Agriculture, Ecosystems and Environment 132 (2009) 98–105

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions

Haifeng Zheng[a], Liding Chen[a,*], Xiaozeng Han[b], Xinfeng Zhao[a], Yan Ma[a]

[a] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, PR China
[b] National Field Research Station of Agroecosystem in Hailun, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, PR China

ARTICLE INFO

Article history:
Received 23 September 2008
Received in revised form 13 March 2009
Accepted 16 March 2009
Available online 15 April 2009

Keywords:
Drought
Soybean
Yield gap
Classification and regression tree (CART)
Adaptation
Northeast China

ABSTRACT

Drought is the most critical environmental factor limiting the productivity of agricultural crops worldwide. Increased frequency and severity of drought are expected to accompany climate change and will negatively impact global food security. Wide yield variability from field to field, and consequently reduced average yield on a regional scale often occur under drought conditions. The reasons for the yield variability are still poorly understood. In this study, we explored sources of soybean yield variability among fields in a rural village of Northeast China associated with a severe drought growing season in 2007. Soil parameter measurements were made on fields following three transects with different distances from homestead. Management data were assembled from household interviews. The relative importance of soil parameters and management practices resulting in yield variability among fields was analyzed with general linear model (GLM) and classification and regression trees (CARTs) models. Our analysis showed that variability in management options, as opposed to variability in soil parameters, caused the majority of yield variability from field to field. The amount of P applied was the most important variable determining yield variability and explained roughly 61% of the variability. Whether or not manure was added into fields was of secondary importance. The classification tree analysis indicated that yield differences among transects was attributed to the content of K nutrient. This might result from variations of long-term management options with distance from homestead. CART models are robust technique for predicting yield variability responses to variations of soil properties and management practices due to its low prediction error. Our study highlights the pressing need to adjust management strategies for narrowing yield variability and increasing crop production in drought years. We recommend that in addition to testing soil, government programs in China should also pay close attention to management practices of farmers.

1. Introduction

Increased population and economic growth are widely acknowledged to drive sustained growth in food demand over the next few decades. Meeting these food needs will be a major challenge because all of the best farm land is already in use and trends of yield stagnation are evident in much of the world (Cassman et al., 2003). In China, yield growth rates of the three major crops, soybean, corn and rice, have slowed since 1995 (FAOSTAT, 2007). What is the likelihood that agriculture can continue to feed the world with increasing population? Many researchers call for a second "Green Revolution" (Mann, 1997; Conway and Toenniessen, 1999; Schiermeier, 2008). This new green revolution emphasizes a prominent role on improving crops management techniques in addition to plant biotechnology. Considerable yield variability often exists among adjacent fields managed by different farmers. Improvement of yields in low yielding fields through adjustment of management practices would make an important contribution in raising world food production to keep up with the food demands.

A few studies have investigated the relative importance of soil parameters and management factors that determine crop yield variability among fields (Lobell et al., 2005; Tittonell et al., 2008). However, effects of extreme climatic events such as severe drought still remain to be explored. Lobell et al. (2005) concluded that causes of field to field yield variability varied among years with different environmental conditions. Drought, caused by decreased precipitation and increased temperatures, has become common

* Corresponding author at: State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, P.O. Box 2871, Beijing 100085, PR China. Tel.: +86 10 62943840; fax: +86 10 62923557.
E-mail addresses: liding@rcees.ac.cn, zhfdd@msn.com (L. Chen).

0167-8809/$ – see front matter
doi:10.1016/j.agee.2009.03.004

over wide areas including Northeast China (Dai et al., 2004; Zou et al., 2005; Schiermeier, 2008) and threatens global food security (Ciais et al., 2005; IPCC, 2007). This emphasizes the urgent need to examine factors contributing to field to field yield variability under drought conditions. Numerous studies have demonstrated that decreased N_2 fixation rate associated with drought resulted in soybean yield reduction (Purcell et al., 2004; Ray et al., 2006; Sinclair et al., 2007) because most of N for protein production arises from N_2 fixation. The largest soybean production areas in China are in the Northeast China where drought often causes a low soybean grain yield (Liu et al., 2008), especially drought stress occurring during flowering and pod filling stages (Xie et al., 1994; Zhao et al., 2006). It is imperative to explore some key factors which can be adjusted to alleviate the negative effects of drought conditions on soybean yield.

To achieve the goals of both higher crop yield and environmental protection, China implemented a project called "Soil Testing for Formulated Fertilization" in 2005. To date, the project has covered roughly 40% of total national cropland area. The project is focused on testing soil parameters with relatively little attention given to management practices of local farmers. Some researchers consider management practices to be a major cause of spatial yield variation (Lobell et al., 2005; Lobell and Ortiz-Monasterio, 2006; Tittonell et al., 2008).

In our study, all soil and management variables believed to have a potential impact on soybean yield were measured during the 2007 growing season when, the study area experienced a particularly severe rainfall deficit associated with higher temperature (Fig. 1). Using these measured variables, our objectives were to: (1) compare the relative importance of soil properties and management practices in determining soybean yield variability, (2) identify the variables which were most important in contributing to soybean yield variability among fields under drought conditions, and (3) investigate whether soybean yield changed with distances of fields from the homestead.

