Investigating Within-Field Variability of Rice from High Resolution Satellite Imagery in Qixing Farm County, Northeast China

Quanying Zhao 1,2,*, Victoria I.S. Lenz-Wiedemann 1,2, Fei Yuan 2,3, Rongfeng Jiang 4, Yuxin Miao 2,4, Fusuo Zhang 4 and Georg Bareth 1,2

1 Institute of Geography, University of Cologne, Cologne 50923, Germany; E-Mails: victoria.lenz@uni-koeln.de (V.I.S.L.-W.); g.bareth@uni-koeln.de (G.B.)
2 International Center for Agro-Informatics and Sustainable Development (ICASD), Cologne 50923, Germany; E-Mail: victoria.lenz@uni-koeln.de
3 Geography Department, Minnesota State University, Mankato, MN 56001, USA; E-Mail: fei.yuan@mnsu.edu
4 College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China; E-Mails: rfjiang@cau.edu.cn (R.J.); ymiao@cau.edu.cn (Y.M.); zhangfs@cau.edu.cn (F.Z.)

* Author to whom correspondence should be addressed; E-Mail: zhaoquanying@gmail.com; Tel.: +49-221-470-1951.

Academic Editor: Wolfgang Kainz

Received: 3 September 2014 / Accepted: 7 January 2015 / Published: 3 February 2015

Abstract: Rice is a primary staple food for the world population and there is a strong need to map its cultivation area and monitor its crop status on regional scales. This study was conducted in the Qixing Farm County of the Sanjiang Plain, Northeast China. First, the rice cultivation areas were identified by integrating the remote sensing (RS) classification maps from three dates and the Geographic Information System (GIS) data obtained from a local agency. Specifically, three FORMOSAT-2 (FS-2) images captured during the growing season in 2009 and a GIS topographic map were combined using a knowledge-based classification method. A highly accurate classification map (overall accuracy = 91.6%) was generated based on this Multi-Data-Approach (MDA). Secondly, measured agronomic variables that include biomass, leaf area index (LAI), plant nitrogen (N) concentration and plant N uptake were correlated with the date-specific FS-2 image spectra using stepwise multiple linear regression models. The best model validation results with a relative error
(RE) of 8.9% were found in the biomass regression model at the phenological stage of heading. The best index of agreement (IA) value of 0.85 with an RE of 13.6% was found in the LAI model, also at the heading stage. For plant N uptake estimation, the most accurate model was again achieved at the heading stage with an RE of 11% and an IA value of 0.77; however, for plant N concentration estimation, the model performance was best at the booting stage. Finally, the regression models were applied to the identified rice areas to map the within-field variability of the four agronomic variables at different growth stages for the Qixing Farm County. The results provided detailed spatial information on the within-field variability on a regional scale, which is critical for effective field management in precision agriculture.

Keywords: rice; FORMOSAT-2; agronomic variable; expert classification; multi-data-approach (MDA); within-field variability; Sanjiang Plain; Northeast China

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important staple food crops, feeding over half of the world’s population. In 2010, the global rice production was approximately 672 million tons from a cultivation area of around 154 million ha, with China contributing to 29% and 19% of the rice production and cultivation area, respectively [1]. Paddy rice management and its irrigation strategy have significant effects on greenhouse gas emissions [2,3]. Globally, rice paddies contribute about 10% of the total methane flux to the atmosphere [4]. In 2000, the soil nitrogen (N), phosphorus (P), and potassium (K) nutrient deficit induced by rice production accounted for 42% of the global deficit amount [5]. Paddy rice agriculture in China is therefore of national and global significance in terms of both food security and sustainable development.

In the past years, remote sensing (RS) as an advanced technology has been used extensively in agriculture to obtain spatial and temporal information about crops [6–9]. Paddy rice areas were well projected using RS techniques [10–12]. Geographic Information System (GIS) data have been proved to be important to enhance the accuracy of land use and land cover classification [13,14]. RS data with coarse and medium resolution are widely used in rice cultivation research [12,15–26]. However, the number of conducted studies on rice using high resolution RS images was limited in the past two decades [23,27,28]. Identification of rice cultivation areas and estimation of agronomic parameters from high resolution images are valuable for improving rice production.

Mapping rice cultivation areas accurately is fundamental for the assessment of agricultural and environmental productivities, the analysis of food security and therefore national and international food trade decisions [29–31]. Many previous studies have mapped rice areas [12]. Various classifiers, including maximum likelihood, artificial neural network, decision tree, and spatial reclassification kernel, have been used in vegetation mapping using RS. However, classification accuracies are mainly determined by the quality and quantity of RS data. No ideal image classifier is uniformly applicable to all tasks [32]. Extensive field knowledge and auxiliary data help to improve the classification accuracy [32]. In the Multi-Data-Approach (MDA) for land use and land cover classifications as well as crop rotation
mapping, multi-temporal RS and GIS vector data are combined to derive as much information as possible. Studies [16,17] demonstrated that considering extensive field knowledge and ancillary GIS data in the post-classification process can improve the classification accuracy by up to 10%.

