 1998), CROPGRO-Soybean (Irmak *et al.*, 2000), ROTASK (Clevers *et al.*, 2000) and STICS (Prévot *et al.*, 2000a,b). Amongst these models STICS stands outs for its flexibility (Brisson *et al.*, 1998). Moreover, in the STICS model the observed data can be used directly (such as LAI, an important parameter reflecting ground biomass cover) as model input. Its adaptability provides an added advantage in linking STICS with remote sensing observations.

Reflectance from an object or a target such as a crop canopy is the consequence of numerous factors, which adds difficulty to the remote sensing technique's differentiation and quantification of objects. To conquer this complexity, more comprehensive measurements are needed, which could be acquired with the use of hyperspectral sensors that offer data acquisition in much narrower wavebands, and which appear to be a more effective tool for data collection than multispectral sensors. Given the complexities of this recent technology, the present research was conceived with a more controlled and defined experimental setup. In this context, first the STICS model was calibrated and validated for the project environment. Next, LAI values were estimated from hyperspectral remote sensing observations using a support vector machine approach. Lastly, the coupling of STICS with hyperspectral RS was investigated by way of using RS-derived LAI values in the STICS model. This investigation was carried out under conditions, reflecting different nitrogen application rates and different water regimes. The project sought to provide a tool for efficient management of agricultural fields and better simulation of crop biomass and yield.

In order to test the performance of the model, it was essential to compare predicted values with observed data. A widely used method for this comparison is the correlation-regression analysis (Teo *et al.*, 1992; Chapman *et al.*, 1993; Mayer *et al.* 1994; Retta *et al.*, 1996; Kiniry *et al.*, 1997; Kobayashi and Salam, 2000), wherein the correlation between estimated and measured data is determined by applying a *t*-test to evaluate prediction accuracy (Teo *et al.*, 1992; Kiniry *et al.*, 1997; Kobayashi and Salam,

2000). The regression analysis can determine the measure of relationship between the estimated and observed data. It can also provide a simultaneous test for a slope value of one and intercept value of zero, which are the values in the case of full agreement between simulated and measured data. For this evaluation, null and alternative hypotheses can be set as:

$$H_0 : Slope = 1 \; and \; Y - Intercept = 0 \qquad 5.1$$

$$H_1 : Slope \neq 1 \; or \; Y - Intercept \neq 0 \qquad 5.2$$

Furthermore, simulated and measured data can be directly evaluated by using various methods that deal with the differences between simulated and observed data. Root mean square error (RMSE), relative RMSE (RRMSE), mean square deviation (MSD), and mean deviation (MD) are some examples of these methods (Retta *et al.*, 1996; Kiniry *et al.*, 1997; Kobayashi and Salam, 2000).

The overall goals of this study were to: (i) to evaluate the performance of STICS, a crop growth model, for monitoring corn growth under different growth conditions in southwestern Quebec, (ii) to estimate crop parameters such as LAI using remote sensing data, and (iii) to link remote sensing data with the STICS model.

5.3 Materials and Methods

5.3.1 Study site and experimental details

Experiments were conducted at the Research Farm of Macdonald Campus, McGill University in Ste-Anne-de-Bellevue, Quebec, Canada from 2000-2002 growing seasons. Corn, the main crop of the region, was selected and grown under different agricultural practices, including different weed control regimes, and different rates of irrigation and nitrogen fertilizer application. Corn was planted at a 0.75 m row spacing and at a density of 76,000 plants ha^{-1}.

Experimental details for the years 2000 and 2002, are presented in sections 3.3.1 and 4.3.1, respectively. In 2001, a field experiment was carried out at the Lods Agronomy Research Center of Macdonald Campus, McGill University, Ste-Anne-de-

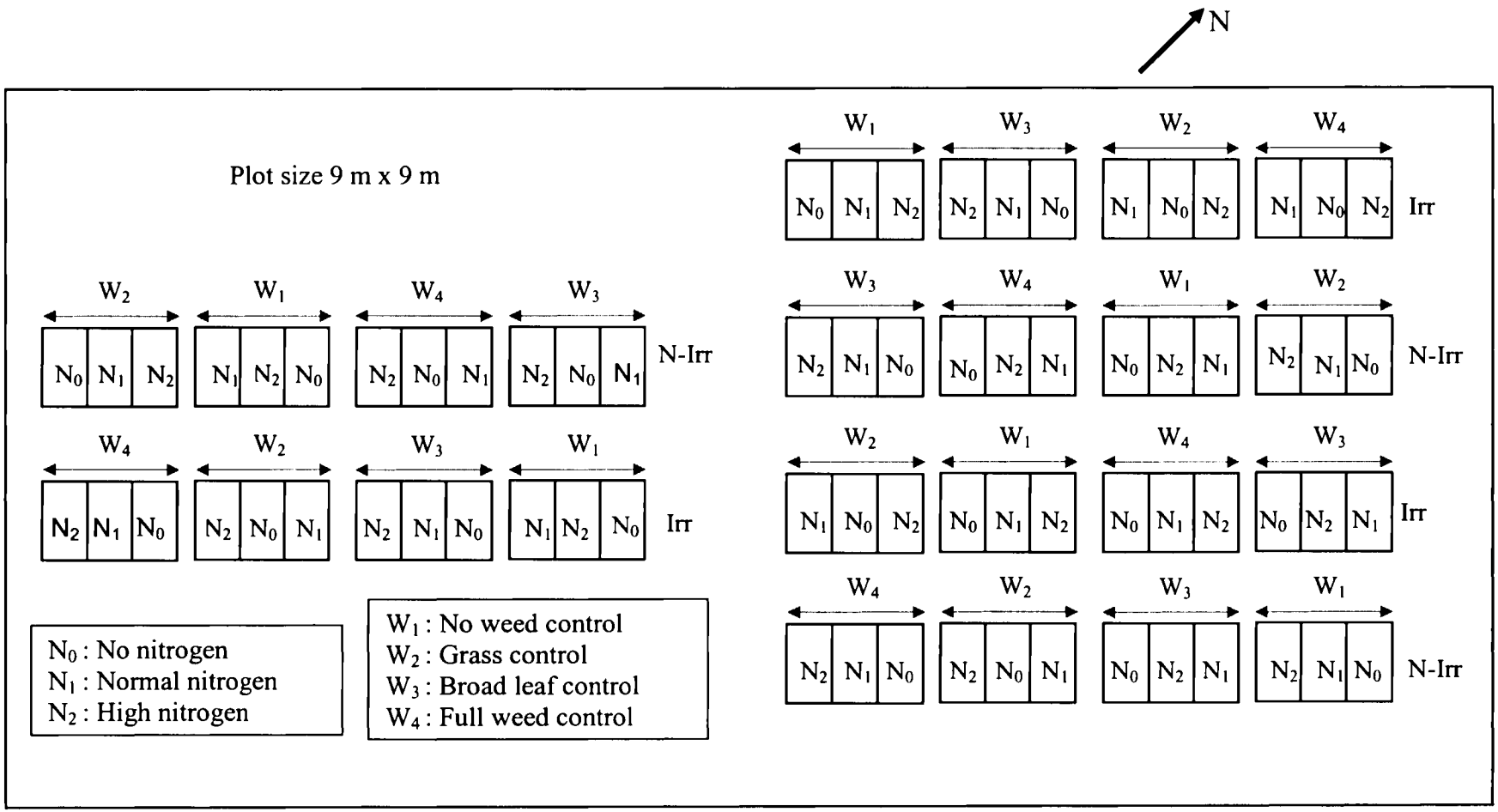


Figure 5.1 Experimental layout, 2001 (Irr, N-Irr are irrigated and non-irrigated treatments, respectively).

Bellevue, Québec, Canada (45°25′45"N lat., 73°56′00"W long.). A schematic diagram of the experimental design is shown in Figure 5.1. Corn plots were arranged in a split-split plot layout, with two irrigation water regimes (with and without), three nitrogen application rates (60, 120, and 250 kg N ha^{-1}), and four weed controls regimes (no control, control of grasses, control of broad-leafs, control of both) as main, sub, and sub-sub-plots, respectively.

5.3.2 Crop parameters

Measurements of various crop canopy and other parameters, including plant height, leaf greenness, LAI, biomass and soil moisture, were taken during the entire growth period. Crop yield and final biomass were measured at season's end. The LAI (m^2 foliage area per m^2 ground area) was measured with a LAI-2000 Plant Canopy Analyzer (Li-Cor, Inc., Lincoln, Nebraska, USA). Biomass was estimated on the basis of five to ten plants per plot that were harvested and weighed. Finally, ten representative plants from each plot were harvested and used to estimate crop yield and final biomass.

5.3.3 Methodology

Different components of the STICS model, describing various mechanisms, were first calibrated, and then validated. Furthermore, considering the ability of the STICS model to use LAI data as input, the LAI values were estimated from aerial remote sensing data, using support vector machine (SVM), a new innovative data mining tool (Vapnik 1995; Burges 1998). Finally, using the estimated LAI values as input data in STICS, crop yield was simulated.

5.3.3.1 STICS model and its modules

STICS simulates daily growth in terms of aboveground biomass (carbon and nitrogen) of crop, and the LAI on the basis of soil parameters and climatic data over a period ranging from emergence to harvest time. In STICS, evaporation can be calculated by the Priestly-Taylor (PT), Penman (PE), or stomatal resistance model (SW). The soil profile is divided into a maximum of five homogenous layers. For simulation purposes, each layer is defined by its water and nitrogen content. Moreover, the soil properties included for each layer are thickness, bulk density, field capacity, and wilting point.

Interaction between plant and soil is done by roots, which are defined by their length distribution in the soil profile. The crop growth simulation is performed on the basis of carbon and nitrogen balance. As a result, agricultural production (biomass, yield) can be computed, and environmental impact (water and nitrogen leaching) of different agricultural conditions can be assessed (Brisson *et al.*, 1998).

The crop growth period is divided into seven stages for the entire growing season (Table 5.1). These stage classifications are slightly different from the ones generally used in agronomy. The stages are defined to facilitate simulating LAI and grain filling. Dividing the growth period into various stages makes it possible to simulate different crops or species. In STICS, the simulation starts by determining the seed germination time, which is related to soil temperature at sowing depth. The germination starts when the sum of degree-days meets a defined threshold value for the specified crop.

Table 5.1 Crop development stages in STICS

Fortran notation	**Description**
ILEV	day-of-the-year of occurrence stage LEV (emergence)
IAMF	day-of-the-year of occurrence AMF (greatest acceleration in leaf growth, end of the juvenile phase)
ILAX	day-of-the-year of occurrence LAX (Maximum LAI value)
IDRP	day-of-the-year of occurrence DRP (beginning of grain filling)
ISEN	day-of-the-year of occurrence SEN (beginning of net senescence)
IMAT	day-of-the-year of occurrence MAT (Physiological maturity)
IREC	day-of-the-year of occurrence REC (Harvest)

To estimate the LAI values, growth period is divided into four stages namely: two growth stages, one stability stage and one senescence stage. Since, the model assumes that the radiation is mainly intercepted by the plant canopy, it is important to have an accurate estimation of LAI. For the stability phase, where interception is at its maximum value, accuracy of LAI estimation has less significance. LAI is estimated with a non-linear equation for the period from emergence to the maximum LAI. A constant

value LAI is assumed for the period between maximum LAI and the beginning of senescence. For the senescence period, the LAI is assumed to decrease linearly. One of the important features of STICS is that measured LAI values can be incorporated in the simulation process. The model also offers non-linear relationships of LAI with time, and interpolates LAI on daily basis. This feature is useful when the user has a series of measured LAI that can replace the simulated LAI.

For the calculation of yield components, a period of variable duration is set to calculate grain number, based on the mean growth rate of the crop canopy during this period. The total dry matter and the nitrogen content of the grain are related to the above ground biomass by some harvest indices (nitrogen and carbon harvest indices). These indices are used to calculate the magnitude of grains and the quantity of nitrogen in grains.

In STICS, root growth is defined in term of its length. Based on the radiation use efficiency, root biomass is calculated from the above ground foliage, and in turn, the root length is estimated from root biomass. The root front development depends on the crop variety, crop temperature, soil depth, and water availability. In the model, the active root front rate is related to soil water conditions and crop temperature.

The crop growth is simulated in the model using seven different modules, namely: development, shoot growth, yield component, root growth, water balance, thermal environment and nitrogen balance. For each module, a number of input variables and parameters are required. While the values of certain parameters are fixed for a particular condition, often variable values (e.g., climatic data) change with time. In addition, there are some structural parameters, which are different from the input parameters, and are independent from the soil, the plant, and the crop management practices. The variables are further grouped as, external, internal driving, and output variables, which are based on climatic data, come from other modules, and are calculated within a module, respectively. A detailed description of STICS is provided in appendix A.

5.3.3.1.1 Inputs and outputs

In the STICS model, all inputs and outputs are arranged in different files. The first input file contains climatic data, which include date, climatic station information,

minimum and maximum temperatures, radiation, and rainfall. A monthly summary of climatic data is provided in Table 5.2. The second input file contains parameters relating to cultural practices. It includes all the data regarding plant species, seeding information, growth stages, fertilization, irrigation, tillage and input of organic residues. Table 5.3 provides necessary information on cultural practices variables for the growth period. The third input file specifies the permanent characteristics of the soil, which are required for the simulation, including texture, soil depth, field capacity, and permanently wilting point of different layers (to a maximum of five).

Some of the models outputs for crop growth are: nitrogen content of the plant and soil, soil moisture, LAI, biomass, yield and grain number, which are calculated on a daily basis. These parameters are formatted and split up into different output files along with other output parameters. A sample set of input and output data files are given in a CD.

5.3.3.1.2 Calibration of the model

The model calibration was performed by adjusting different parameters based on data from the treatments that received 120 - 150 kg N ha^{-1} (the recommended application rate of the region) during the simulation period. Crop growth stages defined in the model depend on agronomical growth stages and the variety of crop. For corn, they are "silking", "onset of grain" and "physiological maturity". These growth stages are redefined in the STICS model, as shown in Table 5.1 (Brisson *et al.*, 1998, 2002). In addition, LAI, biomass and grain yield parameters, which describe the development of the biomass (e.g., optimum temperature, conversion efficiency factor) and grain yield (e.g., maximum and minimum grain numbers), were adjusted to match the observed yield (Table 5.4), along with the range of the parameters that were used for testing. The adjustment procedure was performed by a trial and error method. In each trial, an effort was made to optimize one parameter, assuming that other parameters values remained constant. In this way, the optimum combination of the parameters was obtained.

5.3.3.1.3 Evaluation of the Model Performance

For the evaluation of the STICS model, the yield, biomass, and simulated LAI data used for model calibration were not used. The performance of the model was

Table 5.2 Mean monthly temperature, relative humidity, and total precipitation for the growing season during the simulation period and for a normal year (based on Ste Anne station)

			May	June	July	August	September	October
Year 2000	Precipitation (mm)		132.3	87.0	79.2	123.9	82.5	26.1
	Radiation (MJ m^{-2})		528.1	574.1	618.8	525.5	404.1	247.3
	Temperature (C°)	Min.	7.6	11.2	14.4	14.2	8.2	3.9
		Max.	17.9	21.4	24.1	23.7	18.7	13.0
		Mean	12.7	16.3	19.3	18.9	13.4	8.5
Year 2001	Precipitation		70.5	57.7	36.6	56.9	78.3	77.9
	Radiation (MJ m^{-2})		582.0	617.5	624.0	575.8	416.9	N/A
	Temperature (C°)	Min.	9.3	13.7	14.1	14.7	10.1	6.0
		Max.	20.2	24.8	24.5	27.8	21.6	14.8
		Mean	14.7	19.3	19.3	21.2	15.8	10.4
Year 2002	Precipitation		146.2	68.6	31.9	8.4	64.5	64.6
	Radiation (MJ m^{-2})		623.0	723.1	920.7	710.6	556.7	366.6
	Temperature (C°)	Min.	6.9	12.4	17.0	16.0	12.5	3.3
		Max.	15.2	22.0	26.8	26.6	23.5	11.3
		Mean	11.1	17.2	21.9	21.3	18.0	7.3
30-average (1961-1990)	Precipitation		74.8	68.3	82.5	85.6	100.3	86.5
	Radiation (MJ m^{-2})		N/A	N/A	N/A	N/A	N/A	N/A
	Temperature (C^0)	Min.	0.6	7.3	12.5	15.4	14.1	9.3
		Max.	10.7	18.5	23.4	26.2	24.6	19.8
		Mean	5.7	12.9	18.0	20.8	19.4	14.6

Table 5.3 Details on cultural practices

Year	Operation	Date	Specific details
2000	Sowing	May 30	76000 seeds ha^{-1} 0.75 m row spacing
	Harvest	October 10	
	Fertilization	a. May 30 b. June 01 c. July 12	10-120-50, in all plots ($N-P_2O_5-K_2O$) (kg ha^{-1}) 0 in N_{60}; 10 in N_{120}; 90 in N_{250} (N kg ha^{-1}) 50 in N_{60}; 100 in N_{120}; 150 in N_{250} (N kg ha^{-1})
2001	Sowing	May 10	76000 seeds ha^{-1} 0.75 m row spacing
	Harvest	October 12	
	Fertilization	a. May 10 b. May 10 c. July 03	35-80, in all plots ($N-P_2O5$) (kg ha^{-1}) 0 in N_{60}; 10 in N_{120}; 90 in N_{250} (N kg ha^{-1}) 50 in N_{60}; 100 in N_{120}; 150 in N_{250} (N kg ha^{-1})
	Irrigation	a. July 25 b. August 03	40 mm 50 mm
2002	Sowing	May 22	76000 seeds ha^{-1} 0.75 m row spacing
	Harvest	October 08	
	Fertilization	a. May 22 b. June 01 c. July 12	10-120-50, in all plots ($N-P_2O_5-K_2O$) (kg ha^{-1}) 0, 5, 25, 45, and 65 (N kg/ha) in N_{50}, N_{100}, N_{150}, N_{200}, and N_{250} treatments, respectively 15, 60, 90, 120, and 150 (N kg/ha) in N_{50}, N_{100}, N_{150}, N_{200}, and N_{250} Treatments, respectively
	Irrigation	a. July 19 b. July 29 c. August 01 d. August 04 e. August 08 f. August 12 g. August 16 h. August 21	28 mm 16 mm 18 mm 16 mm 14 mm 16 mm 16 mm 18 mm

Table 5.4 Adjusted model parameters

Parameter	Description	Evaluation Range	Value used
stlevamf	juvenile stage duration (degree-days)	200-450	300
stamflax	duration between the stages AMF and LAX (degree-days)	300-550	450
dlaimax	maximum growth rate of LAI	0.0011-0.002	0.0017
efcroiveg	conversion efficiency (before flowering)	3.4-6.0	5.5
efcroirepro	conversion efficiency (after flowering)	3.0-5.0	4.6
cgrainv0	minimum grain number (grains m^{-2})	100-500	300
pmaxigrain	maximum weight of one grain (% water content)	0.250-0.450	0.300
teopt	Optimal temperature for the biomass growth (^{0}C)	20-30	22.0

evaluated on the basis of regressions between observed and simulated data. Regression analyses were performed for these data, and the intercepts and slopes of regression lines were determined and compared with their ideal values of 0 and 1, respectively. Several statistical methods (Addiscott and Whitmore, 1987; Brisson *et al.*, 2002) were used for a proper evaluation (Table 5.5). For applying these methods, the analysis was done using IRENE (Integrated Resources for Evaluating Numerical Estimates), a statistical software (Fila *et al.*, 2003). From linear regression, bias in the data can be obtained from the intercept and slope; on the other hand, the model's dependability can be calculated with the coefficient of determination (r^2) (Brisson *et al.*, 2002).

Modeling Efficiency (EF), which is considered an indicator of the overall correspondence between the measured and estimated values, explains the goodness of fit of the predicted values with observed values. In the case of biased data, if the data are strongly correlated, EF is supposed to do a better job than $\mathbf{r}^2$. A negative value of EF shows that model prediction to be very poor. Furthermore, RMSE can be used as an indicator of the mean distance between measured and estimated data (Kobayashi and Salam, 2000). Since the numerical value of RMSE is subjected to overall observed data values (mean and the range) and the units of measurement it is advantageous to use

RRMSE over RMSE (Bellocchi *et al.*, 2002). The value of the RRMSE can vary from 0 to infinity. A lower numerical value of RRMSE indicates a better performance of the model. However, it should be noted that RMSE and mean bias error (MBE) display various features of the overall deviation between estimated and measured data. The optimal values of all indices, indicating a full agreement between estimated and measured values are given in Table 5.5.

Table 5.5 Statistical criteria used in evaluation of model performance

Statistical criteria	Optimal value	Definition
Least square regression	1, 0, 1	Slope, Intercept, r^2 = coefficient of determination
Root mean square error	0	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - o_i)^2}$
Relative RMSE (%)	0	$RRMSE = \frac{RMSE}{\overline{O}}$
Mean bias error	0	$MBE = \sum_{i=1}^{n}\frac{E_i - O_i}{n}$
Model efficiency	1	$EF = 1 - \frac{\sum_{i=1}^{n}(E_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$
Modified model efficiency	1	$EF_1 = 1 - \frac{\sum_{i=1}^{n}\lvert E_i - O_i\rvert}{\sum_{i=1}^{n}\lvert O_i - \overline{O}\rvert}$

n: being the number of observation, **O:** observed value, $\overline{O}$ **:** mean of the observed values
E: estimated value by model

5.3.3.2 Estimation of LAI using RS data

The estimation of LAI was accomplished using experimental data collected in 2000. For estimation of LAI from hyperspectral aerial observations, the acquired aerial images were first imported into ENVI 3.5 (Environment for Visualizing Images) software,

which is an image processing system designed to provide comprehensive analysis of airborne and satellite remote sensing data. The mean reflectance values of each experimental unit/plot were determined. Then, for determining the relationship between the reflectance data and LAI values, a SVM method was used. A brief explanation of the SVM approach follows.

