Identifying and characterizing yield limiting soil factors with the aid of remote sensing and data mining techniques

Yi-Ping Wang · Yuan Shen

Abstract Soil provides crop with nutrients, water and root support. But, soils vary a great deal in terms of origin, appearance, characteristics and production capacity. Better understanding of the causality between yield and yield-limiting soil factor(s) is essential for site-specific crop management. The objectives of this study were deriving a spatio-temporal yield trend map of a 144 km^2 paddy rice growing region located at an alluvial plain in southwestern Taiwan from satellite images and exploring the potential yield-limiting soil factor(s) in conjunction with general soil survey data. Due to the complexity of data sets, classification and regression trees analysis (CART) was used to relate soil characteristics to yield classes in the spatiotemporal yield trend map, and followed by comparisons of soil characteristics between those consistently-high and -low yielding areas to explore the interactions between yields and soil properties. Through the above data mining analysis, high soil pH, severe leaching loss of applied nitrogen fertilizers, and excessive reductive root environment were suspected to be the major soil related low-yielding mechanisms spread within studied region. Soil characteristics that induced these low-yielding mechanisms were identified and mapped. Error analysis indicated that 61.8 % of the consistently low-yield areas could be correctly identified by just a few soil characteristics. Improvements of management practices to alleviate the negative effects on yields were also proposed based on the identified low yielding mechanisms. Our study highlighted the pressing need and possible methodologies to adjust management strategies for narrowing yield variability and increasing crop production.

Keywords Paddy rice · Yield limiting factor · Soil survey · Remote sensing · Data mining

Y.-P. Wang · Y. Shen (✉)
Department of Soil and Environmental Sciences, National Chung-Hsing University, Taichung, Taiwan, ROC
e-mail: yshen@nchu.edu.tw

Abbreviations

CART	Classification and regression tree analysis
CV	Coefficient of variation
H	Consistent-high yield class
ID	Internal drainage grade
L	Consistent-low yield class
M	Consistent-average yield class
N	Nitrogen
SD	Standard deviation
SPOT	Satellite Pour l'Observation de la Terre
SY	Standardized yield
T1–T4	Soil texture grade at four layers
V	Inconsistent yield class

Introduction

Due to the limitation of land and water resources available and suitable for crop production, improvements in crop productivity are needed to allow food production in keeping pace with the demand (Van Ittersum et al. 2013). Therefore, quantifying the difference between potential and actual yields and identifying the key causes of the yield gap are very important in order to prioritize efforts in research and extension for management changes that can raise crop productivity (Lobell 2013). However, many challenges existed in quantifying the yield gap and understanding the key causes. For example, scales between the reported actual yields and estimated yield potential are often mismatched in the measurement of yield gaps (Lobell et al. 2009), the fields analyzed for yield limiting causes may not be representative of the entire region when considering the spatial and temporal heterogeneity of agricultural landscapes (Hochman et al. 2012), or the year(s) in which the study was done may not be representative of the full range of weather conditions (Dobermann et al. 2003; Lobell et al. 2007).

Potential yield is the yield of a crop cultivar grown under the environments where water and nutrients are optimally supplied and completely protected against any growth-reducing factors (Van Ittersum and Rabbinge 1997). Therefore, at a given location, crop yield potential is determined only by genetic traits and climatic factors (e.g. incoming solar radiation and temperature). Instead, commonly observed spatial variability of actual yields within fields reflects the effects and interactions of local yield-limiting factors (Lobell 2013; Van Ittersum et al. 2013). Lacking proper irrigation facilities or shortage of fertilizers and other agrochemicals may be the primary yield limiting factors in some part of the world (Kropff et al. 1993; Tittonell et al. 2008; Zwart and Leclert 2010). But, soil, as the medium of root growth and source of essential nutrients and water, must have played an important role in determining the local variations in yield performance in areas having well developed irrigation system and farmers with adequate knowledge and abilities in applying those needed agrochemicals, such as in Taiwan and many other developed countries. Therefore, a clear understanding of the cause–effect relations between soil properties and crop yields is a prerequisite in developing proper site-specific crop management strategies, particularly in those areas where intensive agricultural practices have already been implemented.

The capabilities of airborne or satellite-mounted sensors in providing large coverage, more spectral bands, and frequent visits have made remote sensing a useful tool for dealing with spatial and temporal heterogeneity in comparison with traditional human surveys. Numerous approaches for estimating crop yields with remote sensing have been reviewed by Moulin et al. (1998) and Gallego et al. (2010). Although remote sensed crop yields may be less accurate than those measured at fields, the unprecedented spatial and temporal coverage of remote sensing can often outweigh the negatives for many applications (Lobell 2013). A simple way to gain insight into the key low-yielding causes is to examine the spatiotemporal trend of yield performance over several crop seasons, since factors that are not dominant will tend to cancel out across years. For example, using multiyear yield maps derived from satellite images, Wang et al. (2012) mapped areas of consistent-high, -average, and -low yield, and inconsistent areas, for a 100 ha paddy rice study site in central Taiwan. After collecting soil and plant samples from those consistent-high and -low areas and grouping these data into yield classes, severe leaching loss of applied N fertilizers was suggested to be the major soil limiting mechanism affecting the spatial variation of yield within the study site. The systematic approach developed by Wang et al. (2012) identified and characterized the yield limiting soil factors in great detail for a small interested area. However, the extensive labor and time required for soil and plant samples collection and laboratory analysis may limit its application to much larger areas.

Soil survey examines the basic soil properties (such as parent material, texture, color, internal drainage, pH, salinity, etc.) and their spatial distribution over an extensive region, and delineates boundaries of soils having similar characteristics on a map (Soil Survey Division Staff 1993). Soil surveys, at scales of 1:20 000 to 1:50 000 have already been conducted in many countries to provide comprehensive soil resource information for general agricultural planning purposes. But, in most cases, soil maps at these scales are not considered adequate from a precision agriculture perspective because of the poor purity of mapping units (Franzen et al. 2002) and lack of a simple correspondence between yields and soil properties. Intensive grid samplings that ignore soil mapping units may identify more of the variation in soil types across smaller landscapes and are much more accurate and reliable for making decisions at the farm-level, but may not be cost effective due to the significantly higher expenses for sample collection and analysis (Brevik et al. 2003). However, the assembled information by general soil surveys should still have value in providing insights into soil capabilities and limitations on a regional scale. Therefore, to be economically efficient, any new intensive sampling project should first consult and utilize general soil survey data in order to reduce the number of samples to be collected as well as the items to be analyzed.

As discussed above, consistent-low yielding areas within a region can be influenced by one or more soil factors. Furthermore, the yield limiting soil factor(s) may differ from area to area. For example, an area of consistently low yield may be due to strong soil acidity; although poor internal drainage may be the limiting factor for another area. In Taiwan, fertilization practices are generally recommended based on rice varieties and surface soil texture and pH provided no detailed soil testing data available. But, commonly observed spatial variations in yields indicate better delineation of soil management zones is required. Therefore, the goals of this study were to address the potential value of satellite-based remote sensing for efforts to measure and explain yield performance, and to determine whether patterns of soil profile characteristics assembled by general soil surveys are associated with areas of consistent-low and -high yield areas, specifically. Knowledge of basic soil physical and chemical processes was also applied to elucidate the functioning of suspected yield limiting soil factors. It was authors' intention that the findings of this study

could delineate better soil management zones and serve as guidance for designing more efficient soil sampling and analysis plans and conducting field trials for management practices improvement in subsequent years.

