Journal of Agricultural Science (2014), **152**, 523–533.
doi:10.1017/S0021859613000142

CLIMATE CHANGE AND AGRICULTURE RESEARCH PAPER

Weather effects on maize yields in northern China

B. J. SUN[1,2]* AND G. C. VAN KOOTEN[2]

[1] *College of Economics and Management, Northwest A&F University, No. 3 Taicheng Road, Yangling, Shaanxi 712100, China*
[2] *Department of Economics, University of Victoria, PO Box 1700, STN CSC, Victoria, BC V8W 2Y2, Canada*

(*Received 20 July 2012; revised 8 February 2013; accepted 26 February 2013; first published online 27 March 2013*)

SUMMARY

In the present study, the effect of weather on maize yields in northern China was examined using data from 10 districts in Inner Mongolia and two in Shaanxi province. A regression model with a flexible functional form was specified on the basis of agronomic considerations. Explanatory variables included in the model were seasonal growing degree days, precipitation, technological change (e.g. adoption of new crop varieties, improved equipment, better management, etc.) and dummy variables to account for regional fixed effects. Results indicated that a fractional polynomial model in growing degree days could explain variability in maize yields better than a linear or quadratic model. Growing degree days, precipitation in July, August and September, and technological changes were important determinants of maize yields. The results could be used to predict potential maize yields under future climate change scenarios, to construct financial weather products and for policy makers to incentivize technological changes and construction of infrastructure (e.g. irrigation works) that facilitate adaptation to climate change in the agricultural sector.

INTRODUCTION

China is the second largest maize producing country in the world after the United States (FAO 2010) and accounts for nearly 0·20 of the global population; therefore it is important to study how climate affects maize yields in China. However, few studies have examined the impact of climate factors on crop yields in China, partly because data are scarce, difficult to obtain and of varying quality. Nevertheless, as a basis for understanding current and future climate risks, it is important to increase knowledge about the relationship between existing weather records and crop yields.

Many studies have investigated climate effects on crop yields in various regions using different methods, including plant simulation models (Vučetić 2011). Most statistical approaches, however, have employed simple correlations or estimated a relationship that is either linear or quadratic (Almaraz *et al.* 2008; Chen *et al.* 2011). For example, Li *et al.* (2011) regressed maize yields on climate variables distinguished by whether they affected the planting or growing periods, using both linear and quadratic forms of these variables. Another relevant study by Chen *et al.* (2011) examined the impact of weather on maize yields in three provinces (Heilongjiang, Jilin and Liaoning) located along the northeast coast of China; they found that the minimum temperature deviations for May and September had a strong positive effect on maize yield, where yield in a given year was also measured as differences from average yield over the 44 years from 1965 to 2008 for which data were available. Separate regressions were estimated for each province, and for the three provinces combined, with intercept terms included only in regressions for Heilongjiang and Liaoning provinces and not for Jilin province or the aggregated regression. Not unexpectedly, temperatures in May had a stronger impact on yield than temperatures in September.

Chen *et al.* (2011) sought only to find the climate factors correlated with maize yields and did not attempt to provide agronomic insights into their results. Maize is planted in early May and a high minimum May temperature is desirable in the study region used by Chen *et al.* (2011); moisture is not a limiting factor

* To whom all correspondence should be addressed. Email: baojingsun@gmail.com

Fig. 1. Study area showing 12 districts in two provinces (Source: Data Sharing Infrastructure of Earth System Science from 1977–96). YL: Yulin; YA: Yanan; HL: Hulunbeier; XA: Xinganmeng; TL: Tongliao; CF: Chifeng; WL: Wulanchabu; HH: Huhehaote; BT: Baotou; EE: Eerduosi; BY: Bayannaoer; AL: Alashanmeg.

in May and temperatures are not high enough to adversely impact early growth. Maize is harvested during September (with dates varying according to latitude); since high minimum temperatures are indicative of frost-free days, higher minima contribute to higher quality harvests and yields.

In the present paper, the work of Chen *et al.* (2011) is extended in several directions. First, the focus is on a partially adjacent region to the west, the provinces of Inner Mongolia and Shaanxi, which are part of the main maize growing regions of China (Fig. 1). These provinces, respectively, accounted for 0·083 and 0·030 of China's total maize production in 2010 (National Bureau of Statistics of China 2011). Thus, they provide some indication regarding the impact that climate factors might have on the availability of maize in China under projected climate change. Second, district-level data for the two provinces were employed (10 districts in Inner Mongolia and two districts in Shaanxi). Since data in the present research are disaggregated to a greater extent than those used by Chen *et al.* (2011), only district-level data for the period 1989–99 (11 years) for Inner Mongolia and 1989–2001 (13 years) for Shaanxi could be employed. However, since disaggregated weather data and panel regression were used, regression models in the present paper have 136 observations compared with only 44 observations used by Chen *et al.* (2011).

Third, the data used in the present paper only cover the period following the agricultural reforms that were completed by 1983 (Lin 1988, 1992) and more broad-based economic reforms in place by 1989 (Huang *et al.* 2009). Another structural change occurred around 2001–02 when China joined the World Trade Organization (WTO). Chen *et al.* (2011) covered a period that spanned years on both sides of these important events, but failed to test for potential structural breaks in yields. That is, given the potential impact of socio-economic factors on crop yields, this should perhaps have been taken into account. This problem is avoided in the present paper because the data do not span the events in question.

