PROOF COVER SHEET

Journal acronym:	TRES
Author(s):	Felix Kogan, Luis Salazar and Leonid Roytman
Article title:	Forecasting crop production using satellite-based vegetation health indices in Kansas, USA
Article no:	621464
Enclosures:	1) Query sheet 2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

For detailed guidance on how to check your proofs, please see http://journalauthors.tandf.co.uk/production/checkingproofs.asp.

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Felix	Kogan	
2		Luis	Salazar	
3		Leonid	Roytman	

Queries are marked in the margins of the proofs. Unless advised otherwise, submit all corrections and answers to the queries using the CATS online correction form, and then press the "Submit All Corrections" button.

AUTHOR QUERIES

General query: You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. (Please see http://journalauthors.tandf.co.uk/preparation/permission.asp.) Please check that any required acknowledgements have been included to reflect this.

AQ1 Please provide the department name (if applicable) for affiliation † and also check whether affiliations have been set correctly.

AQ2 Please expand FAO and UNESCO.

AQ3 Please check (°N) (°W).

AQ4 Please validate the website address http://www.usda.gov/nass/. It seems to be not accessible.

AQ5 Please check the usages of A_{max} and A_{min} in the text.

AQ6 Please provide the location details (city, state (if applicable), country) for NOAA.

AQ7 Please define the terms a_0, b_1, and b_2 in the text.

AQ8 Please check the inserted citation of Table 2 in the text.

AQ9 Please provide the page range for Vogel and Bange 1999.

AQ10 The reference 'USHCN 2005' has not been included in the reference list. Please check.

AQ11 Please provide good quality figures. Also, check all the figures for adding any information.

AQ12 Please check the missing term (), missing part label (b) and also the sense of caption of Figure 4.

Taylor & Francis
Author Services

How to make corrections to your proofs using Adobe Acrobat

Taylor & Francis now offer you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly using Adobe Acrobat. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

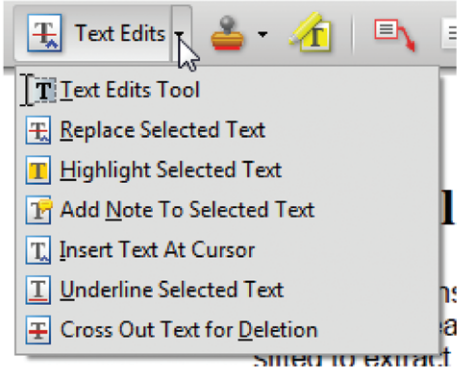
1. Save the file to your hard disk.

2. Check which version of Adobe Acrobat you have on your computer. You can do this by clicking on the "Help" tab, and then "About".
If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

- If you have Adobe Reader 8 (or a later version), go to "Tools"/ "Comments & Markup"/ "Show Comments & Markup".

- If you have Acrobat Professional 7, go to "Tools"/ "Commenting"/ "Show Commenting Toolbar".

3. Click "Text Edits". You can then select any text and delete it, replace it, or insert new text as you need to. It is also possible to highlight text and add a note or comment.



4. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. A full list of the comments and edits you have made can be viewed by clicking on the "Comments" tab in the bottom left-hand corner of the PDF.

If you prefer, you can make your corrections using the CATS online correction form.

TRES_A_621464
C/e: AS C/e QA: AS
International Journal of Remote Sensing
Vol. 00, No. 0, Xxxx 2011, 1–17

Forecasting crop production using satellite-based vegetation health indices in Kansas, USA

FELIX KOGAN*†, LUIS SALAZAR‡ and LEONID ROYTMAN§
†NOAA, National Environmental Satellite Data and Information Services, Camp Springs, MD 20276, USA AQ1
‡Optical Remote Sensing, NOAA CREST Center, City College of New York, New York, NY, USA
§City College of New York, New York, NY 10031, USA

(*Received 27 May 2007; in final form 31 August 2011*)

This article shows the results of early crop yield prediction from remote-sensing data. The study was carried out in Kansas, USA. The methodology proposed allows the estimation of winter wheat (WW), sorghum and corn yields 3–4 months before harvest. The procedure uses the vegetation health (VH) indices (vegetation condition index (VCI) and temperature condition index (TCI)) computed for each pixel and week over a 21-year period (1985–2005) from the Advanced Very High Resolution Radiometer (AVHRR) data. Over this period, a strong correlation was found between crop yield and VH indices during the weather-related critical period of crop development, which controls much final crop productivity. The 3-month advanced yield forecasts were independently compared with official agricultural statistics, showing that the estimation errors for WW, sorghum and corn were 8%, 6% and 3%, respectively. Implementing the 3–4 months lead forecast in operational practice will aid farmers to mitigate weather vagaries using irrigation, diseases/insects control, application of fertilizers and so on during a growing season and will help decision-makers to regulate marketing strategies, import/export and price policies and address food security issues.

1. Introduction

Over the years, large-area assessment of crop production has been based on *in situ* observations such as weather, climate, soil, economic and agricultural technology (e.g. cultivar changes, fertilizers, breeding, crop's protection, mechanization, diseases and pest control), using different modelling methods: stochastic, deterministic (process-based) and their combination (Obuhov 1949). These approaches have been used for many years, and the applications were conditioned by their advantages and shortcomings. One of the principal shortcomings of these approaches is that *in situ* observations are not dense enough in some regions/countries to adequately represent the high spatial variability of many weather parameters, especially precipitation and soil moisture. This is very important in regions/countries with limited and/or variable water supplies where the majority of staple food crops are grown. Besides, the availability of *in situ* observations is often limited due to economic, political and social situations – experienced primarily in non-developed and also in developing countries. Therefore, in the

*Corresponding author. Email: Felix.Kogan@noaa.gov

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2011.621464

recent 20 years, satellite data have been successfully used to monitor crop and pasture growth and development and to assess agricultural production. Crops such as spring and winter wheat (WW), corn, sorghum and soybeans were investigated in the USA, Brazil, Argentina, Greece, Morocco, Zimbabwe, China, Mongolia, India, Poland and other countries. These research works were conducted for the entire country and all administrative divisions (state, district, county and so on); the accuracy of predictions was estimated at 8–15%, with 1–3 months warnings in advance of the harvest (Hayas and Decker 1996, Kogan 1997, McVicar and Jupp 1998, Dabrowska-Zielinska *et al*. 2002, Liu and Kogan 2002, Singh *et al*. 2003, Kogan *et al*. 2003, Domenikiotis *et al*. 2004, Kogan *et al*. 2005, Mo *et al*. 2005).

In our recent research (Salazar *et al*. 2007) it was shown that the vegetation health (VH) indices obtained from the Advanced Very High Resolution Radiometer (AVHRR) flown on the National Oceanic and Atmospheric Administration (NOAA) operational polar orbiting satellites since 1981 can be successfully used to estimate crop yields over large (Kansas State, around 209 920 square kilometres (km^2)), medium (Crop Reporting District (CRD), 20 000–51 000 km^2) and small (county, 500–1300 km^2) administrative regions. Statistical analysis of WW in Kansas indicated that WW growth and productivity can be monitored during the entire growing season and that yield can be estimated 1–3 months in advance of the harvest. Specifically, the WW yield was shown to correlate strongly with AVHRR-based VH indices (Kogan 1997) during the critical period of crop development and that VH indices can be used as yield predictors.

