

CLIMATE CHANGE AND AGRICULTURE RESEARCH PAPER

Maize yield forecasting by linear regression and artificial neural networks in Jilin, China

K. MATSUMURA^{1,2*}, C. F. GAITAN², K. SUGIMOTO³, A. J. CANNON⁴ AND W. W. HSIEH²

¹ *Department of Applied Informatics, School of Policy Studies, Kwansai Gakuin University, Hyogo 669-1337, Japan*

² *Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, 2020-2207 Main Mall, Vancouver, BC V6T 1Z4, Canada*

³ *Graduate School of Environmental Studies, Nagoya University, Furocho D2-1, Chigusa-ku, Nagoya 464-8603, Japan*

⁴ *Pacific Climate Impacts Consortium, University of Victoria, PO Box 3060 Stn CSC, Victoria, BC V8W 3R4, Canada*

(Received 6 March 2013; revised 14 January 2014; accepted 7 April 2014; first published online 20 May 2014)

SUMMARY

Forecasting the maize yield of China's Jilin province from 1962 to 2004, with climate conditions and fertilizer as predictors, was investigated using multiple linear regression (MLR) and non-linear artificial neural network (ANN) models. Yield was set to be a function of precipitation from July to August, precipitation in September and the amount of fertilizer used. Fertilizer emerged as the dominant predictor and was non-linearly related to yield in the ANN model. Given the difficulty of acquiring fertilizer data for maize, the current study was also tested using the previous year's yield in the place of fertilizer data. Forecast skill scores computed under both cross-validation and retroactive validation showed ANN models to significantly outperform MLR and persistence (i.e. forecast yield is identical to last year's observed yield). As the data were non-stationary, cross-validation was found to be less reliable than retroactive validation in assessing the forecast skill.

INTRODUCTION

Changing climate conditions on seasonal and longer time scales influence agricultural production. Inter-annual variations in the Pacific sea surface temperatures (e.g. from the El Niño-Southern Oscillation and the North Pacific Gyre Oscillation) influence Canadian wheat yield (Hsieh *et al.* 1999). Lobell (2007) used the change in the daily temperature range (i.e. difference between the daily maximum and minimum temperatures) to estimate crop yield under climate change. Several rice and maize crops show a negative response in yield to increased daily temperature range, reflecting a non-linear response of yields to temperature. Lobell & Field (2007) suggested that simple measures of growing season temperature and precipitation explained up to 30% or more of year-to-year variations in global average yields for the world's six most widely grown crops. Improvements of soil and fertilizer are strong influencing factors on agricultural

production; however, agricultural production is influenced by climate conditions even in highly developed countries.

Crop yields are influenced by many factors, the most obvious being meteorological (Fortin *et al.* 2011) and soil conditions (Campbell *et al.* 1988, 1997a, b; Alvarez 2009). Topography (elevation, slope and aspect) has significant effects on wheat yield (Yang *et al.* 1998). Social-economic factors were also taken into account in Ghodsi *et al.* (2012), where rainfall, guaranteed purchasing price, area under cultivation, subsidy, insured area, inventory, import, population and value-added agricultural production were used as predictors for wheat production.

Researchers have used two main types of models – mechanistic (i.e. process-oriented) models and empirical models – to study crop growth/yield with respect to environmental variables. Mechanistic models, based on eco-physiological processes for crop growth, are best for understanding the complicated relations between crop, soil, climate and ecology. However, these models generally contain parameters

* To whom all correspondence should be addressed. Email: kanichi1@mbox.kyoto-inet.or.jp



Fig. 1. China's Jilin province, with temperature and precipitation data obtained near the capital Changchun (43°53'N, 125°19'E, 212 m a.s.l.) (map produced from ESRI Data and Maps, ESRI 2008).

that depend on the particular plant species and/or environment, thereby requiring careful calibration. Examples of mechanistic models for maize include Yang *et al.* (2004), Liu *et al.* (2011) and Ma *et al.* (2012). Compared to empirical models, the mechanistic models are more complicated to develop and, although built on an eco-physiological process-based framework, they are not particularly well suited for making accurate regional-scale crop yield forecasts.

Empirical models are based on statistical methods or machine learning methods developed from the field of artificial intelligence. While they are generally inferior to mechanistic models in terms of revealing the underlying eco-physiological processes, they are usually easier to develop and are well suited for making regional-scale crop yield forecasts.

Among empirical models, multiple linear regression (MLR) models are widely used for analysing climate and soil conditions and their impact on crop yield (Qian *et al.* 2009; Lobell & Burke 2010). However, MLR lacks the general ability to model the non-linear relationship between predictor variables and the response variable. The rise of artificial intelligence has introduced non-linear empirical models such as artificial neural networks (ANNs) (Rumelhart *et al.* 1986; Hsieh 2009), which has begun to complement

or replace MLR models in agricultural forecasting during the last decade or so. In general, ANN has been found to give better predictions than MLR (Hill *et al.* 2002; Kaul *et al.* 2005; Ji *et al.* 2007; Dai *et al.* 2011; Guo & Xue 2012; Chantre *et al.* 2014).

Maize is now the largest grain crop (in metric tonnes) produced in the world according to the Food and Agriculture Organization of the United Nations (FAOSTAT 2012). Previous studies relating climate and maize yield involved using MLR (Lobell & Field 2007; Lobell & Burke 2010; Chen *et al.* 2011) regression with a fractional polynomial model (Sun & Van Kooten 2013) and ANN (O'Neal *et al.* 2002; Kaul *et al.* 2005).

The purpose of the current study was to analyse how maize yield relates to climate conditions and fertilizer used and to improve short-range maize yield forecasting by MLR and ANN models for Jilin province, located in the north-eastern part of China (Fig. 1). With its cold harsh climate, Jilin is limited to being a monocropping agricultural region, i.e. farmers must rely on a single crop, hence poor crop production is particularly disastrous. Maize yield was modelled in the current study as a function of the average monthly temperature from April to October, accumulated precipitation from July to August, precipitation in September and the

amount of fertilizer used. The previous year's maize yield was also tested as an alternative predictor to fertilizer.