Finally, there is no reason to think that crop yields are a linear or quadratic function of climate variables. Rather, the relationship is more complicated. For example, Schlenker & Roberts (2006) specified a highly nonlinear relationship between temperature and yields, employing daily temperature information (Schlenker & Roberts 2006, 2008). In the current paper, therefore, a nonlinear flexible functional form is used rather than the linear one employed by Chen *et al.* (2011). Instead of relying on daily maximum and

https://doi.org/10.1017/S0021859613000142 Published online by Cambridge University Press

minimum temperatures, however, growing degree days derived from daily temperature information are employed – the number of observations available on yield in the present study is simply too few to permit the use of non-aggregated daily temperature data. These aspects are discussed further below.

MATERIALS AND METHODS

Methods

The research was conducted at the Department of Economics, University of Victoria, Canada. Schlenker & Roberts (2006, 2008) made the case that crop yields depend on climate factors in a nonlinear fashion and that crop growth depends on cumulative heat throughout the growing season. That is, if soil moisture availability is generally not a constraint during the growing season, then crop growth is not specifically dependent on one or several periods of warm weather but, rather, a function of the total sum of warm days during the growing season. In the present study, tests on the parameters of the interaction between precipitation and temperature turned out to be statistically insignificant. However, extreme temperatures above 36 °C, for instance, may have an adverse impact on crop yield (Dupuis & Dumas 1990), as might lack of soil moisture during the growing season. In the present study, hot days (with maximum temperatures above 36 °C) accounted for $<0{\cdot}03$ of all days in the dataset and $>0{\cdot}80$ of these occurred in one district. Further, although we would expect the associated district dummy variable to account for hot days, it was found that the 'number of hot days' variable was statistically insignificant in any variant of the regression model.

Schlenker & Roberts (2006, 2008) began by postulating that plant growth is cumulative over time and that yield is directly dependent on plant growth. Let T represent temperature and y_{jt} represent the log of plant growth in region j in year t. Then, assuming yield growth is given by $g(T)$, the natural logarithm of crop yield is determined by the following relationship:

$$y_{j,t} = \int_{\underline{T}}^{\bar{T}} g(T)\Phi_{j,t}(T)\mathrm{d}T + \sum_i \alpha_i z_{i,j,t} + \theta_j D_j + \varepsilon_{j,t} \qquad (1)$$

where $\bar{T}$ and $\underline{T}$ refer to the upper and lower bounds that observed temperatures can take; $\Phi_{j,t}(T)$ is the cumulative distribution function of temperatures (heat) over the growing season in region j during year t; $z_{i,j,t}$ are i other factors (precipitation, technology, fertilizers, etc.) that affect crop growth in region j during time t; α_i and θ_j are parameters to be estimated; D_j are time-invariant region-fixed effects; and $\varepsilon_{j,t} \sim N(0,\ \sigma_{j,t})$ are identical independently distributed error terms.

The issue of concern relates to the growth function. Schlenker & Roberts (2006) employed a flexible mth-order Chebychev polynomial evaluated at the m midpoints of the intervals between $\bar{T}$ and $\underline{T}$. Unfortunately, while Schlenker & Roberts (2006) had 87 619 observations on maize yields for the period from 1950 to 2004, data used in the present research were severely limited as discussed below. Therefore, an alternative flexible functional form is employed. When observations are limiting, Schlenker & Roberts (2006) recommended that, at the very least, growing degree days during the growing season and the square of growing degree days be used as explanatory variables in the regression model instead of temperatures *per se* (Schlenker & Roberts 2006). The quadratic function can capture at least one nonlinear aspect, although they found many more nonlinearities using midpoints of intervals or dummy variables representing number of days that temperatures fall within certain bounds during a growing season. As noted, they were afforded this luxury by their large data sets.

Temperature is the primary variable of interest, similar to what was assumed and confirmed by Schlenker & Roberts (2006, 2008), although precipitation may be important at certain times of the year. For example, too much rainfall in September may delay harvests, reduce grain quality, or even reduce overall yield. Likewise, if fields are too wet in the autumn, harvesting may be delayed and yields may actually decline due to decay; or if fields are too wet in spring the delay in planting leads to reduced exposure to heat, or exposure to heat at the wrong time in the growing cycle, and thereby lower yields according to Eqn (1). In contrast to Schlenker & Roberts (2006, 2008), precipitation effects for each month, rather than the growing season as a whole, were considered in the present research, and quadratic effects of monthly precipitation were also considered since it was not obvious that the effect of precipitation on crop yield was linear.

Growing degree days (denoted by G) were employed in the present research and, because the mth-order Chebychev polynomial of Schlenker & Roberts (2006, 2008) was not used to represent the effect of heat on crop yields, the method of fractional polynomials (Royston & Altman 1994) was adopted to model the nonlinear relation between crop yield and G. The use of G and G^2 as explanatory variables is thus considered a special case of the more general mth-order fractional polynomial regression.

Royston & Altman (1994) began by defining a fractional polynomial of degree m written as

$$f_m(X;\ \xi,\ \boldsymbol{n}) = \xi_0 + \sum_{j=1}^{m} \xi_j X^{(n_j)} \tag{2}$$

where the parentheses on the power term on X signify the following transformation:

$$X^{(n_j)} = \begin{cases} X^{n_j}, & \text{if } n_j \neq 0 \\ \ln X, & \text{if } n_j = 0 \end{cases} \tag{3}$$

For $m = 2$ and $\boldsymbol{n} = \{n_1,\ n_1\}$, Eqn (2) can be written as

$$F_2(X;\xi,\ \boldsymbol{n}) = \xi_0 + (\xi_1 + \xi_2)\, X^{(n1)} \tag{4}$$

which is nothing more than a fractional polynomial of degree 1.