Since the results of modelling a single crop were successful, we expanded the application of this approach to two other major crops in Kansas. In this article we compare the results of WW yield response to VH indices with those for sorghum and corn, which have different crop calendars and responses to weather fluctuations. In addition, these crops are fundamentally different, since wheat has C3 photosynthetic type and other two have C4. The specific goals were to (a) investigate whether sorghum and corn yields, similar to WW yield, correlate strongly with VH indices during the critical period (when crop productivity is very sensitive to weather conditions, both moisture and temperature) of their development; (b) investigate whether VH indices can be used as indicators of sorghum and corn yields; (c) build statistical models and investigate their performance; and (d) develop and investigate a scheme for early assessment of regional yield.

2. Study area

Kansas, called the 'granary' of the USA (Shroyer *et al*. 2004), was selected as the study area because it plays a leading role in the country's agricultural economy. Kansas constantly figures among the top states in the production of WW and sorghum, providing 24% and 39%, respectively, of the total value of crop production in the USA during 2003–2005 (table 1, United States Crop Reporting Board (USCRB) 2006). Kansas' corn accounts for 3.7% of the total production of the USA, yet it is important to the state agricultural economy, since its value is 2–3 times more than sorghum. Besides, corn is an interesting crop for investigation because it is less drought resistant than sorghum for the same crop weight.

Kansas's major soil is Harney silt loam, which possesses the ideal qualities of a prairie soil with the best combination of physical and chemical characteristics for growing crops and grasses and producing food and feed (USDA/SCS

Table 1. The average value of crop production (10^6 US$), Kansas, USA (USCRB 2006) during 2003–2005.

Crop	Kansas, USA	USA	Kansas (percentage (%) of total USA crop production)
WW	1263	5157	24.5
Sorghum	330	841	39.2
Corn	882	23 300	3.7

1993). Following ~~FAO/UNESCO~~ classification (http://www.fao.org/ag/agl/agll/wrb/wrbmaps/htm/kastano.htm), Kansas's prevailing soils are kastanozem type; the United States Department of Agriculture (USDA) (USDA/SCS 1993) provides additional information, classifying Kansas soils as prairie of Harney silt loam type. They cover mostly 26 West Central Kansas counties. Crops occupy over 12 million hectares (56% of the total 21 million hectares), while range and pasturelands occupy over 8 million hectares (37%). Kansas soils directly impact the economic well-being of people, providing nearly $7.5 billion annual income (USCRB 2006). Soils under crops in addition to adequate water, light and growing season thermal resources help Kansas lead the USA states in WW and sorghum production (USDA 2006). AQ2

Kansas has a continental climate (Boden *et al*. 2004). The average annual precipitation during 1990–2000 showed a gradual reduction from ~~700 to 800~~ mm in the east to 400 to 500 mm in the west. Most precipitation falls in summer with the monthly average ranging between 110 mm in the east and 70 mm in the west. Winter precipitation forms snow cover, which can reach up to 40–50 cm per year for the most of Kansas area except for the north-west, where snow reaches up to 90 cm. The annual average temperature in Kansas is between 11°C and 13°C. Summer is the warmest period, with the monthly average temperatures of 24–27°C and the maximum reaching 34°C. In general, Kansas's climate is suited to agriculture, especially in the north central part of the state. WW and sorghum grown here are ideally suited to the climate and soils of the area. Corn is generally a successful crop, although sustained periods of hot, dry and windy weather in July and August and maximum temperatures reaching 35–41°C occasionally prove disastrous. Such weather normally leads to excessive evapotranspiration, shortages of water and deterioration of VH, which in turn ~~lead~~ to loss in crop production.

3. Data

(°N) (°W) AQ3

3.1 *Crops*

Crop production in tonnes (t), area in hectares (ha) and yield in tonnes per hectare (t ha^{-1}) were obtained from the USDA National Agricultural Statistics Service (NASS) database site (http://www.~~usda.gov/nass/~~) for Kansas from 1980 through 2005. The NASS conducts the June Agricultural Survey each year. During the first 2 weeks in June, about 2400 interviewers contact more than 125 000 farmers, either by telephone or in person, to obtain information about crop areas that have either been planted or that they intend to plant and the proportion they expect to harvest as grain. Data AQ4

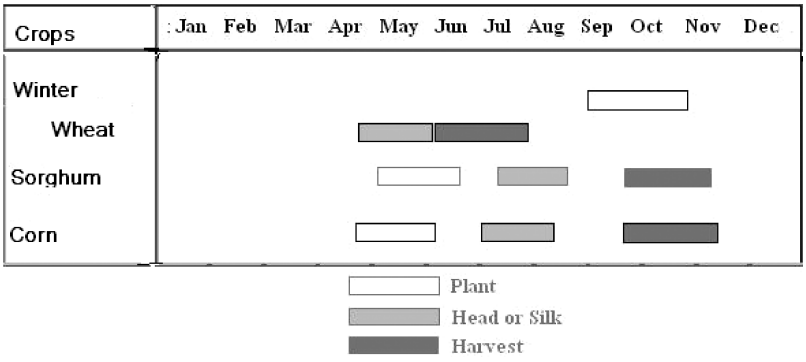


Figure 1. Crop calendar, Kansas, USA (~~USHCN 2005~~). AQ10

AQ11

from this survey are used to estimate the total area planted. The sample design for this survey utilizes two different sampling frameworks. The area framework, which is essentially the entire land mass of the USA, ensures complete coverage of the USA farm population. The list framework, a list of known farmers and ranchers, does not provide complete coverage of all farms, allowing the use of efficient data collection methods (Vogel and Bange 1999).

The calendar of the three investigated crops is shown in figure 1. In Kansas, WW planting starts late August, with the majority of crops being planted from September to early October. Heading of WW occurs the following year late April, with ripening by mid–June, when harvesting starts, and concludes in late July/early August. Sorghum is generally planted in May and June and corn in April and May. The harvest of both crops occurs in September and October (figure 1). The critical period to weather conditions for these crops is from July to August during heading in sorghum and during silk in corn.

