Study on Indication and Monitoring of Transgenic Paddy Rice Cultivation by Hyperspectral Remote Sensing Techniques

LI, Ru

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in

GeoInformation Science

The Chinese University of Hong Kong

June 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/Assessment Committee

Professor ZHUANG Yuanzhi (Chair)

Professor CHEN Jinsong (Supervisor)

Professor LIN Hui (Co-Supervisor)

Professor HE Junxian (Committee Member)

Professor ZHANG Xia (External Examiner)

ABSTRACT

of thesis entitled:

Study on Indication and Monitoring of Transgenic Paddy Rice Cultivation by Hyperspectral Remote Sensing Techniques

Submitted by LI, Ru
for the degree of Doctor of Philosophy
at The Chinese University of Hong Kong
in June 2011

Due to the stochasticity, diversity and variability of gene expression, transgenic crop study, is confronted with some uncertainties, such as what kinds of the influence from foreign gene on the transgenic crop, and how to fulfill the monitoring of transgenic crop growth real-/ near real-time efficiently. The influence of foreign gene could be treated as a special source of stress to vegetation. Therefore, it is promising to detect the difference between transgenic and contrast group and so as to monitor the growth of sample to assist to fulfill sample screening work, focusing on the plant biophysical traits or responses to stress by spectral techniques. Hyperspectral remote sensing technique is a kind of practical and field spectroscopy technique, which is simple, rapid, real-/ near real-time, user friendly and cheap. In this study, this technique was employed to indicate the differences between transgenic crop samples and their parents, and to monitor their growth. By the proposed approach, fine spectra of transgenic paddy rice were obtained, and the growth of samples were monitored the by their biophysical traits, finally the screening of cultivars were fulfilled in contrast controlled experiments. The biophysical traits or bio-process were concentrated on rather than on micro-structure or components of proteins. It will be implemented to monitor the growth of the samples real-/ near real-time, helping researchers know their samples clearly and screen samples efficiently.

In order to develop and validate this approach, 6 experiments in different fields were conducted, including three kinds of genomes and their transgenic samples. They were classified as the experiment-repeat experiments and the gene-repeat experiments. Moreover, a three-month experiment was also conducted for evaluating the capability of the approach to monitor the sample growth under the condition of an artificial stress (herbicide stress). Morphologic and parameterized

features of foliar spectra of samples were applied to indicate the growth of the samples.

The results proved this approach proposed was not a substitute to the popular methods for gene detection and crop assessment, but an important, helpful and efficient complement to make the crop breeding study under control and efficient as much as possible. By the approach, the researcher could know their samples clearly and real-/near real- time.

In the future, more factors should be considered. They are mainly: much more effective communication with biological researchers should be conducted; more research methods should be introduced, the study scope should be extended to the whole bands (350-2500nm) and more foliar chemicals should be involved as indicators of the growth status of the samples, etc.

摘要

基因表达过程充满随机性多样性和可变性，因而转基因作物研究，就会回到很多不确定性问题，诸如转入基因对受体到底有何种影响，如何有效的实时准实时的监测其生长等。转入基因对受体作物的影响可以看作对一种特殊的胁迫。因而可以借助光谱技术，着重考虑样本生物性状和对胁迫的反应，探测转基因样本与其母本间的差异，监测它们的生长状况，辅助科研人员完成样本的选育工作。高光谱遥感技术是一种实用光谱技术，具有简单、易用、快速，实时准实时显示目标探测结果等特点。故本研究应用该技术，探测、指示转基因作物与其母本间的差异，实现对其生长的监测。本研究着眼于样本生物形状或生化过程，而不是作为微观结构或组成的改变。在控制对照试验中，应用该技术，获得转基因水稻叶片精细光谱，可以实现样本生长的实时、准实时的监测，辅助研究人员掌握样本情况，完成样本选育。在这项研究中，总计开展 6 次的野外实验。这些实验涉及不同转基因水稻品种，分别设计为实验重复实验和基因重复实验。同时，有一项持续时间长为 3 个月的样本监测实验。在该实验中，样本生长在人工胁迫状态下（除草剂）。光谱特征，包括形态学特征和参数化特征都被引入该研究，用以指示样本的生长状态。实验结果显示，本研究开发的方法是一种重要的实验室基因检测和评估专业方法有效补充。应用本方法可以帮助研究人员在转基因作物研究中，如育种，更清晰的实时、准实时的了解样本状态，加强实验控制、节约成本。在将来的研究中，为了更进一步的改进本方法，应加强与生物研究人员的有效沟通，应用更多的研究方法，将研究扩展到全波段（350- 2500nm），并引入更多的叶片生物化学物质成分作为样本生长状态的指示因子。

致谢

光阴荏苒，已经要博士毕业了，当年来港前的博士宣言还犹如昨日所讲一般，清晰可见：做博士不是读博士，是事业不是工作，要转变不是被动，努力升级不是保守退缩。

再次回首这三年的博士生活，不容易，幸有家人亲人的支持，有太多的良师益友的帮助、鼓励：

感谢奶奶、爸爸、妈妈、姨、姨夫、弟弟、妹妹、表弟、表妹的支持，有他们在家里的辛勤，我才能安心的完成博士工作。

感谢我的导师林珲、陈劲松教授和香港中文大学张渊智、何军贤教授，有他们的支持、指导和帮助，我才能成功的开展实验，顺利的完成博士工作。

感谢中国科学院遥感应用研究所及对地观测与数字地球中心的郑兰芬、王普年、张兵、张霞、卫征、陈正超、张文娟、张立福、方俊勇、刘学、邓滩研究员，有他们的鼓励和支持，我才得以有更坚定的信念一气呵成完成博士工作。

感谢中国科学院亚热带农业生态研究所的王克林、肖国樱、陈洪松、宋同清、何旬阳，周传社、王敏研究员，岳跃民、赵森、张宇喆、彭晚霞、张明阳、曾朝霞、杨青青、祁向坤、凡飞得、聂跃鹏、付小青、陈芬博士，付伟、张伟师兄、舒适燕、杨珊、谭薇、邓彭艳等师弟师妹。正是有诸位的鼎力支持，还有其他默默无闻的支持者，我才得以在贵所持续实验，获得长时间序列的第一手观测数据完成博士论文，感谢诸位身体力行的帮助！

感谢湖南杂交水稻工程中心袁定阳、唐俐、谭炎宁博士等，感谢诸位提供实验样本，并三伏天和我们一起在田地里参与实验、采集数据，正是有诸位的鼎力支持，我才能获得丰富的实验数据，顺利完成博士论文。

感谢环保部南京环境科学研究所刘波博士，大家一起实验，一起讨论，相互支持。

感谢江西省科学院于一尊，九江林业局谭铭策老先生，鄱阳湖保护区高云、罗浩等诸位！鄱阳湖实验大家合作愉快，我也从中学习到很多东西。

感谢中国科学院寒旱所戴礼云博士，谢谢她对我实验工作的支持和在兰州学习期间的款待。

感谢中国科学院长春地理所赵凯、金吉南、张世轶、李晓峰研究员，郑兴明、杜嘉、姜涛、汤光旭、武彦清、雷晓春博士等。数九寒天，与诸位一起在雪地里做一天实验，嗯，别有一番豪情！

感谢香港中文大学太空与地球信息科学研究所 Chloris Yip，Wendy Wong，Isacc Chan，Anthony Wong，Hazel Sin，May Tam，Yeung Kin 是诸位每日默默的辛勤工作，支持着全实验同仁的工作，有了诸位的努力，我们才得以安心的进行论文工作。感谢我的师兄师妹们：陈富龙博士、江利明、胡亚、胡明远、李蓉蓉、刘婷婷、陈旻、王丹、赵卿、王贵武、席红艳、刘彪、高斯、闫肃、张佩瑶、王宇飞、武磊、王盛校、胡显志、程世来、谭露，张春晓、罗

卿莉，顾艳镇、Stephen Lau、王红说、张鸿生、孙玉娟、车伟涛、孙贵博、林天鹏博士等，和大家在一起工作的时光很有趣，谢谢诸位，也感谢其他没有提到名字的同仁的支持！

感谢香港中文大学刘杰鹏、窦亚平、王凯、夏衡、路晋诚博士、殷伟欣小弟（呵呵），一个宿舍 3 年，大家相互帮助相互照应，非常开心！

感谢香港中文大学人间仙境 BBS 里 Acmilier 兄及其他诸位的热情，三年里和大家一起 hiking、户外：翻过了香港一座座山，踏过了一条条径，冒险爬悬崖，涉水过礁石、无畏的在灌木高草中强行开路……呵呵，甚是豪迈。

特别感谢邓飚博士、岳跃民博士、刘波博士、张宇喆博士、王敏博士，诸位不仅在实验开展、论文写作、项目进展等方面，在生活、工作上、发展上都予以我最大的支持，这些都是一种默契！

还有其他没有提到名字但在默默无闻的支持过我的人，非常感谢诸位的支持！

鞠躬致意！

三年时间的磨练，有诱惑，有煎熬，有苦难，有悲凉，更多的是坚信、坚持，于是欢乐、向上是主调。整个人也实现了升级：思想有徘徊进而升华，肉体有磨练进而强健，精神更加饱满，信念更加坚定。从这个角度讲，升级也是三年的主调。

在博士答辩结束后，历时 28 天，从成都出发一路骑单车到了拉萨，总计翻过了有名字的大山 14 座，其中海拔 5000 米以上的两座，累积爬升远远超过 2 万米。有时候，一出门就想着什么时候是个头，一出门就想着推车吧，推车吧，自己不停的在与自己斗争，好在，所有的山都坚持骑，扛过来了。等回头再看，淡淡的一句话：坚持了，就过来了。这算是个顶大的收获。

此行，算是对三年博士的总结，也是未来生活的开始，汇到一句话，即：百废待兴。没用承前，只有启后。就像我们每天骑车，看好前方的路，过去的就过去了。套用邓飚博士常讲的：以史为鉴，面向未来！

Study on Indication and Monitoring of Transgenic Paddy Rice Cultivation by Hyperspectral Remote Sensing Techniques

TABLE OF CONTENT

ABSTRACT I

摘要 III

致谢 V

LIST OF FIGURES XIII

LIST OF TABLES XV

CHAPTER 1 PROBLEMS IN TRANSGENIC CROP STUDY 1

1.1. INTRODUCTION 1

1.1.1 Indication and monitoring of gene expression 1

1.1.1.1 The definition of Gene expression at molecular level 1

1.1.2 Problems in indication and monitoring of gene influence 2

1.1.2.1 Polymorphism of gene expression 2

1.1.2.2 Over- and suppressed- expression of gene 3

1.1.2.3 Gene silencing and Co-suppression 4

1.1.2.4 Unknown influences on receptor from foreign genes 4

1.1.3 Comments on foreign gene monitoring 4

1.2. INDICATION AND MONITORING OF CULTIVATED TRANSGENIC PADDY RICE GROWTH BY HYPERSPECTRAL REMOTE SENSING TECHNIQUES 5

REFERENCES 7

CHAPTER 2 FUNDAMENTALS FOR THE NEW APPLICATION OF HYPERSPECTRAL TECHNIQUES TO TRANSGENIC CROP STUDY 8

2.1. INFLUENCE ON RECEPTOR FROM FOREIGN GENE 9

2.1.1 Influence on receptor from foreign gene 9

2.1.2 Laboratory approaches of indication and monitoring of gene expression 9

2. 1.3 Plant responses to transferred gene 10

2.2 HYPERSPECTRAL TECHNIQUES FOR DETECTING INFLUENCES OF TRANSGENE AND STRESS 11

2.2.1 Introduction to spectroscopy 11

2.2.2 Optical characteristics of plant leaf 16

2.2.3 Bands response to stress 17

2.3. HYPERSPECTRAL REMOTE SENSING TECHNIQUES 18

2.3.1 Concept of hyperspectral remote sensing techniques 18

2.3.2 The historical, current and future development of hyperspectral remote sensing 20

2.3.3 Hyperspectral remote sensing of vegetation 21

REFERENCES 22

CHAPTER 3 EXPERIMENT AND ITS UNCERTAINTIES 25

3.1 GUIDELINE OF THE EXPERIMENT 25

3.2 COLLABORATING INSTITUTES, SAMPLES AND DEVICES IN THE EXPERIMENTS 26

3.2.1 Collaborating institutes and samples 26

3.2.2 The gene-repeat and experiment-repeat experiment 28

3.2.2.1 The gene-repeat experiment 28

3.2.2.2 The experiment-repeat experiment 28

3.2.3 Devices 28

3.3 UNCERTAINTY PROBLEMS 30

3.3.1 The definition and classification of Uncertainties in remote sensing 30

3.2.2 Uncertainties in experiments 31

3.2.2.1 Time of the Measurement 31

3.2.2.2 Interval of the measurement 31

3.2.2.3 Noise from the reference black background 32

3.2.2.4 Canopy-level and fine leaf-level spectra 33

3.2.2.5 Mean spectra of the group and spectra of the individual sample 33

3.2.2.6 Uncertainties in model applied 33

3.2.2.7 Problems in measurement 34

REFERENCES 34

CHAPTER 4 INDICATORS OF THE FOLIAR CHEMICALS AND THE SAMPLE GROWTH 34

4.1 FOLIAR PHOTOSYNTHETIC PIGMENTS ESTIMATION 35

4.1.1 Introduction 35

4.1.2 Methodology 36

4.1.2.1 Spectral indices developed as indicators of chlorophyll 36

4.1.2.2 Spectral indices developed as indicators of carotenoids 38

4.1.2.3 Spectral indices developed as indicators of anthocyanin 39

4.1.2.4 Spectral reflectance data and in situ data 39

4.1.2.5 Statistical analysis model 40

4.1.3 Results and discussion 41

4.1.3.1 Assessment of correlation of spectral indices with photosynthetic pigments 41

4.1.3.2 Correlation between spectral indices 51

4.1.3.3 Assessment of spectral index sensitivity 53

4.1.4 Conclusion 54

REFERENCES 55

4.2 NEW SPECTRAL INDICES AS INDICATORS OF THE PHOTOSYNTHETIC PIGMENTS OF PADDY RICE 57

4.2.1 Pigment absorption 57

4.2.2 New spectral indices as indicators of chlorophyll pigments 59

4.2.2.1 New spectral index as indicators of chlorophyll pigments 59
4.2.2.2 Results and discussion 61
4.2.3 New spectral index as indicator of carotenoids *71*
4.2.3.1 Basic assumption 71
4.2.3.2 New spectral index as indicator of carotenoids 72
4.2.4 Conclusion *74*
REFERENCES 75
4.3 ESTIMATION OF THE CONTENT OF WATER, NITROGEN AND LIGNIN 76
4.3.1. Nitrogen and lignin content estimation methods *76*
4.3.2. Bands selected based on stepwise regression analysis *77*
4.3.3 Results and discussion *78*
4.3.4 Conclusion *82*
REFERENCES 82

CHAPTER 5 MONITORING AND SCREENING TRANSGENIC PADDY RICE UNDER CONTROLLED CONTRAST EXPERIMENTS BY FIELD-HYPERSPECTRAL MEASUREMENTS 83

5.0.1 Assumption of gene expression *83*
5.0.2 Basic idea and function *83*
5.0.3 Spectral analysis methods *84*
5.0.3.1 Analysis based on spectral morphological features 84
5.0.3.2 Analysis based on parameterized features 85
5.1 COMPARISON ANALYSIS BASED ON SPECTRAL REFLECTANCE EDGES WITHIN SAMPLE GROUPS 85
5.1.1 Data acquisition system *85*
5.1.2 Calculation of the edges *87*
5.1.3.1 Samples 89
5.1.3.2 Index selection 89
5.1.4 Results and discussion *92*
5.1.5 Conclusion *97*
REFERENCE 98
5.2 COMPARISON ANALYSIS BASED ON SPECTRAL SHAPE AND PARAMETERIZED FEATURES AMONG SAMPLE GROUPS --A CASE OF THE GENE REPEAT EXPERIMENT 100
5.2.1 Experiment design *100*
5.2.2 Samples and data acquisition *101*
5.2.2.1 Samples 101
5.2.2.2 Spectral data acquisition 102
5.2.3 Methodology *102*
5.2.3.1 Spectral angle between groups 102
5.2.3.2 Continuum removal spectra 103
5.2.3.3 Spectral indices as indicators of chlorophyll and carotenoids 103
5.2.4 Result and discussion *104*

5.2.4.1 Analysis on spectral morphological features 104
5.2.4.2 Analysis of parameterized features 105
5.2.5 Conclusion *107*
REFERENCES 108
5.3 COMPARISON ANALYSIS OF TRANSGENIC PADDY RICE GROWTH BASED ON LONG TIME-SERIES SPECTRAL DATA --A CASE OF THE EXPERIMENT REPEAT EXPERIMENT 109
5.3.1 Introduction *109*
5.3.2 Transgenic samples of long time series *109*
5.3.3 Growth analysis within individual sample *110*
5.3.4 Growth analysis among sample groups *115*
5.3.5 Growth analysis among transgenic group and contrast group *118*
5.3.6 Conclusion *120*
REFERENCES 121
5.4 ANALYSIS OF DIFFERENCE RESPONSE OF TRANSGENIC SAMPLES TO STRESS --A CASE OF THE FOREIGN GENES ACTIVATED BY ARTIFICIAL INDUCTION 121
5.4.1 Samples and experiment *122*
5.4.1.1 Samples 122
5.4.1.2 Experiment 122
5.4.1.3 Equipment 123
5.4.2 Quantitative analysis based on spectral morphological characteristics *126*
5.4.2.1 Spectral angle between samples 127
5.4.2.2 Spectral difference of continuum removal 129
5.4.3 Qualitative analysis based on parameterized features *135*
5.4.3.1 Parameter selection 135
5.4.3.2 Quallitative analysis based on parameterized features 138
5.4.4 Conclusion *143*
REFERENCE 144
5.5 PRE-TEST OF DIFFERENCE ANALYSIS WITHIN SINGLE SAMPLE GROUP --A CASE OF INDIVIDUAL SAMPLE SCREENING. 146
5.5.1 Outlier detection based on Grubbs Test *146*
5.5.2 Analysis based on morphological features *148*
5.5.3 Growth status analysis based on statistic methods *149*
5.5.4 Conclusion *150*
REFERENCE 151
CHAPTER 6 DATA PROCESS AND ANALYSIS SYSTEM **151**
6.1 DEVELOPMENT ENVIRONMENT 151
6.2 SYSTEM STRUCTURE DESIGN 152
6.2.1 Data process modules *152*
6.2.2 Data analysis modules *154*

6.2.2.1 Integrated data process module 155
6.2.2.2 Outlier within group detection module 156
6.2.2.3 Analysis of features among groups 156
6.2.2.4 Analysis of sample growth trends 157
6.2.3 The interface of the system *157*
6.2.3.1 Menu 157
6.2.3.2 Windows buttons 157

CHAPTER 7 SUMMARY **159**

7.1 CONCLUSION 160
7.2 PROSPECTIVE 162

LIST OF FIGURES

Figure 1-1. A model of the expression of a single gene 2
Figure 1-2. Gene expression polymorphism 2
Figure 1-3. Over- and suppressed- expression of gene 3
Figure 1-4. Over- and suppressed- expression of gene 3
Figure 1-5. Gene silencing: RNA degradation and DNA methylation 4
Figure 1-6. The frame of indication and monitoring of cultivated transgenic paddy rice growth by hyperspectral remote sensing techniques 6
Figure 2-1. The tools applied to detect genes 10
Figure 2-2. Generalized diagram of Electromagnetic bands and theory of spectroscopy 13
Figure 2-3. Three kinds of spectrum discussed 14
Figure 2-4. 44 absorption features 16
Figure 2-5. Generalized diagram of a leaf' structure and its reflectance characteristics at visible and near IR wavelength 17
Figure 2-6. The typical vegetation spectrum 18
Figure 2-7. Spectral signatures from hyperspectral vs. multispectral sensors 19
Figure 2-8. Application of Hyperspectral remote sensing 21
Figure 3-1. Guideline of the controlled experiment in this study 26
Figure 3-2. China national hybrid rice R&D center 27
Figure 3-3. Institute of subtropics agriculture 27
Figure 3-4. Devices 28
Figure 3-5. The contact-probe and the incident angle of light 29
Figure 3-6. Spectral reflectance of the reference background 30
Figure 3-7. Photosynthetic mode of plant in daytime 31
Figure 3-8. Heat effect caused by probe light 32
Figure 3-9. Two kinds of spectral reflectance 33
Figure 4-1-1. Spectra of samples and morphological details of photosynthesis sensitive bands 36
Figure 4-1-2. Correlation between spectral indices as chlorophyll indicators 51
Figure 4-1-3. Correlation between spectral Indices and carotenoids indicators 52
Figure 4-2-1. Typical photosynthetically active radiation (PAR) action spectrum, shown beside absorption spectra for chlorophyll a, chlorophyll b, and carotenoids 58
Figure 4-2-2. Red edge shift 59
Figure 4-2-3. Sensitivity analysis for reference indices to chlorophly concentrantion 64
Figure 4-2-4. Spectral reflectance curve mixed with different normally distributed random numbers. 70
Figure 4-2-5. Relationship between the content of Caretonoinds (mg/g, per fresh weight) and new spectral index. 74
Figure 4-3-1. Strong water absorptions under natural illumination conditions. 78
Figure 4-3-2. Correlation coefficients between varibles and nitrogen concentration. 79
Figure 4-3-3. Correlation coefficients between varibles and lignin concentration. 81
Figure 5-0-1. Early Indication based HRST for transgene plant cultivation 84

Figure 5-0-2. Analysis based on spectral morphological features 85
Figure 5-0-3. Analysis based on spectral parameterized features 85
Figure 5-1-1. a. Observation with probe-leaf-clip system; b. Field Spectrometer 86
Figure 5-1-2. Spectra curves and the morphological details of photosynthesis sensitive bands. 86
Figure 5-1-3. Sample spectra and their first derivative curves. 88
Figure 5-1-4. Positions of the edges 92
Figure 5-1-5. Comparison of spectral parameters between Parent, T1, T2, T3, and T4 94
Figure 5-1-6. Spectral indicators for specific pigments. 96
Figure 5-2-1. General Design of the Experiment 101
Figure 5-2-2. Observation with probe-leaf-clip system 102
Figure 5-2-3. Spectral Angles of the samples 104
Figure 5-2-4. Spectral Continuum Removal Curve 105
Figure 5-2-5. Spectral indices correlated to foliarChl a+b 106
Figure 5-2-6. Spectral indices correlated to foliar Carotenoids 106
Figure 5-2-7. Dispersion of spectral indices between groups taking datt and CARI2 as examples 107
Figure 5-3-1. Track of Individual sample group growth 111
Figure 5-3-2. local condition of temperature and weather. The temperature in the green house was marked "In-door" while the other was marked "Out-door". 112
Figure 5-3-3. The track of Individual sample group growth 119
Figure 5-4-1. ASD Spectrometer with probe; b. Data acquisition 123
Figure 5-4-2. Spectral angles among samples by time 129
Figure 5-4-3.Dispersion of spectral continuum removal results (compared with the contrast groups by time, *100%) 132
Figure 5-4-4.Dispersion of spectral continuum removal with groups (compared with the status at -0h, *100%) 135
Figure 5-4-5. Dispersions of spectral indices within groups of time series (compared with -0h' which represents the status before herbicide sprayed, *100%) 139
Figure 5-4-6. Dispersions of spectral indices within groups (compared with the contrast samples, *100%). 140
Figure 5-4-7. Dispersions of spectral indices among groups (compared with the contrast samples, *100%) 140
Figure 5-5-1. The Spectral continuum removal image 149
Figure 6-1. GUIDE: the MATLAB graphical user interface development environment 152
Figure 6-2. Input data for data pre-process: mass raw spectral data for data pre-process 154
Figure 6-3. A special case for calculating the mean spectrum 154
Figure 6-4. Input directory for data pre-process 155
Figure 6-5. The log file (a) and the results (b) of integrated data process module 156
Figure 6-6. Menu of the system 157
Figure 6-7. User interface window 158
Figure 6-8. The results of "Analysis of features among groups" 158
Figure 6-9. The results of "Analysis of sample growth trends" 159

LIST OF TABLES

Table 3-1. Statistics of the spectral reflectance of the background reference panel. 32
Table 3-2. Spectral indices with difference equation type and wavelength 32
Table 4-1-1. Spectral indices developed as indicators of chlorophyll a& b 37
Table 4-1-2. Spectral indices developed as indicators of carotenoids 38
Table 4-1-3. Regression model applied 41
Table 4-1-4. Regression results between spectral indices and content of chl a 44
Table 4-1-5. Regression results between spectral indices and content of chl b. 45
Table 4-1-6. Regression results between spectral indices and content of chl a +b 46
Table 4-1-7. Regression results between spectral indices and content of carotenoids 47
Table 4-1-8. Regression results between spectral indices and ratio of carotenoids to Chlorophyll a 48
Table 4-1-9. Pearson correlation analysis between spectral Indices 48
Table 4-1-10. Pearson correlation analysis between spectral Indices 49
Table 4-1-11. Sensitivity of spectral indices (based on normalized data) 53
Table 4-2-1. Regression results between spectral indices and content of chl a 65
Table 4-2-2. Regression results between spectral indices and content of chl b 66
Table 4-2-3. Regression results between spectral indices and content of chl a+b 67
Table 4-2-4. Comparison of regression results between the spectral indices based on the original signal and polluted one 68
Table 4-2-5. Pearson correlations analysis between new spectral Indices 71
Table 4-3-1. Pre-selected potential absorption features 77
Table 4-3-2. Model summary for nitrogen concentration estimation 80
Table 4-3-3. Model summary for lignin concentration estimation 81
Table 5-1-1. Inner-cc of sample groups 88
Table 5-1-2. Spectral Indices developed as chlorophyll (a+b) indicators 90
Table 5-1-3. Spectral indices developed as carotenoids indicators: 91
Table 5-1-4. Spectral indices developed as anthocyanin indicators: 91
Table 5-2-1. Spectral indices developed as chlorophyll (a+b) indicators 103
Table 5-2-2. Spectral indices developed as carotenoids indicators: 103
Table 5-3-1. Index selected and its relationship with foliar chemicals 112
Table 5-3-2. The relationship (interaction) between foliar chemicals 114
Table 5-3-3. The correlation coefficients between the cc of samples 116
Table 5-3-4. The significant test of the difference between T4 and P (α=0.01) 120
Table 5-4-1. Pre-test of consistency within sample groups 124
Table 5-4-2. Spectral Indices developed as chlorophyll (a+b) indicators 136
Table 5-4-3. Spectral indices developed as carotenoids indicators 137
Table 5-4-4. Spectral indices developed as anthocyanin indicators 137
Table 5-4-5. Spectral indices developed as water content indicators 138
Table 5-4-6. The correlation coefficients among transgenic groups 140
Table 5-4-7. The significant test of the difference between T4 and P (α=0.05) 142

Table 5-5-1. The outliers in the samples 147
Table 5-5-2. The Spectral angles of sample spectra (in degree) 148
Table 5-5-3. The significant test (sig.) of the difference (α=0.01) 150
Table 5-5-4. The correlation coefficients among spectra 150

Chapter 1 Problems in transgenic crop study

1.1. Introduction

Biotechnology has received much attention nowadays, which was identified as one of the three most important emerging and evolving fields along with nanotechnology and geo-technology by the US Department of Labor (Gewin, 2004). As a big progress of biotechnology, transgenic techniques are marked as one of the greatest technologies. By introducing foreign genes and making them to express functionally in plant in less than two decades, super transgenic crops (species) can be developed with improved resistance to insect and disease, seeds and fruits with enhanced nutritional qualities, and plants that are better adapted to adverse environmental conditions (Herrera-Estrella, et al., 2005). With the development of transgenic techniques, some problems or uncertainties are arising, such as problems of the detection of foreign genes (transferred) expression (gene silencing), problems of the monitoring of functional influences on the objects (both on the target objects and other uncertain objects) caused by transgenic expression (gene over- or suppressed expression).

In some sense, these problems are the issues of how to find and assess the influences of foreign gene on transgenic plants and how to monitor the growth of transgenic plant. They are significant to assess the uncertainty of transgenic products, and are crucial to the study, development and application of these techniques and their products, especially for the studies such as crop breeding.

1.1.1 Indication and monitoring of gene expression

1.1.1.1 The definition of Gene expression at molecular level

Gene expression is a technical term to describe how active a particular gene is, namely, how many times it is expressed or transcribed, to promote it encodes. It is "*a process of the translation of the information encoded in a gene into an RNA transcript. Expressed transcripts include message RNAs translated into proteins, as well as other types of RNA, such as transfer RNA, ribosomal RNA, micro RNA, and non-coding RNA, which are not translated into protein. Gene expression is a highly specific process by which cells switch genes on and off in a timely manner, according to their state*" (Marchionni, et al., 2008). In summary, Gene expression at molecular level could be described as produce proteins according to the encoding message from the transgene.

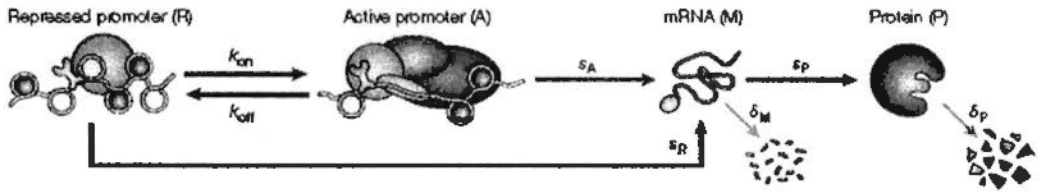


Figure 1-1. A model of the expression of a single gene (K rn, et al., 2005)

1.1.2 Problems in indication and monitoring of gene influence

Gene expression is susceptible thus it shows the characteristics of uncertainty (Sanderson, 2007) and stochasticity (Kaern, et al., 2005, Raser, et al., 2005) including gene expression polymorphism(Kaern, et al., 2005, Raser and O'Shea, 2005), over- and suppressed-expression(Baulcombe, 2004) and gene silencing and co-suppression (Van Blokland, et al., 1994) etc.

1.1.2.1 Polymorphism of gene expression

Expression polymorphisms can be identified as the significant difference at the expression level (Huang, et al., 2006). It means that the same gene unit or genome without external influences such as stresses (water, temperature, chemical injury and so on) would also have different expression, namely different phenotypes. The object the polymorphism happens in has no relation with the fact whether the organisms/ cells have transgene.

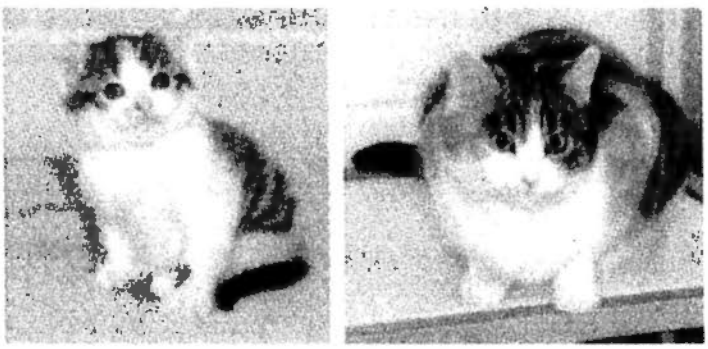


Figure 1-2. Gene expression polymorphism

"Cc, the first cloned cat (left) and Rainbow, Cc's geneticmother (right), display different coat patterns and personalities. Photo credit, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University" (Raser and O'Shea, 2005).

1.1.2.2 Over- and suppressed- expression of gene

Over-express: in biology, is indentified to make too many copies of a protein or other substance. Over-expression of certain proteins or other substances may play a role in cancer development[1]. Over-expression (Figure 3) could introduce many effects. Suppressed expression (gene suppression) is an opposite term to over-expression, and could be described as one gene suppresses the expression of other genes. As following figure 4 showed, GFP had been suppressed (Baulcombe, 2004).

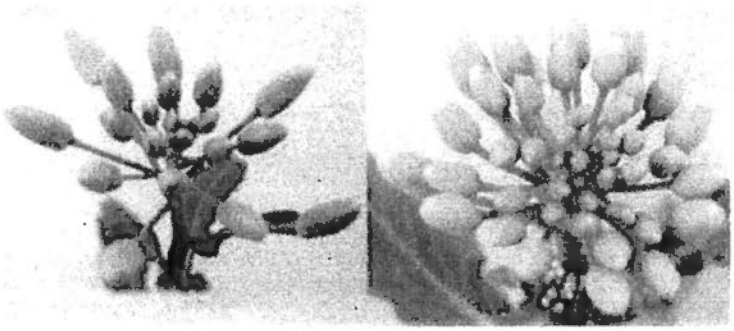


Figure 1-3. Over- and suppressed- expression of gene

"The canola plant with an over-expressed HSD gene (right) has more flower buds than the unmodified plant (left). More flowers produce more seeds and therefore higher oil yields"[2].

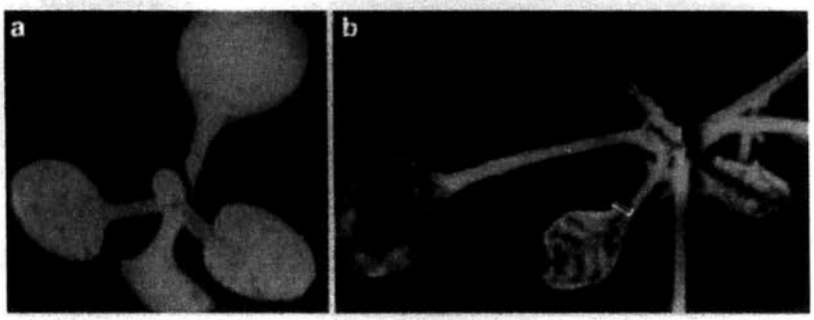


Figure 1-4. Over- and suppressed- expression of gene

(a)" *In the wild-type plant, a GFP transgene is constitutively transcribed, but the GFP fluorescence is suppressed. b, This plant has a mutation in RDR6 and the silencer signal is able to act only in the cells that are that are further than about 20 cells from the phloem and consequently they appear green under ultraviolet light The GFP transgene is not silenced in cells that are further*"

(Baulcombe, 2004).

---

[1] National Cancer Institute, http://www.cancer.gov/dictionary/?CdrID=45812

[2] http://www.nrc-cnrc.gc.ca/eng/news/nrc/2007/12/07/genetic-boost.html. National Research Council Canada, 2007

1.1.2.3 Gene silencing and Co-suppression

"In genetically modified plants, the introduced transgenes are sometimes not expressed, namely being silenced Transgenes integrate at different chromosomal locations. If they become inserted into euchromatin, in a transcriptionally active region, expression may be influenced by regulatory sequences of nearby host genes. If they insert in or near repetitive DNA or heterochromatin, they can be inactivated. Another important factor associated with gene silencing is the number of transgenes per integration site" (Stam, et al., 1997). Transgenes can also cause the silencing of endogenous plant genes if they are sufficiently homologous, a phenomenon known as co-suppression (Stam, et al., 1997).

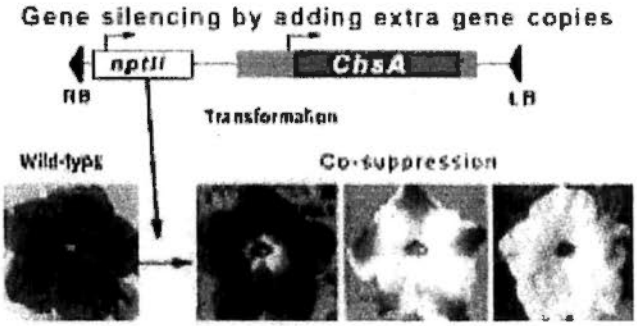


Figure 1-5. Gene silencing: RNA degradation and DNA methylation[3] (Van Blokland, et al., 1994)

1.1.2.4 Unknown influences on receptor from foreign genes

There would be also some unknown influences caused by foreign gene. These influences could be thought as by-products. In some sense, it is a serious issue. To find and monitor these influences is important to many studies, especially for crop breeding.

The expression of gene is a process which is easily influenced by stochastic effects/noise (K rn, et al., 2005) and results in variability. Not only environment and history would contribute to variability in cellular phenotype but also organisms with same genes in same environment, with the same history, display variations in form and behavior that can be subtle or dramatic (Raser and O'Shea, 2005). In summary, there are full of uncertainties during the expressions of gene, and we need stable, reliable, fast, efficient and sensitive enough approaches to detect and monitor gene expression and influences quantitatively and qualitatively in the study.

1.1.3 Comments on foreign gene monitoring

Testing for a single bio-trait or genetically modified organ event may require only a simple method, whereas testing for presence of multiple events, possibilities for identification and quantification may require use of combinations of methods (Christianson, et al., 2008,

[3] http://www.bio.vu.nl/genetica/Research-projects-new/Projects_Kooter/Projects_J.Kooter.htm

Holst-Jensen, 2007, 2009). Moreover no matter on DNA or protein the techniques based, they are applied at laboratory. They are quantitative, qualitative, stable and reliable enough to fulfill detection of gene expression and deduce itself influences on receptor. These techniques are at a micro view to confirm whether genes or special structures exist and then deduce or try to reveal the potential influences caused by the foreign things on the receptor. They are efficient when the objective genes are known with indication of prior knowledge. However, because the gene expressed at molecular level is so complex that, as presented above, current dominated techniques are reliable and efficient with help of prior knowledge, but if no prior knowledge, it would be time and labor consuming and hard to accomplish the engaged task. A PCR method cannot detect a completely unknown gene modified organism (GMO), since prior knowledge of the DNA sequence is necessary for the primer design (Michelini, et al., 2008). The optimal DNA chip has been designed to solve this problem (Tengs, et al., 2007), but for large amount of samples, it is still embarrassed.

Furthermore, taking into account the phenomena/ special problems, sometimes though the gene was transferred into receptor successfully it may not be expressed or expressed but researcher could not make sure whether this foreign thing has influences on the object organism or only on this organism. In this situation, these approaches are helpless because of a huge amount of work, expensive cost both at time and material consuming, and other uncontrolled problems. In addition, at different growth stage, the transgene may have different expressions and influences on different organism, real-time detection and monitoring of it are very important, especially for experiment lasting for a long time, such as crop breeding.

1.2. Indication and Monitoring of cultivated Transgenic Paddy Rice Growth by Hyperspectral Remote Sensing Techniques

Is there any technique, which is rapid, real-time or near real-time, stable, sensitive, easy to use and cheap, could satisfy to be early indicating of information of gene influences (e.g. transgenic expression) on receptor? Spectroscopy would be a candidate, which is non-destructive, fast, without pollution and no requirement of sample pre-treatment (Blanco, et al., 2002).

In this study, the author proposed to employ hyperspectral remote sensing techniques, a kind of practical and field spectroscopy technique, to obtain fine spectra of transgenic plant, by monitoring the real-/ near real-time growth of sample, to fulfill early indication of possible the differences between transgenic crops and their contrast in the controlled contrast experiment. The idea is an approach from a macro view rather than the one focusing on the molecular level. It compares the differences between transgenic samples and their contrast which are cultivated in the controlled contrast environment. It will be implemented to monitor the growth of the samples real-/ near real-time, assist to screen samples and help researchers clearly know their

samples.

Hyperspectral technique could be applied to separate matter if it has unique spectral features (diagnostic absorption features by certain chemical bonds or molecular structure intact) with contiguous narrow spectral bands (Goetz, et al., 1985b). It is simple, practical, rapid, real-/ near real-time, user friendly and cheap. In the process, Hyperspectral remote sensing techniques play a role of detection and monitoring of gene by an indirect way from a macro-view. To be emphasized, the approach proposed is not the one replacing the current traditional methods for gene detection and crop assessment, but an important, helpful and efficient complement to make the study such as crop breeding under control and efficient as much as possible.

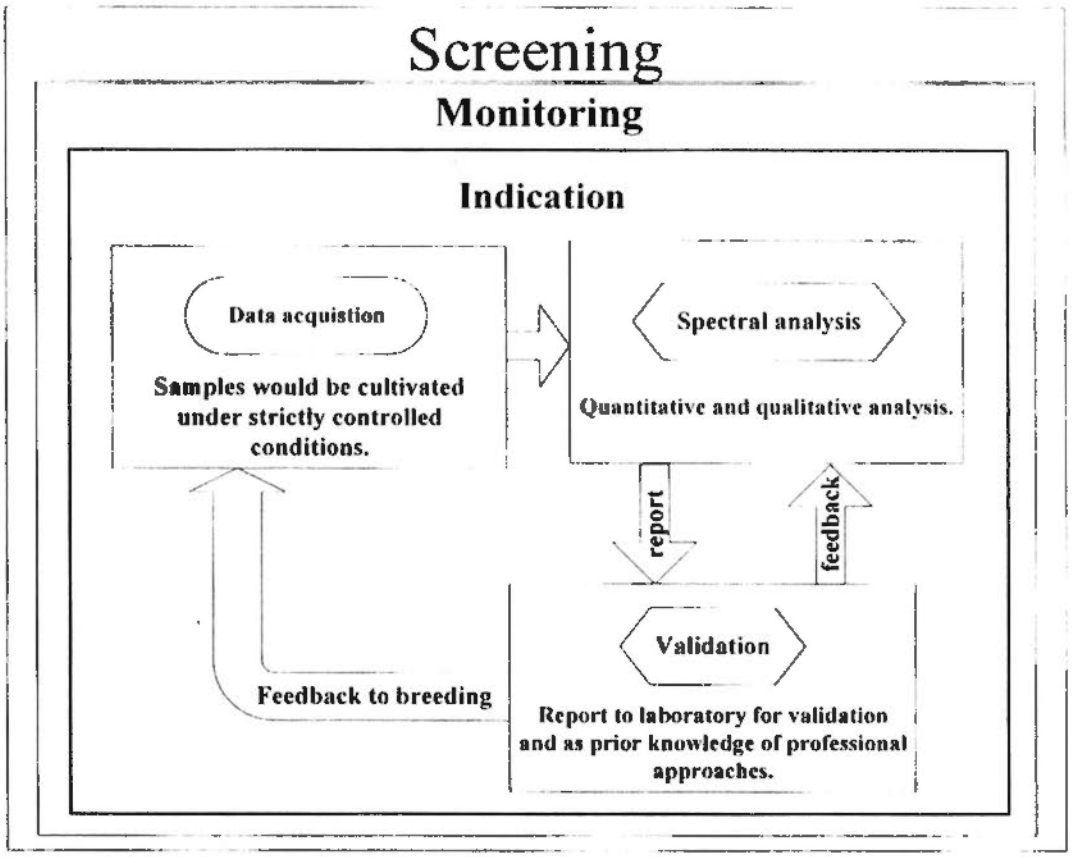


Figure 1-6. The frame of indication and monitoring of cultivated transgenic paddy rice growth by hyperspectral remote sensing techniques

Rice is a vital food crop of the world. In a few years, GM rice will be ready for commercialization, including varieties with higher yields, greater tolerance of biotic and abiotic stresses, resistance to herbicides, improved nutritional quality, and novel pharmaceutical proteins. Therefore, we take transgenic rice as target samples to develop and validate our proposed approaches.

The primary objectives of the proposed research are to:

1. To investigate hyperspectral properties of transgenic rice paddy at growing stages.
2. To assess the sensitivity of the existed vegetation spectral indices when applied to transgenic paddy rice study and design more stable indices according to spectral features of paddy rice.

3. To locate and quantify the changes of transgenic rice paddy compared with their parents both on morphologic and parametric characteristics.
4. To develop a practical method for indication and monitoring gene expression and its influences as Early warning system based HRST for transgene plant cultivation

The research has the significances as follow:

1. To fulfill systematic study on spectral analyses transgenic rice paddy, including analysis of spectral characteristics, sensitive bands to transgenic foliar chemicals. It is the fundamental work for remotely sensed rice plant.
2. To analyze spectral characteristics of transgenic rice plant and develop a feasible workflow for indicating and monitoring of possible difference between transgenic samples and their contrast which would be useful to reveal the gene expression and influences on receptor, and assist crop breeding. It is of significance for transgenic crop breeding and it would be a new promising application area of hyperspectral remote sensing techniques.
3. It is also significant to breeding works of other plants, transgenic or non-transgenic. The thesis proved the feasibility of application of hyperspectral techniques to assist studies in biological field. The approach explored and developed in the thesis is a practical approach and could be applied to assist plant breeding, not limited to paddy rice, directly with help of hyperspectral remote sensing techniques.

References

Baulcombe, D. (2004). Rna silencing in plants. *Nature*, 431(7006), 356-363

Blanco, M., & Villarroya, I. (2002). Nir spectroscopy: A rapid-response analytical tool. *Trends in analytical chemistry*, 21(4), 240-250.

Christianson, J., McPherson, M., Topinka, D., Hall, L., & Good, A. (2008). Detecting and quantifying the adventitious presence of transgenic seeds in safflower, carthamus tinctorius l. *Journal of Agricultural and Food Chemistry*, 56(14), 5506-5513.

Gewin, V. (2004). Mapping opportunities. *Nature*, 427(6972), 376-377.

Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for earth remote sensing 228(4704): 1147-1153.

Herrera-Estrella, L., Simpson, J., & Martinez-Trujillo, M. (2005). Transgenic plants: An historical perspective. Pena, L., 286(3-5).

Holst-Jensen, A. (2007). Sampling, detection, identification and quantification of genetically modified organisms (gmos). *Food Toxicants Analysis. Amsterdam, Elsevier*, 231-268.

Holst-Jensen, A. (2009). Testing for genetically modified organisms (gmos): Past, present and future perspectives. *Biotechnology Advances*, 27(6), 1071-1082.

Huang, Y., Li, L., Chen, Y., Li, X., Xu, C., Wang, S., & Zhang, Q. (2006). Comparative analysis of gene expression at early seedling stage between a rice hybrid and its parents using a cdna microarray of 9198 uni-sequences.

Science in China Series C: Life Sciences, 49(6), 519-529.

K rn, M., Elston, T., Blake, W., & Collins, J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451-464.

Marchionni, L., Wilson, R., Wolff, A., Marinopoulos, S., Parmigiani, G., Bass, E., & Goodman, S. (2008). Systematic review: Gene expression profiling assays in early-stage breast cancer. *Annals of Internal Medicine*, 148(5), 358-369.

Michelini, E., Simoni, P., Cevenini, L., Mezzanotte, L., & Roda, A. (2008). New trends in bioanalytical tools for the detection of genetically modified organisms: An update. *Analytical and bioanalytical chemistry*, 392(3), 355-367.

Raser, J., & O'Shea, E. (2005). Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743), 2010-2013.

Sanderson, C., (2007). *Understanding genes and gmos*. World Scientific Pub Co Inc.

Stam, M., Mol, J., & Kooter, J. (1997). Review article. The silence of genes in transgenic plants. *Annals of Botany*, 79(1), 3-12.

Tengs, T., Kristoffersen, A., Berdal, K., Thorstensen, T., Butenko, M., Nesvold, H., & Holst-Jensen, A. (2007). Microarray-based method for detection of unknown genetic modifications. *BMC biotechnology*, 7(1), 91-98.

Van Blokland, R., Van der Geest, N., Mol, J., & Kooter, J. (1994). Transgene-mediated suppression of chalcone synthase expression in petunia hybrida results from an increase in rna turnover. *Plant journal*, 6(6), 861-877.

Chapter 2 Fundamentals for the new application of hyperspectral techniques to transgenic crop study

As mentioned in the chapter 1, the approach had been proposed to apply hyperspectral remote sensing technique to indicate outliers among samples, assess and screen them quantitatively and qualitatively, and finally report the results to laboratory for validation by specific professional approaches to fulfill screening of cultivars from a macro view. Why could hyperspectral technique be applied to fulfill the task of indication and monitoring of transgenic crop growth and help crop breeding? In this chapter, the relevant techniques and their theory would be introduced and interpreted. For articulating the internal logical relationship within this proposed approach, mainly two kinds of problems should be answered. One is what (kind of) is the influence caused by foreign (or modified) gene on receipt, the other is why and by what hyperspectral technique could fulfill the expected task.

2.1. Influence on receptor from foreign gene

2.1.1 Influence on receptor from foreign gene

Foreign (or modified) gene may cause two kinds of influences to the receptor: one is influence because of expression of gene and its associated effects (e.g. gene expression polymorphism), and the other is influence because of non-gene-expression but with associated effects (e.g. gene silencing). When Gene transferred into receptor is expressed, sequence of amino acids is specified by coding region of the gene, these sequences of amino acids will make up the proteins (Sanderson, 2007). By the proteins made up, difference, bio-physiological or biochemical, would be made, no matter new material, nor changes of content of material, internal structure of cell or organism, intensity of biophysical process. However, because gene expression is susceptible, it shows the characteristics of uncertainty (Sanderson, 2007) and stochasticity (Kaern, et al., 2005, Raser and O'Shea, 2005). Therefore, in some situations, through the gene is not expressed, it still makes influences on the receptor.

For limiting our study scope, we should get some general judgment or assumption about samples in the study: If foreign genes are expressed in plant cell/ organism, there would made some differences comparing to their parent, these differences are, not limited: components of cell are changed such as new material generated or the content of existing component increasing/ decreasing; structure of cell would be changed; biochemical or biophysical processes would be influenced. If changes happen, component or structure of leaf organism, they would be discovered by direct or indirect approaches of spectroscopy equipments with certain sensitivity and spectral resolution.

2.1.2 Laboratory approaches of indication and monitoring of gene expression

In general, the tools applied for detecting genes (Figure 1) are primarily bioassays: protein based (mainly immunological) assays and DNA based assays (mainly applying the polymerase chain reaction [PCR] technology) (Holst-Jensen, 2009). Some researchers also classified the main categories of detection strategies as polymerase chain reaction (PCR)-based methods, DNA-based approaches (which may or may not involve the use of PCR) (Deisingh, et al., 2005). The high sensitivity and specificity of PCR allows it to be the best choice in detecting transgenes. The quality and quantity of target analyte will also influence the method of choice and again PCR methods are generally chosen (Deisingh and Badrie, 2005, Holst-Jensen, et al., 2003). These are the dominating techniques to identify and detect gene existence and other relevant information.

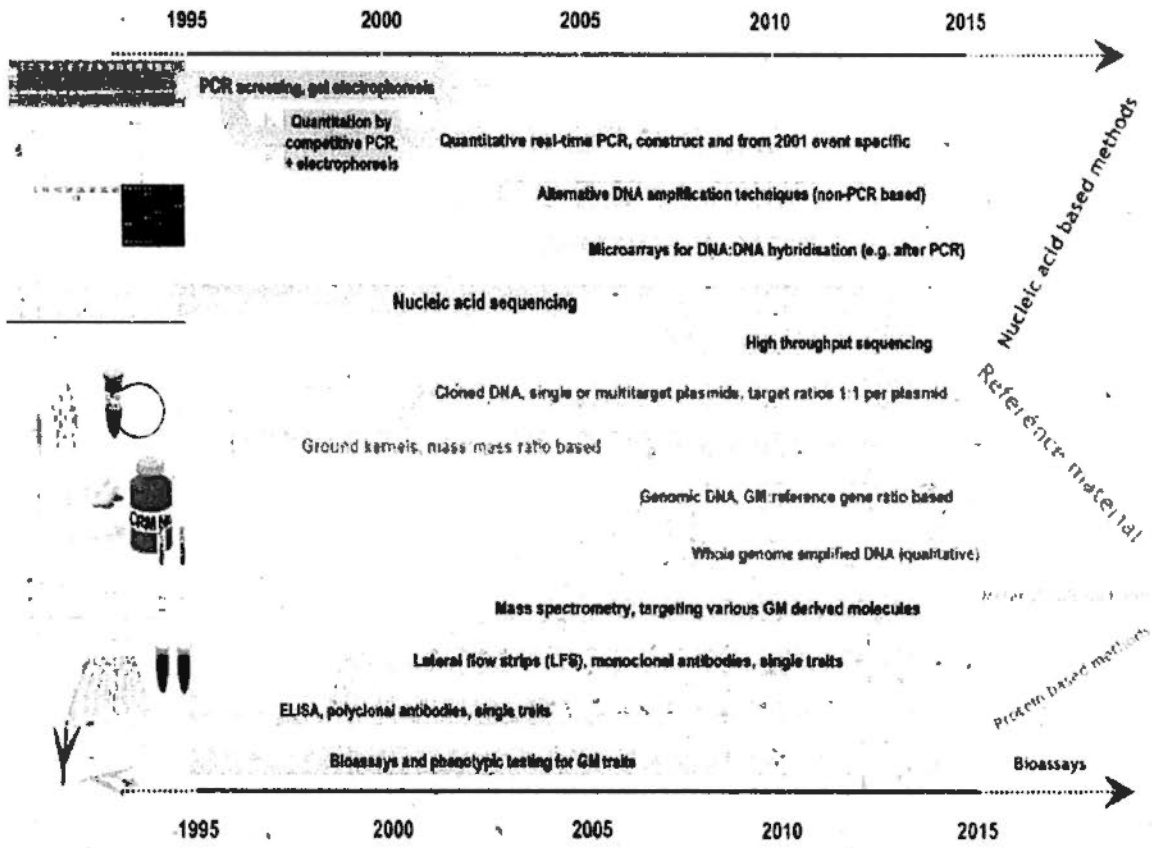


Figure 2-1. The tools applied to detect genes (Holst-Jensen, 2009)

Besides DNA- or protein- based methods, other technologies have been developed too, such as NIR Spectroscopy (Hurburgh, et al., 2000, Roussel, et al., 2001) and Chromatography (Byrdwell, et al., 1996). Microchip electrophoresis (ME), with its high separation efficiency, short analysis time, and low sample and solvent consumptions, was investigated for the analysis of genetically modified organisms in maize (Kumar, et al., 2007). Wavelength-dispersive X-ray fluorescence (WDXRF) was also proposed for the determination of phosphorus in transgene food samples (Jastrzebska, et al., 2003). Fluorescence is the favored signaling technology and several techniques relying on energy transfer between a fluorophore and a proximal quencher molecule (Deisingh and Badrie, 2005, Whitcombe, et al., 1999). Now Green Fluorescence Protein has been widely applied as indicator of gene expression (Chalfie, 2009, Chalfie, et al., 1994, Harper, et al., 1999, Misteli, et al., 1997).

2. 1.3 Plant responses to transferred gene

When we discuss the response of plant to gene influence, we could not avoid the response of plant to stress. The influence caused by foreign genes could be treated as a special internal stress. They are also involved to respond stress, such as signaling, protection of proteins (NDong, et al., 2002, Wang, et al., 2000). There would be changes within plant as a result of stress, such as physiological, anatomical, morphological, biochemical, and even molecular ones(Jackson, 1986, Wang, et al., 2003, Wang, et al., 2000).

Jackson (1986) defined stress *"as any disturbance that adversely influence growth"*. Lichtenthaler et al. (1988, 1996) gave extension that stress should be divided into Eu-stress and dis-stress. Eu-stress *"is an activating and simulating stress and a positive element for plant development, whereas dis-stress is a severe and a real stress that cause damage and thus has a negative effect on the plant and its development"* (Lichtenthaler and Rinderle, 1988, Lichtenthaler, 1996). No matter how to define it, in common sense, stress makes influence on plant growth and productivity (Wang, et al., 2000). Therefore, to detect and monitor stress is also important to assist transgenic crop breeding and cultivating.