Rewrite Eqn (4) as

$$f_2(X;\ \xi,\ \boldsymbol{n}) = \xi_0 + \xi_1 X^{(n1)} + \xi_2 X^{(n1)} X^{(n2)-(n1)} - \xi_2 X^{(n1)} + \xi_2 X^{(n1)} \tag{5}$$

Rearranging gives

$$f_2(X;\xi,\boldsymbol{n}) = \xi_0 + \xi_1 X^{(n1)} + \xi_2 X^{(n1)} (X^{(n2)-(n1)} - 1) + \xi_2 X^{(n1)} \tag{6}$$

$$\Rightarrow f_2(X;\ \xi,\ \boldsymbol{n}) = \xi_0 + (\xi_1 + \xi_2) X^{(n1)} + \xi_2 (n_2 - n_1) X^{(n1)} \times \left[\frac{(X^{(n_2-n_1)} - 1)}{(n_2 - n_1)}\right] \tag{7}$$

As $n_2 \to n_1$, the last term in parentheses in Eqn (7) becomes $\ln X$. Then, upon rearranging:

$$f_2(X;\ \xi,\ \boldsymbol{n}) = \xi_0 + (\xi_1 + \xi_2)\, X^{(n1)} + \xi_2 (n_2 - n_1)\, X^{(n1)} \ln X \tag{8}$$

Letting $\zeta_0 = \xi_0$, $\zeta_1 = \xi_1 + \xi_2$ and $\zeta_2 = \xi_2(n_2 - n_1)$, Eqn (8) can be written as

$$f_2(X;\ \xi,\ \boldsymbol{n}) = \zeta_0 + \zeta_1 X^{(n1)} + \zeta_2\, X^{(n1)} \ln X \tag{9}$$

This can then be generalized in the same way to $m > 2$ (Royston & Altman 1994):

$$f_m(X;\ \xi,\ \boldsymbol{n}) = \zeta_0 + \zeta_1 X^{(n1)} + \sum_{j=2}^{m} \zeta_j X^{(n_1)} \ln X^{j-1} \tag{10}$$

Notice that, much like a Taylor series expansion occurs about a particular point, the flexible functional form Eqn (10) is obtained by expanding around the power n_1. This is clear from the way that Eqn (9) is derived. Therefore, the fractional polynomial function is generalized further:

$$f_m(X;\ \xi,\ \boldsymbol{n}_k) = \zeta_0 + \sum_{k=1}^{K} \left[\zeta_{1,k} X^{(n_k)} + \sum_{j=2}^{m} \zeta_{j,k} X^{(n_k)} \ln X^{j-1}\right] \tag{11}$$

where the number of potential powers equals K. Essentially there are an infinite number of functions that can be considered, but, in practice, the powers are limited to n_k values that are integers between -2 and $+3$, with $\pm 0{\cdot}5$ ($1/\sqrt{X}$ and $\sqrt{X}$) included, although higher powers are not ruled out.

The notation and method are illustrated for identifying a regression model with several examples. Consider the fractional polynomial for variable X and let K denote powers of X and m the maximum number of terms of a particular power of X; in these examples, an intercept term and terms involving other variables are excluded (although treated as a single variable in the text, explanatory variable X could just as well be considered a vector of variables). For example, if $K = 3$ and $m = 3$, there would be three powers of X with one power having three terms, although there might be other powers of X whose terms do not exceed m. Consider $\{X,\ (-0{\cdot}5,\ 0,\ 1,\ 1,\ 1)\}$. This leads to the following regression equation:

$$f_3(X;\ \xi,\ \boldsymbol{n}_3) = \zeta_0 + \zeta_1 \frac{1}{\sqrt{X}} + \zeta_2 \ln X + \zeta_3 X + \zeta_4 X \ln X + \zeta_5 X \ln X^2 \tag{12}$$

Likewise, $\{X,\ (0,\ 0,\ 1,\ 1,\ 1,\ 2,\ 3)\}$ leads to

$$F_3(X;\ \xi,\ \boldsymbol{n}_4) = \zeta_0 + \zeta_1 \ln X + \zeta_2 \ln X^2 + \zeta_3 X + \zeta_4 X \ln X + \zeta_5 X \ln X^2 + \zeta_6 X^2 + \zeta_7 X^3 \tag{13}$$

One final note regards the definition of power zero, which is set equal to $\ln X$ as is clear from Eqns (12) and (13).

In the current application, the regression model is specified as

$$y_{j,t} = \beta_0 + \sum_{k=1}^{K} \left[\beta_{1,k} G_{i,t}^{(n_k)} + \sum_{j=2}^{m} \beta_{j,k} G_{i,t}^{(n_k)} \ln G^{j-1}\right] + \alpha_i z_{i,j,t} + \theta_j D_j + \varepsilon_{j,t} \tag{14}$$

where the dependent variables ($y_{j,t}$), $z_{i,j,t}$, α_i, θ_j, $\varepsilon_{j,t}$ and D_j were defined in conjunction with Eqn (1), and, importantly, the second term in square brackets in Eqn (14) disappears if $m = 1$. In Eqn (14), βs are parameters to be estimated and $G_{i,t}$ refers to total degree days in region i during growing season t. Further, only functional forms with powers $\boldsymbol{n_8} \in \{-2,\ -1,\ -0{\cdot}5,\ 0,\ 0{\cdot}5,\ 1,\ 2,\ 3\}$ were explored, although Schlenker &

Roberts (2006, 2008) used a sixth power in their function but did not employ fractions and natural logarithms. Higher polynomials and fractions could also be employed, but the powers in the present paper are limited to those indicated. Finally, technological change is represented by a time variable.