3.2 *Satellite*

The satellite data represented by AVHRR-measured solar radiation reflected or emitted from the land surface were collected from the NOAA global vegetation index (GVI) data set available from 1981 through present. The GVI data set was developed by sampling the 4-km^2 global area coverage (GAC) data (in 10-bit counts) to 16-km^2 spatial resolution and compositing from the daily afternoon observations to seven-day composite (Kidwell 1997). The GVI digital counts in the visible (VIS, 0.58–0.68 μm, Ch1), near-infrared (NIR, 0.72–1.1 μm, Ch2) and infrared (IR, 10.3–11.3 μm, Ch4) spectral regions were used in this research. The VIS and NIR counts were converted into reflectance using pre-launch calibration coefficients, and the resulting values were post-launch calibrated following Rao and Chen (1995, 1996), Kidwell (1997), Rao and Chen (1999) and Heidinger *et al*. (2003), normalized by the cosine solar zenith angle (SZA) and corrected for the Sun–Earth distance. No corrections for water vapour absorption, aerosol scattering, bidirectional reflectance, sun–sensor geometry, satellite positioning (pitch, roll, yaw), orbit degradation and random errors were applied. The normalized difference vegetation index (NDVI) was calculated from the corrected VIS and NIR values as (NDVI = ((NIR) – (VIS)) ~~((NIR) + 1/(VIS))~~. The Ch4 counts were converted into brightness (radiative) temperature (BT) following Kidwell (1997). The

resulted annual NDVI and BT time series were smoothed statistically by applying a 5-span median filter and the least-squares technique (Kogan 1997) in order to remove high frequency noise in the data set due to the previously mentioned uncorrected factors.

4. Methodology

The research strategies used here were (a) to extract the weather component from the annual crop yield time series and from the NDVI and BT weekly time series values and (b) to correlate the weather-related components of crop yield with NDVI and BT components. The goal was to investigate the strength of the relationship and to determine whether the strongest correlation coincides with the period (critical) when the crop productivity is highly sensitive to weather conditions.

4.1 *Yield time series*

WW, sorghum and corn yields during 1980–2005 are presented in figure 2. As seen, these crops have exhibited positive yield trends since 1980 due to the improvement in agricultural technology (e.g. cultivar advances, fertilizers, breeding, crop protection). These yields were approximated by equation (1) (Brockwell and Davis 2000):

$$Y_i = T_i + \mathrm{d}Y_i, \tag{1}$$

where T is a slowly changing function representing the deterministic component (trend) regulated by agricultural technology, $\mathrm{d}Y$ is a random component regulated by weather fluctuations and i is the year or coded year number.

A trend of the form $T_i = a_0 + a_1 t_i$ was fitted to the data by choosing the parameters a_0 (intercept) and a_1 (slope) and by minimizing $\sum(Y_i - T_i)^2$ from the first to the last year; finally, the slopes (a_1) for WW, sorghum and corn were estimated as 0.0213, 0.0258 and 0.0662 (t ha^{-1} year^{-1}), respectively.

The random component ($\mathrm{d}Y$) can be approximated by either a difference or a ratio of the actual yield and the trend yield (estimated from equation (1)). If the trend slope is small, the difference is an appropriate method of approximation (Obuhov 1949), which we used in this study:

$$\mathrm{d}Y_\mathrm{i} = Y_i - T_i. \tag{2}$$

4.2 *AVHRR-based VH indices*

The principle for constructing VH indices stems from the properties of green vegetation to reflect/emit solar radiation in the VIS, NIR and IR wavelengths. If the vegetation is healthy, it reflects little radiation in the VIS (due to high chlorophyll absorption of solar radiation), much in the NIR (due to scattering of light by leaf internal tissues and the low absorption by pigments, water and other leaf constituents) and emits less thermal radiation in the IR spectral bands (because most of the available energy is partitioned into the latent heat flux compared with the sensible heat flux resulting in a relatively cooler canopy compared with the case when the canopy is water stressed). As a result, healthy vegetation has a higher NDVI and a lower BT. Conversely, unhealthy vegetation has a lower NDVI and a higher BT.

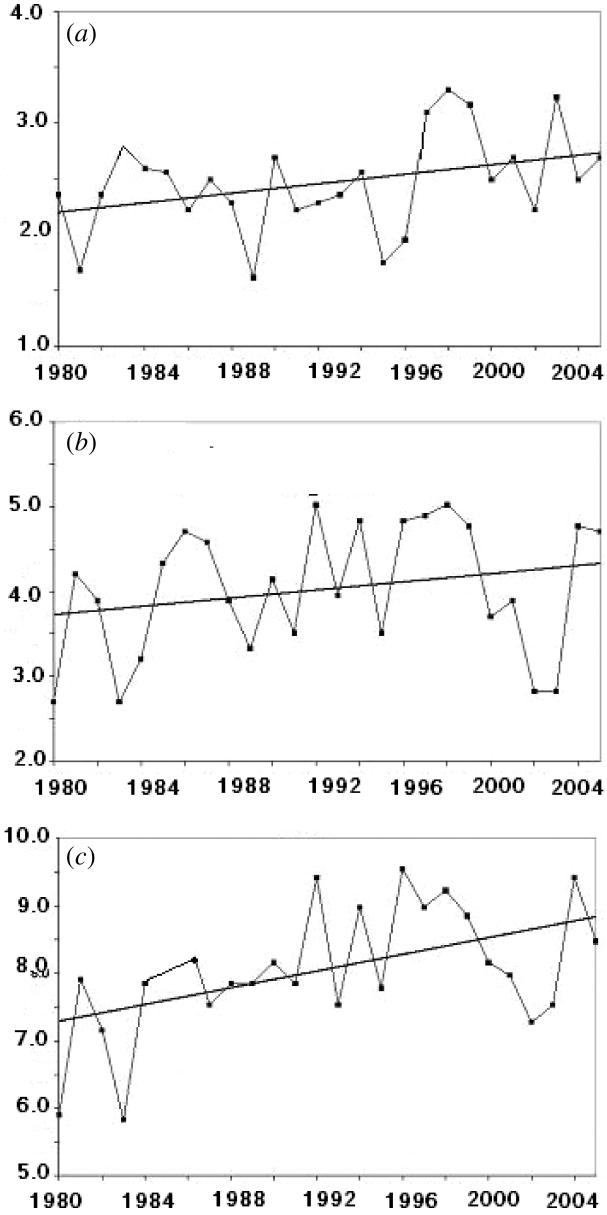


Figure 2. The average annual crop yield time series, Kansas, USA: (*a*) WW, (*b*) sorghum and (*c*) – corn.

The VH indices were calculated from the NDVI and the BT (equations (3) and (4)). The details of the algorithm are presented in Kogan (1997). The vegetation condition index (VCI) characterizes greenness and vigour, and through them, the chlorophyll and moisture contents. The temperature condition index (TCI) characterizes how hot the land surface is, including the vegetation canopy. Besides, the TCI characterizes the moisture availability through the near-surface radiation and aerodynamic conditions