5.3.3.2.1 Support vector machine

The SVM algorithm is a machine learning technique based on statistical theory (Vapnik, 1995) that can be used for both classification and regression purposes. In a classification problem, the aim is to train a SVM model to develop a function to be used for separating treatments into different classes using unseen data. The best classifier would be the optimum function that can maximize its distances with nearest neighbor points in each class (the optimal separating hyperplane). The SVMs have emerged as a popular technique to perform data mining tasks such as classification, regression and automated detection of outliers. In this section, a brief overview of the theory of SVMs, and an explanation of the motivation for their use in this study is given (as compared to other classification techniques). For more detail on the mathematical concepts readers are referred to Burges (1998) and Vapnik (1995).

5.3.3.2.2 Theory of SVMs

SVMs can be considered as "maximal margin classifiers." For a given training sample belonging to two different classes, SVMs derives a classification function, (a hyper-plane) which is at a maximum distance from the closest points belonging to both the classes. These closest points are called as the "support vectors". It is solely the support vectors (and not the other training samples) that affect the actual decision function. According to Vapnik's theory on statistical learning (Vapnik, 1995), the probability of classification errors from decision functions thus obtained are dependent only on the number of support vectors and the number of training samples, but not on the dimensionality of the training samples. Such a decision function is called an "optimal separating hyper-plane" (OSH).

Assuming that the two classes to be distinguished are linearly separable, and denoting the training samples as x_1 to x_M, where M is the number of training samples. Let each sample be labeled by $y = \{+1,-1\}$ depending upon which class it belongs to. The equation of the separating hyper-plane can be written as follows:

$$f(x) = \sum_{i=1}^{M} \alpha_i y_i (x_i \bullet x) + b \qquad 5.3$$

The coefficients α_i and b are obtained upon solving a quadratic programming problem [which is internally accomplished by the SVM package (Chang and Lin, 2004)]. It is interesting to note that all training examples apart from the support vectors have alpha values equal to zero (and thus do not affect the equation of the OSH). An example under testing is classified by finding its signed distance to the OSH. The farther the sample is from the OSH, the more the reliability of the classification result.

The discussion so far has been restricted to binary classification in those cases where the data are linearly separable. For training samples that are not linearly separable, the data have to be transferred onto a space of higher dimensionality (called the "feature space") where a reliable linear separation can be computed. The additional attributes to map the data onto the feature-space have to be non-linear functions of the already existing attributes. Let us denote such a mapping onto feature-space by Φ, such that $z = \Phi(x)$ is the feature-point corresponding to a data item x. The equation of the OSH in feature-space can then be written as follows:

$$f(x) = \sum_{i=1}^{M} \alpha_i y_i (\Phi(x_i) \bullet \Phi(x)) + b \qquad 5.4$$

However, computing the mapping Φ is often not feasible in practice. The way to circumvent this difficulty is to employ a so-called "kernel function" $K(x, y)$ which is equal to the dot product of $\Phi(x)$ and $\Phi(y)$. Substituting the dot product in the above equation by the kernel, the OSH becomes:

$$f(x) = \sum_{i=1}^{M} \alpha_i y_i K(x_i, x) + b \qquad 5.5$$

While designing a non-linear SVM classifier, the user has to choose an appropriate kernel. The most commonly used kernels are the radial basis function kernel (RBF), the sigmoid kernel and the polynomial kernel. (the RBF kernels given as $K(x, y) = e^{-\gamma(x-y)^2}$ are known to be the most popular). Furthermore the user also has to choose the appropriate parameter such as gamma for the RBF kernel, and a parameter C which is essentially a tradeoff explaining how severely classification errors are to be penalized. The quality of the classification can be greatly affected by C, in that a very large value to this parameter can overfit the data.

The above binary classification scheme can be easily extended for K classes. Such schemes follow two approaches: one-against-all and one-against-one. In the former approach, K SVMs are trained with examples of one particular class being trained against those of all other classes. In the latter, $K(K-1)/2$ SVMs are created in a binary tree-like fashion. While the former strategy is positively more efficient in terms of training time, the latter strategy has been empirically shown to yield more desirable results (Hsu and Lin, 2002).

SVM was also extended to regression problems by introducing a loss function. Four typical loss functions that could be used are: Quadratic, Laplacian, Huber, and ε-insensitive (Gunn, 1998). Quadratic loss function's principles are similar to those of the conventional least square error. The quadratic loss function is more sensitive to outliers than Laplacian. The Huber loss function is assumed to be robust loss function for the case of data with an unidentified distribution. Gunn (1998) reported that the three loss functions, quadratic, Laplacian, and Huber could eliminate the sparseness in the support vector. The ε-insensitive loss function was introduced by Vepnik (1995) to explain the sparse set of the support vector machine that is an approximation of Huber's loss function. In the present study, SVM regression models were developed based on the ε-insensitive loss function. To find the optimal epsilon (ε) value, different SVM models were trained using various epsilon values. Better estimations were generally obtained with lower epsilon values.

The SVM is trained with a set of data containing both independent and dependent variables. Therefore, SVM was trained using a portion of the data set (75%) containing

the dependent and independent variables. The remaining 25% of data was used for testing purposes. Considering the smaller size of our data set, the evaluation of the model was performed using leave-one-out cross-validation. For this procedure the existing data set is split into n pair of separate sets. Using different combinations of n-1 sets, SVM models are trained, and tested with the remaining set.

5.3.3.3 Coupling STICS with remote sensing

The STICS model can be linked to remote sensing data by direct use of LAI values estimated from RS observations. Most of the crop models need continuous LAI values (e.g. daily base) for the entire simulation period. However, the availability of the remote sensing data is limited to a few measurement dates, so estimated LAI values must be interpolated/extrapolated for the simulation period.

Since the remote sensing data in the year 2000 was available for three dates only during the crop growing season, the SVM method was used to estimate LAI values from hyperspectral data. Next, the STICS model was run for the nitrogen treatment in weed-free plots and the simulated LAI compared with the corresponding LAI values estimated from the remote sensing data for the three dates. The LAI parameters were adjusted so that the simulated and estimated (from RS) were close to each other. The applicability and accuracy of this approach was evaluated by comparing the simulated results with the observed yield and biomass data. Such a comparison was done for different crop management conditions to evaluate the procedure.

5.4 Results and Discussion

The results of the analyses, aimed at calibration and evaluation of STICS model performance, are presented and discussed in section 5.4.1. Discussions on the estimation of the LAI values using hyperspectral observations are given in section 5.4.2. Finally, in the section 5.4.3 a discussion on the use of estimated LAI values on the STICS model is provided.

5.4.1 Calibration and Evaluation of STICS Model

The analysis was performed using data collected from different nitrogen and water treatments (weed-free) during the years 2000 to 2002. The simulated and the mean measured treatment crop yields, total aboveground biomass, and LAI are shown in Figures 5.2, 5.3, and 5.4, respectively. Generally, for all three parameters, the simulated values are in close agreement with measured values.

The basic statistics for model validation are given in Table 5.6. It is evident from the table that the values of all parameters for the simulated data are comparable to the ones for the observed data. The results of the regression analysis are given in Table 5.7. The correlation between observed and simulated yield was 0.95, which indicated that the model established a very good relationship between observed and simulated yield. The following regression equation was derived:

$$Y = 0.92\ X + 0.82 \qquad 5.6$$

where Y and X were the simulated and mean measured yield values (Mg ha^{-1}), respectively. The regression parameters, intercept and slope were not significantly different from their ideal values of 0 and 1 ($P \leq 0.05$), respectively, which clearly indicated that the model was able to simulate crop yield very well. The MBE was low which implied that the model did not under- or over-predict the yield values. The RMSE and RRMSE of yield were 0.85 (Mg ha^{-1}), and 11%, respectively. These values are also considered to be low. The EF was calculated to be 0.89, which, once again confirms a good model performance.

The basic statistics of biomass for model validation are also given in Table 5.6. Similar to yield, for biomass the parameters obtained from the simulated data are comparable to those obtained from the observed data. The correlation between observed and simulated biomass was 0.96. The following regression equation was derived:

$$Y = 1.06\ X - 0.65 \qquad 5.7$$

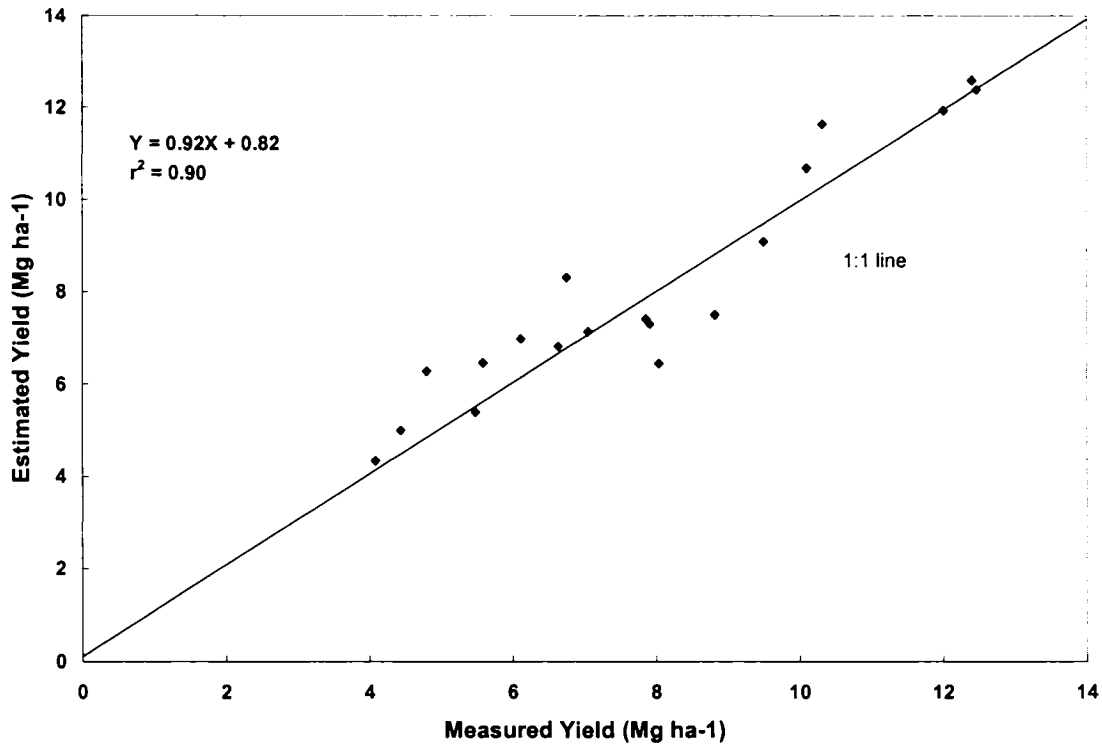


Figure 5.2 Comparison of measured and estimated yield (for model evaluation).

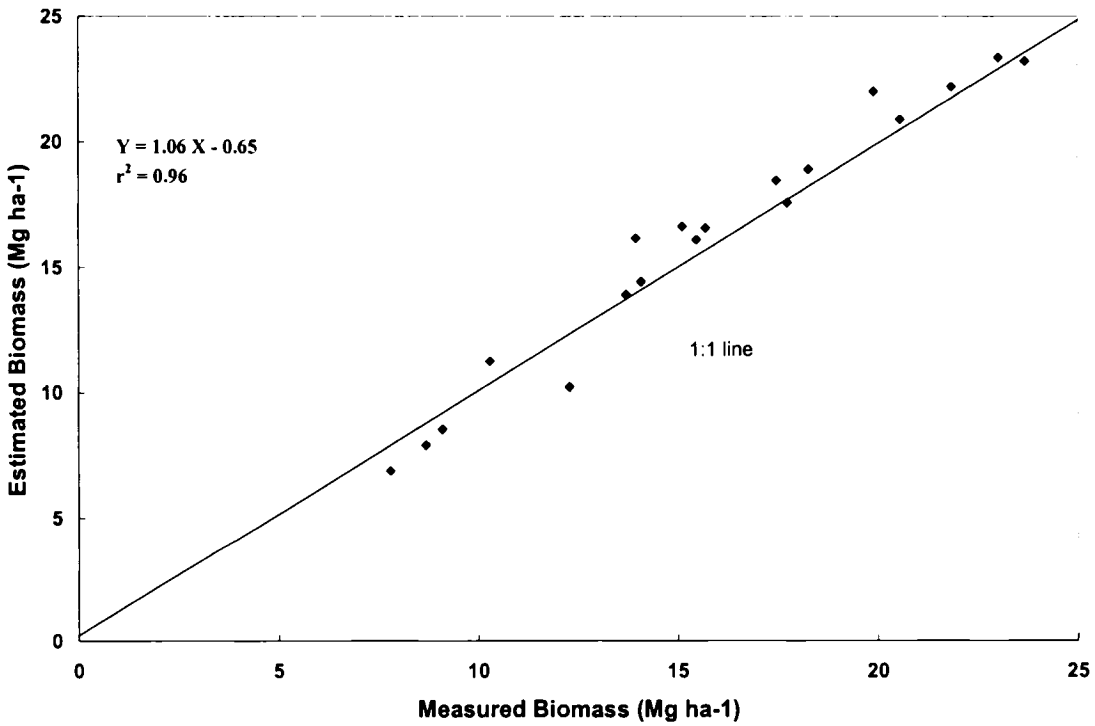


Figure 5.3 Comparison of measured and estimated biomass (for model evaluation).

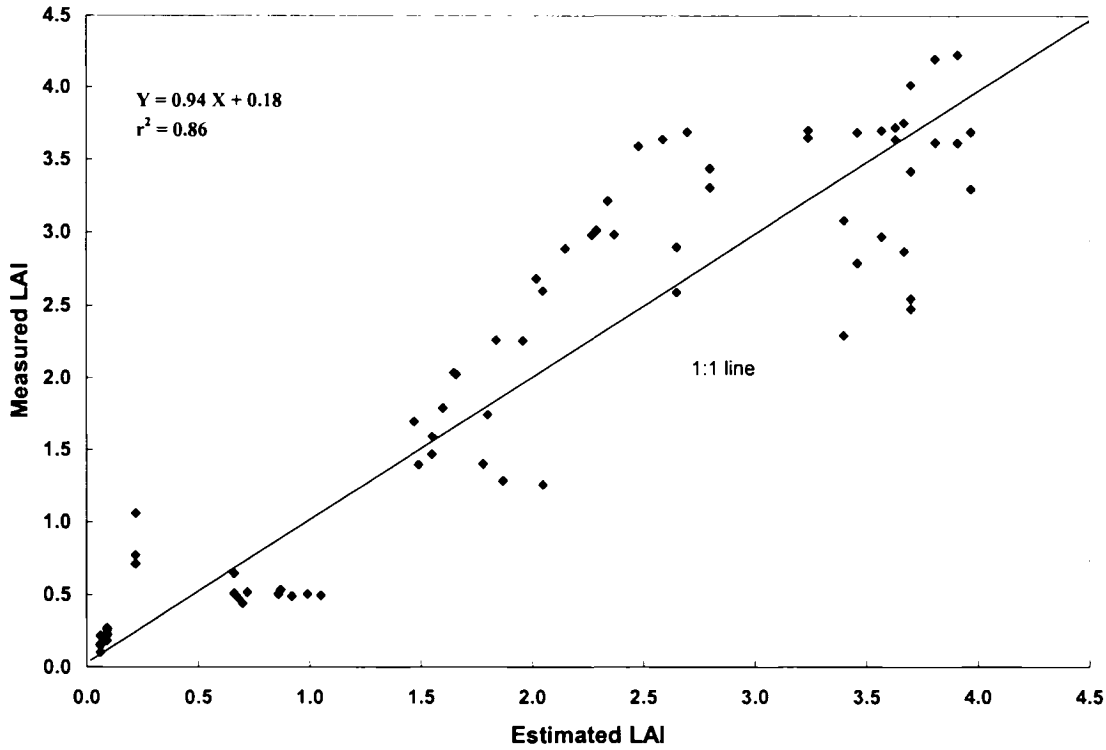


Figure 5.4 Comparison of measured and estimated LAI (for model evaluation).

Table 5.6 Basic statistics for estimated and measured corn crop parameters

Statistic	Yield*		Biomass*		LAI*	
	Estimated	**Measured**	**Estimated**	**Measured**	**Estimated**	**Measured**
Number of Data	19	19	19	19	72	72
Minimum	4.34	4.09	6.89	7.80	0.06	0.11
Maximum	12.59	12.47	23.34	23.69	3.97	4.23
Mean	8.09	7.91	16.06	15.72	2.03	2.09
Median	7.3	7.86	16.56	15.48	2.04	2.28
Standard Deviation	2.567	2.648	5.207	4.811	1.333	1.352
Standard Error	0.589	0.608	1.195	1.104	0.157	0.159
Coefficient of variation (CV)	0.317	0.335	0.324	0.306	0.657	0.646

* yield and biomass are in Mg ha^{-1}, LAI is unitless

Table 5.7 Statistical comparison of measured and estimated corn crop parameter values

Index *	Yield	Biomass	LAI
RMSE	0.850	1.068	0.514
RRMSE	0.107	0.068	0.246
MBE	0.172	0.333	-0.059
EF	0.891	0.948	0.853
EF1	0.691	0.778	0.602
R	0.946	0.981	0.927
intercept	0.82	-0.65	0.18
slope	0.92	1.06	0.94

* **RMSE**= Root mean square error (Mg ha $^{-1}$), **RRMSE** = Relative RMSE, **MBE** = Mean bias error, (Mg ha $^{-1}$) **EF** = Model efficiency, **EF1** = Modified EF

where Y is the simulated biomass and X is the mean measured biomass values (Mg ha^{-1}). Once again, the intercept and slope were not significantly different from their ideal values of 0 and 1 ($P \leq 0.05$), respectively. The RMSE and RRMSE of biomass were 1.07 (Mg ha^{-1}) and 7%, respectively. The EF also showed a high value of 0.95. All these parameters testify to good model performance. The MBE was 0.33, which could be considered to be low, given the range of biomass values from 7.8 to 23.69 (Mg ha^{-1}). Thus, the model neither under- nor over-predicted biomass.

The performance of the model in the simulation of LAI appears to be good given comparable statistics for simulated and observed LAI data (Table 5.6). The correlation between the simulated and observed LAI was quite satisfactory (r=0.93), however, it is lower than those for yield and biomass (Table 5.7, Figs 5.1, 5.2, 5.3). The following regression equation for measured and simulated LAIs was derived:

$$Y = 0.940\ X + 0.180 \qquad 5.8$$

where Y is the simulated LAI and X is the mean measured LAI values. The intercept and slope were not significantly different from their ideal values of 0 and 1 ($P \leq 0.05$), respectively, indicating satisfactory model performance. The RMSE of LAI was 0.51,

and RRMSE was 25%, which also suggest that the model performance was good. The EF was 0.85, which further proves that the simulation of LAI values was satisfactory. The closeness of MBE value to its optimal value indicates that the model performed very well, without under- or over-prediction.

5.4.2 Estimation of LAI values

LAI values, estimated from the hyperspectral measurements with SVM models for various epsilon (ε) values, were constructed using 75% of the data for training, and the model performance was tested using the remaining 25% of the data. A summary of the regression analysis of estimated LAI values using the SVM algorithm and the observed LAI is given in Table 5.8 for training and testing data set for the three different flight dates. The results demonstrate that lower ε values generally provide a better estimation. Although, high correlations were obtained for the training data ($r^2 > 0.80$), the same was very low for testing data. Moreover, the performance was poor for the second and third fights for which the best-case r^2 value for the testing data was 0.38.

Ideally, the regression parameters, intercept and slope should be close to 0 and 1, respectively for a perfect match. However, it was observed that in most cases, including training and testing data sets, the intercepts and slopes were significantly different from 0 and 1 ($P \leq 0.05$), respectively. This indicates that the model did not perform that well. This is not surprising given the strict nature of this statistical test. Moreover, it should be noted that poorer performance is not uncommon with smaller data sets, which is true in our case. In the case of smaller data sets, the leave-one-out method can be applied. A summary of the regression parameters for this method is shown in Table 5.9 for different ε values. A comparison of Tables 5.8 and 5.9 indicates that better results were obtained with the leave-one-out method. This also confirms that with more data, SVM can do a better job. In the case of training, the best correlation coefficients were obtained using ε values of 0.01 (first and second flight data set) and 0.10 (third flight data set), however, for testing Data an ε value of 0.01 provided the highest correlation coefficient for all the three flights. Thus, it again appears that the SVM models constructed using lower ε values yielded better results.

For better visualization of the relationship between the observed and estimated LAI, the results for optimal value of ε are plotted for both training and testing cases (Figures 5.5 to 5.7). The data are well distributed around the regression lines. The statistics calculated for the best cases are given in Table 5.10. Although the intercept and slope were significantly different from 0 and 1 ($P \leq 0.05$), it appears that there remains a relationship between the estimated and measured LAI values. In the case of testing data, for the first flight data the correlation was slightly less than the training one, however, in the second and third flights, the correlations for testing data were comparable to those for the training data. This again explains the prediction capability of SVM model. Lower RMSE and RRMSE values also indicate that the model performance could be considered to be acceptable. The lowest and highest RMSE of 0.17 (Mg ha^{-1}) and 0.528 (Mg ha^{-1}) were obtained for the training data set of first flight and second flight data set, respectively. The same was true for RRMSE: 7% for the training set of the third flight and 25% for the testing data of the first flight. These values indicate that the performance difference between the best and the worst models was reasonably low, and the models were reliable.