Data

As already noted, the study area consists of 12 districts in two provinces in northwestern China (Fig. 1). The northernmost district (Yulin (YL)) in Shaanxi Province is adjacent to Inner Mongolia and constitutes 4·35 million ha (Government of Yulin 2012), while the one to the south (Yanan (YA)) is 3·7 million ha (Government of Yanan 2012); these are part of China's Loess Plateau. Inner Mongolia spans three distinct Chinese administrative units, i.e. Northeast China, North China and Northwest China. Inner Mongolia is the third largest province, with a land base of 118·3 million ha which accounts for 0·12 of China's total land area. Of this, 0·53 of the land is plateau, 0·21 is mountainous and 0·008 is covered by water (Government of Inner Mongolia 2010). Inner Mongolia consists of 12 districts, but because requisite data were missing, only 10 districts were considered in the analysis. District-level details are provided in Table 1. Maize yield data were collected from available annual year books for the Inner Mongolia Autonomous Region and for Shaanxi province; yield data were available for Inner Mongolia only for the period 1989–99 (Inner Mongolia Bureau of Statistics 1990–2000) and for Shaanxi for the period 1989–2001 (Shaanxi Bureau of Statistics 1990–2002). Weather data were from the Ecological Environment Database of the Loess Plateau. There were 50 weather stations in Inner Mongolia for which data were available, but only 38 had weather data for a period comparable with the period for which yield data were available. In addition, records from seven weather stations in northern Shaanxi were available from the database. The maize growing season in the study area is from late April or early May to September. Therefore, weather records from May to September were used.

While yield data were available beyond 2001 for Shaanxi province, the same richness of weather station data for the period after 2001 was lacking; no maize yield data were available for Inner Mongolia after 1999. However, as noted earlier, data prior to 1989 or after 2001 might be affected by socio-economic factors that are not likely to be a significant explanatory factor in the period considered in the present paper.

For each district, the weather stations in the district were used to determine the temperature and precipitation associated with the crop yields. If there was no weather station in a given district, then the temperature and precipitation data for the nearest weather station to the central point (centroid) of the district was used as a proxy. If there was only one weather station in the district, the data from that station were used. Finally, if there were two or more weather stations in a district, weighted averages of the precipitation and temperature data were calculated for the centroid of the district.

To make the required calculations, a Geographic Information System (GIS) model was built using the Quantum GIS (QGIS) tool. After plotting district centroids, the distances between centroids and weather stations were measured using the GIS tool, and then the inverse of these distances were used to calculate weighted coefficients for weather readings. The following formula was used:

$$T_j = \sum_{k=1}^{n_j} [T_{k,j} + 0{\cdot}006(e_{k,j} - e_j)] \left[\frac{(1/d_{k,j})}{\sum_{k=1}^{n_j} (1/d_{k,j})} \right] \qquad (15)$$

where T_j is the region j centroid temperature, $T_{k,j}$ is the temperature reading and $e_{k,j}$ is the elevation at weather station k in district j; there are n_j weather stations in region j; e_j is the mean elevation at the centroid of district j, which is used to eliminate elevation differences among stations in the district; 0·006 (measured in °C) corrects for elevation; and $d_{k,j}$ is the distance between the centroid in district j and weather station k, with

$$\sum_{k=1}^{n_j} \left(\frac{1/d_{k,j}}{\sum_{k=1}^{n_j} (1/d_{k,j})} \right) = 1$$

In China, cumulative temperatures in excess of 10 °C are usually used to measure heat conditions for crop growth (Bai *et al.* 2011). The same temperature is also used in the UK as the basis for calculating growing degree days (Bunting 1979). In the present study, therefore, daily growing degree days were calculated as

$$G_{d,j,t}(T_{d,j,t}) = \begin{cases} T_{d,j,t} - 10, & \text{if } T_{d,j,t} > 10\ ^\circ\text{C} \\ 0, & \text{if } T_{d,j,t} \leq 10\ ^\circ\text{C} \end{cases} \qquad (16)$$

and total seasonal growing degree days as

$$G_{j,t}(T_{j,t}) = \sum_{d=1}^{153} G_{d,j,t} \qquad (17)$$

In the preceding equations, $G_{d,j,t}(T_{d,j,t})$ refers to the number of growing degree days on day d in region j in year t, as a function of the average temperature on day

Table 1. *Information regarding districts in Shaanxi Province and Inner Mongolia Autonomous Region*

District	Abbreviation	Size (m ha)	Coordinates	Altitude (m asl)	Website sources*
Shaanxi Province†					
Yulin	YL	4·35	37°07′–39°02′N 107°35′–111°05′E	1200–1800	http://www.yl.gov.cn/site/1/html/zjyl/list/list_18.htm
Yanan	YA	3·7	35°21′–37°31′N 107°41′–110°31′E	1200	http://www.yanan.gov.cn/structure/zjya/zjyax
Inner Mongolia Autonomous Region†					
Hulunbeier	HL	25·0	47°05′–53°20′N 115°21′–126°04′E	200–1000	http://www.hulunbeier.gov.cn/hlbewh/index.asp
Xinganmeng	XA	6·0	46°39′–47°39′N 119°28′–121°23′E	150–1800	http://www.xam.gov.cn/zwgk/zjxam/136359.htm
Tongliao	TL	6·0	42°15′–45°59′N 119°14′–123°43′E	120–320	http://www.tongliao.gov.cn/gaik_text.asp?bid=194
Chifeng	CF	9·0	41°17′–45°24′N 116°21′–120°58′E	300–2067	http://www.chifeng.gov.cn/html/2010–05/259c01bd–a891–4b30–b6dc–40a01acefd31.shtml
Wulanchabu	WL	5·5	39°37′–43°28′N 109°16′–114°49′E	865–2118	http://www.wulanchabu.gov.cn/channel/wlcb/col6722f.html
Huhehaote	HH	1·7	39°35′–40°51′N 110°46′–112°10′E	986–2280	http://www.huhhot.gov.cn/hhht/index.asp
Baotou	BT	2·8	41°20′–42°40′N 109°50′–111°25′E	1000–1500	http://www.baotou.gov.cn/html/btgl/dlqh.html
Eerduosi	EE	8·7	37°35′–40°51′N 106°42′–111°27′E	1000–1500	http://www.ordos.gov.cn/zjeedx/index.html
Bayannaoer	BY	6·4	40°13′–42°28′N 105°12′–109°59′E	1000–2020	http://www.bynr.gov.cn/sqgk/
Alashanmeg	AL	27·0	37°21′–42°47′N 97°10′–106°52′E	900–1400	http://www.als.gov.cn/main/tour/

