International Journal of Robotics and Automation, Vol. 32, No. 3, 2017

PARTITIONING THE WHEAT GRAINS BY THE UNIFORMITY OF PROTEIN QUALITY BASED ON REMOTE SENSING

Tian Weixin,* Chen Chen,* Xiaoyu Song,** Dong Ren,* and Jihua Wang***

Abstract

Separating the wheat grains by its protein quality (PQ) uniformity can add the values of the wheat. This paper proposed an approach to predict the PQ uniformity of wheat when they are growing on field and help partitioning them on PQ uniformity by region. We first partitioned the experimental plot as three scales of $10\,\mathrm{m} \times 10\,\mathrm{m}$, $15\,\mathrm{m} \times 15\,\mathrm{m}$, $20\,\mathrm{m} \times 20\,\mathrm{m}$ and used three different methods to ascertain that the best scale to partition the wheat grains by the PQ uniformity was the $10\,\mathrm{m} \times 10\,\mathrm{m}$, and the best method was the M5 model tree. Secondly, we used the M5 model tree on the $10\,\mathrm{m} \times 10\,\mathrm{m}$ scale, and applied the spatial information to reinforce the partition performance. Compared with the traditional approach based on Kriging, the method proposed in the paper improved the stability performance of the traditional approach without losing accuracy.

Key Words

Protein quality, uniformity, remote sensing, M5 model tree

1. Introduction

On the influence of genetic types, environmental conditions, cultivation techniques and other factors, protein quality (PQ) of wheat in a growing region of certain period may vary remarkably, which brings trouble to the use of wheat, because different wheat products require a different PQ of wheat grains. For example, for making streamed buns and making cakes, the proportion of gluten protein affects the qualities greatly. As it is hard to separate the wheat grains after they are mixed, one feasible approach is to harvest them by regions. Marking the regions by PQ distribution of wheat in growing stage is a key step of this approach. Traditionally, the PQ distribution information is obtained by manual sampling. However, manual sampling is a labour-consuming work and cannot get enough samples to obtain a good partition scheme.

With the development of earth observation technology, many methods and applications are developed to monitor crops for different purposes based on remote sensing [1]. The monitoring methods on crops can be roughly put into two classes: point monitoring method and planar monitoring method. The point monitoring method uses only a few point samples and utilizes geostatistics distribution models or interpolation methods to predict information of other points. For example, in 2014 Kuri *et al.* predicted maize yield using dry dekads derived from remote sensing vegetation condition index [2]. Thomas *et al.* presented a fast method for biomethane potential retrievals using near-infrared spectroscopy and partial least square regression which were developed based on two different sets of spectral reflectance data [3]. Jin *et al.* proposed a crop growth model with remote sensing data to estimate maize yield and found high degree to match the samples [4]. Unlike the point monitoring method, the planar monitoring method does not rely on interpolation methods and can get each point data and predict directly. Many research reported that it could get better result than that of the point monitoring method. For example, Kenneth *et al.* developed and evaluated relationships between hyper-spectral remote sensing and lake water quality parameters – chlorophyll, turbidity, Nitrogen and Phosphorus etc. [5]. Bellvert *et al.* determined the lower and upper baselines for calculating crop water stress index for the Chardonnay *etc.* at different phenological stages by high-resolution remote sensing thermal imagery [6]. The planar monitoring method is a promising method and its application is widening, but current researches seemed to overlook the spatial information while applying it. For crops monitoring, the variation is usually of regional aspects and the plants of spatial neighbour location are usually relational. From this point of view, in this paper we adopted the planar monitoring method to evaluate the PQ uniformity of wheat and took the spatial information of wheat into consideration.