(Jensen 2000, McVicar and Bierwirth 2001, McVicar and Jupp 2002, 1999, Anderson *et al*. 2007), which control the surface energy budget and hence the BT:

$$\mathrm{VCI} = 100 \times \frac{((\mathrm{NDVI}) - (\mathrm{NDVI})_{\min})}{((\mathrm{NDVI})_{\max} - (\mathrm{NDVI})_{\min})}, \tag{3}$$

$$\mathrm{TCI} = 100 \times \frac{((\mathrm{BT})_{\max} - (\mathrm{BT}))}{((\mathrm{BT})_{\max} - (\mathrm{BT})_{\min})}, \tag{4}$$

where NDVI, $\mathrm{NDVI}_{\max}$, $\mathrm{NDVI}_{\min}$, BT, $\mathrm{BT}_{\max}$ and $\mathrm{BT}_{\min}$ are the smoothed weekly NDVI or BT and their 1985–2005 absolute maximum ($A_{\max}$) and absolute minimum ($A_{\min}$). Since the $A_{\max}$ and $A_{\min}$ reflect the ~~lowest and the highest~~ values of NDVI and BT for each week during the 24 years of observation period, they can be considered the extreme thresholds of the annual NDVI and BT fluctuations. The NDVI and BT envelops, outlining these thresholds in the annual cycle, permit one to estimate whether a particular year's NDVI and BT are closer to $A_{\min}$ (the lowest greenness (equation (3)) and the highest BT (equation (4)) indicating vegetation stress, or closer to $A_{\max}$ (highest greenness and lowest BT) indicating healthy conditions, or between $A_{\max}$ and $A_{\min}$. Similarly to climatic thresholds (average, minimum, maximum and so on), we named the $A_{\max}$–$A_{\min}$ envelop as 'climatology' relative to which the NDVI and BT are expressed as VCI, TCI and VHI. VH indices change from 0, quantifying ~~severe~~ vegetation stress, to 100, quantifying favourable conditions (Kogan 1997). The average Kansas spatial values of VH indices for each week and crop during 1985–2005 were calculated for the major areas of each investigated crop location in Kansas (figure 3).

The NOAA-11 AVHRR instrument unexpectedly failed on 13 September 1994. Therefore, the 1995 data for WW and 1994 data for sorghum and corn were excluded from the analysis. This situation caused a gap in the afternoon observations between September 1994 and February 1995, when the NOAA-14 became operational. Moreover, the data from June to September (the period critical for corn and sorghum

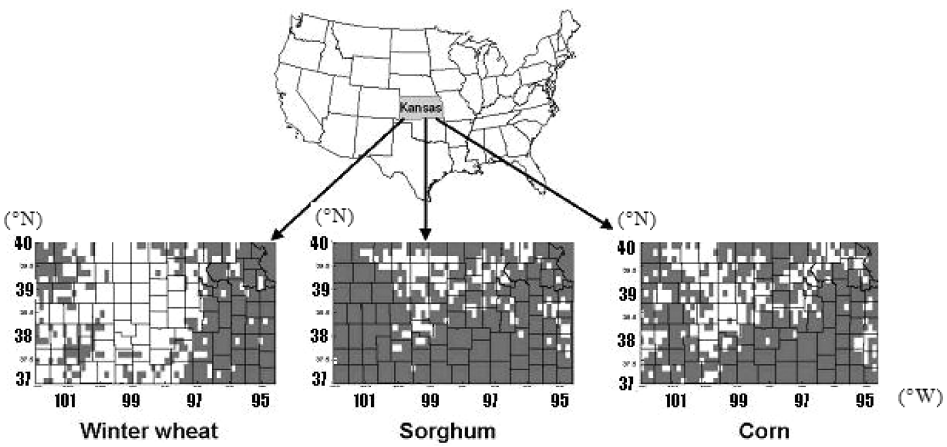


Figure 3. The satellite data collection for the major crop areas (white part). The pixel data were averaged over the white area for each week of 1985–2005 and for each crop, Kansas, USA (USCRB 2006).

production) of 1993 were removed from the statistical analysis because record flooding occurred across North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin and Illinois (Stallings 1994). This flooding produced much standing water, resulting in greatly reduced NDVI values compared with other years (water strongly absorbs radiation in NIR and marginally impacts VIS, resulting in a near-zero and/or negative NDVI).

4.3 *Other methodological considerations*

One of the principles used in the data analysis was to investigate the relationship between the *in situ* and satellite data characterizing weather component. Therefore, the data were expressed as a deviation from some standards: for yield – from the upward trend (expressed as $\mathrm{d}Y$) and for VH (VCI and TCI) – from NDVI and BT climatology. Both correlation and regression analysis of these deviations were performed to investigate the association of $\mathrm{d}Y$ with VCI and TCI. We correlated $\mathrm{d}Y$ of each crop with each week's VCI and TCI during 1985–2005 to investigate whether $\mathrm{d}Y$ correlates strongly with VH indices during the 'critical period' (a period of strong crop response to changes in weather conditions) of the investigated crops and to find the differences and similarities in the investigated crops' response and their correspondence to each crop growth and development. In addition, since the critical period continues for 4–7 weeks in a row, all these weeks might show a strong correlation between $\mathrm{d}Y$ and VH indices. An inclusion of all these weeks in the statistical models is not advisable, since the independent variables (VH indices) are highly collinear (correlate strongly with the neighbouring periods' VH). Therefore, two types of $\mathrm{d}Y$ models were investigated: (a) with the independent variables for the week (during the critical period) with the highest Pearson correlation coefficient (equation (5)) and (b) several weeks indices (during the critical period) with the Pearson correlation coefficient greater than 0.5. In the second case, the mean values for the selected weeks were used as independent variables (equation (6)):

$$\mathrm{d}Y = a_0 + b_1\,(\mathrm{VCI})_i + b_2\,(\mathrm{TCI})_j + e, \tag{5}$$

AQ7

$$\mathrm{d}Y = a_0 + b_1 \sum \frac{(\mathrm{VCI})_i}{n} + b_2 \sum \frac{(\mathrm{TCI})_j}{m} + e, \tag{6}$$

where i and j identify a week number for VCI and TCI, respectively; n and m identify the number of weeks for which the mean VCI and the mean TCI, respectively, are calculated~~; and e is the error~~ (Table 2).

AQ8

An important step in building statistical models is to test model predictions independently, since tests using the same training data would be very optimistic. The approach of reserving part of the data for validation is problematic when the quantity of data is limited as in this case. An alternative approach used here is cross-validation ('leave-one-out'). In this approach, a single year was left out one by one from the data set, a model was built and prediction was made for the eliminated year. As a result, 20 independent comparisons between the model predictions and observations were made.

To estimate the reliability of independent predictions, the corresponding verification model statistics were generated. Test criteria have been separated into two groups:

Table 2. Pearson correlation coefficients of dY with the vegetation condition index (VCI) and temperature condition index (TCI) for the period critical to crop dependence on weather, Kansas, USA.