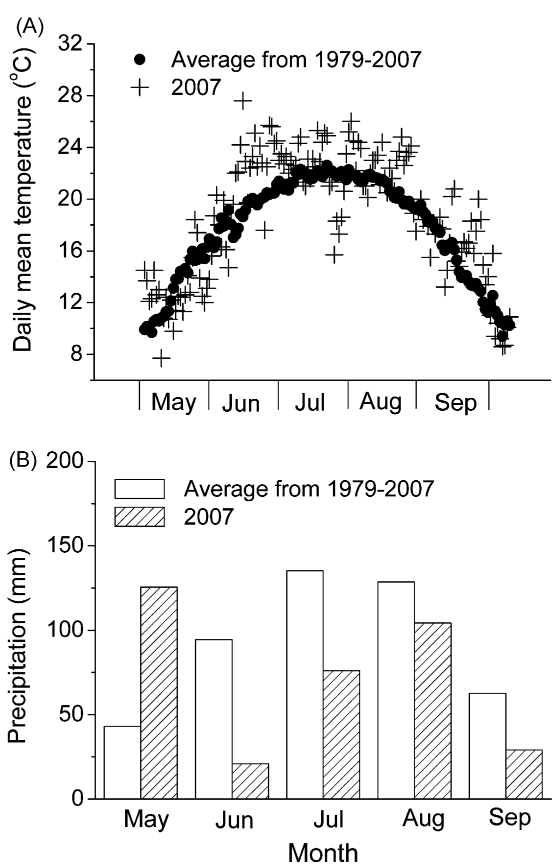

Fig. 1. Monthly mean temperature (A) and precipitation (B) for growing season in 2007, and historical averages from 1979 to 2007.

2. Materials and methods

2.1. Study area

The current study was performed in Hailun County, Heilongjiang Province (47°26′N, 126°38′E), situated in the central part of black soil region in Northeast China (Fig. 2). Fields subject to study were managed in Youjia village, which was about ten kilometers from Hailun County administrative offices. The village was located on a level landscape with an elevation of 240 m. It is in temperate continental monsoon zone with the annual mean temperature of 2.41 °C and annual precipitation of 464 mm since 1979. Rainfall is mainly distributed in May–September. Soil is typically medium layer black soil derived from loam loess with silty clay loam or silty clay surface textures. Growth of crops is highly seasonal and confined to May–September. Soybean (*Glycine max*) and maize (*Zea mays*) are the main staple food crops grown in the village and are normally planted in early May.

Villages in the area are typically a collection of closely spaced homesteads on either side of a road. Farmers live in the villages and walk to surrounding lands to tend to crops. Facilities for housing domestic livestock including swine, chickens, waterfowl, and some cattle are located in the villages.

The study was conducted by sampling soils and crops on existing farms with existing management which both limited both experimental design options, and control over treatments. However, independent management decisions by individual farmers, and existing heterogeneity in fertility of fragmented fields largely resulting from differences in previous practices, resulted in a random distribution of both categorical and continuous factors along the transects and provided us a good opportunity to extract factors which determine yield variability by multivariate statistical methods.

Landholding fragmentation is an important feature of the study area. Each household receives a few fields with different classes of soil fertility. Field sizes are small and varied from 0.2 to 1.0 ha. The length of the fields is large ranging from 200 to 500 m perpendicular to the road, but the width is only about 10–25 m. The term "field" in this paper refers to the one farmer's land, which is managed independently, but is not separated from adjacent fields by fences or other physical barriers.

2.2. Experimental design and sample analysis

Three transects with width of 9 m were established across adjacent fields parallel to the direction of the village extension and perpendicular to the direction of tillage (Fig. 2). The distances from the three transects to the homesteads were 0.3 km for near transect (NT), 0.5 km for middle transect (MT) and 1 km for far transect (FT). A total of 43 fields to be planted soybean and traversed by the three transects were selected, with 11, 19 and 13 fields located in NT, MT and FT, respectively.

At sowing, top-soil (0–20 cm) samples were taken via a manual coring tube from the selected fields within the range of each transect. Nine sub-samples per field in each transect were combined to form a composite sample. Soil chemical properties were analyzed using the standard soil test methods (Lu, 1999). Soil pH was determined in water using a 1:2.5 soil/solution ratio. Soil electrical conductivity (EC) was measured using Mettler Toledo Delta-326 conductivity meter (Mettler Toledo, Shanghai, China). Soil organic carbon (SOC) was measured by the $K_2Cr_2O_7$ titration method after digestion. Soil total nitrogen (TN) was determined by

