



Potential monitoring of crop production using a satellite-based Climate-Variability Impact Index

Ping Zhang*, Bruce Anderson, Bin Tan, Dong Huang, Ranga Myneni

Department of Geography, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA

Received 10 May 2005; accepted 13 September 2005

Abstract

The capabilities of the MODerate resolution Imaging Spectroradiometer (MODIS) present some exciting possibilities for improved and timely monitoring of crop production. A quantitative index is introduced in this paper to study the relationship between remotely sensed leaf area index (LAI) and crop production. The Climate-Variability Impact Index (CVII), defined as the monthly contribution to anomalies in annual growth, quantifies the percentage of the climatological production either gained or lost due to climatic variability during a given month. By examining the integrated CVII over the growing season, this LAI-based index can provide both fine-scale and aggregated information on vegetation productivity for various crop types. Once the relationship between the CVII and crop production is developed based on the historical record, a trained statistical model can be applied to produce homogeneous production forecasts (in which the model is trained and tested for a particular region), as well as heterogeneous forecasts (in which the model is trained in a particular region and applied to a different region). Both the homogeneous and the heterogeneous model predictions are consistent with United States Department of Agriculture (USDA)/FAO estimates at regional scales. Finally, by determining the estimated production as a function of the growing-season months it is possible to determine when in the phenological cycle the predictive value of the CVII plateaus and which months within the phenological cycle provide the greatest predictive capacity. Overall, the high temporal and spatial resolution of the satellite LAI products makes the CVII a useful tool in near real-time crop monitoring and production estimation.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Remote sensing; Leaf area index; Crop yield; Climate impacts; Drought index

1. Introduction

Crop monitoring and early yield assessment are important for agriculture planning and policy making at regional and national scales. Numerous crop growth simulation models are generated using crop state variables and climate variables at the crop/soil/atmosphere interfaces to get the pre-harvest information on crop yields (e.g. Monteith, 1977). However, most of these models are limited to specific regions/periods due

to significant spatial–temporal variations of those variables. Furthermore, the limited network of stations and incomplete climate data make crop monitoring and yield assessment a daunting task (Kogan, 1997). In addition, the meteorological data may miss important variability in vegetation production, which highlights the need for quantification of vegetation changes directly when monitoring climate impacts upon vegetation (Zhang et al., 2004). In this sense, remotely sensed metrics of vegetation activity have the following advantages: a unique vantage point, synoptic view, cost effectiveness, and a regular, repetitive view of nearly the entire Earth's surface (Johnson et al., 1993), thereby making them potentially better suited for crop monitoring and yield estimation than conventional

* Corresponding author. Tel.: +1 617 353 2092;
fax: +1 617 353 8399.

E-mail address: zhping@crsa.bu.edu (P. Zhang).

weather data. For instance, it has been shown that the application of remotely sensed data can provide more accurate crop acreage estimates at national/continental scales (Sharman et al., 1992). Furthermore, numerous field measurements and theoretical studies have demonstrated the utility of remotely sensed data in studies on crop growth and production (Moulin et al., 1998). These two applications suggest the feasibility of large-scale operational crop monitoring and yield estimation.

Empirical relationships between the remotely sensed data and crop production estimates have been developed for monitoring and forecasting purposes since the early 1980s. For instance, Colwell et al. (1977) found a strong correlation between winter wheat grain yield and Landsat spectral data. However, these relationships did not hold when extended in space and time (Barnett and Thompson, 1983). Later, various other vegetation indices generated from Landsat data, such as the ratio of the reflectance at near infrared to red and the normalized difference vegetation index (NDVI) were used in yield estimation of sugarcane (e.g. Rudorff and Batista, 1990), wheat (e.g. Lobell et al., 2003), and rice (e.g. Patel et al., 1985). The Landsat series have a spatial resolution of 30 m and can provide reflectance data from different spectral bands. However, these high-resolution data require enormous processing effort, and may not be applicable for surveys of large-area general crop conditions (Barnett and Thompson, 1983).