	Winter wheat with VCI									
	VCI12	VCI13	VCI14	VCI15	VCI16	VCI17	VCI18	VCI19	VCI20	VCI21
dY	0.8222	0.8645	0.8641	0.8772	0.8724	0.8686	0.8749	0.8722	0.8753	0.8433
	Sorghum with VCI									
	VCI28	VCI29	VCI30	VCI31	VCI32	VCI33	VCI34	VCI35	VCI36	VCI37
dY	0.7873	0.8088	0.8182	0.8682	0.8761	0.8848	0.8792	0.8751	0.8506	0.7649
	Corn with VCI									
	VCI31	VCI32	VCI33	VCI34	VCI35	VCI36	VCI37	–	–	–
dY	0.8539	0.8638	0.8798	0.8528	0.8629	0.8194	0.7771	–	–	–
	Winter wheat with TCI									
	TCI16	TCI17	TCI18	–	–	–	–	–	–	–
dY	0.693	0.7318	0.7205	–	–	–	–	–	–	–
	Sorghum with TCI									
	TCI27	TCI28	TCI29	TCI30	TCI31	TCI32	TCI33	–	–	–
dY	0.7825	0.8416	0.8602	0.8667	0.8283	0.8023	0.7661	–	–	–
	Corn with TCI									
	TCI28	TCI29	TCI30	TCI31	TCI32	TCI33	TCI34	–	–	–
dY	0.7692	0.8497	0.8746	0.8966	0.8834	0.8905	0.821	–	–	–

Notes: Numbers attached to VCI and TCI indicate the week number of the calendar year, where week 1 is 1–7 January. Dashes indicate that the correlation coefficient is less than 0.5.

summary measures and difference measures. The first one includes the mean ($\bar{O}$) of the observed values (O_i) and the mean ($\bar{P}$) of the predicted values (P_i), while the second one describes the quality of simulation and includes the mean bias error (MBE, equation (7)) and the root mean square error (RMSE, equation (8)). They all are calculated according to Willmott (1982) and based on the term of (P_i–O_i):

$$\text{(A) Mean bias error (MBE): MBE} = \sum \frac{((P_i) - (O_i))}{n}, \tag{7}$$

$$\text{(B) Root mean squared error (RMSE): RMSE} = \sum \frac{((P_i) - (O_i))^2}{n}. \tag{8}$$

The summation is done from case 1 ($i = 1$) to case n ($i = n$).

5. Results and discussion

Figure 4 shows the dynamics of correlation coefficients for the end-of-season dY of each crop with each week's (from the first week in January to the last week in December) VCI and TCI during 1985–2005. The goals were to investigate if (a) dY correlates strongly with VH indices during the 'critical period' (a period starting from

2 to 3 weeks before flowering and ending with grain filling) of the investigated crops; (b) dY correlates weakly during non-critical and/or out of the growing season periods; and (c) to find the differences and similarities in crops' response to VCI and TCI and the correspondence of this response to crops' growth and development.

The correlation results (figure 4) emphasize first a distinctive and right difference between an early maturity season of WW, in which flowering and grain filling occur during April–May, and summer crops (sorghum and corn), which reproduce during July–September. As expected, the correlations of WW's dY with VH indices are low during January–March when WW is dormant and the indices are driven by soil and snow rather than by vegetation. WW's dY is highly correlated with VCI (0.82–0.88) during weeks 12–21 (April to May), when WW is heading, flowering and filling grain. Similarly, WW is highly correlated with TCI (0.69–0.73) during weeks 16–18 (late April–early May, see table 3). The indicated period ~~is~~ the most critical for WW because optimal moisture and thermal conditions are needed for maximum WW productivity. Both VCI and TCI reflect this situation quite well correlating strongly with dY. By early summer, when harvest starts and ends a few weeks later, correlations of WW dY with VH indices drop very quickly.

Summer crops have the same bell-shaped correlation curve dynamics reflecting sorghum and corn responses to changes in weather conditions (see figure 4). During July–September when sorghum and corn biomasses are developing, correlations of dY with VCI and TCI are very strong. These crops are very sensitive to moisture and thermal conditions, which are indirectly represented by VCI and TCI, respectively. For sorghum, the highest correlation of dY with VCI (0.76–0.88) occurs during weeks 28–37 (mid-July–mid September) and with TCI (0.77–0.87) occurs during weeks 27–33 (July–mid-August). For corn, the strong correlation of dY with VCI (0.78–0.88) occurs during weeks 31–37 (August–mid-September), and with TCI (0.77–0.90) during weeks

Table 3. Regression of dY on the two independent variables, VCI and TCI, for 1 week with the highest correlation coefficient (model approximated by equation (5)).

Coefficients	Parameter estimate	SE (t ha^{-1})	t-Value	p-Value
Winter wheat*				
a_0	−0.7658	0.1078	−7.1033	0.0000
b_1	0.0105	0.0023	4.6247	0.0002
b_2	0.0033	0.0023	1.4339	0.1697
Sorghum†				
a_0	−1.6317	0.1475	−11.0615	0.0000
b_1	0.0165	0.0034	4.9330	0.0001
b_2	0.0138	0.0032	4.3674	0.0004
Corn‡				
a_0	−1.7025	0.1505	−11.3151	0.0000
b_1	0.0152	0.0037	4.0983	0.0007
b_2	0.0179	0.0038	4.6833	0.0002

Notes: *Residual SE: 0.1853 on 17 degrees of freedom (DF). Multiple $R^2 = 0.7943$. F-statistic: 32.83 on 2 and 17 DF, p-value: <0.0001.
†Residual SE: 0.2598 on 17 DF. Multiple $R^2 = 0.8977$. F-statistic: 74.5855 on 2 and 17 DF, p-value: <0.0001.
‡Residual SE: 0.2465 on 17 DF.
Multiple $R^2 = 0.9014$. F-statistic: 77.68 on 2 and 17 DF, p-value: < 0.0001.

28–34 (mid-July to August) (see table 3). During the indicated periods, sorghum and corn are very sensitive to environmental stress such as a very high temperature, nutrient deficiency and water deficit or excess; any of which may reduce the potential crop yield. High correlations of dY with VCI and TCI during the investigated crops' critical period show the ability of the indices to be used as indicators of crop yield. It is important to note that the correlation of dY with VH indices, during the sorghum critical period, is slightly lower than that during the corn critical period. This is because sorghum is more resistant to drought than corn. It is important to emphasize that, since VCI is an integral or cumulative indicator of crop growth, whereas TCI is not because it is driven by both thermal conditions and moisture availability, the critical period characterized by TCI starts earlier and is shorter than that characterized by VCI (see figure 4).