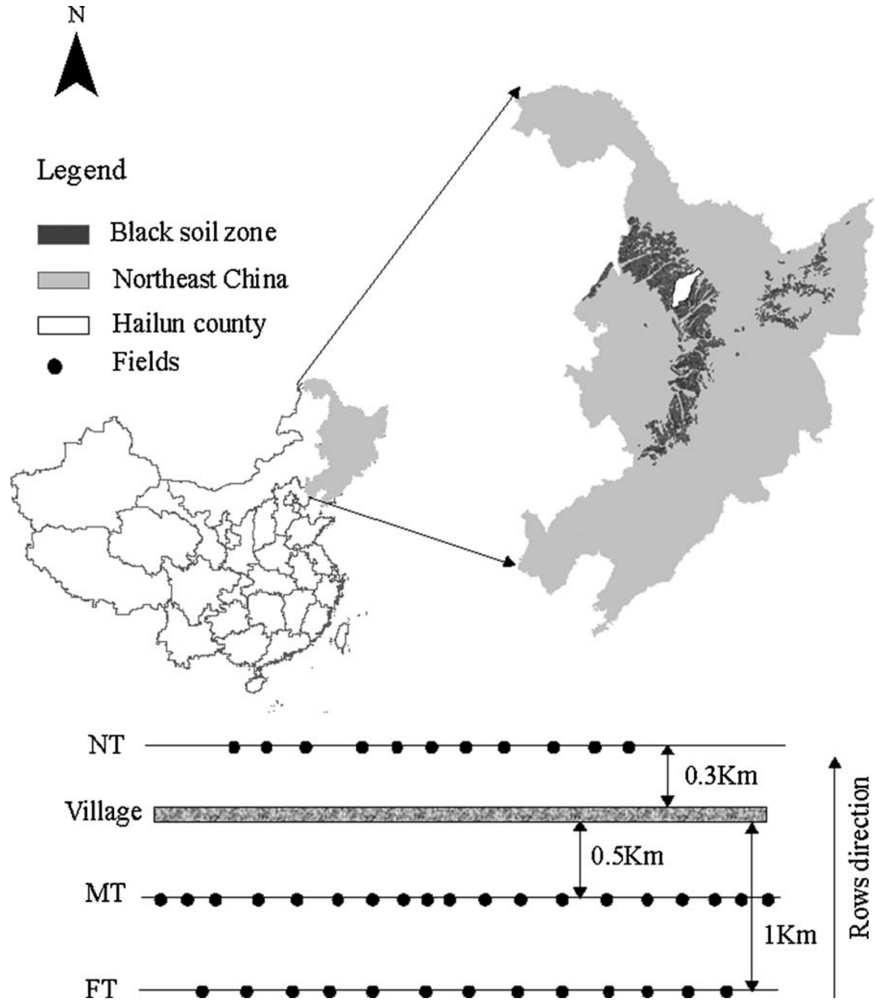

Fig. 2. Map showing location of study area and the sampled fields. NT, MT and FT represent near transect, middle transect and far transect, respectively. Row direction of the fields is also presented.

the semi-micro Kjeldahl method. Available nitrogen (AN) was determined by the Cornfield method (alkaline hydrolysable nitrogen). Total phosphorus (TP) was determined colorimetrically after wet digestion with H_2SO_4 plus $HClO_4$. Total potassium (TK) was determined by atomic absorption spectrometer. Available phosphorus (AP) was extracted with 0.5 mol L^{-1} $NaHCO_3$ solution at pH 8.5. Available potassium (AK) was extracted with 1.0 mol L^{-1} CH_3COONH_4 solution. After filtering, the solution was measured by ICP-AES.

In the late September, nine soybean sub-samples of each field were harvested by hand and each sub-sample was obtained from 1 m segments from the two rows adjacent to where the soil sub-sample was collected. The plants were cut 3 cm above the ground, separated into straw and grain, dried at 70 °C to constant weight. Grain yield was adjusted to a water content of 12.5% wet weight. Row space of all the fields was measured by tapeline for subsequent calculation of soybean yields.

2.3. Household interview

A questionnaire was developed to obtain detailed information on field management practices. Heads of the households for the

Table 1
Agronomic variables used in the CART analysis.

Variables	Unit	Description	Mean	Max	Min	S.D.
FYM	None	Farmyard manure addition (0 = non-manured, 1 = manured)	0.4	1	0	0.5
INSECT	None	Insecticide applied (0 = no, 1 = yes)	0.6	1	0	0.5
HERBICID	None	Herbicide applied (0 = no, 1 = yes)	0.7	1	0	0.5
DTPL	Days	Planting date (days after 1 May)	7.9	11	5	1.8
TT	None	Tillage traction (0 = horse power, 1 = tractor power)	0.8	1	0	0.4
Zn	None	Fertilizer Zn applied (0 = no, 1 = yes)	0.5	1	0	0.5
CROP	None	Crop planted last year (0 = soybean, 1 = corn)	0.7	1	0	0.5
VAR	None	Variety (0 = Beifeng-9, 1 = 9395, 2 = 46-1)	0.9	2	0	0.8
N	kg ha^{-1}	Fertilizer N applied	30.9	42	19.5	5.8
P	kg ha^{-1}	Fertilizer P applied	21.5	31.8	16.2	3.8
K	kg ha^{-1}	Fertilizer K applied	29.9	46.3	20.5	6.4
TILL	None	Tillage practice, autumn ploughing (coded 0) or spring ploughing followed by spring secondary tillage (coded 1)	0.6	1	0	0.5

selected fields were interviewed from late April to late May, questions from the questionnaire were verbally asked, and responses were recorded by the interviewer. The questionnaires were summarized; a subset of the information is given in Table 1. The measured variables included crop rotations; methods of soil preparation before sowing; type of tillage; variety of soybean; number of shovel plough and hand-hoeing operations; planting date; type and amount of insecticide and herbicide applications; whether or not farmyard manure (FYM) was added to fields; type and total amount of fertilizer (including N, P, K and Zn) applied. Amounts of N, P and K applied in each field were calculated from their respective percentages as written on the fertilizer bags, and the bulk application rate. All management practices considered to have potential impact on yield were measured to provide a comprehensive data set from which the most important variables could be identified and extracted. In addition, we also collected information on socioeconomic variables, such as age and education level of the selected heads of the household, area of cropland, household laborers and mean household income.

2.4. Statistical methods

All statistical analyses were performed by SYSTAT 12 (Systat Software, San Jose, CA), and included summary statistics, step-wise linear regression analysis, and classification and regression tree (CART) analysis.