Vegetation indices derived from data from the Advanced Very High Resolution Radiometer (AVHRR) were also used for crop prediction, environmental monitoring, and drought monitoring/assessment (e.g. Kogan, 1990, 1995; White et al., 1997). For example, Rasmussen (1992) found that millet yields in northern Burkina Faso are linearly correlated with the AVHRR NDVI integrated over the reproductive period. Similarly, Hochheim and Barber (1998) found that the accumulated AVHRR NDVI provided the most consistent estimates of spring wheat yield in western Canada. The Vegetation Condition Index (VCI) derived from AVHRR data is widely applied in real-time drought monitoring and is shown to provide quantitative estimation of drought density, duration, and effect on vegetation (Kogan, 1990, 1995). The VCI can separate the short-term weather signals in the NDVI data from the long-term ecological signals. According to Domenikotis et al. (2004), the empirical relationship between VCI and cotton yield in Greece are sensitive to crop condition well before the harvest and provide an indication of the final yield. Unfortunately, the AVHRR data are not ideally suited for vegetation monitoring

applications because of the lack of precise calibration, poor quality of geometric registration, and difficulties in cloud screening (e.g. Goward et al., 1991; Sellers et al., 1994).

The radiometric and geometric properties of the MODerate resolution Imaging Spectroradiometer (MODIS) provide a significantly improved basis for vegetation monitoring and yield predictions with remotely sensed data (Justice et al., 1998; Running et al., 1994; Zhang et al., 2003). Also, given the improved atmospheric correction and cloud screening and the high temporal and spatial resolution of the various MODIS vegetation products (e.g. leaf area index), this sensor seems well suited for near real-time crop monitoring. In general, the daily MODIS products are released 1–2 days after the image is taken. For the 8-day or monthly composite products, a 1–2 weeks processing period is required. Once the data have been archived, they can be downloaded from Earth Observing System data gateway (<http://edcimswww.cr.usgs.gov/pub/imswelcome/>). A valuable metric of vegetation production derived from this sensor are the satellite-based estimates of leaf area index (LAI). Leaf area responds rapidly to abiotic and biotic influences and the variability of LAI can integrate various conditions affecting plant growth and development (Holben et al., 1980). As such, integrated LAI over the growing season is highly correlated with crop yield because both the magnitude and duration of photosynthetic activity is considered (e.g. Tucker et al., 1980). For instance, previous studies have shown that potential yield is a function of leaf area at the beginning of the reproductive state, and final yield is related to the duration of green LAI assuming the absence of significant stresses during the heading/filling stages (see Hatfield, 1983).

We have previously demonstrated that a Climate-Variability Impact Index (CVII) derived from the MODIS LAI product quantifies the percentage of the climatological annual production either gained or lost due to climatic variability and has a potential application in crop monitoring and yield estimations (Zhang et al., 2004). As a continuation of this effort, in this paper we will analyze the relationships between the LAI-based CVII and crop yield for various crop types and locations. With estimates from FAO and USDA, regression models will be trained on past data, and then be applied to predict future yield for a given study region. Finally, we will also evaluate the phenological cycle to determine which months within the growing season provide the greatest predictive capacity, which in turn may allow for more accurate model estimation and early famine prediction prior to the end of the growing season.

2. Data

2.1. MODIS IGBP land-cover map

The MODIS land-cover classification product identifies 17 classes of land cover in the International Geosphere–Biosphere Programme (IGBP) global vegetation classification scheme (Friedl et al., 2002). This scheme includes 11 classes of natural vegetation, 3 classes of developed land, permanent snow or ice, barren or sparsely vegetated land, and water. The latest version of the IGBP land-cover map is used to distinguish croplands from the other biomes in this research.

2.2. MODIS LAI

The retrieval technique of the MODIS LAI algorithm is as follows. For each land pixel, given red and near-infrared reflectance values, along with the sun and sensor-view angles and a biome-type designation, the algorithm uses model-generated look-up tables to identify likely LAI values corresponding to the input parameters. This radiative transfer-based look-up is done for a suite of canopy structures and soil patterns that represent a range of expected natural conditions for the given biome type. The mean value of the LAI values found within this uncertainty range is taken as the final LAI retrieval value. In certain situations, if the algorithm fails to localize a solution either because of biome misclassification/mixtures, high uncertainties in input reflectance data or algorithm limitations, a backup algorithm is utilized to produce LAI values based upon the empirical relationship between NDVI and LAI (Myneni et al., 1997).