DATA

Annual maize yield for Jilin province (kg/ha) from 1961 to 2004 was obtained from China Maize (2012). Monthly temperature and precipitation data at $0.5^\circ \times 0.5^\circ$ resolution were obtained from the CRU (2010) TS3.10 dataset, produced by the Climatic Research Unit, University of East Anglia, with the grid point closest to the provincial capital Changchun used in the current study. Country-based fertilizer consumption data were available from the International Fertilizer Industry Association (IFA 2012). Annual total nitrogen, phosphate and potassium fertilizer consumption for China was obtained from IFA. The amount of fertilizer used for the maize crop was obtained by multiplying the total fertilizer used by the ratio of maize harvested area in China to major crop harvested area in China (FAOSTAT 2012). Fertilizer as a predictor used in the current study was obtained by dividing the amount of fertilizer used for maize by the maize harvested area. This allowed the application of country-based total fertilizer values to smaller, more specific crop regions. Owing to the limitation of the available maize yield data, the data analysis was confined to the period 1961–2004.

Four predictors were considered: the average monthly temperature over the growing season (April–October), accumulated precipitation from July to August, precipitation in September and the amount of fertilizer used. The choice of the precipitation predictors came from considering the correlation between maize yield and precipitation data (both with linear trend removed first), where the correlation was 0.10 (May precipitation), -0.02 (June), 0.24 (July), 0.16 (August), 0.35 (September) and -0.17 (October), with only the September precipitation being significant ($P < 0.05$). However, the correlation between yield and the accumulated July and August precipitation was 0.31 ($P < 0.05$). Moisture was not a limiting factor early in the growing season, while precipitation late in the growing season could have different effects than precipitation during July and August (Sun & Van Kooten 2013), hence the choice of using the accumulated July and August precipitation and the precipitation in September as two separate predictors.

The mean and standard deviation are, respectively, 15.9 and 1.2 °C for the averaged April to October

temperature, 309 and 98 mm for the July to August precipitation, 48 and 26 mm for the September precipitation, 200 and 151 kg/ha for the amount of fertilizer and 4.13×10^3 and 2.13×10^3 kg/ha for the maize yield. The coefficient of variation, defined as the standard deviation divided by the mean, is 0.073 for the April–October temperature, 0.315 for the July–August precipitation, 0.547 for the September precipitation, 0.755 for the fertilizer amount and 0.516 for the maize yield. Figure 2 shows the standardized time series (i.e. all having zero mean and unit standard deviation). Fertilizer and maize yield both increased gradually with time. Temperature also showed a rising trend, while precipitation tended to be low in the final few years.

METHODS

MLR and non-linear regression by ANN are compared in the current study. In MLR, the response variable or predictand y is expressed as a linear function of the m predictor variables x_i ($i = 1, \dots, m$)

$$y = a_0 + \sum_i a_i x_i \quad (1)$$

where a_i ($i = 0, \dots, m$) are the regression coefficients determined by fitting the straight-line solution to the data, i.e. minimizing the mean squared error (MSE) between the modelled y estimate and the observed y value. Since standardizing predictors are crucial for ANN models (Hsieh 2009), all predictors were standardized for the MLR and ANN models. The ANN model used in the current study is the multi-layer perceptron (Rumelhart *et al.* 1986; Hsieh 2004, 2009), with one hidden layer. Here y is expressed as a non-linear function of the form

$$y = \sum_j \tilde{w}_j \tanh\left(\sum_i w_{ji} x_i + b_j\right) + \tilde{b} \quad (2)$$

where \tanh denotes the hyperbolic tangent function, and the parameters or weights, \tilde{w}_j , \tilde{b} , w_{ji} and b_j , are determined by fitting the non-linear function to the data, where a non-linear optimization algorithm is used to minimize the MSE between the modelled y and the observed y values. This empirical function (2) is capable of fitting to any non-linear continuous function to arbitrary accuracy if enough terms are used in the summation over the index j . The number of terms in the summation, HN, is called the number of 'hidden neurons'.

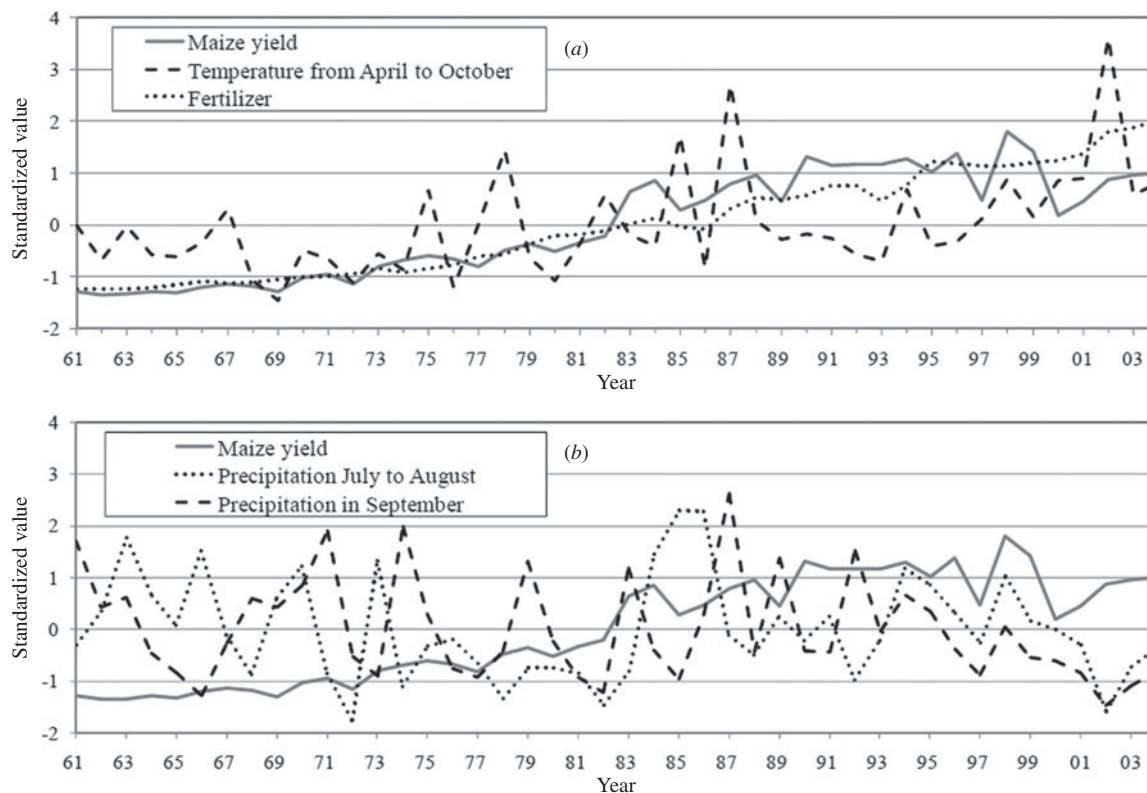


Fig. 2. Standardized annual time series (1961–2004) used in the current study: (a) maize yield of Jilin province (solid), average temperature from April to October (dashed) and fertilizer use (dotted) and (b) precipitation from July to August (dotted) and precipitation in September (dashed).