2.4.1. Exploratory analysis

Summary statistics was run on all the fields to investigate soil variance. All measured soil parameters were used as independent variables in a step-wise regression procedure with against soybean yield as the dependent variable. The statistical tests were considered significant at 0.05 probability level.

2.4.2. Classification and regression tree (CART) analysis

We used classification and regression trees (Breiman et al., 1984; De'ath and Fabricius, 2000) to uncover relationships and interactions between soybean yields and a suite of soil and agronomic variables. CART is a nonparametric modeling approach that can explain the responses of a dependent from a set of independent continuous variables or categorical variables. CART models recursively partition the data to find increasingly homogeneous subsets based on independent variable splitting criteria using variance minimizing algorithms. The dependent data are partitioned into a series of descending left and right child nodes derived from parent nodes (Breiman et al., 1984). Once the partitioning has ceased, the child nodes are designated as terminal nodes. In this study, regression trees were used to predict yield responses from soil and agronomic variables from all fields, and classification trees were used to identify the most important soil and management variables affecting yield from the three transects. We chose the least squares method for regression trees and the Gini index for classification trees as the loss function with a minimum proportional reduction of error (PRE) at any split of 0.05 and minimum of five objects allowed in any node. The PRE represents the proportion of the error in soybean yield that can be explained by the model as a whole, with each split contributing a portion of the PRE. Due to limited sample size, bootstrap resampling (100 replicates) was performed to test the statistical strength of CART.

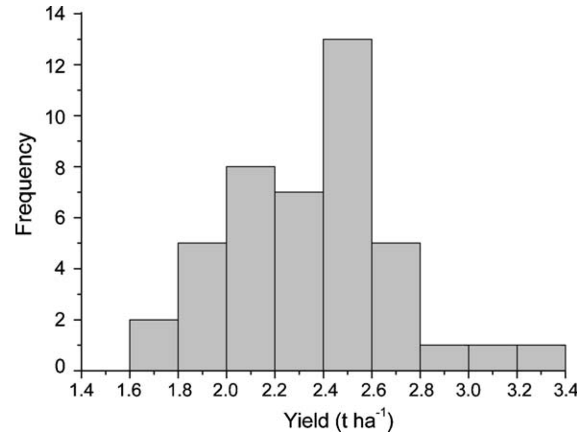

Fig. 3. Variability of soybean yield in all selected fields.

3. Results

3.1. Variability of soil properties and soybean yield

3.1.1. Among all fields

Table 2 summarizes the statistics of the soil variables studied in all the fields from these three transects. EC had the highest variability (47%) and TK content had the lowest variability (3%). The variability of AN (CV = 19%), AP (CV = 37%) and AK (CV = 26%) was high. Other soil variables were relatively stable. There was considerable soybean yield variability among the 43 fields. The highest yield was 3.25 t ha^{-1} and the lowest one was 1.61 t ha^{-1}. Most of the fields' yield ranged from 1.8 to 2.8 t ha^{-1} (Fig. 3).

3.1.2. Among transects

Most of the soil variables did not differ among the three transects (Table 3). Values of TK and EC in FT were lower than those in the other two transects nearest to the village. MT had lower AK compared with the other transects. Wide variation in soybean

Table 2
Summary statistics of soil properties in all fields.

	pH	EC (Ds m^{-1})	SOC (%)	TN (g kg^{-1})	AN (mg kg^{-1})	TP (g kg^{-1})	AP (mg kg^{-1})	TK (g kg^{-1})	AK (g kg^{-1})
Minimum	5.09	7.40	2.81	2.45	156.97	0.56	7.18	15.82	0.30
Maximum	6.39	42.40	3.55	2.98	403.92	1.06	31.21	18.26	0.81
Mean	5.48	16.47	3.25	2.67	256.22	0.80	16.10	17.25	0.43
S.D.	0.29	7.71	0.18	0.12	48.57	0.11	6.01	0.53	0.11
CV	0.05	0.47	0.05	0.05	0.19	0.13	0.37	0.03	0.26

Table 3
Differences of soil variables (mean ± S.E.) among the three transects.

	pH	EC (Ds m^{-1})	SOC (%)	TN (g kg^{-1})	AN (mg kg^{-1})	TP (g kg^{-1})	AP (mg kg^{-1})	TK (g kg^{-1})	AK (g kg^{-1})
NT	5.54 ± 0.12	16.16 ± 2.03	3.27 ± 0.07	2.70 ± 0.05	235.08 ± 10.30	0.80 ± 0.04	13.97 ± 1.34	17.34 ± 0.18	0.49 ± 0.04
MT	5.52 ± 0.07	18.16 ± 2.13	3.23 ± 0.04	2.66 ± 0.03	263.57 ± 13.54	0.79 ± 0.02	14.90 ± 1.34	17.49 ± 0.10	0.39 ± 0.01
FT	5.37 ± 0.04	14.28 ± 1.54	3.25 ± 0.04	2.65 ± 0.03	263.37 ± 10.78	0.83 ± 0.03	19.65 ± 1.73	16.83 ± 0.11	0.45 ± 0.04

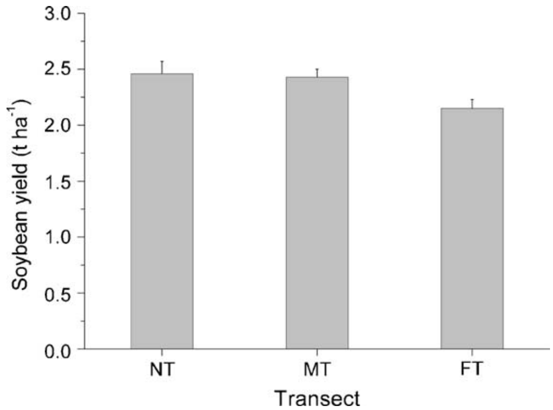

Fig. 4. Variation of soybean yield among the three transects.

yield was detected among the transects. There is "a tendency" that average yield in the FT was 15% lower than NT and MT, and there is no difference of soybean yields between NT and MT (Fig. 4).