Theoretically, it is hard to classify what response within plant is to stress and what is to gene by hyperspectral approach, but the growth conditions could be controlled to eliminate the uncertainties caused by stress (external factors). At least, detecting response to stress or assessing the capability of tolerance of transgenic sample is also helpful to breeding. It could provide messages from the cultivators and assist to make the experiment under control. So in some sense, the influence caused by transgene could be treated as a type of stress, therefore the same spectral techniques can be applied to detect and monitor them.

2.2 Hyperspectral techniques for detecting influences of transgene and stress

Whatever the influences or the effects caused by foreign gene or stress, the responses are firstly at a micro level, such as photosynthetic pigment content changes(Carter, GA, et al., 2001), denaturation of functional and structural proteins(Smirnoff, 1998). By hyperspectral technique, it is hard to detect directly. However, by monitoring the growth of sample real-/ near real-time by spectral techniques in contrast conditions, it is promising to apply this technique to indicate the growth statues of the samples, assess and screen them quantitatively to help to fulfill screening of samples from a macro-view.

As we know, hyperspectral technique is a kind of spectroscopy that is able to detect the existence of material and even internal structure of specific object. Moreover, if we have a time series data set, we could fulfill the work of monitoring the development of the plant. Vegetation has a perfect mechanism to respond the stress, internal or external. When this mechanism starts, the activity of biochemical or physiological process would be changed. Taking photosynthetic activity as an example, when plant encounters hot temperature, is kind of external stress, plant photosynthetic process would die down for self-protection. Therefore, by finding and tracking the stress encountered by plant, useful information of plant growth changes could be found.

2.2.1 Introduction to spectroscopy

The fundamental of hyperspectral technique is spectroscopy (Tong, et al., 2006), especially absorption spectroscopy. *"Absorption spectroscopy refers to spectroscopic techniques that*

measure the absorption of radiation, as a function of frequency or wavelength, due to its interaction with a sample." [4] It is a kind of electromagnetic spectroscopy (figure 2). *"Electromagnetic spectroscopy involves interactions of matter with electromagnetic radiation."*[5] By studying these interactions, information of atom, molecular and its internal structure could be obtained. Theoretically, there are three types of absorption spectrum (figure 3): atom absorption spectrum, molecular spectrum (absorption, vibration and rotation) and crystal lattice vibration spectrum. According to the electromagnetic theory, the generation of the spectrum of any material has its strictly physical rule (Pu, et al., 2000). Because of Electron transition in atom or molecular from one energy state to another, atom or molecular could absorb or emit certain frequency electromagnetic radiation and form spectrum with unique features. Spectrum because of lattice vibration is related to the structure of crystal lattice. Thus for specific crystal, it has specific spectral characteristics. By detecting the signal of these interactions (absorption or reflection of electromagnetic wave) Spectral Reflectance Curves, also could be defined as Reflectance Spectra would be obtained. Reflectance Spectra quantitatively assesses the percentage of the energy reflected by the object to the incident energy. Because the limitation of equipment, the wavelength range of 350- 2500nm is the study scope, covering visible band (380-760nm), near Infrared-red band (NIR: 760- 1400nm) and part of short-wave infrared-red band (SWIR: 1400-3000nm)[6].

[4] The words in "" were cited from Wikipedia: http://en.wikipedia.org/wiki/Absorption_spectrum

[5] http://en.wikipedia.org/wiki/Spectroscopy

[6] http://en.wikipedia.org/wiki/Near_infrared

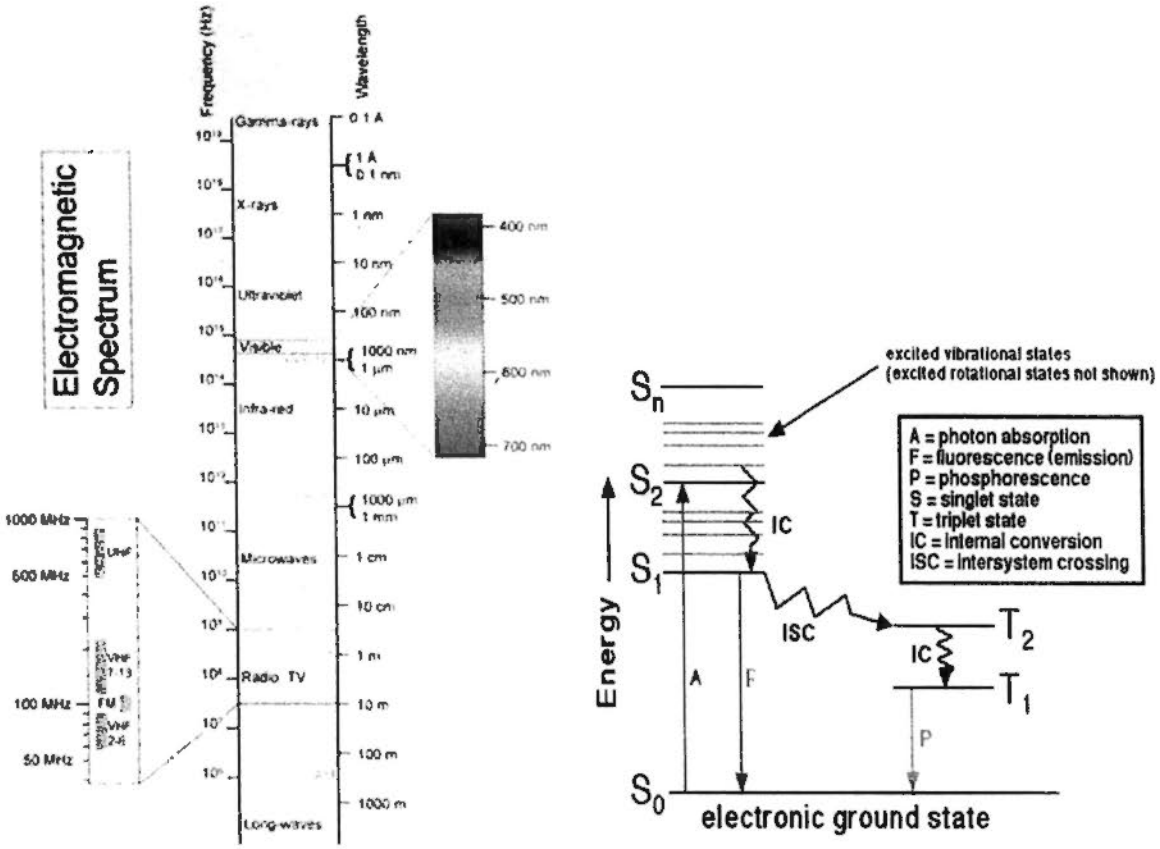
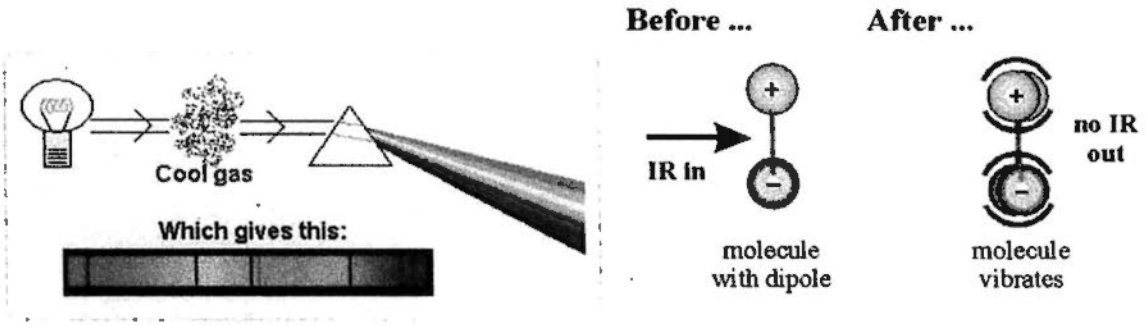


Figure 2-2. Generalized diagram of Electromagnetic bands and theory of spectroscopy

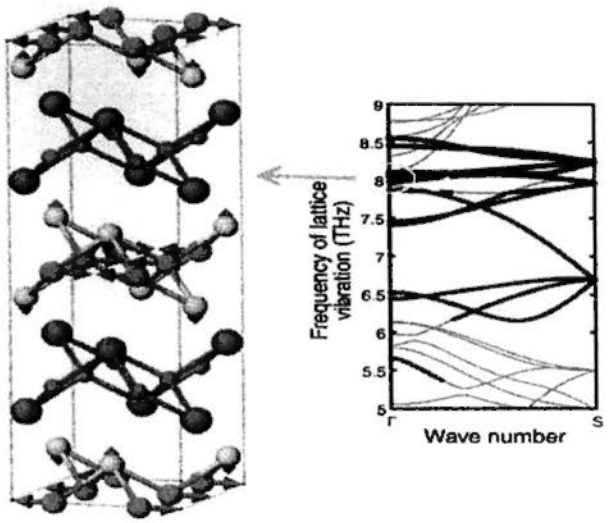
(a).Electromagnetic-spectrum (right)[7]; (b). Jablonski diagram: Interactions of matter with electromagnetic radiation (left).[8]



(a) (b)

[7] http://en.wikipedia.org/wiki/File:Electromagnetic-Spectrum.png

[8] http://www.chemicool.com/definition/jablonski_diagram.html.



(c)

Figure 2-3. Three kinds of spectrum discussed

(a)Atom absorption spectrum (right)[9]; (b) Molecule vibration (left)[10]; (c) Crystal lattice vibration and lattice vibration spectrum: Visualization of the lattice vibration in the iron-based superconductor LaFeAsO and its dispersions[11]

Figure 4 listed 44 absorption features in visible and near-infrared wavebands that had been related to particular foliar chemical concentrations (Curran, 1989).

[9] http://www.green-planet-solar-energy.com/atomic-absorption-spectrum.html

[10] http://qbab.aber.ac.uk/roy/ftir/absorb.htm

[11] http://jolisfukyu.tokai-sc.jaea.go.jp/fukyu/mirai-en/2009/12_3.html

Wave length (μm)	*Electron Transition or Bond Vibration*	*Chemical(s)*	*Remote Sensing Considerations*
0.43	Electron transition	**Chlorophyll a**[a]	Atmospheric scattering
0.46	Electron transition	**Chlorophyll b**	
0.64	Electron transition	**Chlorophyll b**	
0.66	Electron transition	**Chlorophyll a**	
0.91	C—H stretch, 3rd overtone	Protein	
0.93	C—H stretch, 3rd overtone	Oil	
0.97	O—H bend, 1st overtone	**Water**, starch	
0.99	O—H stretch, 2nd overtone	Starch	
1.02	N—H stretch	Protein	
1.04	C—H stretch, C—H deformation	Oil	
1.12	C—H stretch, 2nd overtone	Lignin	
1.20	O—H bend, 1st overtone	**Water**, cellulose, starch, lignin	
1.40	O—H bend, 1st overtone	**Water**	
1.42	C—H stretch, C—H deformation	Lignin	
1.45	O—H stretch, 1st overtone, C—H stretch, C—H deformation	Starch, sugar, lignin, water	Atmospheric absorption
1.49	O—H stretch, 1st overtone	Cellulose, sugar	
1.51	N—H stretch, 1st overtone	**Protein, nitrogen**	
1.53	O—H stretch, 1st overtone	Starch	
1.54	O—H stretch, 1st overtone	Starch, cellulose	
1.58	O—H stretch, 1st overtone	Starch, sugar	
1.69	C—H stretch, 1st overtone	**Lignin**, starch, protein, nitrogen	
1.78	C—H stretch, 1st overtone/ O—H stretch/H—O—H deformation	**Cellulose, sugar**, starch	
1.82	O—H stretch/C—O stretch, 2nd overtone	Cellulose	
1.90	O—H stretch, C—O stretch	Starch	
1.94	O—H stretch, O—H deformation	**Water**, lignin, protein, nitrogen, starch, cellulose	Atmospheric absorption
1.96	O—H stretch/O—H bend	Sugar, starch	
1.98	N—H asymmetry	Protein	
2.00	O—H deformation, C—O deformation	Starch	
2.06	N=H bend, 2nd overtone/ N=H bend/N—H stretch	Protein, nitrogen	
2.08	O—H stretch/O—H deformation	Sugar, starch	
2.10	O=H bend/C—O stretch/ C—O—C stretch, 3rd overtone	**Starch**, cellulose	
2.13	N—H stretch	Protein	
2.18	N—H bend, 2nd overtone/ C—H stretch/C—O stretch/ C=O stretch/C—N stretch	**Protein, nitrogen**	
2.24	C—H stretch	Protein	Rapid decrease in signal-to-noise ratio of sensors
2.25	O—H stretch, O—H deformation	Starch	
2.27	C—H stretch/O—H stretch CH_2 bend/CH_2 stretch	Cellulose, sugar, starch	
2.28	C—H stretch/CH_2 deformation	Starch, cellulose	
2.30	N—H stretch, C=O stretch, C—H bend, 2nd overtone	Protein, nitrogen	
2.31	C—H bend, 2nd overtone	**Oil**	
2.32	C—H stretch/CH_2 deformation	Starch	
2.34	C—H stretch/O—H deformation/ C—H deformation/O—H stretch	Cellulose	
2.35	CH_2 bend, 2nd overtone, C—H deformation, 2nd overtone	Cellulose, protein, nitrogen	

Figure 2-4. 44 absorption features

Absorption features in visible and near-infrared wavebands that have been related to particular foliar chemical concentrations (Curran, 1989)

2.2.2 Optical characteristics of plant leaf

Vegetation spectrum, responding to electromagnetic energy, composed of absorption, reflectance and emission of energy, is decided by vegetation chemical and morphological features which are highly related to the development health and growth conditions of plants (Boochs, 1990). Green vegetation has spectral characteristics different than soil, water and other typical ground objects (figure 1). Vegetation spectrum, responding to electromagnetic energy, composed of absorption and emission of energy, is decided by vegetation chemical and morphological features which are highly related to plant health and growth conditions of plants (Boochs, 1990). In visible bands, kinds of pigments, especially chlorophyll, are the main materials responding to electromagnetic energy and absorb large of incident light at regions which central band are 450 nm (blue light) and 650 nm (red light) respectively (Jago el at., 1999). Between these two absorptions, there is a small reflected peak caused by weaker absorption, we call it *"green peak"*, thus we could find plant is green. When plant is in unhealthy state, absorption by chlorophyll would decrease while reflection increases especially in red bands, thus we would find now plant is yellowish. In near Infrared bands, the responses are mainly controlled by internal cell structure of leaf (Jago el at., 1999; Kumar, 2001). Healthy vegetation has a very high reflection (up to 45%-50%), high transmission (up to 45%- 50%) and low absorption (5% approximately) at near infrared red bands (Philip et al., 1978). Therefore, at region red and near infrared red, at 760 nm approximately, there is a red-shift, namely red-edge would be formed. It is the most significant characteristic of green vegetation (Miller et al., 1991). The spectral region between 1300 and 2500 nm, namely around 1400 nm, 1900 nm, is of interest because water within the leaves absorbs radiation at these wavelengths. Within this region, called the "*mid-infrared*" or the "*water-absorption*" region, leaf reflectance decreases towards long wavelength, with minimums near 1400 and 1850 nm, and becomes negligible beyond 2500 nm. The upper limit of 2500 nm is a result of the decrease of solar radiation with wavelength and the absorption of radiation by atmospheric water vapor (Jackson, 1986). Study shows reflection of leaves at middle infrared bands would decrease associated with decrease of water content of leaves. The specific vegetable spectral curve shape (characteristics) could be found in figure 1.

Previous studies indicated that by vegetable spectrum, we could obtain information both of surface and inner leaf. In the past decades, researchers have had many achievements with these vegetation spectral characteristics and built lots of models/ relationship between spectral responses and vegetable bio-physical indices revealing internal regulations such as plant growth,

environmental stress, and most of these have a wide and successful application.

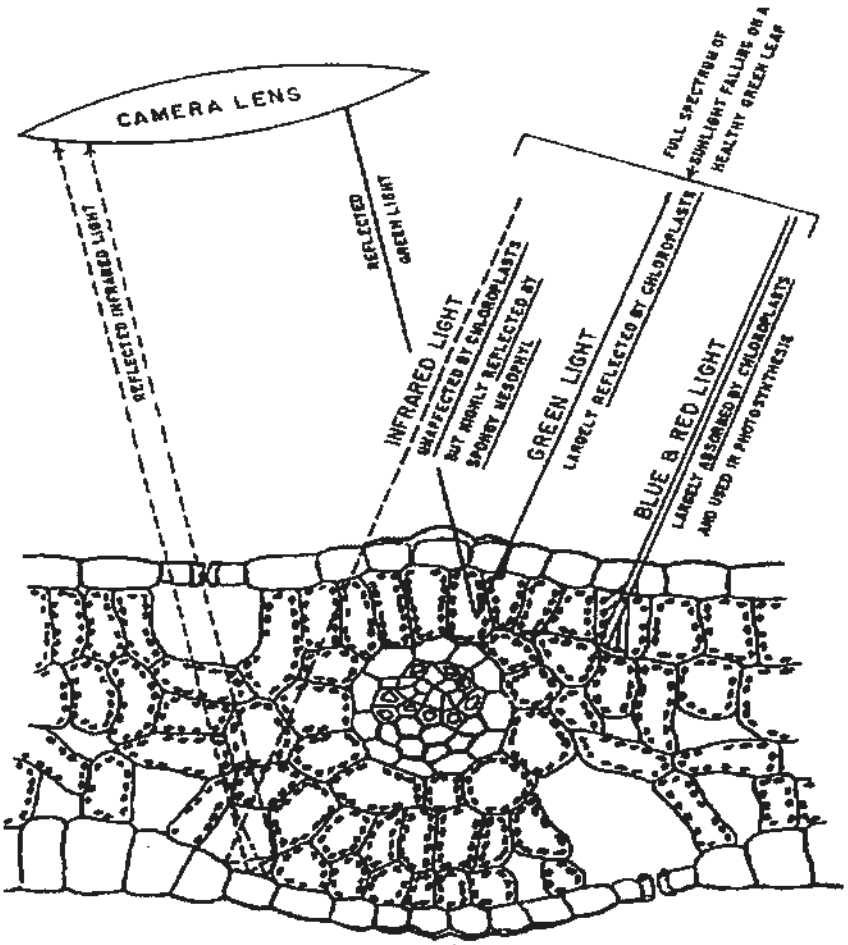


Figure 2-5. Generalized diagram of a leaf' structure and its reflectance characteristics at visible and near IR wavelength[12]

2.2.3 Bands response to stress

In some sense, influence caused by transgene on the receptor could be treated as special stress within plant. Thus we could detect and monitor these two influences by the same spectral techniques. In fact, by current hyperspectral technique, it is impossible to reveal influences mentioned by stress. From a macro view, the stress effect could be responded by spectral technique (Jackson, 1986). Under different stresses, plant would have a different healthy state which represents distinguished characteristics responded by spectral signature. Leaf reflectance responses to environmental conditions that inhibit growth generally involved increased reflectance in the visible (380- 760 nm) or infrared (760- 2500 nm) spectra (Carter, 1993). Carter (1994, 1993) pointed out about sensitive bands and indicated that leaf reflectance altered by stress was more consistent at visible bands (400- 720 nm) rather than it after band of 730nm. A

[12] http://www.geog.ucsb.edu/~jeff/115a/remote_sensing/remotesensing.html

further result showed that leaf optical properties in a relatively narrow spectral band near 700 nm are crucial for plant stress detection (Carter, GA and Knapp, AK, 2001). Because reflectance generally increases at wavelengths near 700 nm with plant stress, the steep slope of the reflectance curve in the far-red to near-infrared transition (namely red shift) spectrum tends to shift toward the blue spectrum, namely the blue shift of the reflectance curve red edge. But to specific stress, sensitive bands may be not same.

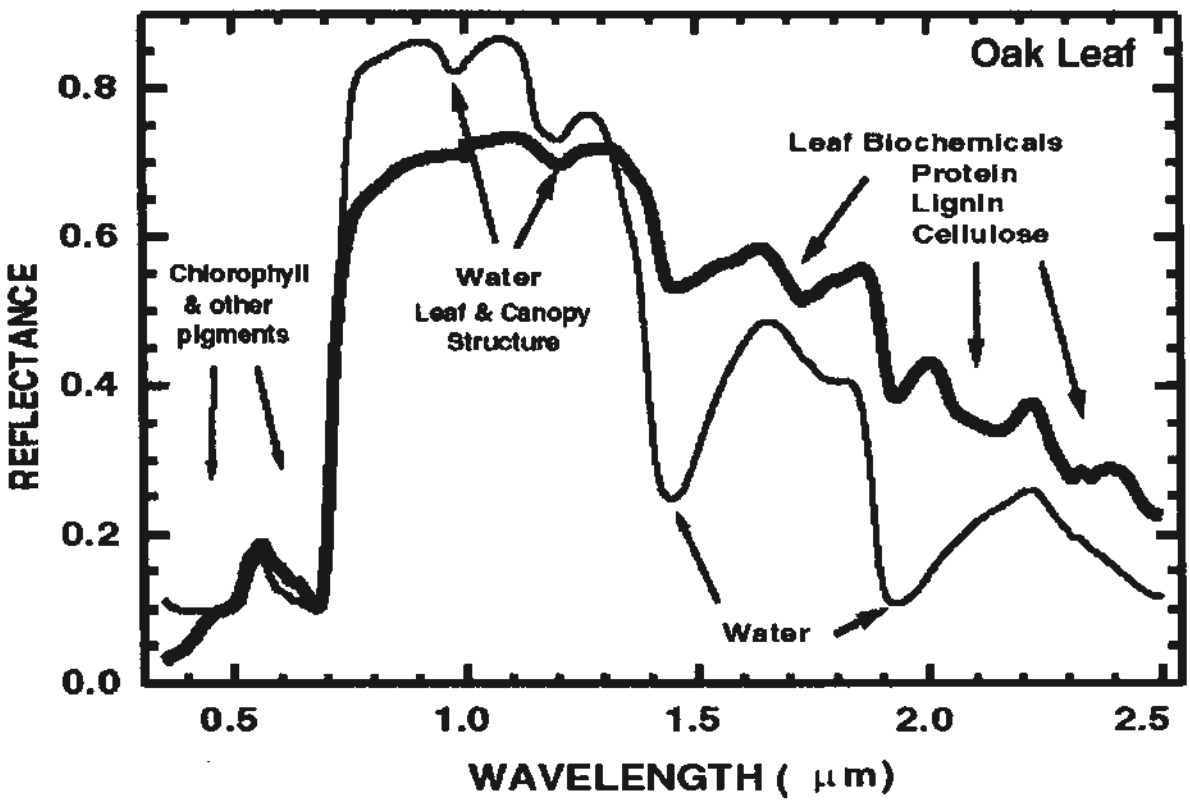


Figure 2-6. The typical vegetation spectrum

Laboratory reflectance spectra of an oak leaf in a fresh state (thin line) and after being dried (thick line). Because the strong absorptions due to water are absent, the dried leaf spectrum shows the protein, lignin and cellulose absorption features in the 1.5-2.5 micron region (Kokaly, et al., 1998).

2.3. Hyperspectral remote sensing techniques

2.3.1 Concept of hyperspectral remote sensing techniques

Hyperspectral remote sensing is a kind of remote sensing technique which obtains data with high-spectral resolution, and its theoretic fundamental is spectroscopy (Tong, et al., 2006). It was discussed formally in 1985 when disusing the technique of imaging spectrometry (Goetz, et al., 1985a). Hyperspectral remote sensing is closely related to "*imaging spectrometry*" at first. As described by Goetz (Goetz, et al., 1985a), imaging spectrometry "*consists of the acquisition of*

images in many narrow contiguous spectral bands throughout the visible and solar-reflected infrared spectral bands simultaneously", "for each pixel a radiance spectrum can be derived." The data set acquired by imaging spectrometry could be called hyperspectral images. This kind of is a three-dimensional data cube, both with spatial and spectral information of the observed object. Compared with multi-spectral remote sensing, the significant character of hyperspectral remote sensing is the data acquired simultaneously in many, hundreds of, narrow contiguous spectral bands for directly indentify material with diagnostic absorption or reflected features. Thus, by spatial features, the distribution and amount of the objects could be found; by spectral information, one pixel one spectrum, the attribute of the material in pixel could be detected. Theoretically, the spectral range could be extended from 400nm to 2500nm (Goetz, et al., 1985a). Hyperspectral remote sensing is a definition compared to multi-spectral remote sensing. The former one concentrates the bands of image are contiguous (more number of spectral bands), even overlapping so that the diagnosed spectral features of the material could be detected (Goetz, 2009, Sankaran, et al., 2010). The figure 3 showed the two kinds of remote sensing data clearly, the one former one was hyperspectral data, the other was multispectral.

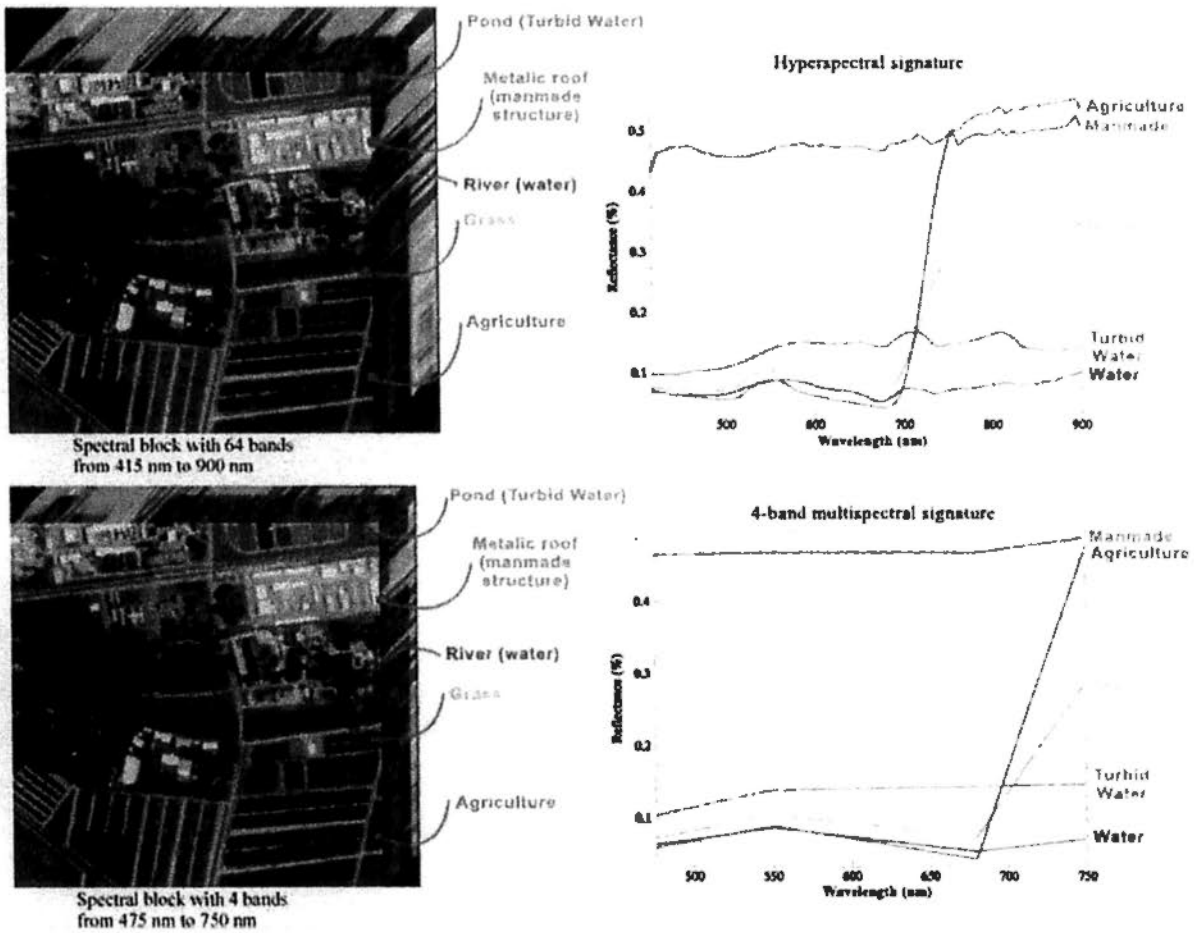


Figure 2-7. Spectral signatures from hyperspectral vs. multispectral sensors[13]

With development of technique and scientific demand, hyperspectral remote sensing has to be

[13] https://www.e-education.psu.edu/geog883/l7_p4.html

multi-platforms: satellite platform (MODIS: the Moderate Resolution Imaging Spectroradiometer), airborne platform (AVIRIS: Airborne Visible Infra-Red Imaging Spectrometer) and ground-based spectrometry (FieldSpec® 3 Hi-Res Portable Spectroradiometer). The first two devices are imaging spectrometer, and could obtain a real spatial-spectral-cube data, while the last only acquire spectral data.

2.3.2 The historical, current and future development of hyperspectral remote sensing

The first truly hypsepectral remote sensing image acquired by AIS-1 in 1983(Tang, 2004, Vane, et al., 1984). However the first formal definition was given by Goetz in 1985 (Goetz, et al., 1985a). AIS covers the spectral range of [1200, 2.400nm] in 128 contiguous spectral bands with 9.3nm bandwidth. After AIS, a new spectrometer had been designed as Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS). The spectral coverage of AVIRIS is 400nm to 2400nm in 224 bands 10nm wide. Compared with AIS, AVIRIS data attracted a wide studies and applications (Carder, et al., 1993, Kruse, et al., 1993, Roberts, et al., 1993). After AVIRIS, hyperspectral remote sensing entered a new era; lots of hyperspectral sensors have been developed, such as the Compact Airborne Spectrographic Imager (CASI) (Babey, et al., 1989, Zarco-Tejada, et al., 1999), hyperspectral digital imagery collection experiment sensor (HYDICE) (Resmini, et al., 1997, Zhang, et al., 2006), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Friedl, et al., 2002, Huete, et al., 2002), High Resolution Hanging Spectrometer(HIRIS)(Martin, et al., 1997, Rock, et al., 1988), the Hyperion Imaging Spectrometer (Pearlman, et al., 2003). In China, there are also some hyperspectral sensors such as Modular Airborne Imaging Spectrometer (MAIS), Operational Module Imaging Spectrometer (OMIS), Push-broom Hyperspectral Imager (PHI), Wide Angle Push-broom Hyperspectral Imager (WHI), Large Aperture Spectral Imaging System (LASIS) and HJ-1A and CE-1.

Ground-based spectrometer also has a good development. The first truly portable field spectrometer (PFRS) was designed in 1970' covered spectral range from 400nm to 2500nm (Goetz, 2009). This device needs a long time up to 30 seconds for a measurement. Thus new generation of ground spectrometers have been designed, such as the products from Geophysical Environmental Research (GER) in New York, the most widely applied spectrometers nowadays are products of Analytical Spectral Devices (ASD) Company. The new spectrometer could cover the spectral range of solar reflected radiance, 350nm to 2500nm and one measurement within 100ms. And the spectral resolution is up to 3nm for some specific wavelength, and band interval is up to 1nm. Thus, more fine and precise spectral reflectance data could be obtained and applied to build and calibrate model for studies of various principles including calibrate satellite image data. Many remote sensing studies begin based on spectral reflectance data acquired by ground-based spectrometer and then extend to image data.

2.3.3 Hyperspectral remote sensing of vegetation

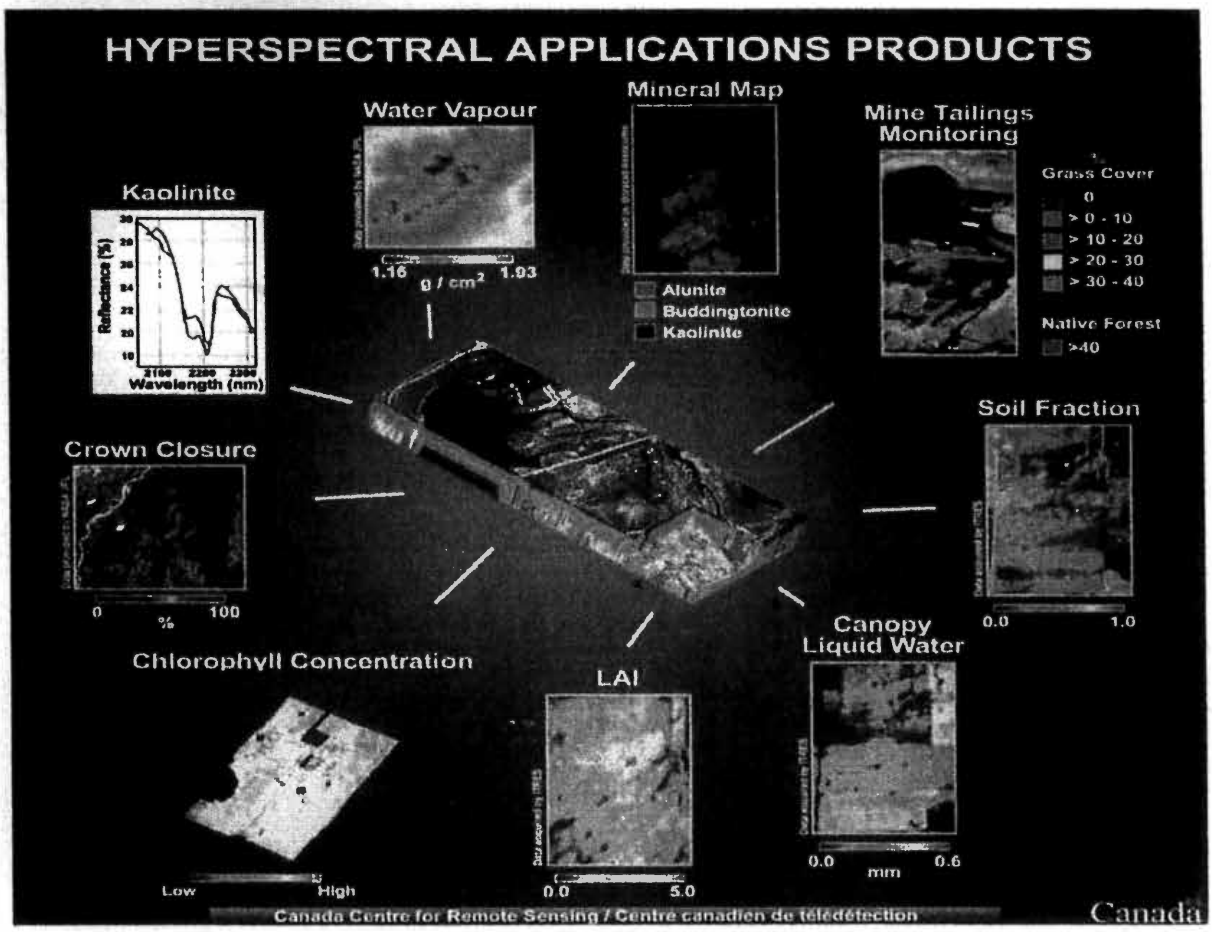


Figure 2-8. Application of Hyperspectral remote sensing (Miglani, 2007.7)

Because of precise contiguous bands of spectral reflectance data, diagnosed features of material could be detected. Thus, hyperspectral remote sensing technique is widely applied to various fields. Figure 4 showed some application fields of this technique. Since the special relationship between electromagnetic energy and relevant response of plant leaf in spectral range of 350nm to 2500nm, lots of studies had been concentrated on vegetation based on hyperspectral remote sensing technique.

Curran (1989) reported 44 absorption features in visible and near-infrared wavebands that have been related to particular foliar chemical concentrations. It makes it possible to predict changes of vegetation chemical compounds and monitoring plant growth. These chemical compounds are highly related to plant growth, yields including photosynthetic pigments (Chlorophylls: Chl; Carotenoids: Car; Anthocayains: Anth), water content, nitrogen, lignin, cellulose and protein etc. Plant pigments, mainly related to photosynthesis, are very sensitive to stresses and important indices to plant health. They are also the principal factors influencing plant spectral feature in visible and near infrared red bands. By studying relationship between pigment content and its spectral reflectance, we could not only detect and monitor plant health status but also obtained more information related to leaf structures, plant developmental stages (Sims, DA, et al., 2002).

To solve this problem, pigment content should be estimated by spectral technique. Previous studies proved many high related bands to pigments content and built correlation models. Nitrogen is an important indicator of photosynthetic rate and overall nutritional status (Curran, 1989). Leaf lignin concentration is an important factor to control plant growth (Serrano, et al., 2002). These all important foliar chemicals could are promising to be used as indicators of plant growth. The study details of the relevant foliar chemicals would be reviewed in chapters 4.

References

Babey, S. K., & Anger, C. D. (1989). A compact airborne spectrographic imager (casi). 2, 1028-1031.

Byrdwell, W., & Neff, W. (1996). Analysis of genetically modified canola varieties by atmospheric pressure chemical ionization mass spectrometric and flame ionization detection. *Journal of liquid chromatography & related technologies*, 19(14), 2203-2225.

Carder, K. L., Reinersman, P., Chen, R. F., Muller-Karger, F., Davis, C. O., & Hamilton, M. (1993). Aviris calibration and application in coastal oceanic environments. 44(2-3), 205-216.

Carter, G. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, 15(3), 697-703.

Carter, G., & Knapp, A. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684.

Carter, G. A. (1993). Responses of leaf spectral reflectance to plant stress. *American Journal of Botany*, 80(3), 239-243.

Chalfie, M. (2009). Gfp: Lighting up life. *Proceedings of the National Academy of Sciences*, 106(25), 10073-10080.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., & Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. *Science (New York, NY)*, 263(5148), 802-805.

Curran, P. (1989). Remote sensing of foliar chemistry. *Remote Sensing of environment*, 30(3), 271-278.

Deisingh, A., & Badrie, N. (2005). Detection approaches for genetically modified organisms in foods. *Food Research International*, 38(6), 639-649.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., & Cooper, A. (2002). Global land cover mapping from modis: Algorithms and early results. 83(1-2), 287-302.

Goetz, A. (2009). Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of environment*, 113, S5-S16.

Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for earth remote-sensing. *Science*, 228(4704), 1147-1153.

Harper, B., Mabon, S., Leffel, S., Halfhill, M., Richards, H., Moyer, K., & Stewart Jr, C. (1999). Green fluorescent protein as a marker for expression of a second gene in transgenic plants. *Nature*

biotechnology, 17(11), 1125-1129.

Holst-Jensen, A. (2009). Testing for genetically modified organisms (gmos): Past, present and future perspectives. *Biotechnology Advances*, 27(6), 1071-1082.

Holst-Jensen, A., R nning, S., L vseth, A., & Berdal, K. (2003). Pcr technology for screening and quantification of genetically modified organisms (gmos). *Analytical and bioanalytical chemistry*, 375(8), 985-993.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the modis vegetation indices. 83(1-2), 195-213.

Hurburgh, C., Rippke, G., Heithoff, C., Roussel, S., & Hardy, C. (2000). Detection of genetically modified grains by near-infrared spectroscopy. 12-17.

Jackson, R. D. (1986). Remote sensing of biotic and abiotic plant stress. 24(1), 265-287.

Jastrzebska, A., Brudka, B., Szyma¨½ski, T., & Sz yk, E. (2003). Determination of phosphorus in food samples by x-ray fluorescence spectrometry and standard spectrophotometric method. *Food Chemistry*, 83(3), 463-467.

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451-464.

Kokaly, R., Clark, R., & Livo, K. (1998). Mapping the biology and mineralogy of yellowstone national park using imaging spectroscopy. 7, 97-21.

Kruse, F. A., Lefkoff, A. B., & Dietz, J. B. (1993). Expert system-based mineral mapping in northern death valley, california/nevada, using the airborne visible/infrared imaging spectrometer (aviris). 44(2-3), 309-336.

Kumar, K. S., & Kang, S. H. (2007). Ultra-fast simultaneous analysis of genetically modified organisms in maize by microchip electrophoresis with lif detector. *ELECTROPHORESIS*, 28(22), 4247-4254.

Lichtenthaler, H., & Rinderle, U. (1988). The role of chlorophyll fluorescence in the detection of stress conditions in plants. *CRC Crit. Rev. Anal. Chem*, 19(Suppl I), S29-S85.

Lichtenthaler, H. K. (1996). Vegetation stress: An introduction to the stress concept in plants. 148(1), 4-14.

Martin, M. E., & Aber, J. D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. 7(2), 431-443.

Miglani, A. (2007). Hyperspectral remote sensing - an overview. *The Global Geospatial Magazine*, 11(7), 41-43.

Misteli, T., & Spector, D. (1997). Applications of the green fluorescent protein in cell biology and biotechnology. *Nature biotechnology*, 15(10), 961-964.

NDong, C., Danyluk, J., Wilson, K., Pocock, T., Huner, N., & Sarhan, F. (2002). Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins. Molecular characterization and functional analyses. *Plant physiology*, 129(3), 1368-1381.

Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., & Carman, S. L. (2003). Hyperion, a

space-based imaging spectrometer. 41(6), 1160-1173.

Pu, R., & Gong, P. (2000). Hyperspectral remote sensing and its applications. *Higher Education, Beijing*.

Raser, J., & O'Shea, E. (2005). Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743), 2010-2013.

Resmini, R. G., Kappus, M. E., Aldrich, W. S., Harsanyi, J. C., & Anderson, M. (1997). Mineral mapping with hyperspectral digital imagery collection experiment (hydice) sensor data at cuprite, nevada, USA. 18(7), 1553-1570.

Roberts, D. A., Smith, M. O., & Adams, J. B. (1993). Green vegetation, nonphotosynthetic vegetation, and soils in aviris data. 44(2-3), 255-269.

Rock, B. N., Hoshizaki, T., & Miller, J. R. (1988). Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline. 24(1), 109-127.

Roussel, S., Hardy, C., Hurburgh, C., & Rippke, G. (2001). Detection of roundup ready(tm) soybeans by near-infrared spectroscopy. *Applied Spectroscopy*, 55(10), 1425-1430.

Sanderson, C., (2007). *Understanding genes and gmos*. World Scientific Pub Co Inc.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, 72(1), 1-13.

Serrano, L., Pe uelas, J., & Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in mediterranean vegetation from aviris data:: Decomposing biochemical from structural signals. *Remote Sensing of environment*, 81(2-3), 355-364.

Sims, D., & Gamon, J. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of environment*, 81(2), 337-354.

Smirnoff, N. (1998). Plant resistance to environmental stress. *Current Opinion in Biotechnology*, 9(2), 214-219.

Tang, Y. (2004). Study on the hyperspectral characteristics and simulating and estimating models about biophysical and biochemical parameters of rice. 35-45.

Tong, Q., Zhang, B., & Zheng, L., (2006). *Hyperspectral remote sensing*. Beijing: Higher Education Press.

Vane, G., Goetz, A. F. H., & Wellman, J. B. (1984). Airborne imaging spectrometer: A new tool for remote sensing. 22(6), 546-549.

Wang, W., Vinocur, B., & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. 218(1), 1-14.

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2000). Biotechnology of plant osmotic stress tolerance physiological and molecular considerations. 285-292.

Whitcombe, D., Theaker, J., Guy, S., Brown, T., & Little, S. (1999). Detection of pcr products using self-probing amplicons and fluorescence. *Nature biotechnology*, 17, 804-807.

Zarco-Tejada, P. J., & Miller, J. R. (1999). Land cover mapping at boreas using red edge spectral parameters from casi imagery. 104(D22), 27921-27933.

Zhang, J., Rivard, B., S, chez-Azofeifa, A., & Castro-Esau, K. (2006). Intra-and inter-class spectral variability of tropical tree species at la selva, costa rica: Implications for species identification using hydice imagery. 105(2), 129-141.

Chapter 3 Experiment and its uncertainties

3.1 Guideline of the experiment

The procedure of gene expression is complex, and its influences on receptor are of stochasticity, diversity and variability. Therefore, in this study the expression of foreign gene and its influences on receptor were not explored directly. The author focused on the bio-traits or responses to stress of samples to detect spectral differences among samples. By analyzing these differences, the samples could be monitored real-/near real-time, the interested space of samples can be optimized and finally priori knowledge for laboratory work can be extracted. By this study, the author wanted to develop an operational and efficient approach based on hyperspectral remote sensing techniques to assist transgenic crop study (e.g. crop breeding) in large sample space condition.

Figure 1 showed the guideline of the study. All samples should be cultivated in contrast conditions controlled by professional biological technician. Based on fine spectra of sample leaves, indicative parameters (e.g. the content of the foliar chemicals) would be obtained. Then with help of these indicators, the following information could be confirmed:

- ✓ Whether there are any spectral differences or outliers (of spectral morphology parameter) among samples (transgenic one and their counterpart) exist?
- ✓ If yes, the spectral differences would be described quantitatively, located where the responding bands are, and deduced what caused differences.

Finally a report would be submitted to the laboratory for further study and validation. Spectral monitoring, laboratory study and field cultivation would respond feedback to each other to make crop breeding and screening efficient with low cost. Thus, the researcher would know his sample clearly real-/near real- time even with mass samples.

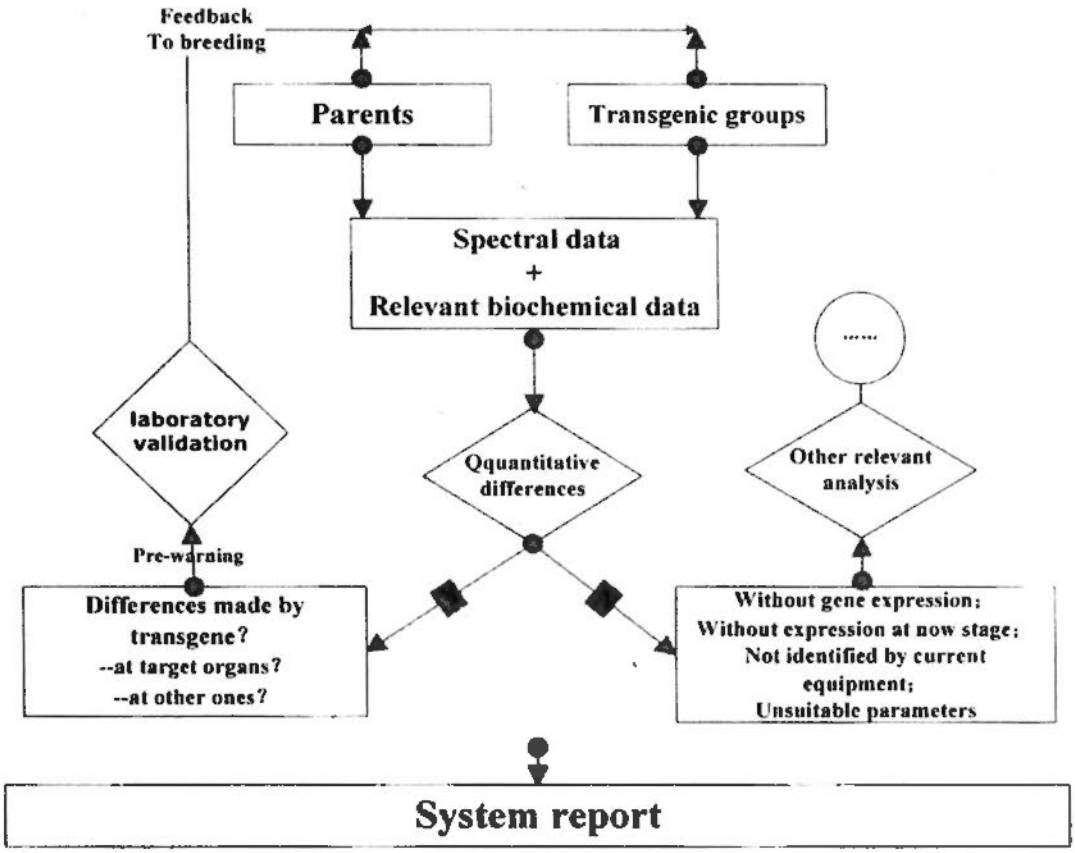


Figure 3-1. Guideline of the controlled experiment in this study

3.2 Collaborating institutes, samples and devices in the experiments

3.2.1 Collaborating institutes and samples

For validating and revising the approach proposed, the data of different types of transgenic samples and their contrast are required. Generally speaking, three kinds of data are needed, the data of same gene in different year and place, the data of different gene and the data from different growth stages. These procedures are recognized as "*the gene-repeat experiment*" and "*the experiment-repeat experiment*".

The study was supported by China National Hybrid Rice R&D Center (CNHRRDC) and Institute of Subtropics agriculture (ISA), Chinese Academy of Sciences (CAS) both in Hunan Province. CNHRRDC (figure 2) is a research institute of professional paddy rice steered by academician Yuan Longping. They have many professional breeding fields in Hunan and Hainan. ISA (figure 3) has a standard greenhouse so that study works could be done whole year and the crop growth condition could be controlled easily. 6 times experiments of more than two years were conducted

totally. The longest one lasted three months, and a long growth period data of paddy rice were obtained. And the samples of the 6 experiments were also the objects of professional breeding studies in CNHRRDC and ISA, and they had a strict requirement about the samples and their growth conditions. Therefore, it ensured that the proposed approach is matched with the study, such as crop breeding work.

(a) (b)

Figure 3-2. China National Hybrid Rice R&D Center(CNHRRDC)

CNHRRDC: (a) Hainan National Rice Breeding Field in Sanya, Hainan Province; (b) Transgenic paddy rice field in Changsha, Hunan Province.

(a) (b)

Figure 3-3. Institute of Subtropics agriculture (ISA)

(a) ISA, CAS; (b) Professional greenhouse in ISA, Changsha

The Samples were cultivated in two conditions, one was in the field in natural condition, and the other was in pot (20cm in diameter and 30 cm in height approximately) in the greenhouse or natural condition. In field, before flowering stage, the samples would be separated from the surrounding.

3.2.2 The gene-repeat and experiment-repeat experiment

3.2.2.1 The gene-repeat experiment

In the gene-(sample) repeat experiment, the same kind of transgenic paddy rice with their contrast parents was cultivated in different years to validate weather the proposed approach was useful and stable enough. The samples were transferred into different gene unit of phycocyanin genome. And by laboratory validation the genes had been successfully transferred. This genome was forecasted to promote the receptor's photosynthetic efficiency and help to produce rice of high quality. These samples were cultivated by China National Hybrid Rice R&D Center three times in Changsha, Hunan Province and Sanya, Hainan Province. We obtained spectral data at the same growth stage in these experiments.

3.2.2.2 The experiment-repeat experiment

To assess the sensitivity and stability of the proposed approach, we had done the experiment-repeat experiment. In these experiments, different kinds of transgenic samples were selected including phycocyanin gene, fluorescence protein gene units and BT & BAR gene. The last two genes ware forecasted to enhance resistance of herbicide and insects. The samples were transferred different gene units and cultivated in independent experiments, respectively.

3.2.3 Devices

(a) (b)

Figure 3-4. Devices

(a). an observation with probe-leaf-clip system; (b) a Field Spectroradiometer.

Considering that the differences between transgenic samples and their parent would be slight, fine spectra of leave were acquired. In the study, an integrated system was employed consisting of an Analytical Spectral Devices (ASD) FieldSpec 3 Spectrometer, the contact-probe and the leaf-clip (Figure 4). ASD is an instrument with a spectral range of 350-2500nm and a rapid data collection time of 0.1 second per spectrum, which is compact, field portable and accuracy. It has

a spectral resolution of 3 nm at 700nm and 10 nm at 1400/2100 nm while spectral interval is 1.4 nm for 350-1050 and 2nm for 1000-2500nm. In the study we set the view of field of ASD to 25 degrees. The Noise Equivalent Radiance (NEdL) is UV/VNIR 1.1*10-9 W/cm2/nm/sr @700 nm, NIR 2.4*10-9 W/cm2/nm/sr @1400 nm and NIR 4.7*10-9W/cm2/nm/sr @2100 nm[14]. More details about the ASD could be read from its website. The equipment is easy to use especially convenient for field data collection. High intensity Contact probe[15] is excellent for mineral, leaf, grain, and granule applications (ASD Document 600544 Rev. C, p26). Its view spot is in size of 10nm, Halogen bulb color temperature is 2901 +/- 10º% degree K and specular reflectance is 5% max off flat first surface mirror. Because of the design of the contact probe, the incident light to sensor is not perpendicular to the leaf surface but a constant angle (approximate 30° , Figure 5).

Figure 3-5. The contact-probe and the incident angle of light

Leaf-clip interfaces with the ASD high intensity probe. The contact probe and the leaf-clip consist of the pre-head of data collection. This unit had two kinds of reference panels. They are the white and the black standard background which are replaceable. The white background is 0.120 mm thick x .935" OD Gor-tex white PTFE reflector material. The black background is 0.004" thick x 0.935 OD black painted Vinyl. Figure 6 showed reflectance curves of background (standard reference panel) associated with leaf-clip and samples with white and black background respectively. Baseline-reference panel as standard reference is used to calibrate ASD spectrometer and calculate the spectral reflectance. In the ranges before 400nm, and after 2000nm, the black background is sensitive to noise. Except for those bands, the max reflectance is no more than 0.034, and the changes range from 0.031 to 0.034 approximately. For the white background, though the curve is smoother than the black one, but it causes high reflected shoulder (left figure) because of multi-reflection. These multi-reflections are complex and select to enhance reflected energy since the unique characteristics of leaf responding to incident electromagnetic wave. Thus, if the effect caused by background was acceptable, the black panel is an ideal choice for data acquisition. In this study, we took baseline-reference panel and black background as reference to calibrate reflectance data.

[14] http://www.asdi.com/products/fieldspec-3-portable-spectroradiometer

[15] http://www.asdi.com/accessories/high-intensity-contact-probe

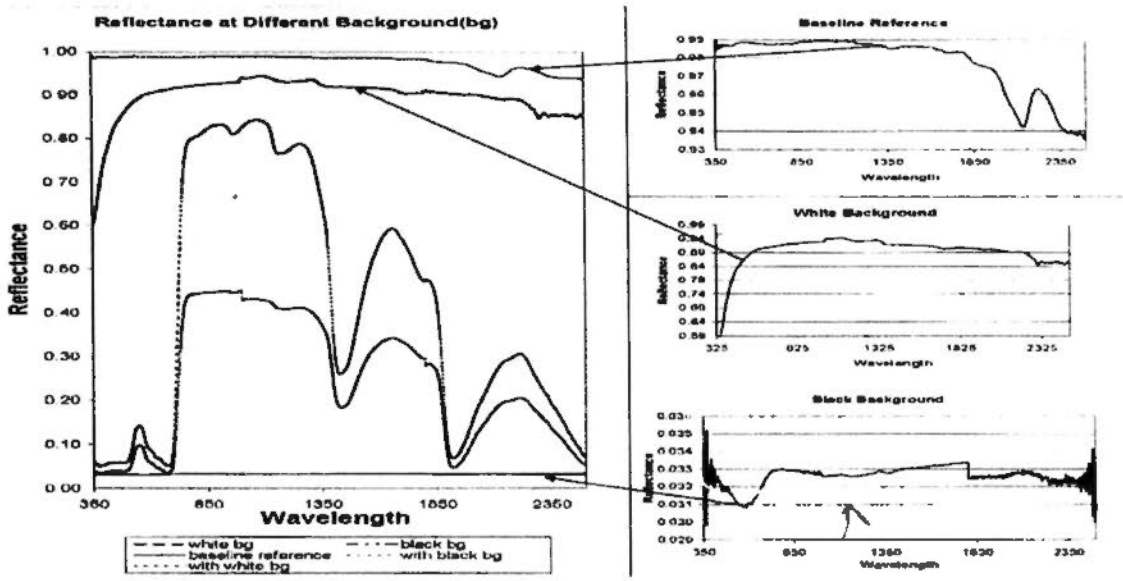


Figure 3-6. Spectral reflectance of the reference background (reference panel)

3.3 Uncertainty problems

3.3.1 The definition and classification of Uncertainties in remote sensing

"*Uncertainty: The lack of certainty, a state of having limited knowledge where it is impossible to exactly describe existing state or future outcome, more than one possible outcome.*"---Wikipedia[16]

It is an indicator to describe the certain influences caused by the factors which are not controlled or avoided under current techniques, or could be controlled in certain region (not solved) but unnecessary to be paid much attention to specific application which is accurate. Therefore, the uncertainty of remote sensing is certain, but this uncertainty could not be controlled certainly, mostly we could only give a qualitative description about it. For strictly contrast, the experimental uncertainty should be considered seriously during the experiment.