* Websites were all verified on 8 February 2013.

† Districts for Shaanxi are listed from North to South in Fig. 1, while those for Inner Mongolia are listed from East to West.

Table 2. *Summary statistics for variables**

Variable	Observed	Mean	S.D.	Min	Max
Yield (kg/ha)	136	5600	2027·4	1020	11 525
G (°C)	136	1465	376·9	772	2590
P_5 (mm)	136	25	20·1	0	122
P_6 (mm)	136	48	31·8	1	135
P_7 (mm)	136	94	51·6	4	216
P_8 (mm)	136	77	45·4	7	203
P_9 (mm)	136	338	23·2	1	147

* G: growing degree days; P: precipitation, with subscripts indicating the month (e.g. 5 = May).

d in region j ($T_{d,j,t}$); and $G_{j,t}(T_{j,t})$ is total growing degree days in region j during the growing season of 153 days (May through September) in year t.

Evidence suggests that when pollinated spikelets are exposed to temperatures above 36 °C, maize yields may be negatively impacted (Dupuis & Dumas 1990). In the current study, average temperatures were used. These rarely exceeded 32 °C, although this did not preclude some periods of maximum temperatures exceeding 36 °C. From a statistical standpoint, therefore, the possibility that temperatures might be too high is assumed to be captured by the district dummy variables (as discussed earlier). For the response variables, the distribution of unadjusted yields was found to be closer to a normal distribution than either the distributions of the logarithm of yields or the square root of yields. Summary statistics for the variables used in the model are provided in Table 2. Correlations among monthly growing degree days were strong, as shown in Table 3; thus, to avoid multicollinearity problems, growing degree days accumulated over the entire growing season were employed in the regressions rather than separate monthly values. This was not the case for precipitation (Table 4), so monthly precipitation values were employed.

RESULTS

Estimation results are provided for the linear, quadratic and the four higher order terms for seasonal growing degree days (G); these are represented by Eqns (1)–(6), respectively (Table 5). In each of the models, dummy variables were used to capture district fixed effects. With some exceptions, unexplained differences among districts were found to be important in explaining differences in maize yields across the study region.

Table 3. *Correlations of monthly growing degree days (*G*)*

	G_5	G_6	G_7	G_8	G_9
G_5*	1				
G_6	0·8666	1			
G_7	0·8505	0·9280	1		
G_8	0·7434	0·8126	0·8531	1	
G_9	0·5799	0·6514	0·6621	0·7076	1

* Subscripts refer to the month (e.g. 5 = May).

Table 4. *Correlations of monthly precipitation (*P*)**

	P_5	P_6	P_7	P_8	P_9
P_5	1				
P_6	0·3736	1			
P_7	0·3030	0·5463	1		
P_8	0·0959	0·2338	0·4556	1	
P_9	0·0929	0·3070	0·3258	0·1707	1

* Subscripts denote the month in the growing season (e.g. 5 = May).

For Eqns (1), (2) and (3) (models with linear, quadratic and degree − 1 terms of G), the estimated parameters were not statistically significant. As the degree on the growing degree days variable increased from 1 to 4 for Eqns (3)–(6), the model fit improved as indicated by $\bar{R}^2$, while the level of statistical significance of the deviance difference statistic equalled 0·01 for Eqn (3) and 0·01 and 0·10 for Eqns (4) and (5), respectively; however, for Eqn (3), the estimated parameters on G were not statistically significant. For Eqn (6), the deviance difference was not statistically significant, which indicated that the functional form led to an 'over fitting' of the model; this was also indicated by the lack of statistical significance on any of the estimated parameters on G.

Based on deviance differences that are statistically significant at $P \leqslant 0{\cdot}1$, the best models are, therefore, Eqns (4) and (5). These explain 80·3 and 82·3% of the variation in maize yields, respectively. A modification to the Akaike information criterion (AIC), namely, the quasi-likelihood independence model criterion (QIC), was used to test the goodness-of-fit of the generalized estimation equations represented by the two models (Pan 2001). The QIC criterion corresponding to Eqn (5) is smaller than that of Eqn (4), so that, from a statistical standpoint, the functional form of Eqn (5) provides the best fit of the data.