Agricultural RS refers to the method of non-contact measurements of electromagnetic radiation reflected or emitted from plant materials or soils in agricultural fields [33]. Different vegetative covers can be distinguished according to their unique spectral behavior in relation to overall ground elements [34]: Visible radiation in the red (630–690 nm) is absorbed by chlorophyll while radiation in the near infrared (760–900 nm) is strongly reflected by leaf cellular structures. Vegetation Indices (VIs) are developed to qualitatively and quantitatively evaluate vegetative characteristics by combining spectral measurements from different wavelength channels [35]. Theoretically, the VIs should be particularly sensitive to vegetative covers, insensitive to non-vegetation factors such as soil properties, atmospheric effects and sensor viewing conditions [36]. In practice, factors of soil characteristics, atmosphere and sensor radiometry degradation, as well as differences in the spectral responses and bidirectional effects all have considerable effects on vegetation indices. Therefore, many VIs have been developed to enhance the vegetative cover signal while minimizing the background response [35]. Hansen and Schjoerting [37] discussed the optimized NDVI from different band centers and band widths to represent wheat parameters such as biomass, leaf area index (LAI), chlorophyll, and N status. In their study, the partial least square regression algorithm was applied. Yao et al. [38] successfully explored empirical models based on ground-based RS techniques to estimate N status and improve N use efficiency in rice. Yu et al. [39] investigated the potential of hyperspectral band combinations in interpreting canopy N status in rice. Despite the lack of interpretations of the physical mechanisms and interactions between the target properties and the measured signals, these empirical models are fundamental materials for further research in order to represent the physical reality.

Remotely sensed photosynthetically active radiation has been used to evaluate the primary production of crops [40,41]. The most sophisticated method commonly combines RS data with dynamic crop growth models [42,43]. However, these methods require numeric parameters and are sensitive to soil background conditions.

Precision agriculture (PA) emerged in the middle of the 1980s [33]. Later, in 1991, satellite data were firstly used in PA [44]. Spatial and temporal variability of soil and crop factors within a field is the essential base of PA [45]. Measurement of various crop canopy variables during the growing season provides an opportunity for improving grain yield and quality by site-specific fertilizer applications. Due to the ability of providing high temporal, spatial and spectral resolution images, satellite RS has a significant potential in PA [33,46]. The multi-temporal within-field information on crop status captured by satellite RS is invaluable in PA.

Because of its operational and economical uses over large areas, satellite RS technology has been widely used to conduct in-season crop yield forecasting for decision making on marketing intervention and policy support on regional or global scales [29,31]. Satellite RS is also an essential technique for agro-ecosystem studies on regional scales [47–49].

The overall aim of this study is to investigate the possibilities and accuracies of within-field variability of rice status monitoring during the entire growing season on a county scale (Qixing Farm) using satellite RS data. The term “within-field variability” in this study refers to the spatial variability of agronomic variables (biomass, LAI, N concentration, N uptake, etc.) within a rice crop field defined by enclosed
boundaries. The size of a rice field in this study ranges from 0.5–100 ha. First, a map of rice cultivation areas was produced using multi-temporal RS FORMOSAT-2 (FS-2) images, coupled with auxiliary GIS data (MDA) to improve the classification accuracy. Then, empirical regression models were developed to relate RS spectra with field measurements of the four agronomic variables. One advantage of this research is the construction of specific regression models for different growth stages. Based on a stepwise regression analysis, the optimized predictors from the satellite images were identified to investigate the dynamics of crop canopy characteristics at different stages. The specific objectives of this study are to (1) identify rice cultivation areas with high classification accuracies based on multi-temporal RS and GIS data; (2) develop regression models for deriving rice crop variables from the FS-2 imagery; (3) validate the ability of the regression models to estimate rice parameters; (4) apply the regression models to the entire Qixing Farm County study area.

The presented method of extracting high resolution within-field variability information from the FS-2 imagery will assist farmers in their site-specific rice crop management and strategy planning.

2. Study Area

The study site Qixing Farm (47.2°N, 132.8°E), is located in the Sanjiang Plain (SJP) in North-eastern China (see Figure 1). The SJP is an alluvial plain formed by the Songhua River, the Heilong River and the Wusuli River. The administrative area of the Qixing Farm County is about 120,000 ha. The climate in the SJP is temperate sub-humid, with a mean annual precipitation of 500–650 mm. Rainfall mainly occurs from May to September during the growing season of crops. The accumulated \( \geq 10 \, ^\circ C \) temperature all across the year is about 2300–2500 \( ^\circ C \) and only single-season crops are planted. The topography is rather flat with an average elevation of 60 m and is characterized by broad alluvial plains and low terraces formed by the rivers. The SJP, which covers an area of more than 100,000 km\(^2\), exceeding the size of the Netherlands almost by three times, is one of the major agricultural areas of China. In 2009, arable land in the SJP accounted for nearly 60% of the total land, being dominated by paddy rice (57%) [50]. Since the irrigation constructions (water channels, raised ridges) for paddy rice are reusable year by year, the field boundaries are mostly stable. In recent years, for better economic profit, there has been moderate land use change from dryland to paddy rice. Compared with Europe’s highly fragmentated agricultural landscapes which prevent the utility of coarse resolution RS data for quantitative crop monitoring and yield forecasting [32], SJP’s large homogenous landscapes provide an ideal site for monitoring crops using satellite RS.
Figure 1. Location of the study area in Northeast China (the upper left corner shows a subset of the FS-2 image acquired on 9 August 2009).