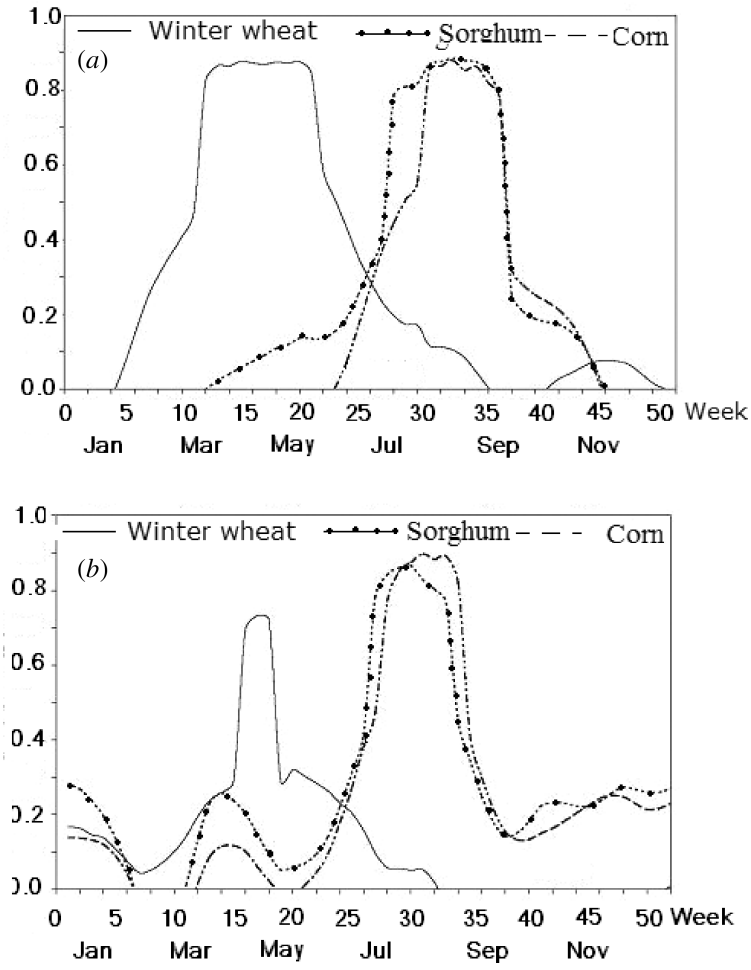


Figure 4. Dynamics of the Pearson correlation coefficient ~~()~~ of dY for WW, sorghum and corn with ~~(a) and~~ Kansas, USA (small negative values on the Y-axis are not shown). AQ12

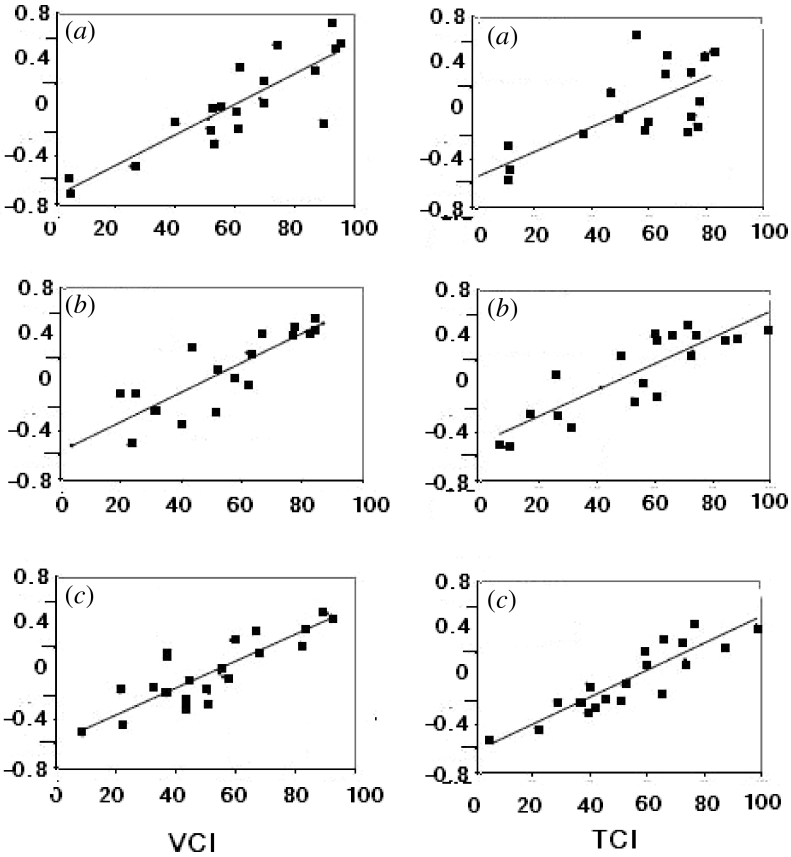


Figure 5. Scatter plots of dY with VH indices for the weeks with the highest correlation, Kansas, USA: (*a*) WW, (*b*) sorghum and (*c*) corn.

Figure 5 shows the scatter plots of dY with VH indices for the weeks with the highest correlation coefficients. Positive correlation of dY with VCI indicates that the transition from moisture stress to no stress (an increase in vegetation greenness) is accompanied with dY change from below to above the trend line; positive correlation with TCI indicates that dY changes from below to above trend when BT has similar to VCI stress-to-no stress changes (reduced from hot to cool, see equation (4)). Except for a few cases, extreme deviations in dY are associated with VCI and TCI below 20 and above 80. It is important to note that standing water and/or soil saturation reduce NDVI producing false signal in dY estimation. Such a situation occurred in 1993 when excessive early spring rainfall produced flood in the USA's Midwest, including a part of Kansas. As a result, WW yield was reduced, which is observed in figure 5 (week 15), where one point showed below trend dY for a very high (90) VCI value. Fortunately, such conditions do not occur frequently and/or rarely cover a large area.

Table 3 shows that the highest correlation of dY with VCI for WW, sorghum and corn was for weeks 15, 33 and 33 and with TCI for weeks 17, 30 and 31, respectively. If VCI and TCI for these weeks (see figure 5) are used as the only independent variables in a simple linear regression model, these would account for 77%, 78% and 77% (VCI) and 54%, 75% and 80% (TCI) of the variation in dY for WW, sorghum and corn, respectively (figure 5).

Table 4. Regression of dY on the two independent variables, mean VCI and mean TCI (model approximated by equation (6)).

Coefficients	Parameter estimate	SE (t ha^{-1})	t-Value	p-Value
Winter wheat*				
a_0	−0.7402	0.0993	−7.4524	0.0000
b_1	0.0123	0.0024	5.1110	0.0001
b_2	0.0013	0.0023	0.5498	0.5896
Sorghum†				
a_0	−1.6343	0.1515	−10.7905	0.0000
b_1	0.0161	0.0035	4.5779	0.0003
b_2	0.0144	0.0037	3.9025	0.0011
Corn‡				
a_0	−1.5873	0.1306	−12.1542	0.0000
b_1	0.0154	0.0031	4.9481	0.0001
b_2	0.0168	0.0030	5.5801	0.0000

*Residual SE: 0.1763 on 17 DF. Multiple $R^2 = 0.8140$. F-statistic: 37.19 on 2 and 17 DF, p-value: <0.0001.
†Residual SE: 0.2667 on 17 DF. Multiple $R^2 = 0.8922$. F-statistic: 70.34 on 2 and 17 DF, p-value: <0.0001.
‡Residual SE: 0.2285 on 17 DF. Multiple $R^2 = 0.9152$. F-statistic: 91.78 on 2 and 17 DF, p-value: < 0.0001.

Although there was only a 1–2% difference between the R^2 (coefficient of determination) values, we developed two regression models (*a*) and (*b*) using dY as a dependent variable and VCI and TCI as independent variables. Model (*a*) uses VCI and TCI for 1 week only with the highest correlation coefficient between dependent and independent variables (figure 5, equation (5)) and model (*b*) uses the average VCI and TCI for several weeks with the highest correlation (equation (6)). Tables 3 and 4 show the results of fitting an ordinary least squares (OLS) regression model given by equations (5) and (6) designed to estimate dY. The regression results support our preliminary conclusions from the analyses of figures 4 and 5. Table 3 shows that there is a strong relationship between dY and the independent variables (weeks with the highest correlation coefficient of each crops' dY with VCI and TCI, figure 5). The R^2 values are 0.79, 0.90 and 0.90 for WW, corn and sorghum, respectively. The tests of the composite hypothesis showed that all regression coefficients are statistically significant with F-statistic of 32.83, 74.59 and 77.68, respectively.