As there are generally multiple minima in a non-linear optimization problem, the optimization is commonly performed a number of times starting from different random initial weights, yielding an ensemble of ANN models. The final predicted value for y from the ensemble is simply the average of all the individual model predictions for y .

With a flexible non-linear model, it is easy to over-fit the data during model training, i.e. fitting to the noise in the data. An over-fitted model may fit the training data very well, but predicts poorly when given new data. A Bayesian regularization approach has been successfully used in ANN models to avoid over-fitting (MacKay 1992; Hsieh 2009). In the current study, the code `trainbr` from the neural network toolbox for MATLAB was used, which was based on the Bayesian ANN model of Foresee & Hagan (1997).

Besides MLR and ANN, the persistence model was also used as a benchmark or reference model. In the persistence model, the predicted value for y at time t is simply the observed y value of the previous year ($t - 1$). Data from 1961 to 2004 were used in our study, giving predictions for the maize yield y from 1962 to 2004.

To test a model's forecast skill, the traditional validation methodology (O'Neal *et al.* 2002; Kaul *et al.* 2005) used part of the data record for model training and the remaining part for validation (i.e. for testing the model forecast performance). Here the cross-validation procedure (Hsieh 2009), which uses the entire data record for validation, was used instead, since in recent years, cross-validation has increasingly become the new standard procedure for validation, especially when the data record is not long. The current data record was divided into four approximately equal segments. One segment was withheld for forecast validation, while data from the other three segments were used for training the model, i.e. the data used in forecast validation were never used in training the model. Next, a different segment was withheld for validation and the three remaining segments used for training. This process was repeated until all four segments of the data record had been used for forecast validation, i.e. the root mean squared error (RMSE) between the model forecasted maize yield and the observed yield over the whole data record was calculated by this cross-validation procedure.

Table 1. Predictors used in the multiple linear regression (MLR) and artificial neural network (ANN) models. Precipitation from July to August is denoted by $prcp7+8$ and precipitation in September by $prcp9$, while $t-1$ denotes the previous year

Model	$prcp7+8$	$prcp9$	$prcp7+8(t-1)$	$prcp9(t-1)$	Fertilizer	Yield($t-1$)
MLR1	+	+			+	
MLR2	+	+				+
ANN1	+	+			+	
ANN2	+	+				+
ANN3			+	+		+
ANN4	+	+	+	+		+

The forecast validation was assessed using the RMSE skill score (SS) with the persistence model serving as the reference model, i.e.

$$RMSESS = 1 - RMSE/RMSE_{\text{persist}} \quad (3)$$

where $RMSE_{\text{persist}}$ is the RMSE of the persistence model forecasts. The RMSE SS is 1 for perfect model forecasts and is negative when the model forecasts are worse than those from the persistence model.

RESULTS

A preliminary test was performed using an MLR model of maize yield with four predictors—the averaged monthly temperature over the growing season (April–October), accumulated precipitation from July to August, precipitation in September and the amount of fertilizer used.

Owing to cross-validation dividing the data record into four segments, the forecast validation was done separately over four data subsets (1962–1972, 1973–1983, 1984–1994 and 1995–2004). Since the P value (the significance probability for rejecting the null hypothesis) for temperature was above 0.3 in each data subset, temperature was rejected as a predictor.

Table 1 lists the predictors used in the subsequent MLR and ANN models. The model MLR1 tested maize yield as a function of precipitation from July to August ($prcp7+8$), precipitation in September ($prcp9$) and fertilizer. The regression coefficients and the P values in Table 2 show fertilizer to be a much more important predictor than the two precipitation inputs. While the Pearson correlation between the modelled and observed yield appeared relatively high in Table 2, the RMSE SS for MLR1 was actually -0.877 , indicating the MLR to be under-performing the persistence model.

To make real-time maize yield forecasts, the predictors need to be available in real time.

Table 2. Regression coefficients (with corresponding P -values underneath) for the various predictors in model MLR1, during each of the four validation periods, with the correlation between the observed and the model values presented in the rightmost column. Note the listed years are the validation years, with the model trained using data from the non-validation years

	$prcp7+8$	$prcp9$	Fertilizer	Correlation
1962–1972	0.068	0.052	1.314	0.760
P	NS	NS	<0.001	
1973–1983	0.053	0.164	0.861	0.935
P	NS	<0.05	<0.001	
1984–1994	0.186	0.208	0.971	0.910
P	<0.05	<0.05	<0.001	
1995–2004	0.202	0.200	0.764	0.855
P	<0.05	<0.05	<0.001	

NS, not significant.

Unfortunately, fertilizer data are only available years later. As the correlation between yield and previous year's yield was 0.912, the possibility of using previous year's yield ($yield(t-1)$) as a predictor was investigated. Next, the model MLR2 which was identical to MLR1 except that the fertilizer predictor had been replaced by the previous year's yield (Table 1) was run. The RMSE SS improved from -0.877 (MLR1) to -0.097 (MLR2) when yield ($t-1$) replaced fertilizer as predictor. This means the MLR2 forecast performance was almost as good as the persistence model, which was to be expected since the very small regression coefficients and relatively large P values found for the precipitation predictors (Table 3) indicated that the MLR was essentially relying only on yield ($t-1$).

To see the difference between non-linear regression and linear regression, the ANN1 model was built with

Table 3. Regression coefficients and P-values in model MLR2 computed for the four validation periods, with the forecast correlation score given in the rightmost column

Years	Yield($t-1$)	prcp7+8	prcp9	Correlation
1962–1972	0.976	–0.028	0.033	0.954
<i>P</i>	<0.001	NS	NS	
1973–1983	0.900	0.011	0.048	0.902
<i>P</i>	<0.001	NS	NS	
1984–1994	0.917	0.005	0.035	0.913
<i>P</i>	<0.001	NS	NS	
1995–2004	0.771	0.017	0.063	0.823
<i>P</i>	<0.001	NS	NS	

NS, not significant.