Table 5. *Fractional polynomial regression in different degrees†*

Variable	1 Linear	2 Quadratic	3 Degree 1	4 Degree 2	5 Degree 3	6 Degree 4
$G^{-0·5}$				60 409***		
				(20442·0)		
$G^{-0·5} \times (\ln G)$				33 288***		
				(11040·5)		
G	−2	3288				
	(1·3)	*(3880·1)*				
G^2		−1494				
		(1089·2)				
G^3			−193		6040***	−1974
			(119·7)		*(1767·0)*	*(5795·9)*
$G^3 \times \ln G$					−12 612***	10 198
					(3304·7)	*(16097·9)*
$G^3(\ln G)^2$					6790***	−18 603
					(1704·3)	*(17822·5)*
$G^3(\ln G)^3$						10 234
						(7249·5)
P^5	342	637	571	389	−234	−247
	(1241·7)	*(1175·8)*	*(1199·3)*	*(1146·7)*	*(1094·2)*	*(1078·3)*
P_5^2	106	−90	−54	109	705	776
	(1310·6)	*(1241·0)*	*(1266·7)*	*(1210·0)*	*(1152·2)*	*(1135·9)*
P_6	481	402	485	2	−234	−133
	(1513·8)	*(1435·4)*	*(1463·2)*	*(1399·0)*	*(1315·1)*	*(1289·5)*
P_6^2	−65	27	−34	333	446	338
	(1059·9)	*(1006·0))*	*(1023·1)*	*(981·9)*	*(921·6)*	*(903·7)*
P_7	1420	2007**	1884**	1666*	1424	1857**
	(1008·9)	*(962·2)*	*(923·3)*	*(937·2)*	*(890·0)*	*(875·4)*
P_7^2	−423	−632	−591	−494	−406	−581
	(409·4)	*(390·3)*	*(374·5)*	*(380·7)*	*(361·6)*	*(356·8)*
P_8	1464*	1517*	1525**	1732**	1736**	1594**
	(809·9)	*(774·4)*	*(757·2)*	*(757·0)*	*(711·2)*	*(697·6)*
P_8^2	−669*	−702*	−697*	−838**	−851**	−767**
	(380·8)	*(362·8)*	*(360·7)*	*(355·2)*	*(333·3)*	*(327·4)*
P_9	−2315**	−2285**	−2314**	−2144**	−1986**	−2053**
	(1141·1)	*(1080·5)*	*(1098·6)*	*(1052·7)*	*(989·9)*	*(972·4)*
P_9^2	935	954	969	817	729	808
	(932·3)	*(882·8)*	*(898·2)*	*(860·2)*	*(808·7)*	*(794·7)*
Time	2601***	2442***	2438***	2543***	2681***	2546***
	(675·0)	*(650·5)*	*(543·5)*	*(631·8)*	*(595·4)*	*(583·7)*
D_2	−1781***	−1646***	−1656***	−1763***	−1895***	−1836***
	(454·9)	*(431·2)*	*(426·9)*	*(420·6)*	*(396·8)*	*(389·0)*
D_3	2154**	2867**	2595**	2738***	3642***	4910***
	(990·2)	*(1107·2)*	*(1117·9)*	*(928·4)*	*(1042·9)*	*(1364·9)*
D_4	3104**	4327***	4066***	4166***	4546***	5083***
	(1246·1)	*(1297·5)*	*(808·2)*	*(1247·4)*	*(1199·9)*	*(1185·0)*
D_5	693	1016	990*	711	347	541
	(648·5)	*(618·3)*	*(556·4)*	*(604·7)*	*(581·4)*	*(570·2)*
D_6	−834	−649	−640	−904	−1117*	−927
	(644·6)	*(614·8)*	*(540·1)*	*(604·8)*	*(578·2)*	*(569·3)*
D_7	398	575	581	395	138	244
	(577·5)	*(550·5)*	*(489·1)*	*(538·7)*	*(516·4)*	*(506·3)*
D_8	−4074***	−2762**	−3103***	−2218*	−1440	−1307
	(1236·7)	*(1353·4)*	*(702·4)*	*(1325·5)*	*(1266·0)*	*(1241·7)*
D_9	−12	50	71	−69	−253	−234
	(509·5)	*(483·1)*	*(465·3)*	*(471·6)*	*(446·8)*	*(438·8)*
D_{10}	−2173**	−1044	−1292**	−1277	−946	−456
	(984·3)	*(1002·7)*	*(650·3)*	*(964·0)*	*(930·0)*	*(932·4)*

Table 5. (*Cont.*)

Variable	1 Linear	2 Quadratic	3 Degree 1	4 Degree 2	5 Degree 3	6 Degree 4
D_{11}	−2518***	−2155***	−2167***	−2518***	−2914***	−2703***
	(*759·6*)	(*726·4*)	(*639·2*)	(*710·7*)	(*684·1*)	(*670·8*)
D_{12}	−33	325	294	77	−150	67
	(*708·7*)	(*682·7*)	(*566·2*)	(*666·0*)	(*638·1*)	(*627·6*)
Cons	6393***	6213***	6204***	6439***	6757***	6663***
	(*526·7*)	(*501·4*)	(*437·1*)	(*491·8*)	(*475·2*)	(*466·1*)
D.F.	112	111	112	111	110	109
$\bar{R}^2$	0·776	0·795	0·788	0·803	0·824	0·832
Res. S.D.			918·7	875·9	823·5	807·4
Deviance	2224·6	2208·5	2215·4	2201·2	2183·2	2176·6
Dev. dif.			42·7	28·5	10·5	3·9
Prob.			0·000	0·001	0·074	0·221

* $P<0·1$, **$P<0·05$, ***$P<0·01$.

† *G*: growing degree days; *P*: precipitation, with subscripts indicating the month (e.g. 5 = May). *D*: district dummy variable, with subscripts indicating each of the 11 districts out of 12. Time is the variable representing technological change. *G*, *P* and time are standardized in the regressions. Coefficients and standard errors (in parentheses) are listed for each variable.