Materials and methods

Description of studied region

A major paddy rice (*Oryza sativa* L.) producing region was selected for this study (Fig. 1). The studied region located within an alluvial plain in southwestern Taiwan and covered an area of 12 km × 12 km, of which 47.9 % are used for paddy rice cultivation. The parent materials from which soils formed were mostly sandstone-shale alluvium and schist diluvium. Haplaquepts, Ochraualfs, Udifluvents are major soil great groups existed within the region (Sheh and Wang 1989). The climate was classified as humid subtropical climates (Cwa) by Köppen system (Chiou et al. 2004).

Conventional two-season paddy rice cropping systems are practiced in the studied region. Transplantation generally begins in early February and late July, with harvesting in late June and late November for the first and the second crop seasons, respectively. Midterm drainage at late tillering stage followed by intermittent irrigation at 2 week intervals until harvest is the common water management practices adopted in the region.
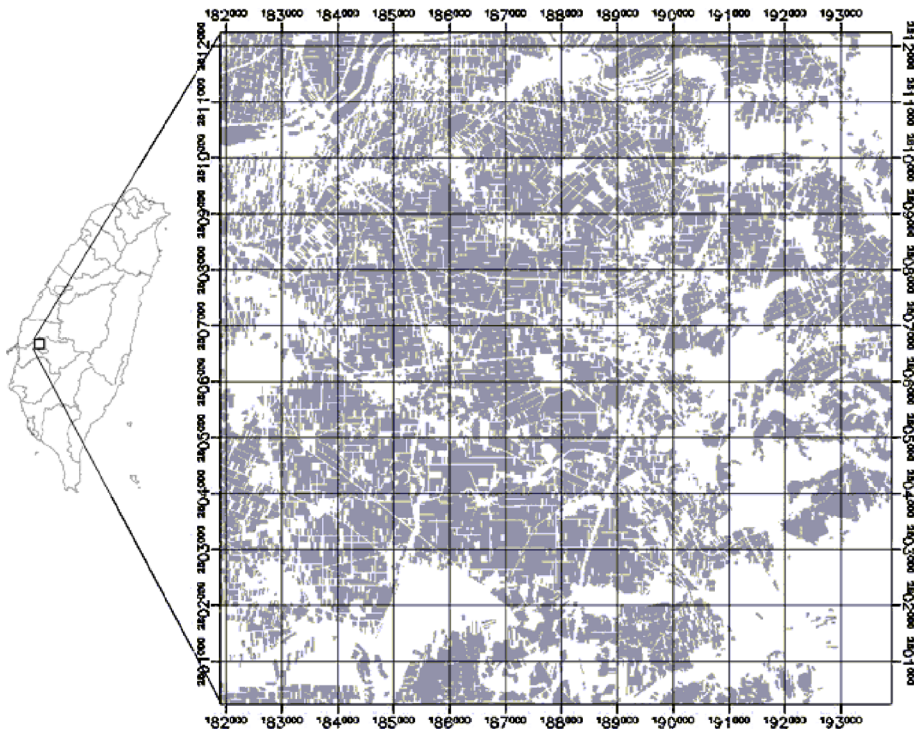

Fig. 1 Schematic diagram, with Universal Transverse Mercator (UTM) coordinates, indicating the studied region. The *gray color* illustrates the areas planted with paddy rice

Fertilizers, particularly the nitrogen fertilizers, applied per crop season usually exceed the local recommendation ($N:P_2O_5:K_2O = 120:70:70$ kg ha^{-1}). Commonly, about 40 % of nitrogen fertilizer (ammonium sulfate or urea), 100 % of phosphate fertilizer (superphosphate) and 20 % potassium fertilizer (muriate of potash) are applied as preplanting basal, whereas the remaining nitrogen and potassium fertilizers are split into 2–3 doses and top dressed before primordia formation. Insects, diseases, and weeds were controlled by regularly spraying of appropriate agrochemicals throughout the growth period to avoid yield loss.

Spatiotemporal rice yield trend maps of studied regions

To obtain the multiyear yield maps for the studied region, level-3 SPOT-2 and SPOT-4 images ranging from panicle initiation and the heading stages in every 16 crop seasons from 2004 to 2011 (two crop seasons in each year) were purchased from the Center for Space and Remote Sensing Research, National Central University, Taiwan. Procedures for atmospheric correction of acquired SPOT images and derivation of rice yield maps from corrected images followed those described by Wang et al. (2010). The 16 historical yield maps, representing each crop season from 2004 to 2011, retrieved from satellite imageries were then subjected to spatial and temporal trend analysis similar to those outlined in Wang et al. (2012).

First, yield data were standardized based on the mean yield of corresponding year and season using Eq. 1.

$$SY_i = (y_i/\tilde{y}) \tag{1}$$

where SY_i is the standardized yield at location i, y_i is the estimated yield derived from satellite image, and $\tilde{y}$ is the regional mean. Yield data from the first or second crop seasons were not distinguished in this study because Pearson's correlation analysis (data not shown), computed for pixels that had complete 16 yield dataset, suggested that the overall averaged standardized yield value was a better yield predictor than averages based on crop seasons for each pixel.

Considering the yield data in some locations might be missing due to cloud obscuration, or be damaged by meteorological disasters, and the fields could be planted with other crops or be used for other purposes, a discriminating process based on normalized difference vegetation index (NDVI) values was applied to determine whether pixels were actually in rice paddies and suitable for yield estimation (Wang et al. 2012). The pixels which had at least eight yield data were considered as effective pixels for further analysis. The eight (or more) standardized yield values of each effective pixel were then averaged to create a mean standardized yield map, while the coefficient of variation (CV) was utilized to determine its yield temporal variability. Pixels with CV ≤ 30 % were classified as consistent in yield performance whereas those with CV > 30 % were classified as inconsistent (Wollenhaupt et al. 1997).

In order to remove the high frequency variations within the derived mean standardized yield map and preserve the low frequency components of the spatial trend, a low pass filter was implemented on each effective pixel by replacing its value with the average of all kernel elements which were also effective pixels. The kernel size was 3 × 3. For each yield consistent pixel, a yield class was then assigned based on its ranking in the smoothed mean standardized yield values. The pixels with yield values in the top quartile were

considered consistent-high (H), values in the lowest quartile were considered consistent-low (L), and values in the middle two quartiles were considered consistent-average (M).