3. Data

3.1. Satellite RS Images and GIS Data

FORMOSAT-2 (FS-2) collects multispectral images with a ground pixel resolution of $8 \times 8$ m$^2$ over a swath of 24 km. The FS-2 images used in this study are optical images with 4 bands of blue (450–520 nm), green (520–600 nm), red (630–690 nm), and near-infrared (760–900 nm). Three tiles of high quality images covering the main arable land area (~56,000 ha) of the Qixing Farm were captured on 24 June, 6 July, and 9 August, in 2009. Thus, both the vegetative phase (24 June and 6 July) and the reproductive phase (9 August) of rice are well represented in these images.

GIS vector data of Qixing Farm field boundaries were provided by the Qixing Modern Agriculture Research Center. Information on crop field boundaries, irrigation wells, water drainages, and shelter forests edges are given at a fine field unit scale. Additional information such as crop type of each field is given in the GIS attribute table.
3.2. Ground Truth Data Collection

Field campaigns for the agronomic data collection were carried out during the entire rice growing season in 2009. In total, 42 sample sites covered by the FS-2 images were selected for this study. All these 42 sites were located in seven farmers’ fields being spatially separated. Each site was represented by one plot covering approximately 0.1–0.3 ha. The final plant samples collected from each site were a mixture of three or four spatially separated samples, taken from the same plot. As ground truth data, the areas of sample sites were mapped using a Trimble™ Global Positioning System (GPS) receiver. Field management calendars of transplanting, N topdressing, irrigation, application of insecticides, and harvest dates, were recorded. Several field campaigns were carried out to collect samples from the tillering stage, booting stage, heading stage, 20 days after heading, and the harvest stage. For each site, biomass, LAI and plant N concentration were measured and plant N uptake was calculated as well.

After the field sampling, the plants were first cleaned, and then separated into different organs (leaves, stems, panicles) to measure the biomass values. LAI was measured using a sub-sample of the leaf biomass. One sub-sample consisted of 10–20 leaves, randomly selected among the youngest fully developed leaves. All fresh samples were processed in the oven at 105 °C for half an hour to stop enzyme activity. After that, they were dried at 75 °C for at least 72 h until a constant weight was reached before they were finally weighted. N concentration was measured using the Kjeldahl-N method. The plant N uptake was calculated as the aboveground dry mass multiplied by the N concentration. Detailed information is listed in Table 1.

In the study area, rice cultivation technologies in high latitude are relatively sophisticated and rice cultivation regulations developed by the government are applied by most of the farmers. The rice seedlings are first grown in greenhouses and are then transplanted into the paddy fields. In the regulations, a date window of 15–25 May is suggested for transplanting in order to capture the maximum accumulated temperature throughout the whole year. In this research, we divided our ground truth data sets into two groups according to the transplanting dates. Sample sites with seedlings transplanted from 15–25 May were used to construct empirical regression models between the agronomic parameters and the RS data. Sample sites with seedlings transplanted beyond those dates were used as validation data sets to evaluate the performance of the regression models.
Table 1. Agronomic parameters of the sample sites in 2009.

<table>
<thead>
<tr>
<th>Field</th>
<th>Number of Sample Sites</th>
<th>Variety</th>
<th>Number of Foliage</th>
<th>Transplanting Date</th>
<th>Quantity of N Topdressing (kg/ha)</th>
<th>Plant Density (plants/m²)</th>
<th>Soil Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K (mg/kg)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P (mg/kg)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N (mg/kg)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SOM (g/kg)</td>
</tr>
<tr>
<td>Field 1</td>
<td>4</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 10</td>
<td>104, 104, 104, 104</td>
<td>109</td>
<td>6.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Kongyu 131</td>
<td>11</td>
<td>May 17</td>
<td>0, 94, 141</td>
<td></td>
<td>75.7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Xixuan 1</td>
<td>13</td>
<td>May 15</td>
<td>132</td>
<td>101</td>
<td>6.22</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Longdun 104</td>
<td>12</td>
<td>May 16</td>
<td>118</td>
<td></td>
<td>134.5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 17</td>
<td>118</td>
<td>133</td>
<td>6.07</td>
</tr>
<tr>
<td>Field 2</td>
<td>4</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 13</td>
<td>84, 84, 84, 84</td>
<td>115</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 13</td>
<td>0, 67, 101</td>
<td></td>
<td>103.8</td>
</tr>
<tr>
<td>Field 3</td>
<td>4</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 12</td>
<td>67, 67, 67, 67</td>
<td>117</td>
<td>6.13</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Longjing 21</td>
<td>12</td>
<td>May 12</td>
<td>0, 54, 81, 67</td>
<td></td>
<td>126.5</td>
</tr>
<tr>
<td>Field 4</td>
<td>4</td>
<td>Kendao 6</td>
<td>11</td>
<td>May 20</td>
<td>83, 83, 83, 83</td>
<td>158</td>
<td>6.29</td>
</tr>
<tr>
<td>Field 5</td>
<td>4</td>
<td>Longjing 24</td>
<td>11</td>
<td>May 16</td>
<td>95, 95, 95, 95</td>
<td>166</td>
<td>5.84</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Longjing 24</td>
<td>11</td>
<td>May 16</td>
<td>0</td>
<td></td>
<td>73.4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Longdun 249</td>
<td>12</td>
<td>May 20</td>
<td>102</td>
<td></td>
<td>37.8</td>
</tr>
<tr>
<td>Field 6</td>
<td>1</td>
<td>Chaoyou 949</td>
<td>11</td>
<td>May 18</td>
<td>106</td>
<td>144</td>
<td>6.08</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Jinxuan 1</td>
<td>11</td>
<td>May 20</td>
<td>0, 86, 129</td>
<td></td>
<td>98.4</td>
</tr>
<tr>
<td>Field 7</td>
<td>3</td>
<td>Kongyu 131</td>
<td>11</td>
<td>May 17</td>
<td>0, 82, 123</td>
<td>122</td>
<td>37.08</td>
</tr>
</tbody>
</table>
4. Methods