3.2. Effects of soil properties on yield variability

Step-wise regression between soil variables and soybean yields using the probability F for entry (0.05) and removal (0.1) selected only SOC and EC. The regression equation was: $Y = 0.73\text{SOC} + 0.01\text{EC}$ ($R^2 = 0.242$, $P = 0.004$). The result indicated that only 24.2% of the variability in yield could be explained by the measured soil parameters. Judging from the regression coefficients, it was apparent that high soybean yield was related to high amounts of SOC and EC.

The regression tree model for soybean yield as a function of soil variables is shown in Fig. 5. The results indicated that SOC was the most important variable determining yield variability. There were 20 fields with SOC more than 3.27% and average yield of these fields was 2.52 t ha^{-1}. EC was the second most important variable and significantly impacted only those fields with high SOC. Yield of fields with both SOC greater than 3.27% and EC more than 16.6 Ds m^{-1} averaged 2.66 t ha^{-1}. The overall model explained roughly 32.7% of yield variability using the two variables. Compared with the general linear model (GLM), CART model explained a greater amount of yield variability with the same number of independent variables. Our results were consistent with the previous findings that CART showed significantly higher prediction accuracy than general linear model (GLM) (Lobell et al., 2005; Park et al., 2005).

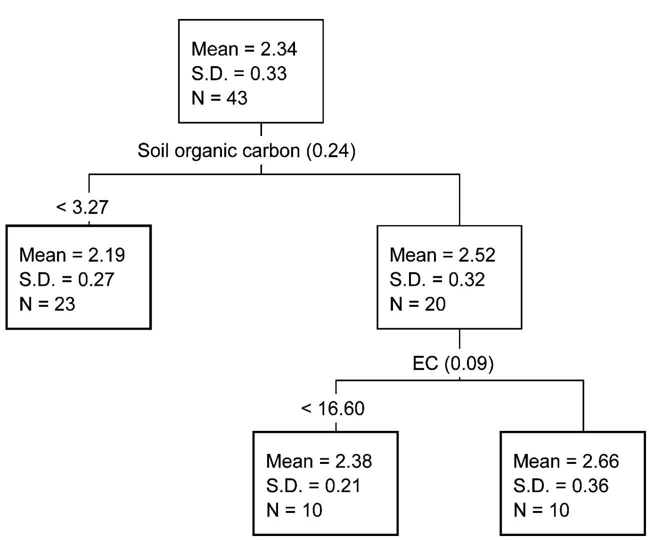

Fig. 5. Regression tree predicting soybean yields from soil variables (PRE = 0.33). Each node (square) is labeled with average yield (mean), standard deviation (S.D.) and the number (n) of fields in that group. The model is read from top down until terminal nodes appear. Partial PRE values are presented in parentheses at each root node to split.

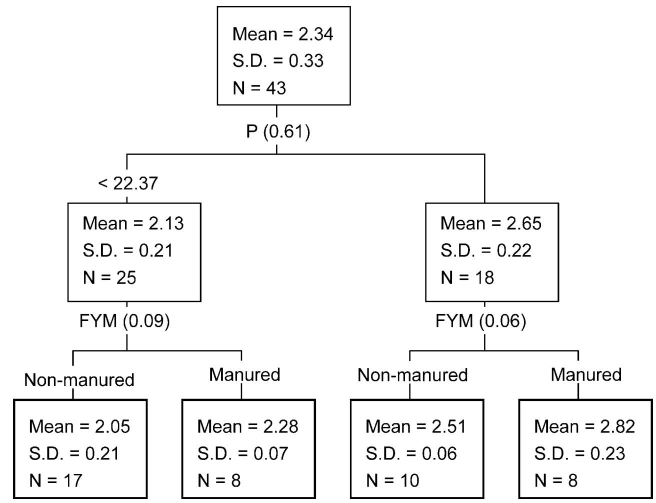

Fig. 6. Regression tree predicting soybean yield from agronomic management variables (PRE = 0.76).

3.3. Effects of agronomic practices on yield variability

To determine contribution of management practices and identify important agronomic variables, we fitted a regression tree for soybean yields against all management variables (Fig. 6). This tree explained a large part of the variation in yields, with 76% for proportional reduction in error (PRE). The optimum regression tree had three splits and four terminal nodes. The first split in the tree occurred at a P rate of 22.4 kg ha^{-1}, which suggested that the amount of P application was the most important factor in determining yield. This split produced two groups of data: one was lower P rate group with average yield of 2.13 t ha^{-1}, and the other was higher P rate group with average yield of 2.65 t ha^{-1}. This single split in the data accounted for 60.6% of the total variation in yield. Each of these two groups was further split again on the basis of FYM addition, which explained 9.3% and 6.2% of the