While applying remote sensing to crops monitoring, Vegetation Indices (VIs) are almost the indispensable

* College of Computer and Information Technology, Three Gorges University, Yichang 443000, People's Republic of China; email: t_wxin@hotmail.com, 382972312@qq.com, rendong5227@163.com

** National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, People's Republic of China; email: songxy@nercita.org.cn

*** Beijing Research Center for Agrifood Testing and Farmland Monitoring, Beijing 100097, People's Republic of China; e-mail: wangjh@nercita.org.cn

Corresponding author: Tian Weixin

Recommended by Dr. Chaomin Luo

(DOI: 10.2316/Journal.206.2017.3.206-4839)

elements. The VIs are combination of some bands of spectrum. Their effectiveness has been shown by many studies. In 2012, Hazratkulova *et al.* examined genotypic variations for Normalized Difference Vegetation Index (NDVI) at different growth stages and its relationship to yield in winter wheat under terminal heat stress [7]. In 2013, Fatiha *et al.* analysed the spatio-temporal variations of the vegetation dynamics from a series of available multi-date satellite images and described the context of overall vegetation indices such as NDVI, Soil-Adjusted Vegetation Index (SAVI) and Transformed Soil-Adjusted Vegetation Index (TSAVI), and chose the best index [8]. In 2014, Hongxiao Jin and Lars Eklundh used plant phenology index to characterize terrestrial vegetation canopy green leaf area dynamics [9]. In 2015, Liao compared the chlorophyll content prediction quality by various red-edge-based vegetation indices based on the compact airborne spectrographic imager (CASI) data and obtained a high accuracy [10]. NDVI has linear correlation with the vegetation distribution density, and can reflect the influence of plant canopy, such as soil, damp ground, snow, withered leaves, and roughness. NRI (Nitrogen Reflectance Index) can be used to get the inversion of protein content of winter wheat grain. OSAVI (Optimized Soil-Adjusted Vegetation Index) is based on SAVI, and can not only reflect the growing situation of crop, but can also eliminate the disturbance of soil backgrounds. NPCI (Normalized Pigment Chlorophyll Ratio Index) can be used to analyse the changes of canopy chlorophyll content. GNDVI (Greenness Normalized Difference Vegetation Index) can reflect the greenness of canopy. SIPI (Structure Insensitive Pigment Index) can be used to reduce sensitivity and monitor vegetation health. PSIR (Plant Senescence Reflectance Index) can maximum increase the sensitivity of carotene and chlorophyll. We used those vegetation indices as features for the PQ uniformity prediction.

In contrast to crops yield monitoring, the quality uniformity monitoring is more delicate and sensitive. Even if the methods for the crop quality monitoring can be modified and adopted to the crop quality uniformity monitoring, there are many issues deserving to figure out. The key issue of the quality uniformity monitoring is to decide which partition for fields can be suitable for measuring. To solve the problem, we adopted three classical classification model to predict the uniformity of wheat grain protein over data of different scales. The scale that the best average accuracy was obtained on over all the three classifiers would be looked as the best partition scale. The best classifier of the three along with the best partition scale would be further used for the spatial protein uniformity partition. The classifiers used in the paper are the Naïve Bayes model, the M5 model tree and the logistic regression model, which are the most popular methods in the data classification domain. For example, Zhai *et al.* proposed an automatic classification method of cotton blind stinkbug hazard level [11]. Results showed that the Naïve Bayes model had advantages in accuracy and speed with average rate of correct recognition as 90%, which was better than support vector machine and back propagation neural network model. Jaroonrut and Charnchai proposed a system to locate white blood cells in the microscopic blood smear images, segmenting them into nucleus and cytoplasm regions, then extracting suitable features and finally classified them into five types [12]. The overall correction rate of the classification phase was about 94% for Naïve Bayes models. Kisi investigated the accuracy of M5 model tree in modelling pan evaporation [13]. The M5 model tree is also used to achieve steam consumption characteristics [14]. Zhang *et al.* summarized the characteristics of micro-blog search results, and proposed a method using a sort of decision model-binary logistic model, tested the confidence level of the model and estimated the weight of the variable in the model collecting the real samples [15]. Results showed the feasibility of applying model for detecting the relation between user's decision and the factors from each individual micro-blog search as well as ranking. Jun *et al.* used the logistic regression model to identify different kinds of fruits in images based on the pixel classification algorithm [16].

This paper aims at predicting the PQ uniformity of wheat grains before they are harvested using remote sensing. Firstly, we determined that the best partition scale was $10\,\mathrm{m} \times 10\,\mathrm{m}$ and the best classifier was the M5 model tree for the remote sensing image of 0.6 m resolution. Then, we used M5 model tree over the $10\,\mathrm{m} \times 10\,\mathrm{m}$ scale to predict the PQ uniformity with the additional spatial information. Result showed that the approach was better than the traditional interpolation method on stability and can be an alternative for the traditional method. The rest of the paper is organized as follows. In Section 2, we introduced the experimental material processes and some basic methods. In Section 3, we detailed the experimental procedures, results and analysis. In Section 4, we presented the conclusion of this paper and possible future work.