4.1. Satellite Image Pre-Processing

Pre-processing of satellite images prior to vegetation information extraction is essential to remove noise and increase the interpretability of image data [51]. Solar radiation reflected by the earth surface to the satellite sensors is significantly affected by its interaction with the atmosphere. Atmospheric correction is an important pre-processing step for satellite RS [52,53]. The uncertainties resulting from atmospheric effects in agriculture applications of satellite RS have been well discussed during the past decades [18,21,54,55]. In this study, atmospheric correction was performed using ENVI FLAASH, version 5.1. Table 2 lists the main parameters used in the atmospheric correction process. The sub-arctic summer model for rural region was selected.

<table>
<thead>
<tr>
<th>Date</th>
<th>Visibility (km)</th>
<th>Zenith Angle</th>
<th>Azimuth Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 24</td>
<td>50</td>
<td>146°10'0.84&quot;</td>
<td>−115°13'19.56&quot;</td>
</tr>
<tr>
<td>July 6</td>
<td>50</td>
<td>152°15'58.34&quot;</td>
<td>−45°34'58.34&quot;</td>
</tr>
<tr>
<td>August 9</td>
<td>50</td>
<td>155°12'2.90&quot;</td>
<td>−83°58'22.08&quot;</td>
</tr>
</tbody>
</table>

Geometric distortion is another important factor affecting the results of image processing, especially when combing geospatial data from different dates or multiple sources. Dai and Siamak [56] noted that the geometric error even on a sub-pixel level can significantly affect the accuracy of land use classification from satellite images. The precision of geometric correction depends on the number, distribution, and accuracy of the Ground Control Points (GCPs) [57]. To avoid labor intensive work of in-situ GCP collection, Zhao et al. [58] developed a geometric correction method for georeferencing multi-source geodata by using TerraSAR-X data as a reference. The same method was applied in this study. Specifically, a stacked TerraSAR-X image produced from five dates served as the reference to georectify the FS-2 images. The main georeferencing parameters are shown in Table 3. The positional error (PE) was less than 6 m for all three images. More details about the method can be found in Zhao et al. [58].

<table>
<thead>
<tr>
<th>Image Capture Date</th>
<th>Number of Control Points</th>
<th>Number of Check Points</th>
<th>PE of Control Points (m)</th>
<th>PE of Check Points (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 24</td>
<td>100</td>
<td>20</td>
<td>2.991</td>
<td>5.901</td>
</tr>
<tr>
<td>July 6</td>
<td>104</td>
<td>20</td>
<td>4.143</td>
<td>5.032</td>
</tr>
<tr>
<td>August 9</td>
<td>101</td>
<td>20</td>
<td>3.353</td>
<td>5.666</td>
</tr>
</tbody>
</table>

The mean reflectance and VIs for each sample plot were calculated using the “Zonal Analysis” tool in ERDAS IMAGINE 2013. For each of the three FS-2 images, correspondingly 42 RS samples were extracted and used to develop the empirical regression models for crop status monitoring.
4.2. Mapping Rice Cultivation Areas

A rice area map of the Qixing Farm in 2009 was produced using aforementioned MDA. In particular, multi-temporal RS and GIS data were integrated to improve the accuracy of rice mapping. Based on the properties/attributes of the different datasets, a knowledge base was constructed and implemented, using the Knowledge Engineering tool mounted in ERDAS IMAGINE 2013. A series of logical rules were created in the Knowledge Engineering tool to integrate the RS and GIS data in order to achieve higher classification accuracy.

The general steps for delineating rice areas are summarized as follows: (1) a supervised classification based on the maximum likelihood algorithm was carried out. In order to extract more unique spectral signatures, more than five subclasses were classified firstly (for the image of June 24: nine subclasses in total with one of them being rice; for the image of July 6: 16 subclasses in total with three of them representing rice; for the image of August 9: 11 subclasses in total with three representing rice); (2) the resulting subclasses were further combined to five main classes, including rice, dryland, forests, residential areas, and “other areas” for each image; (3) isolated pixels were eliminated using the “clump and eliminate” functions embedded in ERDAS IMAGINE 2013; (4) an expert classification system based on knowledge rules was implemented to integrate and improve the rice classification results from multiple dates; and (5) the vector GIS data were used as auxiliary dataset in the post-classification process to further improve the accuracies of rice classification.