5.4.3 Coupling STICS model with RS

The STICS simulation results for both yield and biomass, obtained from the model using the default LAI values generated by the model during the process of model validation, and those obtained when the LAI values were derived from RS observations are shown in Table 5.11. Although, STICS performed well with the simulated LAI values, the model performed much better with LAI values estimated from RS observations. It should be noted that the availability of the remote sensing observations was limited to only three flights, and the performance could be improved even further, had there been a more frequent collection of RS data. In the validation processes (without RS information) the STICS model overestimated yield for all three nitrogen application rates by an average of 10.4%. With RS-derived LAI values there was an underestimation for the normal nitrogen application rate and overestimation for low and high nitrogen

Table 5.8 Regression prediction results using SVM software

a) Training data set

ε Value	First Flight			Second Flight			Third Flight		
	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*
0.01	0.807	0.714	0.26	0.862	0.570	1.67	0.737	0.462	2.03
0.02	0.807	0.704	0.27	0.879	0.731	1.02	0.848	0.774	0.86
0.05	0.803	0.692	0.29	0.881	0.727	1.03	0.851	0.759	0.92
0.10	0.778	0.646	0.34	0.880	0.721	1.06	0.845	0.728	1.04
0.20	0.722	0.526	0.46	0.875	0.699	1.13	0.829	0.671	1.30
0.50	0.332	0.184	0.88	0.862	0.570	1.67	0.737	0.462	2.03

b) Testing data set

ε Value	First Flight			Second Flight			Third Flight		
	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*
0.01	0.562	0.467	0.40	0.334	0.189	3.02	0.275	0.240	2.91
0.02	0.557	0.458	0.41	0.303	0.281	2.56	0.333	0.354	2.54
0.05	0.571	0.447	0.43	0.317	0.280	2.56	0.354	0.369	2.49
0.10	0.554	0.420	0.47	0.341	0.279	2.56	0.380	0.382	2.44
0.20	0.533	0.363	0.55	0.382	0.276	2.57	0.373	0.372	2.49
0.50	0.606	0.164	0.86	0.334	0.189	3.02	0.275	0.240	2.91

Table 5.9 Regression prediction results using SVM software with the leave-one-out method

a) Training data set

ε Value	First Flight			Second Flight			Third Flight		
	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*
0.01	0.844	0.759	0.218	0.827	0.592	1.663	0.691	0.490	1.854
0.02	0.843	0.757	0.222	0.813	0.657	1.227	0.791	0.709	1.132
0.05	0.848	0.748	0.241	0.813	0.652	1.252	0.801	0.704	1.147
0.10	0.839	0.715	0.295	0.813	0.639	1.312	0.810	0.689	1.195
0.20	0.760	0.599	0.432	0.818	0.612	1.409	0.807	0.654	1.324
0.50	0.462	0.222	0.859	0.799	0.543	1.830	0.691	0.477	1.907

b) Testing data set

ε Value	First Flight			Second Flight			Third Flight		
	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*	r^2	*Slope*	*Inter*
0.01	0.575	0.587	0.392	0.827	0.592	1.663	0.69	0.490	1.854
0.02	0.507	0.530	0.466	0.258	0.289	2.736	0.356	0.406	2.262
0.05	0.504	0.516	0.485	0.275	0.293	2.722	0.360	0.403	2.273
0.10	0.482	0.482	0.529	0.299	0.298	2.702	0.358	0.392	2.306
0.20	0.383	0.384	0.638	0.314	0.294	2.718	0.371	0.379	2.348
0.50	0.163	0.116	0.966	0.254	0.229	3.083	0.327	0.294	2.606

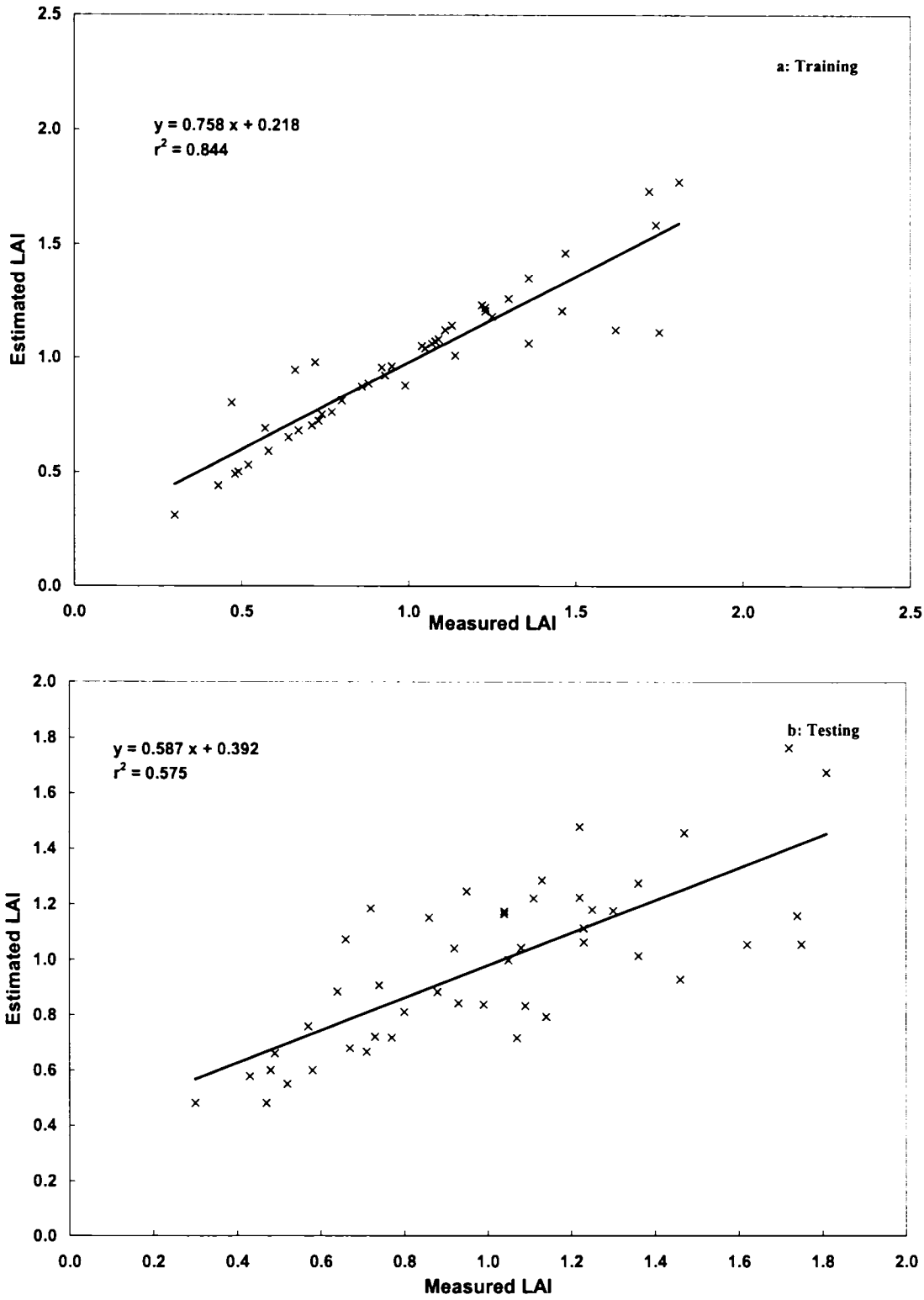


Figure 5.5 Measured and predicted LAI value based on airborne hyperspectral observation at the initial stage (first flight, June 30, 2000).

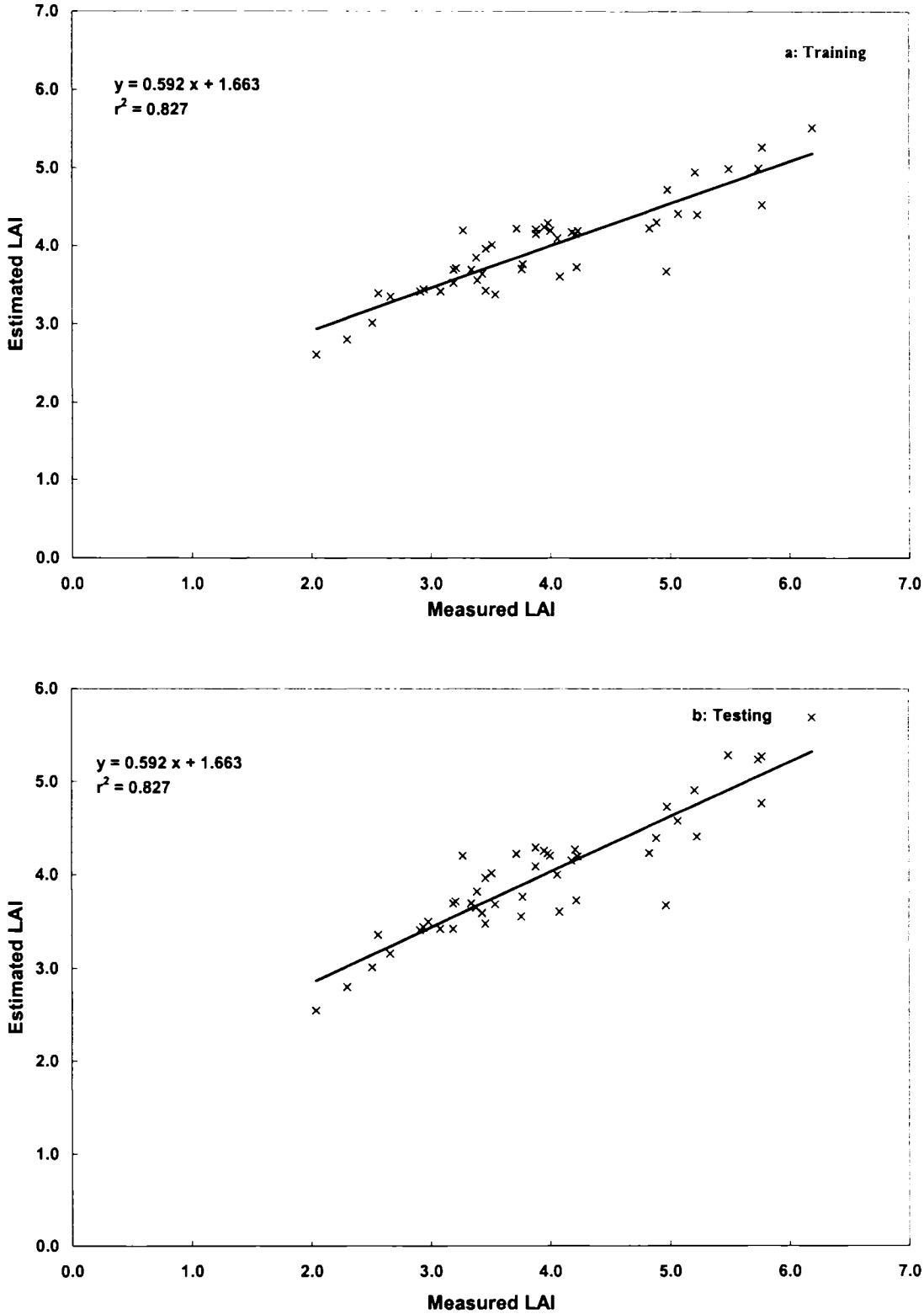


Figure 5.6 Measured and predicted LAI value based on airborne hyperspectral observation at the tasseling stage (second flight, August 05, 2000).

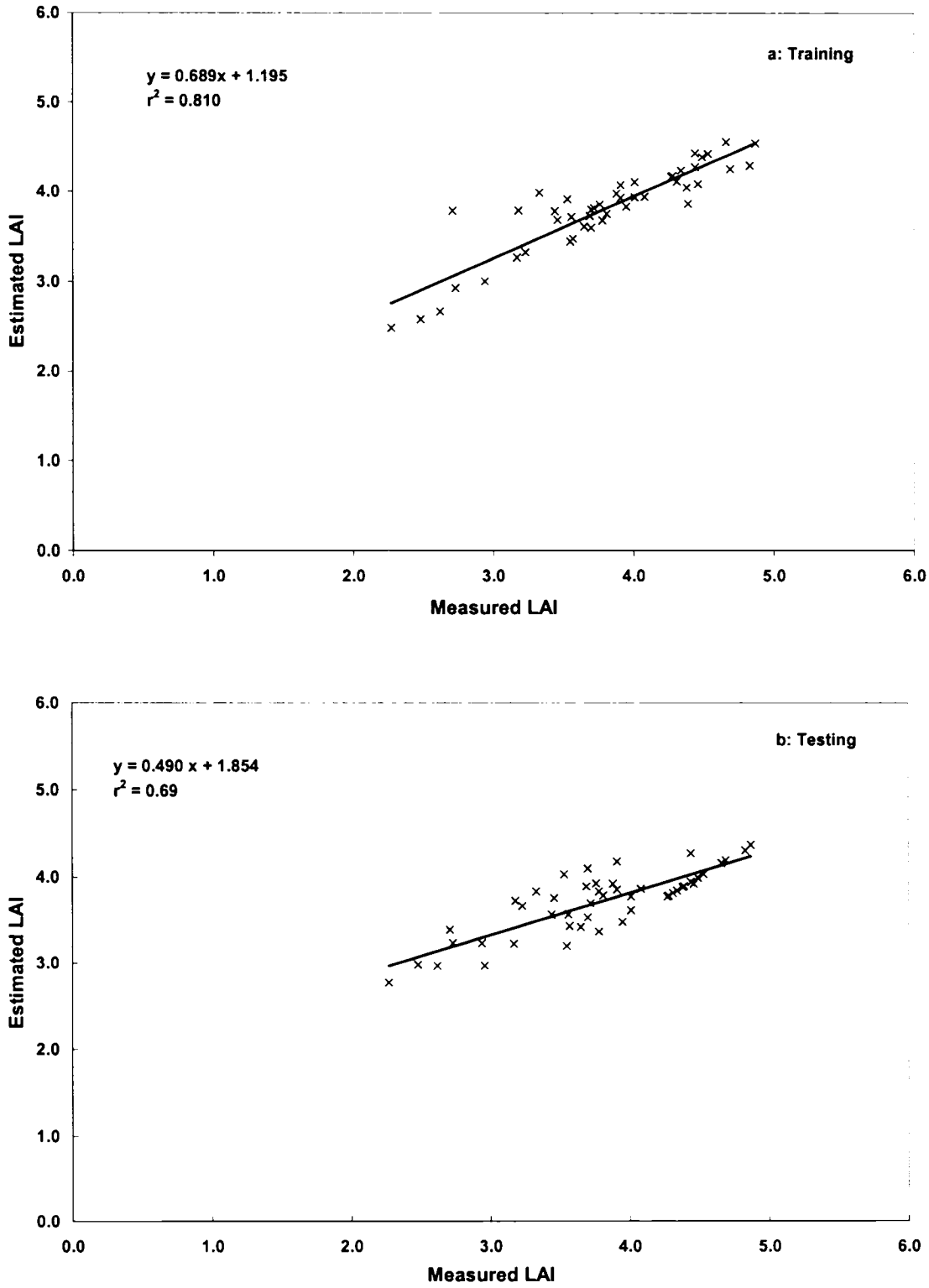


Figure 5.7 Measured and predicted LAI value based on airborne hyperspectral observation at the full maturity stage (third flight, August 25, 2000).

Table 5.10 Statistical comparison of measured and estimated LAI values using the SVM model

Index *	First Flight		Second Flight		Third Flight	
	Test	Train	Test	Train	Test	Train
RMSE	0.249	0.156	0.489	0.528	0.389	0.285
RRMSE	25.4	16.0	12.3	13.3	10.5	7.5
MBE	0.570	0.831	0.755	0.716	0.618	0.014
EF	0.417	0.754	0.478	0.442	0.318	0.791
EF1	-0.024	-0.025	0.068	0.038	-0.075	0.595
r	0.758	0.918	0.910	0.894	0.831	0.900
intercept $^{\varepsilon}$	0.392	0.218	1.663	1.830	1.854	1.195
slope $^{\varepsilon}$	0.587	0.758	0.592	0.543	0.490	0.689

* **RMSE**= Root mean square error (Mg ha $^{-1}$), **RRMSE** = Relative RMSE (%), **MBE** = Mean bias error (Mg ha $^{-1}$), **EF** = Model efficiency, **EF1** = Modified EF,

$^{\varepsilon}$ The parameters are significantly different ($p < 0.05$).

Table 5.11 Summary of measured and simulated (with and without using remote sensing data) yield and biomass for year 2000 data (Mg ha $^{-1}$)

	Yield			Biomass		
	Observed	Validation	Simulated-RS	Observed	Validation	Simulated-RS
Low nitrogen	4.80	6.28	5.23	15.12	16.63	14.29
Normal nitrogen	6.11	6.97	5.92	17.47	18.45	15.59
High nitrogen	6.75	8.31	7.24	19.91	22.00	18.9
Mean	5.89	6.74	6.13	17.50	19.03	16.61

application rates, resulting in an overall of 4.1% overestimation of yield. This indicates an overall 6.3% improvement in yield prediction. In the case of biomass, the evaluation processes resulted in 8.7% overestimation, while in the coupled process the results showed an underestimation of 5 %, which still shows a small improvement in model performance. Thus, it can be concluded that with the coupling of a crop growth model such as STICS with RS observations, better yield prediction could be expected.

5.5 Conclusions

This study demonstrates that the STICS model shows promise for simulating corn growth in southwestern Quebec. The correlation coefficient between the observed and simulated values of grain yield, biomass and LAI were 0.96, 0.98, and 0.93, respectively. The corresponding modeling efficiencies were 0.93, 0.95, and 0.85. All these values are indicative of good model performance.

Although, the performance of the SVM model for the estimation of LAI values from a limited number of RS data were less than desirable, the relationship between the estimated and measured LAI values revealed the potential usefulness of the SVM model for this purpose. It can be concluded that SVM models could be successfully constructed for the prediction of LAI values from larger RS data sets.

Despite of the relatively poorer LAI estimations from RS, the coupling of these values with the crop growth model still improved its performance.

PREFACE TO CHAPTER 6

The application of the STICS model to simulate corn yield under different water regimes and nitrogen application rates, and further coupling of the model with the hyperspectral information was described in Chapter 5. In this Chapter, the applicability of the STICS model will be studied for yield prediction in weed-infested areas. This study was performed with data collected in 2000. A procedure was proposed to incorporate the effect of weeds in crop growth modeling. Various steps associated with this procedure will be described in this Chapter.

Research papers based on the chapter:

Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, and P. K. Goel , 2004. Incorporating the effects of weeds in crop growth modeling. *Agricultural Systems* (Submitted for publication).

CHAPTER 6

INCORPORATING THE EFFECT OF WEEDS INTO CROP GROWTH MODELING

6.1 Abstract

Currently, the accuracy of field-scale crop growth simulation models under weed-infested conditions is limited due to the complexity of weed-crop interactions. A procedure taking into account the effect of weed stresses on crop growth, in a manner that can be incorporated into most crop models, is proposed and demonstrated for a corn crop with the crop growth model STICS. A correction factor is applied to the observed/estimated leaf area index (LAI) values to obtain an LAI corrected for the weed infested areas. Applicability of the method was tested by using data obtained from a field experiment, carried out in the 2000 growing season, at the research farm of McGill University, Ste Anne de Bellevue, Quebec, Canada. The experiment was laid out in a split-plot design, with four-weed control management strategies (no control, control of grasses, control of broadleaf plants, and control of both) in the main plots, and three different nitrogen application rates (60, 120, and 250 kg N ha^{-1}) in subplots, with four replications. Extensive observations were made of weedy conditions in test plots and many other crop-growth influencing parameters. Using corrected LAI values as input to the crop growth model, it was possible to accurately estimate yield on the weed-infested areas of the field.

6.2 Introduction

The adverse effect of weeds on crop yield is well understood and has been widely documented by various researchers (Wilcut *et al.*, 1987; McLennan *et* al., 1991; Donald and Khan, 1992; Knezevic *et al.*, 1994; Mamolos and Kalburtji, 2001). However, reliable methods for quantitative estimation of weed effects on crop growth are still lacking. Many field studies have been conducted to develop models to assess the effect of weeds on crop yield losses. Many researchers have developed empirical regression models to

estimate crop yield losses by relating yield to weed characteristics such as density, biomass, etc. (Dew, 1972; Cousens, 1985a,b; Cousens *et al.*, 1987; Kropff and Lotz, 1992; Knezevic *et al.*, 1997; Ngouajio *et al.*, 1999a; Canner *et al.*, 2002, Lemieux *et al.*, 2003). However, the extrapolation of these results is limited due to the spatial and temporal variability of input parameters.

Cousens (1985b) developed a simple two-parameter hyperbolic model to describe yield loss as a function of weed density. The model was also extended to include crop density as an additional parameter (Cousens, 1985a). In comparison to previously published models, both methods were able to provide better yield loss estimations. Ngouajio *et al.* (1999b) and Lemieux *et al.* (2003) used relative leaf area (the ratio of the weed leaf area to the total leaf area) based models to develop a decision support system for post-emergence herbicide applications. However, they pointed out that accurate estimation of crop losses due to weed effects is the most complicated component of any decision support system because of the complex relationship between weeds and crops. Using multiple regression and path analysis, Mamolos and Kaburtji (2001) studied the effect of Canada thistle density, biomass, and N concentration on wheat yield. The wheat yield was more affected by N concentration as compared to Canada thistle biomass or density.