Soil survey data

Data of detailed soil survey on cultivated lands, carried out successively by counties from 1960s to 1970s and digitized by Taiwan Agriculture Research Institute (TARI) in 1990s (Guo et al. 2009) to a GIS database, were used in this study. The surveys, at scale of 1:25 000, used soil series as the lowest soil taxonomic unit. The survey also utilized soil phase and/or soil types and/or soil associations and/or soil complexes as the mapping units. The minimum size of a mapping unit was about 1 ha (100 m × 100 m) which is not only larger than a paddy field (∼0.2 ha = 80 m × 25 m) but also much larger than a pixel (0.04 ha = 20 m × 20 m). If a pixel happened located across mapping units, soil characteristics of the mapping unit contributing most of the area were assigned to that pixel. The digitized soil survey data includes four categorical data, i.e. parent material, soil appearance, soil geomorphology and soil reaction, and six ordered data, namely, acidity of surface soil (pH), soil textures at depths of 0–30 cm (T1), 30–60 cm (T2), 60–90 cm (T3), and 90–120 cm (T4), and internal drainage (ID) (Table 1). The internal drainage conditions were modified by using a classification scheme based on the shallowest depth at which rusty-colored soil mottles were observed (Sun and Yang 1989) to avoid the inconsistent classification systems used between counties when the surveys were conducted. Although the soil surveys were conducted approximately 40 years ago, the intrinsic soil properties should have not changed easily by time. Additionally, these were the most complete and detailed soil survey data available to the authors. Soil properties for every effective pixel in the spatiotemporal yield trend map were retrieved from the spatially joined soil GIS database and used in the following data mining analysis.

Data mining analysis

We used classification and regression trees (CART) analysis to uncover relationships and interactions between the soil physical and chemical properties to yield classes. CART is a nonparametric statistical approach that recursively partition the dependent data into a series of descending left and right child nodes to find increasingly homogeneous subsets based on splitting criteria using variance minimizing algorithms (Breiman et al. 1984). CART can be used to split highly skewed or multi-modal numerical data, as well as categorical data with either ordinal or non-ordinal structure. In addition, CART uses only one predictor variable to split data at a node, whereas the variables used may be selected again for other nodes to generate the classification trees of hierarchical nature.

Statistical analyses were performed by Classification TREE module (STATISTICA Version 10, StatSoft, Inc.) to generate binary classification trees of hierarchy in nature with yield classes (i.e. H, M, L) as dependent variables. All the categorical and ordered soil properties listed in Table 1 were used as predictors. The split selection method used was C&RT-style exhaustive search for univariate splits, with goodness-of-fit estimated by Gini measures (Breiman et al. 1984). Estimated prior probabilities and equal misclassification costs were the specified options in the analysis. A node would stop splitting if it contained more than one yield class and had less than five cases (minimum N specified). The classified tree was then pruned by the build-in misclassification error minimization function.

Table 1 Attributes of the categorical and ordered soil properties existed within the studied region and corresponding codes

Code	Ordered data		Categorical data			
	pH	Soil texture	Parent material	Soil appearance	Soil geomorphology	Soil alkalinity
1	<4.5	coS, S	Schist diluvium	No description	Alluvial soil	Carbonates not existed
2	4.6–5.0	fS, LcoS, LS	Sandstone -shale alluvium	Sedentary soil	Sand hill	Have lime tuberculosis
3	5.1–5.5	LfS, coSL, SL, fSL		Red and brown soil color	Red soil (Fe/Mn nodules existed within top 60 cm)	Alkaline soil
4	5.5–6.0	vfS, LvfS, vfSL		Red and yellow soil color		
5	6.1–6.5	Si, SiL				
6	6.6–7.3	L				
7	7.4–7.8	SCL				
8	7.9–8.4	CL, SiCL				
9	8.5–9.0	SiC				
10	>9.1	C				

coS coarse sand, *S* snad, *fS* fine sand, *LcoS* loamy coarse sand, *LS* loamy sand, *LfS* loamy fine sand, *coSL* coarse sandy loam, *SL* sandy loam, *fSL* fine sandy loam, *vfS* very fine sand, *LvfS* loamy very fine sand, *vfSL* very fine sandy loam, *Si* silt, *SiL* silty loam, *L* loam, *SCL* sandy clay loam, *CL* clay loam, *SiCL* silty clay loam, *SiC* silty clay, *C* clay

Generally, each terminal node classified contained noises, i.e., different dependent variables of various numbers of cases existed in the same terminal node. In addition, a terminal node was named conventionally after the dependent variable which had the largest number of cases in the node. In this study, numerous classified terminal nodes were named as M because half of the pixels were classified as the M yield class based on the yield ranking. Sometimes, a terminal node would be named as H but also contained a significant amount of L cases, and vice versa. In order to reduce the complexity in identifying the major yield limiting soil factors due to the ambiguity caused by the naming convention, the following three rules were applied in screening the terminal nodes to be used for searching the systematic patterns of soil properties for those consistent-high and -low yielding areas with significance. (1) The number of total cases in the node must be higher than a threshold value (set at 50 in this study) to exclude those fragmented areas. (2) For terminal nodes named as H or L: the cases of the H (or L) class must be at least twice that of the L (or H) class to make sure that the H (or L) was the dominant yield class. (3) For terminal nodes named as M: in order to increase the database in discussing possible yield limiting mechanisms, a M terminal node would be grouped together with those H (or L) terminal nodes if the cases of H (or L) class in this particular M node were more than half of the cases in M class and the cases of H (or L) class were at least twice that of the L (or H) cases. Since each pixel covered an area of 0.04 ha, setting 50 as the threshold value for rule #1 would theoretically exclude areas smaller than 2 ha if all the cases classified into the node were spatially well connected. But detailed location examination of pixels excluded by rule #1 indicated that they generally spread far apart and the largest patches were usually smaller than 1 ha. The addition of rule #3 increased the area used for soil characteristics analysis nearly five times, which provided great assistances in sorting out the possible yield limiting factors.

The soil properties utilized to split into those selected terminal nodes, which were called as effective H and L terminal nodes, were compiled together in a table by tracing back along the hierarchical order of the classified tree. From the compiled table, possible yield limiting soil factors and corresponding mechanisms were explored by comparing the differences of soil properties between those effective H and L terminal nodes.

Accuracy and error analysis

The accuracy of the proposed yield limiting soil factors and mechanisms as well as the corresponding omission and commission errors were evaluated by Eqs. 2a–2c via overlapping maps of areas which corresponded to the speculated yield limiting soil factor(s) onto the derived spatiotemporal yield trend map.

$$\text{Accuracy} = L_{included} / L_{total} \tag{2a}$$

$$\text{Omission Error} = L_{omitted} / L_{total} \tag{2b}$$

$$\text{Commission Error} = H_{included} / H_{total} \tag{2c}$$

where L_{total} and H_{total} were total numbers of pixels assigned as L or H yield class in the studied region, $L_{included}$ and $L_{omitted}$ were the numbers of L yield class pixels located within or outside the expected low-yield areas, and $H_{included}$ was the numbers of H yield class pixels located within the expected low-yield areas, respectively.

Results and discussion

Temporal changes of yields in studied region

During the study period from 2004 to 2011, the median estimated grain yield values ranged from 5.81 t ha^{-1} in 2011 to 8.37 t ha^{-1} in 2006 for the first crop season, and from 4.31 t ha^{-1} in 2009 to 7.67 t ha^{-1} in 2008 for the second crop season (Fig. 2). Although large yield variations were observed as the result of yearly changes in weather conditions, the average estimated grain yields for the first and the second crop season were all about 7.2 and 5.7 t ha^{-1}, respectively. Yield differences between the two crop seasons are considered normal in Taiwan because of different climatic environments in the first and second crop season (Hsieh and Liu 1979).

