Europ. J. Agronomy 77 (2016) 10–19

Contents lists available at ScienceDirect

European Journal of Agronomy

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

An approach to understanding persistent yield variation—A case study in North China Plain

Yi Zhao[a,*], XinPing Chen[b], David B. Lobell[a]

[a] Department of Environmental Earth System Science and Center on Food Security and the Environment, Stanford University, Stanford, CA 94305, United States

[b] Center for Resources, Environment and Food Security, China Agricultural University, Beijing 100193, China

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form 16 March 2016
Accepted 20 March 2016
Available online 31 March 2016

Keywords:
Maize
China
Survey
Remote sensing

ABSTRACT

Large gaps between maize yields on average farmers' fields and the highest yields achieved by either experiment or farmers are typical throughout the developing world, including in the North China Plain (NCP). Understanding the underlying causes to this yield gap is important for prioritizing strategies for shrinking this gap and improving food security. Quzhou county in Hebei province is typical of the winter-wheat summer-maize system in NCP where the average plot size is only 0.25 ha. To analyze this cropping system amidst the challenge of substantial heterogeneity, we identified fields that were either persistently higher or lower yielding according to remote sensing yield estimates, and then conducted detailed field surveys. We found irrigation facility to be a major constraint to yield both in terms of irrigation water quality and farmers' access to wells. In total, improving the access to unsalty water would be associated with a 0.32 t/ha (4.2%) increase in multi-year average yield. In addition, farmers' method of choosing cultivar, which likely relates to their overall knowledge level, significantly explained yield variation. In particular, those choosing cultivars according to technician advice, personal experiences and high yielding neighbors' advice had on average higher yield than farmers that either followed seed sellers' advice or collectively purchased seeds.

1. Introduction

Improving yield of the lowest yielding smallholder farmers in developing countries is essential to food security and livelihood of farmers. Yield gap, which can be empirically defined as the difference between the average and the highest (95 or 98 percentile) farmer's yield, is one way to quantify the yield variation among farmers and the potential for improving average yields through management changes (Lobell et al., 2009). While major wheat and maize in the developed countries have typically reached 80% or greater of model simulated yield potential, yield gap remains high in the developing world (Neumann et al., 2010; van Ittersum et al., 2013).

One of these large yield gap cases is maize in China, where the average farmers' yield is 48% of model simulated potential, 64% of experimental potential, and 51% of historically recorded highest yield (Meng et al., 2013). Most farmers in China are cultivating small land areas. In particular, North China Plain (NCP), whose average farmers' yield are at 41% of potential, has an average cultivated area of 0.7 ha per household (Meng et al., 2013). This large yield gap is likely driven by heterogeneity in both management practices and biophysical conditions faced by smallholder farmers. A key question in any particular region is to identify the handful of factors that are the most important drivers of yield variation.

Previous studies have investigated various factors, including biophysical, agronomic, and socio-economic factors (Lobell et al., 2009; Liang et al., 2011; Affholder et al., 2013). Biophysical factors mainly include soil conditions along with irrigation water quality for irrigated systems. Specifically among soil conditions, cation exchange capacity (Miao et al., 2006), organic matter content (Cai and Qin, 2006; Zhen et al., 2006; Gong et al., 2009), and soil salinity (Ma et al., 2008; Kang et al., 2010) have been suggested as important factors to yield gap in NCP. For management practices, Liang et al. (2011) conducted a survey for 2004–2005 season that covered 362 farmers and six counties of Hebei province. They qualitatively identified nutrient management, irrigation facility and mechanization issues as well as high opportunity cost of labor, small farm size and lack of technical service as major causes to the yield gap.

A major challenge to yield gap analysis is the significant amount of temporal heterogeneity in yield and yield-controlling factors.

* Corresponding author.
E-mail addresses: yizhao@stanford.edu, jeanzhaoyi@gmail.com (Y. Zhao).

http://dx.doi.org/10.1016/j.eja.2016.03.006

For instance, a two-year survey of wheat in Yaqui valley of Mexico found the most significant factor to be irrigation timing in one year but fertilization amount in another year that was cooler than usual (Lobell et al., 2005). Ideally one would have a time series of survey data that comprehensively represent both spatial variation in yield and temporal variation in weather. However, that is often infeasible given substantial cost in time and labor.

In this paper, we assess causes of the yield gap for maize in Quzhou county in NCP. Our study is unique from prior work in that we (i) focused on average yields over a five year period rather than yields in individual years, in order to avoid confounding of results by temporal variability, and (ii) ensured an adequate sample of both low and high yielding fields by using remote sensing to identify persistently high or low yielding fields. We conducted field surveys that asked farmers for their yield over the past five years as well as various potential explanatory biophysical, management, and socio-economic factors. Relationships were analyzed using ordinary linear regression. These results were also compared with those using one-year surveyed yield as the response variable to test the importance of a multi-year perspective. And we further compared the ordinary linear regression results with analyses using nonlinear methods (Random forests and Multivariate adaptive regression splines (MARS)) to test for potential nonlinear relationships.