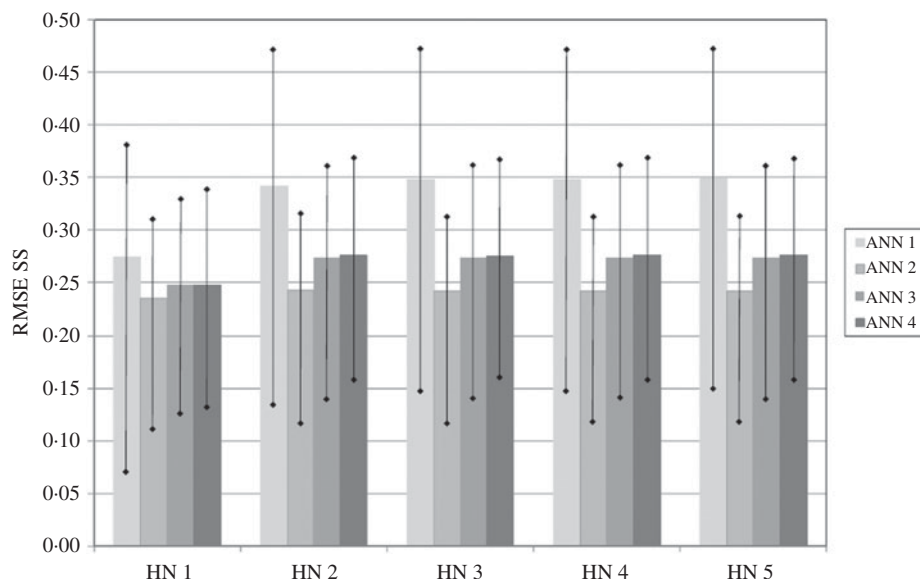


Fig. 3. The root mean squared error skill score of the four artificial neural network (ANN) models as the number of hidden neurons (HN) varied from 1 to 5 (HN1–HN5), with the 95% confidence intervals obtained from bootstrapping.

exactly the same predictors as MLR1, i.e. prcp7+8, prcp9 and fertilizer. The number of ensemble members was varied up to 300:200 was found to be enough for very stable results. The number of HN was varied from one to five in Fig. 3, showing a maximum RMSE SS at HN=3, although the SS was almost constant for HN=2–5. Hence ANN1 used HN=3 and 200 ensemble members. The 95% confidence intervals for the RMSE SS in Fig. 3 were calculated using the bootstrap method of Gilleland (2010).

Figure 4(a) compares the cross-validated yield forecasts by MLR1 and ANN1, and the observed yield. The RMSE SS for ANN1 was 0.348, clearly outperforming persistence, and MLR1 (SS=–0.877).

The poor performance of MLR was due to its tendency to over-predict the yields in the early and latter parts of the data record, and under-predict the yields during the middle part. An explanation of this behaviour is given in the Discussion section.

Additional tests were made by varying the predictors used in the ANN (with HN=3 and the ensemble size kept at 200), and the RMSE SS were examined: test A using only fertilizer gave SS=0.308, test B using prcp7+8 and fertilizer gave SS=0.341, test C using prcp9 and fertilizer gave SS=0.304, compared with the original ANN1 using prcp7+8, prcp9 and fertilizer, giving SS=0.348. These results suggest that precipitation in July and August was more relevant

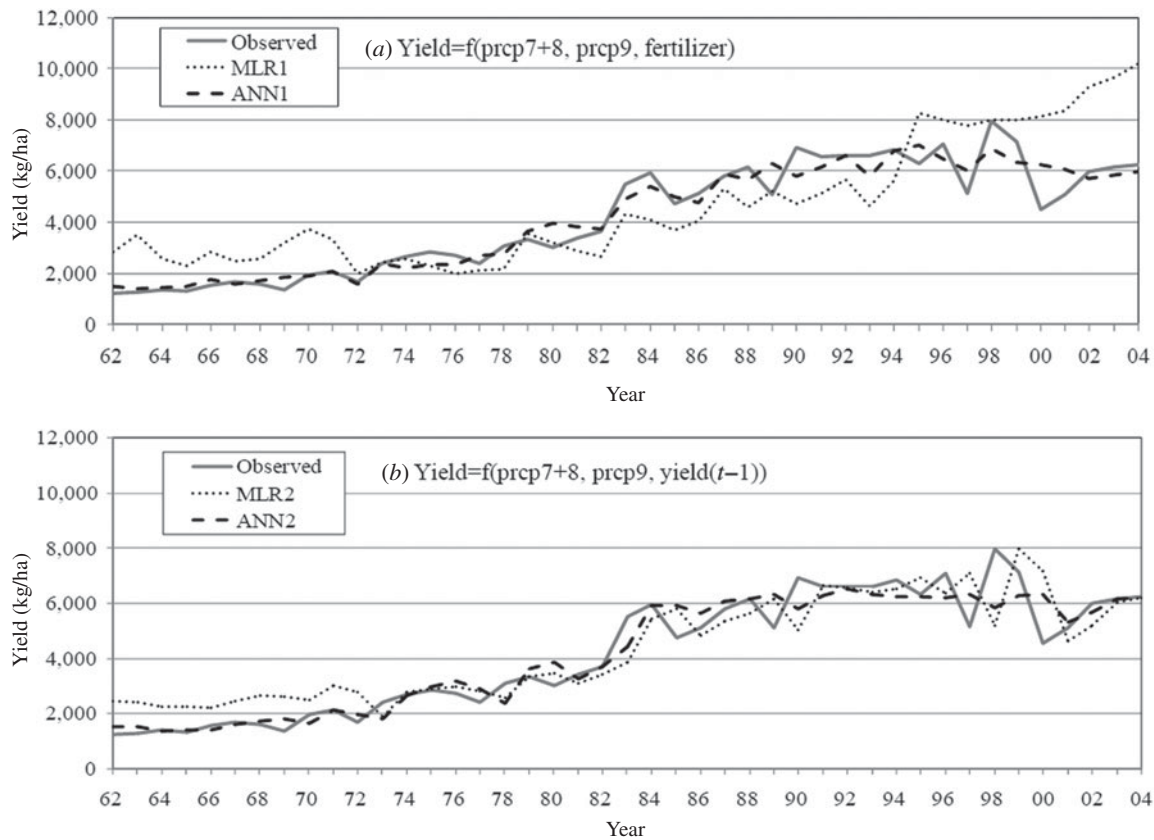


Fig. 4. Maize yield forecasts by multiple linear regression (MLR) and ANN v. observed yield (1962–2004), where in (a) the predictors used were prcp7 + 8, prcp9 and fertilizer and in (b) with fertilizer replaced by the previous year's yield.

than that in September. Additional tests using only precipitation as predictor(s) all gave negative SS.