Most of the above models were developed on the basis of a single crop type and a single weed species at a specific location, which limits the applicability of these models. Moreover, the collection of necessary input parameters for these models, like weed density and relative leaf cover, is time-consuming and labor-intensive. Also, estimation of crop yield by mathematical models in the presence of weeds is rather difficult. No current conceptual model can take into account the effect of weeds due to the lack of a clear-cut relationship between the weeds and crop growth and yield.

In the case of crops grown under normal field conditions (i.e. without any weed stress) higher values of LAI will represent better vegetative growth, which will often translate into higher yields. In weed-infested fields, however, higher LAI values may not imply higher vegetative growth of the crop. The higher LAI values would be the result of both weeds and the main crop.

Recent research in precision agriculture has demonstrated that crop growth parameters, such as LAI, leaf N content and plant height, can be determined accurately by remote sensing under a range of stresses (Foody *et al.*, 1989; Cibula and Carter, 1992; Pearson *et al.*, 1994; Saha and Jonna, 1994; Clevers, 1997; Moran *et al.*, 1997; Lelong-Camille *et al.*, 1998; Anderson *et al.*, 1999; McNairn *et al.*, 2001; Pacheco *et al.*, 2001; Haboudane *et al.*, 2002; Singh *et al.*, 2002; Strachan *et al.*, 2002). With remote sensing, it is possible to identify not only the stress-free areas of the field but also those that are under water, nitrogen and weed stresses (Borregaard *et al.* 2000, Cho *et al.*, 2002; Goel *et al.* 2002, Karimi *et al.*, 2004). Once the weedy and non-weedy areas of the field are located, it should be possible to isolate the impact of weeds on LAI values and, consequently, use the "corrected" LAI values in modeling. Although in most mathematical models, LAI values are simulated, some models do allow the use of observed LAI values at different stages of crop growth as input (Brisson *et al.*, 1998). The latter gives one an opportunity to "adjust" the LAI values as per the prevailing growth conditions in the field.

The overall goal of this study was to develop a method that would allow crop growth models to simulate yield in weedy areas. The applicability of the method is demonstrated using the STICS crop growth model.

6.3 Materials and Methods

6.3.1 Experimental detail

The experiment was conducted at the Lods Agronomy Research Center of Macdonald Campus, McGill University, Ste-Anne-de-Bellevue, Québec, Canada (45°25′45"N lat., 73°56′00"W long.) in 2000. Corn was planted on May 30 at a density of 7.6 plants m^{-2}. The experiment was laid out in a split-plot design, with four-weed management strategies as the main treatment units and three nitrogen application rates as the sub treatment, in quadruplicate. Each plot (20 m × 20 m) consisted of 26 corn rows. The four weed treatments were: no weed control (W1), control of grass species (W2), control of broadleaf species (W3), and full weed control (W4). The three nitrogen application rates consisted of: low nitrogen (60 kg N ha^{-1}, N_{60}), normal nitrogen (120 kg N ha^{-1}, N_{120}), and high nitrogen (250 kg N ha^{-1}, N_{250}). Weeds were controlled by selective

use of herbicides on June 26. At sowing time, minimal amounts of N were applied, namely at rates of 10, 20, and 90 kg N ha^{-1} for the N_{60}, N_{120} and N_{250} treatments respectively, and the remainder broadcast on July 12th.

6.3.2 Crop parameters

Measurements of various crop canopy and other parameters, including plant height, leaf greenness, LAI, biomass and soil moisture, were taken. Crop yield and final biomass were measured at the season's end. The LAI (m^2 foliage area per m^2 ground area) was measured with a LAI-2000 Plant Canopy Analyzer (Li-Cor, Inc., Lincoln, Nebraska, USA). Biomass was estimated on the basis of five to ten plants per plot that were harvested and weighed. Finally, ten representative plants from each plot were harvested and used to estimate crop yield and final biomass. Table 6.1 shows the variations in plant parameters as a function of different weed and nitrogen treatments.

6.3.3 Weed survey

To determine weed populations, a weed survey was performed on July 14 and again on August 15, 2000. In each plot, weeds were counted at three and four randomly chosen 0.5 x 0.5 m quadrates during the first and second survey, respectively. The collected data were: weed types, density, plant height, and ground coverage. The percent of weed cover was also estimated visually in each quadrate. The most prevailing grassy weeds were Yellow Nutsedge (*Cyperus esculentus* L.), Barnyard grass [*Echinochloa Crus-galli* (L.) Beauv.], and crabgrass [*Digitaria ischaemum* (Schreb.) Mühl], whereas the dominant broadleaf weeds were Canada thistle [*Cirsium arvensis* (L.) Scop.], Sow-thistle (*Sonchus oleraceus* L.), Redroot pigweed (*Amaranthus retroflexus* L.), and lamb's quarters (*Chenopodium album* L.). For more details, readers are referred to Goel *et al.* (2003b).

Verification of weed populations in different treatments indicated that the most dominant weeds were grassy weeds. Also, there were minimal broadleaf weeds present in the field. Therefore, the plots were re-categorized into no-weed (consisting of W2 and W4), and weedy treatments (including W1 and W3).

Table 6.1 Details of various crop parameters based on different treatments during entire growth season

Date	Weed	No-weed control			Grass control			Broad leaf control			Full-weed control			Experimental	
	Nitrogen	N_{60}	N_{120}	N_{250}	N_{60}	N_{120}	N_{250}	N_{60}	N_{120}	N_{250}	N_{60}	N_{120}	N_{250}	Min	Max
LAI	6-Jul-00	1.139	1.124	1.498	1.099	0.784	1.399	0.586	0.853	1.125	0.771	0.714	1.056	0.300	2.175
	27-Jul-00	1.986	3.064	3.779	1.635	1.976	2.684	1.634	2.223	2.724	1.473	1.785	2.256	0.775	4.455
	11-Aug-00	3.970	4.450	5.166	2.883	3.155	4.233	3.884	4.626	4.463	2.835	3.444	3.698	2.035	6.185
	26-Aug-00	4.035	4.259	4.574	2.960	3.667	3.831	3.677	4.336	4.271	2.901	3.314	3.646	2.270	5.310
SPAD	30-Jun-00	34.9	35.7	41.7	32.5	36.6	40.6	32.6	36.2	41.7	33.3	34.7	39.3	26.1	43.7
	18-Jul-00	28.3	30.7	40.6	33.1	37.0	46.5	28.5	34.3	41.0	32.5	33.3	37.6	23.4	52.6
	25-Jul-00	33.1	36.4	43.1	41.5	44.0	48.7	34.7	39.9	46.1	38.6	42.1	46.3	27.8	54.4
	8-Aug-00	34.9	43.2	45.1	42.4	46.7	51.6	42.7	46.5	44.1	42.3	46.0	43.4	30.8	56.8
	9-Aug-00	38.9	44.9	52.6	46.0	53.1	55.8	42.2	51.1	54.4	47.1	53.0	55.4	30.2	57.0
	25-Aug-00	34.4	44.6	51.8	44.2	48.1	53.1	34.8	46.4	52.0	39.9	45.1	45.8	28.3	55.2
Plant height (m)	19-Jun-00	7.2	7.4	7.5	7.4	8.1	7.7	8.1	7.6	7.7	6.9	7.9	8.2	6.1	9.6
	27-Jun-00	18.8	20.1	21.3	18.8	19.9	20.1	18.7	19.5	21.2	18.3	19.3	20.6	14.9	23.7
	18-Jul-00	61.7	68.1	89.2	66.7	73.2	92.4	58.1	70.0	83.5	56.1	65.4	74.4	38.5	98.0
	25-Jul-00	88.7	95.2	121.3	98.6	112.5	129.1	82.7	102.3	118.1	80.5	102.1	113.8	61.6	139.7
	8-Aug-00	165.4	173.7	204.4	169.9	194.0	217.6	152.8	183.2	206.5	159.0	182.7	199.6	114.4	230.5
Soil Moisture (% gravimetric)	26-Jun-00	36.0	36.0	32.7	36.4	33.8	35.7	37.3	37.2	38.3	36.1	38.3	37.9	30.4	42.0
	29-Jun-00	33.0	34.6	31.4	31.8	29.7	31.5	33.5	31.7	31.2	29.5	32.9	32.0	24.0	39.9
	30-Jun-00	33.1	33.8	32.9	32.1	33.9	34.0	34.3	34.4	34.0	34.3	35.6	35.3	27.3	39.5
	4-Jul-00	36.6	37.6	35.4	38.1	36.5	38.0	39.0	41.0	37.8	37.6	35.8	39.2	25.4	44.5
	25-Jul-00	26.9	27.5	27.8	26.3	27.2	25.9	29.1	27.0	27.5	27.6	27.8	27.5	21.5	33.8
	5-Aug-00	32.1	34.6	30.9	33.0	31.8	31.1	32.9	32.1	30.6	33.8	31.7	30.9	27.5	38.3
	25-Aug-00	34.0	32.5	33.7	32.2	32.2	31.4	34.4	33.8	32.7	34.3	33.0	32.6	27.1	40.2

6.3.4 Methodology

The STICS crop growth model was selected because of its comprehensive capabilities for simulating crop growth under varying geohydrologic and cropping conditions (Brisson *et al.*, 1998). The model simulates daily growth over the entire growing season on the basis of aboveground biomass of the crop, leaf area index, and soil parameters, according to climatic data. In the model, soil is divided into different homogenous layers. For simulation purposes, each layer is defined by its own water and nitrogen content. Roots modulate plant-soil interactions, which are in part defined by root distribution in the soil. The model computes crop production (biomass and yield) and environmental impact (water and nitrogen leaching) under different agronomic practices. The main advantage of the model is that the information related to LAI can be incorporated easily into the model. So, the field variability can be accounted for in the model by incorporating the parameters affecting crop growth (Brisson *et al.*, 1998).

The LAI measurements/estimations in weedy areas cannot be used directly in a model because they would be greater in weedy areas merely as a result of the presence of both crop and weed canopies, and thus would not be a reflection of a stronger vegetative crop. Actually, the crop would be suffering in weedy areas and the "true" crop LAI values would always be less than those in non-weedy areas.

The LAI measurements could be made in the experimental field in both weedy and non-weedy areas with the help of a Plant Canopy Analyzer but such measurements are impractical on a field scale. As stated earlier, optical remote sensing measurements can be used to estimate LAI values in the field (Pacheco *et al.*, 2001; Strachan *et al.*, 2002; Goel *et al.*, 2003a). However, one still needs to account for the weedy and non-weedy areas of the field as the LAI estimations in weedy areas may not be correct. We propose a methodology which can overcome this problem. Since ground measurements of LAI values were available to us, they were used to demonstrate the proposed methodology. The method should work equally well when only remote sensing observations are available.

In this regard, the weedy plots' mean measured LAI values were calculated and related to the corresponding values in non-weedy plots. A linear relationship existed between the two sets of LAI values. However, as stated before, this relationship would

not be of much use as the LAI values in weedy areas would be the result of both the crop and the weeds, and although these values will be higher than the corresponding values in the non-weedy areas, the "true" LAI values in weedy areas would actually be less because of the competition between the crop and the weeds.

An alternate strategy was developed to incorporate the effect of weeds into crop growth modeling. First, the mathematical model was run iteratively with different LAI values to bring the simulated crop yield closer to the observed crop yield in weedy areas. The final set of LAI values would represent the "true" LAI values of the crop (only) in weedy areas. This "true" set of LAI values can be related to the measured LAI values in weedy areas, representing both the crop and the weeds. If this relationship turns out to be satisfactory, then this will allow us to correct the estimated/measured LAI values in weedy areas, and hence the "true" LAI values could be used with mathematical modeling to estimate crop yields. As mentioned before, remote sensing can be used to not only locate the weedy areas on the field but also provide good estimates of their LAI values.

To demonstrate this method, the STICS crop growth model, which does not take into account the effect of weeds, was run iteratively by changing the input parameters for LAI (i.e. the relevant growth stage parameters and dlaimax, a parameter for defining the maximum rate of LAI increase) so that the simulated crop yield matched with the observed crop yield. This was done for the weedy plots only. Once a good set of LAI values were determined, they were plotted against the LAI values measured/estimated in the field, and a linear relationship between the two was observed. This linear relationship was used to transform the measured/estimated LAI values in weedy areas to the "correct" crop LAI values.

As stated earlier, due to the prevailing weedy conditions in our test plots, the weed treatments were reclassified as weedy and no-weed treatments, each consisting of twenty-four test plots. For each weed treatment, there were eight plots per nitrogen treatment. For each nitrogen treatment in the weedy plots, the STICS model was run iteratively with the mean data for that treatment, and a set of LAI values was obtained. In this way, three sets of LAI values were obtained for the weedy plots. These LAI values represent the "true" crop LAI values in the weedy plots as they were obtained

by matching the simulated crop yields with the observed ones. These LAI values were plotted against the corresponding mean of the measured LAI values, and a linear relationship between them was developed. Next, this relationship was used to convert the observed LAI values in the weedy plots to "true" LAI values in each individual plot. The "true" LAI values were further used to simulate yield in weed-plots using the STICS model.

One of the weaknesses of this is that the same data source was used for model development and validation. Therefore, to overcome this problem, the parent data set was randomized and divided into two subsets: 75% for the LAI relationship development and 25% for its validation. The mean values in the developmental data set were used to obtain the LAI relationship, which was used to obtain the "true" LAI values for each plot in the 25% validation data set. The "true" LAI values, thus obtained, were used in the STICS model to estimate crop yield. To check for the robustness of the LAI relationship, a five-fold cross-validation procedure was used where the data was randomized five times and the above steps were repeated.

6.4 Results and discussion

The variation in the measured mean LAI from the weed and no-weed treatments, along with the best fit line ($r^2 = 0.97$), are shown in Figure 6.1. As expected, the LAI values in the weed plots were higher than those in the no-weed plots. This is because of the presence of both weed and main crop canopies in weedy areas. Furthermore, in both weedy and weed-free plots, the LAI values were proportional to nitrogen applications, causing higher LAI values in higher nitrogen application treatments (Figures 6.2 and 6.3). In the case of crop yield, although yield was proportional to LAI values in both weedy and weed-free plots, as expected, the yield was lower in the weedy plots. The difference between the yields of weedy and weed-free plots was greater at lower LAI values, under the lower nitrogen application rates, as compared to the ones at high LAI values. This also implies that the effect of weed stress is more pronounced in the presence of nitrogen stress due to a greater

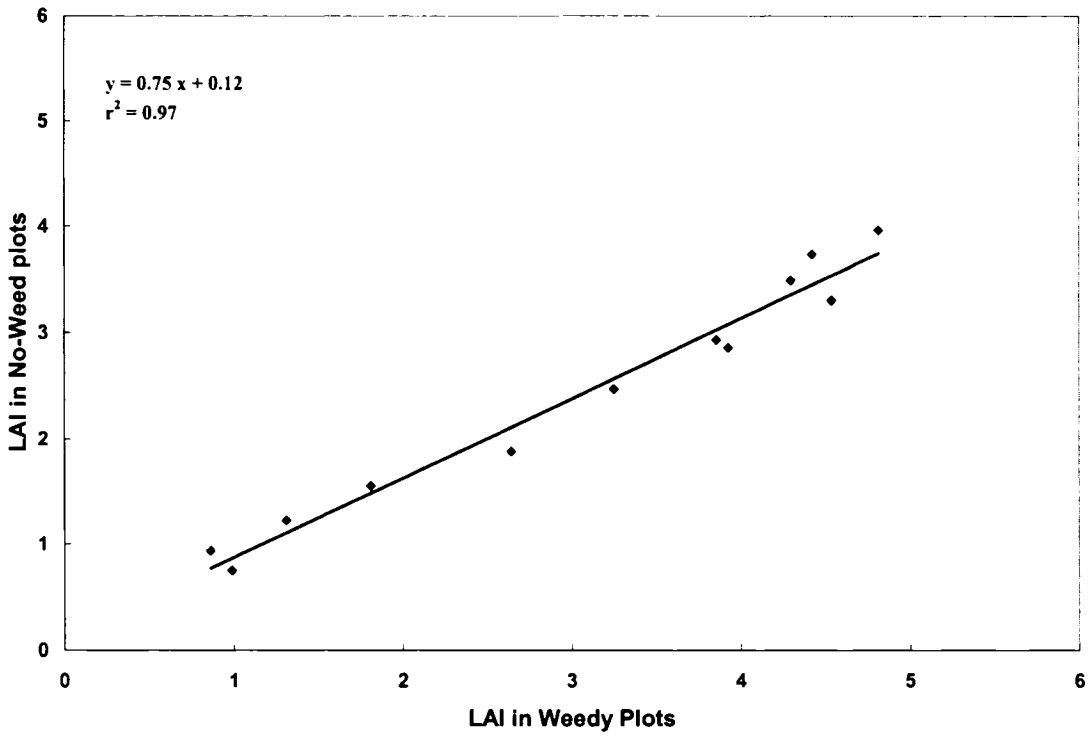


Figure 6.1 Measured LAI values in weedy and no-weed plots.

competition between weeds and crop for nutrients and water uptake (Figure 6.4).

As stated earlier, no mathematical model can take into account the effect of weeds in a logical way, and the STICS model is no exception. Following our approach, we need to correct the LAI values of weedy plots to find their "true" values. So, first the STICS model was run iteratively by adjusting the LAI parameters to match the simulated yields with the measured ones in weedy plots. Next, the LAI values obtained from the model simulations were plotted against the mean measured LAI values in weedy plots (Figure 6.5). The relationship between the two was described by the following regression equation:

$$Y = 0.59\,X + 0.06 \qquad (r^2 \text{ of } 0.98;\ P \leq 0.05) \qquad 6.1$$

where Y is the simulated ("True") LAI by the model and X is the mean measured LAI value in the weed plots. The high r^2 value confirms a very good linear relationship between the estimated and measured LAI values.

The applicability of the overall approach was tested by applying the STICS model to each individual weedy plot. The "true" LAI values for each plot were estimated from equation 6.1, and the latter were used in the STICS model to simulate crop yield. Figure 6.6 shows the estimated versus the measured yield values for the weedy treatments. It is evident from the best fit line (equation 6.2) and the 1:1 line that the model did an excellent job of simulating crop yield in the weedy plots.

$$Y = 0.98\,X - 0.01 \qquad (r^2 = 0.94;\ P \leq 0.05) \qquad 6.2$$

where Y and X are the simulated and measured yields, respectively. A *t*-test was performed and both the slope and the y-intercept were found to be not significantly different from one and zero, respectively. This proves that not only the model performed very well, but also the proposed approach was successful. The root mean square error (RMSE) was 0.43 Mg ha^{-1}, and the relative RMSE (RRMSE), a ratio of RMSE to the mean observed yield, was 0.08, which, again, indicates an excellent agreement between

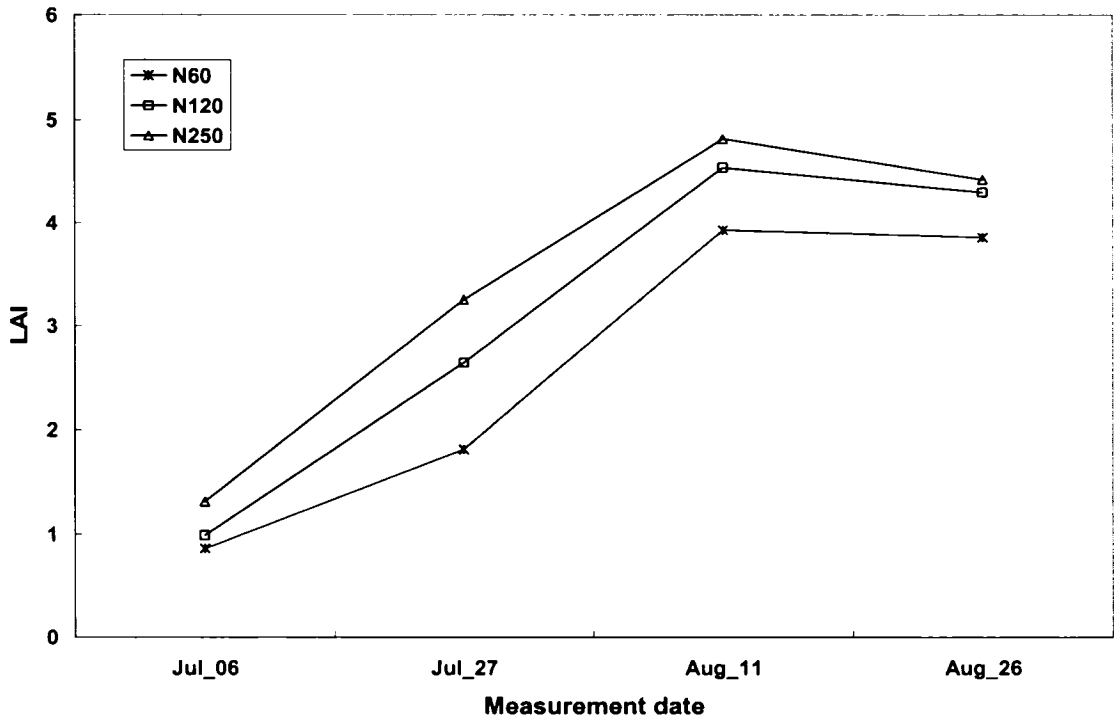


Figure 6.2 Variation in the average LAI values in weedy treatments (N_{60}, N_{120}, and N_{250} are representing low, normal, and high nitrogen application rates, respectively).