2. Materials and Methods

2.1 The Experimental Plot

The experiment was carried out in Xiaotangshan town, Beijing, located in north latitude of $40°10'$, and east longitude of $116°26'$. The region has temperate continental monsoon climate. The dominant wind direction is north, northwest. Annual rainfall is 570 mm, and the annual evaporation is 2,000 mm. The experimental cultivation variety was selected as the wheat of Jingdong8. The location of experimental plot was a rectangular region from north latitude of 40.1829°, east longitude of 116.4402° to north latitude of 40.1832°, east longitude of 116.4422°. Figure 1 is the remote sensing image, in which the experimental plot was marked with a rectangle.

2.2 Image Processing

The remote sensing image was taken by QuickBird in the stage of heading. QuickBird is a sun-synchronous satellite launched by the DigitalGlobe company in the United States. It carried a high-resolution multi-spectral bands camera. The bands include blue band (450–520 nm), green band (520–630 nm), red band (630–690 nm) and near-infrared band (760–900 nm). The resolution of panchromatic sensor is 0.61 m, and the multi-spectral is 2.4 m. An

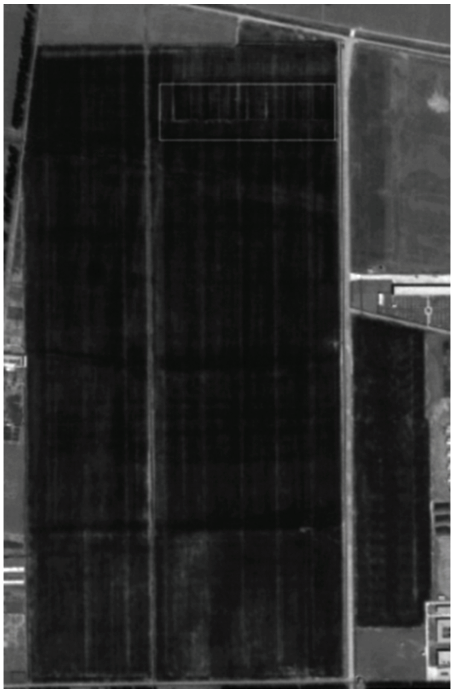

Figure 1. Remote sensing image of the experimental plot.

experiential linear method was taken for atmospheric correction. The field for drying grains was chosen as the bright target, and water area was chosen as the dark target [17]. A reference GPS point was set in the remote sensing image [18]. Then, we used a high-precision Differential GPS to do the geometric correction. Through images fusion, we finally obtained an image of four bands with a resolution of 0.61 m.

2.3 Algorithms

We choose the M5 model tree to predict the uniformity of wheat grains. The M5 model tree is a sectional linear regression model, which was proposed by Wang in 1997 on the basis of Quinlan's version [19]. Since then, it has been successfully applied to solve many prediction problems such as rainfall-runoff, flood forecasting and wave spectrum [20]. The M5 model tree divides the samples along a binary tree. The division criterion of the middle nodes follows the principle of maximizing the difference. For a middle node with n samples, the model calculating *n-1 SDR* (standard deviation reduction) of possible partition positions on every attributes, then the maximum SDR is selected as the threshold [19]. SDR can be calculated as follows:

$$SDR = sd(T) - \frac{|T1|}{|T|} sd(T1) - \frac{|T2|}{|T|} sd(T2) \tag{1}$$

where T is the set of total samples on the current node, $T1$, $T2$ are the two sample sets after partition. Function $sd\,(T)$ calculates the standard deviation of T.

In the leaf nodes of the model, there are linear regression models generated from the allocated subset of samples. The M5 model tree also proposed method for pruning and smoothing the tree. Equation (2) shows the condition for pruning tree:

$$ER = |T|R_{mse} - |T1|R_{mse1} - |T2|R_{mse2} \tag{2}$$

where R_{mse} denotes the mean square root error of the set of total samples on the node, R_{mse1}, R_{mse2} denotes the mean square root error of the two subset on the left and right child nodes, respectively.