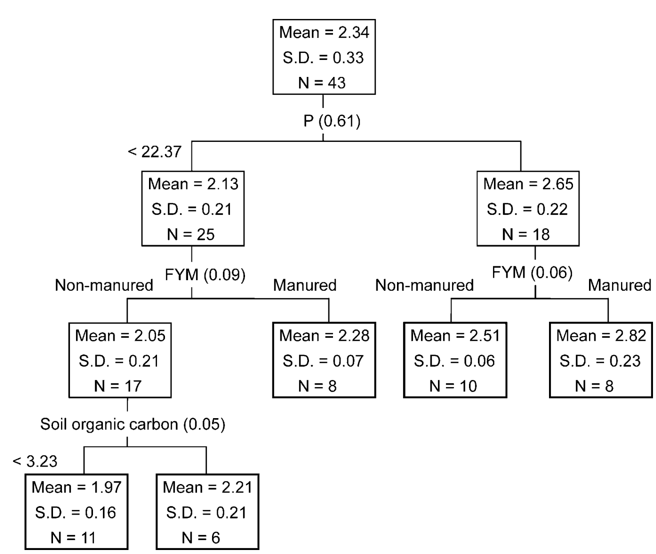

Fig. 7. Regression tree predicting soybean yield from soil variables plus agronomic management variables (PRE = 0.81).

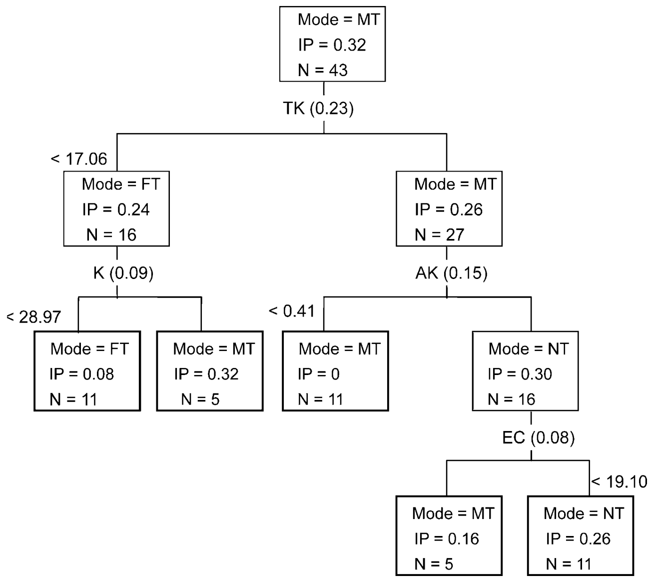

Fig. 8. Classification tree diagram of transects (PRE = 0.55). Number of fields in each group and the impurity (IP) of classification are presented at each node. Partial PRE values are presented in parentheses at each root node to split.

variation, respectively, for low P and high P. The fact that no additional splits were performed indicated that the amount of P applied and FYM were the main factors influencing soybean yield for these fields. Fields with higher P rate and FYM added obtained the best average yield; whereas, fields with lower P rate and without FYM addition had the lowest average yield, while low P and FYM, and high P and no FYM had intermediate yields.

We further developed the regression tree to explore the importance of all variables by integrating soil properties and management practices on the variability of soybean yield. The final regression tree identified three main functional properties which explained nearly 81% of yield variability (Fig. 7). The first two splits in the tree structure were the same as the first two splits in the model with only management variables (Fig. 6). The bootstrap resampling analysis confirmed the reliability of the tree topology; the bootstrap probabilities are 100% for P applied and 55% for FYM addition. SOC, the only soil property variable selected by the analysis, caused a third split in the low P, non-manured branch, but only explained 5% of yield variation.

3.4. Causes of yield variability among transects

Causes of yield variability among the transects were assessed by CART analysis using the transect as a categorical dependent variable. Yield variability among transects was largely determined by variations in soil and management. Thus, TK, AK, amount of K applied and EC which caused the variations among the three transects were selected as determinant factors of the yield variability (Fig. 8). TK was the primary splitting variable in the classification tree, explaining 23% of data variation. In most of the fields, FT had lower TK (TK < 17.06 g kg^{-1}) and the impurity of the classification was 24%. Misclassification (8%) for FT was reduced after the data were further split by K rate of 28.97 kg ha^{-1}. The remaining fields are separated into two groups dominated by MT and NT, based on AK threshold of 0.41 g kg^{-1}. In the left-hand branch, 11 fields of MT with low AK were classified as a terminal group and no misclassification was observed. On the other hand, the right portion of the tree was dominated by NT with impurity of 30%. Along the right branch, further partitions were presented based on EC of 19.1 Ds m^{-1}. The overall model provided a total PRE of 55% in yield variation among the transects.

4. Discussion

During 2007, the area experienced a severe drought with June temperatures up to 2.3 °C above the means since 1979, and June precipitation deficits up to 75 mm, 80% below the average (Fig. 1). According to the CART models, the most important factors affecting yield variability were amount of P applied, FYM application (yes or no) and SOC (Fig. 7). In this drought year, over 80% of the soybean yield variation was explained by these three variables. The first most explanatory split was into branches with high and low P rate, which explained roughly 61% of the yield variability. Phosphorus and nitrogen are the most important limiting nutrient elements for plant growth and production (Elser et al., 2007). A plant-soil nutrient model study showed that enhanced plant growth due to increased N should conserve P because of more dissolved organic P loss (Perring et al., 2008). For soybean, symbiotic N_2 fixation has been observed to be sensitive to drought and reduced in fixation is associated with decreased yield capacity (Sinclair et al., 2000). More application of P increased amount of N accumulated in grain indicating stimulation of biological N_2 fixation (Graham and Vance, 2000; Ogoke et al., 2003). This suggests that more P rate can alleviate yield reduction due to decline of N_2 fixation resulting from water deficit. In addition, P fertilizer application to soybean could improve the root morphology, P uptake and consequently grain yield when drought stress occurred at the reproductive stage (Jin et al., 2005). However, farmers interviewed did not have enough knowledge to make effective decisions on fertilizer application. According to our interviews, roughly 85% of household heads never noticed the ratios of N:P_2O_5:K_2O labeled on the bags of fertilizer. Surprisingly, we found that there were about 20 types of fertilizer with different ratios of N:P_2O_5:K_2O used in the selected fields. The percentages of N, P_2O_5 and K_2O ranged from 11 to 18%, 14 to 25% and 10 to 16%, respectively. The diversity of fertilizer type caused notable differences in nutrients input among the fields. Farmers were very confused for selecting the right types of fertilizer for their fields.