The first kind of uncertainty which is not controlled or avoided for remote sensing generally involves: (1) system errors or flaws, such as uncertainty of remote sensing including signal-to-noise ratio (SNR) of device, uncertainty of kinds of resolutions (spatial, temporal for time series data, spectral and radiation); (2) uncertainty of data itself such as uncertainty of time series data which also includes model uncertainty problems; (3) model uncertainty both empirical and physical models since the inaccuracy description or acceptance & refusal of variations. The second kind of uncertainty of remote sensing could be controlled in a certain region: (4) uncertainty during specific applications (processed) which mainly is caused by personal preference during model running; it's a kind of subjective and post- uncertainty. If before running model, a unique rule has been made to ensure standardized operation,

[16] http://en.wikipedia.org/wiki/Uncertainty

uncertainty problems could be finally controlled in the same group for the contrast experiment. Sometime, for some reasons, such as time or material consuming cost, manager would sacrifice some accuracy for saving, and this scarification also brings uncertainty which could be avoided.

3.2.2 Uncertainties in experiments

3.2.2.1 Time of the Measurement

Plant has its own mode in biochemical and physiological process to fit environment and stress such as high temperature and extra-light condition. Thus, plant is not active all the time (Figure 7). Taking photosynthesis as example, the following figure 7 showed the changes of photosynthesis in daytime. Thus for long time monitoring or contrast experiment, when to measure is very important. Observation time of different experiments must be matched. Otherwise, unreasonable results would be obtained and mislead further study. Especially for contrast experiments, spectral data measurements are required at the same time at the same growth conditions to suppress external uncertainties. Generally speaking, spectral measurement could be conducted during 10:30- 11:30 under a stable weather condition (natural condition).

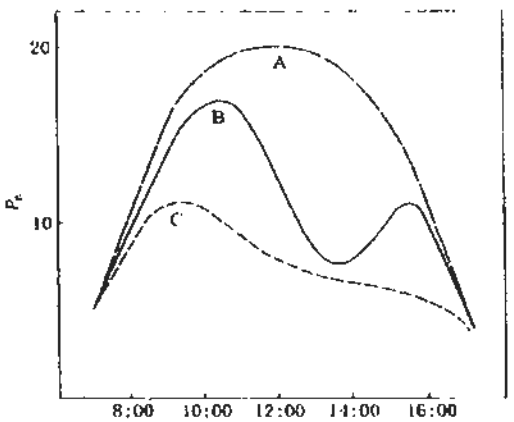
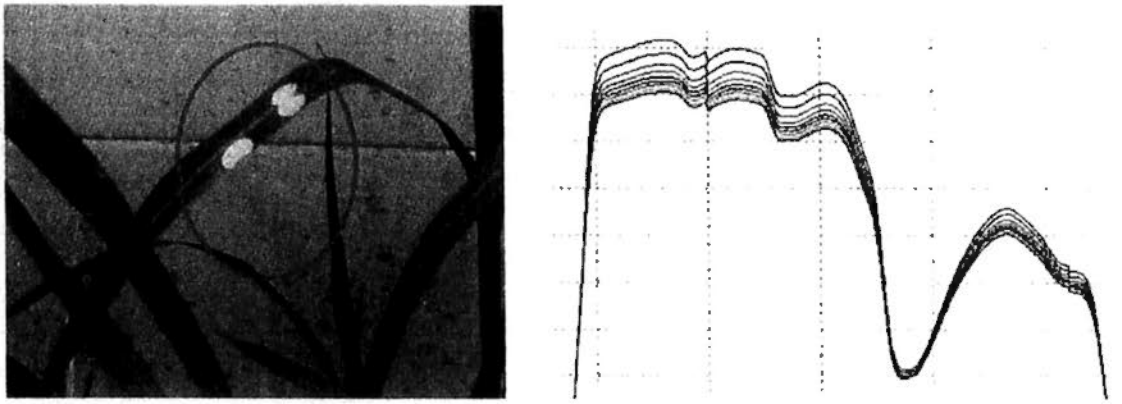


Figure 3-7. Photosynthetic mode of plant in daytime

A. One peak mode; B. two peak mode; C. Special one peak mode(Xu, 2002)

3.2.2.2 Interval of the measurement

For filtering the random noise in the measurement, ASD spectrometer collects several spectral data one time as one measurement. However because of heating effect of the probe light, the leaf spot located in the probe view field would be hurt (Figure 8). Also some bands are sensitive to the water change going with temperature change caused by the probe. Thus suitable sampling times should be considered. In this study, for every measurement, 5 pieces of spectra were collected within 5 seconds.



(a) (b)

Figure 3-8. Heat effect caused by probe light

(a). Leaf damaged by high temperature; (b). Spectral curve under temperature fast change condition.

3.2.2.3 Noise from the reference black background

The black background (reference panel) of the pre-head of data collection would also contribute to the leaf spectral. Table 1 showed the statistic information of spectral reflectance of the black background (420-2400nm).

Table 3-1. Statistics of the spectral reflectance of the background reference panel

	N	Mean	Std. Deviation	Minimum	Maximum
background	1979	.032495	.0005412	.0307	.0333

For assessing the influence made by background, an assumption was made: the final energy from background passing the leaf fits to the regulation described by the typical leaf spectral reflectance. Then the final energy was added to normal spectra as noise, and the spectral indices were calculated. Four spectral indices were chosen to represent kinds of types.

Table 3-2. Spectral indices with difference wavelength

Index Description	source
Chapp=Chlorophyll aR675/R700	(Chappelle, E. W., et al., 1992)
datt=0.0236*[R672/(R550*R708)]$^{0.7954}$	(Datt, B., 1998)
Red edge position: max(first derivate in red edge)	(Pu and Gong, 2000)
Photochemical reflectance index (PRI) PRI=(R531-R570)/(R531+R570)	(Gamon, et al., 1992)
Rdaa=sum(R700-730)	section 4-2
NDWI1240=(R860-R1240)/(R860+R1240)	(Chen, et al., 2005, Zarco-Tejada,

NDWI1640=(NIR858-SWIR1640)/(NIR858+SWIR1640)	et al., 2003)
NDWI2130=(NIR858-SWIR2130)/(NIR858+SWIR2130)	

The two values of the spectral indices were paired. Then the paired sample T-test was done to test if there was any significant difference within them. The alpha (significance) was set as 0.01. The result showed that all selected spectral indices had no significant difference between the original and noised values in strict statistic condition. It revealed that though the black background brought uncertainties to the measurement results, these influences were acceptable.

3.2.2.4 Canopy-level and fine leaf-level spectra

In this study, the fine leaf-level spectra were applied. Because the real differences between the isogenous paddy rice would be very slight. It is easy to be overlaid by external absorption such as vapor absorption. It will bring lots of uncertainties to make the study without reliability. The Fine leaf-level spectrum is obtained under stable artificial illumination condition, and the atmosphere influence could be avoided totally. Figure 9 showed the spectral reflectance of the two levels.

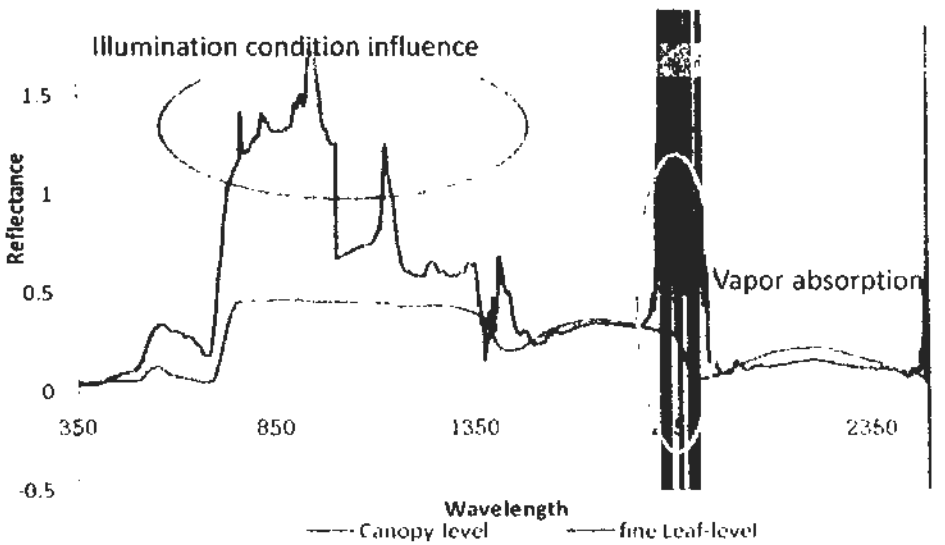


Figure 3-9. Two kinds of spectral reflectance

3.2.2.5 Mean spectra of the group and spectra of the individual sample

When the stable features at the group (class) level are focused, the mean spectra would be applied. However, when calculating the average, features of individual would be neglected. To suppress uncertainty in this procedure, a pre-test would be set to ensure the data are consistent.

3.2.2.6 Uncertainties in model applied

Many models, instant of spectral index, are empirical based on statistic regression. It has lots of

uncertainty itself. Thus, when applying them as indicators of sample growth, it would not be reliable. For overcoming this problem, more spectral indices should be assessed and used together for cross-validation.

3.2.2.7 Problems in measurement

For eliminating uncertainties caused by data collecting, the center of FOV located at the center of the middle front of the second leaf counted from the core of paddy rice to the out. And the spot of the probe view located at the middle of the leave to suppress misunderstanding caused by differences of leave. The principal vein of leaf was vertical to the view line from the fiber. Two measurements were taken at one leaf by micro moving along the principal vein. By this measurement, the consistency of data collected could be ensured.

References

Chappelle, E. W., Kim, M. S., & Mcmurtrey, J. E. (1992). Ratio analysis of reflectance spectra (rars) - an algorithm for the remote estimation of the concentrations of chlorophyll-a, chlorophyll-b, and carotenoids in soybean leaves. *Remote Sensing of Environment*, 39(3), 239-247.

Chen, D., Huang, J., & Jackson, T. (2005). Vegetation water content estimation for corn and soybeans using spectral indices derived from modis near-and short-wave infrared bands. *Remote Sensing of Environment*, 98(2-3), 225-236.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111-121.

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote sensing of environment*, 41,35-44.

Pu, R., & Gong, P. (2000). Hyperspectral remote sensing and its applications. *Higher Education, Beijing*.

Xu, d., (2002). *Photosynthetic efficiency*. Shanghai: Shanghai scientific & Technical Publishers.

Zarco-Tejada, P., Rueda, C., & Ustin, S. (2003). Water content estimation in vegetation with modis reflectance data and model inversion methods. *Remote Sensing of Environment*, 85(1), 109-124.

Chapter 4 Indicators of the foliar chemicals and the sample growth

Spectral reflectance of vegetation is high correlated with vegetation chemical contents, biophysical parameters (Broge, NH, et al., 2001) and even physiological stresses (Blackburn, et al., 2008). Broge and Leblanc (2001) declared that spectral reflectance in the visible (VIS) bands was characterized mainly by chlorophyll pigments and formed special shape such as red edge. Curran (1989) listed 44 absorption features in visible and near-infrared (NIR) wavebands that have been related to particular foliar chemical concentrations and could be detected by fine spectrometers such as field spectrometer. Thus, it is promising and helpful to take spectral indices as the

indicators of the foliar chemicals to respond the changes and differences in the samples from the stress. These indicators are also used to describe the growth status of the sample. In this chapter, spectral indices indicating the contents of the 8 foliar chemicals were selected to assessed, revised or developed for paddy rice, namely chlorophyll a, chlorophyll b, carotenoids, anthocyanin, water, nitrogen and lignin. By these spectral indices could be used to parameterize the spectral features of transgenic sample which was helpful to assess the sample quantitatively.

4.1 Foliar photosynthetic pigments estimation

4.1.1 Introduction

There are two kinds of models: one is based on physical mechanism such as parameters of edge; while the other is based on statistic analysis or half statistic and half physical. These parameters are also called spectral index.

Figure 1 showed the first kind of parameters namely edges. Edges related to photosynthesis include red edge, red absorption, blue edge, green peak and yellow edge (figure 1). They cover the whole photosynthesis bands, and are comprehensive parameters to indicate photosynthesis of leaves. All of them include parameters of the edge position, the reflectance at the edge and area. Edge position, is generally defined as the largest changing rate of the curve at a certain band. The reflectance value at the edge position is the reflectance of the edge while the sum of first derivative of spectral curve is the area of the edge. The red edge is around 680- 760 nm, while the red absorption around 650- 690nm, the yellow edge around 560- 640 nm, the blue edge around 490- 530 nm, and the green peak around 510- 560 nm (Filella, et al., 1994, Gitelson, et al., 1999, Gong, et al., 2002, Horler, et al., 1983). Carter et al. (1992) hypothesized that the increase of reflectance in photosynthetic bands may be a result of decreased chlorophyll content. This hypothesis had been proved later (Bauerle, et al., 2004, Carter, G. A., et al., 2001). These parameters chosen have good correlations to photosynthetic pigments and nitrogen components (K, N and P) no matter on Experiment or theoretical analysis (Nagendra, 2001, Stamps, et al., 1987, Tang, 2004).

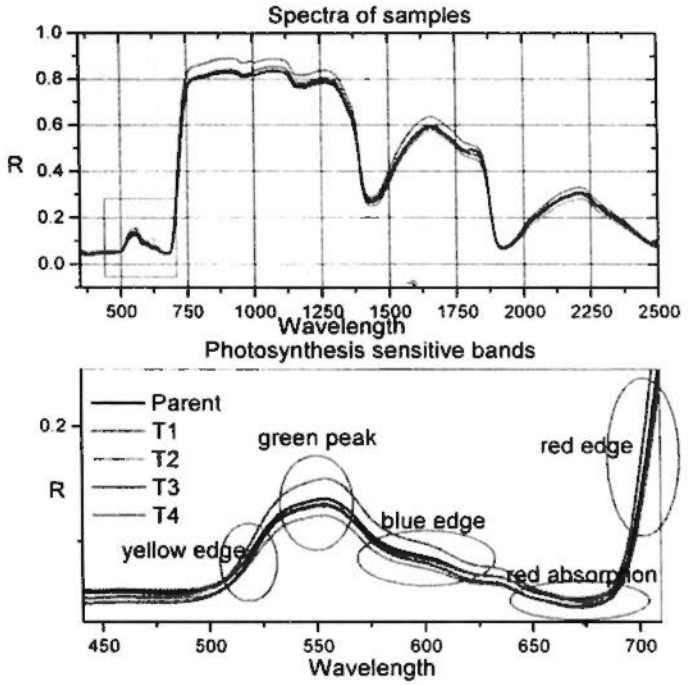


Figure 4-1-1. Spectra of samples and morphological details of photosynthesis sensitive bands

There are lots of studies on the relationship of vegetation spectral reflectance edges and the vegetation relevant biophysical process. All of these study show that edges are powerful indicators to study vegetation when remote sensing approaches are applied to most vegetation studies. In this section, the second kinds of parameters were focused to assess their sensitivity in a universal criterion based on same data set. Totally 9 parameters about chlorophyll a+b, 8 about carotenoids were chosen to be assessed.

4.1.2 Methodology

4.1.2.1 Spectral indices developed as indicators of chlorophyll

Most studies of pigment by remote sensing focused on the relationship between reflectance spectral indices and chlorophyll a & b, because chlorophyll is the dominated pigment during photosynthesis and easily be detected by remote sensing approaches.

Chappelle et al. (1992, Datt, B, 1998) found out R675/R700, short for Chapp, was a good indicator of chlorophyll a concentration in soybean leaves. This parameter was proved that at low chlorophyll concentration, was negative to chlorophyll concentration, while was positive at high concentration (Gitelson, AA, et al., 2003). After assessing several models about prediction of chlorophyll a+b, Datt(1998) developed new index, short for datt. It was combined sensitive photosynthetic bands reported. Gitelson et al. (2003) showed a new model, short for Git03, the basic form of which was $\left[(R_\lambda)^{-1}-(R_{NIR})^{-1}\right]\times R_{NIR}$, and he pointed out that the reciprocal reflectance R^{-1} during [520, 580nm] and [695, 740nm] related closely to the total chlorophyll

concentration in leaves of all species and was a very good linear positive correlation to the total pigment and a high accuracy (Gitelson, AA, et al., 2003). Chlorophyll Absorption in the Reflectance Index (CARI) was first constructed by kim et al. (1994b). Bannari et al. (2007a) reported that it could be used to "*reduce the variability of photosynthectically active radiance due to the presence of diverse non-photosynthetic materials*". Because of so sensitive to the background reflectance properties, the modified CARI (mCARI) was developed (Daughtry, CST, et al., 2000). This new model introduced the ratio between the reflectance in narrow bands around 700nm and 670nm aiming to suppress "*the combined effect of the underlying soil reflectance and the canopy non-photosunthetic materials*" (Bannari, et al., 2007a). Haboudane et al. (2002) presented a new CARI index Transformed Chlorophyll Absorption in the Reflectance Index (TCARI). They reported that TCARI could minimize the soil background reflectance while was sensitive to chlorophyll concentration. Peuelas et al. (1994) pointed out that the Normalized difference between dR and dG (EGFN), a kind of derivative analysis indices, was positive linear correlated with chl and N content. Triangle Vegetation Index (TVI) was firstly developed by Broge and Leblanc (2001). "*The index wascalculated as the area of theriangle defined by the green peak, the chlorophyll absorption minimum, and the NIR shoulder in spectral space. It is based on the fact that both chlorophyll absorption causing decrease of red reflectance and leaf tissue abundance causing increased NIR reflectance would increase the total area of the triangle*" (Broge, NH and Leblanc, E., 2001). Chlorophyll Absorption Ration Index (CARI2) was defined by kim (1994b). This index was based on the idea that the ratio of 550 and 700nm reflectance to be constant at the leaf level regardless of the differences in chlorophyll concentrations.

Table 4-1-1. Spectral indices developed as indicators of chlorophyll a& b (Bannari, et al., 2007a, Gitelson, AA, et al., 2003)

Indices Description	source
Chapp=Chlorophyll aR675/R700	(Chappelle, E. W., et al., 1992)
datt=0.0236*[R672/(R550*R708)]$^{0.7954}$	(Datt, B., 1998)
Git03= $(R_{750-800})/(R_{695-740})$-1	(Gitelson, A. A., et al., 2003)
Chlorophyll Absorption in the Reflectance Index (CARI) CARI =(R700-R670)-0.2*(R700-R550)	(Kim, et al., 1994a)
Modified CARI (mCARI) mCARI =[(R700-R670)-0.2*(R700-R550)]*(R700/R670)	(Daughtry, C. S. T., et al., 2000)
Transformed Chlorophyll Absorption in the Reflectance Index (TCARI) TCARI =3*[(R700-R670)-0.2*(R700-R550)*(R700/R670)]	(Haboudane, D., et al., 2002)
Normalized difference between dR and dG (EDGN)	(Pe uelas, et al., 1994)

EGFN=(Dr-Rg)/(Dr+Rg) Dr: max(first derivate in red edge), Rg: max(first derivate in green peak) (related to Chl a+b and Nitrogen content)	
Triangle Vegetation Index (TVI) TVI=0.5|det(AB,AC)|=0.5(120(Rnir-Rgreen)-200(Rred-Rgreen)) A=(550nm,Rgreen), B=(670nm,Rred), C=(750nm,Rnir)	(Broge, N. H., et al., 2001)
Chlorophyll Absorption Ration Index (CARI2) CARI2=CAR(R700/R670), CAR=|(a*670+R670+b)|/(a^2+1)^0.5 a=(R700-R500)/150, b=R550-(a*550)	(Broge, N. H. and Leblanc, E., 2001)

$R_{750-800}$= Average(Sum(Reflectance(750:800)));

RNIR=Average(Sum(Reflectance(700:750)));

R_{red}=Average(Sum(Reflectance(650:690))), the red absorption feature;

R_{green}=Average(Sum(Reflectance(510:560))) the green peak feature.

So do the other abbreviations in the follow tables.

4.1.2.2 Spectral indices developed as indicators of carotenoids

Structure Insensitive Pigment Index (SIPI) is an empirical index to estimate the ratio of carotenoids to chl a which could be best described using a logarithmic model (Bannari, et al., 2007a, Blackburn, GA, 1998, Penuelas, et al., 1995). However, SIPI lacks sensitivity when the ratio of carotenoids to chla s low (Bannari, et al., 2007a). To remove the chlorophyll contribution from inverse reflectance in the green edge, Carotenoids Reflectance Index (CRI) was developed and this model showed a very high R^2 in regression equation (Gitelson, AA, et al., 2002). CRI550 represents the reciprocal reflectance would be affected by both Car and Chl at 550, while CRI700 is the one only affected by Chl. Based on CRI, Gitelson (2006) developed a new three-band model, Modified CRI (mCRI) to estimate carotenoids and anthocyanin contents in high plant leaves. Datt developed an empirical spectral index, datt_car, which was linear positive related to total carotenoids content (Datt, B, 1998). Photochemical reflectance index (PRI) was applied to estimate to photosynthetic light use efficiency and was related to epoxidation state of the xanthophylls cycle pigments such as carotenoids. Chlorophyll Absorption Ration Index (CARI2) could be used as indicator of carotenoids.

Table 4-1-2. Spectral indices developed as indicators of carotenoids (Ustin, et al., 2009)

Indices Description	Source
Structure Insensitive Pigment Index (SIPI) SIPI =(R800–R445)/(R800–R680)	(Penuelas, et al., 1995, Sims, D. A., et al., 2002)

Carotenoid Reflectance Index (CRI) $CRI550=R^{-1}_{510}-R^{-1}_{550}$ $CRI700=R^{-1}_{510}-R^{-1}_{700}$	(Gitelson, A. A., et al., 2002)
Modified CRI (mCRI) $mCRIgreen=(R^{-1}_{510\cdot 520}-R^{-1}_{560-570})\times R_{NIR}$ $mCRIredge=(R^{-1}_{510\cdot 520}-R^{-1}_{690-710})\times R_{NIR}$	(Gitelson, A. A., et al., 2006)
$datt\_car=0.0049*[R672/(R550\times R708)]^{0.7488}$	(Datt, B., 1998)
Photochemical reflectance index (PRI) PRI=(R531-R570)/(R531+R570)	(Gamon, et al., 1992)
Chlorophyll Absorption Ration Index CARI2=CAR(R700/R670), CAR=|(a*670+R670+b)|/(a^2+1)^0.5 a=(R700-R500)/150, b=R550-(a*550)	(Broge, NH and Leblanc, E., 2001)

4.1.2.3 Spectral indices developed as indicators of anthocyanin

Anthocyanin Reflectance Index (ARI) was developed by Gitelson (2001) and then the author gave a modified form based on the idea of the three-band model (Gitelson, AA, et al., 2006). Red:Green Ratio (RGR) was related to the ratio of anthocyanin to chlorophyll(Sims, DA and Gamon, JA, 2002). And Gamon (1999) also pointed out that this index is highly linear related to the total anthocyanin content and in the index, "red" refers to bands in [600, 699nm] while "green" is in [500, 599].

4.1.2.4 Spectral reflectance data and in situ data

In this section, the data set of Leaf Optical Properties Experiment 93 (Lopex93) was applied. The data set includes spectral and relevant biochemical data of 70 leaf samples representative of more than 50 species. Spectral data were obtained by a 19 double-beam Perkin Elmer Lambda spectrophotometer over the 400-2500nm wavelength with an interval of 1nm The spectral resolution were 1 to 2nm in 400-1000nm and 4 to 5 nm in 1000-2500nm. The properties of spectral reflectance data are much closed to the data obtained by ASD spectrometer. For each sample, 5 measurements would be conducted in different area of the leaf to overcome the leaf to leaf variability.

Meanwhile reflectance data of transgenic paddy rice were used too. These data were obtained by ASD spectrometer with the contact probe and leaf-clip (Details explained in Chapter 3). The samples of paddy rice were in reviving stage. Biological contents (photosynthetic pigments) measurements were measured by spectrophotometer. The samples were first put in the solution

1:1 (alcohol: acetone) for 24 hours. And then spectral absorption at specific bands was obtained for calculating contents of pigments by the equations (Li, et al., 2005, LICHTENTHALER, 1987) as follow:

Chlorophyll a: chl a=11.24*A661.6-2.04*A644.8 (4-1-1)

Chlorophyll b: chl b=20.13*A644.8-4.19*A661.6 (4-1-2)

Total carotenoids: car=(1000*A470-1.9*chl a-63.14*chl b)/214 (4-1-3)

4.1.2.5 Statistical analysis model

In this section, linear statistical model would be chosen to describe the relationship between spectral features and foliar chemicals. The dependent variable is the measured contents of chl a, chl b, chl a+b and car, while the independent variable is spectral index. The basic model is given by the following equation (LI, 2006):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (4\text{-}1\text{-}4)$$

Where $\varepsilon_i \sim N(0,\sigma^2)$, i=1,2,...,n

The errors ε are assumed normally distributed with mean zero and variance σ^2. The linear fit model is as follow:

$$\hat{Y} = \beta_0 + \beta_1 X \quad (4\text{-}1\text{-}5)$$

Thus, the linear regression equation could be formed as:

$$Q(\beta_0,\beta_1) = \sum_{i=1}^{n}(y_i - E(y_i))^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \quad (4\text{-}1\text{-}6)$$

According to The least square estimation, the estimated values of β_0, β_1 are ones which could minimize the results of the equation 7, that is:

$$Q(\hat{\beta}_0,\hat{\beta}_1) = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \min_{\beta_0,\beta_1}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \quad (4\text{-}1\text{-}7)$$

So the coefficients β_0, β_1 would be obtained.

$$\beta_0 = Y - k\bar{X} \quad (4\text{-}1\text{-}8) \qquad \beta_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - Y\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \quad (4\text{-}1\text{-}9)$$

Where $\bar{Y}$ and $\bar{X}$ are the mean values of variable of Y and X.

Two parameters were calculated to evaluate the regression model, one is RMSE which "*indicates*

the magnitude of the average error produced by a model", and the other is "*the coefficient of determination (R2) to evaluate the strength of linear relationship between the observed and predicted valises*" (Bannari, et al., 2007a).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{Y} - \bar{Y}\right)^2}{n}} \quad (4\text{-}1\text{-}10) \qquad R^2 = \frac{\sum_{i=1}^{n}\left(Y_i' - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} \quad (4\text{-}1\text{-}11)$$

Also curve estimation approaches would be applied as reference for specific spectral index. The details of these approaches show in the table 3.

Table 4-1-3. Regression model applied

Model	Regression equation
linear	$y_i = \beta_0 + \beta_1 x_i$
Quadratic	$y = \beta_0 + \beta_1 x + \beta_2 x2$
Logarithmic	$y = \beta_0 + \beta_1 ln(x)$
Exponential	$y = \beta_0 e^{\beta_1 x}$
Power	$y = \beta_0 (x^{\beta_1})$

4.1.3 Results and discussion

4.1.3.1 Assessment of correlation of spectral indices with photosynthetic pigments

In this study, six statistic parameters were selected to indicate the relationships. Column "*Pearson correlation*" represents the correlation coefficient of Pearson correlation statistics to describe the strength of correlation. Pearson coefficients range from -1 to 1. Large absolute value indicates stronger relationships of the two variables. The first line of "*Fit equation*" column gives the linear regression equation. "R^2"represents the proportion of variance in the dependent variable explained by the regression model. Small value means the model does not fit the data well. "*Std. Error of the Estimate*" shows the predicted error caused by the model. If R^2 were two small, less than 0.5 means the linear regression is not fit the data. Then the second line of "*Fit equation*" would show other regression model results and the same as "R^2"and "*Std. Error of the Estimate*" (Std. Error) column. The results of estimated residual were also assessed and the standard deviation of it was given. Also Durbin-Watson test was used to test whether the residuals are

auto-correlated. It "*ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation, a value toward 0 indicates positive autocorrelation, and a value toward 4 indicates negative autocorrelation.*"[17] If the residuals are auto-correlated, that implies that regress equation do not explain the regulation in the variable enough or the regression model is not fit to the data. These three parameters were used to assess the fit equation.

The ambiguity of correlation between spectral indices mentioned previously and contents of photosynthetic pigments were found. Thus, in this study, correlations between contents of chlorophyll a (chl a), chlorophyll b (chl b), chlorophyll a+b (chl a+b) and carotenoids and spectral indices were assessed respectively. Table 4 showed the regression results of the content of chlorophyll a and spectral indices. The content of chlorophyll a was dependent variable while spectral index were an independent one. All spectral indices showed correlations with chlorophyll a, with the least Pearson correlation coefficient was larger than 0.6. Spectral index datt showed the largest relationship with chl a up to 0.789. While the R^2 of linear fit equation was the largest one, the Std. Error of the Estimate and standard deviation of residual were the least one. These parameters proved that the datt were the best index in the selected items to indicate the contents of chlorophyll a in vegetation leaf. Chapp also showed good performance. From the results, linear regression was found that it could not explain the relationship between mCARI and chl a with R^2 in 0.454. However, when power model was applied the R^2 was up to the largest, 0.626. So mCARI would also show good performance when power model applied. The worst performances were shown by Git03 and EFGN though these two indices also had large Pearson correlation coefficients. The fact that Durbin-Watson test result (1.846, 1.626)of the fit equation of the two were very near to 2 means lower performance would be not because of the model. A possible reason was the smaller data set. When more data joined the regression, the results of correlation analysis may be acceptable.

Table 5 showed the results of correlation between spectral indices and the content of chlorophyll b. The content of chlorophyll b was a dependent variable while spectral index was an independent one. The results showed that spectral indices had a relative good correlation with chlorophyll b. Except for Git03, mCARI and CARI2, the Pearson correlation coefficient of the others were more than 0.6, and the lowest value was also up to 0.513(Git03). However, except datt (0.577), the R^2 of linear regression were not good, lower than 0.5. Spectral index datt showed good performance again in the mode estimated chlorophyll b with the least Std. Error of the Estimate and Std. deviation of residual. Meanwhile, the result of Durbin-Watson test of datt was also the least that meant the regression model of datt and chl b could be improved to be much better. When other models were applied, the other spectral indices, chap, CARI, mCARI,

[17] SPSS 13.0 for Windows, help- topics: "Durbin-Watson Significance Tables"

TCARI and TVI performance of estimation for chl b were improved. Git03 also showed the worst performance. EFGN and CARI were also not suitable to built model for Chl b prediction However, the Pearson correlation coefficient of CARI2 was not the least and with large Durbin-Watson test near to 2; it implied the bad performance may be caused by lack of enough data.

Table 6 showed the results of correlation between spectral indices and the content of chlorophyll a+b. The content of chlorophyll a+b was dependent variable while spectral index was independent one. To be noted, the chlorophyll a+b could not be calculated directly because of data missing, thus here the content of chlorophyll a+b was replaced by the sum of the content of chl a+the content of chl b approximately (Zhang, 1985). Like correlation with chlorophyll a and b, all spectral indices had a good Pearson correlation coefficient. The mode estimated chlorophyll a+b built with datt were the best one with the least Std. Error of the Estimate and Std. deviation of residual. The model built with Chapp was also good. When non-linear regression models were applied, the performance of mCARI, TVI and CARI2 were improved. Git03 also had not good performance and then was EGFN. Especially for Git03, all regression models for chl a, chl b and chl a+b, had a large Durbin-Watson test results near to 2. Besides the influence contributed by small data set, it was also an important reason. From these results, it could be found that the complicated spectral indices maybe could not enhance the performance of the index to predict chlorophyll. Complicated index took account of more factors however it limited the applicable the index.

Table 7 showed the results of correlation between spectral indices and the content of carotenoids. The content of carotenoids was a dependent variable while spectral index was an independent one. CARI500, CARI700, mCARIgreen, mCARIred and PRI have good Pearson correlation coefficient all more than 0.6, while SIPI, Datt_car and CARI2 were very lower. The Durbin-Watson tests results of these three were large, near to 2. It implied there was no or very slight relationship with carotenoids and the indices with data set used. Table 7 showed the results of correlation between spectral indices and the ratio of the content of carotenoids to the content of chlorophyll a. The results showed there was almost no relationship between the two variables.

Table 4-1-4. Regression results between spectral indices and content of chl a

Y=content(chl a) mg/g(per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Chapp	0.777	Y=-0.622+8.860x	0.604	0.6319	0.6213	1.267
datt	0.789	Y=0.578+53.981x	0.623	0.6163	0.6059	1.525
Git03	0.620	Y=0.035+3.996	**0.384**	0.7876	0.7743	1.846
		$y = 3.925x^{0.999}$	**0.423**	/	/	/
CARI	-0.720	Y=5.126-30.912x	0.518	0.6970	0.6853	1.710
mCARI	-0.674	Y=4.238-6.414x	**0.454**	0.7419	0.7294	1.663
		$y = 1.187x^{-0.47}$	0.626	/	/	/
TCARI	-0.713	Y=5.070-9.732	0.509	0.7036	0.6917	1.723
EGFN	0.602	Y=-1.460+7.791	**0.363**	0.8013	0.7879	1.626
		$y = 6.960x^{1.583}$	**0.396**	/	/	/
TVI	-0.691	Y=8.932-0.215x	0.477	0.7261	0.7139	1.828
		$y = 0.011x^2 - 0.847x + 17.64$	0.512	/	/	/
CARI2	-0.673	Y=4.865-4.458x	**0.453**	0.7424	0.7230	1.755
		$y = 1.422x^{-0.72}$	0.588	/	/	/

The number in bold means lower correlation coefficient or R^2. Shading means the best results of non-linear regression model. "##" means no or very slight relationship between the

two variables. It is the same as other tables.

Table 4-1-5. Regression results between spectral indices and content of chl b

Y=content(chlb) mg/g(per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Chapp	0.694	Y=-0.225+2.920x	**0.482**	0.2664	0.2619	1.286
		$y = 9.076x^2 - 4.798x + 1.348$	0.535	/	/	/
datt	0.760	Y=0.103+19.154x	0.577	0.2408	0.2367	1.270
Git03	0.513	Y=0.069+1.219x	**0.263##**	0.3178	0.3124	1.717
CARI	-0.655	Y=1.680-10.370x	**0.429**	0.2798	0.2751	1.573
		$y = 144.4x^2 - 32.72x + 2.432$	0.533	/	/	/
mCARI	-0.581	Y=1.365-2.041x	**0.338**	0.3012	0.2961	1.543
		$y = -0.48\ln(x) + 0.072$	0.511	/	/	/
TCARI	-0.661	Y=1.673-3.326x	**0.437**	0.2777	0.2731	1.540
		$y = -0.73\ln(x) - 0.224$	0.516	/	/	/
EGFN	0.610	Y=-0.708+2.910x	**0.372**	0.2933	0.2884	1.334
		$y = 0.175e^{2.861x}$	**0.381**	/	/	/
TVI	-0.684	Y=3,127-0.079x	**0.468**	0.2700	0.2655	1.730

		$y = 0.006x^2 - 0.443x + 8.149$	0.556	/	/	/
CARI2	-0.596	Y=1.578-1.456x	**0.356**	0.2971	0.2922	1.626
		$y = 3.488x^2 - 5.07x + 2.357$	**0.490**	/	/	/

Table 4-1-6. Regression results between spectral indices and content of chl a +b

Y=content(chl a+b) (per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Chapp	0.767	Y=-0.847+11.780x	0.589	0.8668	0.8522	1.204
datt	0.794	Y=0.681+73.135x	0.631	0.8210	0.8072	1.358
Git03	0.601	Y=0.104+5.215x	**0.361**	1.081	1.0620	1.789
		$y = 5.155x^{0.972}$	**0.396**	/	/	/
CARI	-0.714	Y=6.806-41.282x	0.509	0.9466	0.9306	1.635
mCARI	-0.660	Y=5.603-8.455x	**0.435**	1.0159	0.9988	1.595
		$y = 1.572x^{-0.47}$	0.617	/	/	/
TCARI	-0.711	Y=6.742-13.058	0.505	0.9506	0.9346	1.635
EGFN	0.614	Y=-2.168+10.700x	**0.377**	1.0633	1.0484	1.515
		$y = 9.315x^{1.607}$	**0.404**	/	/	/

TVI	-0.700	Y=12.059-0.293x	**0.490**	0.9648	0.9486	1.777
		$y = 0.017x^2 - 1.290x + 25.78$	0.539	/	/	/
CARI2	-0.663	Y=6.443-5.914x	**0.440**	1.0115	0.9945	1.690
		$y = 1.885x^{-0.72}$	0.577	/	/	/

Table 4-1-7. Regression results between spectral indices and content of carotenoids

Y=content(carotenoids) mg/g(per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
SIPI	**0.395##**	Y=-12.973+13.799x	**0.156**	0.2141	0.2105	1.726
CRI550	0.809	Y=-0.064+0.107x	0.654	0.1370	0.1347	1.514
CRI700	0.773	Y=0.266_0.055x	0.597	0.1478	0.1454	1.806
mCRIgreen	0.684	Y=-0.056+0.549x	**0.467**	0.1701	0.1672	1.492
		$y = 0.226e^{0.781x}$	**0.483**	/	/	/
mCRIred	0.648	Y=0.215+0.317x	**0.420**	0.1775	0.1745	1.576
		$y = 0.334e^{0.448x}$	**0.429**	/	/	/
Datt_car	**0.300##**	Y=0.465+31.736x	**0.09##**	0.2223	0.2186	1.665
PRI	-0.637	Y=0.653-1.1213x	**0.406**	0.1796	0.1766	1.837

		$y = 0.619e^{-1.73x}$	**0.424**	/	/	/
CRAI2	**-0.492##**	Y=1.148-1.067x	**0.242##**	0.2028	0.1994	1.821

Table 4-1-8. Regression results between spectral indices and ratio of carotenoids to Chlorophyll a

Y=content(carotenoids/ chl a) mg/g(per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
SIPI	**0.246##**	##	**0.061##**	##	##	##
CRI550	**0.216##**	##	**0.047##**	##	##	##
CRI700	**0.116##**	##	**0.013##**	##	##	##
mCRIgreen	**0.329##**	##	**0.108##**	##	##	##
mCRIred	**0.336##**	##	**0.113##**	##	##	##
Datt_car	**-0.217##**	##	**0.047##**	##	##	##
PRI	**-0.012##**	##	**0.000##**	##	##	##
CRAI2	**0.156##**	##	**0.024##**	##	##	##

Table 4-1-9. Pearson correlation analysis between spectral indices

Item	chapp	datt	mcari	cari	tcari	tvi	cari2	git03	egfn

chap p	1								
datt	.932(**)	1							
mcari	-.859(**)	-.832(**)	1						
cari	-.851(**)	-.902(**)	.964(**)	1					
tcari	-.851(**)	-.917(**)	.952(**)	.996(**)	1				
tvi	-.638(**)	-.752(**)	.581(**)	.685(**)	.718(**)	1			
cari2	-.819(**)	-.845(**)	.985(**)	.989(**)	.977(**)	.610(**)	1		
git03	.719(**)	.765(**)	-.794(**)	-.829(**)	-.815(**)	-.330	-.823(**)	1	
egfn	.800(**)	.903(**)	-.810(**)	-.888(**)	-.897(**)	-.528(**)	-.850(**)	.792(**)	1

"**" Correlation is significant at the 0.01 level (2-tailed).

Table 4-1-10. Pearson correlation analysis between spectral indices

Item	sipi	cri550	cri700	mcrigreen	mcrired	datt_car	pri	cari2
sipi	1							
cri550	.378	1						
cri700	.457	.916(**)	1					
mcrigreen	.213	.903(**)	.735(**)	1				
mcrired	.284	.826(**)	.822(**)	.875(**)	1			

datt_car	.470	.256	.591(*)	-.031	.207	1		
pri	-.475	-.722(**)	-.938(**)	-.502(*)	-.711(**)	-.806(**)	1	
cari2	-.411	-.476	-.726(**)	-.192	-.317	-.911(**)	.849(**)	1

** Correlation is significant at the 0.05 level (2-tailed).

*** Correlation is significant at the 0.01 level (2-tailed).

4.1.3.2 Correlation between spectral indices

The spectral indices were applied to describe the relationship between vegetation spectral reflectance properties and its photosynthetic pigment contents. Thus, when they describe the same pigment content, they should be correlated to each other. The intensity of relationship between spectral indices could be used as an indicator of consistency and reliability within the indices. And it also showed an indirect transfer of relationship between pigments content and other spectral indices.

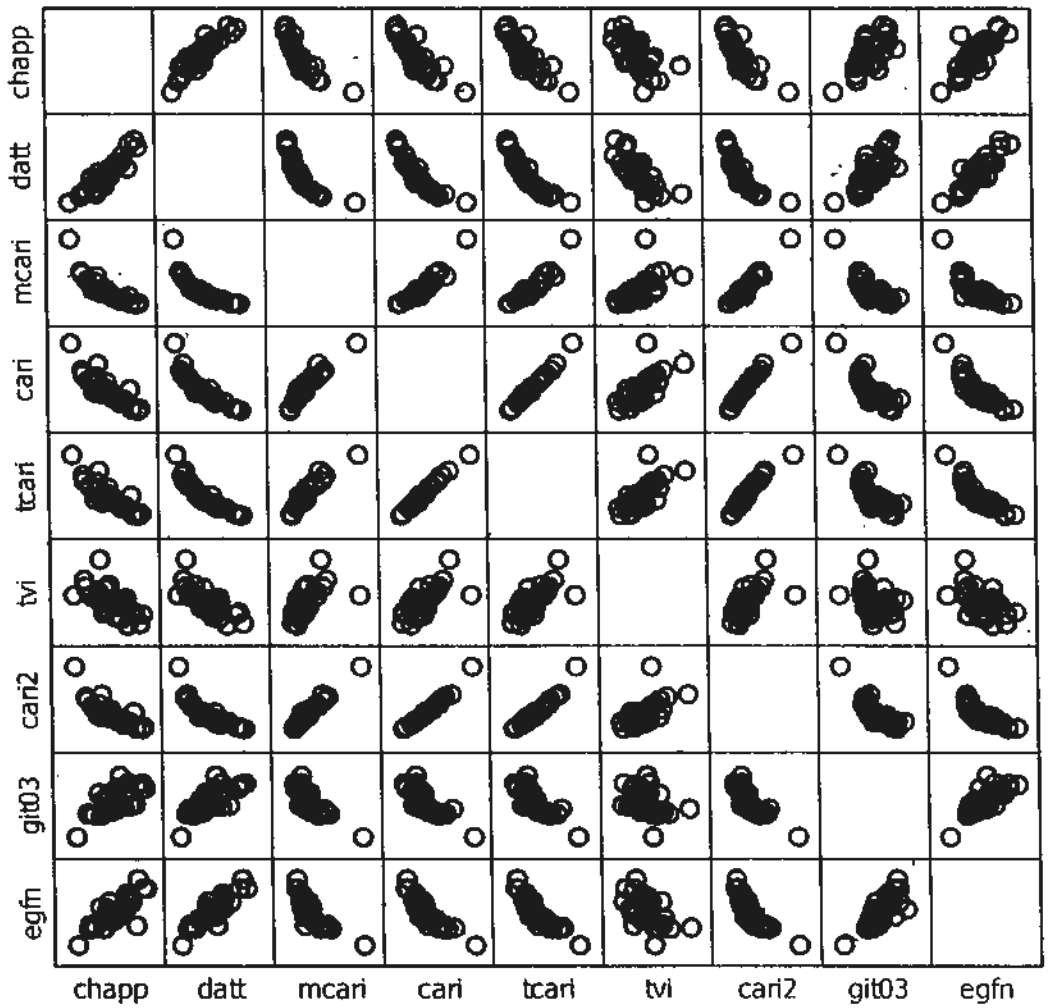
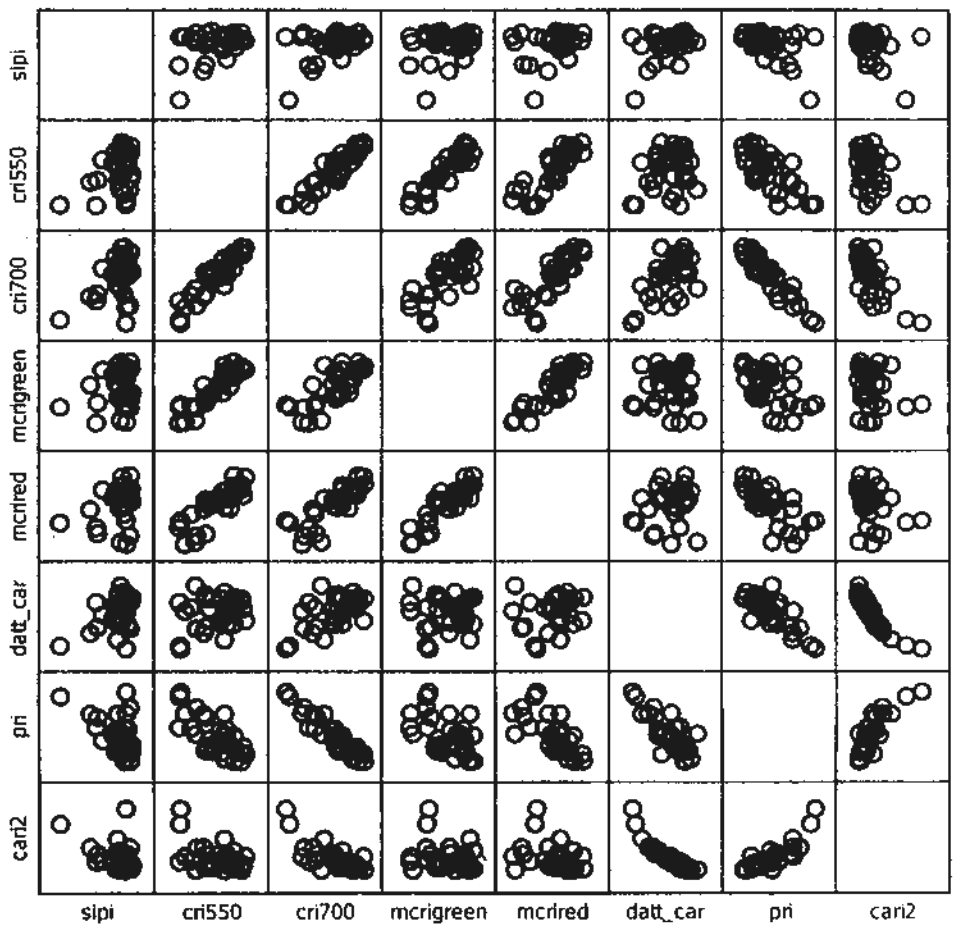


Figure 4-1-2. Correlation between spectral indices as chlorophyll indicators

Table 9 and 10 listed Pearson correlation coefficients between selected spectral indices. As indicators of the content of chlorophyll, these indices were highly related each other except git03 and TVI. There were 9 the correlation coefficients which was more than 0.9, namely coefficients of datt and chapp, datt and CARI, datt and EGFN, mCARI and CARI, mCARI and TCARI, mCARI and CARI2, CARI and TCARI, CARI and CARI2, and CARI2 and TCARI. CARI, mCARI and TCARI were the model built based on the same idea, and TCAI and mCARI were revised version of CARI. Thus, they were highly correlated, and in the test data set, these three model could be replaced each

other. However, CARI2 was a totally diff model, compared with cari, mcari and tacri, it was most strongly related with the follow ones, more thán 0.97. It revealed as the indicators of consistency and reliability of chlorophyll content, they were highly consistent and reliable. Datt still had good relationship with other indices. It coincided with correlation analysis about spectral indices with the content of chlorophyll in table 4-6.

Git03 and TVI showed no relationship, the correlation coefficient was only -0.330. Also, figure2 showed TVI with EFGN, chap and datt were also clustered as much as it with others. TVI was a complicated index (table 1). More factors were considered when built the model for detailed description. However, it made the model over-sensitive and narrows its application scope. Figure 2 also showed what kind of relationships the indices had, positive, negative, linear or exponential regression, etc.



(b)

Figure 4-1-3. Correlation between spectral indices and carotenoids indicators

It was much more complex about the relationship between spectral indices as carotenoids indicators. In figure3, most indices showed no relationship. Because PRI was an index widely proved, it could be taken as reference. PRI was strongly related with CAR500, CARI700, mCRIred, datt_car and CARI2. Though, the previous analysis (table 8 and 9) implied that same indices had no relationships with the content of carotenoids and the ratio of carotenoids/ chl a, there were

some kind of relationships, more complicated, between them. When we do not have more accurate data or data missed, they could be still as indicators for carotenoids in some sensing. The poorest spectral index as an indicator for carotenoids was SIPI (table 7, 8 and 10). Figure 3 showed it almost had no relationships with others. By censoring the data of SIPI, the factor that the values of SIPI were very clustered. It implied SIPI would be not very sensitive to the target content.

4.1.3.3 Assessment of spectral index sensitivity

Index Sensitivity is also an important feature when the index was applied to predict or distinguish different level of target content. There are kinds of the responses encountered, one is to the target changes, and the other is to noise. Here the former is the one of interested. A good sensitive index could amplify the differences or changes of target content, namely in a scatter-plot, the normalized values of the index to their counterparts, the plot points should be discrete enough in a certain correlation-ship. This characteristic shows that the index could respond certain changes sensitively. When assessing the sensitivity of a spectral index, the mean, the standard deviation and the coefficient of variation (CV) of the index could be applied to describe the dispersion of the sample data. Coefficient of variation[18] could be defined as the ratio of the mean to the standard deviation.

$$cv = \frac{\sigma}{\mu} \quad (4\text{-}1\text{-}12)$$

Where σ , μ are mean and standard deviation of the data, respectively.

For comparison between parameters, firstly the data should be normalized (standardize) to project them to a universal space without unit.

$$Normalization = \frac{(index - \min)}{(\max - \min)} \quad (4\text{-}1\text{-}13)$$

Table 4-1-11. Sensitivity of spectral indices (based on normalized data)

Item	mean	Std. deviation	CV
Chapp	0.5586	0.2398	**0.4292**
datt	0.467	0.2436	0.5216
Git03	0.1998	**0.1967**	[illegible]
CARI	0.2889	0.2109	0.73
mCARI	0.3036	0.2221	[illegible]

[18] http://en.wikipedia.org/wiki/Coefficient_of_variation

TCARI	0.3469	**0.2055**	0.5925
EGFN	0.2306	**0.1985**	0.861
TVI	0.5812	0.2069	**0.3559**
CARI2	0.5092	0.2152	0.4226
SIPI	0.8059	0.2089	**0.2592**
CRI550	0.5559	0.3001	0.5399
CRI700	0.6134	0.2718	0.4431
mCRIgreen	0.5819	0.3006	0.5166
CRIred	0.58	0.2637	0.4546
datt_car	0.5837	0.2537	0.4347
PRI	0.3698	0.2671	0.7223

In the table 11, red meant largest three items while black with shading meant the lowest three. CARI550, mCRIgreen and CRIred had largest standard deviations which implied they were more dispersing than others. Git03, TCARI and EGFN had smallest standard deviation. Low dispersion meant lower sensitive. CV is other parameter which could be applied to describe discreteness of data. The larger values meant lower discreteness of data. From the table, git03 and EGFN had very large cv. Results of Standard deviation and CV was consistent with these two indices. However, CV also would become not reliable. Taking SIPI as an example, it had lowest CV which implied it was dispersing. However, it was much more clustered than others. Larger mean value made CV of SIPI was lowest. Thus, CV was just for reference.

4.1.4 Conclusion

In this section, 9 parameters about chlorophyll a+b and 8 parameters about carotenoids had been assessed. The 9 spectral indices are good related to chlorophyll a and a+b, also most of them can be as indicators as chlorophyll b. Chapp and datt are the best indicators. For caretonoids, the spectral indices are not as good as indicators of chlorophyll. However, most of them had a good relationship to PRI which was widely accepted. When we do not have more accurate data or data missed, they could be still as indicators for carotenoids.

An efficient index should be dispersing enough to ensure sensitivity to aimed content or changes. Through analysis, complicated model was found to be not the best one, a good model as an indicator should be simple while have enough resistance to noise. However, most spectral indices are based on several bands. Thus, in section 4.2 and 4.3, new models should be developed to overcome these flaws.

References

Bannari, A., Khurshid, K. S., Staenz, K., & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(10), 3063-3074.

Bauerle, W. L., Weston, D. J., Bowden, J. D., Dudley, J. B., & Toler, J. E. (2004). Leaf absorptance of photosynthetically active radiation in relation to chlorophyll meter estimates among woody plant species. *Scientia Horticulturae*, 101(1-2), 169-178.

Blackburn, G. (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., & Ferwerda, J. G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of environment*, 112(4), 1614-1632.

Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2), 156-172.

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684

Carter, G. A., Mitchell, R. J., Chappelka, A. H., & Brewer, C. H. (1992). Response of leaf spectral reflectance in loblolly-pine to increased atmospheric ozone and precipitation acidity. *Journal of Experimental Botany*, 43(249), 577-584.

Chappelle, E. W., Kim, M. S., & McMurtrey Iii, J. E. (1992). Ratio analysis of reflectance spectra (rars): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sensing of environment*, 39(3), 239-247.

Curran, P. (1989). Remote sensing of foliar chemistry. *Remote Sensing of environment*, 30(3), 271-278

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111-121.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229-239.

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459-1470.

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote sensing of environment*, 41, 35-44.

Gamon, J., & Surfus, J. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, 143(1), 105-117.

Gitelson, A., Gritz , Y., & Merzlyak, M. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271-282.

Gitelson, A., Merzlyak, M., & Chivkunova, O. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74(1), 38-45.

Gitelson, A., Zur, Y., Chivkunova, O., & Merzlyak, M. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3), 272-281.

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Gitelson, A. A., Schalles, J. F., Rundquist, D. C., Schiebe, F. R., & Yacobi, Y. Z. (1999). Comparative reflectance properties of algal cultures with manipulated densities. *Journal of Applied Phycology*, 11(4), 345-354.