Similarly, table 4 shows that there is also a strong relationship between dY and the independent variables (average VCI and TCI for several weeks with significant correlation, table 3). The R^2 values are 0.81, 0.89 and 0.92, respectively, for WW, corn and sorghum. The test of the composite hypothesis showed that all regression coefficients are statistically significant with F-statistic of 37.19, 70.34 and 91.78, respectively. This is a very important result indicating that VH indices can be applied to many crops with different physiology and requirements to weather conditions.

Figure 6(*a*) and (*b*) display the observed (*O*) versus the independently simulated (predicted, *P*) crop yield. The R^2 values of 0.78, 0.87 and 0.87 (equation (5)) and 0.80, 0.86 and 0.89 (equation (6)) show that in most years, WW, sorghum and corn yields in Kansas can be modelled by the VH variables during the critical period (equations (5) and (6)). These results are slightly better for sorghum and corn, whose growing seasons start and end in the same year, than for WW, whose growing season spans

Table 5. Statistics (t ha^{-1}) of an independent testing of simulated (P) versus observed (O) yields for models approximated by equations (5) and (6), Kansas, USA.

		Equation (5)				Equation (6)			
Crop	$\bar{O}$	P	MBE	Var	RMSE	P	MBE	Var	RMSE
WW	2.5286	2.5278	−0.0008	0.0425	0.2010	2.5267	−0.0019	0.0389	0.1923
Sorghum	4.2056	4.2041	−0.0015	0.0703	0.2584	4.2072	0.0016	0.0744	0.2658
Corn	8.3767	8.3728	−0.0039	0.0648	0.2482	8.3728	−0.0039	0.0648	0.2275

from autumn to early summer of the following year. Residual values in individual years typically range between 0.12 and 0.23 t ha^{-1}year^{-1} for WW, 0.22 and 0.32 t ha^{-1}year^{-1} for sorghum and 0.18 and 0.25 t ha^{-1}year^{-1} for corn.

Table 5 shows the statistics of (P) versus (O) fit for models approximated by equations (5) and (6). We should emphasize first that the difference in the statistics between the two models for all parameters, MBE, Var and RMSE, is negligible. Therefore, both models can be used for prediction. However, the models where a single-week VH index is used as a predictor (equation (5)) are preferable, since these have 1–3 weeks lead time versus multi-week models (equation (6)). Second, corn and sorghum show larger RMSE and Var compared with WW. However, the error of prediction for WW is larger (7.95%) (WW is a three-season crop – fall, winter and spring) compared with 6.14% and 2.96% for single season crops sorghum and corn (equation (5)). Equation (6) models show similar results (7.68%, 6.39% and 2.72%).

In summary, the independent cross-validation suggests that WW, sorghum and corn yields can be forecasted with high accuracy from AVHRR-based VH indices (figure 6). In general, the models, in addition to correctly predicting the yield values, also correctly predict the direction of yield anomalies (deviation from the trend) for most of the years.

6. Conclusions

In this research, two AVHRR-based VH indices characterizing vegetation greenness and vigour (VCI) and also moisture and thermal conditions (TCI) were used as yield predictors of three different crops, WW, sorghum and corn in Kansas, USA. Modelling and independent testing showed that the yield of all three crops can be effectively predicted from the indices 2–4 months before harvest, giving farmers and others in the industry the opportunity to use this information to make more efficient crop management and market decisions.

The current research was limited to three crops of the many grown in the Great Plains, USA. However, the important fact is that these crops were quite different in physiology, crop calendar and requirement to environmental conditions. WW, a C3 crop, spans from autumn through early summer, whereas sorghum and corn (C4) are summer crops. Corn has a low resistance to drought compared with sorghum. Despite the differences, all three crops showed very good results of independent validation (explaining 80–90% of dY variance). The forecast of the simulated yields was independently compared with official agricultural statistics, showing that the errors of the estimates of WW, sorghum and corn are 8%, 6% and 3%, respectively.

This article also showed that there is potential for VH application in modelling other weather-dependent crops. It is important to state that VCI is an integral and

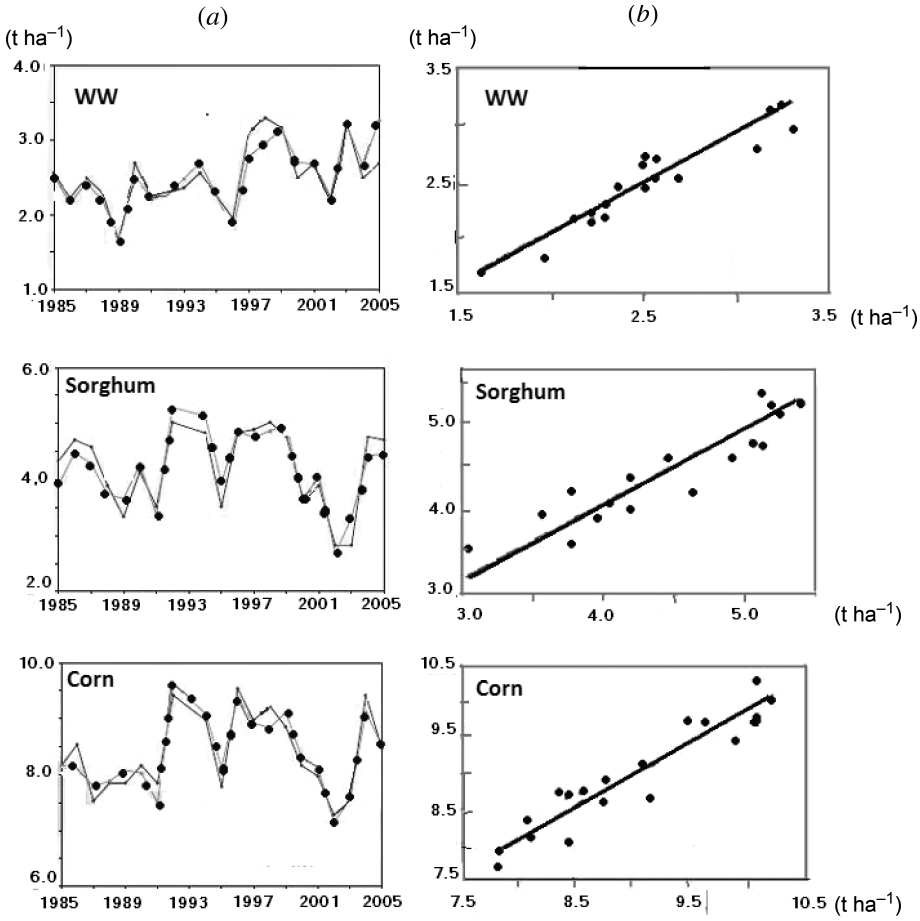


Figure 6. Observed and independently predicted average Kansas (USA) crop yields using models approximated by equations (5) (*a*) and (6) (*b*).

cumulative indicator of crop growth, whereas TCI is not because it is driven by thermal condition and moisture availability; therefore, the critical period characterized by TCI starts earlier and is shorter than that characterized by VCI. Interestingly, the models that used independent variables for 1 week only (with the strongest impact on yield) showed similar results as the models that used independent variables for several weeks with a high correlation. This indicates that single-week models are preferable, since they have 1–3 weeks lead time versus multi-week models.