In a MDA study, Waldhoff et al. [13] reported that the support vector machine (SVM) approach and the maximum likelihood classifier (MLC) yielded similar classification accuracies. They found although the SVM method performed slightly better (up to 3%) in three of the four cases, the MLC had a shorter processing time. Therefore, the MLC was selected in this study to derive multi-temporal high resolution land cover classifications over a large study area. After the classification, the rice classes derived from the three RS images were “combined” by logical rules to take advantage of the specific spectral information corresponding to growth stages. Specifically, in the first step of the supervised classification, to improve the producer’s accuracy of rice classes, relatively “broad” spectral information was selected as rice spectral signatures. Thus, some other land covers such as “dryland” and “bare soils” may have been classified into rice classes as well. However, these misclassified areas from one single image could be different from those areas from the other dates, because spectral differences between rice and other land covers vary with growth stage [59]. In the next step, using the knowledge base, the pixels classified into rice classes in all three RS images were categorized into a new “rice class 1”. The areas classified into rice classes in two RS images were categorized into a new “rice class 2”. After that, these refined RS rice areas were further “combined” with the GIS data (rice area masks). To assign a pixel into the final rice class, the following conditions had to be satisfied: (1) this pixel was in the “rice class 1”; or (2) this pixel was in the “rice class 2” and in the GIS rice mask.

4.3. Ground Truth Data Interpolation

Ground truth data for each site were collected at the tillering stage (28–30 June), the jointing stage (10–12 July), the heading stage (2–8 August), 20 days after heading (22–28 August), and the harvest time. The FS-2 images were acquired on 24 June, 6 July and 9 August, at slightly different dates from
the field campaigns. Therefore, the values of the agronomic variables of biomass, LAI, plant N concentration and N uptake were interpolated for these three FS-2 dates. First, a specific polynomial growing curve for each site was constructed based on the time series of ground truth data. The values of the agronomic variables for the three RS-2 dates were then interpolated. These interpolated values were used as the new ground truth data to explore their relationships with the satellite image reflectance and VIs.

### 4.4. Development of Regression Models for Deriving Agronomic Variables

A multiple linear regression method was constructed for each agronomic variable based on reflectance values and the VI derived from the FS-2 images and the corresponding field measurement. The VI used in this method was treated as an equivalent factor of reflectance. Correlation and regression analyses were performed in SPSS 21 (SPSS, Inc., Chicago, IL, USA). The format of the multiple linear regression models was as follows:

$$Y_E = \beta_0 + \beta_{Blue} \times R_{Blue} + \beta_{Green} \times R_{Green} + \beta_{Red} \times R_{Red} + \beta_{NIR} \times R_{NIR}$$

$$+ \beta_{NDVI} \times NDVI \hspace{1cm} (1)$$

The NDVI is calculated by the following function:

$$NDVI = \frac{(R_{NIR} - R_{Red})}{(R_{NIR} + R_{Red})} \hspace{1cm} (2)$$

$R$ stands for the reflectance value at the subscripted satellite band. In Equation (1), $Y_E$ stands for the estimated agronomic variable; $\beta_i$ is the coefficient of the reflectance at the corresponding band/vegetation index. The performance of the regression model was evaluated by the coefficient of determination ($R^2$).

Biomass and LAI are essential crop physiological variables which determine the crop yield. LAI refers to the ratio of leaf surface area to ground area. It is a fundamental canopy parameter in agronomy and RS since it drives absorption of solar radiation and evapotranspiration for carbon assimilation, and thus primary production. In this research, both biomass and LAI were related to the FS-2 band reflectance based on the aforementioned Equation (1).

N is one of the most remobilizable elements during the reproductive stage in rice plants [60]. Since its remobilization causes leaf senescence, it is directly related to crop productivity [61]. Accurate plant N status detection to develop site specific N management strategies for rice in the SJP is of importance regarding both agricultural and environmental aspects [39]. In this study, plant N concentration and N uptake were also derived from the FS-2 images using the regression model represented by Equation (1).

### 4.5. Validation of the Regression Models

The regression models were evaluated in a validation analysis. The feasibility of the model was quantified by the statistical measures of relative error (RE) and index of agreement (IA). In a further step, scatterplots were generated to assess the performance of the regression models. The RE is the ratio of the Root Mean Square Error (RMSE) to the mean of observed values, describing the differences between the predicted and the observed values relative to the mean of the ground truth values. The IA represents the degree of agreement between the model estimations and observed values [62]. It is calculated as:
\[
IA = 1 - \frac{\sum_i (O_i - P_i)^2}{\sum_i (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}
\]  

(3)

where \(O_i\) is the observed value, \(P_i\) is the model-simulated value, and \(\bar{O}\) is the mean of observed values. The denominator in Equation (3) was defined as a “potential error” by Willmott [62]; therefore the IA represents the ratio between the mean square error and the “potential error”. Although the IA is sensitive to extreme values [63], it can be interpreted straightforwardly since it ranges from 0–1.

5. Results

5.1. Accuracy of Rice Area Classification

To compare the classification accuracies from different data sources, 800 random points were created in the study area. More than 80 random points were distributed in each class. Error matrices were generated to quantify the classification accuracies. The kappa coefficients of agreement and overall accuracies for the rice class are shown in Table 4.