Gong, P., Pu, R., & Heald, R. (2002). Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia. *International Journal of Remote Sensing*, 23(9), 1827-1850.

Haboudane, D., Miller, J., Tremblay, N., Zarco-Tejada, P., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of environment*, 81(2), 416-426.

Horler, D., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4, 273-288.

Kim, M., Daughtry, C., Chappelle, E., McMurtrey III, J., & Walthall, C. (1994a). The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation. Proceedings of the ISPRS Sixth International Colloquium on Physical Measurements and Signatures in Remote Sensing; 1994a 17- 21 January 1994; Val d'Isere, France: European Space Agency; 1994a. p. 299-306.

Li, D., Guo, Y., Yun, H., Zhang, M., Gong, X., & Mu, F. (2005). Determined methods of chlorophyll from maize *Chinese Agricultural Science Bulletin*, 21(6), 153-155.

Li, Z., (2006). *Spectral analysis technology and its application*. Beijing: Science Press.

LICHTENTHALER, H. K. (1987). Chlorophyll and carotenoids: Pigments of photosynthetic biomembranes. *Methods in enzymology*, 148(350-382.

Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22(12), 2377-2400.

Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2), 221-230.

Peuelas, J., Gamon, J., Fredeen, A., Merino, J., & Field, C. (1994). Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sensing of environment*, 48(2), 135-146.

Sims, D., & Gamon, J. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of environment*, 81(2), 337-354.

Stamps, J., Buechner, M., & Krishnan, V. (1987). The effects of edge permeability and habitat geometry on emigration from patches of habitat. *The American Naturalist*, 129(4), 533-552.

Tang, Y. (2004). Study on the hyperspectral characteristics and simulating and estimating models about biophysical and biochemical parameters of rice. 35~45.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, 113(S67-S77.

Zhang, q. (1985). Approaches for the content of chlorophyll measurement(测定叶绿素的几种方法). *Chinese Bulletin of Botany*, 3(5), 60-64.

4.2 New spectral indices as indicators of the Photosynthetic pigments of paddy rice

Photosynthesis is one of most important biochemical processes for plants. The chlorophyll pigments which *"are integrally related to the physiological function of leaves"* (Sims, DA and Gamon, JA, 2002) are *"essential in the process of photosynthesis"*(Blackburn and Ferwerda, 2008) and they are *"of tremendous significance in the biosphere"* and *"necessary for photosynthesis"* (Blackburn, GA, 2007). Chlorophylls are the most important pigments from a physiological perspective which is related to the amount of solar radiation and total leaf nitrogen and provides valuable information about physiological status of plants (Gitelson, AA, et al., 2003). By censoring of chlorophyll, insights into plant-environment interactions could be obtained (Richardson, et al., 2002). Carotenoids are also important pigments which *"composed of carotenes and xanthophylls which can absorb incident radiation and contribute energy to photosynthesis"*. Thus some scholars acclaimed *"chemical concentrations of foliage are important indicators of ecosystem processes"*(Huang, et al., 2004) and important indicators of vegetation physiological stresses, leaf development, senescence and plant nutrient status etc(Blackburn and Ferwerda, 2008).

Remote sensing estimation of photosynthetic pigments is promising because these pigments each have different spectral absorption features (Bannari, et al., 2007a). Based on this idea, many spectral indices or model had been developed such as CARI, TVI, PRI, and CRI etc. However, as mentioned in the section 4-1, these spectral indicators are not very ideal, for instant some are individual-band-based (in narrow spectral bands) thus are sensitive to noise. *"To provide sufficient sensitivity to a small variation of chlorophyll, a broad spectral range is required"*(Gitelson, AA, et al., 2003). And most spectral indices were developed based on remote sensing images of which data are seriously polluted by external factors (e.g. water vapor and aerosol absorption) thus they are not sensitive or suitable applied directly to the data acquired by ground-based fine-resolution and Signal-to-noise of the equipment (e.g. ASD Spectrometer). Meanwhile, in this study, spectral indices would be applied as sensitive indicators to monitor sample growth and help to sample and outlier screening, thus the spectral index should be reliable and specific. Also, in the study, all spectral data were acquired by ground-based fine spectrometer therefore it was much more necessary to develop new spectral index as pigments indicators.

4.2.1 Pigment absorption

Figure1 showed photosynthetic pigments absorption position. Chlorophyll pigments have a strong absorption during 400- 700nm. At green band, there is a strong reflected peak, thus plant looks green. In the 550- 680nm (red bands), there is very strong spectral absorption where absorption rate is almost up to 95%. Because of plant leaf special internal structure, in infrared red band (after 680nm approximately), reflected rate has increased rapidly. It is the most

significant and unique characteristic of vegetation. Chlorophyll a mainly has two reflected peak one is located at 400-500nm, the other is in 650-700nm. The former has stronger reflected rate than the latter. However, the former reflected peak of chlorophyll overlaps with the one of chlorophyll b and carotenoids. Near the second position, thought absorption of chlorophyll a still overlaps with it of chlorophyll b, the chlorophyll a is the primary one and more strong than chlorophyll b's. For chlorophyll b and carotenoids, the absorption features in 400-580 could be chosen for prediction of them. However, in this range, different absorptions of pigments overlap with each other complicatedly, thus the results of estimation of chlorophyll b and carotenoids would not be as accurate as the one of chlorophyll a. Chappelle et al.(1992) applied 675nm and 700nm to estimate chlorophyll. Gitelson et al. (1996) found the ratio of reflectance at 750 nm to 700 nm (R750/R700) was directly proportional to chlorophyll concentration. Datt (Datt, B., 1998) concluded that the550, 672 and 708nm were the best bands of indicator of chlorophyll a, chlorophyll a+ b and total carotenoids. Based on these reported bands, many spectral indices have been developed (Blackburn, GA, 1999, Daughtry, C. S. T., et al., 2000, Gamon, et al., 1992, Gitelson, AA, et al., 2003, Gitelson, AA, et al., 2006). Therefore, generally empirical models for estimation of chlorophyll choose band around 500nm and 700nm (Sims, DA and Gamon, JA, 2002).

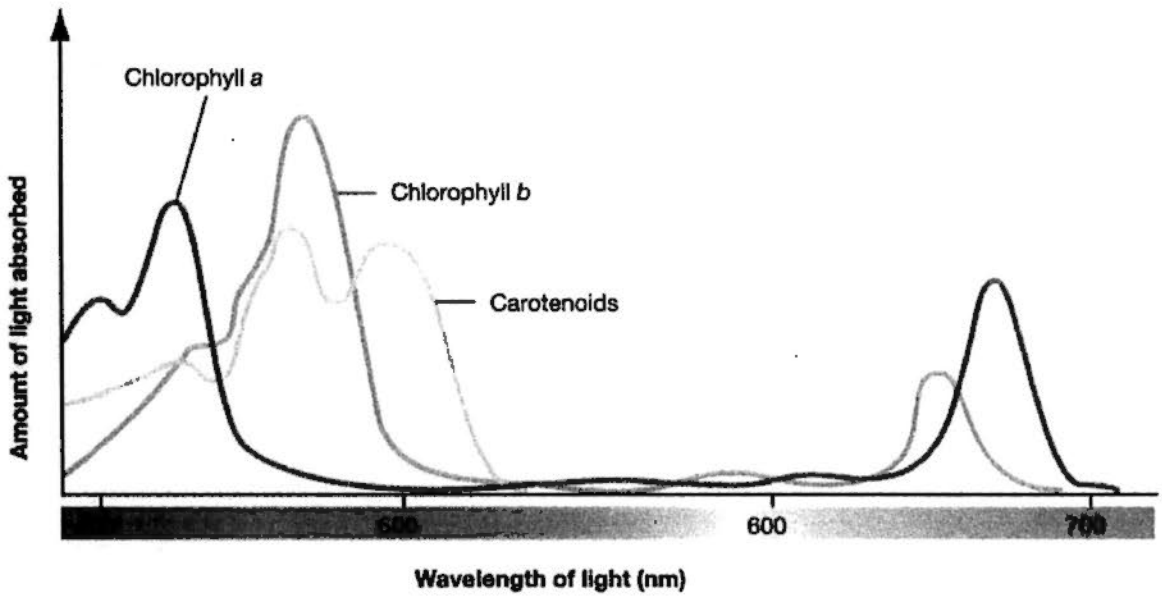


Figure 4-2-1. Typical photosynthetically active radiation (PAR) action spectrum, shown beside absorption spectra for chlorophyll a, chlorophyll b, and carotenoids[19]

In this section, the data set of Leaf Optical Properties Experiment 93 (Lopex93) and the spectral reflectance data of paddy rice acquired were applied. Reflectance data were obtained by ASD

[19] http://www.uic.edu/classes/bios/bios100/lecturesf04am/lect10.htm

spectrometer with the contact probe and leaf-clip. The samples of paddy rice were in reviving stage.

4.2.2 New spectral indices as indicators of chlorophyll pigments

4.2.2.1 New spectral index as indicators of chlorophyll pigments

The previous spectral indices were based on individual band thus were sensitive to noise. Gitelson et al. (2003) ,applying the average of reflectance from 520 to 550 nm and 695 to 705 nm, developed a reciprocal reflectance model which was related closely to the total chlorophyll content in leaf. Another index reported in the same reference used the average of reflectance from 695 to 740 nm and from 750 to 800nm. The general form of this model was given as $\left[(R_\lambda)^{-1}-(R_{NIR})^{-1}\right]$. It is a kind of idea to reduce noise; however, essentially it is still based on narrow spectral bands, when noise was suppressed, useful information would be suppressed. Meanwhile sensitivity of index would be influenced too. Thus, in this section, a new kind of indices based on area around by the axis and spectral reflectance curve will be developed.

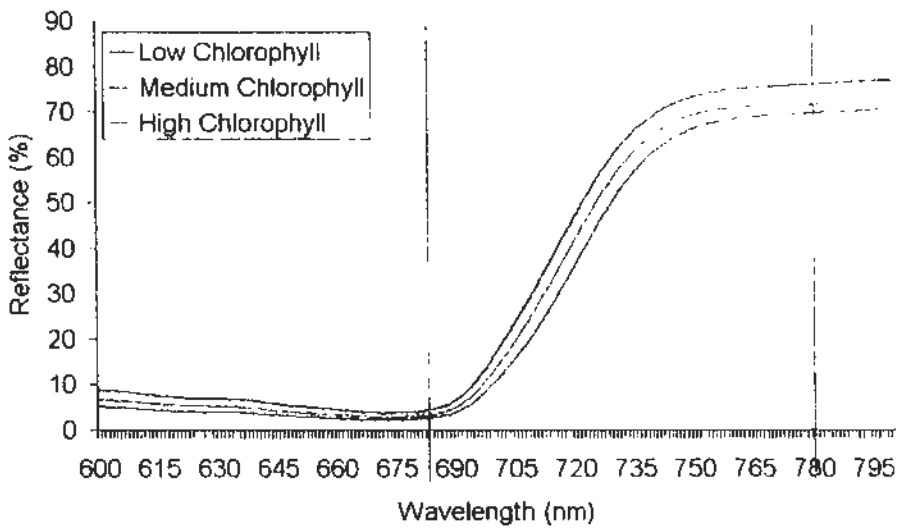


Figure 4-2-2. Red edge shift

"Red shift is due to chlorophyll concentration change. High chlorophyll increases absorption in the red region and pushes the red edge to longer wavelengths"[20]. (Red edge area was marked by red line on the original picture.)

Red edge is a shift of vegetation reflectance from Red bands to Near Infrared bands. It was proved widely as a unique spectral feature of vegetation which was highly related to chlorophyll content in leaf. In the figure2, in the red edge range from 670 to 780nm (Clevers, et al., 2004), when chlorophyll concentration decreases, the red edge moved to short wavelength, the area of red

[20] http://www.seos-project.eu/modules/agriculture/agriculture-c01-s02.html

edge increased dramatically. It is able to reduce noise and responds the changes of chlorophyll sensitive. Thus, the area around by red edge would be an ideal parameter to build the prediction model of chlorophyll.

Red edge position (REP) is generally defined as the largest changing rate (the inflection point) of spectral reflectance curve at red-NIR slop (Clevers, et al., 2004, Dawson, et al., 1998), namely, the band (position) where the maximum of the first derivative of the original curve at the bands is. The REP can be studied by plotting $dR/d\lambda$, as a function of λ. Equation 1 defines REP with 1λ-step. It could be defined with 4λ-steps. Step in derivate equation controls effects of data smooth. The larger it is, the smoother the derivative curve is.

$$\frac{dR(\lambda_i)}{d\lambda} = \frac{\left(\sum_{j=i-1}^{i} R(\lambda_j) - \sum_{j=i}^{i+1} R(\lambda_j)\right)}{2\Delta\lambda} \quad (4\text{-}2\text{-}1)$$

Previous research work defined the area of the red edge (RDa) as the sum of first derivative of spectrum in the red edge (equation 2) and found RDa highly related with chlorophyll (Filella and Penuelas, 1994).

$$Rda = \sum_{680-780nm} sum(dR/d\lambda) \quad (4\text{-}2\text{-}2)$$

According to the definition of Filella et al., RDa is the sum of gradients. In this section the area of red edge (RDaa) was redefined as sum of reflectance in the red edge directly as equation 3.

$$Rdaa = \sum_{680-780nm} sum(R) \quad (4\text{-}2\text{-}3)$$

Plant leaf has a unique characteristic in red and near infared red spectral wavelength. Photosynthetic pigments have strong absorption in red range and even no absortpion in the near infared range. Therefore some researchers proved thay to take non-sensitive bands as reference, the new regresson model could be calibrated and noise could be spuressed. Previous studies found reflectance at 445 (R445, the same below) nm was constant until total chlorophyll content dropped less than 4% of maximal chlorophyll content and R445 would be good reference(Sims, DA and Gamon, JA, 2002). The ratio of R550 and R700nm was constant at the level regardless of the differences between chlorophyll concentrations (Kim, et al., 1994b). 940nm would be another reference band to build spectral index for chlorophyll estimation. Hoel et al. (1998) pointed out that at 940nm, chlorophyll had no absorption. The chlorophyll meter(SPAD 502) is just buiied based on this idea.

Thus, these three band regions were chosen to be constant to chlorophyll. For matching with Rdaa and resistance to noise, the reference indices were extended from a point band to a range of bands.

Therefore, new indices for chlorophyll prediction would be defined spectral area index of Ind as

following:

$$Ind1=Rdaa/(R_{440\text{-}450}) \quad (4\text{-}2\text{-}4)$$

$$Ind2=Rdaa/(R_{540\text{-}560}/R_{690\ 710}) \quad (4\text{-}2\text{-}5)$$

$$Ind3=Rdaa/(R_{930\text{-}950}) \quad (4\text{-}2\text{-}6)$$

For comparison, the index of Inda in which Rdaa was replaced by Rda were calculated too and marked as Inda1, Inda2 and Inda3.

Analyzing the red edge shift in figure 2, the amplitude would be not very big. Generally the amplitude of red shift ranges from several nanometers to decades nanometers. Therefore, when calculating the area of red edge inputting the whole reflectance data with edges, the sensitivity of the area responding to change of chlorophyll would be reduced seriously. In order to overcome it, revised area indices were proposal in this section based on the index of Ind.

Generally, the red edge position locates in the range of [700-730]. It indicates that the spectral responses in this range to chlorophyll changes, if any, are sharply, and the area of this range contributes most to responses of the area in whole the edge responding to chlorophyll. Therefore, by adjusting the scope of the edge for calculating the area, the sensitivity of the area index responding to chlorophyll could be controlled. Thus, the spectral area indices of Ind were redefined as follow:

$$Ind1=Rdaa'/(R_{435\text{-}455}) \quad (4\text{-}2\text{-}7)$$

$$Ind2=Rdaa'/(R_{540\text{-}560}/R_{690\text{-}710}) \quad (4\text{-}2\text{-}8)$$

$$Ind3=Rdaa'/(R_{930\text{-}950}) \quad (4\text{-}2\text{-}9)$$

Where $Rdaa' = \sum_{700}^{730nm} sum(R)$.

For easy to understand, the symbols of these indices were assigned as:

The indices defined in the original red edge [680, 780nm]:

Rda_o was the sum of first derivate spectral reflectance in red edge;

Inda1_o, Inda2_o and Inda3_o were the ones based on Rda_o.

Rdaa_o represents the sum of spectral reflectance in red edge.

The incices defined in the range of [700, 730nm]:

Rda was the sum of first derivate spectral reflectance in [700, 730nm]

Rdaa was the sum of spectral reflectance in [700, 730nm]

Ind1, Ind2 and Ind3 were the ones based on Rdaa.

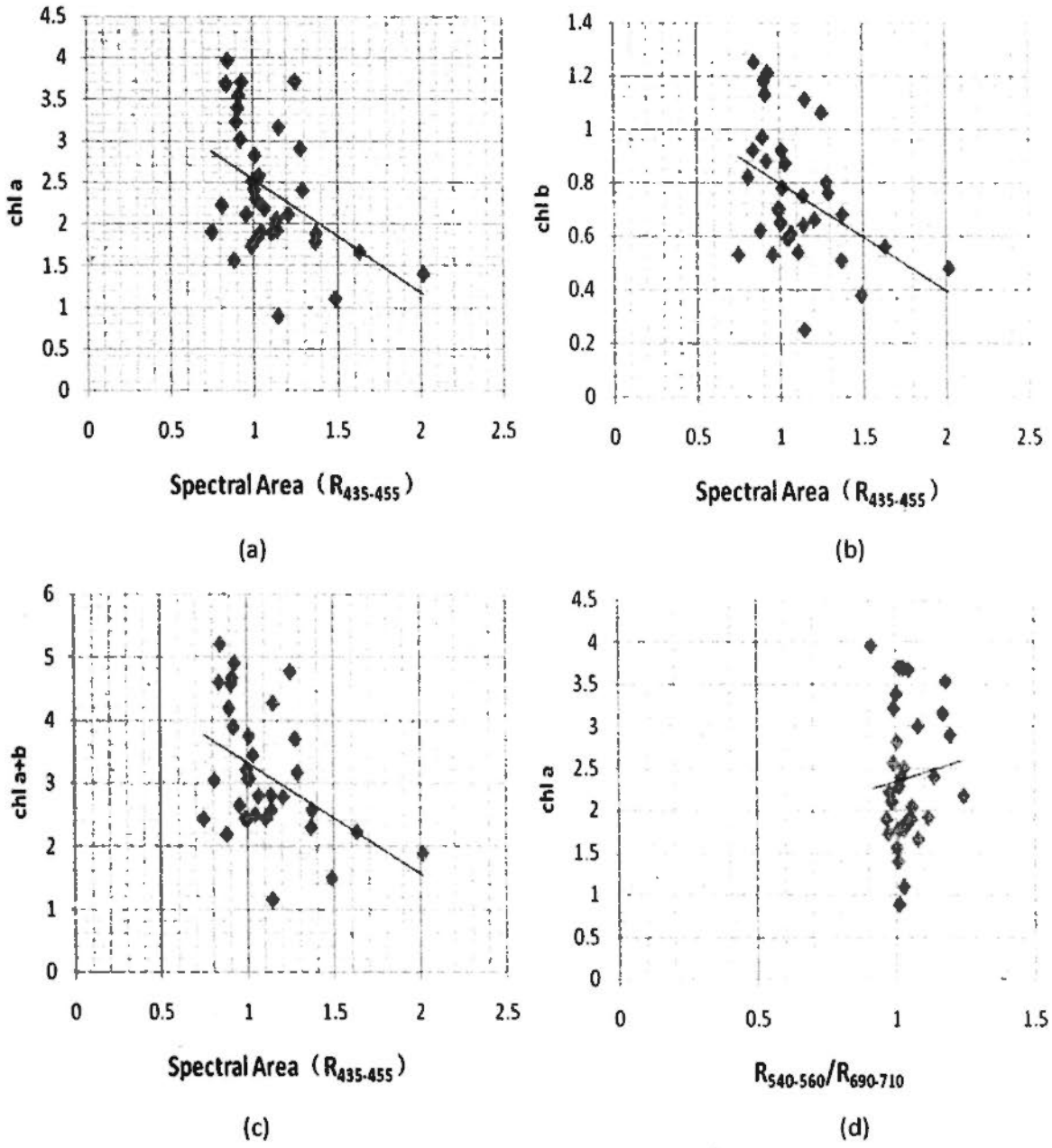
4.2.2.2 Results and discussion

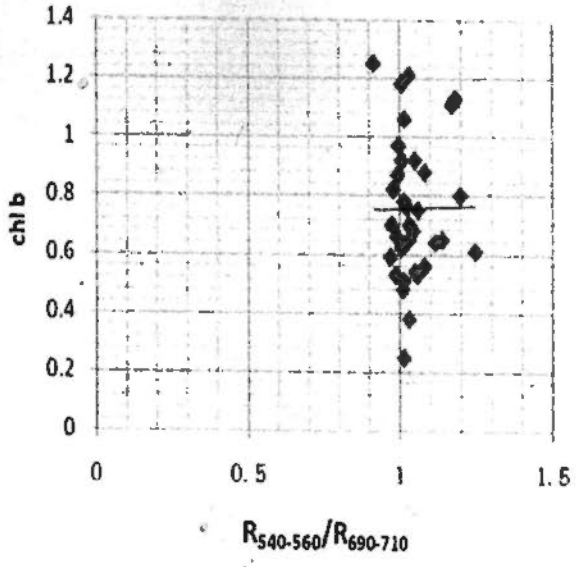
The indices defined with original input data were marked as Ind_o. And in this section spectral index of datt was also obtained for comparison. In the section 4.1, this index had been proved

superior to be indicator as chlorophyll content.

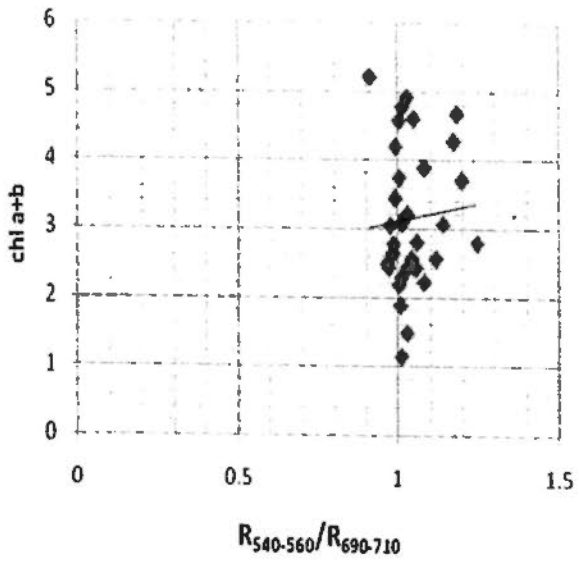
4.2.2.2.1 Sensitivity analysis for reference indices to chlorophly concentrantion

Figure 3 showed the relationship between the content of chlorophyll and the reference indices. These three reference showed no relationship with chlorophyll namely, the reference indices were non-sensitive to chlorophyll.

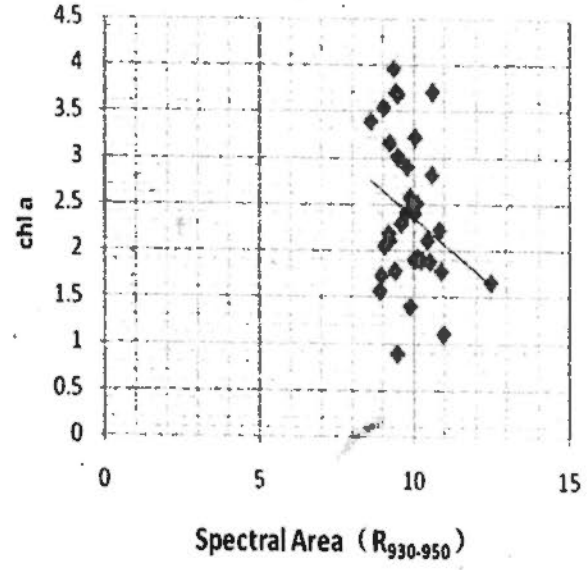



chl b
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
$R_{540-560}/R_{690-710}$

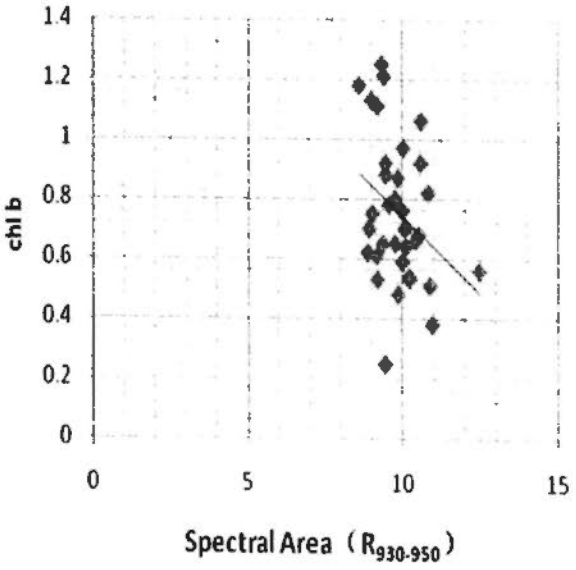
(e)


chl a+b
6
5
4
3
2
1
0
0
0.5
1
1.5
$R_{540-560}/R_{690-710}$

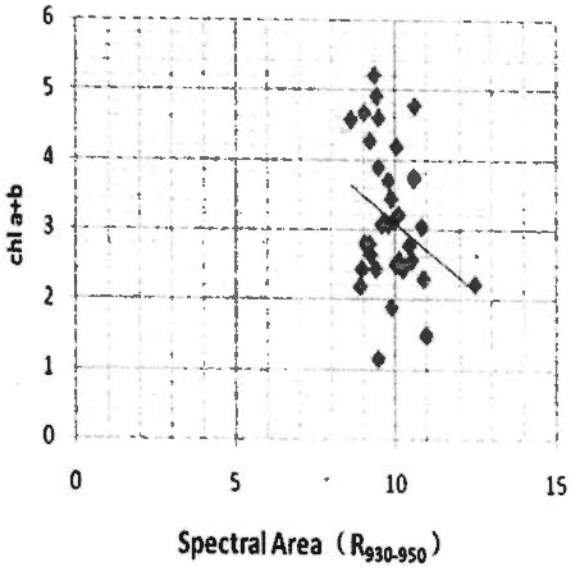
(f)


chl a
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
5
10
15
Spectral Area ($R_{930-950}$)

(g)


chl b
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
Spectral Area ($R_{930-950}$)

(h)


chl a+b
6
5
4
3
2
1
0
0
5
10
15
Spectral Area ($R_{930-950}$)

(i)

Figure 4-2-3. Sensitivity analysis for reference indices to chlorophly concentrantion

4.2.2.2.2 Relationship between new indices and chlorophyll

Table 1, 2 and 3 showed the results of correlation analysis between indices and chlorophyll. Ind2, Ind3, Ind3_o, Rdaa and Rdaa_o were high related to the content of chlorophyll a, chlorophyll b and chlorophyll a+b. Because of data missing, here chlorophyll a+b were replaced by chlorophyll a + chlorophyll b (Zhang, 1985). For prediction of chl a, the best indicator was Ind2, the correlation coefficient was up to -0.81. The result of Ind3 and Rdaa was similar. That was because the reference of Ind3 was concentrated around 1, thus Ind3 was almost equal to Rdaa approximately. The same results for Ind3 and Rdaa were shown when they were used to estimate chl b and chl a+b. Applying the redefined area as indicators of chlorophyll was more efficient, more sensitive and stronger related to chlorophyll than using the orignal one. Taking Rdaa divided by the reference index was kind of re-projection to make the results between different samples to compare each other directly. However, the reference index itself would be with random or systematic noise, thus the correlation coefficients of Ind and chlorophyll were smaller than them of Rdaa and chlorophyll. Also, the stability of the reference, e.g. reasonability of its construction, would affect the result too. In some sensing, it made new index complicated and brought risk. However, if more than one kind species sample were analyzed, this re-projection is necessary. Meanwhile, since Rdaa divided reference index, the relationship between ind_o and it had been strengthened. For example, when estimating chl a, the coefficient of Rdaa_o and Chl a was just 0.574. By re-projected with the reference index ($R_{930,950}$), it was enhanced -0.733, much highly related than the original Rdaa_o. It revealed that the reference index also could help to reduce and suppress the influence from non-related variables.

The results obtained by the new defined indices were highly consistent with the one from datt. However, the new indices contain more physical mean and could be easily explained and understood.

Table 4-2-1. Regression results between spectral indices and content of chl a

Y=content(chl a) mg/g(per fresh weight)

Index (x)	Pearson Correlation	Regression Model			Residual	
		Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Ind1_o	0.160	##	##	##	##	##
Ind1	-0.168	##	##	##	##	##
Ind2_o	-0.368	##	##	##	##	##
Ind2	-0.381	Y=5.292-0.400x	0.463	0.5670	0.5656	1.452
Ind3_o	-0.733	Y=12.297-0.3129x	0.537	0.5339	0.5253	1.470
Ind3	-0.728	Y=6.48-4.956x	0.531	0.5376	0.5288	1.325
Inda1_o	0.342	##	##	##	##	##
Inda1	0.361	##	##	##	##	##
Inda2_o	0.080	##	##	##	##	##
Inda2	0.081	##	##	##	##	##
Inda3_o	0.224	##	##	##	##	##
Inda3	0.269	##	##	##	##	##
Rda_o	-0.068	##	##	##	##	##
Rda	0.075	##	##	##	##	##

Rdaa_o	-0.574	Y=6.932-0.145x	0.380	0.6425	0.6320	1.995
Rdaa	-0.731	Y=5.770-0.449x	0.534	0.5358	0.5271	1.756
datt	0.728	Y=0.176+42.589x	0.530	0.5380	0.5292	1.484

"##" means no or very slight relationship between the two variables. It is the same as other tables. The explanations of the statistic parameters would be obtained in the section 4-1.

Table 4-2-2. Regression results between spectral indices and content of chl b

Y=content(chl b) mg/g(per fresh weight)

Index	Pearson	Regression Model			Residual	
(x)	Correlation	Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Ind1_o	0.130	##	##	##	##	##
Ind1	-0.181	##	##	##	##	##
Ind2_o	-0.331	##	##	##	##	##
Ind2	-0.637	Y=1.554-0.111x	0.405	0.1796	0.1767	1.710
Ind3_o	-0.694	Y=3.533-0.879x	0.481	0.1678	0.1651	1.989
Ind3	-0.696	Y=1.815-1.406x	0.485	0.1672	0.1645	1.924
Inda1_o	0.308	##	##	##	##	##
Inda1	0.312	##	##	##	##	##
Inda2_o	0.102	##	##	##	##	##

Inda2	0.077	##	##	##	##	##
Inda3_o	0.226	##	##	##	##	##
Inda3	0.218	##	##	##	##	##
Rda_o	-0.088	##	##	##	##	##
Rda	0.023	##	##	##	##	##
Rdaa_o	-0.574	Y=2.098-0.043x	0.329	0.1907	0.1876	1.933
Rdaa	-0.717	Y=1.734-0.131x	0.514	0.1624	0.1598	1.983
datt	0.738	Y=0.084+12.814x	0.545	0.5772	0.1546	1.935

Table 4-2-3. Regression results between spectral indices and content of chl a+b

Y=content(chl a+b) mg/g(per fresh weight)

Index	Pearson	Regression Model			Residual	
(x)	Correlation	Fit equation (Unstandardized)	R^2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Ind1_o	0.155	##	##	##	##	##
Ind1	-0.173	##	##	##	##	##
Ind2_o	-0.364	##	##	##	##	##
Ind2	-0.679	Y=8.846-0.511x	0.461	0.7374	0.7254	1.486
Ind3_o	-0.733	Y=15.830-4.008x	0.538	0.6832	0.6721	1.572

Ind3	-0.730	Y=7.963-6.362x	0.533	0.6862	0.6750	1.437
Inda1_o	0.339	##	##	##	##	##
Inda1	0.354	##	##	##	##	##
Inda2_o	0.087	##	##	##	##	##
Inda2	0.081	##	##	##	##	##
Inda3_o	0.227	##	##	##	##	##
Inda3	0.261	##	##	##	##	##
Rda_o	-0.074	##	##	##	##	##
Rda	0.064	##	##	##	##	##
Rdaa_o	-0.582	Y=9.030-0.189x	0.338	0.8173	0.8040	1.973
Rdaa	-0.737	Y=7.504-0.580x	0.543	0.6792	0.6682	1.789
datt	0.740	Y=0.261+55.404	0.547	0.6760	0.6650	1.551

Table 4-2-4. Comparison of regression results between the spectral indices based on the original signal and polluted one

Y=content(chl a) mg/g(per fresh weight)

Index	Pearson	Regression Model			Residual	
(x)	Correlation	Fit equation (Unstandardized)	R2	Std. Error of the Estimate	Std. deviation	Durbin-Watson test
Rdaa	-0.731	Y=5.770-0.449x	0.534	0.5358	0.5271	1.756

Rdaa_1k	-0.729	Y=5.894-0.462x	0.531	0.5488	0.5401	2.027
Rdaa_10k	-0.729	Y=5.897-0.462x	0.531	0.5485	0.5399	2.026
datt	0.728	Y=0.176+42.589x	0.530	0.5380	0.5292	1.484
datt_1k	0.660	Y=0.364+39.763x	0.436	0.6019	0.5925	2.020
datt_10k	0.656	Y=0.383+39.487x	0.431	0.6045	0.5950	2.055

Rdaa_1k represents it was calculated based on the spectral reflectance data which were added 0.001*random noises, while Rdaa_10k were based on 0.0001*random noises. The same as others. The noises were generated by the function of randn under Matlab software. They obey normal distribution which mean and standard deviation was 0 and 1 respectively.

4.2.2.2.3 Analysis of sensitivity to noise for new spectral indices

In this section, new spectral indices were designed based on the idea that the area was able to reduce and suppress noises better than the one based a narrow bands (individual band). For further evaluation, the noise was added to the original spectral reflectance. The spectral reflectance data were added to random noises which were generated by the function of "*randn*" under Matlab software. They fitted to normal distribution of which mean and standard deviation was 0 and 1 respectively. The figure 4 showed the details of the curves. Table 4 gave the regression analysis results. Here "*Rdaa* "and "*datt*" were chosen as examples. Rdaa was the index based on the area of the curve while datt was based narrow bands. The correlation coefficients of Rdaa and chl a had very slight changes for different noise level. For datt, when noises added, the correlation coefficients decrease obviously. The R2 of regression equation of datt also dropped seriously, while Rdaa' almost had no changes. These statistic results revealed the index based on the area could help to reduce and suppress noise well.

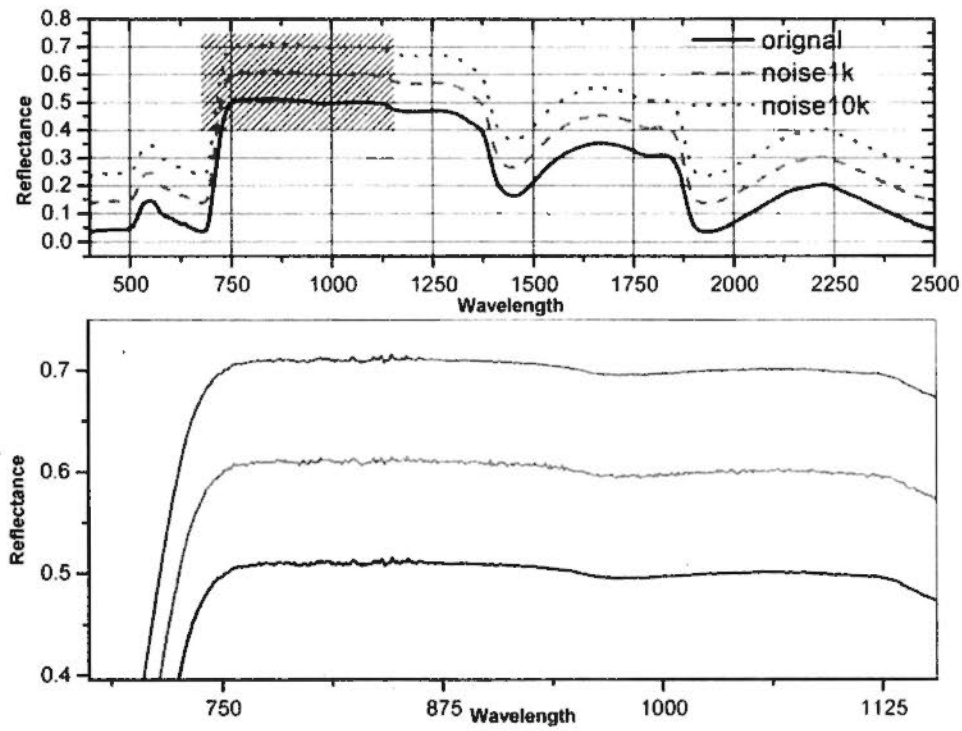


Figure 4-2-4. Spectral reflectance curve mixed with different normally distributed random numbers.

The black curve was the orignal data, the red curve was added 0.001*random noises and the blue one was add 0.0001*random noises.

4.2.2.2.4 Analysis of correlation within spectral indices and their sensitivity

Table 4-2-5. Pearson correlations analysis between new spectral indices

item	Ind2	Ind3	Inda3_o	Rdaa	Rdaa_0	datt
Ind2	1					
Ind3	.839(**)	1				
Inda3_o	.829(**)	.989(**)	1			
Rdaa	.947(**)	.873(**)	.884(**)	1		
Rdaa_0	.826(**)	.560(**)	.598(**)	.891(**)	1	
datt	-.712(**)	-.863(**)	-.843(**)	-.839(**)	-.620(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

In table 5, spectral reference could help to reduce and suppress the influence from non-related variables. Ind3 and Inda3_o were built with same spectral reference but with different variables of the area. However, by re-projection the non-sensitive proportion spectral reflectance in Rdaa_o were suppressed, and therefore the coefficient increased to 0 989. Rdaa_o was less sensitive to chlorophyll. Taking datt as a reference, it was found the correlation coefficient of it and datt was only 0.620 while the one with Rdaa was up to 0.839. The correlation coefficient of Ind2 and Rdaa was up to 0.947. It revealed the spectral reference of ($R_{540\text{-}560}/R_{690\text{-}710}$) was more stable than the others. The correlation coefficient of (datt, Rdaa) and the coefficient of (datt, Ind3) was very close to each other. The correlation coefficient of (Rdaa_o, Inda3_o) was less than 0.6 while the one of (Rdaa, Inda3_o) was more than 0.8, namely Inda3_o was closer to Rdaa. These results indicated it was reasonable to choose the range of [700, 730nm] as the red edge position location which was sensitive and strong related to chlorophyll.

4.2.3 New spectral index as indicator of carotenoids

4.2.3.1 Basic assumption

By reviewing the pattern of carotenoids absorption, it could be found that the features of carotenoids absorption overlapped with the features of chlorophyll's (in figure 1) in wavelength ranging from 450 to 550 nm approximately. For estimating properties of carotenoids accurately, firstly the effects from chlorophyll absorption should be suppressed or reduced.

In this section, an assumption of absorption in [450, 550nm] had been proposed, that is:

The spectral reflectance features in this range are comprehensive effects of chlorophyll (chl) and carotenoids (car) absorption. And the effects of the two kinds of pigments are multiplicative. Thus, the spectral reflectance in this range could be expressed by the equation 10:

$$R_i(\lambda) = \beta \Omega_i(chl) * \Phi_i(car) + \xi \quad (4\text{-}2\text{-}10)$$

Where R is a comprehensive spectral response, i is the wavelength, $\Omega_i(chl)$, $\Phi_i(car)$ are spectral response to chlorophyll and carotenoids respectively. β and ζ are the coefficient and error of function. Therefore, spectral response of carotenoids would be given as follow:

$$\Phi_i(car) = \alpha \frac{R_i(\lambda)}{\Omega_i(chl)} + \varepsilon \quad (4\text{-}2\text{-}11)$$

Where α and ε are the coefficient and error of function.

4.2.3.2 New spectral index as indicator of carotenoids

Based on the equation of 11, two input parameters should be obtained first, namely sensitive bands for carotenoids in [450, 550nm] and features of spectral reflectance to chlorophyll absorption.

4.2.3.2.1 Sensitive bands for carotenoids

To carotenoids absorption pattern, the sensitive bands range was located during [450, 550nm]. Considering the resolution (3nm @ASD spectrometer) and the characteristic of signal-to-noise (SN) of equipment, the step was set as 10 nm. [450-459, 460-469... 560-569, 570-579]. "450-459"represents the sum of spectral reflectance from 450nm to 459 nm, the same as others. Backward of linear regression approach had been applied to finish sensitive bands selected task. The result of regression analysis showed the following bands were relatively sensitive to carotenoids: R530_539, R540_549, R550_559, R520_529, R560_569 and R570_579. This result coincided with the previous study (Barton, et al., 2001, Gamon, et al., 1997). In fact, Photochemical Reflectance Index (PRI) was just defined with reflectance at 531 and 570 nm. In this section, R530_539 and R560_569 were chosen as input variables.

4.2.3.2.2 Chlorophyll absorption responding bands.

Backward of linear regression approach had been applied to assess correlation-ship between the bands selected and chlorophyll. The result showed that the best significant regression model was combined with R480_489, R500_509, R530_539, and R570-579. Here, R480_489 was selected as input variable. And the spectral responses at R480_489 and R500_579 are additive assumed in equation 12.

$$\Omega(chl) = \frac{(R_{480\_489} + R_{500\_509})}{R_{930\_950}} \quad (4\text{-}2\text{-}12)$$

If when assuming the responses in the two bands ranges are multiplicative, the equation 12 could be redefined as:

$$\Omega\ (chl) = \frac{(R_{480\_489} * R_{500\_509})}{R_{930\_950}} \quad (4\text{-}2\text{-}13)$$

4.2.3.2.3 New spectral index for carotenoids estimation

Thus, the proportion of spectral response caused by Carotenoids could be defined as equation 14 or 15:

$$\Phi\ (car) = \frac{R_{530\_539}}{\Omega} = R_{530\_539} * \frac{R_{930\_950}}{(R_{480\_489} + R_{500\_509})} \quad (4\text{-}2\text{-}14)$$

$$\Phi\ (car) = \frac{R_{530\_539}}{\Omega} = R_{530\_539} * \frac{R_{930\_950}}{(R_{480\_489} * R_{500\_509})} \quad (4\text{-}2\text{-}15)$$

When assuming additive relationship for responses of chlorophyll (figure 4a), the two variables were obviously high related, and the correlation coefficient of them was up to -0.779.

If changing R480_489 and R500_509 from additive relationship to multiplicative one to calculate $\Omega\ (chl)$, $\Phi\ (car)$, the results showed there was no relation between $\Phi\ (car)$ and the content of carotenoids (figure 4b). The correlation coefficient of the variables based on idea was -0.368 which implies the variables were not related.

Therefore, the assumption about carotenoids was reasonable and acceptable for description of the spectral features responding to carotenoids absorption while the proportion of spectral response to chlorophyll was described as additive relationship. New index calculated by equation 14 was accepted to describe and estimate the content of carotenoids .

(a)

(b)

Figure 4-2-5. Relationship between the content of Caretonoinds (mg/g, per fresh weight) and new spectral index.

(a) the proportion of spectral response to chlorophyll is described as additive relationship(equation 12,14); (b) the proportion of spectral response to chlorophyll is described as multiplicative relationship(equation 13,15).

4.2.4 Conclusion

The area of red edge defined as the sun of the first derivate spectral reflectance had been pointed out related with chlorophyll, however the statistic results showed it was very weak related to chlorophyll. Therefore, in this section, 15 spectral parameters as chlorophyll indicators were designed. The results showed that Ind1_o and Ind2_o were not related to chlorophyll. Taking account that the red edge position locates in the range of [700-720nm] generally, Rdaa was revised and limited within [700, 720nm] and Ind1, Ind2 and Ind3 based on Rdaa. The Rdaa was proved much more sensitively and strongly related to chlorophyll. Ind1 was not related to all chlorophyll parameters while Ind2 was opposite. Because the spectral reference for Ind3 was close to 1 approximately, thus its characteristic was similar to it of Rdaa. Then the sensitivity of the spectral reference had been assessed. The results showed the Rdaa, Ind2 and Ind3 were all no-related to chlorophyll. Meanwhile, the results showed that by diving spectral reference, a kind action of re-projection, the influence from non-related variables could be reduced and suppressed.

In this section, an assumption was proposed. For estimating carotenoids, that was the effect of chlorophyll and carotenoids were multiplicative in [450, 550nm] and a new spectral index $\Phi(car)$ was defined. $\Phi(car)$ was strongly related to carotenoids and could be applied as

the indicator of carotenoids content.

However, these new spectral indices were developed based on data of Lopex93 and paddy rice. Its application in wide range should be evaluated.

References

Bannari, A., Khurshid, K. S., Staenz, K., & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(10), 3063-3074.

Barton, C., & North, P. (2001). Remote sensing of canopy light use efficiency using the photochemical reflectance index:: Model and sensitivity analysis. *Remote Sensing of environment*, 78(3), 264-273.

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

Blackburn, G. (2007). Hyperspectral remote sensing of plant pigments. *Journal of experimental botany*, 58(4), 855.

Blackburn, G. A., & Ferwerda, J. G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of environment*, 112(4), 1614-1632

Chappelle, E., Kim, M., & McMurtrey, J. (1992). Ratio analysis of reflectance spectra (rars): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sensing of environment*, 39(3), 239-247.

Clevers, J., Kooistra, L., & Salas, E. (2004). Study of heavy metal contamination in river floodplains using the red-edge position in spectroscopic data. *International Journal of Remote Sensing*, 25(19), 3883-3895.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111-121.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229-239.

Dawson, T., & Curran, P. (1998). Technical note a new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing*, 19(11), 2133-2139.

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459-1470.

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote sensing of environment*, 41, 35-44.

Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types, and

nutrient levels. 112(4): 492-501.

Gitelson, A., Gritz , Y., & Merzlyak, M. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271-282.

Gitelson, A., Keydan, G., & Merzlyak, M. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Gitelson, A. A., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148(3-4), 494-500.

Hoel, B. O., & Solhaug, K. A. (1998). Effect of irradiance on chlorophyll estimation with the minolta spad-502 leaf chlorophyll meter. *Annals of Botany*, 82(3), 389-392.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., & Foley, W. J. (2004). Estimating foliage nitrogen concentration from hymap data using continuum removal analysis. *Remote Sensing of environment*, 93(1-2), 18-29.

Kim, M., Daughtry, C., Chappelle, E., McMurtrey III, J., & Walthall, C. (1994). The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (apar). 299-306.

Richardson, A. D., Duigan, S. P., & Berlyn, G. P. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist*, 153(1), 185-194.

Sims, D., & Gamon, J. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of environment*, 81(2), 337-354.

Zhang, q. (1985). Approaches for the content of chlorophyll measurement(测定叶绿素的几种方法). *Chinese Bulletin of Botany*, 3(5), 60-64.

4.3 Estimation of the content of water, nitrogen and lignin

Chemical concentrations of vegetation leaf are important indicators of internal bio-processes. Besides photosynthetic pigments (chlorophyll a and b, carotenoids), nitrogen concentration, water content, lignin are also included. They are good indicators for vegetation growth monitoring.

4.3.1. Nitrogen and lignin content estimation methods

Curran (1989) pointed that nitrogen was an important indicator of photosynthetic rate and overall nutritional status. Previous' studies focused on extraction of nitrogen information from dried ground leaves (Dury, et al., 2002, Grossman, et al., 1996). There were also many attempts of extraction nitrogen information from fresh leaves (Dury and Jia, 2002, Huang, et al., 2004). However, the serious problems for direct estimation of nitrogen were the masking effects of leaf

water absorption. *"Leaf water absorption and the overlapping of other chemical absorption features tend to mask subtle nitrogen absorption feature"*(Clark, et al., 1984, Huang, et al., 2004), and they also pointed out that by continuum removal technique, some of these absorption features of no interest could be suppressed. Some researchers reported an indirectly inversion method for nitrogen content that nitrogen content was highly related to chlorophyll content in leaf, and then the prediction of nitrogen status was transformed to estimate chlorophyll content(Sun, et al., 2005).

Leaf lignin concentration is an important factor to control plant growth in that it is related to litter decomposition rates and nitrogen and lignin have same sensitive absorption bands (Serrano, et al., 2002). Therefore, these two studied together.

In this section, the absorption features of spectral reflectance or related responding wavelength to nitrogen and lignin were chosen and assessed, according to the references.

4.3.2. Bands selected based on stepwise regression analysis

Nitrogen and lignin related wavelength ranges were censored in previous studies (Curran, 1989, Huang, et al., 2004, Okuyama, et al., 1998, Serrano, et al., 2002, Song, et al., 2009). Considering the strong absorption of leaf water (figure 1), the related wavelength ranges were relocated (table1). Meanwhile, the individual band was replaced by the sum of a certain wavelength range to suppress noises. The range (step) was set as 20nm, and a total of 39 regression variables almost covered the whole near and short-wave infrared bands excluding water strong absorption. The input data could be expressed as the following form:

$$Input_k(i) = \sum_{j=i-step}^{i-1} d_j \quad (4\text{-}3\text{-}1)$$

Where i and j is the wavelength of i and j nm, k represents the k^{th} paramieter for regression. d_j is the j^{th} value in the data series.

These 39 variables were calculated based on the original spectral reflectance. The First derivative and the second derivative was able to enhance subtle absorption features of leaf bio-chemicals and overlaying features (Huang, et al., 2004). Thus these variables were also calculated based on the first and second derivative of spectral reflectance. These data were analyzed with the approach of stepwise regression.

Table4-3-1. Pre-selected potential absorption features

Wavelength ranges (nm)	Potential absorption features(variables)
580-700	R580_599,R600_619,R620_639,R640_659,R660_679,R680_699
1000-1080	R1000_1019,R1020_1039,R1040_1059,R1060_1079

1500-1800	R1500_1519,R1520_1539,R1540_1559,R1560_1579,R1580_1599,R1600_1619,R1620_1639,R1640_1659,R1660_1679,R1680_1699,R1700_1719,R1720_1739,R1740_1759,R1760_1779,R1780_1799
2100-2200	R2100_2119,R2120_2139,R2140_2159,R2160_2179,R2180_2199
2230-2410	R2230_2249,R2250_2269,R2270_2289,R2290_2309,R2310_2329,R2330_2349,R2350_2369,R2370_2389,R2390_2409

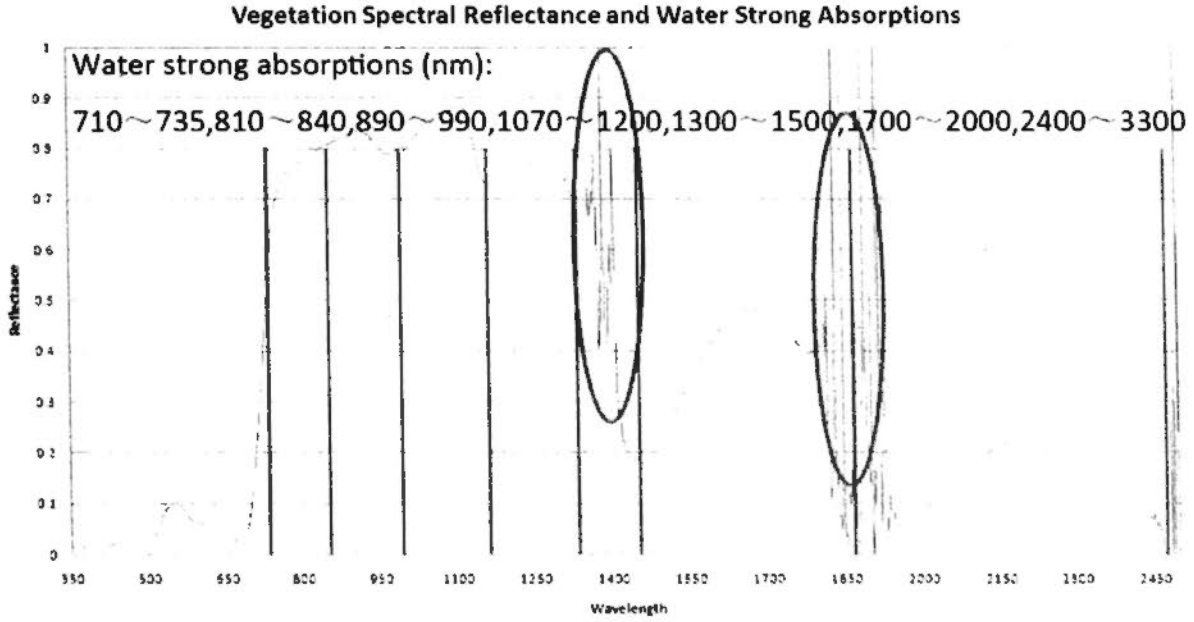


Figure 4-3-1. Strong water absorptions under natural illumination conditions.

The ranges in green circles are seriously influenced by vapor in atmosphere. However, when with the contact probe and the leaf-clip unit, these effects of vapor absorption could be avoided.

Serano et al. (2002) developed Normalized Difference Nitrogen Index (NDNI) and Normalized Difference Lignin Index (NDLI) for estimation of nitrogen and lignin content respectively. In this section these two indices were also obtained. It is necessary to assess both NDNI and NDLI when they are used as indicators of foliar chemicals in this study. Also, by comparison, a cross-validation for the absorption features each other can be done.

$$NDNI = \frac{\log(1/R_{1510}) - \log(1/R_{1680})}{\log(1/R_{1510}) + \log(1/R_{1680})} \quad (4\text{-}3\text{-}2)$$

$$NDLI = \frac{\log(1/R_{1754}) - \log(1/R_{1680})}{\log(1/R_{1754}) + \log(1/R_{1680})} \quad (4\text{-}3\text{-}3)$$

4.3.3 Results and discussion

Figure 2 showed the correlation coefficients of the pre-selected absorption variables and nitrogen. For variables based on the original spectral reflectance, except for the ones located in (1000, 1080nm), others were highly related to nitrogen. And the strongest was the one of R1500_1519

with nitrogen, up to 0.736. This result is consistent with the previous studies (Huang, et al , 2004, Song, et al., 2009). (580, 700nm) was chlorophyll sensitive bands. In the figure 2, the variables in this range also were highly related with nitrogen. Derivative transform could enhance absorption features. However, the derivative results represent change rate of absorption. It is not sensitive to broad bands and small changing rate of the reflectance. Thus the coefficients between variable and nitrogen decreased dramatically, especially for the variables based on the second derivative spectral reflectance. R1720_1739 and R1760_1779 were the most strongly related to nitrogen, the coefficients were more than 0.8. A possible explanation is these ranges were strongly related by water strong absorption, the absorption intensity changes quickly. When the second derivative was applied, there were only 5 variables were related to nitrogen. NDNI was also highly related to nitrogen.

Figure 4-3-2. Correlation coefficients between varibles and nitrogen concentration.