The suggested approach can be improved through combining VH indices with weather parameters, especially those measuring winter conditions (snow cover, minimum temperature). The potential for improvements can come from addressing the non-linear statistical methods and alternative statistical approaches to the multiple linear regression.

Acknowledgements

The statements contained within the manuscript/research article are not the opinions of the USA government, but reflect the authors' opinions. We express our deepest appreciation of the suggestions made by Professor Arthur Roberts and two independent reviewers, which improved this article considerably.

References

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A. and Kustas, W.P., 2007, A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. *Journal of Geophysical Research*, **112**, D11112.

Boden, T.A., Williams, C.N., Quinlan, F.T. and Karl, T. and Carbon Dioxide Information Analysis Center (U.S.), 2004, *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data* (Oak Ridge, TN: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory).

Brockwell, P.J. and Davis, R.A., 2000, *Introduction to Time Series and Forecasting*, 2nd ed., p. 238 (New York: Springer).

Dabrowska-Zielinska, K., Kogan, F., Ciolkosz, A., Gruszczynska, M. and Kowalik, W., 2002, Modeling of crop growth conditions and crop yield in Poland using AVHRR-based indices. *International Journal of Remote Sensing*, **23**, pp. 1109–1123.

Domenikiotis, C., Spiliotopoulos, M., Tsiros, V. and Dalezios, N.R., 2004, Early cotton yield assessment by the use of the NOAA/AVHRR derived Vegetation Condition Index (VCI) in Greece. *International Journal of Remote Sensing*, **25**, pp. 2807–2819.

Hayas, M.J. and Decker, W.L., 1996, Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing*, **17**, pp. 3189–3200.

Heidinger, A.K., Sullivan, J.T. and Rao, C.R.N., 2003, Calibration of visible and near-infrared channels of the NOAA-12 AVHRR using time series of observations over deserts. *International Journal of Remote Sensing*, **24**, pp. 3635–3667.

Jensen, J.R., 2000, *Remote Sensing of the Environment: An Earth Resource Perspective*, p. 154 (New Jersey: Prentice Hall).

Kidwell, K.B., 1997, *Global Vegetation Index User's Guide* (Camp Springs, MD: US Department of Commerce, NOAA, National Environmental Satellite Data and Information Service, National Climatic Data Center).

Kogan, F., 1997, Global drought watch from space. *Bulletin of the American Meteorological Society*, **78**, pp. 621–636.

Kogan, F., Bangjie, Y., Guo, W., Pei, Z. and Jiao, X., 2005, Modelling corn production in China using AVHRR-based vegetation health indices. *International Journal of Remote Sensing*, **26**, pp. 2325–2336.

Kogan, F., Gitelson, A., Zakarin, E., Spivak, L. and Lebed, L., 2003, AVHRR-based spectral vegetation indices for quantitative assessment of vegetation state and productivity: calibration and validation. *Photogrammetric Engineering & Remote Sensing*, **69**, pp. 899–906.

Liu, W.T. and Kogan, F., 2002, Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. *International Journal of Remote Sensing*, **23**, pp. 1161–1179.

McVicar, T.R. and Bierwirth, P.N., 2001, Rapidly assessing the 1997 drought in Papua New Guinea using composite AVHRR imagery. *International Journal of Remote Sensing*, **22**, pp. 2109–2128.

McVicar, T.R. and Jupp, D.L.B., 1998, The current and potential operational uses of remote sensing to aid decisions on drought exceptional circumstances in Australia: a review. *Agricultural Systems*, **57**, pp. 399–468.

McVicar, T.R. and Jupp, D.L.B., 1999, Estimating one-time-of-day meteorological data from standard daily data as inputs to thermal remote sensing based energy balance models. *Agricultural and Forest Meteorology*, **96**, pp. 219–238.

McVicar, T.R. and Jupp, D.L.B., 2002, Using covariates to spatially interpolate moisture availability in the Murray-Darling Basin: a novel use of remotely sensed data. *Remote Sensing of Environment*, **79**, pp. 199–212.

Mo, X.G., Liu, S.X., Lin, Z.H., Xu, Y.Q. and McVicar, T.R., 2005, Prediction of crop production, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. *Ecological Modeling*, **183**, pp. 301–322.

Obuhov, V.M., 1949, *Urozhainost I Meteorologicheskie Factoru (in Russian): Yield and Meteorological Factors*, p. 314 (Moscow: Gosplanisdat).

Rao, C.R.N. and Chen, J., 1995, Inter-satellite calibration linkages for the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-7, -9 and -11 spacecraft. *International Journal of Remote Sensing*, **16**, pp. 1931–1942.

Rao, C.R.N. and Chen, J., 1996, Post-launch calibration of the visible and near infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft. *International Journal of Remote Sensing*, **17**, pp. 2743–2747.

Rao, C.R.N. and Chen, J., 1999, Revised post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft. *International Journal of Remote Sensing*, **20**, pp. 3485–3494.

Salazar, L., Kogan, F. and Roytman, L., 2007, Use of remote sensing data for estimation of winter wheat yield in the United States. *International Journal of Remote Sensing*, **28**, pp. 3795–3811.

Shroyer, J.P., Whitney, D. and Paterson, D., 2004, *Wheat Production Handbook* (Manhattan, KS: K-State research & Extension).

Singh, R.P., Roy, S. and Kogan, F., 2003, Vegetation and temperature condition indices from NOAA-AVHRR data for monitoring drought over India. *International Journal of Remote Sensing*, **24**, pp. 4393–4402.

Stallings, E.A., 1994, *Hydrometeorlogical Analysis of the Great Flood of 1993* (Silver Spring, MD: Department of Commerce, NOAA, National Weather Service).

United States Crop Reporting Board [USCRB], 2006, *Crop Production* (Washington, DC: Crop Reporting Board, Statistical Reporting Service, US Dept. of Agriculture).

United States Department of Agriculture [USDA], 2006, Outlook for U.S. Agricultural Trade. USDA, Washington, DC. Available online at: http://www.ers.usda.gov/publications

United States Department of Agriculture and Soil Conservation Service [USDA/SCS], 1993, *Soil Survey Manual. U.S. Department of Agriculture Handbook 18* (Washington D.C: United States Department Of Agriculture and Soil Conservation Service).

Vogel, F.A. and Bange, G.A., 1999, Understanding crop statistics. In *Office of the Chief Economist*, National Agricultural Statistics Service and World Agricultural Outlook Board (Eds.) (Washington, D.C: US Department of Agriculture). **AQ9**

Willmott, C.J., 1982, Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, **63**, pp. 1309–1313.