In ANN2, the non-linear counterpart of MLR2, fertilizer was dropped as input, but $\text{yield}(t-1)$ and prcp7 + 8 and prcp9 were used as predictors (Table 1). Again with an ensemble of 200 members, the HN was varied from one to five (Fig. 3). The RMSE SS peaked at 0.243 (when HN=2) and stayed near this value as HN increased further to five. With HN=2, the cross-validated yield forecasts are compared with those from MLR2 and observed values in Fig. 4(b).

Using $\text{yield}(t-1)$ as a predictor has the advantage of indirectly supplying fertilizer information to the model since fertilizer usage tends not to vary greatly from 1 year to the next. However, $\text{yield}(t-1)$ is also affected by the precipitation in the previous year which is misleading for the current year's forecast. Hence two more ANN models (ANN3 and ANN4) were run to see if this undesirable effect could be alleviated. In ANN3, instead of using the precipitation of the current year as predictors, precipitation of the previous year was used (Table 1) to test whether the ANN model would be

capable of correcting for the irrelevant precipitation information imbedded in $\text{yield}(t-1)$. Finally in ANN4, precipitation information was supplied for both the current and the previous year, testing whether the ANN model would be capable of alleviating the effects of previous year's precipitation and incorporating current year's precipitation. Again the highest SSs were attained at HN=2 for both ANN3 and ANN4 (Fig. 3). Compared with the SS of 0.243 for ANN2, ANN3 had 0.274 and ANN4, 0.276, suggesting that the ANN model was capable of alleviating the effects of previous year's precipitation imbedded in $\text{yield}(t-1)$. Since the 95% confidence intervals for all the ANN models in Fig. 3 were above 0, all the ANN models significantly outperformed the reference model (i.e. persistence forecasts).

As the non-linear relations found by an ANN model are generally much harder to understand than the linear relations found by MLR, ANN has often been described as a 'black box' approach. Fortunately, for the current problem with relatively few predictors and only one response variable, some of the simpler non-

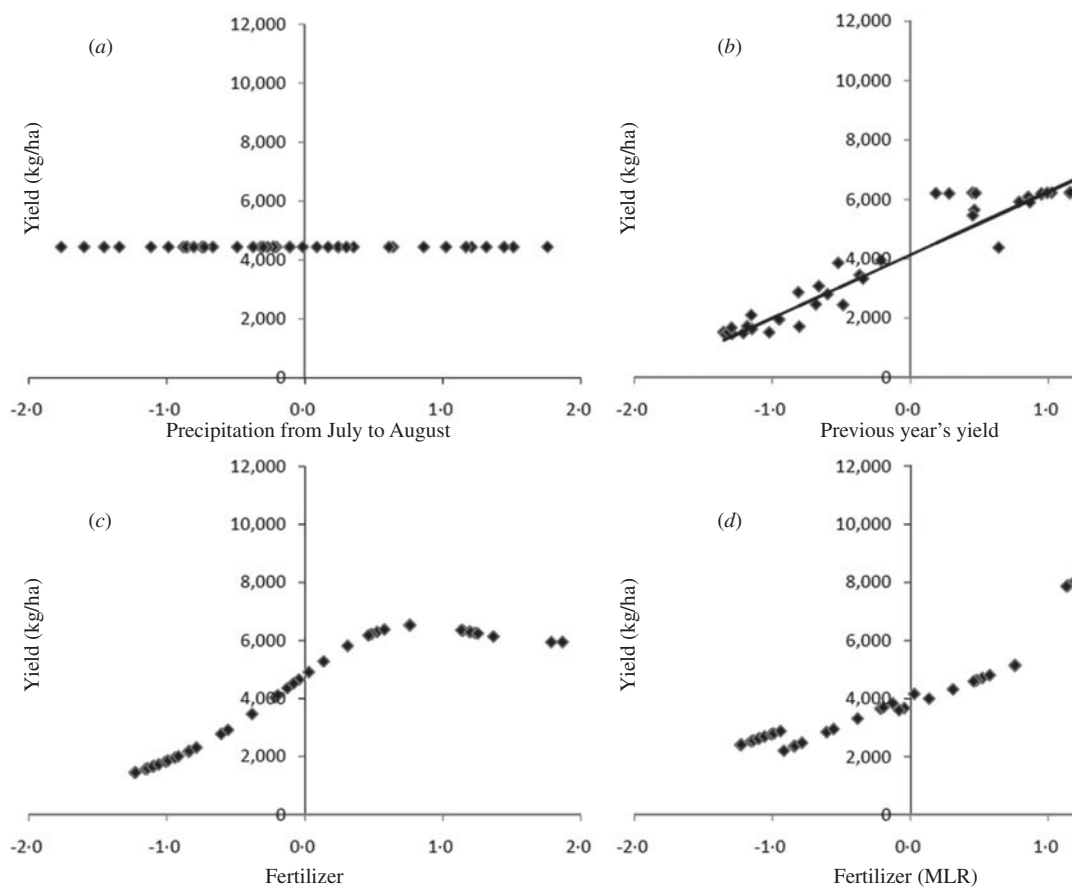


Fig. 5. Maize yield from ANN as a function of a single predictor (as other predictors are held constant at their mean values), with the varying predictor being (a) $prcp7+8$, (b) $yield(t-1)$ and (c) fertilizer. In (d), the MLR1 yield is shown as a function of fertilizer. ANN1 was used in (a) and (c) and ANN2 in (b). The straight line in (b) is when the current yield equals the previous year's yield.

linear relations can be visualized – in particular, how the yield varies as a function of a single predictor while the other predictors are held constant at their mean values. Of course, non-linear interactions involving two or more predictors cannot be visualized by this simple approach.