"origial" represents variables obtained bases on the original spectral reflectance data; "First derivative" represents variables based on the first derivative spectral reflectance data; "Second derivative" represents variables based on the second derivative spectral reflectance data. The variables (X axis) marked 1 to 40 were R580_599, R600_619, R620_639, R640_659, R660_679, R680_699, R1000_1019, R1020_1039, R1040_1059, R1060_1079, R1500_1519, R1520_1539, R1540_1559, R1560_1579, R1580_1599, R1600_1619, R1620_1639, R1640_1659, R1660_1679, R1680_1699, R1700_1719, R1720_1739, R1740_1759, R1760_1779, R1780_1799, R2100_2119, R2120_2139, R2140_2159, R2160_2179, R2180_2199, R2230_2249, R2250_2269, R2270_2289, R2290_2309, R2310_2329, R2330_2349, R2350_2369, R2370_2389, R2390_2409 and NDNI.

Table 2 showed the regression model based on different variables. All models had been done collinearity statistics test. The tolerance of collinearity is a statistic parameter describing whether

the variables are multi-collinear between them. The tolerance ranges from 0 to 1, the closer it is to 1, and the less collinear the variables are. Model 2 had largest R^2, up to 0.754, the two input variable almost had no collinearity. However, the input variable, R2160_2179 was no related to nitrogen, the absolute value of the coefficient less than 0.3. Thus this model was not optimal in practice. Model 3 was the same as Model 2. Model 1 was built by R1500_1519 and R640_659 which correlation coefficients with nitrogen were -0.736 and 0.674. The R^2 of it was up to 0.639 thus it was the optimal one. Model 4 was also acceptable.

Table 4-3-2. Model summary for nitrogen concentration estimation

Model	Data type (spectral reflectance)	Predictors	coefficients	R^2	Collinearity Statistics (Tolerance)
1	Original	(Constant)	23.948	.639	
		R1500_1519	-2.845		.679
		R640_659	7.742		.679
2	First derivative	(Constant)	17.358	.754	
		R1760_1779	-14.811		.993
		R2160_2179	-14.028		.993
3	Second derivative	(Constant)	19.374	.648	
		R600_619	-13.219		.883
		R2140_2159	4.770		.894
		R640_659	-10.561		.980
4	Original	(Constant)	4.459	0.536	
		NDNI	107.720		1

Dependent Variable: Nitrogen.

B. Regression analysis for lignin

Figure 4-3-3. Correlation coefficients between varibles and lignin concentration.

The labels were the same meaning as those in figure 2. And the varibels (X axis) marked 1 to 39 were the same as varibles in figure2, while the 40th is NDLI.

Figure 3 showed that the strongest related bands between the variables first derivative and lignin located in (1500, 1660nm). For the variables based on the original data, the strongest sensitive bands located in (2100, 2400nm). These two ranges are where water absorption was relatively weak. The variables based on second derivative were not obviously related to lignin. Tables 3 summarized regression model for lignin estimation. Model based on first derivative had largest R^2, up to 0.749, and the input variable R1760_1779 was the most strongly related to lignin concentration while R1680_1699 was almost had no relation with lignin. Model 3 was the same case. Though the input variables of model 1 were so strong multi-collinear, the Variance Inflation Factor (VIF) was not very big (3.656), much less than 10. Thus this model was not optimal. Besides model 1, model 4 was also satisfied.

Table 4-3-3. Model Summary for lignin concentration estimation

model	Data type (spectral reflectance)	Predictors	coefficients	R^2	Collinearity Statistics (Tolerance)
1	original	(Constant)	20.764	0.560	
		R2390_2409	8.238		0.274
		R1720_1739	-4.748		0.274
2	First derivative	(Constant)	20.965	0.749	
		R1760_1779	15.453		0.972

		R1680_1699	7.600		0.972
3	Second derivative	(Constant)	7.059		
		R640_659	14.078	0.492	0.982
		R600_619	7.979		0.982
4	original	(Constant)	23.558	0.534	
		NDLI	-440.159		1

Dependent Variable: Lignin.

4.3.4 Conclusion

In this section, 39 potential absorption wavelengths were selected for nitrogen and lignin concentration estimation. Based on the original, the first and the second derivative spectral reflectance, totally 1190 variables were calculated. By stepwise regression analysis, the sensitive bands were selected, and optimal model for nitrogen and lignin concentration estimation were obtained. Meanwhile, both NDNI and NDLI were assessed. They were highly related to nitrogen and lignin respectively and sensitive enough to be indicators of the two leaf chemicals.

By assessment in this section, models for nitrogen concentration estimation and lignin were selected as indicators of nutritional status in paddy rice growth monitoring. The models are:

Y_nitrogen=23.948 -2.845R1500_1519 +7.742R640_659 (4-3-4)

Y_nitrogen=4.459+107.720NDNI (4-3-5)

Y_lignin=20.764+8.238 R2390_2409-4.748 R1720_1739 (4-3-6)

Y_lignin=23.558-440.159NDLI (4-3-7)

Where Rz_x represents the sum of the spectral reflectance ranging from z to x nm.

However, these models were the results of statistics, thus they are significant in statistic level and could not explain the real nitrogen and lignin status, e.g. Model 2 in table 2 and Model 3 in table 3. The statistic results were limited with the samples, sample quality, sample numbers and its distribution. They are just indicators.

References

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329-6340.

Curran, P. (1989). Remote sensing of foliar chemistry. *Remote Sensing of environment*, 30(3), 271-278.

Dury, S., & Jia, X. (2002). Use of a hybrid supervised and unsupervised classification model to determine nitrogen concentration of eucalypt tree foliage using hymap data. 2, 640-642.

Grossman, Y., Ustin, S., Jacquemoud, S., Sanderson, E., Schmuck, G., & Verdebout, J. (1996). Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of environment*, 56(3), 182-193.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., & Foley, W. J. (2004). Estimating foliage nitrogen concentration from hymap data using continuum removal analysis. *Remote Sensing of environment*, 93(1-2), 18-29.

Okuyama, T., Takeda, H., Yamamoto, H., & Yoshida, M. (1998). Relation between growth stress and lignin concentration in the cell wall: Ultraviolet microscopic spectral analysis. *Journal of Wood Science*, 44(2), 83-89.

Serrano, L., Pe uelas, J., & Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in mediterranean vegetation from aviris data:: Decomposing biochemical from structural signals. *Remote Sensing of environment*, 81(2-3), 355-364.

Song, S., Li, P., Gong, W., Zhang, L., Zhu, B., Lv, L., & Zhang, D. (2009). A narrow band combination model to determine leaf nitrogen and water content in rice. 4, 280-283.

Sun, X. m., Zhou, Q.-f., & He , Q.-x. (2005). Hyperspectral variables in predicting leaf chlorophyll content and grain protein content in rice. *Acta Agronomica Sinica*, 31(7), 844-850.

Chapter 5 Monitoring and screening transgenic paddy rice under controlled contrast experiments by field-hyperspectral measurements

5.0.1 Assumption of gene expression

In the proposed study, we had an important assumption:

Namely: if foreign genes were expressed in plant, there would made some differences comparing with their parent, these differences were, not limited, component of cell would be changed such as new matter generated or the content of existing component increasing/ decreasing because of gene encoding; structure of cell would be changed; biochemical or biophysical processes would be influenced because of foreign gene expression.

If changes happened, no matter associate with component or structure, they would be discovered by direct or indirect approaches of spectroscopy with equipments of certain sensitive and resolution

5.0.2 Basic idea and function

To be noted, by present resolution/ sensitivity of equipment (ASD), it could not to accomplish that transgene has been expression just by hyperspectral remote sensing technique.

What can be done to gene expression by hyperspectral approach?

By fine spectral data which are sensitive and stable to plant biochemical and biophysical process, we could extract spectral absorption and reflectance and construct relationship or model between these spectral characteristics and plant parameters such as pigments, biochemical and biophysical processes. Just as mentioned above, the changes of these parameters are sensitive to plant growth status, thus we could deduce relevant information from these changes. When

putting all samples under strictly controlled contrast conditions during whole growth stages, excluding or making samples suffering from the same external influences such as temperature, fertilizer and management, we could locate changes causing influences and inverse where changes happen in plant or what caused these changes. These results then would be given feedback to professional researchers having further analysis of pertinency and make conclusion about the expression and influences to receptor by laboratory approaches. In summary, hyperspectral remote sensing techniques play a role of detection and monitoring of gene by an indirect way from a macro-view.

Figure 5-0-1. Early Indication based HRST for transgene plant cultivation

5.0.3 Spectral analysis methods

5.0.3.1 Analysis based on spectral morphological features

This is a qualitative method focusing on spectral morphologic characteristics.

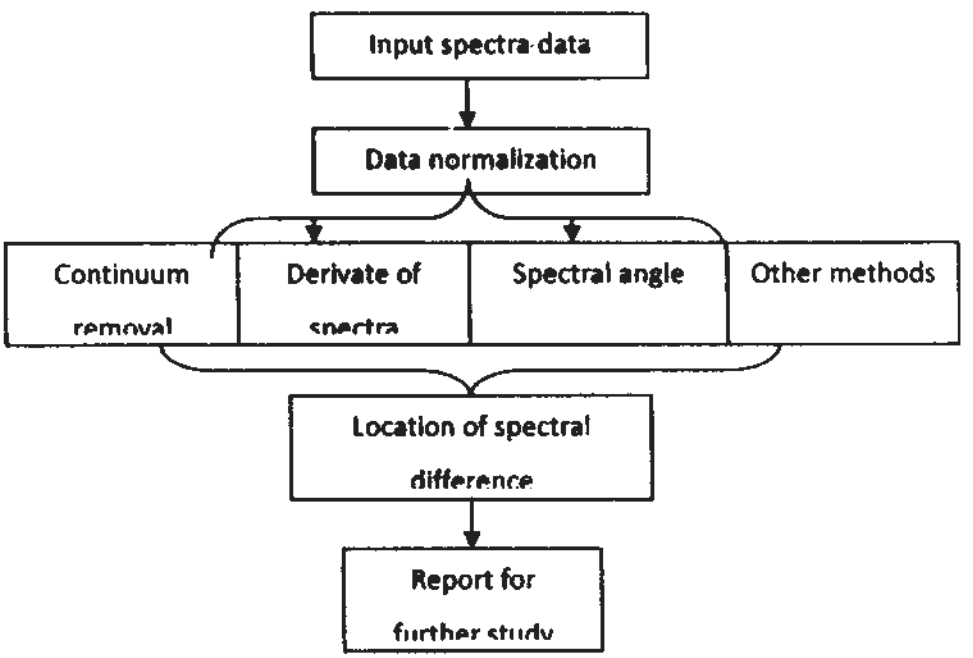


Figure 5-0-2. Analysis based on spectral morphological features

5.0.3.2 Analysis based on parameterized features

This is a quantitative method concentrating on diagnosing spectral characteristics such as chlorophyll absorption, red edge and red shift. By characteristic analysis could locate and qualify the changes compared with the parent directly.

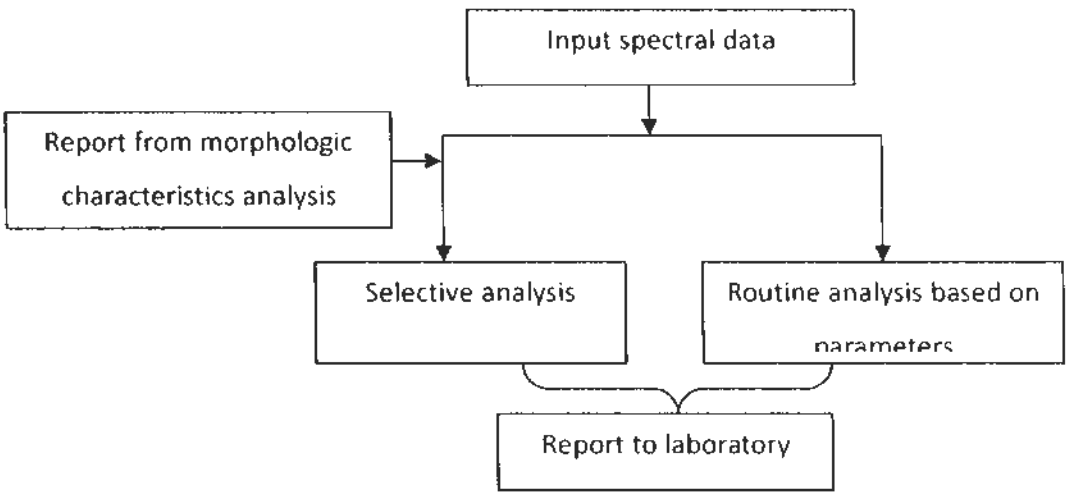


Figure 5-0-3. Analysis based on spectral parameterized features

5.1 Comparison analysis based on spectral reflectance edges within Sample groups[21]

Spectral parameters (indices) of the edges are highly related to foliar chemicals thus they are good indicators for vegetation growth monitoring. In this section, they would be applied to describe and asses the difference of the growth of the samples.

5.1.1 Data acquisition system

Fine spectra of samples were collected to avoid external noise such as atmospheric influence and incident stray from background by an Analytical Spectral Devices (ASD) FieldSpec 3 Spectrometer (Figure1a, the details of ASD could be found in chapter3). With help of high density contact probe (Figure1b) which had self-lightening system and white reference board (white panel), we built a closed data collection units which could ensure to avoid external noise clearly. The spectral data ranged from 350 to 2500nm.

[21] Parts of the results in this section had been published in the journal of Spectroscopy and Spectral Analysis (2010,30(1):202-205) and reported at 30th Asia Conference on Remote Sensing (2009), Beijing, China.

(a) (b)

Figure 5-1-1. a. Observation with probe-leaf-clip system; b. Field Spectrometer

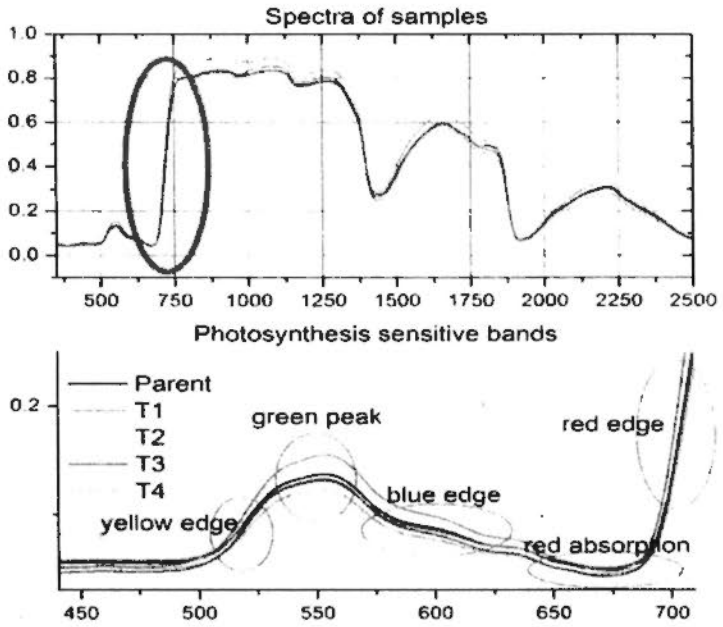


Figure 5-1-2. Spectra curves and the morphological details of photosynthesis sensitive bands.

In figure 2, at near infrared bands, the reflectance of leaves was higher than normal, up to 0.8. It was because of the bi-/multi-reflection. In the data collection system, a white reference board was used. When incident light in near infrared bands reached leaf surface or internal structure, some had been reflected directly and captured by ASD detector, a little had been absorbed, and the left reached to the white board and was reflected back because of the special structure and spectral characteristics of the leaves in near-infrared band, and then was captured. So the reflectance at near infrared bands had been selected enhancement. The fewer the incident light

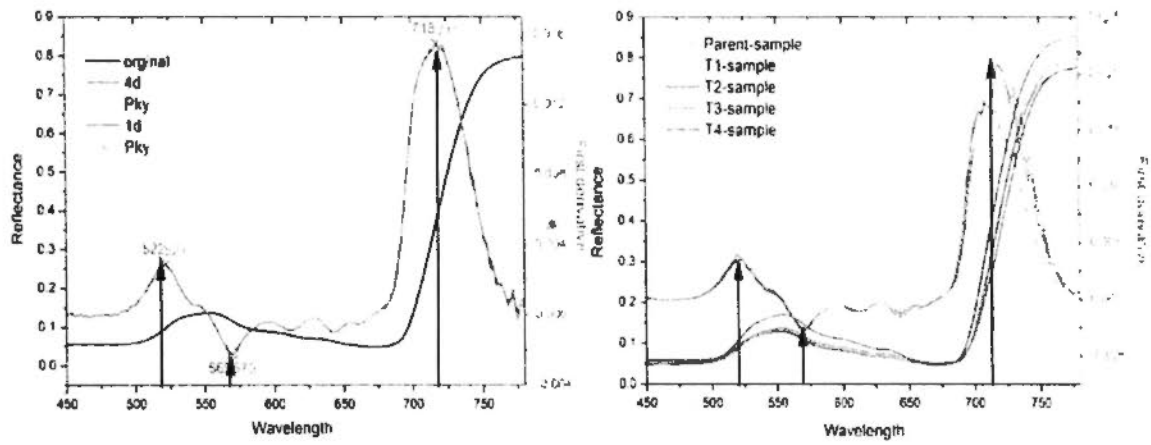
has been absorbed by leaf is, the more the reflection has been enhanced. Considering the spectral characteristics of leaf, it was not a linear relationship between reflection enhancement and spectral bands. It amplified the reflectance with different intensity among bands at a physical level with little information loss which is just what we want to.

5.1.2 Calculation of the edges

Most of the edges positions were defined as the largest changing rate of the curve at a certain bands, namely the maximum of the first derivative.

Edge position, such as red edge, blue edge, is generally defined as the largest changing rate (inflection) of the curve at a certain wavelength range, namely, the wavelength (position) where the maximum of the first derivative of the reflectance is. Because the derivative is a kind of changing rate, even a very small noise might cause a big peak in derivative curve. So 4-step derivative equation was used to calculate the first derivative (equation 1). The step in derivative equation controls effects of data smooth. The larger it is, the smoother the derivate curve is. Large step would bring side effect that is the movement of peak position. Blue curve in Figure 3a showed the derivative curve with 4-step while the red one with 1-step. It could be found that there were peaks positions movements with 1nm for 4-steps curve, comparing to the 1-step result. It was acceptable or could be treat as 1 nm accuracy about edge positions. After all, this strategy helps to suppress pseudo peaks efficiently. Figure 3b showed smooth derivative curves, and the different samples' edge positions had slight differences. This result indicated differences in samples which respond to the changes within samples or photosynthetic ability when the external influences were excluded.

$$\frac{dR(\lambda_i)}{d\lambda} = \frac{\left(\sum_{j=i+1}^{i+3} R(\lambda_j) - \sum_{j=i-3}^{i-1} R(\lambda_j)\right)/3}{4\Delta\lambda} \qquad (5\text{-}1\text{-}1)$$



a b

Figure 5-1-3. Sample spectra and their first derivative curves.

In the left figure, blue line showed the derivative curve calculated with 4-step (equation4), and the red one with 1-step. Pky meant the peak point of the curve.

In this section, the difference or growth status at group level was focused. Within a certain inner clustering (to make sure enough representation of the same sample group), we calculated the mean spectra for the same part of leaves in the same groups which could be treated as representative of the class spectra. Moreover, the coefficient of variation (cv) of data by band, here named Inner-clustering Coefficient for understanding easily, had been calculated to assess the stability and inner clustering of mean spectrum of one class. Using mean spectra as input data was to find differences at class level, thus it was necessary to avoid interferences from the specific random differences in individual which would make the study complicated. By mean spectrum, the random diversity which could not be expressed in the class stably in individual cultivator could be weakened even filtered.

Inner-clustering Coefficient (Inner-cc, cv) is defined as the ratio of standard variation of reflectance to its mean of the identical part of the same class samples under same observing conditions (equation 2). It represents the stability and inner clustering of mean spectrum of the class.

$$Inner-cc(i) = \frac{Stdv(R_{i,j})}{Mean(R_{i,j})} \quad (5\text{-}1\text{-}2)$$

$R_{i,j}$ is the reflectance of the j^{th} observation for the i^{th} band, or the reflectance of the i^{th} individual sample in the j^{th} class. Stdv represents standard variation. The smaller the value of it is, the better the stability and clustering of mean spectrum is, and the more reliable mean spectrum as the representation of the class is. To a set of data, mean and standard variation of inner-cc could be used to describe data acceptance, namely inner-cc(mean(cc), stdv(cc)) where mean and stdv represented the mean and standard variation of the Inner-cc. According to experience, when Inner-cc(mean, stdv)<(0.2,15%), there was enough inner clustering and stability of mean spectrum, and its error was acceptable, it could be used as the representation as the class. Moreover, this parameter also could use to assess data quality during original spectrum data pre-treatment and now available value range for Inner-cc was an empirical one, more precise data would be discussed later. Table1 showed that the maximum of inner-cc is T1's, it was still acceptable, and we could treat mean filtered spectra as the class representation.

Table 5-1-1. Inner-cc of sample groups (classes)

sample	Inner-cc @ [450,780nm]		
	max	min	Inner-cc(mean(cc),stdv(cc))
parent	0.225888	6.71E-05	(0.10, 6%)

T1	0.238021	0.041218	(0.13, 7%)
T2	0.158381	0.005211	(0.09, 4%)
T3	0.066849	0.001715	(0.05, 2%)
T4	0.193529	0.021221	(0.10, 4%)

5.1.3 Samples and methodology

5.1.3.1 Samples

5 samples groups' (classes) spectra had been collected in late September in 2009. These samples were cultivated by China National Hybrid Rice Research and Development Center and grouped: Parent, T1, T2, T3 and T4. Except for the parent, the others had been transferred into different gene unit of phycocyanin genome, and by laboratory validation the genes have been transferred into successfully. This genome was predicted to promote the receptor's photosynthetic efficiency and to produce rice of high quality which means the transferred gene would cause the difference of spectral features between these samples. All samples were planted under a strictly contrast environment to exclude external noise and ensure data's reliability. The samples in first experiment were all at maturity stage, slight yellow-green color but green mainly, few ones were yellow, and most samples grew exuberantly. Since the target genes were predicted to mainly affect the plant photosynthetic system, spectral data of the middle front of leaves were chosen which were actively responsive part to photosynthesis in photosynthetic bands [450nm, 780nm] (Blackburn, G. A., 2007, Gitelson and Merzlyak, 1996).

5.1.3.2 Index selection

Gitelson et al. (1996) found the ratio of reflectance at 750 nm to 700 nm (R750/R700) was directly proportional to chlorophyll concentration. Datt (1998) found that the index R672/(R550 × R708) was the best indicator of chlorophyll a, chlorophyll a+ b, and total carotenoids contents, and R672/R550 was the best indicator of chlorophyll b. There are also many other ones, such as the Chlorophyll Absorption in Reflectance Index (CARI) (Daughtry, C. S. T., et al., 2000). However, in this section, the edge/ peak parameters (Figure 3) sensitive to photosynthesis were focused. These parameters chosen have good correlations to photosynthetic pigments and nitrogen components (K, N and P) no matter on experiences or theories analysis (Nagendra, 2001, Stamps, et al., 1987, Tang, 2004). They cover the whole photosynthesis bands, and are comprehensive parameters to describe photosynthesis of leaves. Besides those above, Photochemical Reflectance Index (PRI) (Penuelas, et al., 1997) and Simple Ratio PRI (SI-PRI) (Wu, et al., 2008) were given too.

The Red edge is around 680- 760 nm, while the Red absorption around 650- 690nm, the Yellow edge around 560- 640 nm, the Blue edge around 490- 530 nm, and the Green peak around 510-560 nm(Filella and Penuelas, 1994, Gitelson, et al., 1999, Gong, et al., 2002, Horler, et al., 1983) All of them include parameters of the edge position, the reflectance there and area. Carter et al. (1992) hypothesized that the increase of reflectance in photosynthetic bands might be a result of decreased chlorophyll content. This hypotension was proved later (Bauerle, et al., 2004, Carter, G. A. and Knapp, A. K., 2001). In this study, the area of edge was defined as the sum of reflectance included in edge ranges, different to the sum of first derivative of spectrum in edge (Filella and Penuelas, 1994). The parameter area represents a comprehensive absorption and reflection process in photosynthetic sensitive bands. It is much easier to understand the relation between area and photosynthesis.

For comparison, some spectral indices were also chosen which were related to specific photosynthetic pigments (chlorophyll a+b, carotenoids and anthocyanin) respectively for cross-validation (table 2, 3 and 4). Most of these parameters were assessed in chapter 4 and proved to have good performance as the foliar chemicals.

Table 5-1-2. Spectral indices developed as chlorophyll (a+b) indicators (Bannari, et al., 2007b, Gitelson, A. A., et al., 2003):

Indices Description	source
Chlorophyll (a+b)=R675/R700	(Chappelle, E. W., et al., 1992)
Chlorophyll (a+b)= (R800–R700)/(R800+R700)	(Gitelson, et al., 1994)
Chlorophyll (a+b)=0.0236*[R672/(R550*R708)]$^{0.7954}$	(Datt, B., 1998)
Chlorophyll (a+b)= ($R_{750-800}$)/($R_{695-740}$)-1	(Gitelson, A. A., et al., 2003)
Structure Insensitive Pigment Index (SIPI) SIPI =(R800-R445)/(R800-R680)	(Bannari, et al., 2007b, Penuelas, et al., 1995)
Chlorophyll Absorption in the Reflectance Index (CARI) CARI =(R700-R670)-0.2*(R700-R550)	(Kim, et al., 1994a)
Modified CARI (mCARI) mCARI =[(R700-R670)-0.2*(R700-R550)]*(R700/R670)	(Daughtry, C. S. T., et al., 2000)
Transformed Chlorophyll Absorption in the Reflectance Index (TCARI) TCARI =3*[(R700-R670)-0.2*(R700-R550)*(R700/R670)]	(Haboudane, D., et al., 2002)
Simple Ratio Pigment Index (SRPI) SRPI =R430/R680	(Bannari, et al., 2007b, Blackburn, G. A., 1999)

$R_{750-800}$= Average(Sum(Reflectance(750:800)));

RNIR=Average(Sum(Reflectance(700:750)));

R_{red}=Average(Sum(Reflectance(650:690))), the red absorption feature;

R_{Green}=Average(Sum(Reflectance(510:560))) the green peak feature;

R^{-1}_{510} is the reciprocal of reflectance at 510nm;

$R^{-1}_{510\ 520}$ is the reciprocal of the average reflectance ranging from 510nm to 520nm;

So do the other abbreviations in the follow tables

Table 5-1-3. Spectral indices developed as carotenoids indicators (Ustin, et al., 2009):

Indices Description	Source
Ratio Analysis of Reflectance Spectra (RARS)	(Blackburn, G. A , 1998,
RARS=R760/R500	Chappelle, E. W., et al., 1992))
Structure Insensitive Pigment Index (SIPI)	(Penuelas, et al., 1995, Sims, D.
SIPI =(R800-R445)/(R800-R680)	A. and Gamon, J. A , 2002)
Carotenoid Reflectance Index (CRI)	(Gitelson, A A., et al , 2002)
$CRI550=R^{-1}_{510}-R^{-1}_{550}$	
$CRI700=R^{-1}_{510}-R^{-1}_{700}$	
Modified CRI (mCRI)	(Gitelson, A. A., et al., 2006)
$mCRIgreen=(R^{-1}_{510\ 520}-R^{-1}_{560\ 570})\times R_{NIR}$	
$mCRIredge=(R^{-1}_{510\ 520}-R^{-1}_{690\ 710})\times R_{NIR}$	
Eucalyptus Pigment Indices (EPI)	(Datt, B., 1998)
$EPI=0.0049*[R672/(R550\times R708)]^{0.7488}$	

Table 5-1-4. Spectral indices developed as anthocyanin indicators (Ustin, et al., 2009):

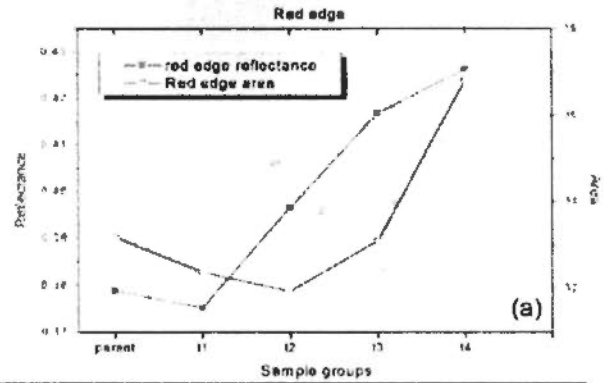
Indices Description	source
Anthocyanin Reflectance Index	(Gitelson, A. A., et al., 2001)
$ARI=R^{-1}_{550}-R^{-1}_{700}$	
Modified Anthocyanin	(Gitelson, A A., et al., 2006)
$mARI=(R^{-1}_{530\ 570}-R^{-1}_{690\ 710})\times R_{NIR}$	
Red:Green Ratio	(Gamon, J. A., et al , 1999, Sims,
$RGR=R_{red}/R_{green}$	D. A. and Gamon, J. A., 2002)

5.1.4 Results and discussion

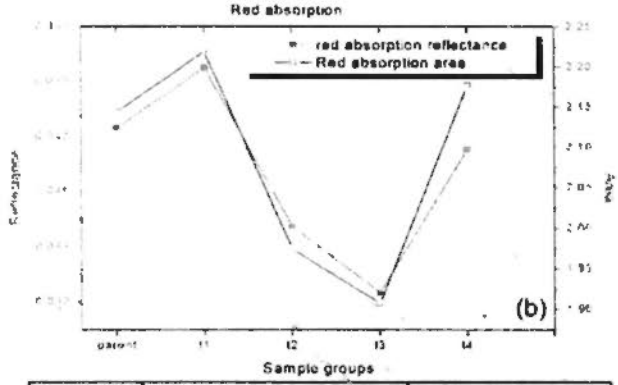
Figure 5-1-4. Positions of the edges.

Figure 4 showed that the position of red edge had a relatively bigger change than others, the biggest was up to 3nm at class level. Except T4, the other samples' red edge moved towards to long-wave band. The Red edge highly related to chlorophyll content (Blackburn, G. A., 2007, Cho, et al., 2006, Sims, D. A. and Gamon, J. A., 2002) would move to long or short wavebands direction according to different change of chlorophyll in plants, biomass, and phenology regulations (Curran, et al., 1990, Filella and Penuelas, 1994). Chlorophyll content had a direct positive relationship to photosynthetic capacity (Murchie, et al., 1997b). Therefore, this result indicated the samples except T4 had more strongly photosynthetic mechanism compared to their parents. The positions of red absorption, yellow edge, blue edge and green peak were stable, largest movement was no more than 1nm compared to sample parents'. At yellow edge, all samples with trans-genes moved to short-wave band together. The same things happened at the green peak too, and the amplitude was up to 1 nm. The Yellow edge is an indicator of representing information of xanthophyl concentration. Xanthophyl is highly related to photosynthesis (Gamon, et al., 1992). When plant is in health with high content of chlorophyll in the period of growth activity, the Green peak would move to blue band direction, and its amplitude would reduce (Gitelson and Merzlyak, 1996, Gitelson, et al., 1999). Gitelson (1996) indicated further that the band near 550nm (around green peak) was determined by total carotenoids, Chl a and Chl b. Near 550nm, the two strong absorption (around 520 and 570) processes (blue and yellow edges) reached their minimum, producing the monotonous relationship with a high sensitivity to Chl a concentration. Combining the changes of reflectance shown following at green peak, it could be deduced that higher content of chlorophyll in sample plant comparing to their counterparts. It was noticed that all these changes/ movements happened at class level. Moreover, in figure4, the

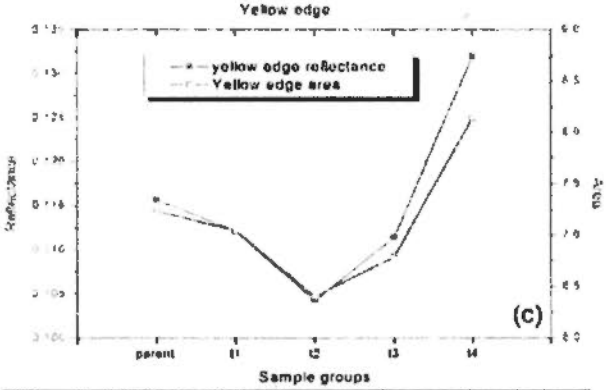
features of T4 had an inconsistent tendency of changes to the other transgenic sample groups.



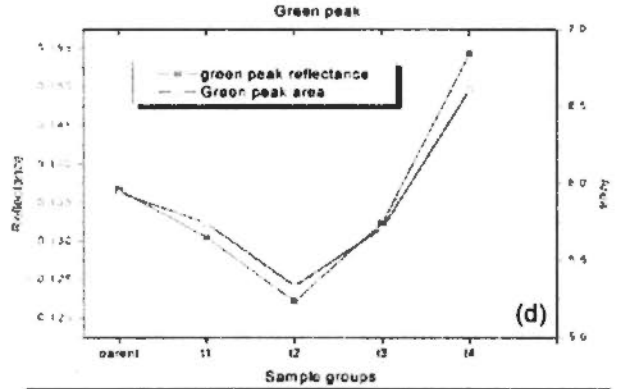
	red edge reflectance	Red edge area
Parent	0.37883	33.18908
T1	0.37520(1%)	32.37767(2%)
T2	0.39662(5%)	31.94781(4%)
T3	0.41686(10%)	33.09989(0.3%)
T4	0.42636(13%)	36.81535(11%)



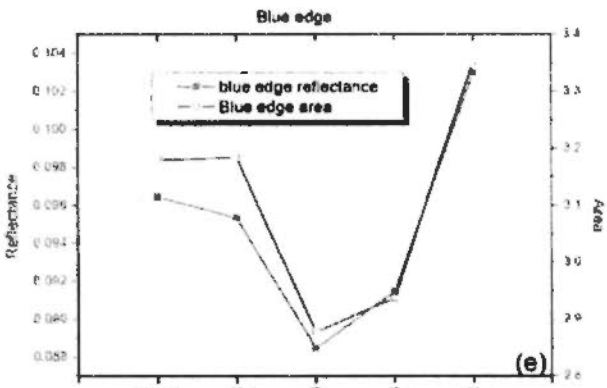
	red absorption reflectance	Red absorption area
Parent	0.04831	2.14394
T1	0.05049(5%)	2.21903(4%)
T2	0.04473(7%)	1.97426(8%)
T3	0.04234(12%)	1.90766(11%)
T4	0.04752(2%)	2.17660(2%)



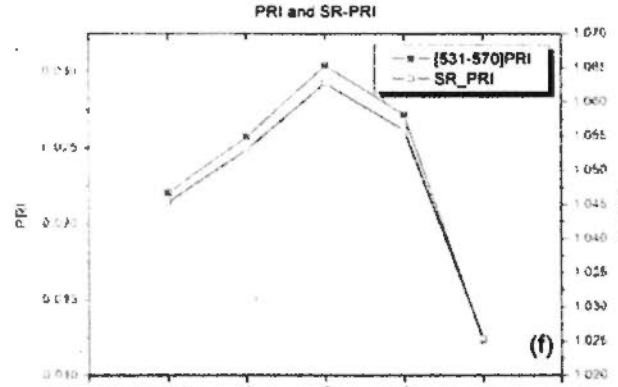
	yellow edge reflectance	Yellow edge area
Parent	0.11569	7.23354
T1	0.11211(3%)	7.04328(3%)
T2	0.10428(10%)	6.40071(12%)
T3	0.11133(4%)	6.78647(6%)
T4	0.13199(14%)	8.13683(12%)



	green peak reflectance	Green peak area
Parent	0.13669	5.94115
T1	0.13048(5%)	5.74127(3%)
T2	0.12221(11%)	5.33551(10%)
T3	0.13236(3%)	5.71177(4%)
T4	0.15437(13%)	6.60884(11%)



	blue edge reflectance	Blue edge area
parent	0.09644	3.17928
T1	0.09534(1%)	3.18458(0.2%)
T2	0.08847(8%)	2.87749(9%)
T3	0.09145(5%)	2.93705(8%)
T4	0.10297(7%)	3.35162(5%)



	[531-570]PRI	SR_PRI
Parent	0.02202	1.04534
T1	0.02570(17%)	1.05300(1%)
T2	0.03041(38%)	1.06273(2%)
T3	0.02716(23%)	1.05589(1%)
T4	0.01225(44%)	1.02540(2%)

Figure 5-1-5. Comparison of spectral parameters between Parent, T1, T2, T3, and T4

Reflectance and area of edges and PRI results, below the figures are the data tables showing the details of difference between samples and their parents. In the table the change percent of parameters were also calculated in bracket.

Figure 5(a- e) showed information of the reflectance and area of edges. These two parameters for the samples had consistent tendency except at red edge, when the reflectance of edges increases, the area of edges we defined increases synchronously. The areas' changes indicated relevant information about photosynthetic intensity of green plant. It was comprehensive description of ability of incident light energy absorption. In a certain band range (e.g. red edge, the same as others), when there was high level of chlorophyll content (responding to a strong photosynthetic process in the same plant species in some sense, as mentioned above) (Murchie and Horton, 1997b), more light energy toward long-wave bands would be absorbed by chlorophylls and less reflected, it made the edge moved to long-wave bands. Thus, it would cause lower reflectance and edge movement at these blocks in a certain energy intensity range. So excluding external noise, the higher the chlorophyll content was, the lower the reflectance and area of edges (reflectance of red edge foreclosed) was. At the red edge, because there was a high reflected shoulder, thus in this range, the reflectance value would change to be greater when moving towards long-wave bands. In the figure 4 (a) at the red edge, the differences between samples showed no clearly regulation of changes. The abnormal samples were located at T1 and T2. Red edge of samples except T1 moved to long-wave bands, according to greater reflectance value compared to sample parent according to the special characteristics of vegetation reflectance at near-infrared bands. Reflectance of sample T3 was also greater than its parent, but its area of edge was smaller than parent's which was not consistent to other samples. Sample T4 showed biggest discrimination both to the other samples and its parent. At the red, yellow and blue edges, the changes of reflectance were up to 13% even to 14%, but these changes had no regulations, some were positive and the others were negative. Compared to T4, except at red edge, sample T2 had a stable direction and amplitude of changes. Comparing to red edge, including the red absorption, the samples had stable regulations of change. And the changes of reflectance and area were consistent there. From figure 5(a- e), we could have a qualitative sort about the chlorophyll content between samples, that was T2, T1, T3, parent and T4 by descending generally. Meanwhile, parameter area was more stable than reflectance to indicate this difference because of resisting random noise.

Figure 5(f) showed PRI and SR-PRI of samples. SI-PRI is a revised one of PRI which has clearer physical meaning. These two parameters closely positively related to xanthophyll cycle pigment content, and could be used to estimate leaf photosynthetic light use efficiency (LUE) (Penuelas, et al., 1997). From these two indicators, it could be found the slight change of anthophyll cycle

pigment content. And PRI seems to be much more sensitive to the difference between samples than SR-PRI. From the figure 5 (f), it could be found that sample T2 had greatest PRI than others, then T3, T1, parent and T4. The curves shapes of PRI and SR-PRI were the reverse to the ones of edges fitly. Also sample T4 had unique features different from the other transgenic ones. It was consistent with the result which obtained from the analysis of the samples edges. In some sense, there were negative influences on sample T4. After these analyses, we finished one time screening. These results would be useful for breeding and assess samples' characteristics.

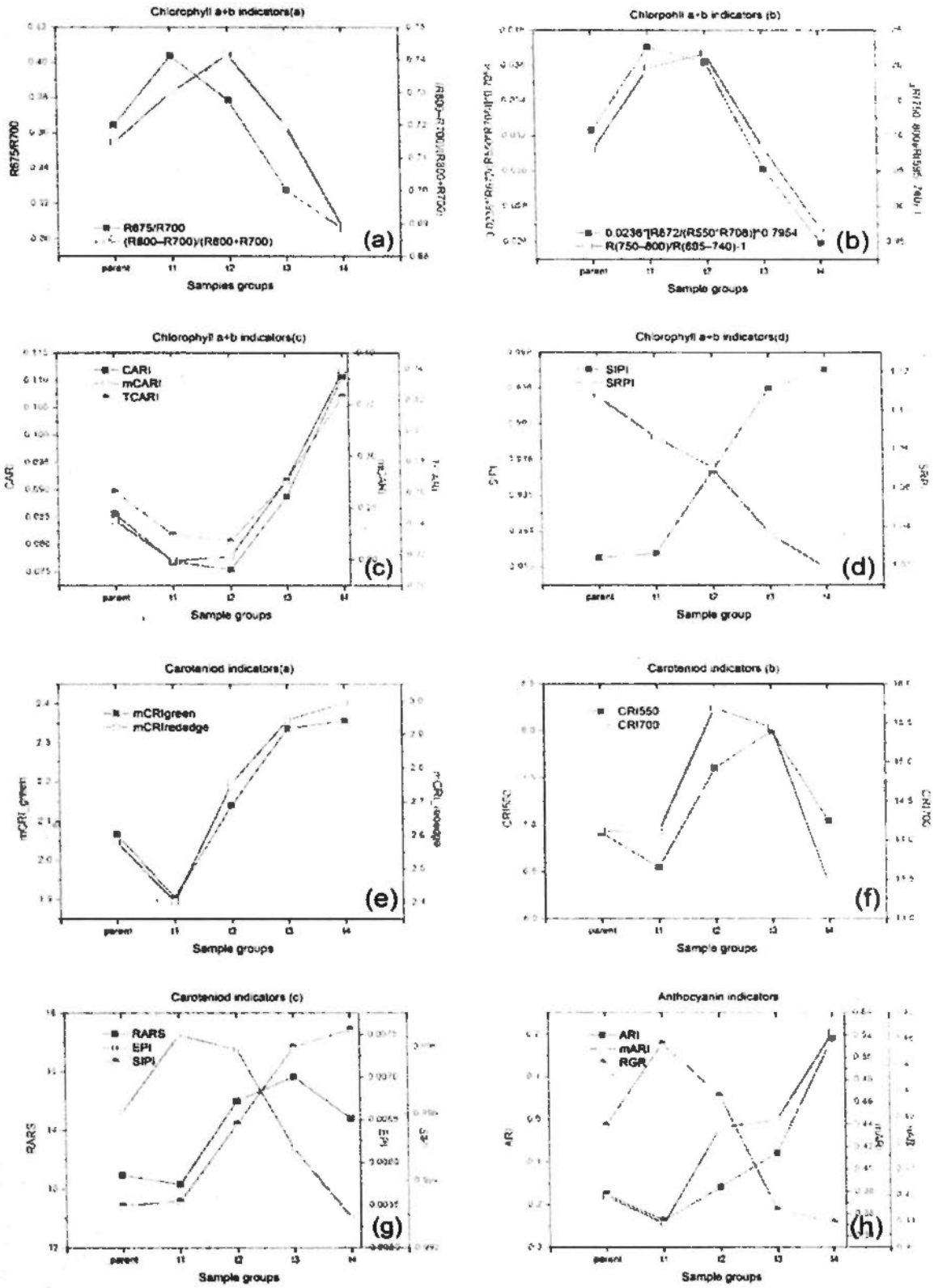


Figure 5-1-6. Spectral indicators for specific pigments.

The information shown in (a- h) was consistent with that in figure 4 and 5.

Because all samples were cultivated in the completed same conditions, and the data were fast collected under identical situation, thus, external noise (including photo-inhibitory and photo-protective response) could be excluded and the changes or differences of the parameters

given above could be treated as the indicators of different photosynthesis ability which would indicate much possibly the inner changes of sample plants organs for photosynthesis. Between samples, there was only one difference was the existence of transgenes, namely other all properties or characteristics of the samples should be the same generally in the strictly controlled contrast experiment. Thus, if there was any difference or change about photosynthesis in samples, it was possibly made by phycocyanin genes and of significance to be validated by laboratory approaches. By current analysis, we could have some assumptive interpretation that the correct expression of phycocyanin genome enhance the light-harvesting system of rice plant which caused positive effect to photosynthesis, namely in the photosynthetic pigment light-harvesting system which was composed of chlorophyll, it was generated a new non-complete approach which were supported by phycocyanin. Phycocyanin had different characteristics to other chromoprotein. The interaction of these two approaches changed the plant ability of photosynthesis then affected its spectral characteristics. The negative influences caused by trans-genes seemed to also affect the plant the light-harvesting system and then caused difference. Though this interpretation lacked relevant support from laboratory, it at least provides messages as prior knowledge for laboratory approaches for further study. ,

5.1.5 Conclusion

In this study, fine spectra collected by ASD field spectrometer were used, and by quantitative analysis, differences (at class' level, not at individual') between samples with trans-gene and their counterpart at spectral level had been found.

(1) The differences at all edges and absorption peak chosen in certain band range. It indicated that there was some matters bring changes to the transgenic samples stably. These parameters had high relationship with kinds of photosynthetic pigments, thus it could be deduced the differences of pigments content in samples. This information could be used to assess the photosynthetic ability of samples. (2) The discovered differences between samples comparing to their parent, some were positive and the other were negative to photosynthesis. (3) From sample T4, it could be found that the factor need to further study would also influence other organism not only photosynthetic ones. In water sensitive bands T4 had higher reflectance than all other ones. The approach and the report helped transgenic species researchers to avoid amount of workload, monitor their cultivars in time and under control.

From all these results, it was deduced that it was significant to study on the monitoring and screening transgenic paddy rice by field hyperspectral techniques, especially in transgenic plant breeding. This approach could give a cheap, easy operation and real- or near-real time monitoring the influence of gene transferred into to samples, and helped researchers to easily control and maintain their long period experiment and was promising to transgenic plant breeding and other

relevant study.

Because phycocyanin genome mainly possibly influences plant photosynthetic system, in this section just photosynthetic sensitive bands were focused. However, it is still unknown whether the genome would affect other organs is unknown in fact. Thus, it is necessary and important to extend the study to the whole bands (350- 2500nm). Also, we would collect biophysical data of plant in future work synchronously, to build a quantitative relationship between spectral parameters changes and photosynthesis of plant.

Reference

Bannari, A., Khurshid, K. S., Staenz, K., & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *Ieee Transactions on Geoscience and Remote Sensing*, 45(10), 3063-3074.

Bauerle, W. L., Weston, D. J., Bowden, J. D., Dudley, J. B., & Toler, J. E. (2004). Leaf absorptance of photosynthetically active radiation in relation to chlorophyll meter estimates among woody plant species. *Scientia Horticulturae*, 101(1-2), 169-178.

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.

Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4), 855-867.

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684.

Carter, G. A., Mitchell, R. J., Chappelka, A. H., & Brewer, C. H. (1992). Response of leaf spectral reflectance in loblolly-pine to increased atmospheric ozone and precipitation acidity. *Journal of Experimental Botany*, 43(249), 577-584.

Chappelle, E. W., Kim, M. S., & Mcmurtrey, J. E. (1992). Ratio analysis of reflectance spectra (rars) - an algorithm for the remote estimation of the concentrations of chlorophyll-a, chlorophyll-b, and carotenoids in soybean leaves. *Remote Sensing of Environment*, 39(3), 239-247.

Cho, M. A., & Skidmore, A. K. (2006). A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, 101(2), 181-193.

Curran, P., Dungan, J., & Gholz, H. (1990). Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology*, 7(1), 33-48.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111-121.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229-239.

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content,

biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459-1470

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote sensing of environment*, 41, 35-44

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, 143(1), 105-117

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271-282.

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Gitelson, A. A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra : experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B-Biology*, 22(3), 247-252.

Gitelson, A. A., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148(3-4), 494-500

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74(1), 38-45.

Gitelson, A. A., Schalles, J. F., Rundquist, D. C., Schiebe, F. R., & Yacobi, Y. Z. (1999). Comparative reflectance properties of algal cultures with manipulated densities. *Journal of Applied Phycology*, 11(4), 345-354.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3), 272-281.

Gong, P., Pu, R., & Heald, R. (2002). Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia *International Journal of Remote Sensing*, 23(9), 1827-1850.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture *Remote Sensing of Environment*, 81(2-3), 416-426

Horler, D., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4, 273-288.

Kim, M., Daughtry, C., Chappelle, E., McMurtrey III, J., & Walthall, C. (1994). The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation. Proceedings of the ISPRS Sixth International Colloquium on Physical Measurements and Signatures in Remote Sensing; 1994 17- 21 January 1994; Val d'Isere, France: European Space Agency; 1994, 299-306.

Murchie, E. H., & Horton, P. (1997). Acclimation of photosynthesis to irradiance and spectral quality in british plant species: Chlorophyll content, photosynthetic capacity and habitat preference. *Plant Cell and Environment*, 20(4), 438-448.

Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22(12), 2377-2400.

Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2), 221-230.

Penuelas, J., Llusia, J., Pinol, J., & Filella, I. (1997). Photochemical reflectance index and leaf photosynthetic radiation-use-efficiency assessment in mediterranean trees. *International Journal of Remote Sensing*,

18(13), 2863-2868.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2-3), 337-354.

Stamps, J., Buechner, M., & Krishnan, V. (1987). The effects of edge permeability and habitat geometry on emigration from patches of habitat. *The American Naturalist*, 129(4), 533-552.

Tang, Y. (2004). Study on the hyperspectral characteristics and simulating and estimating models about biophysical and biochemical parameters of rice. 35~45.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009) Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, 113(S67-577.

Wu, C., Niu, Z., & Tang, Q. (2008). Sensitivity study of a revised leaf photochemical reflectance index (pri) *Spectroscopy and Spectral Analysis*, 28(9), 2014.

5.2 Comparison analysis based on spectral shape and parameterized features among Sample groups --a case of the gene repeat experiment[22]

In this section, the same foreign genes of phycocyanin genome (chapter 5.1) were transferred into paddy rice samples. These samples were calculated in the other field block in the other season of the after year (2010). This is principal requirement of breeding, the experiment-repeat experiments, to prove the growth stability of samples.

5.2.1 Experiment design

The procedure of gene expression is complex, and its influences on receptor is of stochasticity (Kaern, et al., 2005), diversity and variability (Raser and O'Shea, 2005). Therefore, in this study the expression of gene and its influences on receptor were not explored directly. biophysical traits and responses to the stress (taking influences of foreign gene as stress) were focused on by detecting spectral differences among samples. By analyzing these differences, the samples would be monitored real-/ near real-time, interested space of samples was optimized and suppressed and finally priori knowledge and support information for laboratory work was extracted. By this study, an operational and efficient approach based on hyperspectral remote sensing technique to assist transgenic crop breeding with large number of samples would be developed, revised, validated further.

Figure 1 showed the guideline of the study. All samples should be cultivated in contrast conditions which should satisfy the demands of crop breeding. Based on fine spectra of leave,

[22] This section had been published in the Journal of Spectroscopy and Spectral Analysis (2011, 31(6)).

indicative parameters would be obtained. Then with help of these parameters, it could be obtained useful information that if any spectral differences or outliers (of spectral morphology parameter) among samples (transgenic ones and their counterpart) exist; if any, they would be assessed quantitatively and located where the responding bands were and interpreted what caused to them. Finally a report would be formed to the laboratory for validation and further study. By this proposed approach, spectral monitoring, laboratory study and field cultivation could give feedback to each others to make crop breeding efficient with low cost.

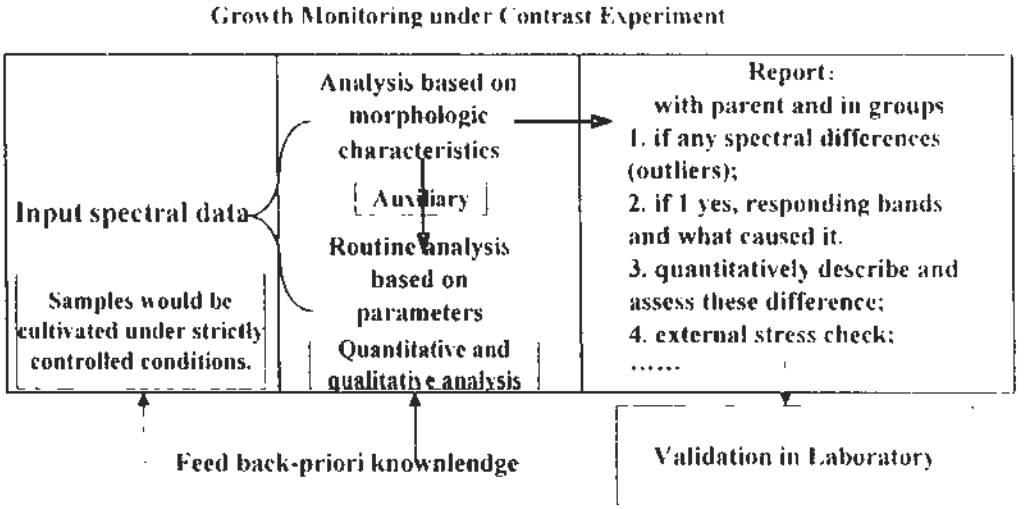


Figure 5-2-1. General Design of the Experiment

5.2.2 Samples and data acquisition

5.2.2.1 Samples

In this section, all samples, paddy rice, were cultivated by China National Hybrid Rice R&D Center/ Hunan Hybrid Rice Research Center in Changsha Hunan province. Fine spectra of foliage were acquired around 11 am Aug. 4, 2010. During this time range, samples were active and had stable biophysical and biochemical processes relatively. Except for the contrast samples, the others had been transferred into different gene unit of phycocyanin genome and by laboratory validation the genes have been transferred successfully. This genome was predicted to promote the receptor's photosynthetic efficiency and help to produce rice of high quality. The researchers wanted by cultivating these samples to find out influences of single gene unit in the genome on the plant, then to screen superior sample for next step breeding. All samples were planted satisfying the demanding of crop breeding, namely cultivated as block in a very small field to ensure they grow in a consistent condition of water, temperature, management to exclude external noise and ensure data's reliability. The samples grew exuberantly in stage of yellow ripeness when data collected. Totally spectra data of 9 groups were obtained. Transgenic groups

were assigned a number T1, T2... T8 and their contrast one was the Parent. Except for the genes transferred, there was no other difference among these samples theoretically.

5.2.2.2 Spectral data acquisition

In this section, the same devices were employed. However the white reference panel was replaced by leaf-clip (figure 1). Leaf-clip has both white and black reference panels. In this study the black one was used, thus there was no effect caused by higher reflectance shoulder.

Figure 5-2-2. Observation with probe-leaf-clip system

For eliminating uncertainties caused by data collection, the center of FOV located at the center of the middle front of the second leaf counted from the core of paddy rice. The principal vein of leaf was vertical to the view line from the fiber. Two measurements were taken at the leaf with micro moving along the principal vein. By this measurement, the consistency of data collected could be ensured. Considering the characteristics of the genes transferred (predicted to mainly affect the plant photosynthetic system) and the noise sensitive bands of the equipment, in this study spectral data at band range from 420- 800nm was chosen.