Table 4. Accuracies of rice maps produced from different data sources.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>User’s Accuracy of Rice</th>
<th>Producer’s Accuracy of Rice</th>
<th>Kappa Coefficient</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS single date (August 9)</td>
<td>94.1%</td>
<td>82.6%</td>
<td>0.733</td>
<td>80.8%</td>
</tr>
<tr>
<td>RS multiple dates (3 dates)</td>
<td>89.4%</td>
<td>91.7%</td>
<td>0.781</td>
<td>85.0%</td>
</tr>
<tr>
<td>RS and GIS data combined</td>
<td>94.2%</td>
<td>92.7%</td>
<td>0.881</td>
<td>91.6%</td>
</tr>
</tbody>
</table>

With auxiliary GIS data, both the Kappa and the overall accuracy values for the rice class were improved by 0.148 and 10.8%, respectively. While a similar user’s accuracy was obtained after combining the three RS classification maps and the GIS ancillary data, the producer’s accuracy increased from 82.6% to 92.7%, implying a decrease of 10% in omission errors. These results are in line with the conclusions made by Shrestha and Zinck [64] and Rozenstein and Karnieli [14].

5.2. Empirical Regression Models

5.2.1. Model Development

The parameters of the best qualified regression models for all agronomic variables are shown in Table 5. The coefficient of determination \(R^2\) was used to evaluate the regression models.
### Table 5. Parameters of empirical regression models.

<table>
<thead>
<tr>
<th>Agronomic Variable</th>
<th>Phenological Stage</th>
<th>Number of Model Construction Sites</th>
<th>Regression Model Parameters</th>
<th>Model Coefficient</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\alpha_0$</td>
<td>$\alpha_{Blue}$</td>
<td>$\alpha_{Green}$</td>
</tr>
<tr>
<td>Biomass (kg/ha)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tillering</td>
<td>17</td>
<td>−316.063</td>
<td>52,440</td>
<td>−25,250</td>
<td></td>
</tr>
<tr>
<td>Booting</td>
<td>23</td>
<td>7280.379</td>
<td>51,620</td>
<td>−121,530</td>
<td>21,270</td>
</tr>
<tr>
<td>Heading</td>
<td>23</td>
<td>5791.901</td>
<td>−148,910</td>
<td>31,710</td>
<td></td>
</tr>
<tr>
<td>LAI</td>
<td></td>
<td></td>
<td>−1.04</td>
<td>140</td>
<td>−50</td>
</tr>
<tr>
<td>Tillering</td>
<td>17</td>
<td>−5.463</td>
<td>−270</td>
<td>380</td>
<td>30</td>
</tr>
<tr>
<td>Booting</td>
<td>23</td>
<td>−0.451</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heading</td>
<td>23</td>
<td>−5.463</td>
<td>−50</td>
<td>−270</td>
<td>380</td>
</tr>
<tr>
<td>N Concentration (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tillering</td>
<td>17</td>
<td>4.371</td>
<td>−101.548</td>
<td>71.492</td>
<td></td>
</tr>
<tr>
<td>Booting</td>
<td>23</td>
<td>4.895</td>
<td>−22.079</td>
<td>−23.972</td>
<td>−4.718</td>
</tr>
<tr>
<td>Heading</td>
<td>23</td>
<td>0.193</td>
<td>−35.742</td>
<td>−44.725</td>
<td>105.14</td>
</tr>
<tr>
<td>N Uptake (kg N/ha)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tillering</td>
<td>17</td>
<td>−21.022</td>
<td>−370</td>
<td>290</td>
<td>260</td>
</tr>
<tr>
<td>Booting</td>
<td>23</td>
<td>−18.834</td>
<td>810</td>
<td>−1420</td>
<td>280</td>
</tr>
<tr>
<td>Heading</td>
<td>23</td>
<td>167.251</td>
<td>−4270</td>
<td>550</td>
<td></td>
</tr>
</tbody>
</table>

Significance test: ** $p < 0.01$, * $p < 0.05$, # $p > 0.05$. 
A stepwise multiple linear regression was applied to select the best RS predictors for the regression models at different phenological stages.

For the biomass regression models, the best $R^2$ values, ranging from 0.519–0.765, were achieved. At the tillering stage, only green and red bands were applied to represent the biomass; while at the booting stage, all four bands as well as the NDVI value were included in the model. At the heading stage, only green and NIR bands were used. Across all three dates, the model for the booting stage (second date), near the height of the growing season, gives the highest $R^2$ value of 0.765.

The stepwise regression analysis for the LAI estimation chose red and green bands as the predictors in the tillering stage, NDVI as the only predicting variable in the booting stage, and all four bands as predictors in the heading stage. The LAI estimation model with the best performance was found at the heading stage when panicles grow outside of the rice stem, with an $R^2$ of 0.65.

Variations in N concentration are more difficult to detect than biomass or LAI development. The N concentration is a weak absorber of spectral radiation and occurs in very small quantities in leaf tissue (0.5–3%) [65]. The red and green bands of the FS-2 image were included in the regression model in the tillering stage, while all four bands were used in both the booting and heading stages. At the tillering and booting stages, when the plant N concentration is at a high level, higher associations were found between the plant N concentration and the RS spectra.

Stepwise regression analysis for N uptake included all four bands as predictors in the tillering and booting stages, whereas only green and NIR were needed in the heading stage. The highest $R^2$ was 0.483 at the heading stage.

5.2.2. Validation of the Regression Models

Model validation was conducted using independent ground truth data sets as mentioned in Section 3.2. Table 6 shows the statistics of the model validation.