Due to the flat land feature, the synoptic weather conditions should be considered the same at different locations within the study region. In addition, the agronomical characteristics of cultivars grown and management practices used (such as planting date, irrigation, and fertilization rate and timing) were all very similar (data not shown). There was also no apparent occurrence of plant diseases due to the regular pest and disease controls practiced. Therefore, soil factors not easily affected by management practices should have contributed to the observed yield range mentioned above.
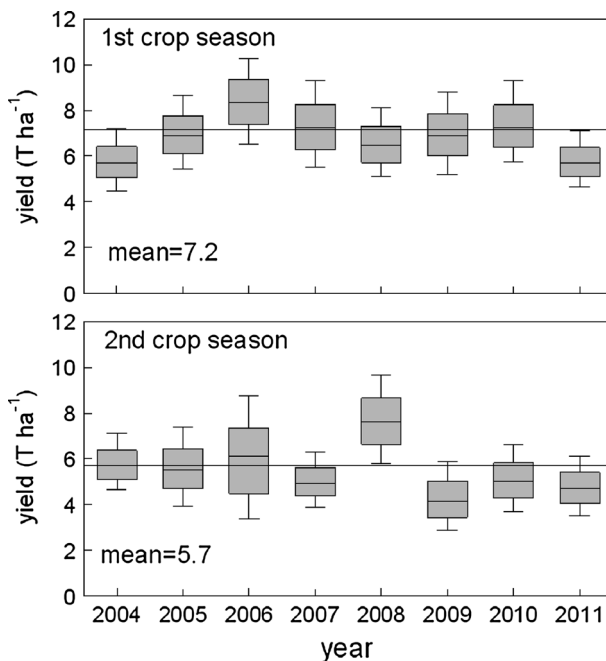

Fig. 2 Box-and-whisker plots of estimated yields for the 16-year/crop season at the study site. The caps at the end of each *box* indicate the extreme values (minimum and maximum), the *box* is defined by the lower and upper quartiles, while the *line* in the center of the box is the median. The length of the *box* indicates the interquartile range

Spatiotemporal distribution of the mean standardized yield in each studied regions

Of the total 161 929 effective pixels in the spatiotemporal yield map for the studied region, 36.0 % were classified as inconsistent (Fig. 3). To evaluate the temporal stability of the pixels having consistent yield in the derived spatiotemporal yield map, statistics of the yield class changes from the derived spatiotemporal yield map were compiled. In most cases, more than 40 % of the pixels retained the same yield classification in different year/ season (Table 2). The downgrade from class H and upgrade from class L were generally within one level. On average, only about 10 % pixels had yield class changes greater than
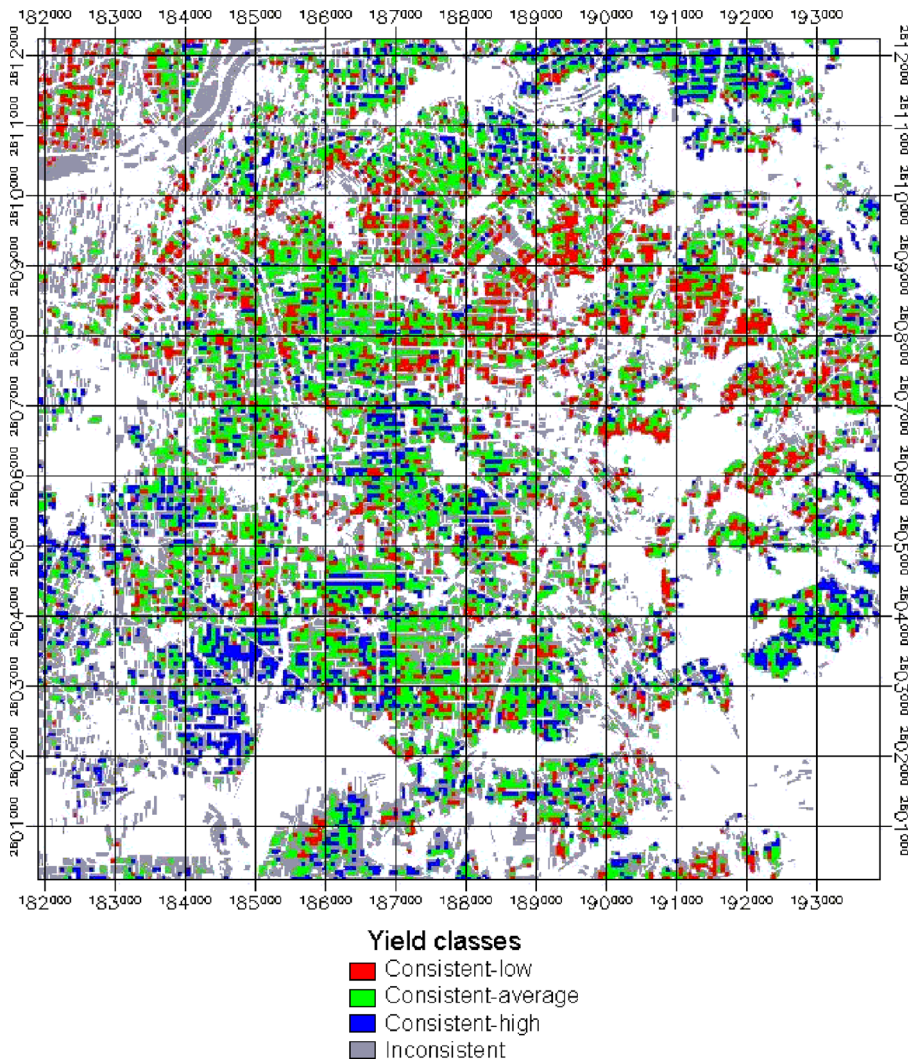

Fig. 3 Spatiotemporal yield trend map for the study site, with UTM coordinates. *Uncolored* pixels were either non-paddy rice pixels or severely contaminated by nearby non-paddy rice pixels (Color figure online)

Table 2 Percentage distribution of yield classes changes (YCC) from the derived spatiotemporal yield map for each yield class and individual year/crop season

YCC[a]	F_04	F_05	F_06	F_07	F_08	F_09	F_10	F_11	S_04	S_05	S_06	S_07	S_08	S_09	S_10	S_11	Mean
HH	46.8	44.3	46.2	49.3	38.9	42.8	39.7	40.3	47.0	41.8	42.1	41.0	48.9	48.4	44.6	51.1	44.6
HM	42.9	45.4	47.4	42.3	47.5	46.6	45.5	47.5	40.5	44.9	46.0	47.1	44.4	40.8	42.7	41.0	44.5
HL	10.3	10.3	6.4	8.4	13.6	10.6	14.8	12.2	12.5	13.3	11.9	11.9	6.7	10.8	12.7	7.9	10.9
MM	52.8	55.0	55.3	54.9	53.1	54.4	51.6	53.3	55.4	54.0	53.7	53.6	55.8	56.2	54.8	57.1	54.4
MH	25.4	22.9	20.5	22.7	25.4	21.6	24.6	23.7	23.5	23.5	23.3	24.5	21.0	19.6	21.4	20.1	22.8
ML	21.8	22.1	24.2	22.4	21.5	24.0	23.8	23.0	21.1	22.5	23.0	21.9	23.2	24.2	23.8	22.8	22.8
LL	39.1	44.5	48.7	45.6	43.6	44.6	36.4	42.7	43.8	40.7	42.2	42.1	48.2	46.0	42.5	48.2	43.7
LM	50.6	46.1	42.0	48.0	46.5	46.3	52.0	45.7	49.6	48.3	47.6	46.7	44.1	49.0	49.2	45.9	47.3
LH	10.3	9.4	9.3	6.4	9.9	9.1	11.6	11.6	6.6	11.0	10.2	11.2	7.7	5.0	8.3	5.9	9.0

F first crop season, *S* second crop season, *_XX* year

[a] The first letter represents the yield class in the derived spatiotemporal yield map; the second letter represents the yield class in each corresponding year/season

one level. Therefore, the derived spatiotemporal yield map might be considered to be temporally stable.