2. Method

2.1. Study site

Our study area is Quzhou County, which is located in Hebei Province of NCP and has an area of 67,700 ha. Major Crops grown at this county include summer maize and winter wheat, cotton, spring maize, and various other vegetables or cash crops. It is typical of NCP in that individual households hold multiple plots that are spread out within their village with area ranging from 0.5 to 30 mu, or 0.03–2 ha, with an average of 0.25 ha per plot. The total area cultivated per household is 0.6 ha on average. This phenomenon began since the land reform in early 1980s when each household was allocated a strip or square of land parcel from fields at multiple locations in its village. The rational was to ensure fairness in that each household should be allocated the land of both higher and lower than average fertility level. As a result, each field is managed by multiple farmers, each with individual plots of land within the field.

As typical of NCP, almost all households have access to some irrigation facilities, which include water pumps as well as adjacent wells or water canals. The majority of the county is irrigated by wells, but the depths of wells vary, and each well is shared by five to 50 farmers. Farmers thus have to wait for their turn to irrigate, unless they own their well or have access to a close-by canal for irrigation. The wait time for using a well ranges from one to three weeks, whereas those using water from a canal or a river do not have to wait. There are two major rivers running through Quzhou county, and soils have alluvial diluvia parent material that is representative of fluvial materials in NCP. More specifically, there are five major categories of soil texture characteristics, which include clay, light loam, medium loam, salt-affected and sandy loam. Quzhou county used to suffer from severe salinization, but land productivity improved significantly since effective salt flushing in the 1980s (Sheng and Xiuling, 1997).

2.2. Identification of survey fields

We sampled fields with a goal of better representing the highest and lowest percentiles in yield, similar to some previous spatial analyses (Lobell et al., 2007; Dang et al., 2011). The sampling scheme requires prior information of multiple years' yield distribution of the study area. We used remote sensing estimated yield of Quzhou county for 2007, 2009, 2010, 2012 and 2013(Zhao et al., 2015). Estimates for 2008 and 2011 were not made because of the lack of cloud-free satellite images. Since yield varied from year to year, we needed a metric that was uniform across years in evaluating within year relative yield performance. For each year, we categorized the five year yield estimates into ten yield decile groups (0–10%, 11–20%,. . ., 90–100%) for each soil type, where 10 denoted the highest and 1 denoted the lowest percentile group. Then, we temporally aggregated this metric for each pixel and kept only those that had missing values in no more than two years. Among those pixels, we chose those that had a mean decile of 3.5 or below as the lowest yielding fields and a mean decile of 8 or above as the highest yielding fields. Table 1 summarized the population size for sampling.

Table 1
Soil parent material sampling size in the survey.

Soil Parent Material	Total Area (km^2)	Low Yielding (km^2)	High Yielding (km^2)
Salt-affected	914.6	86.0	30.2
Clay	1343	122	66.5
Light-loam	3870	556	281
Medium-loam	1315	125	78.5
Sandy	421.2	30.5	31.2

Table 2
Number of fields surveyed by soil type and yield categories.

Soil Parent Material	Low Yielding	High Yielding	Less Persistent
Salt-affected	3	2	7
Clay	2	6	4
Light-loam	2	9	9
Medium-loam	3	6	4
Sandy	3	2	2
Total	13	25	26

We randomly sampled fields that encompassed those persistently high or low yielding pixels for our survey. In addition to the persistently high and low yielding fields, there were also less persistent fields sampled to fully represent the region. In total, we surveyed 64 fields, among which 25 were high yielding, 13 were low yielding and 26 were less persistent. Table 2 summarized the number of fields within each soil type. There were not as many low yielding fields identified as high yielding fields because the persistently low yielding pixels were mostly scattered, especially at the edges of fields, whereas high yielding pixels were mostly agglomerated. Since the survey was field based, fields were first located using a GPS, and then corresponding households that managed those fields were identified with the guidance of local farmers. According to the size of a field, three to eight households were randomly selected. In total, 217 households were surveyed from 64 fields in 47 villages. The locations of surveyed fields were illustrated in Fig. 1.

2.3. Survey contents

The survey was conducted in the summer of 2014. Since the focus of the survey was on persistent yield gaps, farmers were asked about their yield in each of the past five years, or 2009–2013. The average of those five years' yield would better reflect their yield performance than a temporal snapshot of single-year yield. For simplicity of notation, the five-year average yield response variable was denoted as *Ym*. The reported yield of the most recent maize season was for 2013, which was denoted as *Ys*. In addition to yield, survey questions cov-