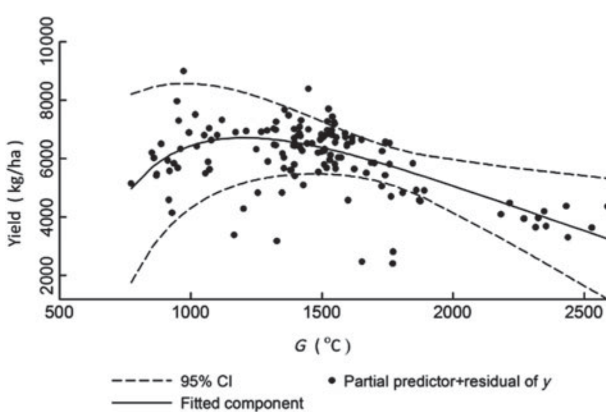

Fig. 2. Fractional polynomial of *G*, of degree 2 with powers (−0·5, −0·5).

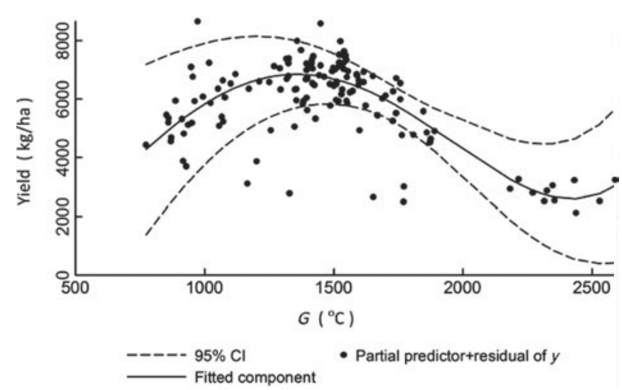

Fig. 3. Fractional polynomial of *G*, of degree 3 with powers (3, 3, 3).

For Eqn (4), fractional polynomial terms for growing degree days consisted of combinations of $1/\sqrt{G}$ and $\left(\frac{1}{\sqrt{G}} \times \ln G\right)$, with the relationship shown in Fig. 2. For Eqn (5), fractional polynomial terms of growing degree days consisted of G^3, $G^3 \times \ln G$, and $G^3 \times (\ln G)^2$, with this estimated function plotted in Fig. 3. As indicated in the figure, growing degree days were found to be an essential determinant of maize yield, but other factors not captured in the figure also came into play, and too much heat caused yields to decline, all other things being equal.

In Eqns (4) and (5), higher levels of precipitation in July (Eqn 4) and August (Eqns 4 and 5) have a positive effect on maize yields, but too much precipitation in any given month would reduce yields as indicated by the negative coefficient on the rainfall squared term (although it is insignificant for July precipitation). Precipitation in September negatively affects crop yields, probably because this is the harvest period; additional rainfall is no longer needed for crop growth and, indeed, rainfall could disrupt harvesting operations causing some crop loss or cause maize yields to decline as precipitation might damage the crop. As expected, the time variable has a strong positive impact on maize yields. This indicates that farmers were adopting new technologies, whether these were improved varieties of maize, more fertilizer, better or newer equipment, or some other improvement.

https://doi.org/10.1017/S0021859613000142 Published online by Cambridge University Press

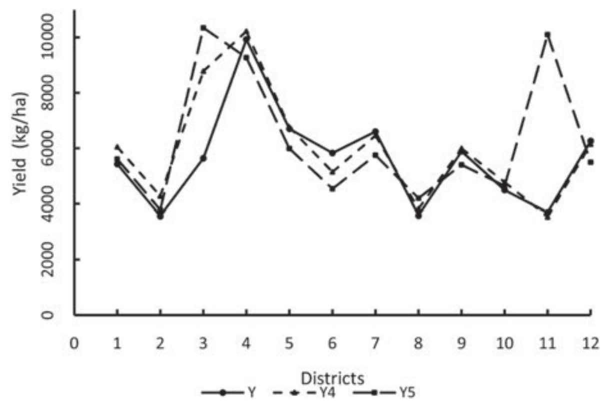

Fig. 4. Actual average (Y) and Monte Carlo simulated yields for Eqn (4) (Y_4) and (5) (Y_5) for the 12 districts.

As a final check on the model, the estimated parameters for Eqns (4) and (5) were employed in a Monte Carlo simulation (with sampling from distributions about the estimated parameters using the estimated standard errors as well as the overall standard error of the estimated model) to determine average maize yield for each of the 12 districts. These are provided in Fig. 4. In the figure, estimated yields (Y_4) derived from Eqn (4) are close to actual yields (Y), with the exception of district 3, and estimated yields (Y_5) from Eqn (5) are close to actuarial yields (Y), except for districts 3 and 11. Overall, in contrast to the QIC criterion, Eqn (4) appears to better predict maize yields than Eqn (5) and thus also the other models.

DISCUSSION

In the present study, the impact of climate variables on maize yields in northern China was investigated. The most important result was that accumulated heat throughout the growing season (as measured by seasonal growing degree days) was probably the most important weather variable influencing maize yields. However, the relationship between growing degree days and maize yields was subtle and could not be captured adequately by a linear or quadratic functional relation. Rather, the relationship was much more complicated, requiring the use of a flexible functional form, and had to be estimated as a highly nonlinear regression model.

Not surprisingly, precipitation was also important but added to crop yields primarily during the peak of the growing season, indicating that it was of less importance than heat units, i.e. moisture was important for the study region, but there was probably enough soil moisture that rainfall in mid-summer simply provided a boost to yields that declined rapidly with higher levels of rainfall. Given the size of the estimated parameters, district fixed effects and adopted technical advances were also important factors explaining crop yields.