Yield performance resulted from complex and dynamic interactions among many biotic and abiotic factors. Soil physical and chemical properties could influence plant growth by affecting the supply of nutrients and water (buffering capacity and availability), the uptake of nutrients and water (plant roots development), and possible losses of applied fertilizers (leaching and volatilization). It might be very difficult to sort out the yield limiting factors for those yield inconsistent areas. However, for those yield consistent areas, differences in soil properties between those consistent-low and -high yielding areas should have played determining roles.

Soil characteristics for delineating low-yielding areas

Results of data mining analysis were listed in Table 3. By comparing the differences of soil characteristics between those effective H and L terminal nodes, we noted that the areas having the following characteristics tended to have more low-yield pixels than high-yield pixels.

$$pH\, code = 8 \tag{3a}$$

$$T1\, and\, T2\, code < 6\, and\, ID\, code = 1\, or\, 2 \tag{3b}$$

$$T1\, and\, T2\, code > 5\, and\, ID\, code = 3 \tag{3c}$$

$$T1\, or\, T2\, code > 5\, and\, ID\, code = 4 \tag{3d}$$

Error analysis indicated that among the total 23 895 consistent low-yield pixels in the studied region, 61.8 % correctly located in the areas identified by these characteristics. The omission and commission errors were 38.2 and 32.0 %, respectively (Fig. 4). Areal statistics indicated that soils under conditions (3a) and (3b) occupied more than 40 % of the paddy rice cultivated area in the studied region.

Only dominant soil types were shown on soil maps due to the limitation of map scales; therefore, each polygon on the map often contained small areas of dissimilar soils. Consequently, Although omission and commission errors were not small, these errors scattered over the entire region and no apparent wrongly classified large cluster of H or L pixels was observed. In addition to the limitations of map scale, the boundaries of the soil map polygons implied that there were abrupt changes in soil types within the landscape. In reality, however, soil varied continuously across the landscape and the lines on the map were merely approximations based on the surveyor's experience. Besides, the digitized information was a simplified summarization of observed internal properties and laboratory analysis data of typical pedons.

Suspected yield limiting mechanisms and possible alleviating field cultivation practices

In order to provide guidance for designing more efficient soil sampling and analysis plans and conducting field trials for cultivation practices improvement in subsequent years, knowledge of basic soil physical and chemical processes was applied to elucidate the functioning of suspected yield limiting soil factors. Since the size of a paddy field (~0.2 ha) is generally much smaller that a soil map polygon, only field-to-field variations may be captured. Therefore, the discussion below were mainly for delineating better soil management zones in the studied region and not for addressing the within-field variations.

Table 3 Node numbering, name, cases of yield classes, and soil characteristics of those effective terminal nodes screened for exploring the low yielding mechanisms

Node numbering/name	Pixels in yield class			pH	Soil texture				Internal drainage	Parent material	Soil appearance	Soil geomorphology	Soil alkalinity
	H	M	L		1	2	3	4					
07/L	7	24	214	8	5,6,8	4,5,9	5,8,9	5,8,9	1,2,3	2	2	1	1,2
14/L	294	1 235	1 555	7,8	3,4,5,6	4,5,6	2,4	2,4,5	1	2	2	1	2
67/L	0	49	80	4	8	5	8	8	4	2	2	1	1,3
70/L	0	78	194	5	4,5	5	5	4	1,2	2	2	1	2
76/L	5	25	64	4	5	5	8	8	2	2	2	1	1,3
12/M	1 711	5 571	4 319	4	3,4,5,6	5	4	4	2,3	2	2	1	1,3
19/M	74	466	389	7	8,9	9	9	9	4	2	1	3	1,2
23/M	84	445	355	7,8	5,6	4,5	4	4	2,3	2	2	1	1,2
55/M	514	2 138	1 532	5	5,8	8	8	8	3,4	2	2	1	1,3
56/M	19	65	55	6	3,4	3,5,8	8	8	2	2	2	1	1,2
84/M	134	591	435	5	5	5	5	4	4	2	2	1	2
SUM	2 842	10 687	9 192										
04/H	449	328	80	2,3	3,4,5,6	5,6,8	2,4	2,4	3,4	1,2	2,3	1,2	1,3
43/H	1 070	1 052	202	5,6	6,8	5,8	5	5	2,3,4	2	2	1	1,2,3
46/H	401	368	53	7	5	5	5,8	5,8	3,4	2	2	1	1,2,3
26/M	2361	2 404	573	3,5,6	5,6	4,5,8,9	5,8,9	5,8,9	1,4	1,2	1,3	2,3	1,3
32/M	512	1 045	165	5,6	3,5,8	5,8	8	8	1	2	2	1	1,2,3
39/M	74	88	0	7	5,6	5,8	5,8	5,8	4	2	2	1	2
57/M	291	554	91	6	5,8	5,8	8	8	3,4	2	2	1	1,2,3
58/M	1 553	1 746	424	7	5,8	8	5,8	5,8	1,2	2	2	1	1,2
62/M	831	919	162	5	4,5	5	5	4,5	1,3	2	2	1	1,3
71/M	81	106	0	5	5	5	5	5	2	2	2	1	1,2
77/M	289	448	113	4	3,5	5	8	8	4	2	2	1	1,3

Table 3 continued

Node numbering/name	Pixels in yield class			pH	Soil texture				Internal drainage	Parent material	Soil appearance	Soil geomorphology	Soil alkalinity
	H	M	L		1	2	3	4					
79/M	525	637	139	*7*	*5,8*	*8*	*5*	*5*	*3,4*	2	2	*1*	*1,2*
81/M	258	264	44	*7*	*5*	*8*	*8*	*8*	*4*	2	2	1	*1,2*
86/M	526	536	125	*7*	*5*	*8*	*5*	*5*	*4*	2	1	*3*	*1,2*
SUM	9 221	10 495	2 171										