The M5 model tree is essentially composed of many linear regression models, and each linear regression model fits only a part of local samples. So in most cases, the performance of the M5 model tree exceeds that of single linear regression models. As a comparison, the Naïve Bayes model and logistic regression model were chosen to process the same data as the M5 model tree in the paper. All the data process experiments were carried out on the *weka*.

2.4 Extracting VIs

To facilitate the classification process, we use some combination of the bands as features instead of single band of the QuickBird remote sensing image. As mentioned in the previous section, the vegetation indices can express the plant growth status. The light of red or blue can be absorbed by the plant leaves, while the light of green or near-infrared band will be well reflected. VI utilizes the operation of different bands boosting some signals or weakening the others to find the intrinsic properties of objects. The NDVI, OSAVI, NRI, GNDVI, SIPI, PSIR and NPCI were chosen in the paper as features to represent the status of the wheat growth. Table 1 shows how to extract VIs from QuickBird remote sensing image.

Table 1
VIs for QuickBird Image

Vegetation index	Bands	Definition
NDVI	B3,B4	$(B_4 - B_3)/(B_4 + B_3)$
OSAVI	B3,B4	$(1 + 0.16)(B_4 - B_3)/(B_4 + B_3 + 0.16)$
NRI	B2,B3	$(B_2 - B_3)/(B_2 + B_3)$
GNDVI	B2,B4	$(B_4 - B_2)/(B_4 + B_2)$
SIPI	B1,B3,B4	$(B_4 - B_1)/(B_4 - B_3)$
PSIR	B1,B3,B4	$(B_3 - B_1)/B_4$
NPCI	B1,B3	$(B_3 - B_1)/(B_3 + B_1)$

Table 2
Scheme of Generating Small Squares

(A) $2u \times 2u$ (Total 30 Small Squares)			
(1)1,2,21,22	(2)3,4,23,24	$\cdots$	(10)19,20,39,40
(11)41,42,61,62	(12)43,44,63,64	$\cdots$	(13)59,60,79,80
$\vdots$	$\vdots$	$\vdots$	
(21)81,82,101,102	(22)83,84,103,104	$\cdots$	(30)99,100,119,120
(B) $3u \times 3u$ (Total 12 Small Squares)			
(1)1,2,3,21,22,23,41,42,43	(2)4,5,6,24,25,26,44,45,46	$\ldots$	(6)18,19,20,38,39,40,58,59,60
(7)61,62,63,81,82,83,101,102,103	(8)64,65,66,84,85,86,104,105,106	$\ldots$	(12)78,79,80,98,99,100,118,119,120

(C) $4u \times 4u$ (Total 10 Small Squares)				
(1)1,2,3,4 21,22,23,24 $\cdots$ 61,62,63,64	(2)5,6,7,8, 25,26,27,28 $\cdots$ 65,66,67,68	(3)9,10,11,12, 29,30,31,32, $\cdots$ 69,70,71,72	(4)13,14,15,16 33,34,35,36 $\cdots$ 73,74,75,76	(5)17,18,19,20 37,38,39,40 $\cdots$ 77,78,79,80
(6)41,42,43,44 61,62,63,64 $\cdots$ 101,102,103,104	(7)45,46,47,48 65,66,67,68 $\cdots$ 105,106,107,108	(8)49,50,51,52 69,70,71,72 $\cdots$ 109,110,111,112	(9)53,54,55,56 73,74,75,76 $\cdots$ 113,114,115,116	(10)57,58,59,60 77,78,79,80 $\cdots$ 117,118,119,120

2.5 Partitioning the Experimental Plot

The experimental plot is a rectangular region composed of 6 rows of units and each row has 20 units. The total 120 units are numbered in row-major order. Considering the total size of the experimental plot, we set the basic unit to the size of $5\,\mathrm{m} \times 5\,\mathrm{m}$, then partitioned the experimental plot into small squares by 2 units $\times$ 2 units, 3 units $\times$ 3 units and 4 units $\times$ 4 units. Table 2 shows the partitioning of the experimental plot.