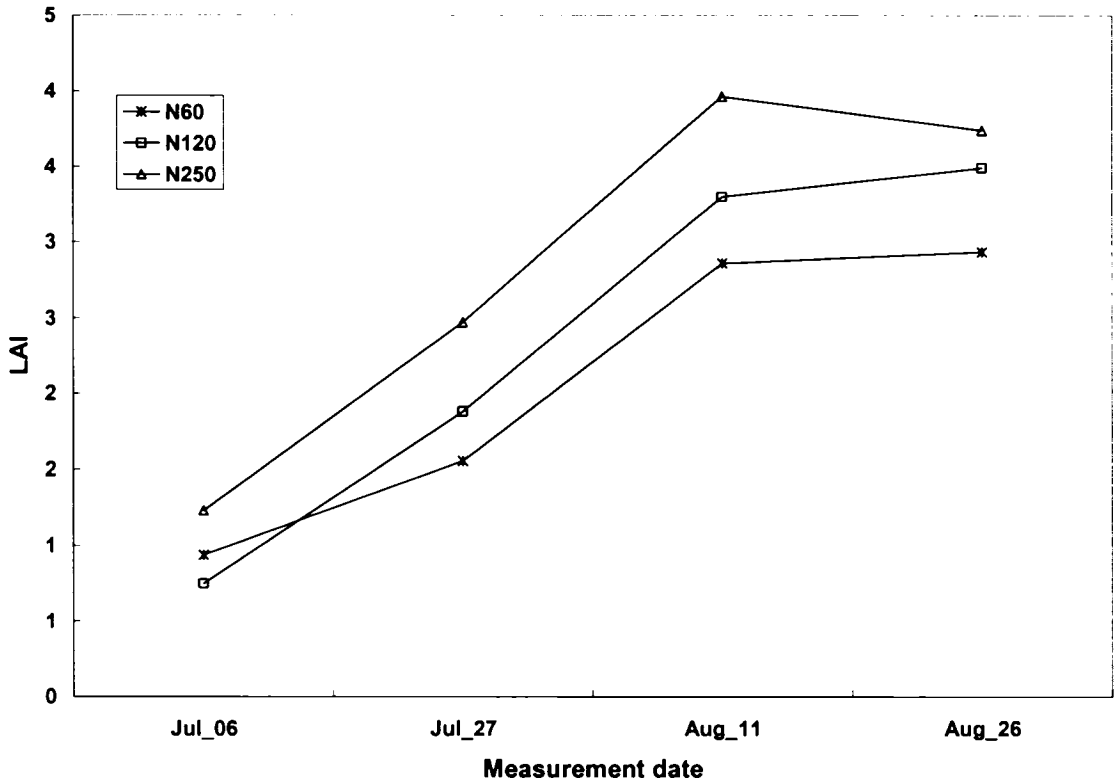


Figure 6.3 Variation in the average LAI values in no-weed treatments (N_{60}, N_{120}, and N_{250} are representing low, normal, and high nitrogen application rates, respectively).

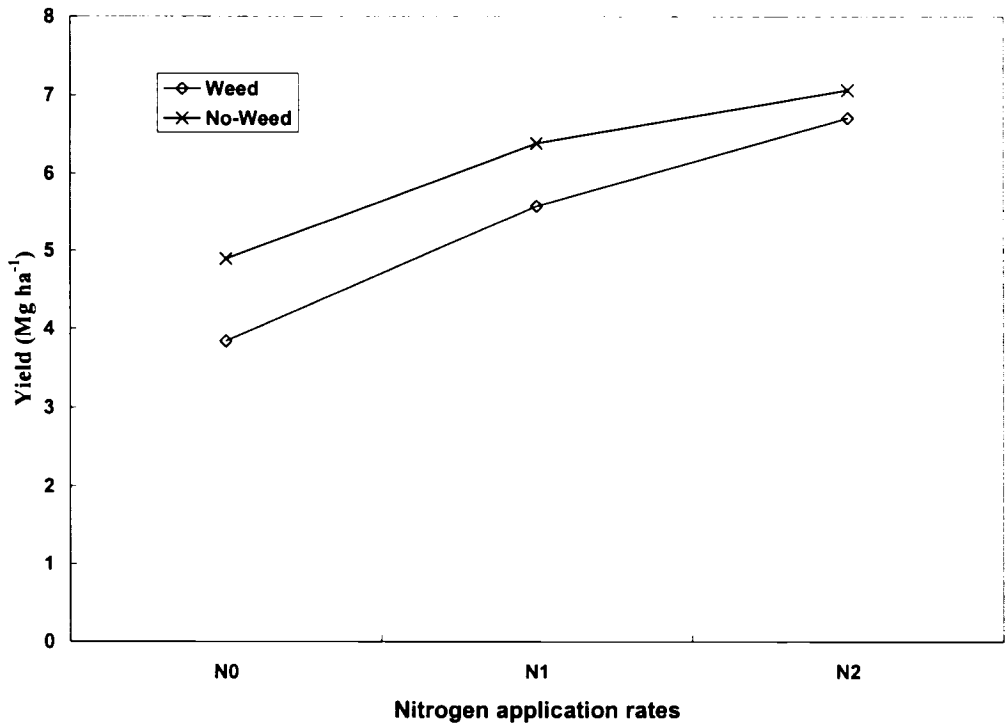


Figure 6.4 Comparison of mean measured crop yield in weedy and no-weed treatments.

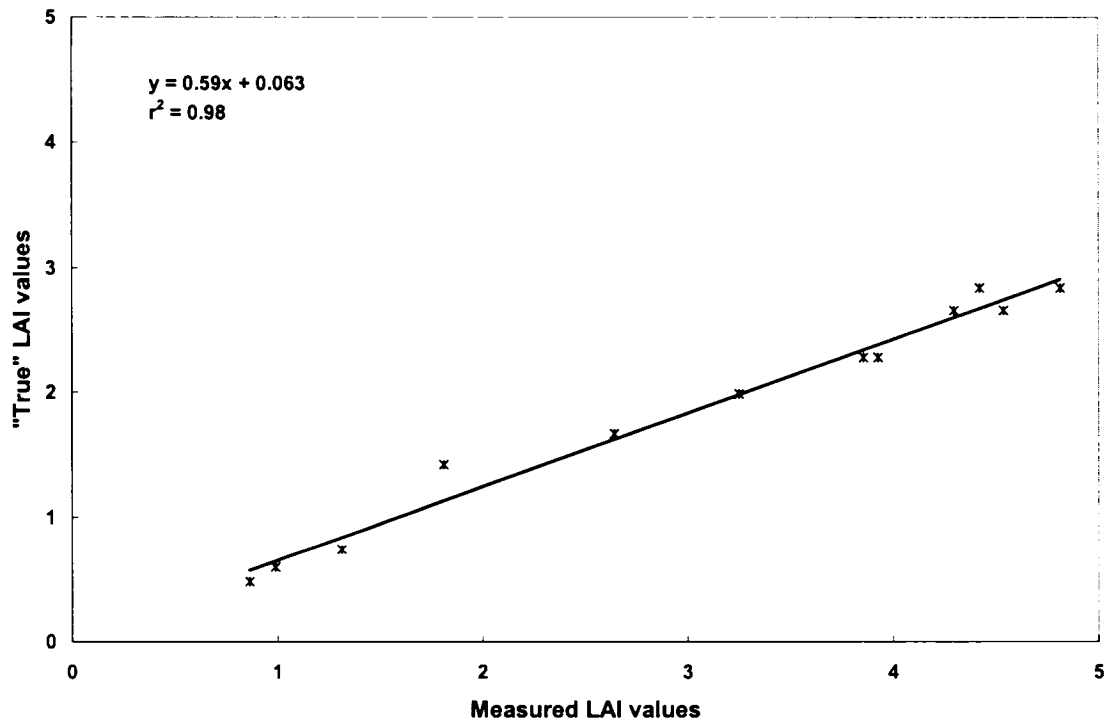


Figure 6.5 Simulated "True" and measured LAI values in weedy plots.

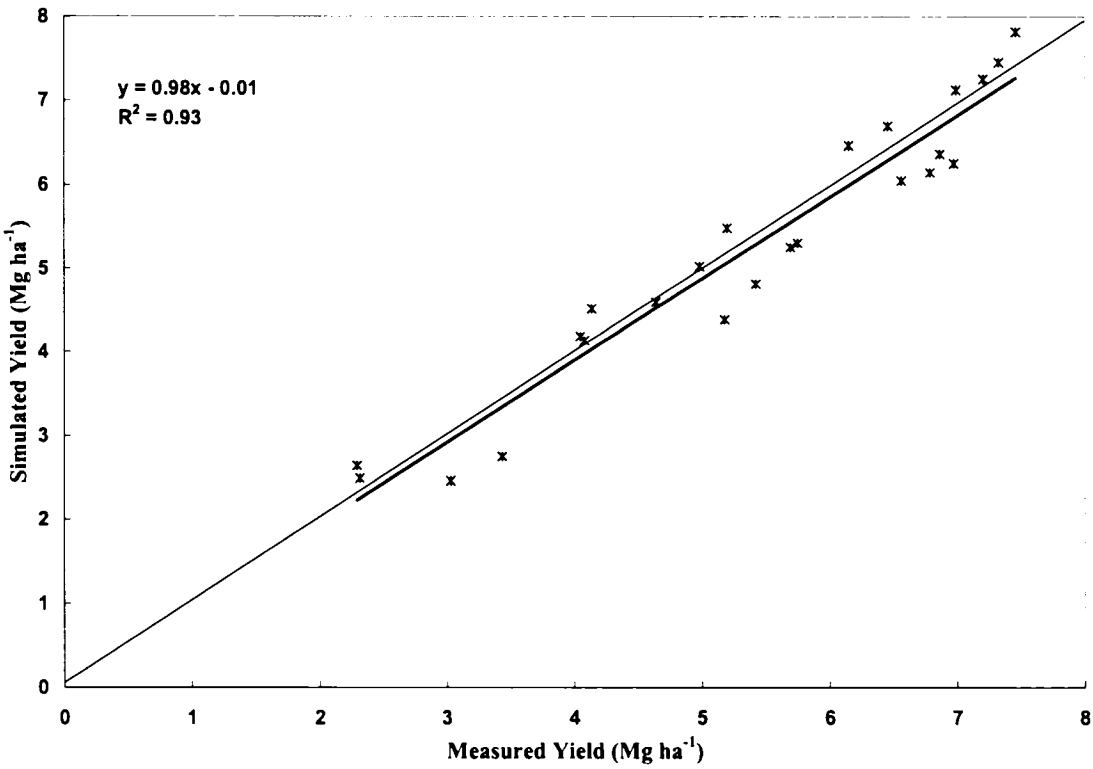


Figure 6.6 Simulated verses measured yield values in weedy plots using all data set (validation).

the observed and simulated yields (Table 6.3). Also, the modeling efficiency of 0.93 proves this point rather well.

One of the weaknesses of this procedure is that the same data set was used for the development and testing, i.e. the average values were used for the development of the linear relationships and the individual plot values were used for model evaluation. Therefore, the evaluation was extended to unseen data. The overall data were divided into two sets, and a five-fold cross-validation was carried out with unseen data. A summary of the regression equations and corresponding r^2 of all five-fold cross-validations is provided in Table 6.2 for both LAI and crop yield. Once again, higher values of r^2 (0.95 to 0.99) for the LAI relationship were found. The r^2 values for crop yield ranged from 0.78 to 0.97, which indicates good model performance. The model performance was also evaluated using the statistical tests on regression parameters. In four out of five repetitions, the regression parameters, intercept and slope, were not significantly different from 0 and 1 ($p \leq 0.05$), respectively, which represents consistently good performance by the model (Table 6.3). In the 4^{th} repetition, although the higher value of r^2 indicates a good model performance, the intercept of the regression line was significantly different from 1 ($p \leq 0.05$). The RMSE varied from 0.90 to 0.26 Mg ha^{-1}, and RRMSE varied between 0.05 to 0.18, which demonstrate a good agreement between the observed and simulated crop yields in all five repetitions (Table 6.3). Also, the EF varied between 0.75 and 0.97, which is another indicator of satisfactory model performance.

Therefore, it appears that the proposed approach can be used for the estimation of crop yield in weed-infested corn fields. Remote sensing may be used to first identify the weedy and weed-free areas in the field and then to estimate LAI values in those areas. However, the proposed approach also required the determination of the "true" LAI values in weedy areas. In this paper, we use the measured crop yields in such areas but more work needs to be done to determine these important input parameters.

6.5 Conclusion

In this study, an attempt was made to develop a procedure to account for the presence of weeds in crop growth models for better crop yield simulations. The procedure involved

the determination of "true" LAI values in weedy areas which can then be used in crop growth models to simulate crop yield. The overall approach was demonstrated by using the STICS model. With the proposed procedure, the crop yield in weedy areas was very well simulated by the model. The high correlation coefficient, low mean difference, low mean absolute deviation, and high modeling efficiency demonstrated good model performance in weedy areas.

Table 6.2 Summary of the regression equations and r^2 obtained using different data sets.

Repetition Number	LAI		Yield	
	Equation	r^2	Equation	r^2
1	Y = 0.60 X + 0.10	0.97	Y = 1.03 X – 0.34	0.93
2	Y = 0.58 X + 0.28	0.95	Y = 0.68 X + 2.17	0.78
3	Y = 0.64 X + 0.01	0.99	Y = 0.74 X + 1.62	0.87
4	Y = 0.62 X + 0.05	0.97	Y = 0.77 X + 1.21	0.97
5	Y = 0.60 X + 0.07	0.97	Y = 0.95 X + 0.28	0.97

Table 6.3 Statistical parameters calculated for measured and simulated yield on weed treatments.

Index *	Complete set	1	2	3	4	5
RMSE	0.43	0.41	1.00	0.67	0.44	0.26
RRMSE	0.08	0.08	0.18	0.13	0.08	0.05
MBE	-0.14	-0.20	0.51	0.28	0.01	-0.001
EF	0.92	0.90	0.69	0.82	0.93	0.97
EF1	0.73	0.68	0.49	0.63	0.74	0.84
R	0.97	0.96	0.88	0.93	0.99	0.99
Intercept	-0.01	-0.34	2.17	1.62	1.21**	0.28
Slope	0.98	1.03	0.68	0.74	0.77	0.95
***P*(F) intercept = 0 & slope = 1**	0.32	0.67	0.43	0.46	0.19	0.88

* **RMSE**= Root mean square error, **RRMSE** = Relative RMSE, **MBE** = Mean bias error, **EF** = Model efficiency, **EF1** = Modified model efficiency

** Intercept is significantly different from zero ($p \leq 0.05$)

CHAPTER 7

SUMMARY AND GENERAL CONCLUSIONS

7.1 Summary

This study investigated the applicability of hyperspectral observations, acquired from airborne sensors or a handheld spectroradiometer, to distinguishing between various types of stresses: weeds, nitrogen and water. In addition, attempts were made to incorporate these stresses and hyperspectral data into crop growth modeling. A new technique to incorporate the effect of weeds in corn fields into crop growth modeling was also developed.

The objectives were met by carrying out three field studies, involving growing corn with different combinations of weed control and nitrogen fertilization rates in 2000, different combinations of weed control, nitrogen application rates and water regimes in 2001, and various combination of nitrogen application rate and water regime in 2002. In 2000, three aerial images were taken during the growing period, i.e., at early growth, tasseling, and full maturity, using a 72-waveband hyperspectral aerial sensor (407-nm to 949-nm range). In the second and third year of the study, the spectral measurements were made using a handheld spectroradiometer (350-nm to 2500-nm with bandwidths of 1-nm). Various crop physiological parameters were also measured.

7.2 Conclusions

A number of general conclusions were reached in this study. Their presentation in this section closely follows the order in which this study was done.

(1) Hyperspectral aerial images taken in the year 2000 were analyzed for their applicability in discriminating between weeds and nitrogen stresses in a corn field. In this study, the ability of a stepwise approach in the selection of highly effective narrow wavebands was examined. The importance of these chosen narrow-waveband hyperspectral aerial observations was demonstrated in identifying weed and nitrogen

stresses. The study also revealed that discriminant functions based on a restricted set of wavebands showed better performance in the detection of stresses in early crop growth stages, which is an important and crucial stage in precision agriculture. Furthermore, the results of the discriminant analysis method were compared to those obtained with ANNs and DT algorithms. Consideration of two target variables at the same time (combined effect of weeds and nitrogen application rates) resulted in higher misclassification rates (25% or greater) for all three classification methods with the cross-validation (unseen) data set. However, when one target at a time was considered (nitrogen application rate, or weed treatment) better results were obtained. Misclassification rates for cross-validation data were below 20%, and the lowest misclassification rate was 11%. A comparison of the results obtained with the above methods showed that discriminant analysis could provide better classification accuracy at the early growth stage for all three classification problems: 75%, 87%, and 83%, for the combined effect of weeds and nitrogen application rates, weeds alone, and nitrogen application rates alone, respectively. At the tasseling stage, once again the discriminant analysis provided higher accuracy for the two target variables at the same time (71%) while ANN models provided higher classification accuracy for the case of one target variable at a time (70%).

(2) The possibility of improving the nitrogen stress detection in conjunction with water stress via the application of very narrow-waveband hyperspectral observations, taken with a handheld spectroradiometer in the year 2002, was explored. Data analysis was also performed using stepwise approaches, which were found previously to be highly effective in the selection of important narrow-wavebands. Various discriminant analysis methods were developed using all wavebands identified by the stepwise procedure, as well as different subsets from selected wavebands. Although when using all selected wavebands various irrigation and nitrogen treatments were not always successfully classified, considerable improvement was achieved when discriminant analysis methods were applied on a well-chosen subset of selected wavebands.

The results of the discriminant analysis method indicated that this method could identify different nitrogen and water treatments, achieving a misclassification rate of only 5% for the 26 August observations, and zero misclassification rates for the 31 July and 10

August observations. Furthermore, the applicability of canonical discriminant analysis (CDA) was tested in this study, and the results showed that the CDA has a good potential to identify different nitrogen and water treatments. The first two canonical functions explained more than 94% variability for the 31 July and 26 August data sets, and about 84% for the August 10^{th} data.

(3) Considering the time, expense and site-specific restriction of field studies, efforts were made to evaluate the performance of STICS, a crop growth model for southwestern Quebec. The applicability of the model was evaluated by comparing the predicted model output with measured values for grain yield, biomass and LAI. Predicted and measured values for these three crop parameters were well correlated with one another. Correlation coefficients of 0.97, 0.98, and 0.93 and model efficiencies (EF) of 0.93, 0.95, and 0.85 for yield, biomass, and LAI, respectively, indicated good model performance. Furthermore, the model was also linked to remote sensing observations by estimating LAI values from these observations. A new method of estimation based on an artificial intelligence system, called SVM, was used. While the model-generated LAI predictions were not so good, coupling the crop growth model with remote sensing observations resulted in an improvement in crop growth simulations.

(4) For more accurate corn yield prediction in weed infested areas, attempts were made to explore the possibility of extending the application of crop growth modeling to such areas. A methodology was proposed that would account for the effect of weeds on crop growth, and thus could be used in crop growth modeling. Since LAI is an indicator of overall crop growth, the proposed procedure was based on differentiating the "true" LAI values of the crop in weedy areas. The "true" LAI values were then used in the model for corn growth simulation. Simulated corn yield in weedy areas using this correction factor correlated well with the measured yield.

CHAPTER 8

CONTRIBUTIONS TO KNOWLEDGE AND SUGGESTIONS FOR FUTURE RESEARCH

8.1 Contributions to Knowledge

The specific contributions to knowledge from this study are:

1. This study is the first of its kind in the application of a discriminant analysis algorithm along with a stepwise approach to simultaneously classify plant stress factors, namely weeds and nitrogen, and nitrogen and water. The stepwise approach allowed us to identify the most useful and effective narrow hyperspectral wavebands among highly complex data sets. The discriminant analysis method, when applied with those selected wavebands, was highly successful in identifying different types of stresses.

2. This study is the first attempt to evaluate the STICS crop growth model for a corn field in southwestern Quebec. The model was also linked with hyperspectral observations.

3. To the author's knowledge, this is the first field study investigating the possibility of applying a new artificial intelligence method, support vector machine (SVM), to estimate the crop parameter (LAI) from hyperspectral aerial images.

4. A new procedure has been proposed to incorporate weed effects into crop growth modeling. This procedure will lead to more precise crop growth monitoring and yield estimation in both weedy and non-weedy areas.

8.2 Suggestions for Further Research

1. To determine the ability of hyperspectral observations to distinguish among various levels of combined effects of weed, nitrogen, and water stresses, it is suggested that a field experiment including all three factors be set up. During the entire growing period, hyperspectral observations should be taken at different growth stages. In addition, various cropping parameters (e.g., LAI, biomass, etc.) should also be collected coincidentally.

2. One of the problems in this study was that the weed treatments were limited to grass and broad leaf type. It is suggested that an experiment with more specific weed combinations be set up, and hyperspectral data in different growth stages be acquired. Also, more accurate spectral signature of various weeds and weed combinations at different growth stages should be assessed. The study should also be extended to other major crops of the region.

3. One of the important steps in incorporating the weed effect in crop growth modeling was the establishment of a reliable technique to find the true crop LAI values in weed-infested areas. Further investigation is required to find more accurate methods for proper development of such relationships.

4. To extend the application of the results of this study to larger areas and also other crop types, more work needs to be done.

5. One of the important steps in incorporating the weed effect in crop growth modeling was the establishment of a relationship to find the true crop LAI values. For proper development of such relationships, further investigation is required to examine the ability of other methods of estimating LAI from hyperspectral data.

6. In this study the general ability of hyperspectral observations in crop growth monitoring was examined. To evaluate each portion of the spectrum., further research must to be performed.