Bold italicized attributes (cf. Table 2) are soil properties and threshold value selected by the classification and regression tree analysis (CART) as classification criteria by tracing back along the hierarchical order of the classified tree, whereas the non-italicized soil attributes are summarizations of soil properties for all cases classified into the corresponding effective terminal node
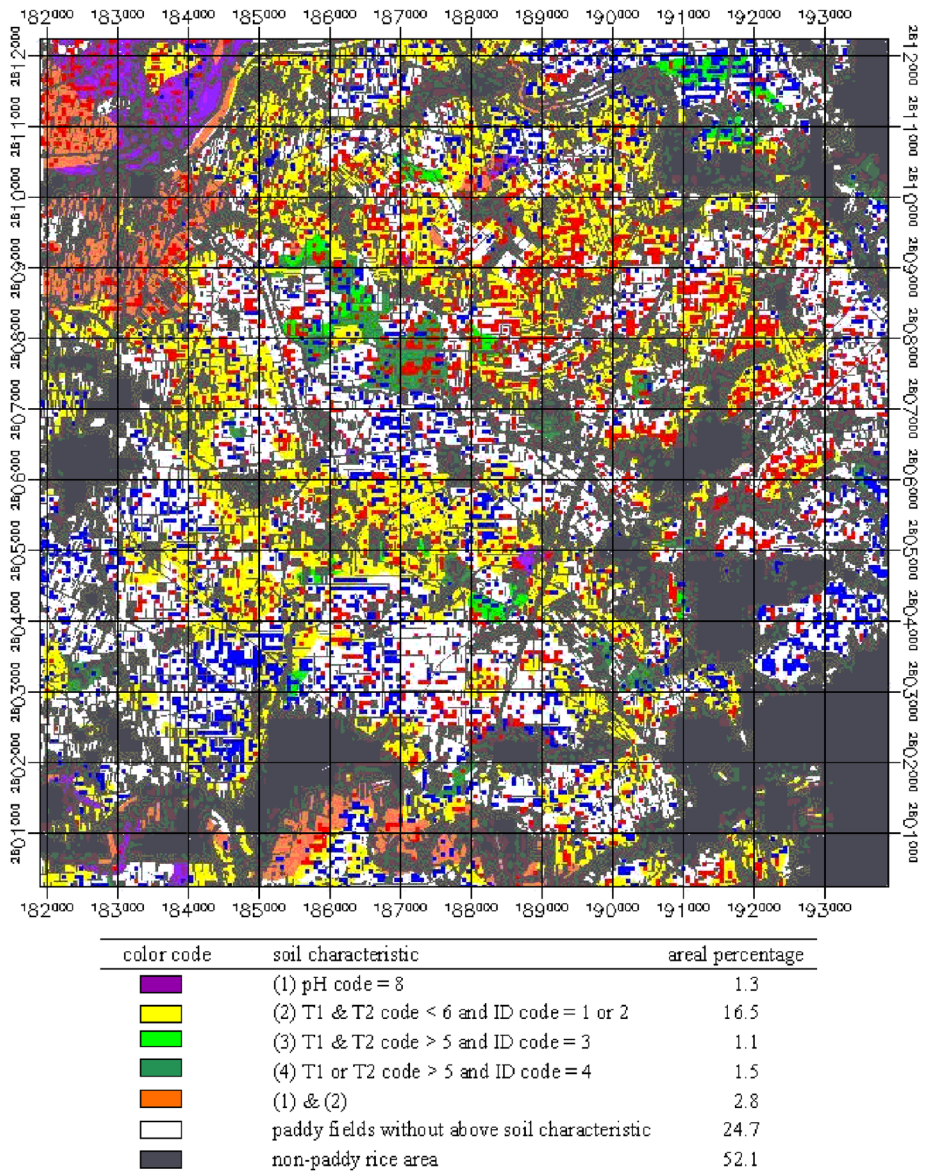
color code	soil characteristic	areal percentage
(purple)	(1) pH code = 8	1.3
(yellow)	(2) T1 & T2 code < 6 and ID code = 1 or 2	16.5
(green)	(3) T1 & T2 code > 5 and ID code = 3	1.1
(dark green)	(4) T1 or T2 code > 5 and ID code = 4	1.5
(orange)	(1) & (2)	2.8
(white)	paddy fields without above soil characteristic	24.7
(gray)	non-paddy rice area	52.1

Fig. 4 Color coded spatial distribution of soil characteristic used to identify low-yielding areas, overlapping with those consistent-high (*blue color*) and -low (*red color*) yield pixels in Fig. 3. Statistics of areal percentages for soils of different characteristic and UTM coordinates were also included (Color figure online)

Flooding drastically altered many soil properties of paddy fields, most notably the redox potential and pH, which had phenomenal effects on plant growth and development. Of all the essential nutrients, nitrogen was required in the largest quantity and was the most frequent limiting factor in rice productivity (De Datta 1981; Yoshida 1981). Therefore, the capacity of soil holding applied nitrogen fertilizers and the inference in preventing nitrogen

uptake were considered first in evaluating the possible low yielding mechanisms. The availability of other essential nutrients and production of phytotoxic substances in relation to changes of root environments under submerged conditions were also considered.

Soils under condition (3a), located mostly in the northwest corner of the studied region (Fig. 4), were calcareous soils which were highly basic, abundant with Ca and Mg carbonates, and coarser in texture (Table 3). For paddy soils, the soil pH generally approached a steady range between 6.5 and 7.0 after weeks of flooding, due to the dissolution of atmospheric CO_2. Unless the soil pH was lower than 4 or higher than 8, the surface soil pH had little effects to rice growth (Ponnamperuma 1972). However, ammonium-form fertilizers might dissociate directly or decompose by catalytic hydrolysis to produce NH_4^+ ions in water. With increasing hydroxyl-ion concentrations in the water, ionized NH_4^+ increasingly converted to non-ionized ammonia, which might escape from the water as gas. It had long been recognized that soil pH influenced ammonia volatilization and that the higher the soil pH, the higher the potential losses (Mikkelsen et al. 1978; Vlek and Craswell 1979).

Phosphorus deficiency occurred widely in calcareous soils because the applied phosphate fertilizers generally precipitated with calcium and transformed into less available form to plants, such as octocalcium phosphate and hydroxyapatite (Shen et al. 2011). Additionally, calcium phosphates precipitated on the coarser soil particles had smaller specific surface activity. Next to nitrogen and phosphorus deficiency, zinc (Zn) deficiency of paddy rice was also widespread on neutral to alkaline calcareous soils (Wissuwa et al. 2006). Following submergence, the increased reduction and solubilization of iron and manganese oxides had an antagonistic effect on the availability and uptake of Zn, and the increased bicarbonate concentration in soil solution inhibited Zn uptake contributed to the increased Zn deficiency (Alloway 2009). However, soil testing conducted by TARI indicated that phosphate fertilization was highly effective in this region, but Zn was not limited (data retrieved from http://taiwansoil.tari.gov.tw/Web.Net2008/index_1/main1-1.aspx).

Therefore, the loss of applied nitrogen fertilizers via ammonia volatilization and fixation of applied phosphorus fertilizers into less soluble calcium phosphates were suspected to be the major yield limiting mechanisms for areas delineated by condition (3a). Choudhury and Kennedy (2005) indicated that deep placement of nitrogen fertilizers; and use of modified forms of urea and slow-release nitrogen fertilizers were effective means to reduce ammonia volatilization losses. Besides, Organic matter amendments significantly increased the recovery of applied phosphate fertilizers by rice in calcareous soils (Saleque et al. 2004; Abolfazli et al. 2012). Kumari et al. (2008) also reported that the low molecular weight organic acids during rice straw decomposition helped solubilization of insoluble phosphates. Therefore, deep placement of basal nitrogen fertilizers and use of slow released nitrogen fertilizers for top dressing at later growth stages; incorporation of crop residues (e.g. rice straws and/or husks) at field preparation stage and amended phosphate fertilizers with organic composts were management practices may improve the yield performance in areas delineated by condition (3a).