2.6 Measuring the PQ of Wheat

The PQ of wheat grains is determined by the total protein content, wet gluten content, dry gluten content, water content, sedimentation value and others. After the wheat was ripe, we harvest them by units and put them into mesh bags and numbered them. After each bag of wheat grains were dried, we recorded their weights, grains numbers and grain qualities. The PQ was measured by NIR Instalab-610, we recorded all the protein content, wet gluten content, and dry gluten content of each unit. Those quality data were then corresponded to the VIs by the position coordinates.

2.7 Computing the Uniformity

The variance was used to express discrete degree of the PQ of wheat. The smaller the variance of the same unit is, the more uniform of the PQ is, and *vice versa*. To maintain the consistency, standard deviation was used to measure the uniformity. There was not a universal definition of protein uniformity of wheat yet. In the experiment, we count every standard deviation of small squares. Table 3 shows the average PQ and its standard deviation of every scales of $1u \times 1u$, $2u \times 2u$, $3u \times 3u$ and $4u \times 4u$.

Table 4 shows the maximum standard deviation, minimum standard deviation and average deviation of the PQ in scales $2u \times 2u$, $3u \times 3u$ and $4u \times 4u$ especially.

We defined the uniformity threshold as the average standard deviation of that in scales of $2u \times 2u$, $3u \times 3u$ and $4u \times 4u$, which is shown in Table 5. The threshold of PQ is defined as follows:

$$\begin{aligned} \mathrm{Thres}_{\mathrm{PQ}} &= \alpha * \mathrm{Thres}_{\mathrm{Water}} + \beta * \mathrm{Thres}_{\mathrm{Protein}} \\ &\quad + \gamma * \mathrm{Thres}_{\mathrm{Wet\ gluten}} + \delta * \mathrm{Thres}_{\mathrm{Dry\ gluten}} \\ &\alpha + \beta + \gamma + \delta = 1, \quad \alpha, \beta, \gamma, \delta \geq 0 \end{aligned} \tag{3}$$

The PQ of a small square would be labelled as uniformity when the standard deviation less than the threshold, otherwise would be labelled as non-uniformity.

3. Experiment and Results

3.1 Scale

The remote sensing method proposed in this paper partitions the target field into small squares like bitmap. The

Table 3
Statistical Properties of PQ Distribution

		Water	Protein	Wet Gluten	Dry Gluten
$1u \times 1u$	Ave(%)	9.95	16.75	39.53	12.02
	Stdev	0.25	0.81	3.21	0.82
$2u \times 2u$	Ave(%)	9.95	16.75	39.53	12.02
	Stdev	0.13	0.44	1.74	0.48
$3u \times 3u$	Ave(%)	9.93	16.76	12.05	39.59
	Stdev	0.08	0.23	0.37	0.77
$4u \times 4u$	Ave(%)	9.96	16.66	39.10	11.94
	Stdev	0.08	0.26	0.81	0.22

Table 4
Standard Deviation of PQ in Different Small Squares

		Water	Protein	Wet Gluten	Dry Gluten
Stdev of $2u \times 2u$	Max	0.43	1.55	5.28	1.12
	Min	0.05	0.10	0.32	0.13
	Ave	0.19	0.57	2.34	0.60
Stdev of $3u \times 3u$	Max	0.36	1.21	4.35	0.97
	Min	0.11	0.27	1.52	0.52
	Ave	0.23	0.76	3.09	0.74
Stdev of $4u \times 4u$	Max	0.33	1.04	4.16	1.15
	Min	0.15	0.40	1.80	0.46
	Ave	0.22	0.71	2.95	0.77

Table 5
The Threshold of Uniformity

	Water	Protein	Wet Gluten	Dry Gluten
Threshold of uniformity	0.21	0.68	2.79	0.70

pixels of the remote sensing image are then mapped to the small squares. The size of each small square is important and depends on the spatial resolution of the remote sensing image and the variation degree of the protein of the wheat. In this section, we tried to find out the best size of the small squares by the experiment. As mentioned in the previous section, there were three ways of partition: $2u \times 2u$, $3u \times 3u$, $4u \times 4u$. For each way of partition, a same process was applied to calculate the protein uniformity of the wheat and the vegetation index such as NDVI, OSAVI and NRI. Naïve Bayes, M5 model tree and logistic regression were used to predict the protein uniformity of wheat growing in the corresponding small square.