Many long-term fertilization experiments showed that a combination of manure and fertilizer resulted in the best soil productivity (Greenland, 1997). We also found that manure addition is an effective management practice for increasing soybean yield. In the regression tree analysis, both the higher and lower P rate groups were further split according to whether or not manure was added (Figs. 6 and 7). The fields with manure addition had average yield about 10% higher than non-manured fields. However, in the study area, more and more farmers have indicated a preference to apply inorganic fertilizer rather than collect manure for application to their fields, presumably, because of reduced labour for inorganic fertilizer application. Furthermore, all of the plant residue is normally removed for fuel and fodder. These two management options tend to reduce SOC and may have long-term adverse effects on soil productivity. Our CART models demonstrated that SOC was the main factor influencing to yield for these fields, and the fields with lower soil carbon content experienced significant yield reduction (Figs. 5 and 7).

Although contribution of soil parameters and management to yield variability can be separated in 1 year, soil variability often interacts with management practices. Within-farm soil heterogeneity is partly a result of the differential management practices for a long time, particularly on farms with level landscape and low inherent natural variability. Farmer interviews indicated that in most cases, more inputs (e.g. organic manure and mineral fertilizer) were preferentially allocated to the fields near homestead (i.e. near transect), which led to decreasing soil fertility and crop yields with increasing distance from homestead on smallholder farms. Similar practices were noted by Zingore et al. (2007)

in Zimbabwe and Tittonell et al. (2005) in Kenya. On the other hand, farmers often adjust their management practices from year-to-year according to their perception of soil fertility. Overall, most of the variability in crop yields can be attributed to the variability of annual and long-term management practices. In our study, the CART model (Fig. 8) showed that soybean yield variation within the transects was affected by variability in K including residual soil K content assumed to result from various long-term various management practices, and applied K. Such a consequence was likely a result of more manure addition for the fields near the village. Farmers do not want to walk to the farthest transect with a load of manure, and would likely preferentially dump it on land nearest to the village where the animals are housed and manure is produced. Manure was applied to 64% of the fields in NT and 42% of the fields in MT while only about 8% of the fields in FT were manured. This probably led to lower soil K content in the fields of FT when compared to those of NT and MT (Table 3). The results imply that yield variability among the three transects can be reduced through changes in K application rate and uniform manure addition among the three transects.

With asymmetric distribution of growing season precipitation and increasing temperature accompanying climate change, long dry summers will likely be the norm in the future. This study attempted to identify the most important soil and management practices factors that distinguish high yielding from low yielding fields under drought conditions. We found that improved application of P rate in the fields with lower yields was the most pressing management need for increasing soybean production; manure addition also provided greater yields in such a drought year. These findings indicate that cropping systems can be adapted to increasing drought condition by altering management options. Our analysis provides information that extension workers can use to coach farmers to change management practice to obtain higher yields. In addition, selection for genotypes with N_2 fixation tolerance to drought is an effective adaptation option for soybean (Sinclair et al., 2000; Purcell et al., 2004). Factors in determining yield variability and the magnitude of their importance depend on year-to-year dynamics of climate (Lobell et al., 2005). While this analysis provides an indication of the predominant factors affecting yield in a drought year, 1 year's data may be insufficient for fully understanding the factors that limit yields in farmers' fields under drought conditions. Further investigation of yield variability and its determining factors is needed to continue this assessment in other climatic conditions.

5. Conclusions

There were large spatial variations in soybean yield, both among all fields and transects. CART was determined to be a suitable analysis tool, and was used to identify the most important yield determining factors, even though the relationships may be non-linear. The results of this study showed that although variation in soil parameters is often considered to be the major contributor to yield variability, only a small part of soybean yield variability could be explained by soil variation. Yield variability from field to field was largely dependent on management practices, which was in line with findings of previous studies, such as maize systems of western Kenya (Tittonell et al., 2008) and wheat systems of Yaqui Valley (Lobell et al., 2005). Although soil TK was the first most important factor in affecting yield variability among transects, its level was most likely a result of long-term management practices, especially manure management. These conclusions indicate that yield variability within fields can be significantly reduced by adjusting agronomic practices in those fields with lower yields. On the basis of these results, we recommend that the scope of government soil fertility programs should be broadened to explore benefits of improved agronomic management and to promote and integrate these with conventional soil tests in the study area. In addition, other effective management practices and drought-tolerant variety should be explored to adapt crops to future climate change.

Acknowledgements

We thank N.B. McLaughlin, Agriculture and Agri-Food Canada, for providing review of earlier drafts of this manuscript and helping with English revisions, and all of the people involved in the experiments. This work was supported by the National Basic Research Program of China (2005CB121107), the National Natural Science Foundation of China (40621061), and the Chinese Academy of Sciences (KZCX2-YW-421).