7. Further exploration of hyperspectral satellite (e.g., Hyperion) observations is needed to test the results from this study across larger regions (e.g., in national level).

CHAPTER 9

REFERENCES

Aase J. K. and F. H. Siddoway. 1981. Assessing winter wheat dry matter production via spectral reflectance measurements. *Remote Sensing of Environment,* 11: 267-277.

Addiscott T. M. and A. P. Whitmore. 1987. Computer Simulation of Changes in Soil Mineral Nitrogen and Crop Nitrogen during Autumn, Winter and Spring. *The Journal of Agricultural Science* (Cambridge). 109: 141–157.

Allen E. M., W. D. Batcher, and T. S. Colvin, 1996. Validation of corn and soybean models in Iowa: Implications for precision farming. *ASAE paper No. 96-1006.* St. Joseph, MI.

Anderson, J. E., R. I. Fisher, and S. R. Deloach. 1999. Remote sensing and precision agriculture: ready for harvest or still maturing. *Photogrammetric Engineering and Remote Sensing,* 65(10): 1118-1123.

Asner, G. P., C. A. Wessman, C. A. Bateson, and J. L. Privette. 2000. Impact of tissue, canopy, and landscape factors on the hyperspectral reflectance variability of arid ecosystems. *Remote Sensing of Environment,* 74(1): 69-84.

Asrar, G., L. E. Hipps, and E. T. Kanemasu. 1984. Assessing solar energy and water use efficiencies in winter wheat: A case study. *Agricultural Forest Meteorology*, 31: 47-58.

Asrar G., E. T. Kanemasu, R. D. Jackson, and P. J. Pinter Jr. 1985a. Estimation of total aboveground phytomass production using remotely sensed data. *Remote Sensing of Environment,* 17: 211-220.

Asrar G., E. T. Kanemasu, and M. Yoshida. 1985b. Estimates of leaf area index from spectral reflectance of wheat under different cultural practices and solar angle. *Remote Sensing of Environment,* 17: 1-11.

Atkinson, P. M. and A. R. L. Tatnall. 1997. Neural networks in remote sensing. *International Journal of Remote Sensing*, 18: 699-709.

Attema, E. P. W. and F. T. Ulaby. 1978. Vegetation modeled as a water cloud. *Radio Science,* 13:357-364.

Bach, H., A. Demircan, and W. Mauser. 1995. The use of AVIRIS data for the determination of agricultural plant development and water content. Proceedings of MAC Europe '91-Final Results. ESA WPP-88 (Paris: ESA), Lenggries, Germany.

Bach, H., W. Verhoef, and K. Schneider. 2001. Coupling remote sensing observation models and a growth model for improved retrieval of (geo) biophysical information from optical remote sensing data. *Remote Sensing for Agriculture, Ecosystems and Hydrology, SPIE* 4171: 1-12.

Baret, F. and G. Guyot. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment,* 35:161-173.

Bastiaanssen, W.G. M. and M. G. Bos. 1999. Irrigation performance indicators based on remotely sensed data: a review of literature. *Irrigation and Drainage Systems,* 13: 291–311.

Bastiaanssen, W.G. M., D. I. Molden, and I. W. Makin. 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. *Agricultural Water Management,* 46(2): 137-155.

Bellocchi, G., M. Acutis, G. Fila and M. Donatelli. 2002. An indicator of solar radiation model performance based on a fuzzy expert system. *Agronomy Journal* 94:1222-1233.

Blackburn G. A. 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing,* 19 (4): 657-675.

Blackburn G. A. 1999. Relationships between spectral reflectance and pigment concentrations in stacks of deciduous broadleaves. *Remote Sensing of Environment,* 70(2): 224-237.

Boochs, F., G. Kufer, K. Docker, and W. Kuhbauch. 1990. Shape of the red edge as vitality indicator for plants. *International Journal of Remote Sensing,* 11: 1741–1753.

Boote K. J. and J. W. Jones. 1998. Simulation of crop growth: CROPGRO model. In: Agricultural Systems Modeling and Simulations, eds. Peart, R.M. and Bruce Curry R. Marcel Dekker, Inc., New York, pp. 651-692

Bork, E. W., N. E. West, and K. P. Price. 1999. Calibration of broad- and narrow-band spectral variables for rangeland cover component quantification. *International Journal of Remote Sensing,* 20 (18): 3641-3662.

Borregaard, T., H. Nielsen, L. Norgaard, and H. Have. 2000. Crop-weed Discrimination by Line Imaging Spectroscopy. *Journal of Agricultural Engineering Research,* 75: 389-400.

Bouman, B. A. M. 1992. Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. *International Journal of Remote Sensing*, 13: 2565-2581.

Brisson, N., B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Durr, G. Richard, N. Beaudoin, S. Recous, X. Tayot, D. Plenet, P. Cellier, J. M. Machet, J. M. Meynard, and R. Delécolle. 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie,* 18: 311-346.

Brisson, B., F. Ruget, P. Gate, J. Lorgeou, B. Nicoullaud, X. Tayot, D. Plenet, M. H. Jeuffroy, A. Bouthier, D. Ripoche, B. Mary, and E. Justes. 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie,* 22: 69–92.

Brown, R. B., J. P. G. A. Steckler, and W. G. Anderson. 1994. Remote sensing for identification of weeds in no-till corn. *Transactions of the ASAE,* 37(1): 297-302.

Brown, R. B. and J. P. G. A. Steckler. 1995. Prescription maps for spatially variable herbicide application in no-till corn. *Transactions of the ASAE,* 38(6): 1659-1666.

Burges, C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2): 121-167.

Burks, T. F., S. A. Shearer, J. D. Green, and J. R. Heath. 2002. Influence of weed maturity levels on species classification using machine vision. *Weed Science,* 50: 802–811.

Canner, S. R., L. J. Wiles, and G. S. McMaster. 2002. Weed reproduction model parameters may be estimated from crop yield loss data. *Weed Science,* 50(6): 763–772.

Carter, G. A. 1994. Ratio of leaf reflectance in narrow wavebands as indicator of plant stress. *International Journal of Remote Sensing,* 15: 697-703.

Carter, G. A. 1998. Reflectance bands and indices for remote estimation of photosynthesis and stomatal conductance in pine canopies. *Remote Sensing of Environment,* 63: 61-72.

Ceccato, P., S. Flasse, S. Tarantola, S.Jacquemoud, and J. M. Grégoire. 2001. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, 77: 22-33.

Chang, C. and C. Lin. 2004. "LIBSVM: a library for support vector machines". Available from: http://www.csie.ntu.edu.tw/~cjlin/libsvm/. Access date, April 20, 2004.

Chapman, S. C., G. L. Hammer, and H. Meinke. 1993. A sunflower simulation model: I. Model development. *Agronomy Journal,* 85:725–735.

Chauki, H, L. Prévot, D. Troufleau, N. Brisson, A. Olioso, F. Baret, and M. Guérif. 1999. Assimilation of ERS and RadarSat SAR observations within the STICS crop growth model. CEOS'99 SAR Workshop, TOULOUSE, France, October 26-29, 1999.

Childs, S. W., J. R. Gilley, and W. E. Splinter. 1977. A simplified model of corn growth under moisture stress. *Transactions of the ASAE,* 20: 858-865.

Cho, S. I., D. S. Lee, and J. Y. Jeong. 2002. Weed–plant discrimination by machine vision and Artificial Neural Network. *Biosystems Engineering,* 83 (3): 275–280.

Christensen, S., E. Nordbo, T. Heisel, and A. M. Walter. 1998. Overview of development in precision weed management, issues and future directions being considered in Europe. In: Precision Weed Management in Crops and Pasture. Medd, R.W. and Pratley J.E. (Eds.), Proceedings of a Workshop, 5-6 May, (CRC for Weed Management Systems, Adelaide), pp. 154.

Cibula W. G., and G. A. Carter 1992. Identification of a far-red reflectance to ectomycorrhizae in slash pine. *International Journal of Remote Sensing,* 13: 925-932.

Clevers, J. G. P. W. 1988. The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sensing of Environment,* 25: 53-69.

Cevers, J. G. P. W. 1989. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote sensing of Environment.* 29: 25-37.

Clevers, J. G. P. W. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment,* 61: 221-228.

Clevers, J. G. P. W. and H. J. C. Van Leeuwen. 1996. Combined Use of optical and microwave remote sensing data for crop growth monitoring. *Remote Sensing of Environment,* 56: 42-51.

Clevers, J. G. P. W., O. W. Vonder, R. E. E. Jongschaap, J. F. Desprats, C. King, L. Prévot, and N. Bruguier. 2000. Monitoring wheat growth by calibrating a crop growth model using optical satellite data. European Geophysical Society (EGS) XXV General Assembly, Nice, France, April, 28, 2000.

Cousens, R., 1985a. An empirical model relating crop yield to weed and crop density and a statistical comparison with other models. *Journal of Agricultural Science, Cambridge,* 105: 513-521.

Cousens, R. 1985b. A simple model relating yield loss to weed density. *Annuals of Applied Biology,* 107: 239-252.

Cousens, R., P. Brain, J. T. O'Donovan, and P. A. O'Sullivan. 1987. The use of biologically realistic equations to describe the effects of weed density and relative time of the emergence on crop yield. *Weed Science,* 35: 720-725.

Curran P. J.. 1985, Aerial photography for the assessment of crop condition: a review. *Applied Geography,* 5: 347-360.

Deck, S. H., C. T. Morrow, D. H. Heinemann, and H. J. Sommer III. 1995. Comparison of a neural network approach. *Transactions of the ASAE,* 39(6): 2319-2324.

Deguise, J. C., K. Staenz, and J. Lefebvre. 1999. Agricultural application of airborne hyperspectral data: weed detection. Presented at the Fourth International Airborne Remote sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada, 21-24 June.

Delécolle, R. and M. Guérif. 1988. Introducing spectral data into a plant process model for improving its prediction ability. Proceedings of the 4th International

Colloquium Signatures Spectrales d'Objets en Télédétection, Aussois, France, 18-22 January 1988, ESA-SP287 (Paris: European Space Agency), pp. 125-127.

Delécolle, R., S. J. Maas, M. Guérif, and F. Baret. 1992. Remote sensing and crop production models: present trends. *ISPRS J. Photogram Remote Sensing,* 47: 145-161.

Denisov, V. V. 2001, Development of the crop simulation system DIASPORA. *Agronomy Journal,* 93: 660–666.

Dew, D. A. 1972. An index of competition for estimating crop loss due to weeds. *Canadian Journal of Plant Science,* 52: 921–927.

Donald, W. W. and M. Khan. 1992. Yield loss assessment for spring wheat (*Triticum aestivum*) infested with Canada Thistle (*Cirsium arvense*): *Weed Science, 40 590–598.*

Eckersten, H. and P. E. Jansson. 1991. Modeling water flow: nitrogen uptake and production for wheat. *Fertilizer Research,* 27: 313-329.

Elvidge C. D. and Z. Chen. 1995. Comparison of broadband and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, 54: 38-48.

ENVI. 2000. *Exploring ENVI Training Course Manual.* Boulder, Colo.: Research Systems Inc. Menges, R. M., P. R. Nixon, and A. J. Richardson. 1985.

Erickson, J. D. 1984. The LACIE experiment in satellite aided monitoring of global crop production. In: The Role of Terrestrial Vegetation in the Global Carbon Cycle: Measurements by Remote Sensing, Woodwell G. M. (ed.) (Chichester: Wiley), pp. 191-217.

Everitt, J. H., R. D. Pettit, and M. A. Alaniz. 1987. Remote sensing of Broom snake weed (*Gutierrezia sarothrae*) and Spiny Aster (*Aster spinosus*). *Weed Science,* 35(2): 295-302.

Everitt, J. H., G. L. Anderson, D. E. Escobar, M. R. Davis, N. R. Spencer, and R. J. Anrascik. 1995. Use of remote sensing for detecting and mapping leafy spurge (Euphorbia esula). *Weed Technology,* 9: 599-609.

Everitt, J. H., D. E. Escobar, M. A. Alaniz, M. R. Davis, and J. V. Richerdson. 1996. Using spatial information technology to map Chinese tamarisk (tamarix Chinesis) infestation. *Weed Science*, 44(1): 194-201.

Fila, G., G. Bellocchi, M. Acutis, and M. Donatelli. 2003. IRENE: a software to evaluate model performance. *European Journal of Agronomy, 18: 369-372.*

Fischer, A., S. Louahala, Ph. Maisongrande, L. Kergoat, and Dedieu, G. 1996. Satellite data for monitoring, understanding, and modeling of ecosystem functioning. In: Global Change and Terrestrial Ecosystems, edited by B. Walker and W. Steffen (Cambridge: Cambridge University Press), pp. 566-591.

Fischer, A., L. Kergoat, and G. Dedieu. 1997. Coupling satellite data with vegetation functional models: review of different approaches and perspectives suggested by the assimilation strategy. *Remote Sensing Review*, 15: 283-303.

Foody, G. M., P. J. Curran, G. B. Groom, and D. C. Munro. 1989. Multi-temporal airborne synthetic aperture radar data for crop classification. *Geocarto International,* 3: 19-29.

Friedl, M. A. and C. E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment,* 61(3): 399-409.

Friedl, M. A., C. E. Brodley, and A. H. Strahler. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions of Geoscience and Remote Sensing,* 37(2): 969-977.

Gamon, J. A., J. Peňuelas, and C. B. Field. 1992. A narrow-waveband spectral index that track diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment,* 41: 35– 44.

Gao, J. 1999. A comparative study on spatial and spectral resolutions of satellite data in mapping mangrove forests. *International Journal of Remote Sensing,* 20(14): 2823-2833.

Gates, D. M., H. J. Keegan, J. C. Schleter, and V. R. Wiedner. 1965. Spectral properties of plants. *Applied Optics*, 4: 11 – 20

Ghazanfari, A., J. Irudayaraj, and A. Kusalik. 1996. Grading pistachio nuts using a neural network approach. *Transactions of the ASAE,* 39 (6): 2319-2324.

Gilabert, M. A., S. Gandia, and J. Melia. 1996. Analysis of spectral-biophysical relationships for a corn canopy. *Remote Sensing of Environment*, 55: 50-58.

Gillett, A. G., N. M. J. Crout, D. T. Stokes, R. Sylvester-Bradley, and R. K. Scott. 2001. An approach to modeling the effect of environmental and physiological factors

upon biomass accumulation in winter wheat. *Journal of Agricultural Science,* 136: 369-381.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment,* 58:289-298.

Goel, P. K., S. O. Prasher, R. M. Patel, D. L. Smith, and A. DiTommaso. 2002. Use of airborne multi-spectral imagery for weed detection in field crops. *Transactions of the ASAE,* 45 (2): 443–449.

Goel, P. K., S. O. Prasher, J. A. Landry, R. M. Patel, and A. A. Viau. 2003a. Hyperspectral image classification to detect weed infestations and nitrogen status in corn. *Transactions of the ASAE,* 46 (2): 539–550.

Goel, P. K., S. O. Prasher, L. A. Landry, R. M. Patel, and A. A. Viau. 2003b. Estimation of crop biophysical parameters through airborne hyper-spectral remote sensing. *Transactions of the ASAE,* 46 (4): 1235–1246.

Goel, P. K., S. O. Prasher, R. M. Patel, J. A. Landry, R. B. Bonnell, and A. A. Viau. 2003c. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture,* 39: 67 -93.

Gong, P., R. Pu, G.S. Biging, and M.R. Larrieu. 2003. Estimationof forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, 41 (6):1355-1362.

Guérif, M., S. de Brisis, and B. Seguin. 1993. Combined NOAA-AVHRR and SPOT-HRV data for assessing crop yields of semi-arid environments. EARSeL *Advances in Remote Sensing,* 2: 110-123.

Guérif, M. and C. Duke. 1998. Calibration of SUCROS emergence and early growth module for sugar beet using optical remote sensing data assimilation. *European Journal of Agronomy,* 9: 127-136.

Gunn, S. 1998. Support vector machines for classification and regression. Technical report, ISIS, Department of Electronics and Computer Science, University of Southampton.

Guyot, G. 1996. Agriculture et statistiques agricoles, in Précis de télédétection: Tome 2, Applications thématiques, F. Bonn (ed.) (Sainte Foy (Québec): Presses de l'université de Queébec), pp. 269-316.

Haboudane, D., J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraz. 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment,* 81: 416–426.

Hamar, D., C. Ferencz, J. Lichtenberg, G. Tarcsai, and I. Frencz-Arkos. 1996. Yield estimation for corn and wheat in the Hungarian Great Plain using LANDSAT MSS data. *International Journal of Remote Sensing,* 17: 1689-1699.

Hansen, S., H. E. Jensen, N. E. Nielsen, and H. Svendsen. 1991. Simulation of nitrogen dynamics and biomass production in winter wheat using the Danish simulation model DAISY. *Fertilizer Research*, 27: 245-259.

Hansen, M., R. Dubayah, and R. DeFries. 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing,* 17: 1075-1081.

Hasegawa, H., D. C. Bryant, and R.F. Denison. 2000. Testing CERES model predictions of crop growth and N dynamics, in cropping systems with leguminous green manures in a Mediterranean climate. *Field Crops Research,* 67: 239–255.

Hatfield, J. K. 1983. Remote sensing estimators of potential and actual crop yield. *Remote Sensing of Environment,* 13: 301-311.

Hatfield, J. L., E. T. Kanemasu, G. Asrar, Pinter P.J., Jr., Reginato, R.J., and Idso S.B. 1985. Leaf area estimates from spectral measurements over various planting dates of wheat. *International Journal of Remote Sensing*, 6: 167–175.

Hayes, M.J. and W.L. Decker. 1996. Using NOAA AVHRR data to estimate maize production in the US corn belt. *International Journal of Remote Sensing*, 17:189-3200.

Holben, B. N. and C-J. Fan. 1980. Spectral assessment of soybean leaf area and leaf biomass. *Photogrammetric Engineering and Remote Sensing,* 4: 651-656.

Hoogenboom, G., H. Lai, and D.D. Gresham. 1993. Spatial yield prediction. *ASAE paper no. 93-3550, St. Joseph, MI.*

Hsu, C. W. and C. J. Lin. 2002. A comparison on methods for multi-class support vector machines, *IEEE Transactions on Neural Networks*, 13: 415-425.

Irmak, A., J. W. Jones, and W. D. Batchelor. 2000. Estimating spatially variable soil properties for application of crop models in precision agriculture. ASAE Paper No. 00-3037, St. Joseph, MI.

Jackson, R. D. and A. R. Huete. 1991. Interpreting vegetation indices. *Preventive veterinary medicine,* 11: 185-200.

Jamieson, P. D., J. R. Porter, J. Goudriaan, J. T. Ritchie, H. Van Keulen, and W. Stol. 1998. A comparison of the models AFRCWHEAT2, CERES-Wheat, SIRIUS, SUCROS2 and SWHEAT with measurements from wheat grown under drought. *Field Crops Research,* 55: 23-44.

JARS, 1999. Remote Sensing Notes. (Chapter 4.) *Japan Association of Remote Sensing.*

Johnson, D.E., 1998. Applied multivariate methods for data analysts. (Chapter 7. pp. 217-285), Brooks/Cole Publishing Company, USA.

Jones, C. A., and J. R. Kiniry. 1986. CERES–Maize: A simulation model of maize growth and development. College Station, Texas: Texas A&M University Press.

Karimi, Y., S. O. Prasher, H. McNairn, R. B. Bonnell, P. Dutilleul, and P. K. Goel. 2004. Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Transaction of the ASAE*, (in press).

Kenkel, N. C., D. A. Derksen, A. G. Thomas, and P. R. Watson. 2002. Review: Multivariate analysis in weed science research. *Weed Science,* 50:281–292.

Khakural, B.R., P.C. Robert, and D.J. Mulla. 1996. Relating corn/soybean yield to variability in soil and landscape characteristics. Proceedings of the Third International Conference on Precision Agriculture. ASA, CSSA, SSSA. Madison, WI. pp. 117-128.

Kiniry, J. R., J. R.Williams, R. L. Vanderlip, J. D. Atwood, D. C. Reicosky, J. Mulliken, W. J. Cox, H. J. Mascagni, S. E. Hollinger, and W. J. Wiebold. 1997. Evaluation of two maize models for nine U.S. locations. *Agronomy Journal,* 89:421–426.

Knezevic, S., S. F. Weise, and C. J. Swanton. 1994. Interference of Redroot Pigweed (*Amaranthus retroflexus*) in corn (*Zea mays*). *Weed Science,* 42: 568-573.

Knezevic, S. Z., M. J. Horak, and R. L. Vanderlip. 1997. Relative time of redroot pigweed (Amaranthus retroflexus L.) emergence is critical in pigweed-sorghum [Sorghum bicolor (L.) Moench] competition. *Weed Science,* 45: 502–508.

Kobayashi, K. and M. U. Salam. 2000. Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal,* 92: 345-352.

Kropff, M. J. and L .A. P. Lotz. 1992. Optimization of weed management systems: the role of ecological models of interplant competition. *Weed Technology,* 6: 462–470.

Kumar, M. and J. L. Monteith. 1981. Remote sensing of crop growth. In: Plants and the Daylight Spectrum, H. Smith (eds.), Academic Press, San Diego, California, 133 – 144.

Kurosu, T., S. Yokoyama, and M. Fujita. 2001. Land use classification with textural analysis and the aggregation technique using multi-temporal JERS-1 L-band SAR images. *International Journal of Remote Sensing,* 22 (4): 595–613.

Lajoie, P.G. 1960. Soil survey of Argenteuil, two mountains and Terrebonne counties, Quebec. Research Branch, Canada Department of Agriculture in Co-operation with Quebec Department of Agriculture and Macdonald Campus, McGill University.