We used *Weka3.8* for data process and analysis. Eighty per cent data was set for training and 20% data for test and cross-validation was used. The experiment results are shown in Table 6.

According to the result, the $2u \times 2u$ was chosen as the best scale for remote sensing monitoring, and the M5 model tree was chosen as the best classifier for the following steps.

3.2 Predicting Uniformity Using Spatial Information

3.2.1 Method Based on Variation Function

Variation function is an important method in the geostatistics to study the spatial variation features of the regionalized variables. It is defined as expectation of the square of the regionalized increments. Variation function combines Kriging method can be used to almost any occasion which needs to estimate the distribution of point data on the ground. Variation function is the traditional method used to predict the distribution of the wheat quality or quantity before the large scale application of remote sensing technique in agriculture. The general form of variation function is:

$$g(x,h) = \frac{1}{2}Var[Z(x) - Z(x+h)] \tag{4}$$

Table 6
The Results of Three Methods

(A) Result of Naïve Bayes			
	2×2	3×3	4×4
Protein	77%	50%	60%
Water	83%	83%	60%
Wet gluten	70%	75%	60%
Dry gluten	57%	42%	40%
(B) Result of M5 Model Tree			
	2×2	3×3	4×4
Protein	90%	50%	60%
Water	87%	75%	60%
Wet gluten	83%	67%	60%
Dry gluten	73%	50%	40%
(C) Result of Logistic Regression			
	2×2	3×3	4×4
Protein	83%	83%	40%
Water	73%	83%	40%
Wet gluten	73%	92%	40%
Dry gluten	67%	67%	40%

where x is the point on a one-dimensional axis, and h is the distance from the current point. Then all uniformity of the wheat on the experimental plot can be estimated on the basis of few samples. In GS+ software, 10 samples were randomly selected from the $2\text{u} \times 2\text{u}$ dataset in each turn as input. The distance h was set to 3 m and the step was set manually. Table 7 shows the result of the accuracy of three runs.

3.2.2 Method Based on M5 Model Tree Using Spatial Information

The uniformity of wheat grown on certain area of field always relates to its spatial location. The PQ uniformity of the crops adjacent to each other usually is approximately the same. That is, the uniformity varies not very sharply. On the map, the distribution of uniformity may seem like some hotspots of non-uniformity spread from the centre to around. To express this property, we add a new attribute to original training and test data. The original data contained seven VIs as attributes. The new attribute is calculated as follows:

$$\theta = \frac{1}{4}(\alpha + \beta + \gamma + \delta) \tag{5}$$

where α, β, γ, δ are the attributes of the four neighbours.

Table 8 shows the result of accuracy of spatial M5 model tree.

Table 7
Accuracy of the Variation Function

	Protein	Water	Wet Gluten	Dry Gluten
1	25% (0.45)	100% (0.70)	50% (0.70)	63% (0.70)
2	88% (0.55)	88% (0.72)	75% (0.52)	50% (0.82)
3	75% (0.80)	88% (0.81)	63% (0.69)	63% (0.80)
Average	63%	92%	79%	58%
stdev	0.272	0.056	0.102	0.061

Note: The figures in the brackets are step lengths.

Table 8
Accuracy of Spatial M5 Model Tree

	Protein	Water	Wet Gluten	Dry Gluten
1	75%	100%	50%	63%
2	75%	75%	63%	75%
3	75%	88%	63%	63%
Average	75%	88%	58%	67%
stdev	0	0.102	0.061	0.057

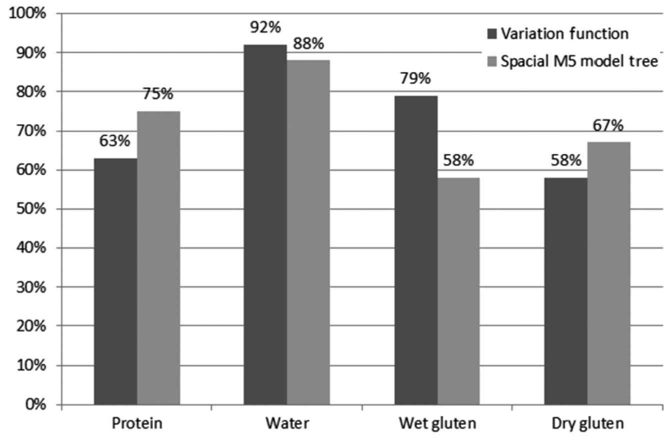

Figure 2. Comparison of spatial M5 model tree and variation function results.