<table>
<thead>
<tr>
<th>Agronomic variable</th>
<th>Phenological Stage</th>
<th>Number of Validation Sites</th>
<th>Model Validation Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RE (%)</td>
</tr>
<tr>
<td>Biomass (kg/ha)</td>
<td>Tillering</td>
<td>12</td>
<td>39.4</td>
</tr>
<tr>
<td></td>
<td>Booting</td>
<td>19</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>Heading</td>
<td>19</td>
<td>8.9</td>
</tr>
<tr>
<td>LAI</td>
<td>Tillering</td>
<td>12</td>
<td>31.4</td>
</tr>
<tr>
<td></td>
<td>Booting</td>
<td>17</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>Heading</td>
<td>17</td>
<td>13.6</td>
</tr>
<tr>
<td>N Concentration (%)</td>
<td>Tillering</td>
<td>12</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>Booting</td>
<td>19</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>Heading</td>
<td>19</td>
<td>27.7</td>
</tr>
<tr>
<td>N uptake (kg N/ha)</td>
<td>Tillering</td>
<td>12</td>
<td>39.9</td>
</tr>
<tr>
<td></td>
<td>Booting</td>
<td>19</td>
<td>32.5</td>
</tr>
<tr>
<td></td>
<td>Heading</td>
<td>19</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Validation analyses (Table 6) showed that the regression model for biomass reached the lowest RE of 8.9% and highest IA of 0.68 at the heading stage. Across all three dates, for the biomass, LAI and N
uptake estimation models, the RE value decreases as rice grows. The LAI model in the heading stage showed the lowest RE and highest IA values of 13.6% and 0.85, respectively. These results demonstrated that the models based on the FS-2 images performed better at later growth stages.

The RE for the N concentration model was the lowest (15.3%) at the booting stage, while the IA values increase slightly from 0.41 at the tillering stage up to 0.48 at the later booting and heading stages. For the N uptake models, the lowest RE (11%) and the highest IA (0.77) also occurred at the heading stage.

Across all three dates, relationships between the FS-2 derived values and the interpolated ground truth data were analyzed by scatterplots (Figure 2). The $R^2$ value was the highest (0.98) for biomass modeling results, followed by the N uptake model (0.93), whereas a moderate $R^2$ value of 0.78 and 0.84 was achieved for LAI and plant N concentration estimation. These results showed that the RS derived values were highly related to the ground truth values, and the linear regression trend lines were ideally close to the 1:1 line (dashed diagonal lines in the figures).

![Figure 2](image_url)

**Figure 2.** Relationships between RS derived values and ground truth values for (a) biomass; (b) LAI; (c) N concentration; and (d) N uptake at the tillering (△), booting (○), and heading (○) stages.
5.3. Regional Application of the Regression Models

Finally, the regression models were applied to the identified rice areas to map the within-field variability of the four agronomic variables for the area covered by the FS-2 images in the Qixing Farm County. As an example, the results for July 6 are shown in Figures 3–6. The within-field variability of the rice status can be easily detected in these maps. In the biomass map (Figure 3), unevenly distributed values and patterns can be identified in most of the field blocks, revealing significant within-field variability of the growing status. Overall, a relatively higher biomass level is shown in the southern part.

![Within-field Variability of Rice Biomass on 6 July 2009](image)

**Figure 3.** Within-field spatial variability of biomass derived from the FS-2 image of 6 July 2009.
The LAI map (Figure 4) reveals higher LAI values in the south, especially the southwest, region of the study area. Variations within the field blocks are also well detected in this map.

**Figure 4.** Within-field spatial variability of LAI derived from the FS-2 image of 6 July 2009.

The N concentration distribution map (Figure 5) shows a general random pattern of spatial variation on the intra-field level. Most of the fields had N concentration values of 2.5–3.5%. However, the within-field variability is still presented.
Figure 5. Within-field spatial variability of N concentration derived from the FS-2 image of 6 July 2009.

Likewise, the N uptake map (Figure 6) displays apparent within-field spatial variation. In the south and southeast part of the image, a relatively higher plant N uptake is detected.
6. Discussion

6.1. Band Selection for Different Growth Stages

The recorded reflectance of optical imagery from a vegetated surface is a function of several physical properties such as vegetation structure, soil type, plant moisture content as well as sensor configuration. Plant pigments have distinct absorption spectra and some have an impact upon the reflectance spectra of leaves. This offers the opportunity of characterizing pigment concentration from reflectance spectra [66],

Figure 6. Within-field spatial variability of N uptake derived from the FS-2 image of 6 July 2009.
and thus to interpret the plant status. Chlorophylls are one of the most important pigments because they control the absorption amount of solar radiation which determines the photosynthesis and, consequently, primary production. Several studies have demonstrated that the plant chlorophyll concentration indirectly indicates the plant nutrient status because the molecular structure of chlorophyll incorporates a large proportion of total leaf nitrogen [67–69].

In our results, different spectral bands were selected for the regression models at different growth stages. The red and green band combination performed well for most of the agronomic variables at the tillering stage, whereas green and NIR band combinations performed best for biomass and N uptake estimation at the heading stage. At the earlier growth (tillering) stage, the rice canopy coverage is lower (much less than 100%), resulting in a lower chlorophyll concentration on the pixel scale. In addition, the chlorophyll concentration on the leaf scale is lower compared to that at the heading stage [70,71]. However, at the later (heading) stage, the canopy cover is higher, and the leaf level chlorophyll concentration is higher as well. The blue band did not frequently show up in the stepwise regression models. This may be attributed to the strong scattering effects of the atmosphere on blue radiation.