Lamb, D.W., 1998. Opportunity for satellite and airborne remote sensing of weeds in Australian crops. In: Precision Weed Management in Crops and Pasture. Medd R.W., Pratley J.E. (Eds.), Proceedings of a workshop, 5-6 May, (CRC for Weed Management Systems, Adelaid), p. 154.

Leblon, B., M. Guérif, and F. Baret. 1991. The use of remotely sensed data in estimation of PAR use efficiency and biomass production of flooded rice. *Remote Sensing of Environment,* 38:147-158.

Lelong-Camille, C. D., P. C. Pinet, and H. Poilvé. 1998. Hyperspectral imaging and stress mapping in agriculture: a case study on wheat in Beauce (France). *Remote Sensing of Environment,* 66: 179-191.

Lemieux, C., L. Valle ´ E., and A. Vanasse. 2003. Predicting yield loss in maize fields and developing decision support for post-emergence herbicide applications. *Weed Research,* 43: 323–332.

Lillesand, T.M. and R.W. Kiefer. 2000. Remote Sensing and Image Interpretation. John Wiley & Sons. Inc. (4th edition).

Lorenzen, B. and A. Jenson. 1989. Change in leaf spectral properties indices in barley cereal powdery mildew. *Remote Sensing of Environment,* 27: 201-209.

Loudjani, P., R. Delécolle, M. Guérif, and P. Vossen. 1995. Combined use of NOAA/AVHRR and SPOT/HRV data for the estimation of European crop yields on a regional scale, In: International Colloquium Photosynthesis and Remote Sensing, G. Guyot (ed.), Montpellier, France, 28-30 August 1995, EARSel-ISPRS (Strasbourg: EARSel), pp. 327-332.

Maas, S. J. 1988a. Use of remotely sensed information in agricultural crop growth models. *Ecol. Model.* 41: 247-268.

Maas, S. J. 1988b. Using Satellite data to Improve Model Estimates of Crop Yield. *Agronomy Journal,* 80: 655-662.

Maas, S. J. 1991, Use of remotely sensed information in plant growth simulation models. *Advances in Agronomy,* 1: 17-26.

Maas, S. J. 1993a. Parameterized Model of gramineous crop growth: I. Leaf area and Dry mass simulation. *Agronomy Journal,* 85: 348-353.

Maas, S. J. 1993b. Parameterized Model of gramineous crop growth: II. Within season simulation calibration. *Agronomy Journal*, 85: 354-358.

Maas, S. J. 1993c. Within-Season Calibration of Modeled Wheat Growth Using Remote Sensing and Field Sampling. *Agronomy Journal,* 85: 669-672.

Maas, S. J., R. D. Jackson, S. B. Idso, P. J. Pinter Jr., and R. J. Reginato. 1989. Incorporation of remotely-sensed indicators of water stress in a crop growth simulation model. Proceedings of the 19th Conference on Agriculture and Forest Meteorology, 7-10 March 1989, Charleston, S. Carolina. pp. 228-231.

Maas, S. J. and P.C. Doraiswamy. 1996. Integration of satellite data and model simulations in a GIS for monitoring regional evaporation and biomass production. Proc., 3rd Int. Conf. on Integrating GIS and Environmental Modeling, Santa Fe, NM. CD-ROM., 1996, (Proc.).

Major, D. J., G. B. Schaalje, C. Wiegand, and B. L. Blad. 1992. Accuracy and sensitivity analyses of SAIL model-predicted reflectance of maize. *Remote Sensing of Environment,* 41: 61- 70.

Makowski, D., D. Wallach, and J. M. Meynard. 2001. Statistical methods for predicting responses to applied nitrogen and calculating optimal nitrogen rates. *Agronomy Journal*, 93:531–539.

Mamolos, A. P. and K. L. Kalburtji. 2001. Competition between Canada thistle and winter wheat. *Weed Science:* 49 (6): 755–759.

Marshall, E. J. P. 1988. Field-scale estimates of grass weed population in arable land. *Weed Research*, 28: 191-198.

Mayer, D. G., M. A. Stuart, and A. J. Swain. 1994. Regression of real-world data on model output: An appropriate overall test of validity. *Agricultural System,* 45:93–104.

McLennan, B. R., R. Ashford, and M. D. Devine. 1991. *Cirsium arvense* (L.) Scop. competition with winter wheat (*Triticum aestivum* L.). *Weed Research,* 31:409–415.

McNairn, H., J. C. Deguise, A. Pacheco, J. Shang, and N. Rabe. 2001. Estimation of crop cover and chlorophyll from hyperspectral remote sensing. 23rd Canadian Remote Sensing Symposium, Sainte-Foy, Québec, Canada, August 21-24, 2001.

Meyer, G. E., T. Mehta, M. F. Kocher, D. A. Mortensen, and A. Samal. 1998. Textural imaging and discriminant analysis for distinguishing weeds for spot spraying. *Transactions of the ASAE,* 41 (4): 1189–1197.

Meyer-Roux, J. 1990. Présentation de Projet Pilote de Télédétection Appliquée aux Satistiques Agricoles. Proceedings of the conference on the application of remote sensing to agricultural statistics, Varese, Italy, October 10-11, 1989, (EUR 12581 EN), pp. 25- 38.

Michaud, N., K. P. B. Thomson, J. Boisvert, and A. Viau. 1998. Extraction of crop information using multitemporal RADARSAT imagery and integration in the SWAP crop growth model. Proceeding of the 2nd International Workshop on the Retrieval of Bio- & Geo-Physical Parameters from SAR Data for Land Applications, ESTEC, Noordwijk, Holland, October, 21-23, p. 149-154.

Miller, J. R., W. U. Jiyou, M.G. Boyer, M. Belanger, and E.W. Hare. 1991. Seasonal patterns in leaf reflectance red edge characteristics. *International Journal of Remote Sensing,* 12: 1509–1523.

Monteith, J. L. 1977. Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London,* B281, 277-294.

Moran, M. S., T. R. Clarke, Y. Inoue, and A. Vidal 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation Index. *Remote Sensing of Environment,* 49: 246-263.

Moran, M. S., A. Vidal, D. Troufleau, J. Qi, T. R. Clarke, P. J. Jr. Pinter, T. A. Mitchell, Y. Inoue, and C. M. U. Neale. 1997a. Combining multi-frequency microwave and optical data for crop management. *Remote Sensing of Environment,* 61: 96-109.

Moran, M. S., Y. Inoue, and E. M. Barnes. 1997b. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment,* 61: 319-346.

Moulin, S., A. Fischer, G. Dedieu, and R. Delécolle. 1995. Temporal variations in satellite reflectances at field and regional scales compared with values simulated by linking crop growth and SAIL models. *Remote Sensing of Environment,* 54: 261-272.

Moulin, S., A. Bondeau, and R. Delécolle. 1998. Review article: Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing,* 19: 1021-1036.

Moulin, S. and M. Guérif, 1999. Impacts of model parameter uncertainties on crop reflectance estimates: a regional case study on wheat. *International Journal of Remote sensing,* 20: 213- 218.

Murai, H., and S. Omatu. 1997. Remote sensing image analysis using a neural network and knowledge-based processing. *International journal of remote sensing,* 18: 811- 828.

Murray, G. D. 1977. A cautionary note on selection of variables in discriminant analysis. *Applied Statistics,* 26 (3): 246–250.

Murthy, C. S., S. Thiruvengadachari, P. V. Raju, and S. Jonna. 1996. Improved ground sampling and crop yield estimation using satellite data. *International Journal of Remote Sensing,* 17: 945-956.

Myneni, R. B., F. G. Sellars, and A. L. Marshak. 1995. optical remote sensing of vegetation: modeling, caveats and algorithms. *Transaction of Geosciences and Remote Sensing,* 33:481-486.

Nagendra H. 2001. Review article; Using remote sensing to assess biodiversity. *International Journal of Remote Sensing,* 22: 2377–2400

Ngouajio, M., G. D. Leroux, and C. Lemieux. 1999a. A flexible sigmoidal model relating crop yield to weed relative leaf cover and its comparison with nested models. *Weed Research,* 39: 329-343.

Ngouajio, M, C. Lemieux, and G. D. Leroux. 1999b. Prediction of corn (Zea mays) yield loss from early observations of the relative leaf area and the relative leaf cover of the weeds. *Weed Science*, 47: 297–304.

Oerke, E.- C., Dehne, H.- W., Schönbeck, F. and Weber, A. 1994. Crop production and crop protection: Estimated losses in major food and cash crops. Elsevier, Amsterdam, The Netherlands. 808 pp.

O'Neill, N. T., F. Zagolski, M. Bergeron, A. Royer, J. Miller, and J. Freemantle. 1997. Atmospheric correction validation of CASI images acquired over the BOREAS Southern study area. *Canadian Journal of Remote Sensing* 23(2): 143-162.

Pacheco, A., A. Bannari, J. C. Deguise, H. McNairn and K. Staenz. 2001. Application of hyperspectral remote sensing for lai estimation in precision farming. 23rd Canadian Remote Sensing Symposium, Sainte-Foy, Québec, Canada, August 21-24, 2001.

Patel, N. K., C. Patnaik, S. Dutta, A. M. Sekh, and A. J. Dave. 2001. Study of crop growth parameters using Airborne Imaging Spectrometer data. *International Journal of Remote Sensing*, 22: 2401–2411.

Paz, J. O., W.D. Batchelor, B.A. Bakcock, T.S. Colvin, S.D. Logsdon, T.C. Kaspar, and D.L. Karlen. 1999, Model based technique to determine variable rate nitrogen for corn. *Agricultural Systems,* 61:69-75.

Paz, J. O. 2000. Analysis of spatial yield variability and economics of precision agriculture: a crop modeling approach. PhD Diss. Iowa State University, Ames, Iowa.

Paz, J. O., W. D. Batchelor, T. S. Colvin, S. D. Logsdon, T. C. Kaspar, and D. L. Karlen, 1998. Analysis of water stress effects causing spatial yield variability in soybeans. *Transactions of ASAE,* 41(5): 1527-1534.

Pearson, R., J. Grace, and G. May. 1994. Real time airborne agricultural monitoring. *Remote Sensing of Environment*, 49: 304-310.

Peñuelas, J. and I. Filella. 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science,* 3: 151–156.

Peñuelas, J., I. P. Fiella, F. Muñoz, and M. Vilajeliu. 1995. Reflectance assessment of mite effects on apple trees. *International Journal of Remote Sensing,* 16:2727-2733.

Perry, C. R. Jr., and L. R. Lautenschlager. 1984. Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment,* 14: 169-182.

Plummer, S. E., A.K. Wilson, and A. K. Jones. 1991. On the relationship between high spectral resolution canopy reflectance data and plant biochemistry. *EARSeL Advances in Remote Sensing,* 1: 27–33.

Porter, J. R. 1984. A model of canopy development in winter wheat. *Journal of Agricultural Sciences, Camb.*, 102: 383-392.

Prévot, L., I. Champion, and G. Guyot. 1993. Estimating surface soil moisture and leaf area index of a wheat canopy using a dual-frequency (C and X bands) scatterometer. *Remote Sensing of Environment,* 46: 331-339.

Prévot, L., H. Chauki, D. Troufleau, M. Weiss, F. Baret, and N. Brisson. 2000a. Assimilating Optical and Radar Data into the STICS Model for Wheat Crops: Preliminary Results. European Geophysical Society (EGS) XXV General Assembly, Nice, France, April, 28, 2000.

Prévot, L., H. Chauki, D. Troufleau, M. Weiss, F. Baret, and N. Brisson. 2000b. Coupling the STICS canopy functioning model to radiative transfer models for assimilating remote sensing data in the solar and microwave domains. European Geophysical Society (EGS) XXV General Assembly, Nice, France, April, 28, 2000.

Price, J. C. and W.C. Bausch. 1995. Leaf area index estimation from visible and near-infrared reflectance data. *Remote Sensing of Environment,* 52: 55-65.

Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya Ser. A,* 26:329–358.

Rastogi, A., N. Kalra, P. K. Agarwal, S. K. Sharma, R. C. Harit, R. R. Navalgund, and V. K. Dadhwal. 2000. Estimation of wheat leaf area index from IRS LISS-III data using Price model. *International Journal of Remote Sensing,* 21: 2943-2949.

Retta, A., R. L. Vanderlip, R. A. Higgins, and L. J. Moshier. 1996. Application of SORKAM to simulate shattercane growth using forage sorghum. *Agronomy Journal*, 88:596–601.

Ritchie, J. T. and S. Otter. 1985. Description and performance of CERES-Wheat: A user-oriented wheat yield model. *USDA-ARS, ARS*-38, p.159-175.

Ritchie, J. T., U. Singh, D. C. Godwin, and T. Hunt. 1989. A User's Guide to CERES Maize-V2.10. International Fertilizer Development Centre, Muscle Shoals, AL.

Robert, P. C., R.H. Rust, and W.E. Larson (Eds). 1995. Site-Specific Management for Agricultural Systems. 27-30 March 1995, Minneapolis, MN, ASA-CSSA-SSSA, Madison, WI, xiii-xiv.

Roerink, G. l., W. G. M. Bastiaanssen, J. Chambouleyron, and M. Menenti. 1997. Relating crop water consumption to irrigation water supply by remote sensing, *Water Resources Management,* 11(6): 445-465.

Ruíz-Nogueira, B., K. J. Boote, and F. Sau. 2001. Calibration and use of CROPGRO-soybean model for improving soybean management under rainfed conditions. *Agricultural Systems*, 68:151-173.

Saha, S. K. and S. Jonna. 1994. Paddy acreage and yield estimation and irrigated crop land inventory using satellite and agro-meteorological data. *Asian Pacific Remote Sensing Journal,* 2: 78-87.

Sakthivadivel, R., S. Thirnvellgadachari, U. Amerasinghe, W. G. M. Bastiaanssen, and D. l. Molden. 1999. Performance evaluation of the Bhakra irrigation system, India, using remote sensing and GIS techniques, Research Report 28, International Water Management Institute, Colombo, Sri Lanka: 22 pp.

Sannier, C. A. D. and J. C. Taylor. 2002. Real-time monitoring of vegetation biomass with NOAA-AVHRR in Etosha National Park, Namibia, for fire risk assessment. *International Journal of Remote Sensing,* 23: 71–89.

SAS. 2003. *SAS User's Guide: Statistics.* Ver. 8.2. Cary, N.C.: SAS Institute, Inc.

Seguin, B., J-P. Lagouarde, and M. Savane. 1991. The assessment of regional crop water conditions from meteorological satellite thermal infrared data. *Remote Sensing of Environment,* 35: 141-148.

Seguin, B., D. Courault, and M. Guérif. 1994. Surface temperature and evapotranspiration: application of local scale methods to regional scales using satellite data. *Remote Sensing of Environment,* 49: 287-295.

Seidl, M. S., J. O. Paz, and W. D. Batchelor. 2000. Integrating remotely sensed images to improve spatial crop model calibration. 93th, *ASAE Paper No. 00-3039*, St. Joseph, MI.

Sharman, M. J., H. de Boissezon, G. Gonzales, and K. Pouss. 1992. Rapid estimates of acreages and potential yield at European scale. Proceedings of the 43rd congress of the International Astronautical Federation. Washington, DC, U.S.A., 28 August- 5 September 1992. IAF, Paper 92-0110 (Washington DC: IAF).

Simane, B., H. van Keulen, W. Stol, and P.C. Struik. 1994. Application of a crop growth model (SUCROS-87) to assess the effect of moisture stress on yield potential of durum wheat in Ethiopia. *Agricultural Systems*, 44 (3) 337-353.

Sinclair, T.R. and R.C. Muchow. 1999. Radiation use efficiency. *Advance in Agronomy*, 65:215-265.

Sinclair, T. R. and N. G. Seligman. 1996. Crop modeling: from infancy to maturity. *Agronomy Journal,* 88:698-704.

Singh, R., D. P. Semwal, A. Rai, and R. S. Chhikara. 2002. Small area estimation of crop yield using remote sensing satellite data. *International Journal Remote Sensing,* 23: 49–56

Slaughter, D. C., W. T. Lanini, and D. K. Giles. 2003. Discriminating Weeds from Processing Tomato Plants using Visible and Near-Infrared Reflectance. *ASAE Paper No: 031108.* Las Vegas, Nevada, USA 27- 30 July 2003.

Soh, L. K. 1999. Segmentation of satellite imagery of natural scenes using data mining. *IEEE Transactions on Geoscience and Remote Sensing,* 37 (2): 1086-1099.

Spitters, C. J. T., H. van Keulen, and D. W. G. van Kraalingen. 1989. A simple and universal crop growth simulator: SUCROS87. In: Simulation and Systems Management in Crop Protection, R. Rabbinge, S. A. Ward, and H. H. van Laar (eds.), Simulation Monographs 32 (Wageningen: Pudoc), pp. 147-181.

Steven, M.D., P.V. Biscoe, and K. W. Jaggard. 1983. Estimation of sugar beet productivity from reflection in the red and infrared spectral bands. *International Journal of Remote Sensing,* 4: 325-334.

Strachan, I. B., H. Pattey, and J. B. Boisvert. 2002. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sensing of Environment,* 80: 213–224.

Teo, Y. H., C. A. Beyrouty, and E. E. Gbur.1992. Evaluating a model for predicting nutrient uptake by rice during vegetative growth. *Agronomy Journal,* 84:1064–1070.

Terawaki, M., T. Kataoka, H.Okamoto, and S. Hata. 2002. Distinction between sugar beet and weeds for development of automatic thinner and weeding machine of sugar beet. *ASAE Paper No. 701P0502.* St. Joseph, MI.

Thenkabail, P. S., R. B. Smith, and E. De Pauw. 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing,* 68 (6): 607-622.

Thenkabail, P. S., E. A. Enclona, M. S. Ashton, and B. Van der Meer. 2003. Accuracy assessments and optimal hyperspectral wavebands for vegetation and agriculture in 400-2500 nanometers. Available from: http://www. Geology.yale.edu/~smith/optimal_hyper_bands.pdf; Date Accessed: January, 2003.

Tolman J. H., D. G. R. McLeod, and C. R. Harris. 2004. Cost of crop losses in processing tomato and cabbage in southwestern Ontario due to insects, weeds and/or diseases. *Canadian Journal of Plant Science,* 84: (3) 915-921.

Tomer, M. D., J. L. Anderson, and J. A. Lamb. 1997. Assessing corn yield and nitrogen uptake variability with digitized aerial infrared photographs. *Photogrammetric Engineering & Remote Sensing, 63 (3): 299-306.*

Towner, M., A. Peters, A. Viña, L. Ji, and B. Waltman. 2000. Evaluating the Use of Multi-temporal High-resolution Airborne Digital Imagery for Optimizing Crop-growth/yield Models for Corn and Soybeans. Report prepared for: the Center for Advanced Land Management Information Technologies, University of Nebraska-Lincoln and Commercial Remote Sensing Program Office.

Travasso, M. I. and G. O. Magrin. 1998. Utility of CERES-Barley under Argentine conditions. *Field Crops Research,* 57: 329–333.

Tuker C. J., B. N. Holben, J. H. Elgin, and J. E. Mcmurtrey III. 1981. Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sensing of Environment* 11: 171-189

Tucker, C. J., C.L. Vanpraet, M. J. Sharman, and G. Van Ittersun. 1985. Satellite remote sensing of total herbaceous production in the Senegalese Sahel 1980-1984. *Remote Sensing of Environment,* 17: 232-249.

Vangessel, M. J., E. E. Schweizer, K. A. Garrett, and P. Westra. 1995. Influence of weed density and distribution on corn (*Zea mays*) yield. *Weed Science*, 43:215-218.

Van Diepen, C. A., J. Wolf, H. Van Keulen, and C. Rappoldt. 1989. WOFOST: a simulation model of crop production. *Soil Management,* 5: 16-24.

Vapnik, V. 1995. The Nature of Statistical Learning Theory, *Springer Verlag.*

Vrindts, E., J. De Baerdemaeker, and H. Ramon. 2002. Weed detection using canopy reflection. *Precision Agriculture,* 3: 63–68.

Wallace, A. 1994. High-precision agriculture is an excellent tool for conservation of natural resources. *Communications Soil Science Plant Analysis,* 25 (1&2): 45-49.

Weir, A. H., P.L. Bragg, J. R. Proter, and J. H. Rayner. 1984. A winter wheat crop simulation model without water or nutrient limitations. *Journal of Agricultural Science,* 102: 371-382.

Welch, S. M., J. W. Jones, G. Reeder, M. W. Brennan, and B. M. Jacobson. 1999. PCYield: Model-based support for soybean production. Proceedings of the

international Symposium on Modeling Cropping Systems, June 21-23, Catalina, Spain, P. 273-275.

Wiegand, C. L., A. J. Richardson, R. D. Jackson, P. J. Pinter Jr., J. K. Aase, D. E. Smika, L. F. Lautenschlager, and H. J. E. McMurtrey. 1986. Development of agrometeorological crop model inputs from remotely sensed information. *IEEE, Transactions on Geoscience and Remote Sensing GE,* 24: 90- 97.

Wiegand, C. L. and A. J. Richardson. 1990. Use of Spectral Vegetation Indices to Infer Leaf Area, Evapotranspiraiton and Yield: I. Rationale, *Agronomy Journal,* 82: 623-629.

Wilcut, J. W., G. R. Wehtje, and R. H. Walker. 1987. Economics of weed control in peanuts (*Aracbis bypogaea*) with herbicides and cultivations. *Weed Science,* 35: 711-715.