Figure 2 shows the comparison of the average accuracies by variation function and special M5 model tree.

From the result, we can find that the content of water is more relevant to the vegetation index than that of the protein, wet gluten and dry gluten. The average accuracy by the M5 model tree is close to that of the variation function method, while the standard deviation is less. It can be inferred that the performance of the M5 model tree is more stable. The two methods are quite different from the theories to applications. The variation function method is an interpolation method, and it does not need any attributes of the unknown samples themselves. It predicts the unknown attributes only basing on the geographic positions of the unknown samples from the known samples. The variation function method is not only a common approach in agriculture, but also a widespread approach in geology, environment and forestry, especially when only a few point samples can be obtained. The M5 model tree is a data-driven method, and it can explore the relationship among the sample attributes. It also utilizes the geographic information of samples, but only looks them as one of attributes. The M5 model tree method can make use of more information than that of the variation function method. As the data acquisition technology in agriculture is increasing, the data-driven method will be widely used. The result of the paper shows that the performance of M5 model tree is parallel to and not better than that of the variation function method. There are mainly two possible reasons accounting for that. One is that the remote sensing image we used is multi-spectral but not hyper-spectral, which causes the accuracy relatively low. The other is that the attributes we chosen are probably not very typical. Those problems will be ascertained and solved in the future research.

4. Conclusion

More and more large-scale planar spectral information are obtained by remote sensing in agriculture, which appeals for the efficient process. The traditional approach to solve those related to spatial information mainly based on the Kriging interposition methods. In this paper, we first partitioned the experimental plot into three scales, and used three different classifier models to ascertain the best scale to partition the wheat grains by the PQ uniformity and the best classifier model. Secondly, we used the M5 model tree on the $10\,\mathrm{m} \times 10\,\mathrm{m}$ scale, and then applied the spatial information to reinforce the partition performance. Compared with the traditional approach based on Kriging, the method proposed in the paper was better in the stability performance, while not any worse in accuracy. Unlike the Kriging approach that depends on the sample selection while putting to real application, the method proposed in this paper is easier to use.

While calculating the uniformity, we self-defined a criterion to evaluate the uniformity of PQ. This criterion was based on the experiment and may be lack of wide applicability. However, the way of defining the uniformity proposed in the paper can be generalized. While applying the method to a different field, what only need to be done is to count the standard deviation of the small squares first. Even so, we looked forward an authorized definition of the uniformity of PQ. The contribution of the paper is to propose a data-driven method that can act as an alternative for the traditional Kriging interposition methods.

In the future, there are two directions of research we hold to extend from the paper. One is to apply the method to the hyper-spectral images. More and more hyper-spectral images from satellites or UAV are accessible presently, and they are more legible and informing. The other is to try more attributes of samples and combine the traditional variation function to utilize the spatial information with the current data-driven method.

Acknowledgement

This paper was supported by NSFC41371349, NSFC 61105025 and National key research and development program (2016YFD0800902).

References

[1] C.J. Zhao, Advances of research and application in remote sensing for agriculture, *Transactions of the Chinese Society for Agricultural Machinery*, *12*, 2014, 277–293.

[2] F. Kuri, A. Murwir, K.S. Murwira, and M. Masochab, Scientific predicting maize yield in Zimbabwe using dry dekads derived from remotely sensed Vegetation Condition Index, *International Journal of Applied Earth Observation and Geoinformation*, *33*, 2014, 39–46.

[3] U. Thomas, D. Philippe, B. Christian, R. Franz, *et al.*, Retrieving the bioenergy potential from maize crops using hyperspectral remote sensing, *Remote Sensing*, *5*, 2013, 254–273.

[4] H.A. Jin, J.D. Wang, and Y.C. Bo, Estimation on regional maize yield based on assimilation of remote sensing data and crop growth model, *Transactions of the CSAE*, *28*(6), 2012, 162–173.

[5] S. Kenneth, J. Gab-Sue, L. Robert, and S.E. John, Long-term agroecosystem research in the Central Mississippi River Basin: Hyperspectral remote sensing of reservoir water quality, *Journal of Environmental Quality*, *44*, 2015, 71–83.