5.2.3 Methodology

5.2.3.1 Spectral angle between groups

In the two dimensional space defined by wavelength X and wavelength Y, the two vectors constituted by spectral data t and r, thus the spectral angle θ of t and r could be defined by equation 1(Sohn, et al., 2002). In this space, the spectral angle could offer the information about the similarity of the two vectors: the more similar the two vectors are, the smaller the angle is. When the angle equals to 0, the two vectors are almost the same.

$$\theta_{t,r} = \cos^{-1}\left(\frac{t \cdot r}{\|t\| \cdot \|r\|}\right) \qquad \theta \in \left[0, \frac{\pi}{2}\right] \quad (5\text{-}2\text{-}1)$$

5.2.3.2 Continuum removal spectra

Continuum removal is a technique for normalization according to the characteristics of the data essentially (Mutanga, et al., 2005). By this approach, the data (vector) is projected into a space of [0, 1] without unit. Thus the data could be compared both by spectral values and shape directly. Also by this transform, the spectral feature could be amplified.

5.2.3.3 Spectral indices as indicators of chlorophyll and carotenoids

Because the genes were highly related to photosynthesis of paddy rice, thus several advanced indices were chosen as indicators for the differences between the samples. These indices were proved highly correlated to photosynthetic pigments content (chlorophyll a and b and carotenoids) which is highly related to photosynthetic capacity of plant(MURCHIE, et al., 1997a). Furthermore, the changes caused by internal or external stresses to the plant could be responded by the pigments content and indicated by these indices. In this study, 10 indices had been chosen to indicate the content of chlorophyll a+ b and carotenoids.

Table 5-2-1. Spectral indices developed as chlorophyll (a+b) indicators(Bannari, et al., 2007a, Gitelson, AA, et al., 2003):

Spectral Indices	Authors
R67=R675/R700	(Chappelle, EW, et al., 1992)
datt=0.0236*[R672/(R550*R708)]$^{0.7954}$	(Datt, B, 1998)
git= $(R_{750-800})/(R_{695-740})$-1	(Gitelson, AA, et al., 2003)
Chlorophyll Absorption in the Reflectance Index (CARI) CARI =(R700-R670)-0.2*(R700-R550)	(Kim, et al., 1994b)
Transformed Chlorophyll Absorption in the Reflectance Index (TCARI) TCARI =3*[(R700-R670)-0.2*(R700-R550)*(R700/R670)]	(Haboudane, D, et al., 2002)

$R_{750-800}$= Average(Sum(Reflectance(750:800))),

R_{NIR}=Average(Sum(Reflectance(700:750))).

R^{-1}_{510} is the reciprocal of reflectance at 510nm.

$R^{-1}_{510\ 520}$ is the reciprocal of the average reflectance ranging from 510nm to 520nm.

The same as others.

Table 5-2-2. Spectral indices developed as carotenoids indicators:

Spectral Indices	Authors

Photochemical reflectance index PRI=(R531-R570)/(R531+R570)	(Gamon, et al., 1992)
Carotenoid Reflectance Index (CRI) $CRI700=R^{-1}_{510}-R^{-1}_{700}$	(Gitelson, AA, et al., 2002)
Chlorophyll Absorption Ration Index CARI2=CAR(R700/R670), CAR=|(a*670+R670+b)|/(a^2+1)^0.5 a=(R700-R500)/150, b=R550-(a*550)	(Broge, NH and Leblanc, E., 2001)
Modified CRI (mCRI) $mCRIgreen=(R^{-1}_{510\cdot 520}-R^{-1}_{560-570})\times R_{NIR}$ $mCRIredge=(R^{-1}_{510\cdot 520}-R^{-1}_{690-710})\times R_{NIR}$	(Gitelson, AA, et al., 2006)

5.2.4 Result and discussion

5.2.4.1 Analysis on spectral morphological features

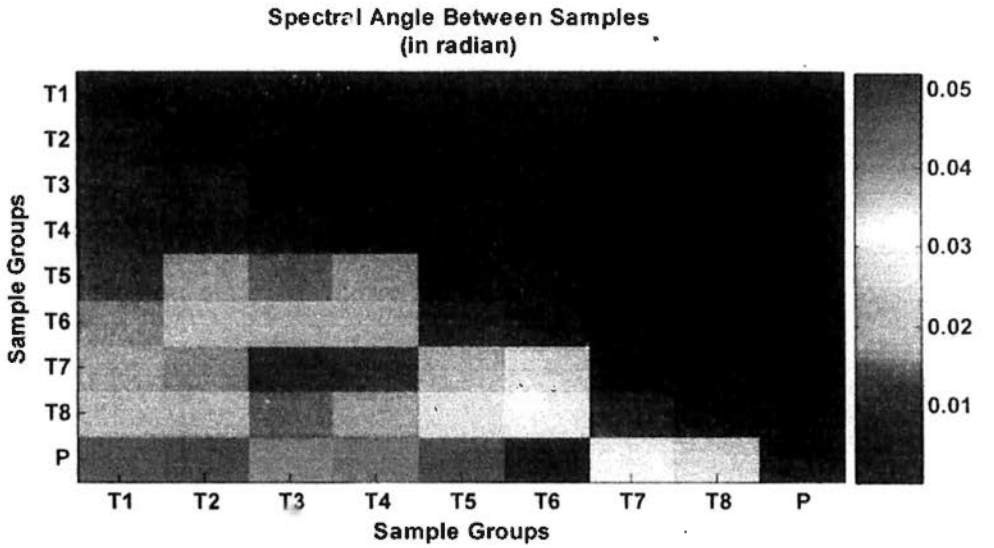


Figure 5-2-3. Spectral Angles of the samples

In figure 3, the spectral angles of each group with the other ones ware calculated. The result showed that, except for transgenic group T7 and T8, the left all had relative big angles with their Parent. T1 had smaller spectral angles with T2, T3 and T4 respectively. It indicated that these groups were more close and similar to each other on spectrum. Meanwhile, they had big angles with the Parent revealed that the transgenic samples had difference on spectrum to the contrast. In transgenic groups, the angles between T2, T3, T4 and T6 were also a little big, but they were less than the angle with the Parent. The transgenic groups could be divided into three classes according to the spectral angles, namely T1-T4, T5-T6 and T7-T8. Considering the values of the angles, when laboratory studying, T1-T4 and T5-T6 should be paid much attention, especially to

T5-T6.

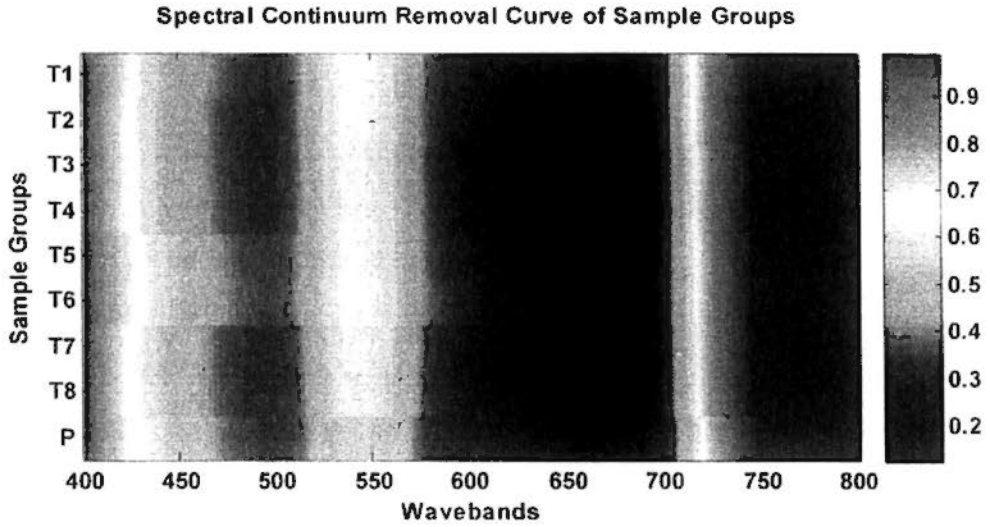


Figure 5-2-4. Spectral Continuum Removal Curve

Figure 4 showed the spectral continuum removal image. In this figure, the spectral differences could be clearly found and assessed qualitatively. By this transform, the values of reflectance data had been projected to [0, 1]. Thus the differences between groups could be shown by the shifts of color (shifts of absorption or reflectance peak poisons) and the changes of color (intensity of absorption or reflectance) at certain band. The figure revealed that the main shifts located around 450nm, 550nm and 720nm. Furthermore, at these shift ranges, the intensity of absorption or reflectance was also of significant difference. So the next step study should focus on these shifts.

5.2.4.2 Analysis of parameterized features

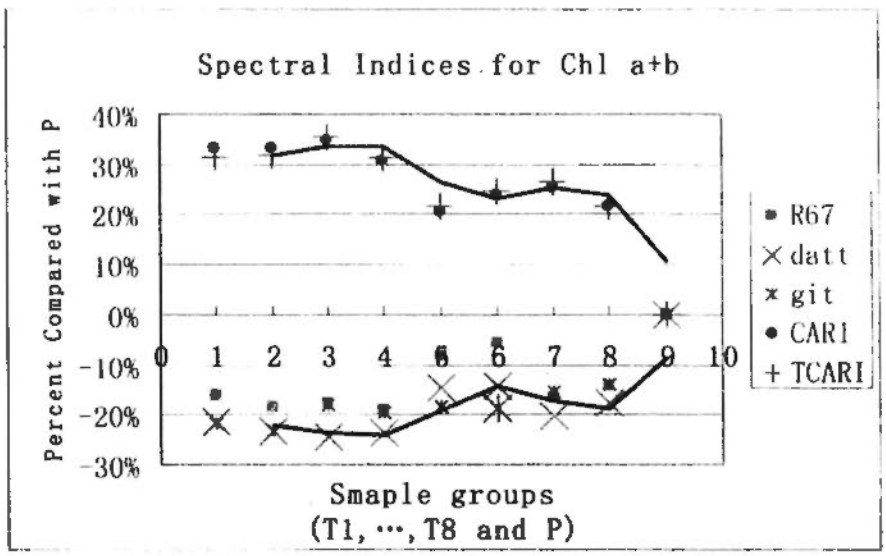


Figure 5-2-5. Spectral indices correlated to foliar Chl a+b

According to equation 3, the ratio of indices difference with the Parent (V) had been obtained and then plotted as figure 5. The figure showed that the differences of the chlorophyll a+b content of transgenic groups were all obvious with the Parent'. The V of T1, T2, T3 and T4 were more similar to each other and significant different compared with the Parent. This result accorded with the one implied by the figure 3. Compared with T7 and T8, T5 and T6 were more close to the Parent. In summary, the 5 chosen indices showed the different chlorophyll a+b content compared with the contrast group consistently, especially T1- T4, the V values were more than 30%. The largest one was T3', up to 40% while the smallest was T5' and T6', however still more than 20%. These results could be applied as auxiliary information to screen the samples.

$$V = \frac{(T(i) - P(i))}{P(i)} \times 100\% \quad (5\text{-}2\text{-}2)$$

T (i) and P(i) represent the i^{th} spectral index value of the transgenic groups and their Parent one'.

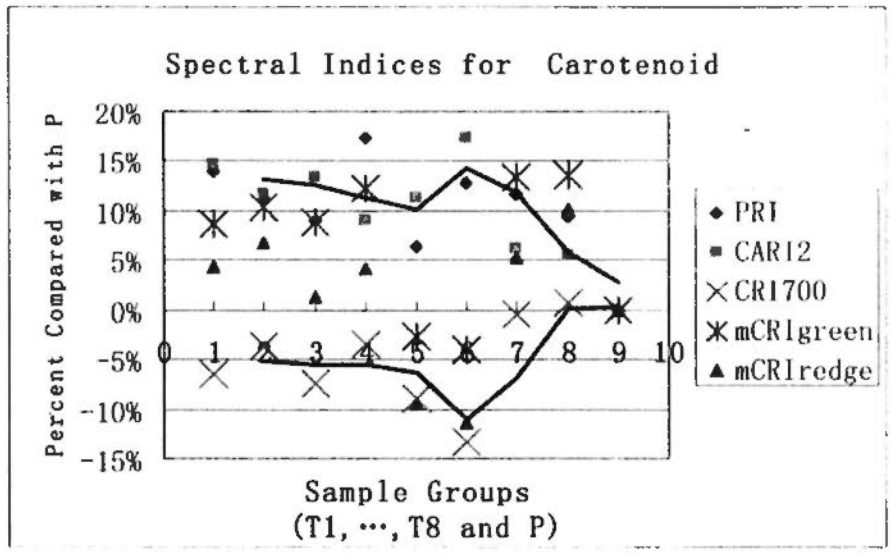


Figure 5-2-6. Spectral indices correlated to foliar Carotenoid

The data shown in figure 6 were calculated followed equation 2. The figure showed that compared with the contrast group, T1-T4 had a good consistence about the content of carotenoids, and also more close to the Parent. The significant difference was in T5 and T6. T7 and T8 was the most close to the Parent.

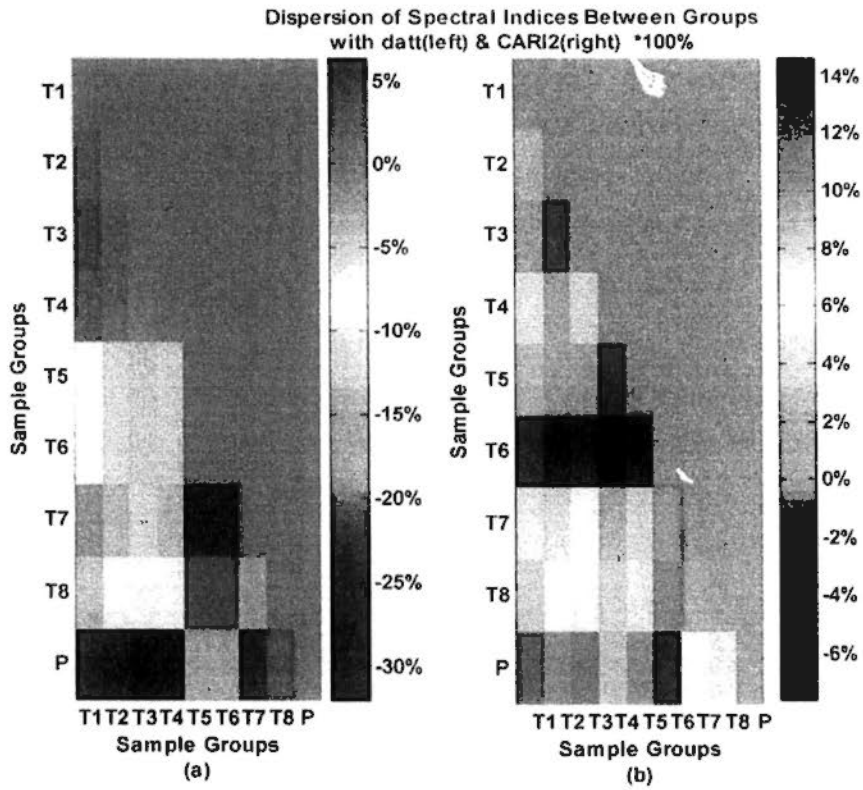


Figure 5-2-7. Dispersion of spectral indices between groups taking datt and CARI2 as examples

Figure 5 and 6 showed the results of differences between transgenic groups and their parent. The results indicated that there were also differences represented by spectral indices among transgenic groups. Therefore, the ratio of indices difference with the Parent (V, by column) had been calculated to assess the differences of pigments contents among transgenic groups which would reveal the capability of photosynthesis indirectly. V is defined as equation 3. In figure7a, T1, T7 and T8 had relative smaller differences, less than 10% than the others. T5 and T6 had a negative difference compared with other groups which accorded with the previous results. Moreover, compared with the Parent, except T5 and T6, all others had a big value, more than 20%, even up to 30%. In general, the differences in transgenic groups were smaller than those against with the Parent obviously, and a good consistency in pigments contents was observed in transgenic groups. In figure 7b, the consistencies were more significant. All these indicate that the pigments contents and the capabilities of transgenic groups existed differences compared the Parent on the chosen indices.

$$V = \frac{(T(i) - T(i+1))}{T(i)} \times 100\% \quad (5\text{-}2\text{-}3)$$

T(i) represents the value of spectral index of the transgenic group i.

5.2.5 Conclusion

In the study, I proposed to use ASD field spectrometer to acquire fine spectra of transgenic samples in contrast experiment conditions, by its biophysical traits and biochemical parameters

to monitor of cultivars from a macro view. The study focused on exploring the spectral differences qualitatively and quantitatively among samples both from spectral shape and parameters and then form monitoring report to assist transgenic crop screening and breeding: by spectral angles, the similarities of samples had been analyzed; by continuum removal, the position of spectral differences had been analyzed qualitatively; by spectral indices high related to photosynthetic pigments contents, the spectral differences among samples had been assessed quantitatively. The results showed, under contrast conditions, the differences between transgenic groups and the Parent could be observed and assessed by hyperspectral remote sensing approach both on spectral morphology and specific indices. Applying our proposed approach, the differences in transgenic groups also could be observed and assessed.

References

Bannari, A., Khurshid, K. S., Staenz, K., & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(10), 3063-3074.

Broge, N., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of environment*, 76(2), 156-172.

Chappelle, E., Kim, M., & McMurtrey, J. (1992). Ratio analysis of reflectance spectra (rars): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sensing of Environment*, 39(3), 239-247.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll azb, and total carotenoid content in eucalyptus leaves. *Remote Sensing of environment*, 66(2), 111-121.

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote sensing of environment*, 41, 35-44.

Gitelson, A., Gritz , Y., & Merzlyak, M. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271-282.

Gitelson, A., Keydan, G., & Merzlyak, M. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Gitelson, A., Zur, Y., Chivkunova, O., & Merzlyak, M. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3), 272-281.

Haboudane, D., Miller, J., Tremblay, N., Zarco-Tejada, P., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of environment*, 81(2), 416-426.

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451-464.

Kim, M., Daughtry, C., Chappelle, E., McMurtrey III, J., & Walthall, C (1994) The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (apar). 299-306.

Murchie, E. H., & HORTON, P. (1997). Acclimation of photosynthesis to irradiance and spectral quality in british plant species: Chlorophyll content, photosynthetic capacity and habitat preference. *Plant, Cell and Environment*, 20(4), 438-448.

Mutanga, O., Skidmore, A., Kumar, L., & Ferwerda, J. (2005). Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, 26(6), 1093-1108.

Raser, J., & O'Shea, E. (2005). Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743), 2010-2013.

Sohn, Y., & Rebello, N. S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing*, 68(12), 1271-1282.

5.3 Comparison analysis of transgenic paddy rice growth based on long time-series spectral data --a case of the experiment repeat experiment

5.3.1 Introduction

By censoring the status of growth on a time point helps to find the growth difference at the same condition, such as sample tolerances to stress. While monitoring the status in time-series data is helpful to find which sample grows stable and it's developing trend.

This work concentrated on three points: growth trend of individual sample groups, comparison of the trends between the transgenic sample groups and comparison of the trends between the transgenic samples and their contrast parents.

5.3.2 Transgenic samples of long time series

Transgenic samples, paddy rice, were cultivated in the greenhouse in Changsha, Hunan Province, China. The greenhouse where temperature could be maintained in a certain degree was managed by Institute of Subtropical Agriculture, Chinese Academy of Sciences. All samples were planted in winter. The first available data were acquired almost 1 month after the samples transferred into the greenhouse, and at that time the samples were in reviving or tillering stage. Because of low temperature, the samples grew slowly. There were total 10 times of data collection, lasting 64 days. Except for the third, fourth and fifth time points, data acquisitions were conducted every 7 days.

These samples were transferred Bar gene for enhancing resistance of herbicide. Before the measurement, all samples had been checked that Bar gene had been transferred into successfully validated by laboratory professional (PCR) approaches. Totally 6 sample groups were selected, 5

were transgenic group and the left was the contrast. Theoretically, except gene transferred, these 6 groups did not have difference; the transgenic samples and contrast one were hereditary; and the transgenic samples were isogenous. The time slices were marked as m01, m02...m10 while the samples were marked as T1, T2...T5 and P, respectively. Every four individual samples were cultivated in one pot since limitation of space. Thus, for eliminating the possible influence caused by growth condition, one pot was treated as one independent sample group.

Spectral data were obtained by ASD spectrometer with contact probe and leaf-clip. The black reference panel was used as background. The center of FOV of the fiber was located at the center of the middle front of the second leaf counted from the core of paddy rice. The principal vein of leaf was vertical to the view line from the fiber. Two measurements were taken at one leaf by micro moving along the principal vein. For reducing heat damage to leaf, 5 pieces of spectra were collected every position as one measurement lasting several seconds. The mean of 5 spectra represents this measurement.

5.3.3 Growth analysis within individual sample

Figure1 showed the individual sample group growth monitoring, expressed by chemicals in leaves. Data in figure was calculated as equation 1:

$$\Delta Index = \frac{(Index_{t_1} - Index_{t_0})}{Index_{t_0}} \quad (5\text{-}3\text{-}1)$$

Where $Index_{t_1}$ represents spectral Index at time t_1.

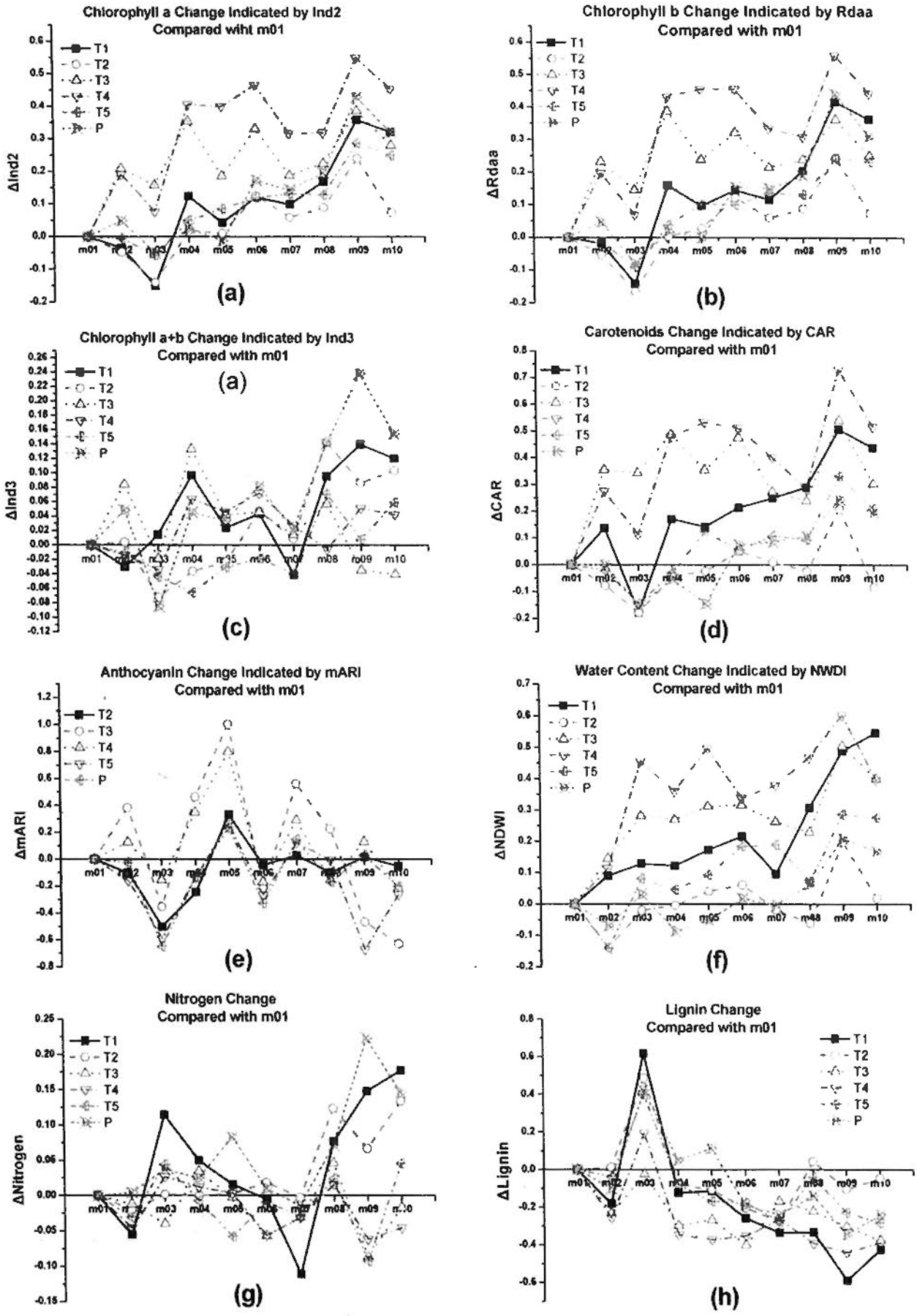


Figure 5-3-1. Track of Individual sample group growth

Expressed by chemicals in leaves and compared with the status at time m01. x axis represents the

10 time slices from m01 to m10.

Table 5-3-1. Index selected and its relationship with foliar chemicals

Index	Equation	
Rdaa	$Rdaa = \sum_{700-730nm} sum(R)$	Chl b (-)
Ind2	Ind2=Rdaa/ ($R_{540-560}/R_{690-710}$)	Chl a(-)
Ind3	Ind3=Rdaa/($R_{930-950}$)	Chl a+b(-)
Car	$\Phi\ (car) = R_{530\_539} * \frac{R_{930\_950}}{(R_{480\_489} * R_{500\_509})}$	Car(-)
mARI	mARI=($R^{-1}_{530-570}-R^{-1}_{690-710}$)×$R_{NIR}$	Anthocyanin (Gitelson, A. A., et al., 2006)
NWDI	$NDWI = \frac{(R_{858} - R_{1640})}{(R_{858} + R_{1640})}$	Water content (Chen, et al., 2005, Zarco-Tejada, et al., 2003)
Nitrogen	Content_nitrogen=23.948 -2.845R1500_1519 +7.742R640_659	Nitrogen content
Lignin	Content_lignin=20.764+8.238 R2390_2409-4.748 R1720_1739	Lignin content

Ch b represents positive correlated with chl b, the same as others.

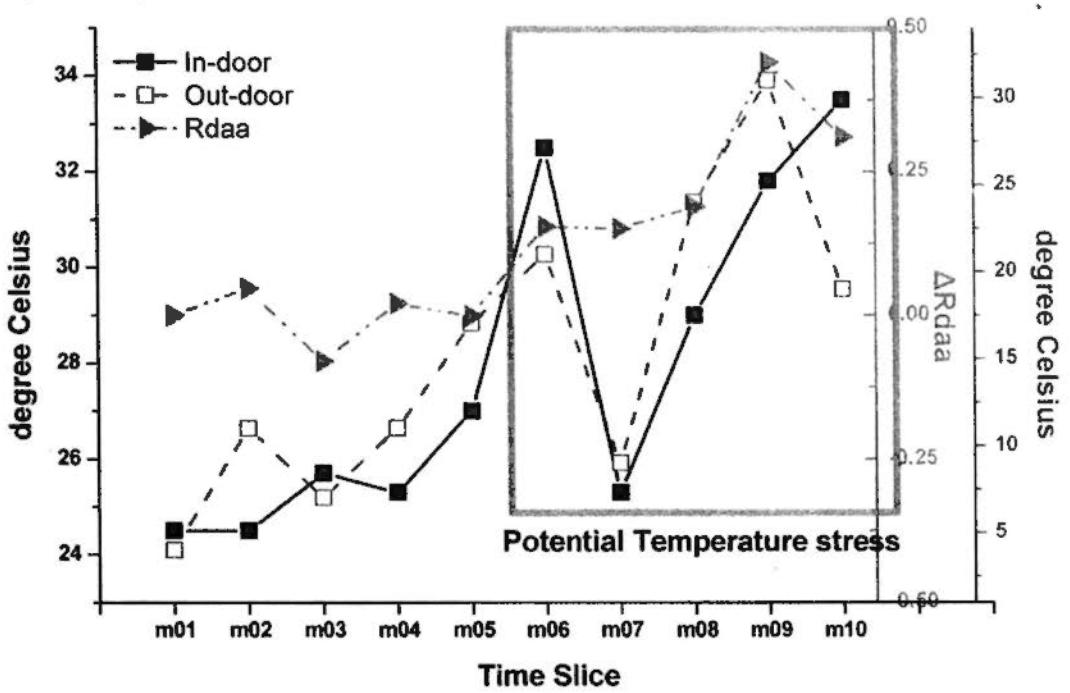


Figure 5-3-2. local condition of temperature and weather. The temperature in the green house was marked "In-door" while the other was marked "Out-door".

Generally speaking, going with grown-up, the photosynthetic ability of samples would be

enhanced. If this trend was expressed by negative related index, the curve of the index should show down-form. However, in figure 1, the results indicated the content of chlorophyll a (figure 1a) and chlorophyll b (figure 1b) in leaves decreased in general except the status at the m01 compared with the status at m01. It might be affected by temperature. Though the greenhouse could maintain temperature, the change of temperature out of the door were violent, the dispersion were up to 10 degree Celsius between day and night outside. The dramatic change also would interrupt the local-environment in the greenhouse. Temperature change is a kind of serious stress to the samples. For avoiding this effect, data acquisition was arranged in the middle of day. However it could not overcome it completely. This stress influences sample growth, on the other hand, it is also a kind of test of stress resistance. Figure 2 showed the relationship between temperature and example spectral index. It was found that when in-door temperature increases, the ΔRdaa (to Rdaa at m01) decreased in general. At time m06, indoor temperature was up to 32.5°, the ΔRdaa increased which revealed the content of chlorophyll pigments decreased. The samples at that time endured temperature stress. And from m06, the samples began entering flowering slowly. Previous studies pointed out that the threshold of high temperature stress was 35°(Jagadish, et al., 2007, Michiels, et al., 1994). Satake et al.(1978) pointed out that if exposed to 35°C more 5 days in flowering stage, rice would be set no seed. However, these data were all acquired under natural condition which would be not in very high humidity. Besides high temperature, very high humidity in the greenhouse was another objective factor which would have influences on rice growth. Meanwhile, many studies also reported that genotypic variation also a factor defining the threshold of high temperature stress (Matsui, et al., 2001, Prasad, et al., 2006, Satake and Yoshida, 1978). Figure 2 showed when the indoor temperature was closed to 33°, temperature stress would happen (figure2 in green block). It was generally consistent with the results reported by Jagadish, et al. (2007). All samples had the same changing trends of chlorophyll pigments. Based on these analyses, Sample T2 and T5 may have stronger resistance to temperature stress. Figure 1f showed NWDI which related to leaf water content. It showed that at m02, m07 and m08, some samples should be watered. Figure 1g showed that T1 maybe consume nitrogen more quickly than the others. And at m02, all samples should be added fertilizer of nitrogen. Figure 1h showed all samples had a stable lignin metabolism. Compared with lignin status at m01, the concentration of leaf lignin decreased in general. Therefore, the greenhouse needed to be cooling and controlled humidity level. These were important findings for professional researcher to discriminate sample changes or external stress and manage the greenhouse.

The foliar chemicals changes reflect the growth pattern of the samples (figure1). If these changes were synchronous, the patterns of these changes would be high related. Moreover, the larger the correlation coefficient of the two chemicals is, the stronger the interaction of them is. It is

important to study of interaction of relevant leaf mechanism. Because of Rdaa, Ind2 and Ind3 are homologous, thus only Rdaa, namely chlorophyll b marked by chl, was chosen to join the correlation analysis. Table 2 showed the relationship (interaction) among foliar chemicals of which input data were the same as data in figure1. The Chl change pattern was stable high related with the track of carotenoids, water content, and lignin content which indicated these chemicals had a strong interaction between each other. They were associated, shown by table 2. It also proved the importance of chlorophyll in leaf from another perspective. It seemed the change pattern of water was slightly related to Nitrogen'. However, this relationship was not stable. Maybe it was because that nitrogen change was more complicated than water and affected by more factors.

Table 5-3-2. The relationship (interaction) between foliar chemicals

Chemicals	Chl	Carotenoids	Anthocyanin	Water	Nitrogen	Lignin
	T1					
Chl	1					
Carotenoids	.959	1				
Anthocyanin	.593	.550	1			
Water	.857	.797	.568	1		
Nitrogen	.490	.297	.367	.761	1	
Lignin	-.857	-.946	-.441	-.580	-.019	1
	T2					
Chl	1					
Carotenoids	.899	1				
Anthocyanin	.537	.505	1			
Water	.699	.746	.294	1		
Nitrogen	.497	.137	.045	.222	1	
Lignin	-.697	-.637	-.720	-.284	-.016	1
	T3					
Chl	1					
Carotenoids	.912	1				
Anthocyanin	.006	-.096	1			
Water	.726	.765	-.356	1		
Nitrogen	-.147	-.414	.514	-.524	1	
Lignin	-.828	-.678	.090	-.648	.142	1
	T4					
Chl	1					

Carotenoids	.982	1				
Anthocyanin	.291	.312	1			
Water	.671	.667	.153	1		
Nitrogen	-.494	-.575	.159	.000	1	
Lignin	-.874	-.821	-.258	-.386	.505	1
	T5					
Chl	1					
Carotenoids	.954	1				
Anthocyanin	-.027	-.079	1			
Water	.788	.715	-.327	1		
Nitrogen	-.316	-.513	-.052	-.098	1	
Lignin	-.835	-.787	-.331	-.498	.456	1
	P					
Chl	1					
Carotenoids	.950	1				
Anthocyanin	.197	.159	1			
Water	.837	.774	-.153	1		
Nitrogen	.756	.560	.089	.823	1	
Lignin	-.852	-.907	-.464	-.490	-.366	1

5.3.4 Growth analysis among sample groups

As mentioned above, the possible difference between the transgenic samples was that they were planted in different pots and put different position in the greenhouse. Theoretically, if ensuring to avoid the influences of external factors, the sample should have same performances (growth status and trends). However, the gene expression is so complicated that the samples may be varied. Therefore, analysis among samples should be conducted.

In this section, the contents of 8 chemicals were chosen as indicators of paddy rice growth monitoring.

Defined a variable vector, chemical change (cc) as:

cc=[chlorophyll_a(Ind2), chlorophyll_b(Rdaa), chlorophyll_ab(Ind3), carotenoids(car), anthocyanin(mARI), water(NDWI), Nitrogen(nitrogen) and Lignin(lignin)] (5-3-2)

Where chlorophyll_a(Ind2) represents that chemical chlorophyll a is indicated by spectral index of the Ind2, the same as the others.

Thus cc represents a comprehensive biochemical status of a sample at certain time. The higher the correlation coefficient between cc of the two samples, the more similar the status of two

samples are. When the samples are isogenous and cultivated at the same condition, the growth difference of thee samples could be reflected by cc. Table 3 lists the correlation coefficients of the cc of all samples. All samples were highly related to each other which indicated that the samples had the similar growth status. However, considering the samples isogenous grew in the same condition, that if any differences between them, it would be very slight. Moreover, the cc is a comprehensive index containing 8 pieces of information, thus the difference of some specific parameter would be overlaid. Therefore, the threshold (expressed by the correlation coefficients between the cc of samples) for the significance of the difference test should be small. In the table 3, the threshold was set to 0.920. If the coefficient was lower than 0.920, the significance of the difference test should be conducted. In the table 3, at m01, the coefficient of T4 and P was less than 0.920. Moreover, the coefficients between transgenic groups and the contrast, and T1 and the others (explained in figure 3), were not as high as the others at m07, m08 and m09. Therefore, growth analysis between transgenic group and contrast group should be conducted.

Table 5-3-3. The correlation coefficients between the cc of samples

Samples	T1	T2	T3	T4	T5	P
	m01					
T1	1					
T2	.999	1				
T3	.967	.963	1			
T4	.958	.952	.999	1		
T5	.995	.995	.985	.977	1	
P	.991	.994	.930	.918	.980	1
	m02					
T1	1					
T2	.987	1				
T3	.990	1.000	1			
T4	.982	.999	.997	1		
T5	.978	.999	.997	.998	1	
P	.999	.982	.986	.977	.972	1
	m03					
T1	1					
T2	.996	1				
T3	.979	.989	1			
T4	.999	.997	.976	1		
T5	.998	.999	.982	.999	1	

P	.980	.986	.997	.975	.980	1
	m04					
T1	1					
T2	.996	1				
T3	.999	.999	1			
T4	.998	.999	.999	1		
T5	.985	.996	.991	.992	1	
P	.998	.999	1.000	.999	.992	1
	m05					
T1	1					
T2	.999	1				
T3	.996	.999	1			
T4	.999	.999	.995	1		
T5	.997	.997	.998	.995	1	
P	.994	.998	.998	.995	.992	1
	m06					
T1	1					
T2	.997	1				
T3	.998	.999	1			
T4	.993	.998	.998	1		
T5	.987	.995	.993	.998		
P	.999	.994	.996	.990	.980	1
	m07					
T1	1					
T2	.988	1				
T3	.945	.972	1			
T4	.962	.987	.997	1		
T5	.986	.998	.983	.993	1	
P	.999	.992	.946	.964	.988	1
	m08					
T1	1					
T2	.967	1				
T3	.959	.999	1			
T4	.977	.996	.994	1		
T5	.962	.997	.997	.988	1	

P	.998	.973	.965	.978	.971	1
	m09					
T1	1					
T2	.964	1				
T3	.943	.997	1			
T4	.976	.998	.993	1		
T5	.955	.994	.997	.995	1	
P	.997	.980	.962	.988	.969	1
	m10					
T1	1					
T2	.952	1				
T3	.971	.996	1			
T4	.978	.990	.999	1		
T5	.985	.986	.997	.999	1	
P	.998	.959	.978	.986	.991	1

5.3.5 Growth analysis among transgenic group and contrast group

To compare chemicals status in leaf the contrast group P, the difference between transgenic sample and its contrast groups at the same time slice following the equation 3.

$$\Delta\, Index \quad = \frac{(Index \quad (T)_{t_i} - Index \quad (P)_{t_i}}{Index \quad (P)_{t_i}} \qquad (5\text{-}3\text{-}3)$$

Where $Index(T)_{t_i}$ represents spectral Index of the sample T at time t_i.

Figure 3 gave the track of Individual sample group growth, expressed by chemicals in leaves and compared with the contrast group (P). The data in the figure 3 were calculated followed equation 3. For chlorophyll a, except for T2, the others had higher chlorophyll a content than the contrast group at time m01. Then the content of T3 and T4 began to decrease while that of T1, T2 and T5 increased at time m02. After m03, the content of all transgenic samples had the same trends increased then decreased. The figure 3a showed that after the time m06, transgenic sample had higher content of chlorophyll a which indicated that the samples were more active photosynthesis than the contrast. The same development patterns of the content of chlorophyll b, Carotenoids were showed in figure 3b, 3d.

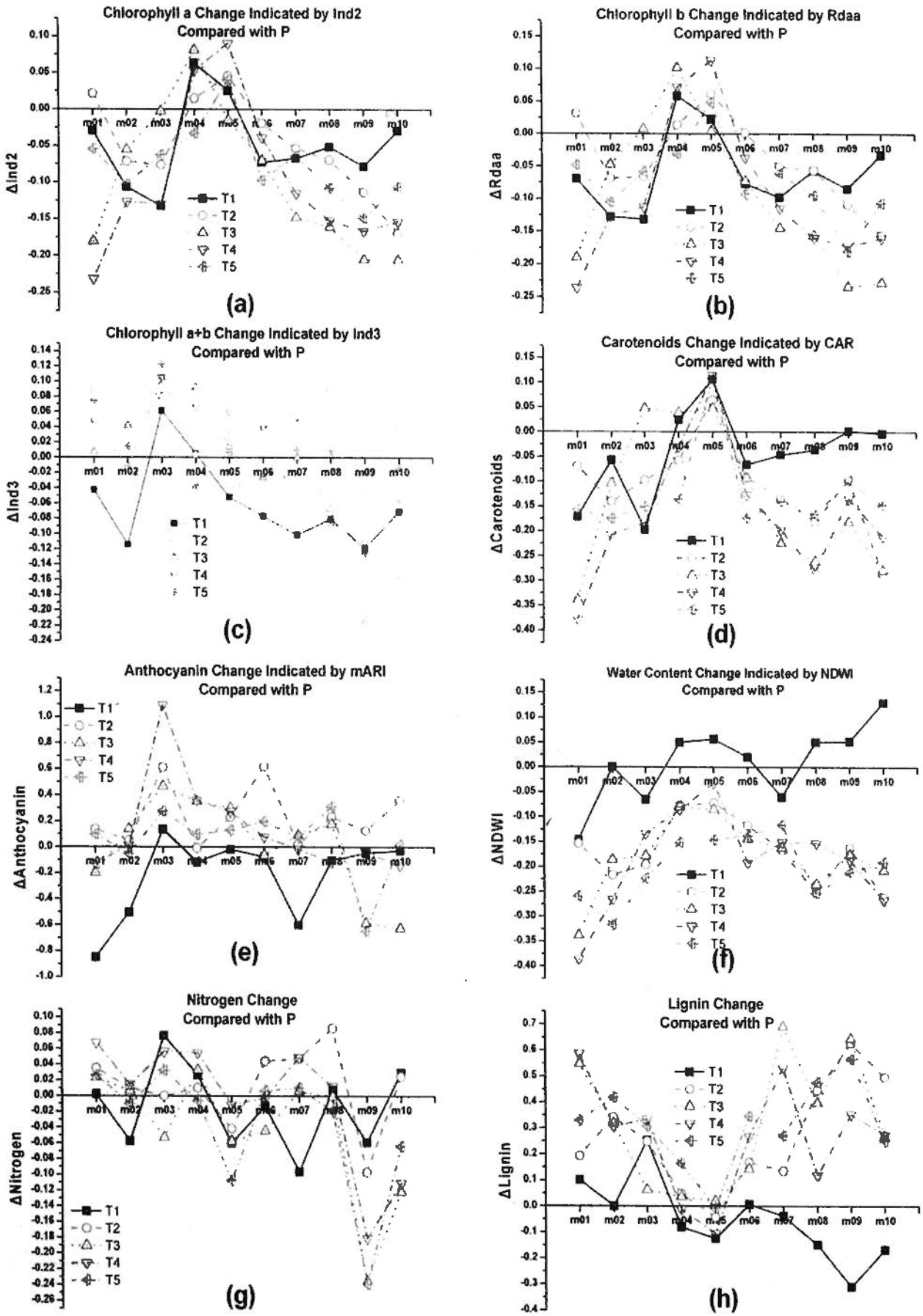


Figure 5-3-3. The track of Individual sample group growth

Expressed by chemicals in leaves and compared with the contrast group (P). The x axis represents

the 10 time slices from m01 to m10.

Figure 2 showed that began m04, the temperature in door increase quickly, and favorable temperature made the sample vigorous. And transgenic samples were more active than the contrast. After m06 (except for m07), the samples suffered temperature stress, however, the transgenic samples still had higher content of chlorophyll a and b. The content of chlorophyll a+b indicated by Ind3 were not consistent with the information showed in figure 3a and 3b in general. Figure 3a and 3b indicated that the content of chlorophyll a and b of transgenic groups were higher than the contrast except for m04 and m05. However, the opposite difference was shown by figure3c. It was possible because of difference of reference bands for spectral index, Ind3. But, the content change of T1 chlorophyll a+b was consistent with the change patterns of chlorophyll a and b. And in table3, the coefficients of the cc of T1 and the other transgenic samples were not as high as the others. One of the reasons might be the abnormal expression of chlorophyll a+b of other samples. Figure 3e indicated the track of the content of Anthocyanin, it was just for reference. The time m04 was a turning point among figure 1a, 1b and 1d. Excluding m05, all other time, transgenic samples, except for T1, had higher content of carotenoids than the contrast. Figure 3g showed the content changes of nitrogen vibrated seriously around zero. From the figure, T3, T4 and T5 were more sensitive, the chemical changes of them varied violent.

The cc reflected that the contrast sample had very similar growth status expressed by foliar chemicals except for with T4 (0.918 in table 3 at m01). Thus, it was necessary to conduct a significant test of the difference between T4 and P. the chemical content changes were set as input data (calculated follow eqaution1).

Table 5-3-4. The significant test of the difference between T4 and P (α=0.01)

stat.	Chl a	Chl b	Chl a+ b	Carotenoids	Anthocyanin	Water	Nitrogen	Lignin
Sig.	0.0993	0.1176	0.8420	0.0091	0.5642	0.00061	0.6926	0.0710
h	0	0	0	1	0	1	0	0

In table 4, sig. is the p-value associated with the t-statistic. In this section, when sig. is smaller than 0.01, the null hypothesis should be rejected and then h equals to 1. This result revealed the two tested variables were significantly different. From the table, the content of carotenoids and water of T4 were significantly different to the contrast group's P's. And they were the principal factors causing the coefficient of the cc of P and T4 lower than 0.920. This result also indicated that the threshold was set to 0.92 was reasonable. Except for water and carotenoids, the other chemicals of T4 and P had no significant difference.

5.3.6 Conclusion

In this section, the long time series data of transgenic groups and the contrast were analyzed. The

samples growth pattern and status were analyzed within individual sample, within sample groups, and among transgenic group and contrast group respectively. By these analyses, some important indicative results were obtained. A potential temperature stress had been found. The Chl change pattern was stable high related with the track of carotenoids, water content, and lignin content. It revealed these chemicals have a strong interaction each other. A new description variable (vector) chemical change (cc) was defined to as indicators of paddy rice growth monitoring. It is a comprehensive biochemical status of a sample at certain time. By censoring the correlation results of samples cc, the status of T4 and P at m01 were not consistent. After statistic test, the change of carotenoids and water contents of them were significantly different. All these results would be useful to assist professional biologists to fulfill crop screening and breeding. It helps them know their samples clearly and help experiment under control.

References

Chen, D., Huang, J., & Jackson, T. (2005). Vegetation water content estimation for corn and soybeans using spectral indices derived from modis near-and short-wave infrared bands. *Remote Sensing of Environment*, 98(2-3), 225-236.

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Jagadish, S., Craufurd, P., & Wheeler, T. (2007). High temperature stress and spikelet fertility in rice (oryza sativa l.). *Journal of experimental botany*, 58(7), 1627-1635.

Matsui, T., Omasa, K., & Horie, T. (2001). The difference in sterility due to high temperatures during the flowering period among japonica-rice varieties. *PLANT PRODUCTION SCIENCE-TOKYO-*, 4(2), 90-93.

Michiels, J., Verreth, C., & Vanderleyden, J. (1994). Effects of temperature stress on bean-nodulating rhizobium strains. *Applied and environmental microbiology*, 60(4), 1206-1212.

Prasad, P., Boote, K., & Allen, L. (2006). Species, ecotype and cultivar differences in spikelet fertility and harvest index of rice in response to high temperature stress. *Field Crops Research*, 95(2-3), 398-411.

Satake, T., & Yoshida, S. (1978). High temperature-induced sterility in indica rices at flowering.

Zarco-Tejada, P., Rueda, C., & Ustin, S. (2003). Water content estimation in vegetation with modis reflectance data and model inversion methods. *Remote Sensing of Environment*, 85(1), 109-124.

5.4 Analysis of difference response of transgenic samples to stress --a case of the foreign genes activated by artificial induction

In sections 5.1-3, the transgenic samples growth under a normal condition without man-made stress, and the characteristics and growth status of the samples were evaluated. In this section, a controlled stress environment had been made for the samples aiming at the personality of the transgene. By these designs, the genes transferred into the receptor would be promoted, and

different characteristics of transgenic samples under this condition would be captured by the proposed approach.

5.4.1 Samples and experiment

5.4.1.1 Samples

In this study the author cooperated with Centre of Crop Ecology in Environmental Stresses, CAS. Because of relative low temperature, maintaining around 25° , samples grew slowly. Until the experiment, after been transferred into pots 90 days, they were still in stage of tillering. For controlling the growth conditions, every 4 same individuals of samples as a group were planted in a pot. They transferred Bar gene for resisting herbicide into samples. It made the sample more complicated during breeding.

In the greenhouse, several kinds of high intensity lights were placed for supplement sunlight in winter. Also air-condition, heater and dehumidifier were employed to automatically change setting to maintain indoor micro-environment for paddy rice growth. For suppressing external noise to samples, soil in and material of pots were the same. Water and fertilizer were controlled too. All pots were arranged on a platform approximately 1.2mters high for heat control. The samples were distinguished into 8 groups according the pots they were. Seven transgenic groups, they are T1 to T7, and one contrast group, parent.

5.4.1.2 Experiment

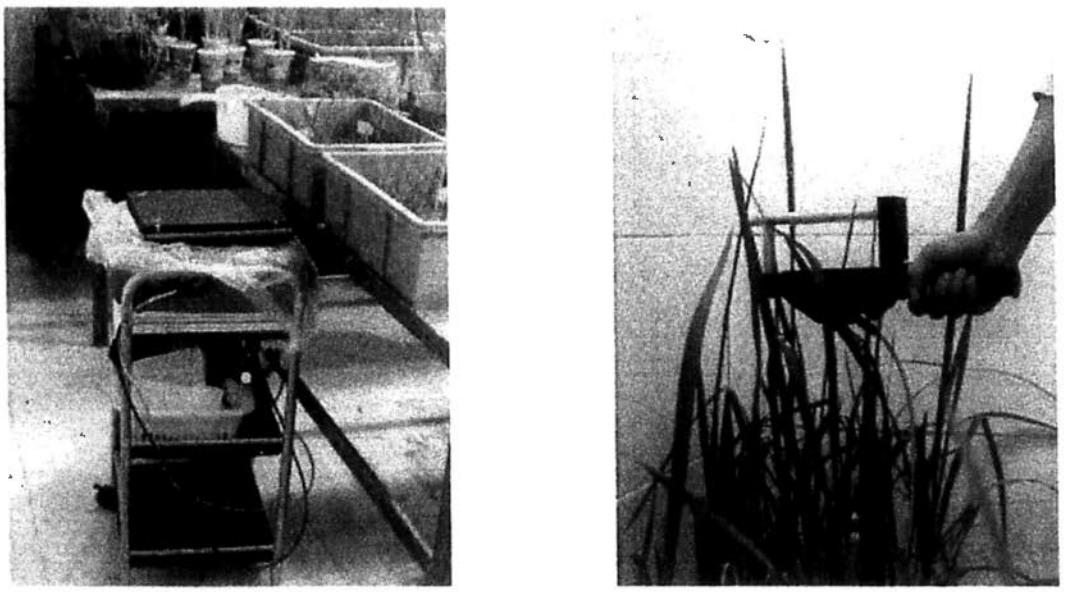
The experiment was done on March, 2010, in Changsha, China. All transgenic samples and their parent as contrast group were cultivated in a professional greenhouse. The author collected spectral reflectance data (-0h) of samples. Then except for the contrast group, all others were sprayed herbicide of 0.3% Basta (C5H15N2O4P). In the following days, the author collected post-24h-spayed spectral reflectance data, post-48h, post-72h and post-168h ones (recognized as -0h, 24h, 48h, 72h, 168h). By compared the transgenic samples with their parents and their brothers in the time series, the author wanted to find out if there is any difference of spectral response of transgenic samples to herbicide; if yes, what intensity of the responses is between sample groups in the same time point and in the groups in the time series respectively. These responses would be the indicators of differences response capability between samples in biophysical and biochemical process. Particularly under controlled experiment, they could be as important reference to assess difference of samples by comparison to fulfill monitoring and screening samples.

Considering that diurnal changes in metabolic actives, growth pattern and cell division were at

day/ night changes, the author collected data in a fixed time, beginning at 2pm and in a certain sorts of samples to suppress the effects of these changes.

5.4.1.3 Equipment

In this experiment, considering the differences between transgenic samples and their parent, the author acquired fine spectra of leave. The selected leaves were the second one of the plant counted from the core leaf to the out. And the spot of probe view located at the middle of the leave to suppress misunderstanding caused by differences of leave. In the study, an integrated system was employed consisting of an Analytical Spectral Devices (ASD) FieldSpec 3 Spectrometer with contact probe and leaf-clip (figure 1). Spectral reflectance data range from 350nm to 2500nm. Leaf-clip interfaced with the ASD high intensity probe.



a b

Figure 5-4-1. ASD Spectrometer with probe; b. Data acquisition

5.4.1.4 Data pre-test

In this study, the author focused on the responding action on a group level, namely the author took the mean spectra to present the group as input data.

Before calculating the mean spectra, the author first had outlier test, Grubbs Test, in groups to ensure the consistency of spectral data in the same group. For this test, the input data should fit to normal distribution, thus before the test, the author first had a Single sample lilliefors hypothesis test of composite normality with command 'lillietest'. For lillietest, namely, the test value equals to 0, the null hypothesis that the samples fit to normal distribution, should be accepted, and otherwise rejected. The results of these two data pre-test could be read from

table1. In the whole bands of ASD, 2151 interpolation points from 350nm to 2500nm was conducted by outlier test, while bands only in 420nm to 2400nm were tested with lillietest since from 350nm to 419nm and 2401nm to 2500nm, 170 interpolation data were easily polluted by noise. The table1 shows some data did not fit to normal distribution. Most of these ones locate in near red infrared and even after bands. These bands (e.g. 1409nm- 1997nm) are water and temperature sensitive. Temperature sensitive in study for foliage is an indirect response of water sensitive. This problem should be paid more attention to in future to avoid temperature effect. In this study, for eliminating its effect, the author choose evaluating factors which do not locate in or take the readings in the range as input data shown in table1 as much as possible. Outlier test showed in some sample groups, there were lots of data were tested as outliers. Since the maximal number of outlier was 2055, compared 2511*8, the outliers were still a small probability. Thus removing the outliers would not affect further studies. And this outlier could be analyzed in the section 5.5. Moreover, the data showed that the number of outliers had no related to its normality.

Standard variation is necessarily supplementary to describe the characteristics of sample data. In these five time slices, the maximum average of standard deviation calculated by band was 0.075 and most of them were around 0.001. It indicated that the data were consistent. In all, from the table1, applying the mean spectra data as input data to represent the group is reasonable and practicable.