Soils under condition (3b) had very coarse texture profiles and good internal drainage capacities (Table 3). In submerged soils, ammonium was the major form of nitrogen available for rice (De Datta 1981; Yoshida 1981). The low cation exchange capacity (CEC) of coarse particles had little buffering capacity for nutrients. The monovalent ammonium could also be easily displaced from the exchange complex to the soil solution by the large quantities of divalent ferrous and manganous ions produced by reduction reactions (Ponnamperuma 1972). The good internal drainage capacity and water head generated by the standing water on the flooded soil could further enhance the leaching loss of

ammonium from root zone. Potassium was another mobile monovalent cation vulnerable to leaching loss in coarse textured sandy soils (De Datta and Mikkelsen 1985). However, soil testing conducted by TARI indicated that potassium was still sufficient in this region (data retrieved from http://taiwansoil.tari.gov.tw/Web.Net2008/index_1/main1-1.aspx).

Therefore, the leaching loss of applied nitrogen fertilizers, resulted from low CEC and fast percolation through coarse textured soils of profiles, was suspected to be the major yield limiting mechanisms for areas delineated by condition (3b). Use of slow-release nitrogen fertilizers, nitrification inhibitors, and puddling of the rice fields were ways to reduce leaching losses (Choudhury and Kennedy 2005). Reducing the amount of basal application and splitting into more number of top dressings could also decrease the leaching loss. Using controlled irrigation, i.e. reduce the height of standing water, could reduce the leaching loss as well (Yang et al. 2013). However, the effectiveness of these possible yield improvement practices need to be refined and compared.

Soils under condition (3c) and (3d) had fine-textured upper layer soils and poor internal drainage capacities (Table 3). The standing water on soil surface was an effective barrier to gas exchange between soil and air. Any oxygen remaining after flooding was depleted by the respiration of roots and soil microorganisms within few hours (Sharma and De Datta 1986). Under the anaerobic conditions, soil microorganisms started using oxidized form of soil constituents and organic metabolites as electron acceptors in their respiration. To avoid suffocation of root tissues in submerged soils, rice plants have developed an aerenchy-matous air-passage system to transport air from shoot to root (Yoshida 1981). When oxygen transported internally within the roots, a certain amount of' oxygen might leak out laterally and diffuse outward into the rhizosphere zone. However, the small pore size and torturous paths in fine-textured soils made the diffusion of oxygen slower than in coarse-textured soils, which could create a more reduced environment in the bulk soil. The poor internal drainage would worsen the situation by reducing the replenishment of dissolved oxygen through percolation (Sharma and De Datta 1986) and would make the excessive reductive environment last longer after midterm drainage.

Armstrong and Drew (2002) reviewed the effects of oxygen deficiency to the growth and functions of roots in flooded soils. They indicated that organic acids (mainly acetic and butyric acids with small amounts of formic, propionic, and lactic acids) and the organic reduction products produced under anaerobic condition could harm the roots of rice plants. Hydrogen sulfide, generated from sulfate reduction and anaerobic decomposition of organic matter, was also phototoxic because it was a strong inhibitor of aerobic respiration after entering the roots. Iron (Fe) toxicity, commonly occurred on acid sandy, acid lato-solic, and acid sulfate soils, should have no contribution to the low yield in the studied region because the soil characteristics did not completely fit the important criteria for occurrence of Fe toxicity in rice (Ponnamperuma 1976).

Therefore, the excessive reductive root environment, which could produce various phototoxic substances to damage the roots and retard the uptake of plant nutrients and rice growth, was suspected to be the major yield limiting mechanisms for areas delineated by conditions (3c) and (3d). Incorporating rice straw or other organic amendments into the soil during the field preparation stage, a practice commonly recommended by local extension specialists, should be avoided to reduce the amount of phytotoxic substances that might be produced under anaerobic condition (Armstrong and Drew 2002). Although drainage (mid-season and multiple drainage) practice could prevent the development of the excessive reductive conditions in the paddy soils, it could also increase the nitrous oxide emissions, as well as reduce the yields (Towprayoon et al. 2005). Water management based on soil redox potential (Eh), which had been proven to be an effective way not only to change the

range of soil Eh to a more positive value but also to increase rice grain yield (Minamikawa and Sakai 2006), may be a practice worth trying in these areas.

Conclusions

This research is important in linking remote sensed yield maps, traditional soil survey data, and data mining techniques to clarify the yield limiting mechanisms for planning proper site-specific crop managements. Our analysis indicated that soil texture profile, internal drainage, and soil acidity were the most important soil properties determining yield variability and explained about 62 % of the yield variability. High soil pH, severe leaching loss of applied nitrogen fertilizers, and excessive reductive root environment were likely to be the major soil related low-yielding mechanisms spread within the studied region.

Although within-field variation were not identified in this study due to limitations in field size, spatial resolution of satellite images utilized and soil survey data available, soil management zones better than those based only on surface soil texture and pH were delineated in the studied region. Besides, the spatiotemporal yield trend map should have values to any new intensive sampling project, if to be conducted, by reducing the number of samples to be collected and the items to be analyzed. The proposed alleviating field cultivation practices could serve as guidance for further field trials in providing more field evidences for those suspected yield limiting mechanisms. Therefore, this study is an important step toward implementing precision agriculture in the studied region.

References

Abolfazli, F., Forghani, A., & Norouzi, M. (2012). Effects of phosphorus and organic fertilizers on phosphorus fractions in submerged soil. *Journal of Soil Science and Plant Nutrition, 12*, 349–362.

Alloway, B. J. (2009). Soil factors associated with zinc deficiency in crops and humans. *Environmental Geochemistry and Health, 31*, 537–548.

Armstrong, W., & Drew, M. C. (2002). Root growth and metabolism under oxygen deficiency. In Y. Waisel, A. Eshel, & U. Kafkafi (Eds.), *PLANT Roots: THE HIDDEN HALF* (3rd ed., pp. 729–761). New York: Marcel Dekker.

Breiman, L, Friedman, J., Olshen. R., & Stone, C. (1984). *Classification and regression trees* (54 pp.). New York, NY: Chapman and Hall (Wadsworth, Inc.).

Brevik, E. C., Fenton, T. E., & Jaynes, D. B. (2003). Evaluation of the accuracy of a central Iowa soil survey and implications for precision soil management. *Precision Agriculture, 4*, 331–342.

Chiou, C. R., Liang, Y. C., Lai, Y. J., & Huang, M. Y. (2004). A study of delineation and application of the climatic zones in Taiwan. *Journal of Taiwan Geographic Information Science, 1*, 41–62.

Choudhury, A. T. M. A., & Kennedy, I. R. (2005). Nitrogen fertilizer losses from rice soils and control of environmental pollution problems. *Communications in Soil Science and Plant Analysis, 36*, 1625–1639.