Finally, the two best-fitted models, Eqns (4) and (5) captured 80% or more of the variation in maize yields. In that case, the estimated regression models could potentially be used as a basis for developing weather-indexed insurance products in this study area. For example, farmers in western and central China have expressed interest in weather-indexed insurance to mitigate weather risks (Turvey *et al.* 2009; Liu *et al.* 2010); therefore, the current work could be extended to develop weather-indexed insurance using the relationship between growing degree days and yields to establish actuarially sound premiums. The model can also be used to predict potential maize yields under future climate scenarios, providing some indication of the potential damage from global warming; however, such scenarios would neither take into account adaptation by landowners (e.g. planting different crops) nor agricultural infrastructure improvements made by government (e.g. irrigation works) and technological changes that lead to greater adaptation to climate changes in the agricultural sector. Further research is needed to examine the effect of climate factors on crop yields, and how this information can be used to develop weather-indexed crop insurance and other financial weather derivatives and strategies for adapting to climate change in the agricultural sector.

The authors are grateful for research support from the LEARN-ERCA network of Agriculture and Agri-Food Canada and the Canadian Social Sciences and Humanities Research Council.

REFERENCES

Almaraz, J. J., Mabood, F., Zhou, X., Gregorich, E. G. & Smith, D. L. (2008). Climate change, weather variability and corn yield at a higher latitude locale: Southwestern Quebec. *Climate Change* **88**, 187–197.

Bai, C. Y., Li, S. K., Bai, J. H., Zhang, H. B. & Xie, R. Z. (2011). Characteristics of accumulated temperature demand and its utilization of maize under different ecological conditions in Northeast China. *Chinese Journal of Applied Ecology* **22**, 2337–2342.

Bunting, E. S. (1979). The relationship between mean temperature and accumulated temperature totals for maize in the central lowlands of England. *Journal of Agricultural Science, Cambridge* **93**, 157–169.

CHEN, C., LEI, C., DENG, A., QIAN, C., HOOGMOED, W. & ZHANG, W. (2011). Will higher minimum temperature increase corn production in northeast China? An analysis of historical data over 1965–2008. *Agricultural and Forestry Meteorology* **151**, 1580–1588.

Data Sharing Infrastructure of Earth System Science (1977–96). *China 1:4,000,000 Basic Data*. Beijing: Institute of Geographic Sciences and Natural Resources Research, CAS.

DUPUIS, I. D. & DUMAS, C. (1990). Influence of temperature stress on *in vitro* fertilization and heat shock protein synthesis in maize (*Zea mays* L.) reproductive tissues. *Plant Physiology* **94**, 665–670.

FAO (2010). *Food and Agricultural Commodities Production*. Rome: FAO. Available online at: http://faostat.fao.org/site/339/default.aspx (verified 20 June 2012).

Government of Inner Mongolia (2010). *Overview*. Inner Mongolia, China. Available online at: http://intonmg.nmg.gov.cn/channel/zjnmg/col6722f.html (verified 20 June 2012).

Government of Yanan (2012). *Overview*. Yanan, China. Available online at: http://www.yanan.gov.cn/structure/zjya/zjyax (verified 28 October 2012).

Government of Yulin (2012). *Overview*. Yulin, China. Available online at: http://www.yl.gov.cn/site/1/html/zjyl/list/list_18.htm (verified 28 October 2012).

HUANG, J., ROZELLE, S., MARTIN, W. & LIU, Y. (2009). China. In *Distortions to Agricultural Incentives in Asia* (Eds K. Anderson & W. Martin), pp. 117–161. Washington, DC: The International Bank of Reconstruction and Development/World Bank.

Inner Mongolia Bureau of Statistics (1990–2000). *Inner Mongolia Statistical Yearbooks*. Beijing: China Statistics Press.

LI, X., TAKAHASHI, T., SUZUKI, N. & KAISER, H. M. (2011). The impact of climate change on maize yields in the United States and China. *Agricultural Systems* **104**, 348–353.

LIN, J. Y. (1988). The household responsibility system in China's agriculture reform: a theoretical and empirical study. *Economic Development and Cultural Change* **36**, S199–S224.

LIN, J. Y. (1992). Rural reforms and agricultural growth in China. *American Economic Review* **82**, 34–51.

LIU, B. C., LI, M. S., GUO, Y. & SHAN, K. (2010). Analysis of the demand for weather index agricultural insurance on household level in Anhui, China. *Agriculture and Agricultural Science Procedia* **1**, 179–186.

National Bureau of Statistics of China (2011). *China Statistical Yearbook*. Beijing: China Statistical Press.

PAN, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.

ROYSTON, P. & ALTMAN, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society Series C: Applied Statistics* **43**, 429–467.

SCHLENKER, W. & ROBERTS, M. J. (2006). Nonlinear effects of weather on corn yields. *Applied Economic Perspectives and Policy* **28**, 391–398.

SCHLENKER, W. & ROBERTS, M. J. (2008). *Estimating the Impact of Climate Change on Crop Yields: The Importance of Non-linear Temperature Effects*. NBER Working Paper 13799. Cambridge, MA, USA: National Bureau of Economic Research. Available online at: http://www.nber.org/papers/w13799 (verified 2 May 2012).

Shaanxi Bureau of Statistics (1990–2002). *Shaanxi Statistical Yearbooks*. Beijing: China Statistics Press.

TURVEY, C. G., KONG, R. & BELLTAWN, B. C. (2009). Weather risk and the viability of weather insurance in western China. In *Annual Meeting of the American Agricultural Economics Association, 26–28 July 2009, Milwaukee, WI* (Ed. AAEA), pp. 1–33. Milwaukee, WI: American Agricultural Economics Association. Available online at: http://purl.umn.edu/49362 (verified 14 February 2013).

VUČETIĆ, V. (2011). Modelling of maize production in Croatia: present and future climate. *Journal of Agricultural Science, Cambridge* **149**, 145–157.