In addition, we found when the plant N concentration was relatively high, green (520–600 nm) and red (630–690 nm) bands were effective for the regression models. This result conforms to the finding of O’Neill et al. [65] that the highest correlation between N concentration and reflectance occurred at 697.8 nm, followed by the region of 536.8–558.5 nm.

6.2. Background Effects in the Early Stage

According to the model validation results of RES, it can be concluded that the regression models for the tillering stage did not perform as well as the ones for the other two stages, regarding all agronomic variables except for N concentration. The highest RE of N concentration occurred at the heading stage, which mainly was caused by the low N concentration in leaf tissue at that stage, as mentioned in Section 5.2.1.

According to the IA, the regression models performed best in the heading stage for all agronomic variables, probably resulting from the higher biomass density and thus lower field soil and water effects compared to the early stages. This result is in conflict with the findings of Gnyp et al. [72], who utilized hyperspectral field data for rice biomass estimation and achieved better results at the early growth stages. However, in the study of Gnyp et al. [72], a specific vegetation index such as NDVI or RVI (ratio vegetation index) was used as the single independent variable in the regression models. They reported that NDVI became saturated at later growth stages before biomass reached 3 t/ha, which might have affected the model performance. They also noted that at the heading stage, the canopy reflectance signal became more complicated since panicles emerged from the sheath. In this study, the original multi-spectral reflectances in addition to NDVI were both used as the descriptive variables in the stepwise regression, which might help to remove the effects of the NDVI saturation problem on the model performance at later growth stages. Nevertheless, this needs to be further confirmed by analyzing the datasets from the two studies. Additionally, different measurement methods, RS instruments, physical conditions of the rice crop, sun angles, and sensor view angles, may also contribute to the discrepancy in these two studies.

In vegetation studies using satellite RS, several caveats have been noted by Myneni et al. [73]. These caveats include bidirectional effects, atmospheric effects, canopy structure effects, background or soil effects, nonlinear effects of scattering, effects of spectral heterogeneity, adjacent effects, nonlinear
mixing, and topographic effects. In this study, the background soil and water effects and the rice canopy structure significantly affect the spectral response. Canopy conditions at different stages are clearly represented in the following photos (Figure 7). At the early stage, the rice canopy coverage was relatively low and therefore more soil and water background effects occurred. As one of the important canopy structure parameters, leaf orientation characterized by a smaller leaf angle at the early stage might result in the weak performance of the regression models as well.

![Figure 7. Rice growth on 24 June (left), 6 July (middle) and 9 August (right).](image)

7. Conclusions

Due to the integration of multi-temporal FS-2 imagery and GIS data, the rice cultivation areas in the Qixing Farm County were more clearly identified. The overall accuracy of the entire classified map was improved remarkably from 81.8% to 91.6%. This highly accurate rice cultivation map provides an ideal basis for further analyses of rice crops in the study area.

This study showed that the performance of the regression models was significantly affected by rice growth stages. Thus, an optimized band selection for every growth stage is important due to the varying spectral reflectance properties. Based on the $R^2$ values, relatively higher goodness-of-fit values were found in the biomass and LAI estimation models than in the plant N uptake and plant N concentration models. In particular, for the estimation of plant N concentration, better model goodness-of-fit occurred at the earliest growth stage (tillering) when the N concentration was relatively high. The RS derived values and the interpolated ground truth values for all three dates were highly correlated. The most accurate models with the lowest REs and the highest IA values were found at the heading stage for three of the four agronomic variables, except for the N concentration. In conclusion, this study provides a framework and example of how high resolution satellite RS can support agricultural field management.
such as fertilizer, irrigation and pesticides management strategies by providing within-field agronomic information on regional scales. The information derived from satellite RS could be further used to study the relations between crop growth and other phenomena such as carbon fixation, climate change, and sustainable management of natural resources.

Acknowledgments

This work was funded by the Natural Science Foundation of China (NSFC, project No. 31071859), the International Bureau of the German Federal Ministry of Education and Research (BMBF, project No. 1DO12013), and the German Research Foundation (DFG, project no. BA 2062/8-1). The authors gratefully acknowledge Yinkun Yao, Shanyu Huang, Minmin Su and Guangming Zhao from the China Agriculture University; Martin Gnyp and Christoph Hütt from the University of Cologne; Liqin Zhao from the Qixing Modern Agriculture Development Center for their supports in ground truth data collection. The authors are extremely thankful to the cooperating officers and farmers for their assistance in field experiments at the Qixing Farm. The Qixing Research and Development Center and the Jiansanjiang Agricultural Research Station, both in Heilongjiang, China, are gratefully acknowledged. The leading author is supported by the China Scholarship Council (November 2010–October 2014), Beijing, China.

Author Contributions

Quanying Zhao conducted the field campaigns, processed and analyzed the satellite images and the agronomic data. She also was responsible for writing the manuscript and incorporating the co-authors’ contributions and comments. Victoria Lenz-Wiedemann and Fei Yuan intensively contributed to the writing of the manuscript and the data analysis. Rongfeng Jiang, Yuxin Miao and Fusuo Zhang designed and supervised the field campaigns, and contributed to the agronomic data analysis and those parts in the manuscripts. Georg Bareth intensively contributed to the structure of the manuscript, the GIS analysis and the satellite image classification. At different stages of the analysis and writing process, all co-authors contributed ideas for the preparation and improvement of the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).