With ANN2, Fig. 5(a) shows the yield to be almost a constant function of $prcp7+8$ (as the other two predictors, $prcp9$ and previous yield, were held constant). Yield was also nearly a constant function of $prcp9$ (not shown), indicating very weak dependence of the ANN model on either $prcp7+8$ or $prcp9$ alone. In Fig. 5(b), the yield was roughly a linear function when the predictor $yield(t-1)$ was low and became nearly constant when $yield(t-1)$ was high. A similar functional shape was found in Fig. 5(c) with fertilizer as the sole varying predictor using ANN1, thus confirming that $yield(t-1)$ could roughly replace fertilizer as input. The very non-linear functional form of Figs 5(b) and (c) confirms that ANN and MLR gave

very different forecast results. In Fig. 5(d), the MLR1 with fertilizer as the sole varying predictor is shown. As the results in Fig. 5 were obtained by cross-validation, i.e. forecasts were made for the four data subsets separately, the MLR1 results in Fig. 5(d) are seen coming from four different straight-line solutions. Compared with Fig. 5(c), the MLR1 forecasted yield values were too high at the high end (and also too high at the low end) as noted before in Fig. 4(a).

While cross-validation is the standard procedure for assessing forecast skills for short data records, the reported skills may still be inflated compared with true out-of-sample forecast validation (Shabbar & Kharin 2007). To mimic a real-world forecasting setup, a retroactive validation was tested in which the first 15 years (1962–1976) were used to train the first model and then yield was predicted for 1977; the window was then extended forward by a year, i.e. the model was trained with the first 16 years (1962–1977), and yield was predicted for 1978; this

process was repeated until forecasts were made for 1977–2004. The four models ANN1–ANN4 were rerun under retroactive validation with $HN=1-5$ and ensemble size of 200. With 20 RMSE SSs computed (for the four models times five HN values), all 20 ANN SSs turned out positive, indicating better performance than persistence – in contrast, all the corresponding MLR SSs were negative. However, the retroactively validated ANN SS were lower than those from the cross-validated SS. For instance, for ANN4, the highest cross-validated SS was 0.276 for $HN=2$, but the highest retroactively validated SS was only 0.112 for $HN=1$. However, the cross-validated SS was calculated over the years 1962–2004, but the retroactively validated SS was over 1977–2004. The highest retroactively validated SS was 0.136 for ANN2 with $HN=1$, with the 95% bootstrap confidence interval being [0.045, 0.199], v. a SS of -0.092 for MLR, with a confidence interval of $[-0.258, 0.007]$. Hence, under retroactive validation, despite the lower SS, ANN continued to significantly outperform persistence and MLR.

DISCUSSION

Forecasting the maize yield of Jilin province, with climate conditions and amount of fertilizer as predictors, was investigated using MLR and ANN models. Datasets from 1962 to 2004 were divided into four segments with three used for model training and one for forecast validation, giving cross-validated forecast skills, where the RMSE SSs were in reference to persistence forecasts. Yield was set to be a function of precipitation from July to August ($prcp7+8$), precipitation in September ($prcp9$) and fertilizer used. The ANN1 SS ($=0.348$), being positive, indicates better performance than persistence, while the MLR1 SS ($= -0.877$), being negative, indicates worse performance than persistence. In ANN1, the dominant predictor was fertilizer and between the two precipitation predictors, $prcp9$ seemed to have less impact.

As fertilizer data are not available for real-time yield forecasting, models using previous year's yield, $yield(t-1)$, instead of fertilizer were tested, with ANN2 SS being 0.243 and MLR2 SS being -0.097 . When the two precipitation variables from the previous year were also added as predictors, the ANN SS increased slightly to 0.276 in ANN4.

When the forecasted yield was plotted as a function of a single varying predictor (with the other predictors held constant), the yield was clearly a very non-

linear function of the fertilizer and $yield(t-1)$, thus explaining why the ANN greatly outperformed the MLR. Future work is to build an operational maize yield forecast model (without using the weak $prcp9$ predictor), so useful forecasts can be issued to farmers/managers in September, well before the harvests in mid-October. Another potential improvement is to add satellite-measured vegetation indices as additional predictors in the ANN model (Jiang *et al.* 2004; Panda *et al.* 2010).

Besides real-time harvest yield forecasting, ANN models will also be valuable in future crop yield projection under climate change. Climate change is commonly studied by complicated global climate models (GCM) under various emission scenarios. Their coarse spatial resolution does not allow them to be used directly in projecting local crop yields in future climate conditions. Instead dynamical downscaling or statistical downscaling techniques are used to project from the coarse resolution GCM results to local climate change and crop yield change. In a comparison of downscaling techniques in the projection of local climate change and wheat yields in New South Wales, Australia, Luo *et al.* (2013) found that statistical downscaling outperformed dynamical downscaling for most of the climate variables. Furthermore, Gaitan *et al.* (2013) found that ANN outperformed MLR in statistical downscaling of temperature in terms of daily variability and annual climate indices over southern Ontario and Quebec, Canada.

There are two lingering questions from the current study: (i) why were the MLR forecasted yield values too high near the end of the data record and also too high near the beginning (Fig. 4(a)) and (ii) why were the SS of the ANN models considerably lower when retroactively validated than when cross-validated?