Table 5-4-1. Pre-test of consistency within sample groups

-0h								
Sample groups	Sample numbers	Outlier test*	Normal Distribution**		Standard variation (stdv)***			
			Normal	Range(nm)	Average	Stdv	Max	Min
T1	8	0	52	643-668	0.009	0.005	0.018	0.001
T2	8	0	1	/	0.009	0.005	0.018	0.003
T3	8	2055	1678	534-566 704-1915	0.048	0.032	0.096	0.004
T4	8	48	44	530-574	0.075	0.018	0.060	0.003
T5	6	920	7	/	0.012	0.011	0.033	0.002
T6	8	0	7	/	0.008	0.004	0.019	0.002
T7	6	0	22	/	0.007	0.007	0.024	0.001
Parent	15	39	451	708-718 742-758 1493-1864	0.013	0.010	0.030	0.002
24h								

Sample groups	Sample numbers	Outlier test*	Normal	Range(nm)	Averge	Stdv	Max	Min
T1	8	0	22	/	0.011	0.008	0.026	0.002
T2	8	0	63	1579-1609	0.008	0.004	0.026	0.003
T3	8	0	170	724-746 1603-1743	0.021	0.018	0.053	0.003
T4	8	0	292	421-508 523-574 1409-1997	0.018	0.013	0.041	0.003
T5	6	0	30	/	0.016	0.010	0.033	0.004
T6	8	0	376	421-493 1010-1295	0.012	0.008	0.027	0.002
T7	6	0	0	/	0.023	0.016	0.054	0.002
Parent	15	1	34	/	0.015	0.007	0.030	0.003

48h

Sample groups	Sample numbers	Outlier test*	Normal	Range(nm)	Average	Stdv	Max	Min
T1	8	920	904	729-1383 1575-1776	0.010	0.005	0.019	0.004
T2	8	66	76	/	0.010	0.007	0.022	0.002
T3	8	155	188	2132-2291	0.034	0.028	0.078	0.002
T4	8	1139	703	421-496 742-781 1388-1593	0.032	0.024	0.068	0.002
T5	6	9	1	/	0.015	0.012	0.039	0.002
T6	8	256	74	1398-1438	0.009	0.004	0.018	0.003
T7	6	14	3	/	0.005	0.004	0.020	0.000
Parent	15	0	29	/	0.013	0.008	0.027	0.002

72h

Sample groups	Sample numbers	Outlier test*	Normal	Range(nm)	Average	Stdv	Max	Min
T1	8	0	35	/	0.010	0.004	0.021	0.003
T2	8	0	79	1495-1530 2081-2105	0.010	0.005	0.021	0.003
T3	8	5	222	1484-1573	0.015	0.010	0.034	0.004

				1889-2003				
T4	8	807	451	1946-2374	0.009	0.005	0.019	0.001
T5	6	985	656	734-1309 1336--1376	0.008	0.005	0.017	0.003
T6	8	206	319	1546-1679 1733-1786 1905-2019	0.007	0.004	0.033	0.003
T7	6	101	0	/	0.025	0.020	0.061	0.003
Parent	15	0	0	/	0.014	0.008	0.031	0.004
168h								
Sample groups	Sample numbers	Outlier test*	Normal	Range(nm)	Average	Stdv	Max	Min
T1	8	135	195	2150-2266	0.012	0.007	0.029	0.003
T2	8	112	240	557-714	0.012	0.005	0.029	0.003
T3	8	12	683	746-1029 1133-1180 1215-1305 1312-1385 1584-1761	0.031	0.025	0.073	0.003
T4	8	6	90	/	0.013	0.011	0.032	0.001
T5	6	0	1	/	0.014	0.007	0.029	0.002
T6	8	0	9	/	0.015	0.005	0.022	0.002
T7	6	10	5	/	0.014	0.007	0.029	0.006
Parent	15	306	95	/	0.013	0.007	0.027	0.002

*The number of samples of which found as outliers.

**The data in 'Normal' column are the numbers of samples which do not fit to normal distribution. The data in 'Range' column means the continuous range of the non-normal distribution samples located in.

***In these columns, the standard variations calculated by the individual spectrum band in the group had been listed, including their averages, standard variations, maximums, and minimums. These data helped to describe the consistency of the samples in group.

5.4.2 Quantitative analysis based on spectral morphological characteristics

In this study, the author had done both qualitative and quantitative analysis of samples based on morphological and biochemical parameter characteristics of spectral reflectance. By spectral

morphological analysis, the author could locate the difference between samples in spectral bands, and then choose parameters orientated. In general the first kind analysis would be described as the spectral feature location. By biochemical parameter characteristics, the author could obtain the quantitative data responding to the differences between or stress to, if any, samples. These data were useful to assess growth status of samples.

5.4.2.1 Spectral angle between samples

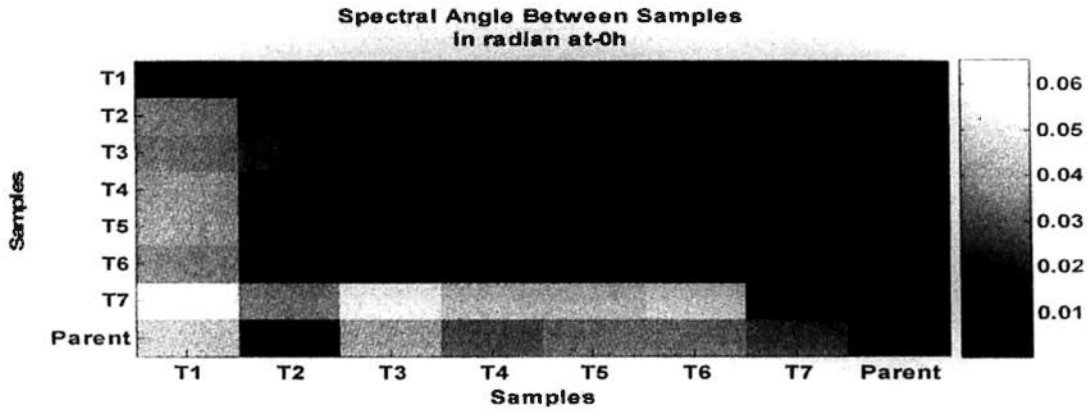
In two-dimensional space defined by band x and band y, spectral signatures could be described as vector, thus the angle between these two spectral signatures, t and r, could be as follow (Sohn and Rebello, 2002). Therefore, spectral angle could be applied to represent the similarity of the two spectral signatures. The less the angle is, the more similar the two signatures are.

$$\theta_{t,r} = \cos^{-1}\left(\frac{\vec{t}\cdot\vec{r}}{\|t\|\cdot\|r\|}\right) \qquad \theta \in \left[0 - \frac{\pi}{2}\right] \quad (5\text{-}4\text{-}1)$$

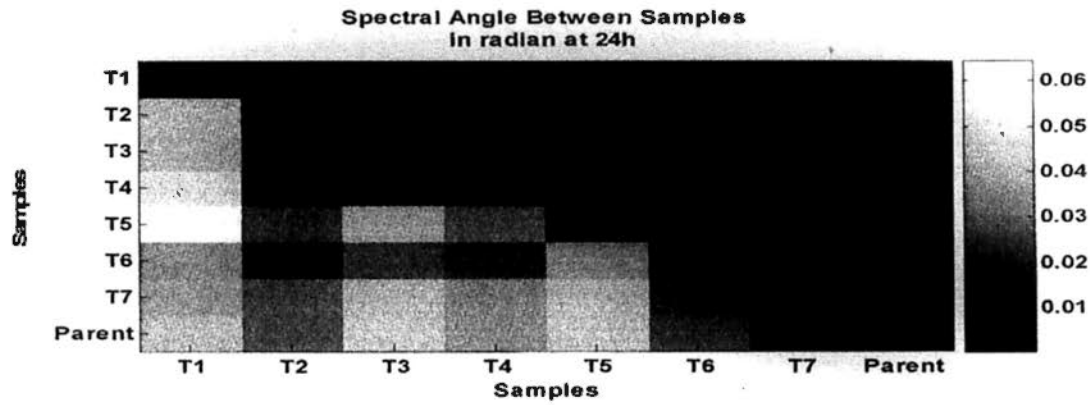
The spectral angles of each two samples at the 5 time points have been calculated in radian. From the figure 2, the spectral angles between theses samples were not very great; the greatest one was no more than 0.07 radian. At -0h, sample T7 had great angles with the others, even great than angles of parent and transgenic ones. The biggest angles were the one of T7 & T1 and T7 &T3. However, the spectral angles between the left groups, just except T7, mainly ranged around 0.03 radian. Comparing the transgenic samples with the contrast group, the spectral angles were not very big, especially for T2, it showed much similarity of spectra to the parent. After 24 hours post sprayed herbicide, the angles of transgenic ones to their parent began to increase gradually. However the values of them between samples were not large in distance. The largest angle was the one between T1 and T5. The others were not different sharply. After 48 hours, the angles with parent were continuing bigger and the largest one became the one between T5 & parent and T5 & T7. After 72 hours, angles between T5 & parent and T5 & T7 were relatively still large. However, the biggest ones were T1 & T5 and T1 & T4. 168 hours after, the plants were of adaptation to the herbicide stresses and had a stable status of growth. Compared with the angles after 72 hours, the angle between of T4 and parent was the largest one, the angles of T4 with T1, T2 and T7, and T5 with T1, T2, T7 and parent were relatively larger too. During these 168 hours, to the angles of T6 with the others were smaller than others, and they were stable. In this procedure, the parent had no stress on, the only differences between it was the influences caused by normal growth. Therefore, it could be as reference to assess the adaptation of samples to stress. Compared with the figures at -0h and 168h, sample T1, T2, T3, T6 and T7 recovered to the original spectral angle values. In this study with existing data, the author could not quantitatively offer what big the spectral angles are significant to indicator the true difference

between samples. However, the angle at last showed a possibility of difference, providing an efficient practical strategy eliminating workload for monitoring samples compared with traditional methods. Thus, from these figures, some questions, as discoveries would be abstracted.

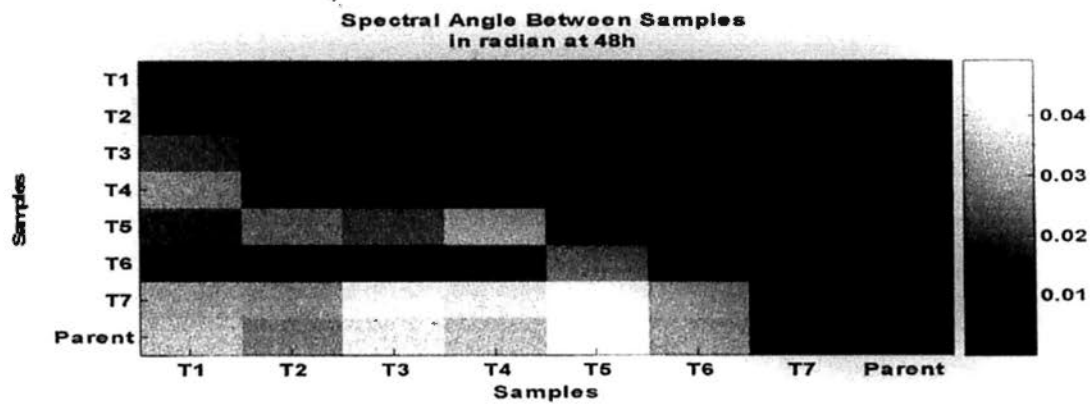
During these 168 hours, sample T1, T3, T4, T5 and T7 were discriminated. The difference existed between transgenic samples was more obvious than that between transgenic ones and their parent. Especially for T7, it was much close to its parent. All these might provide useful information about these samples and be reported to professional technician of biology. And in the following work, T1, T4, T5 and T7 would be focused on.



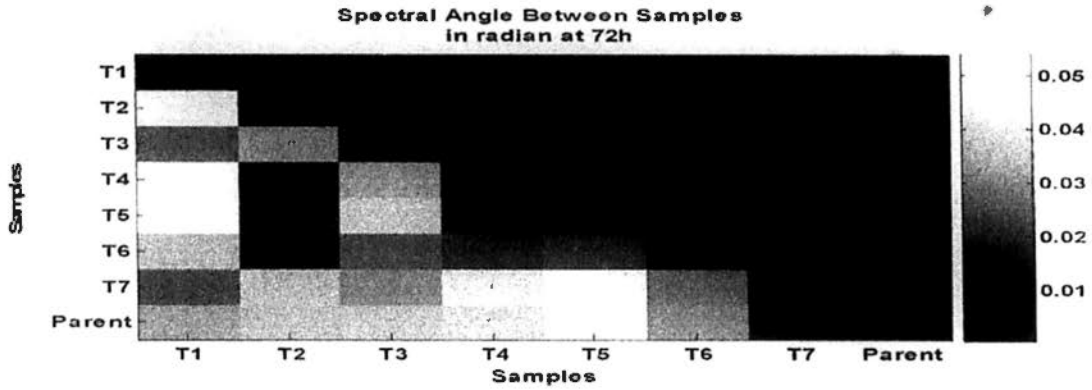
(a)



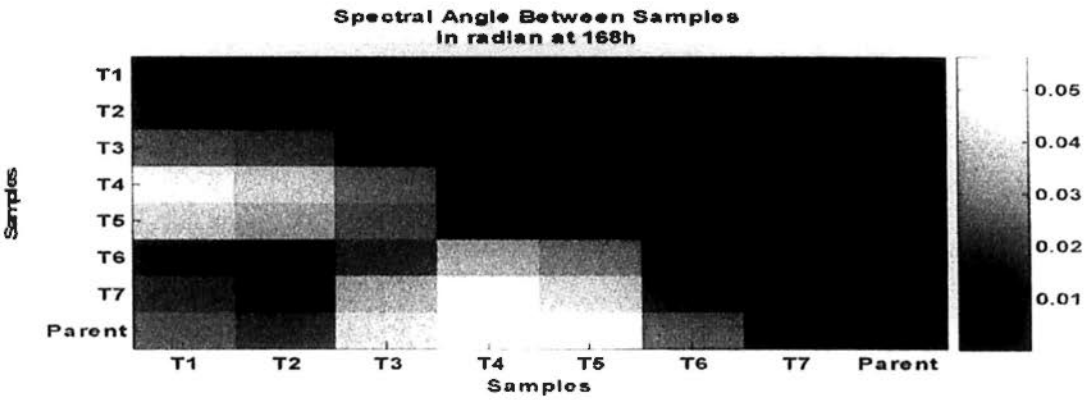
(b)



(c)



(d)



(e)

Figure 5-4-2. Spectral angles among samples by time

5.4.2.2 Spectral difference of continuum removal

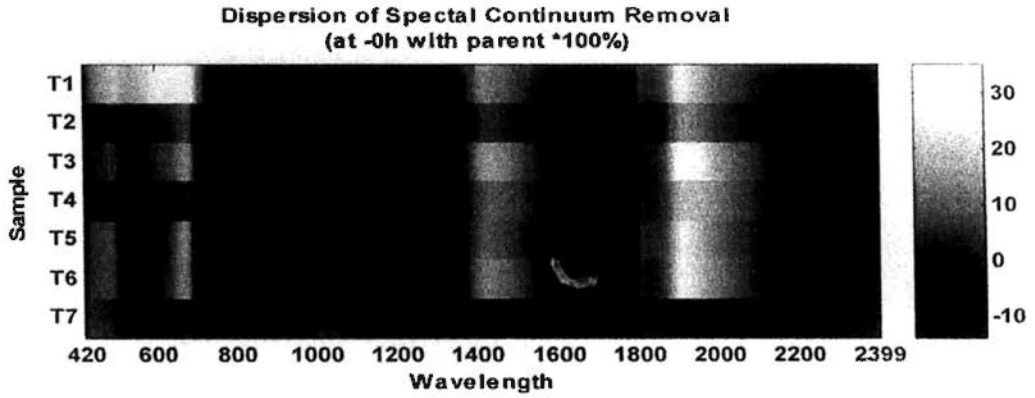
Vegetation spectrum, responding to electromagnetic energy, composed of absorption, reflectance and emission of energy, is decided by vegetation chemical and spectral morphological features which are highly related to the development health and growth conditions of plants (Boochs, et al., 1990, Tong, et al., 2006). In visible bands, kinds of pigments, especially chlorophyll, are the main things responding to electromagnetic energy (Gates, 1965). In near Infrared bands, the responses are mainly controlled by internal cell structure of leaf (Gates, 1965). Previous studies revealed that by vegetable spectrum, information both of surface and inner leaf could be obtained, including biochemical components and biophysical processes (Wessman, et al., 1988). Thus, by comparison the spectral shape, it is easy to locate the difference and deduce the possible reason for this difference. Reflectance is a relative magnitude without unit. However, it is easily affected by external factors such as background. Thus, before comparison, it should be normalized in high-accuracy analysis.

In this study, spectral continuum removal (CV) was employed which project the data to [0, 1]. All components in data contribute to this transform which emphasized the location and depth of

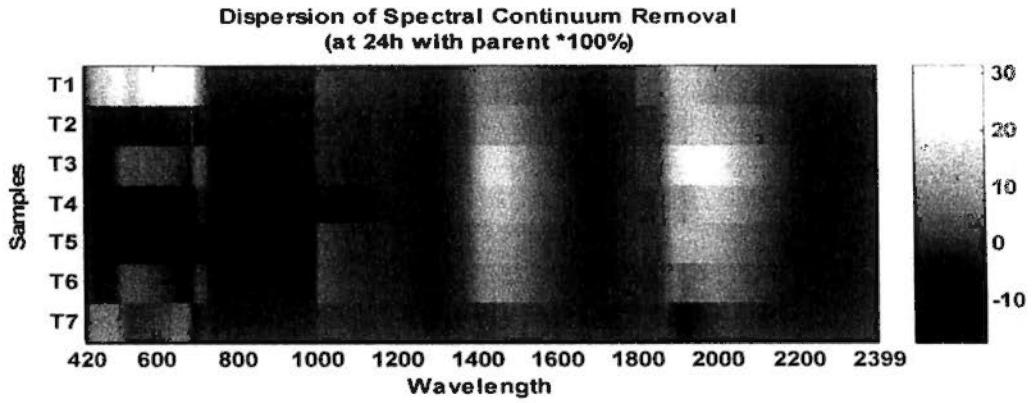
individual spectral absorption features (Clark and Roush, 1984). By this transformation, the data were projected to a space where they could be compared with each other directly associated with wavelength. For clearly describing the spectral the differences between samples, the author chose a baseline of data and subscribes CV of the other from it. There is a simple relationship between CV and spectral reflectance that the greater the CV value is, the greater the spectral reflectance value is.

5.4.2.2.1 Comparison with the parent between groups

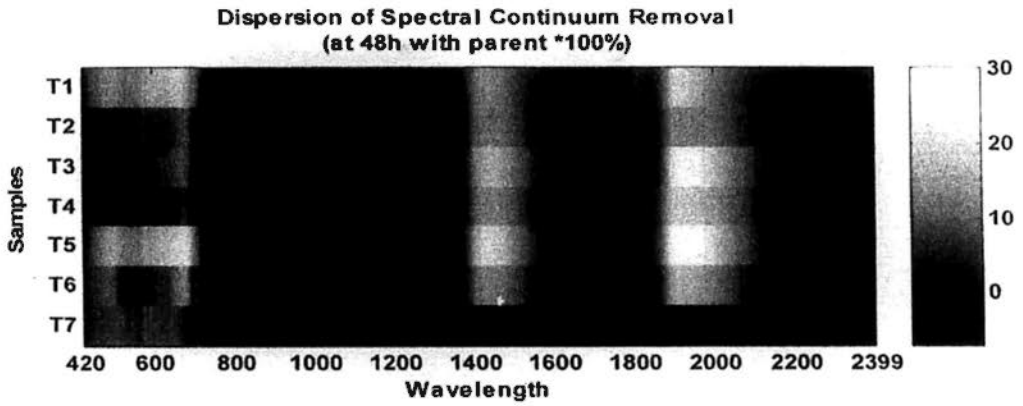
By investigating the figures, it could be found that differences mainly located in range of [420,750nm], [1400, 1600nm] and [1800, 2000nm]. The dispersions compared with parent group could be negative or positive. If the dispersion was negative, that means the transgenic sample had a lower reflectance than its contrast group at that certain bands, else was the opposite. The spectral reflectance which was calibrated represent the capability of certain object reflecting or absorbing light energy (in some situations, emitting is also involved). The absolute biggest dispersion was more than 30%. Before spraying herbicide, the biggest dispersion of transgenic samples with their counterpart was with T1 in [420, 750nm], and with T3 in [1800, 2000nm]. For others these variances were around 0. 24 hours later, transgenic samples responded to stress of herbicide, the dispersions had been enlarge totally, especially after red-infrared ranges. The largest dispersion was still with T1 in [420, 750nm]. But at this time, there was no significant difference between the responses to stress of transgenic samples. 48 hours after, the responses to stress of the samples began to be significantly different (the figure became more violent with clear edges between groups). [1400, 1600nm] and [1800, 2000nm] are bands water sensitive. When light from the high-intensity contact probe's self-lightning system, the surface temperature would be slight changed which would be responded by spectral reflectance of the leaf. By suppressing sampling time to eliminate the surface temperature (water content) change, however, the effect of surface temperature change could not be avoided totally. It was one of the reasons why there would more obvious differences in these two ranges at the figure. Of course the rear difference of water content in samples would also cause the results what the figures show. [420, 750nm] is the range photosynthesis sensitive bands. In this range, T1 at the five time slices had a great positive dispersion, while T2, T4, T5 and T7 had several shifts from negative to positive or the opposite. T3 and T6 were more stable compared with their counterparts. By these comparisons, qualitatively information could be obtained about the capability of the sample responding to the herbicide stress.


Dispersion of Spectal Continuum Removal
(at -0h with parent *100%)
Sample
T1
T2
T3
T4
T5
T6
T7
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
30
20
10
0
-10

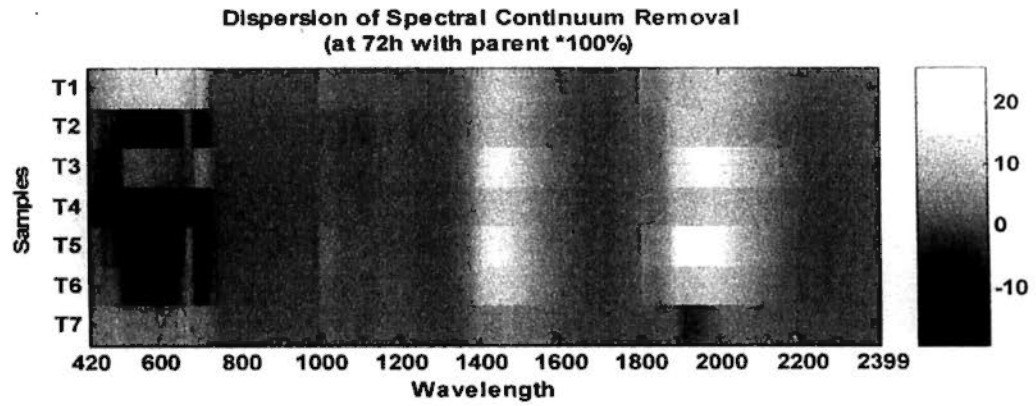
(a)


Dispersion of Spectral Continuum Removal
(at 24h with parent *100%)
Samples
T1
T2
T3
T4
T5
T6
T7
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
30
20
10
0
-10

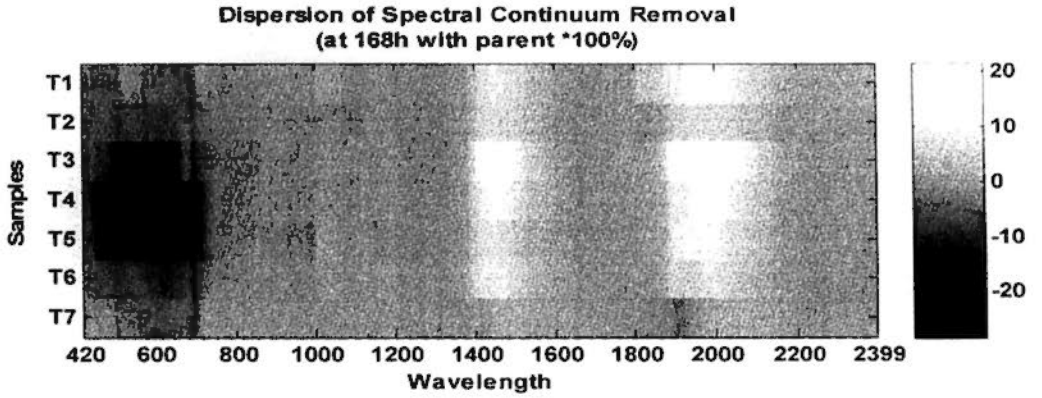
(b)


Dispersion of Spectral Continuum Removal
(at 48h with parent *100%)
Samples
T1
T2
T3
T4
T5
T6
T7
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
30
20
10
0

(c)


Dispersion of Spectral Continuum Removal
(at 72h with parent *100%)
Samples
T1
T2
T3
T4
T5
T6
T7
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
20
10
0
-10

(d)



(e)

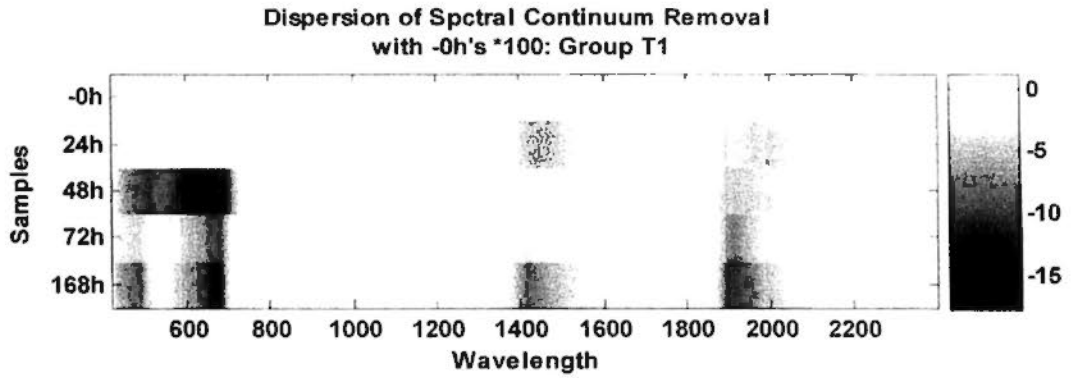
Figure 5-4-3.Dispersion of spectral continuum removal results (compared with the contrast groups by time, *100%)

5.4.2.2.2 Comparison in the inner groups

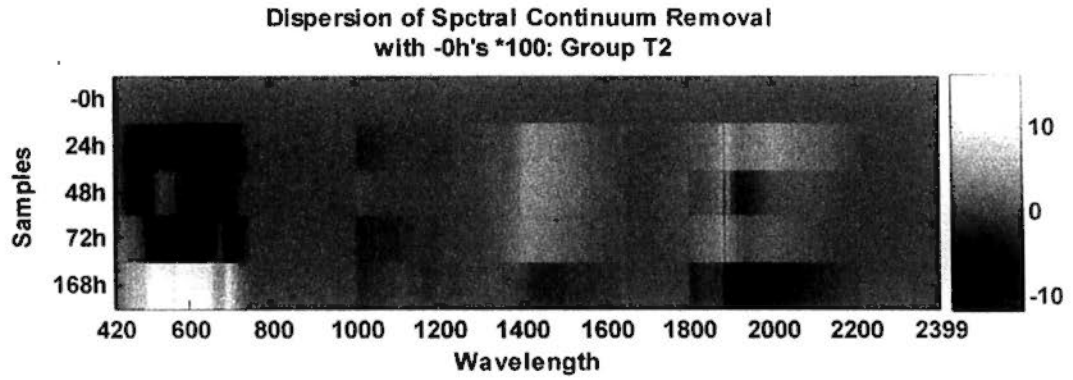
For obtaining the responses to the stress of the same sample, the CV at -0h was taken as baseline, and subtracted by its counterparts at the other time slices. By this analysis, the author could get information of responses during the procedure of stress adaption. From these figure, except for T1, T7 and Parent, the other groups had a shift of dispersion from negative to positive. Compared with the results between groups, the figures showed the dispersions mainly located [1000, 1200nm] besides [420, 750nm], [1400, 1600nm] and [1800, 2000nm]. Also, for T1, T3, T5 and T6, there were some slight differences shown in the figures in ranges [800, 1000nm]. All the samples adapting to the herbicide stress fitted to a same procedure generally: the most sensitive bands were photosynthetic ones. It revealed that herbicide had a very strong stress to the foliar photosynthetic pigments. This stress would positive (figure 4a) or negative (figure 4e) decided by the specific sample's tolerance and responses. And chlorophyll and carotenoids pigments would have different kinds of responses (figure 4c, f and g). 48 hours after, most samples had adapted to herbicide stress (or the stress was over), and the status reflected by spectrum recovered to the -0h level. 168 hours after, the status was all most the normal growth, and the herbicide stress had little influence on the sample. This conclusion could be deduced from the growth pattern of the contrast group. The contrast group had not been sprayed herbicide since they would be destroyed totally. Thus the pattern expressed by the contrast could be set as a reference for comparison. Water content sensitive bands, mainly around 1400nm and 2000nm, also responded to the herbicide stress. Meanwhile these two ranges also were nitrogen and lignin sensitive bands (section 4-3). More data were needed to confirm whether the herbicide or water and nutritional stress made the responses. The different performance to adapt the stress reveals the inner discrimination of samples in general, and all these difference responses would be studied in

the Qualitative analysis based on parameterized features.

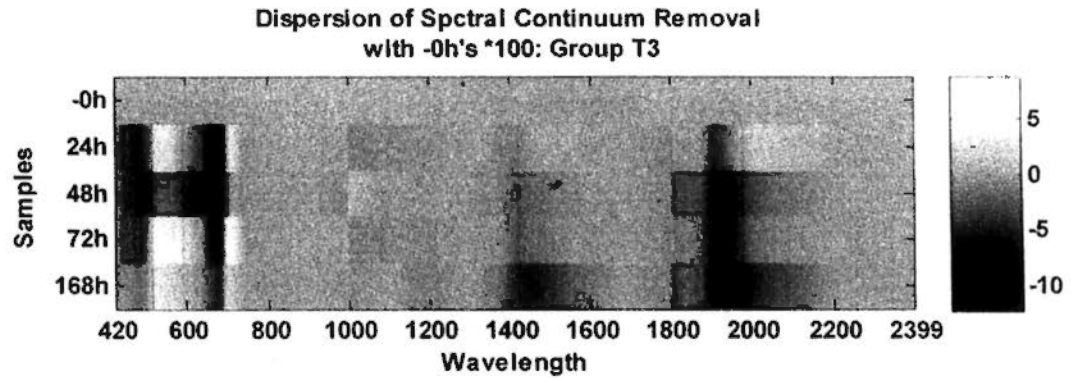
Moreover, from 700 nm to 1000nm (infrared bands), spectral features of transgenic groups had no changes. According to optical properties of leaf, the features in these bands were decided by the leaf inner structure (Carter, GA and Knapp, AK, 2001). It demonstrated that the herbicide had no influence on the internal structure of transgenic leaf.



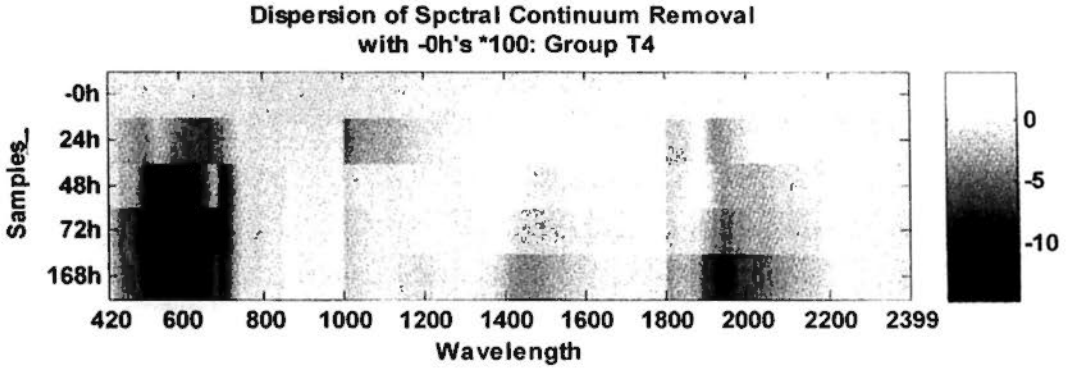
(a)



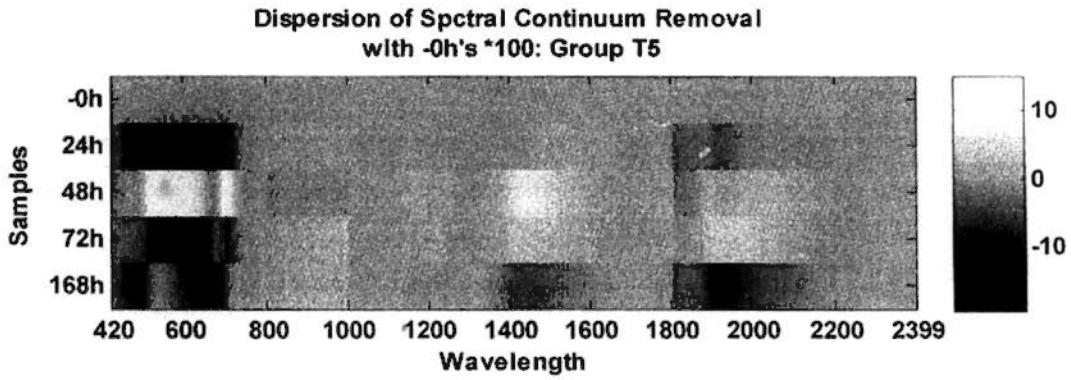
(b)



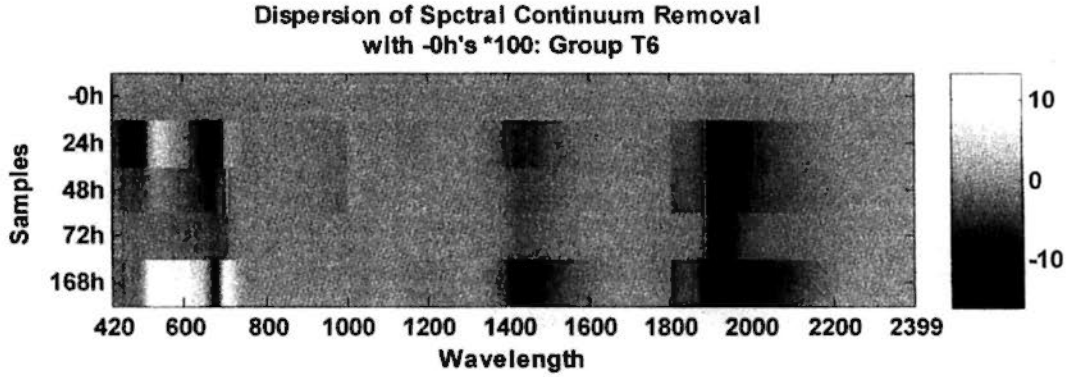
(c)


Dispersion of Spctral Continuum Removal
with -0h's *100: Group T4
Samples
-0h
24h
48h
72h
168h
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
0
-5
-10

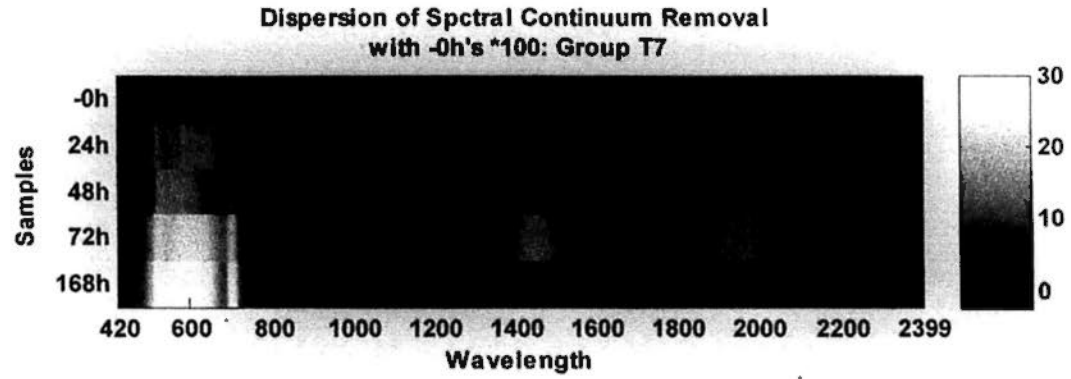
(d)


Dispersion of Spctral Continuum Removal
with -0h's *100: Group T5
Samples
-0h
24h
48h
72h
168h
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
10
0
-10

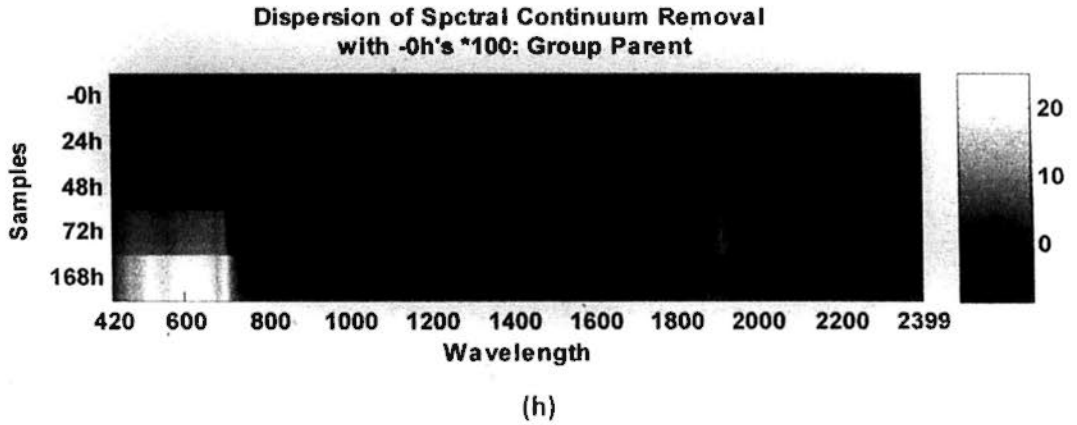
(e)


Dispersion of Spctral Continuum Removal
with -0h's *100: Group T6
Samples
-0h
24h
48h
72h
168h
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
10
0
-10

(f)


Dispersion of Spctral Continuum Removal
with -0h's *100: Group T7
Samples
-0h
24h
48h
72h
168h
420 600 800 1000 1200 1400 1600 1800 2000 2200 2399
Wavelength
30
20
10
0

(g)



(h)

Figure 5-4-4.Dispersion of spectral continuum removal with groups (compared with the status at -0h, *100%)

5.4.3 Qualitative analysis based on parameterized features

5.4.3.1 Parameter selection

5.4.3.1.1 Photosynthetic pigments content indices

According to the analysis above, the principal spectral differences were found, concentrating on the bands ranging in [420, 750nm], [1400, 1600nm] and [1800, 2000nm]. Considering the results of consistent test of the input data, [420, 750nm] is an ideal bands range. Furthermore, this range is photosynthetic bands which are much more sensitive to the external and internal stress to the plant. Through spectral changes here, mainly about the responses to photosynthetic pigment content, a number of information about the changes of biochemical and biophysical process would be revealed.

Photosynthetic pigments content related strongly to the photosynthetic potential of a plant, therefore it could be as an indicator of plant overall physiological state. The changes in these pigments content were sensitive to and indicative of stress of a plant. Therefore, by monitoring the changes of photosynthetic pigments, the slight response to the stress of a plant could be obtained. This stress involves light stress (photo-protection and other responses of plants to high light stress). Stress conditions of plant could be detected by measuring biophysical process as well as content of the photosynthetic pigments. Most stress factors, even if they do not directly affect the composition of the photosynthetic apparatus of its functions, will affect the photosynthetic process in long run(Lichtenthaler, 1996).

In the study, spectra reflectance data, laboratory chlorophyll content were collected. With these two kinds of data, vegetation indices (in table 2-5) involves photosynthesis would be applied to be as indictors of chlorophyll content, indirectly of capability of photosynthesis, to the stresses or influences.

Taking account of the result of consistent test of the input data, it should avoid the bands located in the non-normal distribution ranges (table 1) as much as possible to ensure the representativeness of the sample group. Many indices based on reflectance have been developed to retrieve and predict photosynthetic pigments content of plant at different scale. Retrieving pigment content based on spectral reflectance is a practical method which has been widely used for monitoring plant growth including vegetation stress. Compared with the studies on chlorophyll a and b, there were less works on carotenoids and anthocyanin content with spectral reflectance approaches. However, researchers developed spectral indices as indicator of these components. Thus, in this study, the author applied the previous achievements, spectral indices, as indicators to qualitatively analyze photosynthetic pigment content of the samples. These indices included edges of specific features, complicated indices which could avoid external disturbance or combine several unique spectral features. Considering the spectral resolution of the equipment, the edge position had been given up though it had a very good relationship with chlorophyll content.

Most of these indices listed had been assessed in the section 4-3 and proved sensitive. And for one foliar chemical, several indices were used to indicate its status. The redundant arrangement was for cross-validation to overcome flaws of individual spectral index.

Table 5-4-2. Spectral indices developed as chlorophyll (a+b) indicators (Bannari, et al., 2007b, Gitelson, A. A., et al., 2003):

ID	Indices Description	source
1	Red edge(R)	(Pu and Gong, 2000)
2	Red edge(area)	
3	Red absorption(R)	
4	Red absorption(area)	
5	Blue edge(R)	
6	Blue edge(area)	
7	Green peak(R)	
8	EGFN=(Dr-Rg)/(Dr+Rg) Dr: max(first derivate in red edge), Rg: max(first derivate in green peak)	(Treitz, et al., 1999)
9	Triangle Vegetation Index TVI=0.5|det(AB,AC)|=0.5*(120*(R_{nir}-R_{green})-200*(R_{red}-R_{green})) A=(550nm,Rgreen), B=(670nm,Rred), C=(750nm,Rnir)	(Broge, N. H. and Leblanc, E., 2001)
10	R87= (R_{800}–R_{700})/(R_{800}+R_{700})	(Gitelson and Merzlyak, 1994)

11	git03= $(R_{750-800})/(R_{695-740})$-1	(Gitelson, A. A., et al., 2003)
12	Chlorophyll Absorption in the Reflectance Index (CARI) CARI =$(R_{700}-R_{670})$-0.2*$(R_{700}-R_{550})$	(Kim, et al., 1994a)
13	Modified CARI (mCARI) mCARI =$[(R_{700}-R_{670})$-0.2*$(R_{700}-R_{550})]$*(R_{700}/R_{670})	(Daughtry, C. S. T., et al., 2000)

$R_{750-800}$= Average(Sum(Reflectance(750:800)));

RNIR=Average(Sum(Reflectance(700:750)));

R_{red}=Average(Sum(Reflectance(650:690))), the red absorption feature;

R_{green}=Average(Sum(Reflectance(510:560))) the green peak feature.

So do the other abbreviations in the follow tables.

Table 5-4-3. Spectral indices developed as carotenoids indicators (Ustin, et al., 2009):

ID	Indices Description	Source
14	Photochemical reflectance index (PRI) PRI=$(R_{531}-R_{570})/(R_{531}+R_{570})$	(Gamon, et al., 1992)
15	Chlorophyll Absorption Ration Index CARI2=CAR(R_{700}/R_{670}), CAR=|(a*670+R_{670}+b)|/(a^2+1)^0.5 a=$(R_{700}-R_{500})/150$, b=R_{550}-(a*550)	(Broge, N. H. and Leblanc, E., 2001)
	Carotenoid Reflectance Index (CRI)	(Gitelson, A. A., et al., 2002)
16	CRI550=$R^{-1}_{510}-R^{-1}_{550}$	
17	CRI700=$R^{-1}_{510}-R^{-1}_{700}$	
	Modified CRI (mCRI)	(Gitelson, A. A., et al., 2006)
18	mCRIgreen=$(R^{-1}_{510-520}-R^{-1}_{560-570})\times R_{NIR}$	
19	mCRIredge=$(R^{-1}_{510-520}-R^{-1}_{690-710})\times R_{NIR}$	
20	datt_car=$0.0049*[R_{672}/(R_{550}\times R_{708})]^{0.7488}$	(Datt, B., 1998)

Table 5-4-4. Spectral indices developed as anthocyanin indicators (Ustin, et al., 2009):

ID	Indices Description	source
21	Modified Anthocyanin mARI=$(R^{-1}_{530-570}-R^{-1}_{690-710})\times R_{NIR}$	(Gitelson, A. A., et al., 2006)
22	Red:Green Ratio RGR=R_{red}/R_{green}	(Gamon, J. A. and Surfus, J. S., 1999, Sims, D. A. and Gamon, J. A., 2002)

5.4.3.1.2 Foliar Water Content

Vegetation water content (VWC) is an important parameter to indicate components of plant water status. And help to infer water stress(Penuelas, et al., 1993) and assessment of drought conditions(Tucker, 1980). Detecting vegetation leaf water content using reflectance in the optical domain, however, because most bands which the VWC indices were calculated with dropped into non-normal distribution area, the results of these indices just play auxiliary roles for reference only.

Table 5-4-5 Spectral indices developed as water content indicators

ID	Indices Description	source
	Normalized difference water index (NDWI)	(Chen, et al., 2005, Zarco-Tejada, et
23	NDWI1240=(R860-R1240)/(R860+R1240)	al., 2003)
24	NDWI1640=(NIR858-SWIR1640)/(NIR858+SWIR1640)	
25	NDWI2130=(NIR858-SWIR2130)/(NIR858+SWIR2130)	

5.4.3.2 Qualitative analysis based on parameterized features

In this study, a baseline for reference was set. For the time series analysis within groups, the spectral indices calculated at -0h were set as baseline, and were subscribed by its counterparts at other time slices.

5.4.3.2.1 Dispersions of spectral indices within groups compared with -0h'status

The figure 5 showed the procedure of adaption of sample groups. Most of indices indicated that all transgenic samples had violent changes. The status at 168h also had the same situation including the contrast group. Because only the transgenic samples were sprayed herbicide, thus it could deduce that the changes after 168h were the one caused by the normal growth. It could be deduced by the change pattern of group P.

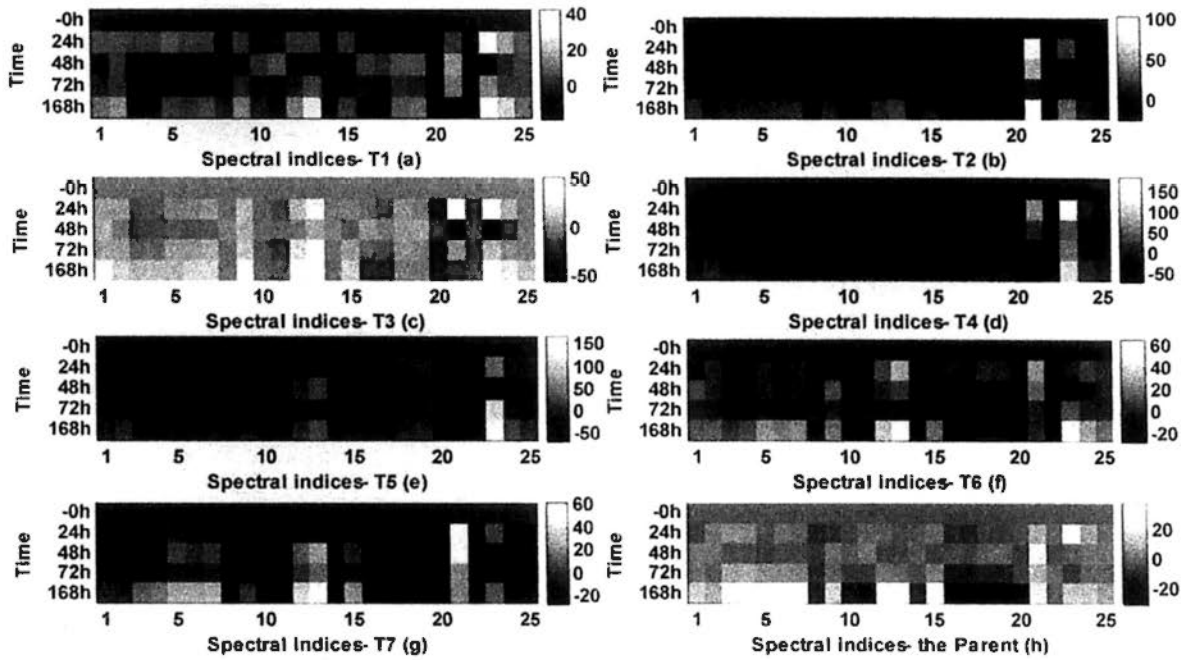


Figure 5-4-5. Dispersions of spectral indices within groups of time series (compared with -0h' which represents the status before herbicide sprayed, *100%)

X axis represents the spectral parameter, and the number relates the ID in the table 3-5, the same as the following figures.

5.4.3.2.2 Dispersions of spectral indices within groups compared with the contrast samples

In the figure 6, the contrast group was set as reference, thus the growth status compared with its parent could be obtained. It is helpful to assess sample growth.

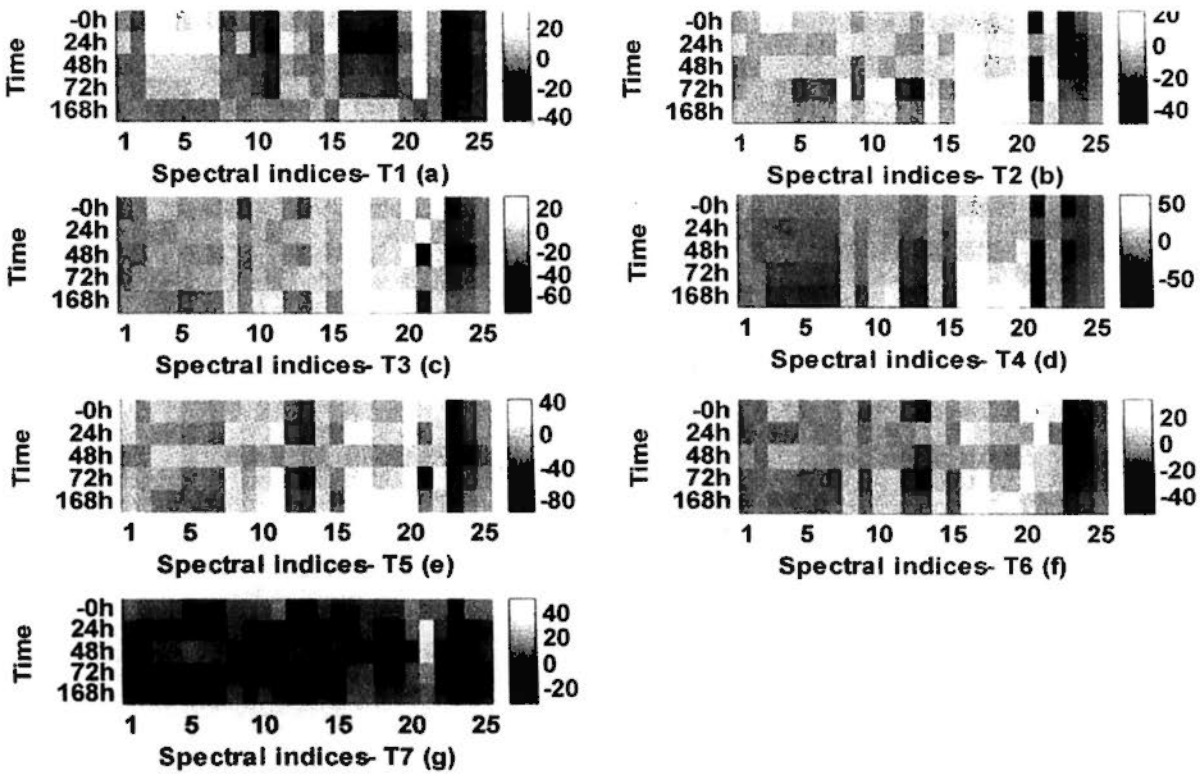


Figure 5-4-6. Dispersions of spectral indices within groups (compared with the contrast samples, *100%)

5.4.3.2.3 Dispersions of spectral indices among groups compared with the contrast samples

Figure 7 gave a comparison among transgenic samples compared with the contrast. The responding differences to herbicide of transgenic samples were shown.

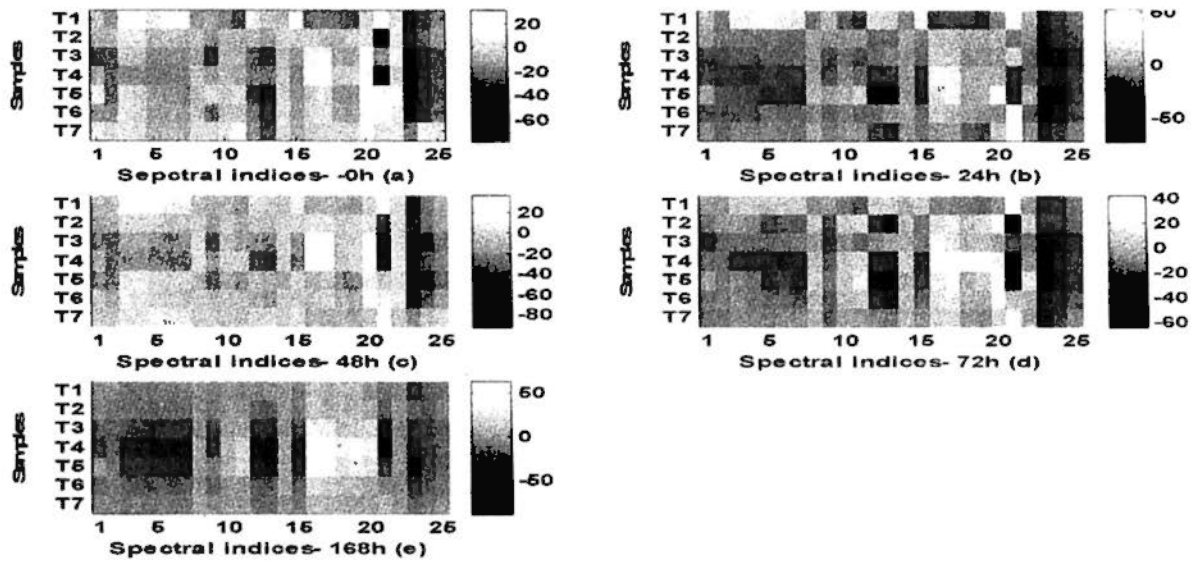


Figure 5-4-7. Dispersions of spectral indices among groups (compared with the contrast samples, *100%)

In this section, the variable vector, chemical change (cc) was re-defined as

cc=[V(parameter$_{1)}$,V(parameter$_2$)...,V(parameter$_{25}$)].

Where V(parameter$_i$)represents the variable decided by the i^{th} spectral indices in the table 3-5. Thus, a comprehensive index was built to describe the sample growth. Here, V (parameters) was the spectral indices of transgenic groups subscribed the one of the contrast group.

In the table 6, before spraying herbicide, the growth statuses of most of transgenic groups were related to each other except for T1, T2 and T7. It indicated the samples had similar growth status. Moreover, because the coefficients were not very high, the samples grew different slightly. T1 and T7 were distinguished from the other groups, especially T1, totally different with the others. After spraying herbicide, T1, T2 and 7 began growing similar to others. Going with time passing, the growth statuses of the samples were more and more strongly related to each other. After 168 hours the growth statuses were all highly related. Herbicide stress made the transgenic sample to grow consistently. It was an important finding.