Wiles, L. J., G. G. Wilkerson, H. J. Gold, and H. D. Coble. 1992. Modeling weed distribution for improved postemergence control decisions. Weed Science, 40: 546-553.

Wilkerson, G. G., J. W. Jones, K. J. Boote, K. T. Ingram, and J. W. Mishoe. 1983. Modeling soybean growth for crop management. *Transactions of the ASAE,* 26: 63-73.

Williams, J. R., C. A. Jones, J. R. Kiniry, and D. A. Spanel. 1989. The EPIC crop growth model. *Transactions of the ASAE,* 32 (2): 497-511.

Woolly, J. T. 1971. Reflectance and transmittance of light by leaves. *Plant Physiology,* 47: 656-662.

Xie, Y., J. R. Kiniry, V. Nedbalek, and W. D. Rosenthal. 2001. Maize and sorghum simulations with CERES-Maize, SORKAM, and ALMANAC under water-limiting conditions. *Agronomy Journal,* 93:1148-1155.

Xu, H., M. D. Steven, and K. W. Jaggard. 1996. Monitoring leaf area of sugar beet using ERS-1 SAR data. *International Journal of Remote Sensing*, 17: 3401–3410.

Yang, C., J. H. Everitt, and J. M. Bradford. 2002. Airborne hyperspectral imaging and yield monitoring of grain sorghum yield variability. *ASAE Paper No. 02-1079.* St. Joseph, MI.

Yang, C.-C., S. O. Prasher, and J. A. Landry. 1999. Use of artificial networks to recognize weeds in corn field. Journée d'information Scientifique et Technique en genie agroalimentaire, Saint-Hyacinthe, QC, Canada, March 3, 1999, pp. 60-65. Conseil des production végétales du Québec, Inc.

APPENDIX A

STICS model and it's modules

The main objective of the STICS model is to simulate the effect of climatic conditions, soil and crop practices on agricultural production (quality and quantity) and the environment. STICS was developed as a generic model, easily adjustable to different plant types. This comes from the generic nature of the formalism that was chosen.

STICS simulates daily growth over a period from emergence to harvest on the basis of crop aboveground biomass (carbon and nitrogen), leaf area index and soil parameters and climatic data. The model simulates the growth of crop state variables (like LAI and biomass) in a daily time step based on the input data set. The soil is divided into different homogenous layers. For simulation purposes, each layer is defined by its water and nitrogen content. Interaction between plant and soil is done by roots, which are defined by their length distribution in the soil profile. In the STICS model, crop growth simulation is performed on the basis of carbon and nitrogen balance. As a result, this will lead to computing agricultural production (biomass, yield) and environmental impact (water and nitrogen leaching) under different agricultural conditions.

To simulate crop growth, the model contains seven different modules, namely: development, shoot growth, yield component, root growth, water balance, thermal environment and nitrogen balance (Brisson *et al.*, 1998). A brief on the modules is presented here, along with a summary of various parameters and variables (and their values) for each module in different tables. The FORTRAN notations, which are used to define the variables, are presented in these tables.

A.1 Development

In the development module of STICS, crop growth period is divided into seven stages (Table 5.1) for the entire growing season. These stage classifications differ from the ones described in agronomy. The stages are defined in a way that is useful for simulating LAI and grain filling and they may differ from agronomical stages. Dividing the growth period to various stages makes it possible to simulate different crops or species.

In this module, the simulation will start by determining the seed germinations, which is related to the soil temperature at sowing depth with the sum of degree-days (equation A.1), and it will occur when this sum meets a defined threshold value for the specified crop. The growth rate of the germinated plant is also related to the soil temperature and sowing depth by equation A.2.

$$STPLTGER = \sum_{I=IPLT}[(TSOL(PROFSEM) - TGMIN) \times PFZ(PROFSEM)] \qquad A.1$$

$$ELONG = ELMAX[1 - \exp(-(BELONG \times \sum_{I=IGER} TSOL(PROFSEM) - TGMIN)^{CELONG})] \qquad A.2$$

where:

BELONG and ***CELONG*** are parameters of the curve of coleoptile elongation (degree-days)

ELMAX is maximum elongation of the coleoptile in dark conditions (cm)

ELONG is coleoptile elongation (cm)

PROFSEM is sowing depth (cm)

PFZ is water status of the soil (0 or 1)

STPLTGER is sum of development units (degree-days) allowing germination

TGMIN is minimum threshold temperature used in the emergence stage

TSOL is temperature of the soil

In earlier crop growth models (i.e. CERES, ARCWHEAT, and SUCROS) air temperature was used to determine the sowing-emergence phase occurrence. Afterward, the duration of the emergence was calculated on the base of soil water status. In STICS, duration of the emergence is simulated by the three factors: temperature, water status of the soil and sowing depth. Previous works showed that replacing the air temperature by a temperature near to the plant could give a better simulation of crop development. So far, STICS development was related to the surface temperature, which makes it possible to link the effect of drought (water deficit) on development to the temperature.

In earlier crop growth models (i.e. CERES, ARCWHEAT, and SUCROS) air temperature was used to determine the sowing-emergence phase occurrence. Afterward, the duration of the emergence was calculated on the base of soil water status. In STICS, duration of the emergence is simulated by the three factors: temperature, water status of the soil and sowing depth. Previous works showed that replacing the air temperature by a temperature near to the plant could give a better simulation of crop development. So far, STICS development was related to the surface temperature, which makes it possible to link the effect of drought (water deficit) on development to the temperature.

A.2 Shoot Growth

In the shoot growth module, the development of LAI is performed in four stages; including two growth stages (ILEV to ILAX), one stability stage (ILAX to ISEN) and one senescence stage (ISEN to IMAT). Since, in the model, the radiation is mainly intercepted by the canopy, it is important to have an accurate estimation of LAI, an indicator of canopy, at various growth phases. For the stability phase, where interception has its maximum value, accuracy has less importance. In the simulation processes, from emergence stage (ILEV) up to maximum LAI stage (ILAX), LAI is calculated with a non-linear equation and for the period between maximum LAI and beginning of senescence it was assumed that LAI has a constant value while for senescence period LAI decrease linearly.

Between ILEV and ILAX the net leaf growth is calculated (equation A.3) using a net leaf development unit (ULAI), which includes values of 1, 2.2 and 3 for the ILEV, juvenile stage (IMAF) and ILAX phase, respectively. A linear interpolation is used to compute the ULAI values between these numbers. A density is a defined target, below which there is no competition between the plants and above which the density will affect the leaf surface resulting in decreased leaf surface.

$$DELTAI = \frac{DLAIMAX}{1+\exp(5.5(2.2-ULAI))} \times (TCULT - TCMIN) \times \min(TURFAC, INNS) \times EFDENSITE \times DENSITE \qquad A.3$$

Where:

DELTAI is daily increase in LAI (m^2 leaves m^{-2} soil surface day^{-1})

DENSITE is sowing density (plants m^{-2})

DLAIMAX is maximum rate of LAI development (m^2 leaves m^{-2} soil degree day^{-1})

EFDENSITE is the density effect acting on the setting up of LAI

INNS is nitrogen stress index (between 0 and 1)

TCMIN is the minimum temperature at which growth will occur (C°)

TCULT is the daily mean surface temperature (C°)

TURFAC is turgescence stress index (between 0 and 1)

ULAI is physiological time units for the calculation of the LAI between ILEV and ILAX

In this module, the accumulation of aboveground biomass (DLTAMS) is related to intercepted radiation using a parabolic equation (equation A.4) with a maximum radiation use efficiency which has different values for various crops during different growing stages. Beer's law is the base for the estimation of the incidental radiation intercepted by the crop (equation A.5).

$$DLTAMAS \times 100 = [EBMAX \times RAINT - 0.0815 \times RAINT^2] \times f(TCULT) \times SWFAC \times INNS \qquad A.4$$

$$RAINT = 0.95 \times 0.50 \times TRG[1 - \exp(-EXTIN \times LAI)] \qquad A.5$$

Where:

DLTAMAS is the crop growth rate (Mg ha^{-1} day^{-1})

EBMAX is maximum radiation use efficiency (Mg ha^{-1} MJ^{-1} m^{-2})

EXTIN is extinction coefficient of photosynthetic active radiation in the canopy

RAINT is photosynthetic active radiation intercepted by the crop canopy (MJ m^{-2})

SWFAC is the stomatal stress index (between 0 and 1)

TRG is global solar radiation (MJ m^{-2} day^{-1})

DLAMS will be affected individually by the water and nitrogen stresses. Final aboveground biomass is the total of daily DLAMS.

A.3 Yield Components Module

In the yield components module, a period of variable duration is set to calculate grain number, based on the mean growth rate of the species canopy during this period (equation A.6). The total dry matter and the nitrogen content of the grain are related to the above ground biomass by applying some harvest indices (nitrogen and carbon harvest indices). These indices are used to calculate the magnitude of grains and the quantity of nitrogen in grains (equation A.7).

$$NBGRAINS = CGRAIN \times VITMOY + CGRAINVO \qquad A.6$$

$$MAGRAIN = IRCARB \times (I - IDRP) \quad AND \quad IRAZO = VITIRAZO \times (I - IDRP) \qquad A.7$$

Where:

CGRAIN is slope of the relationship between grain number and growth rate during the NBJGRAINS before stage IDRP (grains g dry matter $^{-1}$ day^{-1})

CGRAINVO is number of grains produced when the growth rate is zero (grains m^{-2})

I is the running day

IRAZO is nitrogen harvest index (g N grain g N plant^{-1})

IRCARB is carbon harvest index (g C grain g C plant $^{-1}$)

MAGRAIN is the dry matter of the grain (g m^{-2})

NBGRAINS is grain number (grains m^{-2})

VITMOY is mean daily growth rate while setting up the grain number, NBJGRAIN before the stage IDRP (g m $^{-2}$ day $^{-1}$)

VITIRAZO is rate of increase of the nitrogen harvest index (g N grain g N plant $^{-1}$ day $^{-1}$)

Harvest indices are related to the length of time between IDRP and IMAT with an increasing linear function, while the effect of water and nitrogen stress are not considered in the evolution of the harvest index.

A.4 Root Growth

In this module, root growth is defined in term of their length, but is not directly related to the crop's above ground biomass. In fact, root biomass is distributed based on the radiation use efficiency determined, which is calculated on the basis of above ground

foliage with a parameter MSAERO (describing the threshold of above ground dry matter corresponding to the root allocation at the beginning of the cycle) early in the growth period. In the simulation process, water and nitrate are being taken from the soil by the root. The root development front depends on the crop variety, crop temperature, soil depth, and water availability. In the simulation, the active root front rate is related to the soil water conditions and the crop temperature (equation A.8).

$$DELTAZ = CROIRAC \times [\min(TCULT, TCOPT) - TCMIN] \times PFZ(ZRAC) \qquad A.8$$

where:

CROIRAC is the growth rate of the root front (cm degree-day^{-1})

DELTAZ is deepening of root front (cm)

PFZ (Z) is the water status of the soil at layer Z (=0 if HUMIN, otherwise =1)

TCOPT is optimum temperature of the growth (^{0} C)

ZRAC is root depth (cm)

Root growth front begins at sowing depth and limits either at ILAX stages or the maximum soil depth. For each depth the effective root density is determined by the module (equation A.9).

$$LRAC(Z) = \frac{LVOPT}{1 + \exp(-S(Z - ZDEMI))} \qquad A.9$$

Where:

LARC is the effective root density in the Z layer of the soil (cm root cm^{-3} soil)

LVOPT is optimum root density (cm root cm^{-3} soil)

S is a parameter for calculating root density

ZDEMI shows the depth where the root density is half of the surface root density

S and ZDEMI can be defined by equation A.10.

$$S = \frac{-4.6}{ZLABOUR - ZPENTE} \quad and \quad ZDEMI = \max(ZRAC - ZPRLIM + ZPENTE, \frac{1.4}{S}) \qquad A.10$$

Where:

ZLABOUR is plowing depth (cm)

ZPENTE is the depth where the root density is half of the surface root density for the reference profile (cm)

ZPRLIM defines the maximum root density for reference profile (cm)

A.5 Water Balance

In the water balance module, Penman's reference evapotranspiration equation is used to create an equation to calculate evaporation from the soil based on the availability of the energy at soil level and the maximal transpiration. By first using Beer's law, the potential evaporation from the soil is estimated (equation A.11). The water accessibility based on actual evaporation is then calculated.

$$EOS = TETP \times \exp[-(EXTIN - 0.2) \times LAI] \qquad A.11$$

where:

EOS is maximum evaporation flux (mm day $^{-1}$)

EXTIN is extinction coefficient of photosynthetic active radiation in the canopy

TETP is reference evaporation (mm day $^{-1}$)

To calculate the maximum transpiration from the crop, it is first assumed that there is no water scarcity in the system (soil and plant). The evaporation in the crop (EO) is determined by equation A.12, which is related to the LAI by a crop coefficient (KMAX). Maximal transpiration is obtained by subtracting EOS from EO (equation A.13).

$$EO = TETP[1 + \frac{KMAX - 1}{1 + \exp(-1.5LAI - 3)}] \qquad A.12$$

$$EOP = (EO - EOS) \times (1.4 - 0.4\frac{EO}{EOS}) \qquad A.13$$

where:

EO is an intermediary variable for the calculation of evapotranspiration (mm day^{-1})

EOP is the maximum transpiration flux (mm day $^{-1}$)

KMAX is the maximum crop coefficient

The actual transpiration (EP) is related to the maximal transpiration by a variable (TETA), which has a value between 0 and 5 and its value depends on water availability in the soil. The value of TETA is equal to 0 when the soil water content is at the wilting

point. In the daily time step simulation, it is assumed that the transpiration from the plant is equal to the root absorption.

Using these evaporation, irrigation and precipitation values, a water balance is computed, with the assumption that in each layer the soil characteristics (field capacity, permanent wilting point and bulk density) are constant. The soil profile can be defined in five layers of varying depths. Each layer acts as a reservoir with the maximum and minimum capacity equal to field capacity and permanent wilting point respectively. Water balance is calculated on the base of a 1 cm sub-layer, which may not be necessary. However, this method of calculation is accurate for nitrogen and it is important for nitrate transfer. To calculate evapotranspiration based on the energy availability, first the potential evaporation is calculated. In the second step, using the potential evaporation and available water for crop usage, actual evaporation is determined.

A.6 Crop Thermal Environment

Variation in soil temperature not only depends on the surface conditions but also on thermal inertia of the surroundings. This inertia is lowers soil temperature with greater depths. In the calculation of the soil temperature, daily thermal amplitude at the depth (z), which is related to soil surface thermal amplitude, is used. In a thermal environmental module, simulation of the activities inside the soil are based on daily average surface temperature (TCULT) or the temperature of the soil (TSOL) in different layers, which are calculated by equations A.14 and A.15.

$$TCULT = TMOY + \frac{(RN - (EP + ES) \times 2.46)RA}{0.0864 \times 1200} \qquad A.14$$

$$TSOL(Z) = \frac{AMPLZ}{AMPLSURF}(TCULTVEILLE - TMIN) + 0.1 \times (TCULTVEILLE - TSOLVEILLE(Z) + \frac{AMPLZ}{2}) \qquad A.15$$

where:

AMPLSURF is the thermal amplitude at the soil surface (C°)

AMPLZ is the amplitude at the depth Z (C°)

EP is actual transpiration flux (mm day $^{-1}$)

ES is actual soil evaporation flux (mm day $^{-1}$)

RA is aerodynamic resistance (s m^{-1})

RN is net solar radiation (MJ m^{-2})

TCULTVEILLE is daily mean surface temperature on the previous day (C°)

TMIN is minimum temperature (0 C)

TMOY is mean temperature (0 C)

TSOLVEILLE is temperature of the soil at the depth Z, on the day before (C°)

The amount of RN is related to LAI and calculated using equation A.16 and the value of RA, which is used on a daily average base, is assumed constant (30 s m^{-1}).

$$RN = [1 - 0.23 - (0.23 - ALBSOL) \times \exp(-0.75LAI)] \times TRG \times 0.72 - 0.9504 \qquad A.16$$

Where ALBSOL is the bare dry soil albedo and TRG is global solar radiation (MJ m^{-2} day^{-1}).

A.7 Nitrogen Balance

In STICS the nitrogen that is available to the crop comes from two sources: fertilizer applied and mineralized nitrogen. Supplied nitrogen in the form of minerals is artificial nitrogen fertilizer and nitrogen from irrigation and precipitation. Three sources of nitrogen mineralization, which are taken into account, are crop residues (RES), humified organic matter (HUM), and the microbial biomass growing on organic matter (BIOM). The humification will occur up to a depth at least equal to the plowing depth. The potential rate of mineralization (which depends on the clay and limestone content of the soil), the thermal factor, and the water content has an effect on mineralization processes (equation A.17 and A.18) in each soil layer.

$$VMINH(Z) = VPOT(Z) \times FTH(Z) \times FH(Z) \qquad A.17$$

$$VPOT(Z) = K2POT \times NHUM \qquad A.18$$

where:

FH is the soil water content, a correcting factor for the calculation of humus mineralization

FTH is the temperature correction factor for the calculation of humus mineralization

NHUM is the total amount of humus nitrogen in the soil (kg N ha^{-1})

VMINH is the rate of humus mineralization in each soil layer (kg N ha^{-1} cm^{-1} day^{-1})

VPOT is the potential rate of mineralization per layer at the reference temperature and water content (kg N ha^{-1} cm^{-1} day^{-1})

Mineralization of nitrogen and microbial biomass growth is the result of crop residues break down. Thus, the rot of microbial biomass will result in humus. The quality of the residue, temperature and the water availability of the soil determine the rottenness rate of the residue and fatality of the microflora. The C/N ratio of the crop residues defines their quality, which in turn determines the humification rate of residues.

The fraction of the applied nitrogen which remains in its mineral form is called "efficiency of the nitrogen fertilizer" (EFFN). In STICS, it is assumed that EFFN is equal to two thirds of applied nitrogen, while the rest of the fertilizer nitrogen (one third) is lost in its gaseous form.

In each single soil layer, it is assumed that the transferred nitrogen from the upper layer is mixed with the water in the layer and excess water is drained to the next layer. This process will continue until the end of soil profile or where the soil water content is less than its water holding capacity.

Comparing the nitrogen demand of the crop and the nitrogen supply of the soil will result in the nitrogen absorption of the crop. In the case of limiting the nitrogen supply, the actual nitrogen absorption is equal to the nitrogen supply. If there is no shortage in the nitrogen supply, maximum nitrogen demand (equation A.19) of the crop for daily product will be the actual nitrogen absorption rate.

$$\begin{aligned} &\text{if } MASEC \geq 1, \quad DEMANDE = 10 \times ADILMAX \times DLTAMS \times (1 - BDILMAX) \times MASEC^{BDILMAX} \\ &\text{if } MASEC < 1, \quad DEMANDE = 10 \times ADILMAX \times DLTAMAS \end{aligned} \qquad A.19$$

where:

ADILMAX and ***BDILMAX*** are the parameters of the maximum curve of nitrogen demand (g N g dry $matter^{-1}$)

DEMANDE is the daily nitrogen need of plant (kg N ha^{-1} day^{-1})

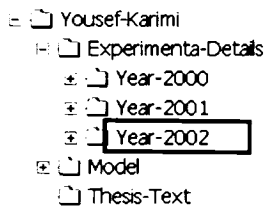
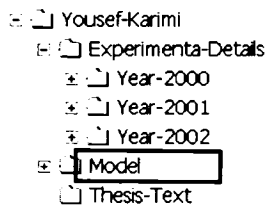
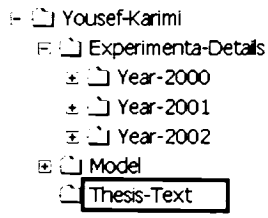
DLTAMAS is the growth rate of the plant (t ha^{-1} day^{-1})

MASEC is above ground dry matter (t ha^{-1})

APPENDIX B

Details of the content of the CD-ROM are as follows:

Directory/Subdirectory	Content
Yousef-Karimi Experimenta-Details Year-2000 Year-2001 Year-2002 Model Thesis-Text	The root directory contains a subdirectory *Yousef-Karimi* which is divided into three subdirectories, *Experiment-Details, Model* and *Thesis-Text.*
Yousef-Karimi Experimenta-Details Year-2000 Year-2001 Year-2002 Model Thesis-Text	This subdirectory is further divided into three sub-subdirectories such as *Year-2000, Year-2001* and *Year-2000,* which respectively contain the detailed experimental data.
Yousef-Karimi Experimenta-Details Year-2000 Year-2001 Year-2002 Model Thesis-Text	In *Year-2000* directory, hyperspectral CASI images, ASD spectral data, and other ground data are given in *CASI-Image*, *ASD-Data*, and *Ground-Data* subdirectories, respectively. Each of these directories is further divided into various sub-subdirectories. The names of these subdirectories are self explanatory with their content. *Readme* file (in *Year-2000* directory) also explain all these details.
Yousef-Karimi Experimenta-Details Year-2000 Year-2001 Year-2002 Model Thesis-Text	In *Year-2001* directory, Ground data are given in a subdirectory *named Ground-Data. Readme* file (in *Year-2001* directory) also explain all these details.

	In *Year-2002* directory, ASD spectral data and other ground data are given in *ASD-Data*, and *Ground-Data* subdirectories, respectively. The *ASD-data* directory is further divided into various subdirectories named based on the measurement dates. R*eadme* file (in *Year-2002* directory) also explain all these details.
	In *Model* directory, an example set of the all necessary files for running the STICS model is listed and further details are in the *readme* file.
	*Thesis-Text* directory contains the thesis document in MS-WORD format.