To understand the MLR behaviour, Fig. 6 illustrates schematically the type of non-linear relation between fertilizer and yield, i.e. the yield becomes saturated with increasing fertilizer. The linear model clearly over-predicts the yield at high and low fertilizer values (corresponding to the later and earlier parts of our data record), and under-predicts for the intermediate fertilizer values, as seen in Fig. 4(a). Furthermore, under the cross-validation procedure used in the current study, to validate the last quarter of the data record where much fertilizer was applied, the earlier data with less fertilizer were used for training the model. During forecast validation, the linear model was actually extrapolating to the high fertilizer regime,

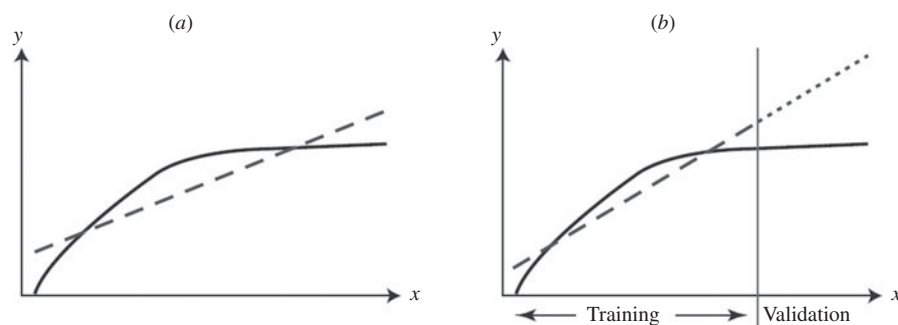


Fig. 6. Schematic diagram illustrating (a) the fitting of a linear regression model (dashed line) to data described by a non-linear relation (solid curve) and (b) the fitting of the linear model to the training data to the left of the vertical line, and then extrapolating to the right (dotted line) for the validation data, resulting in even more excessively high forecasted yields relative to the true relation (solid curve). Here x represents the amount of fertilizer and y the maize yield, and data points (not plotted) are assumed to be scattered around the solid curve.

giving even more excessively high forecasted yields in Fig. 6(b) than in Fig. 6(a).

The ANN model behaves very differently from the MLR model during extrapolation. The $\tanh(z)$ function used by ANN is bounded by the asymptotes ± 1 as its argument $z \rightarrow \pm\infty$. In Eqn (2), the ANN being a finite sum of \tanh functions will remain bounded even if a predictor x_i becomes arbitrary large in magnitude, while the MLR will not.

Hence the ANN extrapolates more gently than a linear function, which means that under the circumstances of Fig. 6(b), ANN will tend to give lower values than the extrapolated MLR solution and hence agree better with the true non-linear relation. Thus, fortuitously, the bounded functional shape of the ANN model gave ANN an advantage over MLR in the current study where extrapolation occurred.

During retroactive validation, extrapolation is almost eliminated since the forecast model is trained with all data available at the time of the forecast. The fortuitous extrapolation advantage of the ANN model was thereby essentially removed, and the retroactively validated ANN SSs were lower than the cross-validated SSs. Hence, the current study serves as a caveat that with non-stationary time series, cross-validated SSs may be inflated when extrapolation is involved during the cross-validation process. In the current case, even with the lower SS from retroactive validation, the ANN models were significantly outperforming persistence and MLR.

While there had been a fair number of studies showing ANN to outperform MLR in agricultural yield forecasts, the current study highlighted a subtle and often overlooked issue of forecast validation. In earlier studies, the standard validation procedure involves using part of

the data record for model training and part for model validation. In recent years, cross-validation, which uses the entire data record for validation, has become the new standard, especially when the data records are not long. However, in the current study, with non-stationary data, cross-validation inflated the SSs of the ANN due to extrapolation. Unlike cross-validation, retroactive validation does not suffer from extrapolation, although it cannot validate forecasts for the entire data record (e.g. in the current case, the first 15 years were used for the initial model training and not validated).

The rapid economic development in China during the current study period made it possible to greatly increase fertilizer use which increased the maize yield. The world's largest agricultural product exporters belong to highly developed countries where fertilizer impact for yield might be saturated and the influence of fertilizer may not be as large as seen in the current study.

ACKNOWLEDGEMENTS

The temperature and precipitation data were kindly provided by the Climatic Research Unit, University of East Anglia. Yield data and cropping calendar information were provided by Jilin University, which also supported K.M.'s trip to interview a maize farmer in Jilin. K.M. is supported by an overseas scholarship from Kwansai Gakuin University and W.H., by a discovery grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ALVAREZ, R. (2009). Predicting average regional yield and production of wheat in the Argentine Pampas by an