Table 5-4-6. The correlation coefficients among transgenic groups

	T1	T2	T3	T4	T5	T6	T7
	-0h						
T1	1						
T2	.076	1					

T3	.223	.595	1				
T4	-.085	.838	.739	1			
T5	.427	.403	.779	.638	1		
T6	.451	.376	.908	.549	.909	1	
T7	-.087	.113	.430	.310	.667	.571	1
	24h						
T1	1						
T2	-.186	1					
T3	-.012	.777	1				
T4	-.587	.843	.712	1			
T5	-.290	.934	.588	.787	1		
T6	.121	.661	.916	.496	.516	1	
T7	.233	.374	.565	.158	.380	.683	1
	48h						
T1	1						
T2	.340	1					
T3	.213	.928	1				
T4	-.098	.749	.913	1			
T5	.850	.726	.677	.394	1		
T6	.786	.605	.657	.482	.926	1	
T7	.834	-.174	-.213	-.448	.519	.516	1
	72h						
T1	1						
T2	-.560	1					
T3	.056	.562	1				
T4	-.564	.903	.734	1			
T5	-.446	.959	.548	.873	1		
T6	-.036	.716	.631	.657	.740	1	
T7	.401	.266	.618	.304	.343	.818	1
	168h						
T1	1						
T2	.854	1					
T3	.797	.981	1				
T4	.744	.947	.977	1			
T5	.807	.947	.934	.904	1		

T6	.845	.915	.905	.845	.923	1	
T7	.672	.722	.684	.575	.824	.862	1

The differences mainly concentrated at photosynthetic and water sensitive bands, the significant difference test was conducted to find out the details. The input data were the time series of relevant chemicals' indicators. Table 7 listed the significant difference test results between transgenic sample groups, 0 represents no significant difference existing in the groups, while 1 is the opposite. Because of lack of data to evaluate the anthocyanin and water content data expressed by mARI and NDWI, thus the results about these two chemicals were just for reference. The results showed that T1, with the other groups, had significant differences in chlorophyll, the T2, T6 and T7 were in Nitrogen, and T5, T6 and T7 were in Lignin.

Table 5-4-7. The significant test of the difference between T4 and P (α=0.05)

	T1	T2	T3	T4	T5	T6	T7
	Chlorophyll						
T1	0						
T2	1	0					
T3	1	0	0				
T4	1	1	0	0			
T5	1	1	0	0	0		
T6	1	0	0	1	0	0	
T7	1	0	0	1	1	0	0
	Carotenoids						
T1	0						
T2	0	0					
T3	0	0	0				
T4	0	0	0	0			
T5	0	0	0	0	0		
T6	0	0	0	0	0	0	
T7	0	0	0	0	0	1	0
	Anthocyanin						
T1	0						
T2	1	0					
T3	1	0	0				
T4	1	0	0	0			
T5	1	0	0	0	0		

T6	0	1	1	1	0	0	
T7	0	1	1	1	1	0	0
	Water						
T1	0						
T2	0						
T3	0	0	0				
T4	0	0	0	0			
T5	1	1	0	1	0		
T6	0	0	0	0	1	0	
T7	0	0	0	1	1	1	0
	Nitrogen						
T1	0						
T2	0	0					
T3	0	0	0				
T4	1	0	0	0			
T5	0	0	0	0	0		
T6	1	1	1	1	1	0	
T7	0	1	1	1	1	1	0
	Lignin						
T1	0						
T2	0	0					
T3	1	1	0				
T4	1	1	0	0			
T5	0	1	1	0	0		
T6	1	1	1	1	1	0	
T7	1	1	1	1	1	1	0

By comprehensive analysis of the tables and figures, especially for table 6 and 7, the time when the significant difference of the chemical content happened could be confirmed.

5.4.4 Conclusion

In this section, the samples transferred Bar gene for resisting herbicide were chosen. For evaluating the transgenic samples' performance to herbicide resistance, an artificial induction experiment was conducted. Except for the contrast group, all others were sprayed herbicide of 0.3% Basta. Post-24h-spayed spectral reflectance data, post-48h, post-72h and post-168h data were acquired. When removing the outliers by band, the mean spectra were calculated to

represent the sample group. Two kinds of analysis were done to assess the samples, namely quantitative analysis based on spectral morphological characteristics and qualitative analysis based on parameterized features within groups and among groups. Spectral angles among samples showed that T7 & T1, T7 &T3, T1 &T5, T5 & T7, T4&T1 had larger angles than others, while T2 showed much similarity of spectra to the parent. In the following work, T1, T4, T5 and T7 would be focused. By investigating the continuum removal results, it could be found that differences mainly located in range of [420,750nm], [1400, 1600nm] and [1800, 2000nm]. And T1 at the five time slices had a great of positive shifts dispersions in the continuum removal figure. The results of the correlation analysis and significant difference test also qualitatively proved the previous founds. T1, T5 and T7 had distinguished differences to the others under herbicide stress. When after 168 hours the growth statuses were all highly related. Herbicide stress made the transgenic sample to grow consistently. It was an important finding. All these results would be good support for professional breeding study.

Reference

Bannari, A., Khurshid, K. S., Staenz, K., & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *Ieee Transactions on Geoscience and Remote Sensing*, 45(10), 3063-3074.

Boochs, F., Kupfer, G., Dockter, K., & K¨uhbauch, W. (1990). Shape of the red edge as vitality indicator for plants. *International Journal of Remote Sensing*, 11(10), 1741-1753.

Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2), 156-172.

Carter, G., & Knapp, A. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684.

Chen, D., Huang, J., & Jackson, T. (2005). Vegetation water content estimation for corn and soybeans using spectral indices derived from modis near-and short-wave infrared bands. *Remote Sensing of Environment*, 98(2-3), 225-236.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329-6340.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111-121.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229-239.

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal

changes in photosynthetic efficiency. *Remote sensing of environment*, 41, 35-44.

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer *New Phytologist*, 143(1), 105-117.

Gates, D. (1965). Energy, plants, and ecology. *Ecology*, 46(1), 1-13.

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271-282.

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11), L11402.

Gitelson, A. A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra - experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B-Biology*, 22(3), 247-252.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3), 272-281.

Kim, M., Daughtry, C., Chappelle, E., McMurtrey III, J., & Walthall, C. (1994). The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation. Proceedings of the ISPRS Sixth International Colloquium on Physical Measurements and Signatures in Remote Sensing; 1994 17- 21 January 1994; Val d'Isere, France: European Space Agency; 1994, 299-306.

Lichtenthaler, H. K. (1996). Vegetation stress: An introduction to the stress concept in plants. 148(1), 4-14.

Penuelas, J., Filella, I., Biel, C., Serrano, L., & Save, R. (1993). The reflectance at the 950-970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*, 14(10), 1887-1905.

Pu, R., & Gong, P. (2000). Hyperspectral remote sensing and its applications. *Higher Education, Beijing*.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2-3), 337-354.

Sohn, Y., & Rebello, N. S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing*, 68(12), 1271-1282.

Tong, Q., Zhang, B., & Zheng, L., (2006). *Hyperspectral remote sensing*. Beijing: Higher Education Press.

Treitz, P. M., & Howarth, P. J. (1999). Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems. *Progress in Physical Geography*, 23(3), 359-390.

Tucker, C. J. (1980). Remote sensing of leaf water content in the near infrared. 10, 23-32.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution

spectroscopy. *Remote Sensing of Environment*, 113, 67-77.

Wessman, C., Aber, J., Peterson, D., & Melillo, J. (1988). Foliar analysis using near infrared reflectance spectroscopy. *Canadian journal of forest research(Print)*, 18(1), 6-11.

Zarco-Tejada, P., Rueda, C., & Ustin, S. (2003). Water content estimation in vegetation with modis reflectance data and model inversion methods. *Remote Sensing of Environment*, 85(1), 109-124.

5.5 Pre-test of difference analysis within single sample group --a case of individual sample screening

Mutation is a very important event to evolutions. The positive mutation is what we want to, while the negative one is what we want to avoid. When gene was transferred into the receptor, if it only made sense to the target organism and bio-process, or if byproducts or mutations caused by this foreign gene existed, are what researchers are interested in. By spectral detection approaches at individual sample level, through the mutation or byproducts could not be found directly, however, it could found the spectral differences between these samples. It could be a kind of efficient pre-selection method for huge amount of sample screening. It is also an important complementarily when mean spectral was applied. Analysis of mean spectra of a sample group aims to the difference at a class level. It would be fulfill the assessment of the status and characteristics of the whole group. Analysis of individual spectra of a sample aims to single spectrum of sample. It is an assessment feature of individual sample. Both a sample class with stable characteristics and a single sample with unique individual superior features are wanted for study and breeding.

Therefore, in this section, analysis based on original individual spectra in one sample group was conducted. It was hoped to answer two questions if any differences existed only in a specific individual sample and if taking mean spectrum as representative of the whole group was reasonable.

5.5.1 Outlier detection based on Grubbs Test

Outlier is defined as an observation that deviates too much from other observations that it arouses suspicions that it was generated by a different mechanism from other observations(Hautamaki, et al., 2004, Hawkins, 1980).

In this study, Grubbs Test(Baksalary, et al., 1990, Zhang, et al., 2010) was applied. The theory of Grubbs Test:

Supposing a set residue of repeated observations, which were sorted by its values: $V(1) \leqslant V(2) \leqslant \cdots \leqslant V(n)$. Grubbs statistic could be calculated according to equation (1):

$$g_n = \frac{V_{(n)} - \overline{V}}{\sigma} \qquad g_n = \frac{\overline{V} - V_{(1)}}{\sigma} \qquad (5\text{-}5\text{-}1)$$

In the formula, $\overline{V} = \frac{1}{n}\sum_{i=1}^{n} V_{(i)}$ σ is standard deviation g_n is a key value to judge if the maxima was outlier, while g_n' to judge whether the minima is outlier. g_n and g_n' have same probability distribution. $g_0(n,\alpha)$ shows the Critical value under the significant α which could be obtained from look-up table.

When $|g_n| \geq g_0(n,\alpha)$ or $|g_n'| \geq g_0(n,\alpha)$, the counterpart of the residue is error, and regards the maxima (or minima) is outlier. The significant α is 0.005 or 0.01 general.

The input data of Grubbs Test need to be normal distribution, thus a pre-test were data first by Lilliefors' composite goodness-of-fit test. Then the Grubbs Test was conducted by bands for outlier detection. Defined the band vector Input as (2):

$$Input_i = [spec_i(1), spec_i(2), \ldots spec_i(j)] \quad (5\text{-}5\text{-}2)$$

Where $Input_i$ represents the vector at the band i, $spec_i(k)$ is the spectral reflectance at the band i of the k^{th} measurements, k ranges from 1 to j.

Normality test were done by Lilliefors' composite goodness-of-fit test, while outlier were detected by *"deleteoutliers.m"* developed by Jaco de Groot. These tests were all taken under Matlab software environment.

The Input data were the ones of T4 at 48h in section 5.3. They were found with lots of outliers. The results of Lilliefors' composite goodness-of-fit test showed most band vectors were fitted to normal distribution (section 5.3 table 1). The results of Grubbs Test by band showed that there were 1139 outliers among the 8 samples, and the outliers concentrated in the spectrum 4, 5 and 6. The outliers were located in the wavelength of 350-450nm, 660-730nm and 1360-2450nm. Since 350-420nm and 2400-2500nm are noise sensitive bands of the spectrometers. Thus, it was hard to confirm whether outliers in these two ranges were the one caused by the equipment or samples. Thus, the two ranges were excluded when the data applied. Therefore, spectrum 4 could be treated as normal one. The range of 660-730 nm is photosynthetic responding bands while 1360-2450nm is sensitive bands to water and nutritional material.

Table 5-5-1. The outliers in the samples

Sample	Number of samples	Number of outliers	Outlier samples	Outliers
5.3-48h-T4	8	1139	4(20)	409-412,419-420,429-431,

		433-441
	5(164)	350-351,353-361,364-497,660-678
	6(955)	368,716-726,1369-2451

When spectrum 5 and 6 were excluded, this sample group could be expressed by its mean spectrum. However, the test was conducted by individual band, the whole curve of the spectra also needed to check, thus, the approaches of spectral angle and continuum removal were applied to describe the shape of the spectra.

5.5.2 Analysis based on morphological features

For spectral reflectance vectors t and r sorted by band, thus the spectral angle θ of t and r could be defined by equaiton(3)(Sohn and Rebello, 2002).

$$\theta_{t,r} = \cos^{-1}\left(\frac{\vec{t}\cdot\vec{r}}{\|t\|\cdot\|r\|}\right) \quad \theta \in \left[0, \frac{\pi}{2}\right] \quad (5\text{-}5\text{-}3)$$

Table 5-5-2. The spectral angles of sample spectra (in degree)

	S1	S2	S3	S4	S5	S6	S7	S8
S1	0.00							
S2	3.63	0.00						
S3	2.22	3.62	0.00					
S4	3.28	6.14	2.91	0.00				
S5	2.34	5.64	2.95	1.41	0.00			
S6	2.91	1.85	2.95	5.36	4.76	0.00		
S7	1.92	4.68	2.03	1.69	1.33	4.04	0.00	
S8	2.09	5.02	2.26	1.38	1.01	4.32	0.38	0.00

The largest spectral angle was the one between S2and S4, up to 6.14° in table 2. S2 and S5 also had a relative bigger angle. The angel between S4 and S6 was the third biggest, up to 5.36° . The table showed an interesting result, that spectrum 2 the one more distinguished to the other spectra rather than spectrum 4, 5 and 6 which had outliers. Therefore, spectrum 2 should be checked too. However, though the largest spectral angle was up to 6.14° , it was still a very small value which indicated that only slight differences between the two samples.

Continuum removal is a useful transform to amplify the slight features of spectrum, and isolate features of interest (Clark and Roush, 1984). By this transform, the data were projected to a space where they could be compared each other directly associated with wavelength in an intuitive way.

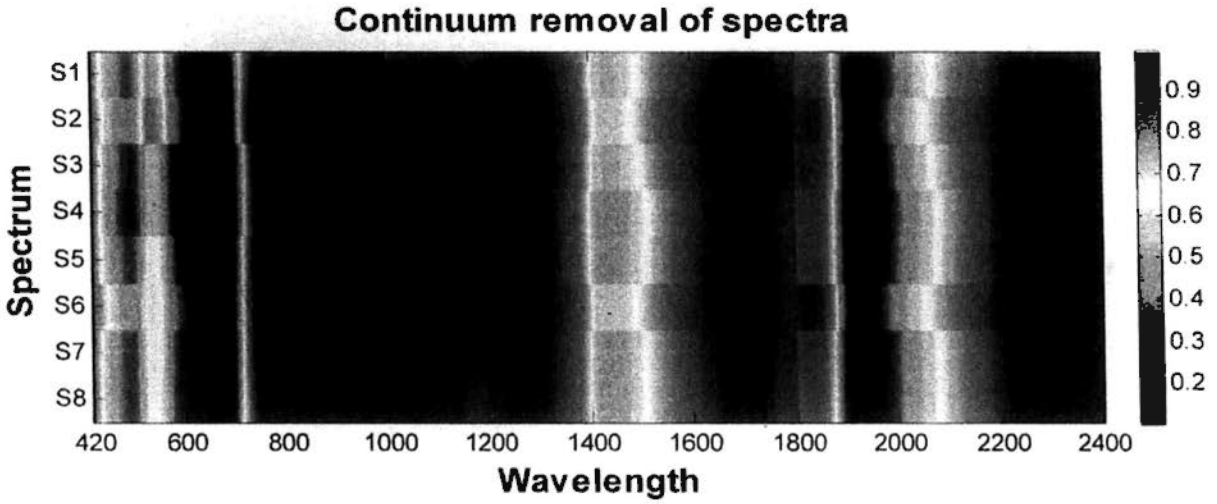


Figure 5-5-1. The spectral continuum removal image

By the transform of continuum removal, the values of reflectance data have been projected to [0, 1]. Thus the differences between groups could be shown by the shifts of color (shifts of absorption or reflectance peak) and the changes of color (intensity of absorption or reflectance) at certain band. The figure revealed that the shifts located around 600nm, 1400nm and 2000nm. Furthermore, at these shift ranges, the intensity of absorption or reflectance was also of significant difference. The first shift bands were sensitive to photosynthetic pigments; the left ones were water sensitive bands. Because of heat effects caused by the light of the contact probe, the water content in leaf would change slightly. However it caused shift at water sensitive bands. In table 1, S1 an S2 had same features, S3 and S4 were more the same, S5-S8 were similar in general. S2 and S6 were much distinguished than the others. The two spectra had obvious shifts of color and changes of color compared with the others. S3 and S4 also were discriminated in 420- 800nm. After 800nm they were more similar with the others. Spectrum 2, 3 and 6 were more different than the others after 1400nm. The shifts around 1400nm and 2000nm were possible caused by heat effect of the probe light. In general, the shifts at the mentioned bands were not significant.

However, S2, S3, S4, S5 and S6 still should be checked in the quantitative analysis.

5.5.3 Growth status analysis based on statistic methods

Here, the description vector cc (section 5.3) was applied to indicate individual sample growth (the chemicals in leaf). The correlation coefficients of the cc within samples were highly consistent to the results of spectral angle. Therefore, it could check significant difference of the cc within sample to indicate the samples' growth status. Two independent sample t-test was applied. This test performs a T-test of the hypothesis that two independent samples, in the vectors S1and S2, come from distributions with equal means, namely to test if significant difference existing within samples. All sig. values were much higher than 0.01, namely there was no significant difference

existing within the spectra of the samples (table 3).

Table 5-5-3. The significant test (sig.) of the difference (α=0.01)

	S1	S2	S3	S4	S5	S6	S7	S8
S1	0.00							
S2	0.97	0.00						
S3	0.82	0.86	0.00					
S4	0.86	0.90	0.96	0.00				
S5	0.96	0.99	0.86	0.90	0.00			
S6	0.92	0.95	0.92	0.96	0.96	0.00		
S7	0.92	0.95	0.90	0.94	0.96	0.99	0.00	
S8	0.93	0.96	0.90	0.93	0.97	0.99	0.99	0.00

The correlation analysis among spectra were done (table 3). All sample spectra were highly related to each other. However, Spectrum 6 was only slightly different (in bold in table 3). All spectra were consistent. No distinguished outlier (mutation) existed in this sample group.

Table 5-5-4. The correlation coefficients among spectra

	S1	S2	S3	S4	S5	S6	S7	S8
S1	1							
S2	.967	1						
S3	.997	.983	1					
S4	.983	.906	.968	1				
S5	.989	.928	.979	.996	1			
S6	.888	.971	.918	.799	.838	1		
S7	.996	.952	.990	.990	.997	.875	1	
S8	.995	.948	.989	.992	.998	.868	1.000	1

5.5.4 Conclusion

In this section the original individual spectra of samples, instead of the mean spectrum of a group, were checked and analyzed. By outlier detection test, outlier was found. Based on spectral morphological features, the angles and the continuum removal of spectra were calculated and compared. The angles were small which revealed that the samples were close to each other in spectrum. The continuum removal showed the factors mainly trivial changes of the water content causing slight angles. The water content change may caused by the light of contact probe which heated the leaf. Statistic analysis found no significant difference among samples. All these results proved that (1) no significant differences existed in this sample group and (2) applying mean spectrum to represent this sample group was reasonable.

Reference

Baksalary, J. K., & Puntanen, S. (1990). A complete solution to the problem of robustness of grubbs's test. *Canadian Journal of Statistics*, 18(3), 285-287.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications *Journal of Geophysical Research*, 89(B7), 6329-6340.

Hautamaki, V., Karkkainen, I., & Franti, P. (2004). Outlier detection using k-nearest neighbour graph 3(430-433 Vol.433.

Hawkins, D. (1980). Identi cation of outliers.

Sohn, Y., & Rebello, N. S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing*, 68(12), 1271-1282.

Zhang, X., Li, R., Yue, Y., Liu, B., & Liu, H. (2010). Improved algorithm for reconstructing vegetation index image time series based on fourier harmonic analysis. *JOURNAL OF REMOTE SENSING* 14(3), 1-5.

Chapter 6 Data process and analysis system

In chapter 5, the exploratory research works demonstrated that the proposed approach was able to monitor growth of the transgenic samples and helping to screen samples from a macro view scope. By the approach, the samples were assessed successfully both from spectral morphological features and parameterized characteristics (foliar chemicals), and long time growth tracks also were extracted. All these were important to assist the cultivation (e.g. breeding) to know the growth status of the samples clearly.

In this chapter, based on the achievements of the previous content, a system of data process and analysis is developed.

6.1 Development environment

MATLAB is a numerical computing environment developed by MathWorks. It is a fourth generation programming language. It allows "*matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages. It is widely used in academic and research institutions as well as industrial enterprises*"[23]. It has simple syntax and is user friendly. It has graphic user interface programming design. GUIDE (figure1), the MATLAB graphical user interface development environment, provides a set of tools for creating graphical user interfaces (GUIs). These tools

[23] http://en.wikipedia.org/wiki/Matlab

greatly simplify the process of designing and building GUIs[24]. Therefore, under GUIDE of MATLAB environment of Version 7.0.1 (R14) Service Pack1, the coding was designed, compiled and implemented.

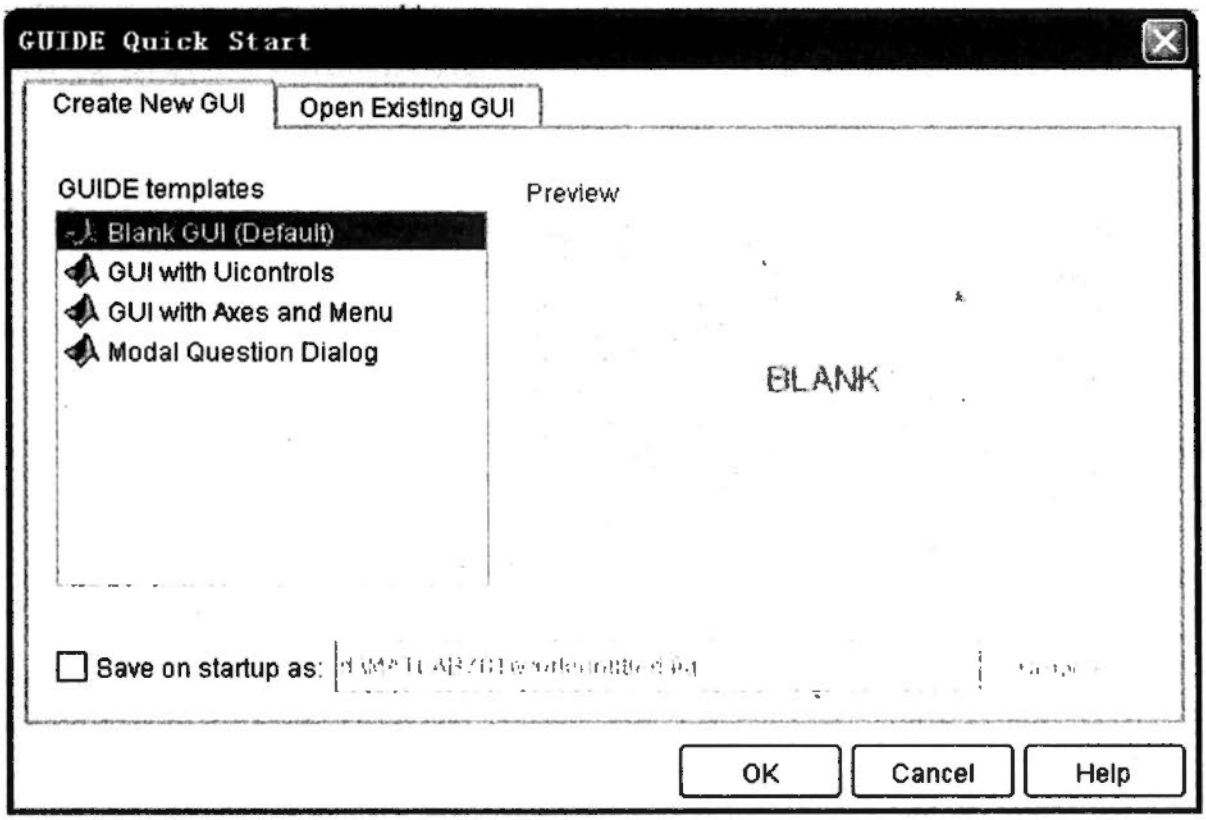


Figure 6-1. GUIDE: the MATLAB graphical user interface development environment

6.2 system structure design

The system should have mainly two modules, data process and results analysis. For providing a simple and handy window interface, the first kind of modules is designed to implement from menu while the second module were implemented on the window interface directly.

6.2.1 Data process modules

Data process module should contain basic data process functions, including:

1) To pre-process raw DN spectral data;
2) To convert original spectral data from raw binary to text file including raw DN data and raw reflectance data;
3) To calculate Reflectance data (for spectral DN data);
4) To draw Spectral curve (for single spectrum);
5) To filter the noise in spectrum (spectra);
6) To calculate the first-order derivative of spectrum (spectra);

[24] MATLAB help: What Is GUIDE?

7) To calculate the second-order derivative of spectrum (spectra);
8) To implement the continuum removal of spectrum (spectra);
9) To calculate the spectral angles;
10) To calculate the parameterized spectral features (spectral indices as indicators of foliar chemicals);
11) To check spectrum containing outliers detected by outlier detect modules;

In these kinds of modules, except for spectral-curve-drawing and check-spectrum-with-outliers, the others could solve both single spectrum and hands of spectra. By running these modules, the functions of data process could be implemented separately. It is facilitated to general data process and analysis rather than the process for the proposed approach.

Moreover, in these modules, an important function would be implemented, that is pre-process raw DN spectral data. The so-called *"pre-process raw DN spectral data"* represents to remove the individual spectrum which is obviously different from the others and calculate the mean of the rest as the spectrum of a measurement. For ASD spectrometer, to suppress noises, a measurement more than 1 spectrum would be acquired. User could decide how many spectra to be collected. The more the number is, the longer the measurement needs. Thus, for application of spectra, the first step is to pre-process data by manual evaluation to get the spectra of measurements.

Generally, for a measurement, ten spectra will be obtained, and for an experiment, hundreds of measurements would be conducted (figure 2). With tandem mass spectra, manpower to complete data pre-process is time-consuming, laborious. For spectra in a measurement, they are very similar to each other. In an ideal case, they should be the same. However because of external influences, such as changes of illumination condition and leaf water content, they are just close to each other rather than the same. When these differences detected by bands are not bigger than the threshold, they could be treated as the same and the mean spectrum could be calculated. In the data process system, an integrated approach for data pre-process based on Inner-clustering Coefficient (coefficient of variation, cv) would be applied to fulfill this task. When cv of at a wavelength was bigger than the threshold, the data at this band (e.g. i^{th} band) would be conducted to outlier detect.

Define R(j) as the reflectance of the j^{th} sample at i^{th} band, avR(j)=mean(R(1),..,R(j)). When abnormal data are found at i^{th} band, for j^{th} sample, if $|avR(k)-avR(j)|>avR(j)*cv$ (k ranged from 1 to sample number), then set the (k,j) was abnormal. When the number of abnormal data for the j^{th} sample is more than the normal, then set the reflectance of the j^{th} sample is abnormal. When all reflectance had been checked follow this method, if the numbers of the abnormal is bigger than half of total samples, convert the abnormal and normal flag. If abnormal reflectance of a spectrum is more then 100, this spectrum was treated as abnormal and removed when mean

spectrum i calculated.

By this algorithm, most problems would be solved, except for the case in figure 3. In practice, to solve the problem showed in figure 3, it was to choose relatively similar spectra to calculate the mean and there was no ideal method. Thus, the solution of the algorithm is acceptable.

When this module run, the user selects the directory of raw data saved, inputs the number of spectra for one time measurement and the threshold of cv, the module could fulfill the data pre-process automatically.

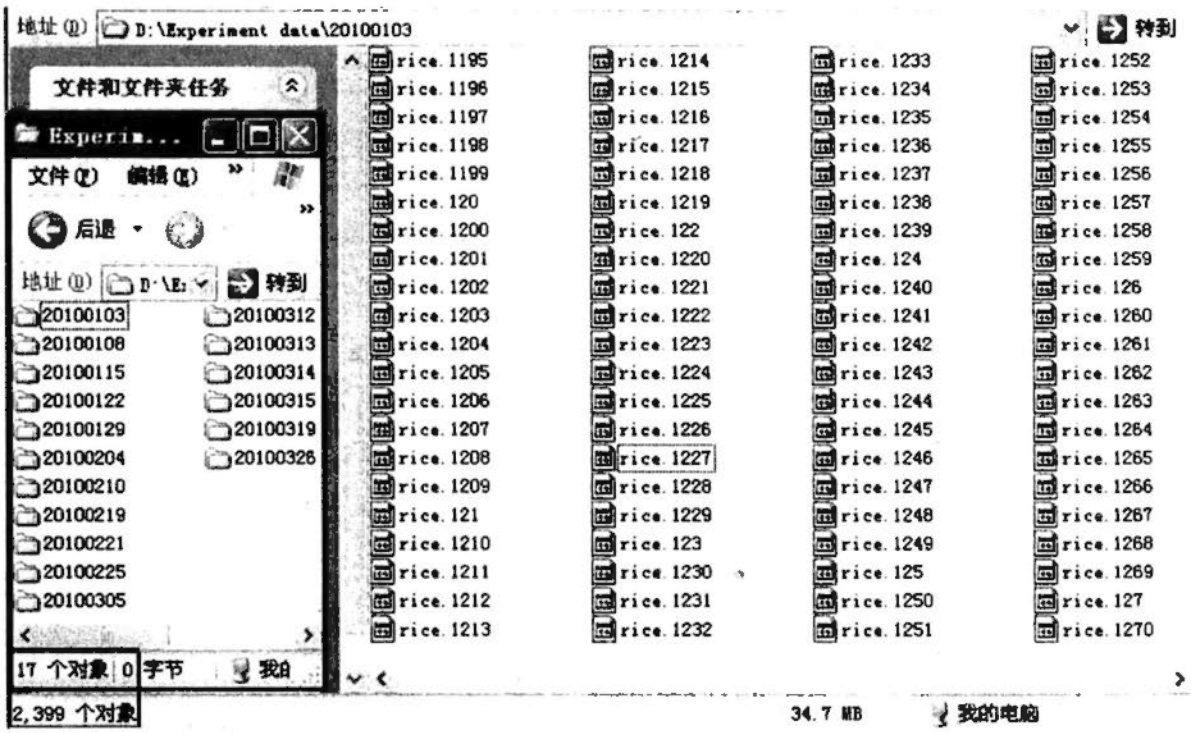


Figure 6-2. Input data for data pre-process: mass raw spectral data for data pre-process

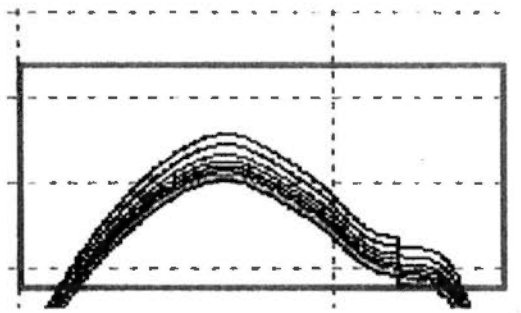


Figure 6-3. A special case for calculating the mean spectrum

6.2.2 Data analysis modules

In section 6.2.1 the modules were introduced for implementation functions separately. These modules could be applied for common spectral analysis and parameter calculation. In this section, an integrated data process and analysis modules were designed to support the approach proposed. Mainly it was consist of integrated data process module, outlier detection module, analysis of features among groups and analysis of sample growth trends.

6.2.2.1 Integrated data process module

The input data for this module would be a directory containing spectral files in the format of '*.txt'. These spectral files should be organized. The same spectra of the same group should be put into a folder. The module could process a folder representing the data of a group or of one experiment. It also could solve a folder containing sub-folders. In this case, the previous folder represents an experiment and the sub folders means different samples groups. In figure 4 9311 represents sample group 9311, while 20100305txt represents an experiment, it contains lots of samples groups, such as 730-1, 730-2. Both 9311 and 20100305txt could be accepted.

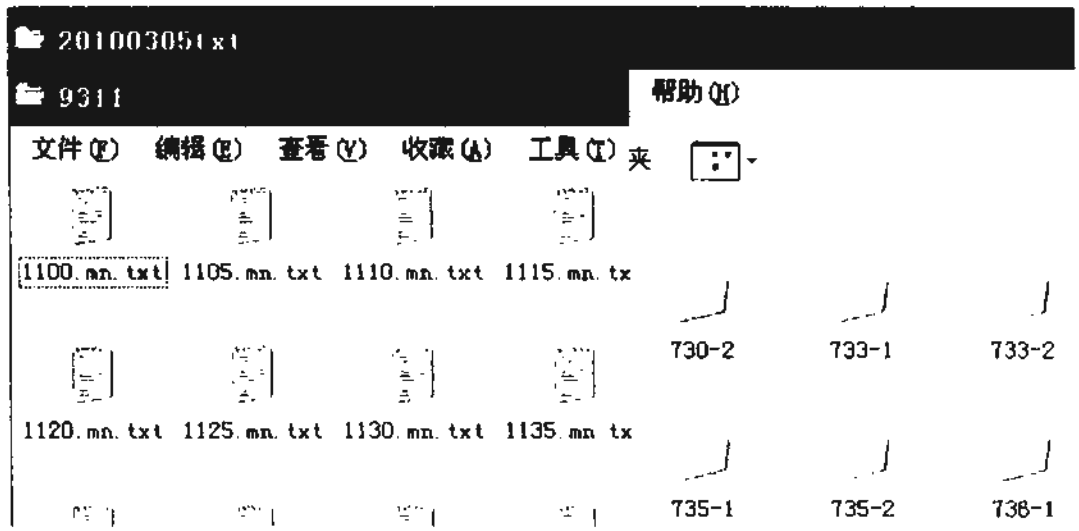


Figure 6-4. Input directory for data pre-process

When the process ended, a log (figure 5a) would be generated to record all data processed. Spectra of the same group would be put together in an Excel file named as group+'.xls', and the information of mean spectra of groups would given, named as group+ '_mean.xls' (figure 5b). All of mean spectra should be gathered in a file, named as experiment+'_total.xls', and the spectral parameters obtained based on mean spectra of group should be calculated and put in a file, named as experiment+'_para.xls'. '_total.xls' and '_para.xls' are also important labels for data reading in following modules. In case of a group, the spectral parameters would be calculated based on original spectrum.

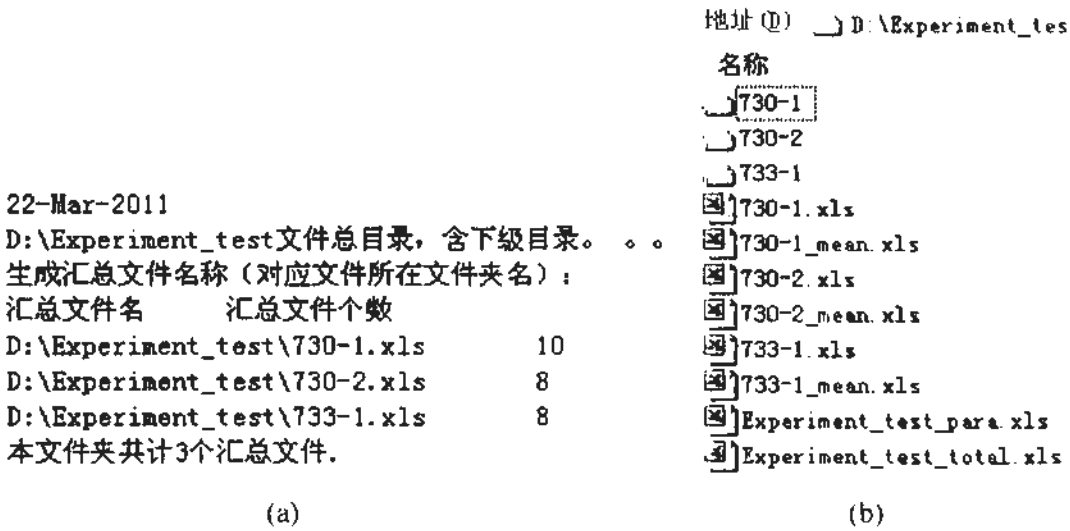


Figure 6-5. The log file (a) and the results (b) of integrated data process module

6.2.2.2 Outlier within group detection module

In this module, the features of individual spectrum of single sample rather than mean spectra on groups' level are focused. Outliers of reflectance will be detected by bands based on Grubbs' test (details in section 5.5). The results indicate the total numbers of outliers of the sample group, the sign of the spectrum containing outliers, the numbers of outliers this spectrum and the wavelength of the outliers locating. After detection, the spectra with outliers will be copied to specific folder for check. And the message of outliers should also be sent to 6.1.1-11) module for user checking. Moreover, a report of outliers will be made. The records in the report would be formatted as follow:

**&&@@##Total of outliers%$* Specific spectrum with outliers& numbers of the outliers in the spectrum@ specific wavelength of the outliers

This module was designed based on section 5.5.

6.2.2.3 Analysis of features among groups

This module aims to single experiment. The input data is the results of integrated data process module, mainly spectral parameters and mean spectra. Based on the two kinds of input data, the spectral edges (e.g. red edge), spectral angles among samples, continuum removal of spectra, information of 8 selected foliar chemicals were compared. Moreover, statistic test also would be conducted to find if any significant difference among samples.

These analyses are based on section 5.1, 5.2 and 5.4. When the process ended, the figures about the edges, 8 selected foliar chemicals, continuum removal of spectrum, spectral angles would be obtained and saved. Meanwhile, the figures about the changes of the parameters mentioned

above compared with the contrast were also obtained. Finally two sample significant difference tests were conducted and the results showed as figures.

6.2.2.4 Analysis of sample growth trends

If time series data of the sample is available, this module will be implemented. The input data would be the same 6.2.2.3's for a time point. The time series should at least contain 3 experiments' data, and '*_para.xls' and '*_total.xls' files should be matched strictly.

This module would be implemented based on 5.1, 5.2, 5.3 and 5.4. The figures of change trends of the edges in time series, chemicals and the figure of continuum removal of spectra and spectral angle among samples were obtained. Then the figures of change trends of parameters mentioned above compared with the data of the first time points (within groups) and with the contrast group (among groups) were obtained. At last the statistic results within and among groups of the parameters changes were obtained.

6.2.3 The interface of the system

6.2.3.1 Menu

Totally there are 7 first level menu items containing the sub-menu to fulfill the functions in 6.2.1 (figure 6).

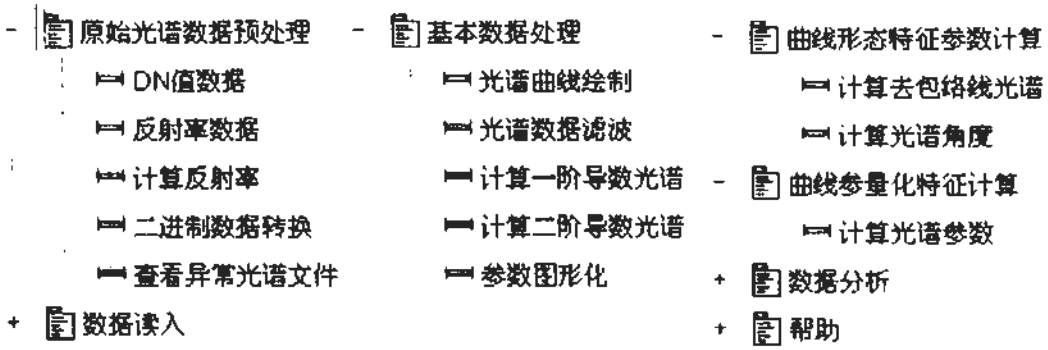


Figure 6-6. Menu of the system

6.2.3.2 Windows buttons

The modules of 6.2.2 fulfills the data process and analysis function of proposed approach, thus they will put on the main window (figure 7). Figure 8 and 9 showed the results (figure generated and saved automatically) of the 6.2.2.4 and 6.2.2.3.

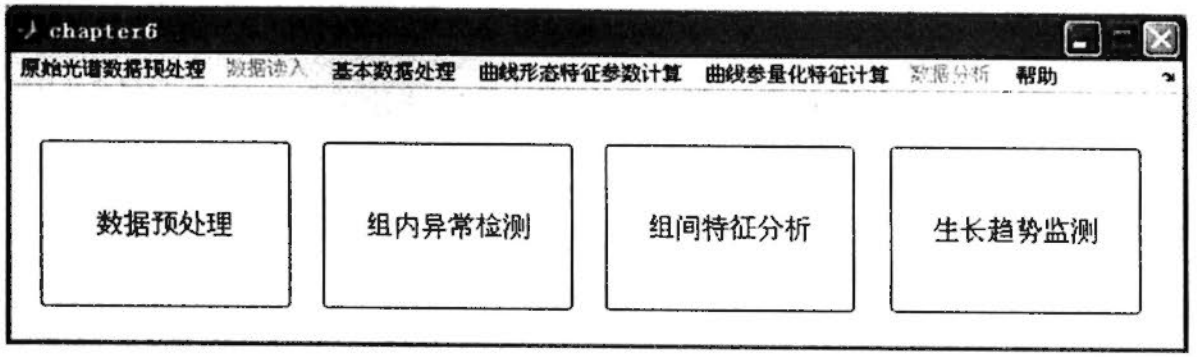


Figure 6-7. User interface window

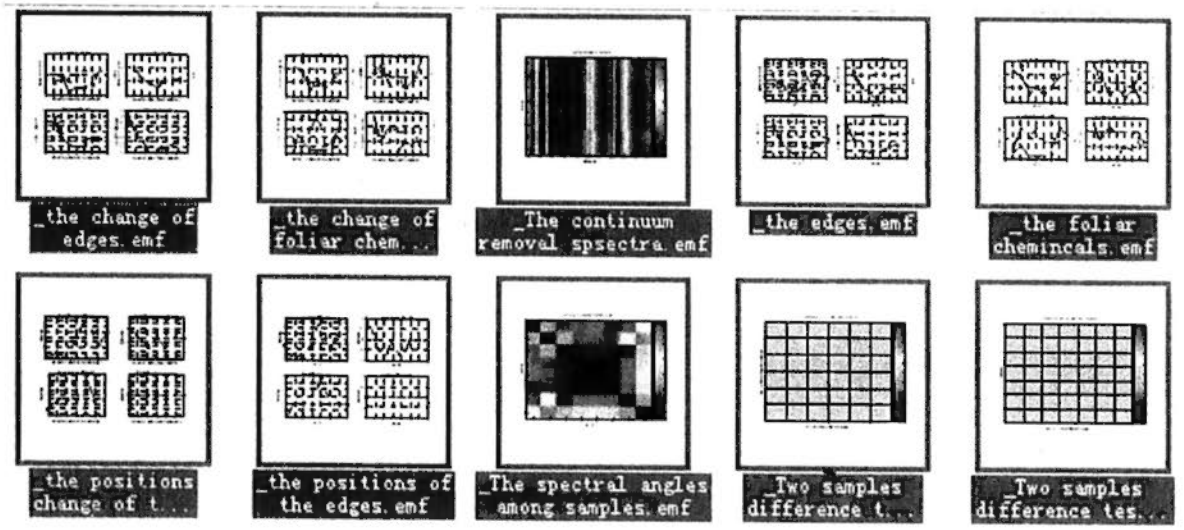


Figure 6-8. The results of "Analysis of features among groups"

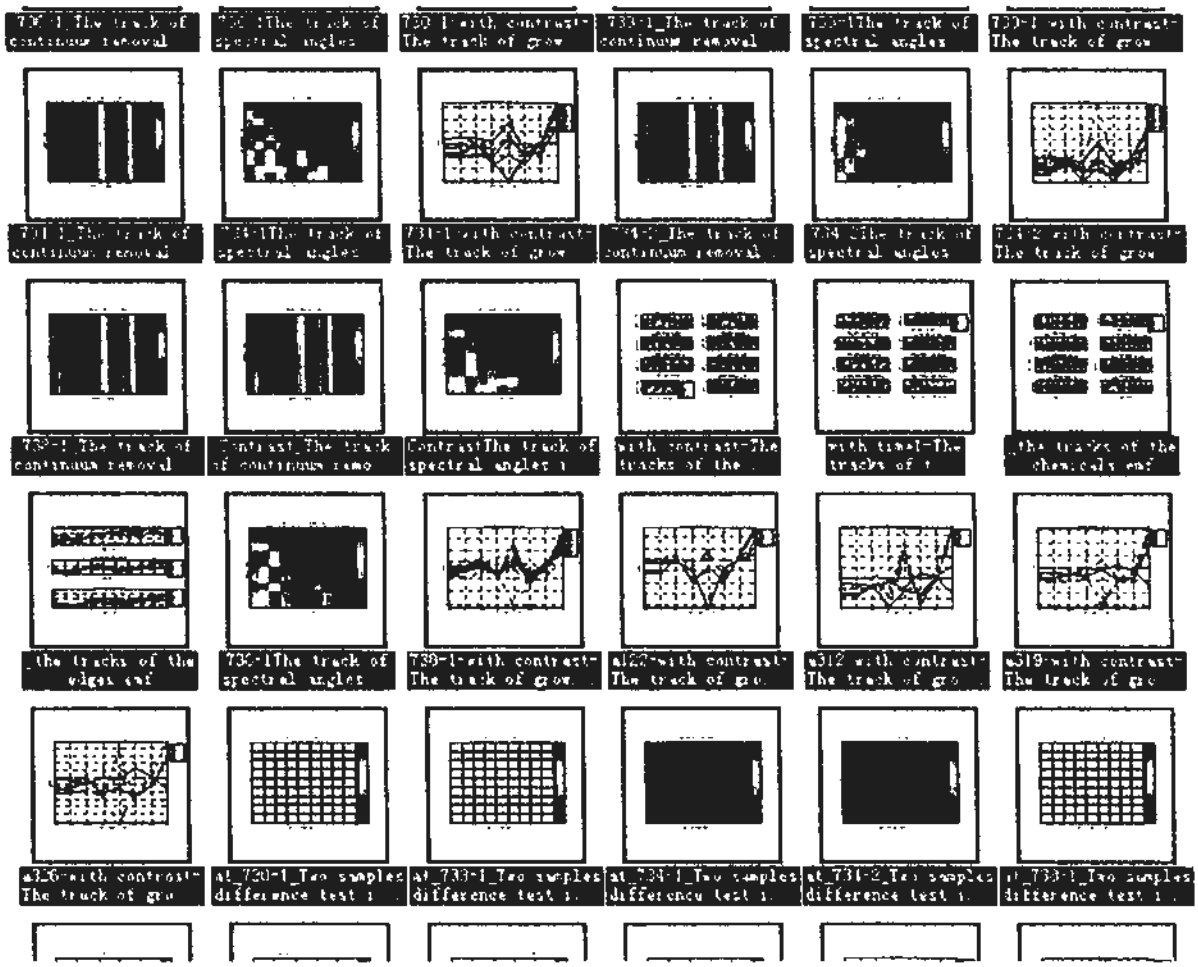


Figure 6-9. The results of "Analysis of sample growth trends"

Chapter 7 Summary

Due to the stochasticity, diversity and variability of gene expression, transgenic crop (e.g. breeding) is confronted with some uncertainties, such as what kinds of the influence from foreign gene on the transgenic crop will be, and how to monitor the growth of transgenic crop in time efficiently. Some professional approaches were developed (e.g. PCR) are at a micro view to confirm some problems. However, prior knowledge is needed in some cases, and if mass samples are to process, these approaches will be helpless to monitor all samples real-/ near real-time because of high cost and some uncontrolled problems. It is very important to monitor and assess the growth of samples, especially for the experiment lasting for a long time, such as plant breeding.

In this study, we proposed to employ hyperspectral remote sensing technique, a kind of practical and field spectroscopy technique, to obtain field fine spectra of transgenic paddy rice and monitor the its growth by its biophysical traits to fulfill screening of cultivars in contrast

controlled experiments. The biophysical traits or bio-process were concentrated on rather than micro-structure or components of proteins. This approach compares the differences between transgenic samples and their contrast which are cultivated in the controlled contrast environment. By monitoring the real-/ near real-time growth of sample, the techniques were applied to fulfill early indication of possible the differences between transgenic crops and their contrast in the controlled contrast experiment. It will be implemented to monitor the growth of the samples real-/ near real-time, will assist to screen samples and help researchers clearly know their samples.

In the monitoring process, hyperspectral remote sensing techniques play a role of indicating and monitoring of the influences from foreign gene by an indirect way from a macro-view. The influences of foreign gene could be treated as a special source of stress to vegetation. Therefore, it is possible to detect the difference between transgenic and contrast group and monitor the growth of sample to assist to fulfill sample screening work, focusing on the plant biophysical traits or responses to stress. In the study, more than 6 times experiments in different fields were conducted, involving three kinds of genomes and their transgenic samples. The experiments were designed as the experiment-repeat experiments and the gene-repeat experiments. Moreover, an experiment lasting for three months was also conducted for evaluating the capability of the approach to monitor the sample growth. Spectral analysis and statistic approaches were applied to assess the samples and their growth.

7.1 Conclusion

Both morphologic and parameterized features of foliar spectra of samples were applied to indicate the growth of the samples: (1) These responses revealed the difference of growth status between transgenic samples and the contrast group. (2) These features had been proved related to the bio-process of vegetation. (3) The spectral angle and continuum removal of spectrum were used as spectral morphologic features to indicate the differences among samples qualitatively. (4) Meanwhile, the status of the eight important foliar chemicals was applied as indicators of sample growth. They were chlorophyll a, chlorophyll b, chlorophyll a+b, carotenoids, anthocyanin, water, nitrogen and lignin. (5) For estimating their contents accurately, current spectral indices were evaluated systemically. (6) And to paddy rice, new spectral indices, for estimation of the contents of photosynthetic pigments, were also designed to overcome the flaws of the current ones, such as noise resistance. (7) Sensitive bands for retrieval of foliar nitrogen and lignin concentration were assessed and new regression models were developed.

These indices are sensitive indicators of sample growth in the approach. And the results proved the new spectral indices or regression models developed in this study for paddy rice were efficient, sensitive and reliable in estimation of foliar chemicals.

By applying these indicators (achievements) and analyzing the kinds of features in different types of experiments, the conclusions could be obtained:

(1) In section 5.1, mainly the spectral indices of the edges were used to assess the ability of photosynthesis of transgenic samples. The results proved the approach was efficient to assess samples. Both the results of the two experiments showed that: (a) the differences at all edges and absorption chosen in certain band range. The results indicated that there were some matters bringing changes to the transgenic samples stably. These parameters had high relationship with kinds of photosynthetic pigments, thus it could be deduced the differences of pigments content in samples. This information could be used to assess the photosynthetic ability of samples. (b) The discovered differences between samples compared with their parents, some were positive and the other were negative to photosynthesis. (c) Because the whole growth processes of samples were cultivated under a strict controlled contrast condition and external and random noise had been weaken by mean spectra, a conclusion based on spectral analysis could be obtained.

(2) In section 5.2, the gene-repeat experiment was conducted and the data were analyzed to assist transgenic crop screening and breeding, based on parameterized features (foliar chemicals) and spectral shape. The Results, the similarities of samples by spectral angles, the position of spectral differences by continuum removal, and the spectral differences by spectral indices high related to photosynthetic pigments contents had been compared, analyzed and assessed qualitatively and quantitatively. It showed that under contrast conditions, the differences between transgenic groups and the contrast could be observed and assessed by hyperspectral remote sensing approach both on spectral morphology and specific indices. Applying our proposed approach, the differences in transgenic groups also could be observed and assessed.

Both the analyses in section 5.1 and 5.2, the target bands were just photosynthetic sensitive ranges since the genes transferred were directly related to photosynthesis.

(3) In the section 5.3, the long time series data of transgenic groups and the contrast were analyzed. The samples growth pattern and status were analyzed within individual sample, within sample groups, and among transgenic group and contrast group, respectively. A potential temperature stress had been found. The Chl change pattern was stable highly related to the track of carotenoids, water content, and lignin content. A new description variable (vector) cc was defined to as indicators of paddy rice growth monitoring. It is a comprehensive biochemical status of a sample at a certain time. By censoring the correlation results of samples cc, it was found that the status of some samples at a certain time were not consistent. After a statistic test, the change of carotenoids and their water contents were significantly different. All these results are useful to assist professional biologists to fulfill crop screening and breeding. It helps them know their samples clearly and make their experiments under control.

(4) In section 5.4, the samples transferred Bar gene for resisting herbicide were chosen. In order to evaluate the transgenic samples' performance to herbicide resistance, an artificial induction experiment and a controlled stress environment were conducted. Except for the contrast group, all others were sprayed herbicide. Two kinds of analysis were done to assess the samples, namely quantitative analysis based on spectral morphological characteristics and qualitative analysis based on parameterized features within groups and among groups. Spectral angles among samples showed that the samples had different response to the stress of herbicide. Bigger angles were found among some samples while than others showed much similarity of spectra to the parent. By investigating the continuum removal results, it could be found that differences mainly located in range of [420,750nm], [1400, 1600nm] and [1800, 2000nm]. And a sample at the five time points had a great of positive shift dispersions in the continuum removal figure. The results of the correlation analysis and significant difference test also qualitatively proved the previous finding. And some samples were found distinguished differences from the others under herbicide stress. These results revealed the different capabilities of the sample responding to herbicide stress which may relate to Bar gene between them directly. It screened some samples should be paid more attention in the following work. When after 168 hours the growth statuses were all highly related. Herbicide stress made the transgenic sample to grow consistently. All these results will be of good support for professional breeding study.

(5) In section 5.1-5.4, the analyses were based on mean spectra at group level, in section 5.5, an study for detecting outliers of single spectral of an individual sample were conducted. These analyses were mainly concentrated on outlier detection, and growth analysis based on spectral morphologic features and statistic test. The study is expected to answer two questions if any differences exist only in a specific individual sample and if it is reasonable to take mean spectrum as representative of the whole group. It is an important complement for the study focusing on group level.

In general, by analyses in the gene-repeat experiments and experiment-repeat experiment, all the results proved that the proposed approach was useful. It could be an important, helpful and efficient complement to make the study under control and efficient.

7.2 Prospective

Cross-application of techniques in different disciplines would bring a creative surprise. In this study, hyperspectral remote sensing techniques were applied in indicating and monitoring the cultivated paddy rice growth to assist professional biologists to fulfill crop screening and breeding. The gene-repeat experiments, the experiment-repeat experiment, stress-induced experiment were conducted to validate and improve on the proposed approach. The results were satisfactory. However, if the following problems were considered, the results would be improved much:

(1) Much more effective communication with biological researchers

This is a common problem faced by cross-application. By full communication, the demand of biological researchers in transgenic crop breeding and screening will be understood clearly, and more efficient approach based on hyperspectral remote sensing techniques can be design and implement, namely, application-orientated study.

(2) Diversity of the research methods

For an efficient communication, more visual results should be obtained and provide to biological researchers. Thus the research methods should be developed. In the study, mainly three kinds of methods were applied to assess the samples, morphologic features of spectrum, parameterized features of spectrum and statistic test. The reliable relationship between the results from the hyperspectral techniques should be built. All these need more research methods to be applied.

(3) Extension of study scope

Current study concentrated on the fine spectrum of leaf-level, totally neglecting the canopy features of sample. Some properties at canopy level are also important for crop breeding, such as LAI which is related to yields. And morphologic characteristics of sample are indices of evaluation in crop cultivation.

(4) More foliar chemicals

In this study, total eight kinds of important foliar chemicals were chosen as indicators of sample growth. Most of these chemicals were photosynthetic pigments. In some sense, they could not reflect all responses of sample to all kinds of stress. Thus, more foliar chemicals should be studied, such as cellulose, pertain. And more sensitive parameters should be introduced, such as fluorescence of photosynthesis.