References

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman and Hall (Wadsworth, Inc.), New York, NY, USA, 254 pp.
Cassman, K.G., Dobermann, A., Walters, D.T., Yang, H., 2003. Meeting cereal demand while protecting natural resources and improving environmental quality. Annu. Rev. Env. Resour. 28, 315–358.
Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grunwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Vesala, T., Valentini, R., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533.
Conway, G., Toenniessen, G., 1999. Feeding the world in the twenty-first century. Nature 402, C55–C58.
Dai, A.G., Trenberth, K.E., Qian, T.T., 2004. A global dataset of palmer drought severity index for 1870–2002: relationship with soil moisture and effects of surface warming. J. Hydrometeorol. 5, 1117–1130.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.
Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B., Smith, J.E., 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol. Lett. 10, 1135–1142.
FAOSTAT, 2007. FAOSTAT Data. http://faostat.fao.org/default.aspx.
Graham, P.H., Vance, C.P., 2000. Nitrogen fixation in perspective: an overview of research and extension needs. Field Crop Res. 65, 93–106.
Greenland, D.J., 1997. Inaugural Russell memorial lecture—soil conditions and plant growth. Soil Use Manage. 13, 169–177.
IPCC, 2007. Climate change 2007: impacts, adaptation and vulnerability. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E. (Eds.), Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, Cambridge.
Jin, J., Wang, G.H., Liu, X.B., Pan, X.W., Herbert, S.J., 2005. Phosphorus application affects the soybean root response to water deficit at the initial flowering and full pod stages. Soil Sci. Plant Nutr. 51, 953–960.
Liu, X., Jin, J., Wang, G.H., Herbert, S.J., 2008. Soybean yield physiology and development of high-yielding practices in Northeast China. Field Crop Res. 105, 157–171.
Lobell, D.B., Ortiz-Monasterio, J.I., Asner, G.P., Naylor, R.L., Falcon, W.P., 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 97, 241–249.
Lobell, D.B., Ortiz-Monasterio, J.I., 2006. Regional importance of crop yield constraints: linking simulation models and geostatistics to interpret spatial patterns. Ecol. Model. 196, 173–182.
Lu, R.K., 1999. Methods for Soil and Agriculture Chemistry Analysis. Chinese Agricultural Science and Technology Press, Beijing (in Chinese).
Mann, C., 1997. Reseeding the green revolution. Science 277, 1038–1043.
Ogoke, I.J., Carsky, R.J., Togun, A.O., Dashiell, K., 2003. Effect of P fertilizer application on N balance of soybean crop in the guinea savanna of Nigeria. Agric. Ecosyst. Environ. 100, 153–159.
Park, S.J., Hwang, C.S., Vlek, P.L.G., 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agric. Syst. 85, 59–81.
Perring, M.P., Hedin, L.O., Levin, S.A., McGroddy, M., de Mazancourt, C., 2008. Increased plant growth from nitrogen addition should conserve phosphorus in terrestrial ecosystems. Proc. Natl. Acad. Sci. U.S.A. 105, 1971–1976.
Purcell, L.C., Serraj, R., Sinclair, T.R., De, A., 2004. Soybean N_2 fixation estimates, ureide concentration, and yield responses to drought. Crop Sci. 44, 484–492.
Ray, J.D., Heatherly, L.G., Fritschi, F.B., 2006. Influence of large amounts of nitrogen on nonirrigated and irrigated soybean. Crop Sci. 46, 52–60.
Schiermeier, Q., 2008. Water: a long dry summer. Nature 452, 270–273.

Sinclair, T.R., Purcell, L.C., Vadez, V., Serraj, R., King, C.A., Nelson, R., 2000. Identification of soybean genotypes with N_2 fixation tolerance to water deficits. Crop Sci. 40, 1803–1809.
Sinclair, T.R., Purcell, L.C., King, C.A., Sneller, C.H., Chen, P.Y., Vadez, V., 2007. Drought tolerance and yield increase of soybean resulting from improved symbiotic N_2 fixation. Field Crop Res. 101, 68–71.
Tittonell, P., Vanlauwe, B., Leffelaar, P.A., Shepherd, K.D., Giller, K.E., 2005. Exploring diversity in soil fertility management of smallholder farms in western Kenya. II. Within-farm variability in resource allocation, nutrient flows and soil fertility status. Agric. Ecosyst. Environ. 110, 166–184.
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. Agric. Ecosyst. Environ. 123, 137–150.
Xie, F.D., Dong, Z., Sun, Y.H., Wang, X.G., 1994. Influence of drought on growth and yield of soybeans at different growth stages. J. Shenyang Agric. Univ. 25, 13–16 (in Chinese).
Zhao, H.W., Li, Q.Z., Wei, Y.X., 2006. Effect of drought at different growth stages on main physiological parameters and yield in soybean. Soybean Sci. 3, 329–332 (in Chinese).
Zingore, S., Murwira, H.K., Delve, R.J., Giller, K.E., 2007. Influence of nutrient management strategies on variability of soil fertility, crop yields and nutrient balances on smallholder farms in Zimbabwe. Agric. Ecosyst. Environ. 119, 112–126.
Zou, X.K., Zhai, P.M., Zhang, Q., 2005. Variations in droughts over China: 1951–2003. Geophys. Res. Lett. 32, L04707, doi:10.1029/2004GL021853.