- artificial neural network approach. *European Journal of Agronomy* **30**, 70–77.
- CAMPBELL, C. A., ZENTNER, R. P. & JOHNSON, P. J. (1988). Effect of crop rotation and fertilization on the quantitative relationship between spring wheat yield and moisture use in southwestern Saskatchewan. *Canadian Journal of Soil Science* **68**, 1–16.
- CAMPBELL, C. A., SELLES, F., ZENTNER, R. P., MCCONKEY, B. G., BRANDT, S. A. & MCKENZIE, R. C. (1997a). Regression model for predicting yield of hard red spring wheat grown on stubble in the semiarid prairie. *Canadian Journal of Plant Science* **77**, 43–52.
- CAMPBELL, C. A., SELLES, F., ZENTNER, R. P., MCCONKEY, B. G., MCKENZIE, R. C. & BRANDT, S. A. (1997b). Factors influencing grain N concentration of hard red spring wheat in the semiarid prairie. *Canadian Journal of Plant Science* **77**, 53–62.
- CHANTRE, G. R., BLANCO, A. M., FORCELLA, F., VAN ACKER, R. C., SABBATINI, M. R. & GONZALEZ-ANDULAR, J. L. (2014). A comparative study between non-linear regression and artificial neural network approaches for modelling wild oat (*Avena fatua*) field emergence. *Journal of Agricultural Science, Cambridge* **152**, 254–262.
- CHEN, C. Q., LEI, C. X., DENG, A. X., QIAN, C. R., HOOGMOED, W. & ZHANG, W. J. (2011). Will higher minimum temperatures increase corn production in Northeast China? An analysis of historical data over 1965–2008. *Agricultural and Forest Meteorology* **151**, 1580–1588.
- China Maize (2012). *Corn Acreage and Yield in Chinese Provinces over Years*. Beijing: Institute of Crop Science, Chinese Academy of Agricultural Sciences (ICS-CAAS) (in Chinese). Available from: www.chinamaize.com.cn/scgk/ (accessed November 2011).
- CRU (Climatic Research Unit) (2010). *CRU TS3.10: Climatic Research Unit (CRU) Time-Series (TS) Version 3.10 of High Resolution Gridded Data of Month-by-Month Variation in Climate (January 1901–December 2009)*. Norwich: Climatic Research Unit, University of East Anglia. Available from: http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_fe67d66a-5b02-11e0-88c9-00e081470265/ (accessed 20 February 2014).
- DAI, X., HUO, Z. & WANG, H. (2011). Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crops Research* **121**, 441–449.
- ESRI (2008). *ESRI Data and Maps: Country-Based Boundary Datasets*, CD/DVD set. Kranzberg: Esri Deutschland GmbH.
- FAOSTAT (2012). *FAOSTAT Database*. Rome: FAO. Available from: <http://faostat.fao.org/site/567/default.aspx#anchor> (accessed February 2014).
- FORESEE, F. D. & HAGAN, M. T. (1997). Gauss–Newton approximation to Bayesian regularization. In *Proceedings of the 1997 International Joint Conference on Neural Networks*, Vol. 3, pp. 1930–1935. Piscataway, NJ: IEEE.
- FORTIN, J. G., ANCTIL, F., PARENT, L. E. & BOLINDER, M. A. (2011). Site-specific early season potato yield forecast by neural network in Eastern Canada. *Precision Agriculture* **12**, 905–923.
- GAITAN, C. F., HSIEH, W. W., CANNON, A. J. & GACHON, P. (2013). Evaluation of linear and non-linear downscaling methods in terms of daily variability and climate indices: surface temperature in southern Ontario and Quebec, Canada. *Atmosphere-Ocean*, doi: 10.1080/07055900.2013.857639.
- GHOOSI, R., YANI, R. M., JALALI, R. & RUZBAHMAN, M. (2012). Predicting wheat production in Iran using an artificial neural networks approach. *International Journal of Academic Research in Business and Social Sciences* **2**, 34–47.
- GILLELAND, E. (2010). *Confidence Intervals for Forecast Verification*, NCAR Technical Note NCAR/TN-479+STR. Boulder, CO: National Center for Atmospheric Research. Available from: <http://nldr.library.ucar.edu/repository/assets/technotes/TECH-NOTE-000-000-000-846.pdf> (accessed 20 February 2014).
- GUO, W. W. & XUE, H. R. (2012). An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Computing and Applications* **21**, 109–117.
- HILL, B. D., MCGINN, S. M., KORCHINSKI, A. & BURNETT, B. (2002). Neural network models to predict the maturity of spring wheat in western Canada. *Canadian Journal of Plant Science* **82**, 7–13.
- HSIEH, W. W. (2004). Nonlinear multivariate and time series analysis by neural network methods. *Reviews of Geophysics* **42**, RG1003, doi: 10.1029/2002RG000112.
- HSIEH, W. W. (2009). *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge: Cambridge University Press.
- HSIEH, W. W., TANG, B. & GARNETT, E. R. (1999). Teleconnections between Pacific sea surface temperatures and Canadian prairie wheat yield. *Agricultural and Forest Meteorology* **96**, 209–217.
- IFA (International Fertilizer Industry Association) (2012). *IFADATA Statistical Databases*. Available from: www.fertilizer.org/En/Statistics/IFADATA.aspx (accessed September 2012).
- Ji, B., SUN, Y., YANG, S. & WAN, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science, Cambridge* **145**, 249–261.
- JIANG, D., YANG, X., CLINTON, N. & WANG, N. (2004). An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing* **25**, 1723–1732.
- KAUL, M., HILL, R. L. & WALTHALL, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems* **85**, 1–18.
- LIU, H. L., YANG, J. Y., DRURY, C. F., REYNOLDS, W. D., TAN, C. S., BAI, Y. L., HE, P., JIN, J. & HOOGENBOOM, G. (2011). Using the DSSAT-CERES-Maize model to simulate crop yield and nitrogen cycling in fields under long-term continuous maize production. *Nutrient Cycling in Agroecosystems* **89**, 313–328.
- LOBELL, D. B. (2007). Changes in diurnal temperature range and national cereal yields. *Agricultural and Forest Meteorology* **145**, 229–238.
- LOBELL, D. B. & FIELD, C. B. (2007). Global scale climate crop yield relationships and the impacts of recent warming. *Environmental Research Letters* **2**, 014002, doi: 10.1088/1748-9326/2/1/014002

- LOBELL, D. B. & BURKE, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology* **150**, 1443–1452.
- LUO, Q. Y., WEN, L., MCGREGOR, J. L. & TIMBAL, B. (2013). A comparison of downscaling techniques in the projection of local climate change and wheat yields. *Climatic Change* **120**, 249–261.
- MA, L., TROUT, T. J., AHUJA, L. R., BAUSCH, W. C., SASEENDRAN, S. A., MALONE, R. W. & NIELSEN, D. C. (2012). Calibrating RZWQM2 model for maize responses to deficit irrigation. *Agricultural Water Management* **103**, 140–149.
- MACKAY, D. J. C. (1992). A practical Bayesian framework for backpropagation networks. *Neural Computation* **4**, 448–472.
- O'NEAL, M. R., ENGEL, B. A., ESS, D. R. & FRANKENBERGER, J. R. (2002). Neural network prediction of maize yield using alternative data coding algorithms. *Biosystems Engineering* **83**, 31–45.
- PANDA, S. S., AMES, D. P. & PANIGRAHI, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing* **2**, 673–696.
- QIAN, B., JONG, R. D., WARREN, R., CHIPANSHI, A. & HILL, H. (2009). Statistical spring wheat yield forecasting for the Canadian prairie provinces. *Agricultural and Forest Meteorology* **149**, 1022–1031.
- RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations (Eds D. E. Rumelhart & J. L. McClelland), pp. 318–362. Cambridge, MA: MIT Press.
- SHABBAR, A. & KHARIN, V. (2007). An assessment of cross validation for estimating skill of empirical seasonal forecasts using a global coupled model simulation. *CLIVAR Exchanges* **12**, 10–12.
- SUN, B. J. & VAN KOOTEN, G. C. (2013). Weather effects on maize yields in northern China. *Journal of Agricultural Science, Cambridge* **152**, 523–533.
- YANG, C., PETERSON, C. L., SHROPSHIRE, G. J. & OTAWA, T. (1998). Spatial variability of field topography and wheat yield in the Palouse region of the Pacific Northwest. *Transactions of the ASAE* **41**, 17–27.
- YANG, H. S., DOBERMANN, A., LINDQUIST, J. L., WALTERS, D. T., ARKEBAUER, T. J. & CASSMAN, K. G. (2004). Hybrid-maize – a maize simulation model that combines two crop modeling approaches. *Field Crops Research* **87**, 131–154.