[6] J. Bellvert, J. Marsl, J. Girona, and P.J. Zarco-Tejada, Seasonal evolution of crop water stress index in grapevine varieties determined with high-resolution remote sensing thermal imagery, *Irrigation Science*, 33, 2016, 81–93.

[7] S. Hazratkulova, R.C. Sharma, S. Alikulov, S. Islomov, *et al.*, Analysis of genotypic variation for normalized difference vegetation index and its relationship with grain yield in winter wheat under terminal heat stress, *Plant Breeding*, *131*(6), 2012, 716–721.

[8] B. Fatiha, A. Abdelkader, H. Latifa, and E. Mohamed, Spatio temporal analysis of vegetation by vegetation indices from multi-dates satellite images: Application to a semi arid area in ALGERIA, *Energy Procedia*, *36*, 2013, 667–675.

[9] H. Jin and L. Eklundh, A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment*, *152*, 2014, 512–525.

[10] Q.H. Liao, Chlorophyll content mapping of crops in Heihe River Basin based on hyper spectral vegetation indices, *Chinese Society of Agricultural Engineering*, *31*, 2015, 159–163.

[11] Z.F. Zhai, Z. Xu, X.Q. Zhou, L.L. Wang, *et al.*, Recognition of hazard grade for cotton blind stinkbug based on Naive Bayesian classifier, *Chinese Society of Agricultural Engineering*, *31*, 2015, 204–211.

[12] P. Jaroonrut and P. Charnchai, Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers, *BioMedical Engineering OnLine*, *14*(1), 2015, 63–81.

[13] O. Kisi, Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree, *Journal of Hydrology*, *528*, 2015, 312–320.

[14] Y.Y. Gu and L.X. Hu, The application of M5 model tree on optimal load distribution in thermal power plants, *Energy Conservation Technology*, *31*, 2013, 125–131.

[15] J. Zhang and W.J. Hou, Research and analysis of method of ranking micro-blog search results based on binary logistic model, *Lecture Notes in Computer Science*, 7719, 2013, 830–842.

[16] S.S. Jun, S.L. Won, and R. Ehsani, Postharvest citrus mass and size estimation using a logistic classification model and a watershed algorithm, *Biosystems Engineering*, *113*(1), 2012, 42–53.

[17] A. Martin and M. Reza Emami, Just-in-time cooperative simultaneous localization and mapping: A robust particle filter approach, *International Journal of Robotics and Automation*, *29*(2), 2014, 66–78.

[18] A. H. Jabbari, O. Giuseppe, and B. Hossein, An adaptive scheme for image-based visual servoing of an underactuated UAV, *International Journal of Robotics and Automation*, *29*(1), 2014, 92–104.

[19] Y. Wang and I.H. Witten, Induction of model trees for predicting continuous classes, *Poster Papers of the Ninth European Conference on Machine Learning*, Prague, Czech Republic, April 23–25, 1997, 128–137.

[20] S. Sakhare and M.C. Deo, Derivation of wave spectrum using data driven methods, *Marine Structures*, *22*(3), 2009, 594–609.

Biographies

Tian Weixin is an associate professor in Computer and Information Technology School, Three Gorges University, China. His research interests are data mining and intelligent algorithm and intelligent information processing.

Chen Chen is a graduate student of Three Gorges University, China. His research domain is data mining. He had worked at National Engineering Research Center for Information Technology in Agriculture as an intern for two years, researching and developing software for processing remote sensing data.

Xiaoyu Song received the Ph.D. degree in geographic information system and remote sensing at Beijing Normal University, Beijing, China, in 2009. She is currently a research associate with the Beijing Research Center for Information Technology in Agriculture, Beijing, China. Her current research interests include remote sensing image process, crop growth monitoring and precision agriculture application by remote sensing.

Dong Ren received his Ph.D. degree in Jilin University. He is a vice professor in the College of Computer and Information Technology at China Three Gorges University, Yichang, China. His research interests include Artificial intelligence, Pattern recognition and 3S technology.

Jihua Wang he received his Ph.D. degree in China Agricultural University. He is a director in Beijing Academy of Agriculture and Forestry. His research interests include integration of agricultural product quality detection and information technology.