De Datta, S. K. (1981). *Principles and practices of rice production*. New York: Wiley.

De Datta, S. K., & Mikkelsen, D. S. (1985). Potassium nutrition of rice. In R. D. Munson (Ed.), *Potassium in agriculture* (pp. 665–699). Madison, Wisconsin USA: ASA, CSSA, SSSA.

Dobermann, A., Ping, J. L., Adamchuk, V. I., Simbahan, G. C., & Ferguson, R. B. (2003). Classification of crop yield variability in irrigated production fields. *Agronomy Journal, 95*, 1105–1120.

Franzen, D. W., Hopkins, D. H., Sweeney, M. D., Ulmer, M. K., & Halvorson, A. D. (2002). Evaluation of soil survey scale for zone development of site-specific nitrogen management. *Agronomy Journal, 94*, 381–389.

Gallego, J., Carfagna, E., & Baruth, B. (2010). *Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. Agricultural survey methods* (pp. 193–211). West Sussex, UK: Wiley.

Guo, H. Y., Liu, T. S., Chu, C. L., & Chiang, C. F. (2009). Development of soil information system and its application in Taiwan. In: *Proceedings of Monsoon Asia Agro-Environmental Research Consortium: A new approach to soil information systems for natural resources management in Asian countries*. October 13–17, 2008, Tsukuba, Japan.

Hochman, Z., Gobbett, D., Holzworth, D., McClelland, T., Van Rees, H., Marinoni, O., et al. (2012). Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia. *Field Crops Research, 136*, 85–96.

Hsieh, S. C., & Liu, D. J. (Eds.). (1979). The causes of low yield of the second crop rice in Taiwan and the measures for improvement. In: *Proceedings of a symposium held at Taiwan Agricultural Research Institute*, June 7–8, 1978. Taiwan, ROC: National Science Council. (In Chinese with English abstract).

Kropff, M. J., Cassman, K. G., van Laar, H. H., & Peng, S. (1993). Nitrogen and yield potential of irrigated rice. *Plant and Soil, 155*(156), 391–394.

Kumari, A., Kapoor, K. K., Kundu, B. S., & Mehta, R. K. (2008). Identification of organic acids produced during rice straw decomposition and their role in rock phosphate solubilization. *Plant, Soil and Environment, 54*, 72–77.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research, 143*, 56–64.

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). Crop yield gaps: Their importance, magnitudes, and causes. *Annual Review of Environment and Resources, 34*, 179–204.

Lobell, D. B., Ortiz-Monasterio, J. I., & Falcon, W. P. (2007). Yield uncertainty at the field scale evaluated with multi-year satellite data. *Agricultural Systems, 92*, 76–90.

Mikkelsen, D. S., De Datta, S. K., & Obcemea, W. N. (1978). Ammonia volatilization losses from flooded rice soils. *Soil Science Society of America Journal, 42*, 725–730.

Minamikawa, K., & Sakai, N. (2006). The practical use of water management based on soil redox potential for decreasing methane emission from a paddy field in Japan. *Agriculture, Ecosystems & Environment, 116*, 181–188.

Moulin, S., Bondeau, A., & Delecolle, R. (1998). Combining agricultural crop models and satellite observations: From field to regional scales. *International Journal of Remote Sensing, 19*, 1021–1036.

Ponnamperuma, F. N. (1972). The chemistry of submerged soils. *Advances in Agronomy, 24*, 29–96.

Ponnamperuma, F. N. (1976). Specific soil chemical characteristics for rice production in Asia. In *International Rice Research Institute Res. Paper Ser. 2*. Manila, Philippines.

Saleque, M. A., Naher, U. A., Islam, A., Pathan, A. B. M. B. U., Hossain, A. T. M. S., & Meisner, C. A. (2004). Inorganic and organic phosphorus fertilizer effects on the phosphorus fractionation in wetland rice soils. *Soil Science Society of America Journal, 68*, 1635–1644.

Sharma, P. K., & De Datta, S. K. (1986). Physical properties and processes of puddled rice soils. *Advances in Soil Science, 5*, 139–178.

Sheh, C. S. & Wang, M. K. (Eds.). (1989). *Soils of Taiwan—explanatory text of the 1988 soil map of Taiwan*. Soil Survey and Testing Center, National Chung Hising University. (In Chinese with English abstract).

Shen, J. B., Yuan, L. X., Zhang, J. L., Li, H. G., Bai, Z. H., Chen, X. P., et al. (2011). Phosphorus dynamics: From soil to plant. *Plant Physiology, 156*, 997–1005.

Soil Survey Division Staff. (1993). *Soil survey manual*. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Sun, T. L., & Yang, T. C. (1989). The occurrence of poor drainage conditions in alluvial soils of Taiwan. *Journal of the Chinese Agricultural Chemical Society, 27*, 38–45. (In Chinese with English abstract).

Tittonell, P., Shepherd, K. D., Vanlauwe, B., & Giller, K. E. (2008). Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—An application of classification and regression tree analysis. *Agriculture, Ecosystems & Environment, 123*, 137–150.

Towprayoon, S., Smakgahn, K., & Poonkaew, S. (2005). Mitigation of methane and nitrous oxide emissions from drained irrigated rice fields. *Chemosphere, 59*, 1547–1556.

Van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance—a review. *Field Crops Research, 143*, 4–17.

Van Ittersum, M. K., & Rabbinge, R. (1997). Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Research, 52*, 197–208.

Vlek, P. L. G., & Craswell, E. T. (1979). Effect of nitrogen source and management on ammonia volatilization losses from flooded rice-soil systems. *Soil Science Society of America Journal, 43*, 352–358.

Wang, Y. P., Chang, K. W., Chen, R. K., Lo, J. C., & Shen, Y. (2010). Large area rice yield forecasting using satellite imageries. *International Journal of Applied Earth Observation and Geoinformation, 12*, 27–35.

Wang, Y. P., Chen, S. H., Chang, K. W., & Shen, Y. (2012). Identifying and characterizing yield limiting factors in paddy rice using remote sensing yield maps. *Precision Agriculture, 13*, 553–567.

Wissuwa, M., Ismail, A. M., & Yanagihara, S. (2006). Effects of zinc deficiency on rice growth and genetic factors contributing to tolerance. *Plant Physiology, 142*, 731–741.

Wollenhaupt, N. C., Mulla, D. J., & Gotway Crawford, C. A. (1997). Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In F. J. Pierce & E. J. Sadler (Eds.), *The state of site specific management for agriculture* (pp. 19–53). Madison, WI: ASA, CSSA, and SSSAJ.

Yang, S. H., Peng, S. H., Xu, J. Z., Hou, H. J., & Gao, X. L. (2013). Nitrogen loss from paddy field with different water and nitrogen managements in Taihu lake region of China. *Communications in Soil Science and Plant Analysis*. doi:10.1080/00103624.2013.803564.

Yoshida, S. (1981). *Fundamentals of rice crop science*. Los Baños, Philippines: IRRI.

Zwart, S. J., & Leclert, L. M. C. (2010). A remote sensing-based irrigation performance assessment: A case study of the Office du Niger in Mali. *Irrigation Science, 28*, 371–385.