The latest version of MODIS global LAI from February 2000 to December 2004 was taken to characterize the crop activity in this study. The 8-day LAI products are distributed to the public from the Earth Observing System (EOS) Data Gateway Distributed Active Archive Center. The 8-day products also provide quality control variables for each LAI value that indicate its reliability. The monthly global product was composited across the 8-day products using only the LAI values with reliable quality. The monthly global products at 1-km resolution with Sinusoidal (SIN) projection are available at Boston University (Yang et al., 2005). In this paper, monthly LAI at 1-km resolution are used to generate our Climate-Variability Impact Index. As these will be compared with estimates of crop production reported at county/state-levels, the vegetation-based CVII fields were aggregated over the corresponding counties/states using the county bound-

aries 2001 map from the National Atlas of the United States (<http://nationalatlas.gov>).

2.3. AVHRR LAI

AVHRR LAI is used as a substitute for the MODIS LAI to examine the temporal characteristics of vegetation activity over longer time periods. The AVHRR LAI is derived from the Global Inventory Modeling and Mapping Studies (GIMMS) NDVI produced by NASA GIMMS group (Tucker et al., in press). Monthly LAI from 1981 to 2002 at 0.25° were derived based on the empirical relationship between NDVI and LAI for different biomes. Literature works show that this empirical relationship might be different for the same biome at different locations (Asrar et al., 1992; Gutman, 1991; Price, 1993). To eliminate this effect, models are generated for each pixel to calculate GIMMS LAI from GIMMS NDVI. The MODIS LAI and GIMMS NDVI overlapped from March 2000 to December 2002, which provides a basis for generating a piecewise linear relationship between these two products. Once the coefficients of the linear model are calculated, the whole range of GIMMS NDVI can be converted into GIMMS LAI, which is consistent with the MODIS products. Our preliminary results indicate a good agreement between GIMMS LAI and MODIS LAI at quarter degree resolution with less than 5% relative difference for each main biome (results not shown).

2.4. GIMMS NPP

In this research, we also use model-generated estimates of Net Primary Production (NPP) from Nemani et al. (2003) as a predictor of crop production. This NPP is a monthly product from 1982 to 1999 at a spatial resolution of half degree. This global NPP product was generated as follows. GIMMS NDVI were first used to create LAI and FPAR with a 3D radiative transfer model and a land-cover map as described in Myneni et al. (1997). Then, NPP was estimated from a production efficiency model (PEM) using the following three components: the satellite-derived vegetation properties, daily climate data, and a biome specific look-up table of various model constants and variables. Further details can be found in Nemani et al. (2003).

2.5. Crop production

Crop production data from several sources are used in this research. We focus upon total production, as

opposed to yield, for instance, because although the two are highly correlated with each other, total production is typically the parameter of interest for crop monitoring and yield prediction. In this paper, we will explicitly refer to “production” when discussing quantitative results, however for simple qualitative statements we sometimes retain the generic term “yield” as synonymous for “production”.

The country-level crop production from 1982 to 2000 in European countries is from FAOSTAT 2004 data set. The county-, district-, and state-level production data in United States are from the National Agricultural Statistics Service (NASS) at United States Department of Agriculture (USDA) (<http://www.nass.usda.gov:81/ipedb/>). USDA provides two independent sets of county crop data: one is a census of agriculture, which is released every 5 years; the other one is annual county crop data, which is based on reports from samples. We used the annual crop estimates in this study. Due to the processing effort required for the fine resolution remotely sensed data, we studied two crops (corn and spring wheat) in two US states (Illinois and North Dakota) at county- and district-scales. At coarser scales, we expanded the regions to include Illinois (IL), Minnesota (MN), Michigan (MI), Iowa (IA), Indiana (IN), and Wisconsin (WI) for corn; to North Dakota (ND), Montana (MT), Minnesota (MN), and South Dakota (SD) for spring wheat; to Kansas (KS), Oklahoma (OK), Colorado (CO), and Nebraska (NE) for winter wheat. The county- and district-level data of Illinois and North Dakota are from 2000 to 2004; the state-level data are from 1982 to 1999.

3. Correlation between CVII and production at different scales

We previously developed an index, the Climate Impact Index (CII), to identify regions that are particularly susceptible to vegetation loss due to climatic variability during the growing season (Zhang et al., 2004). In addition to its use as a diagnostic tool to quantify the difference between ecosystems, given the high spatial and temporal resolutions this index can also capture the temporal variations in each ecosystem. In this paper, a similarly derived Climate-Variability Impact Index will be used for real-time crop monitoring, yield estimations, and climatic impact diagnosis. For a given pixel p , let $L(p, m, y)$ be the LAI in month m and year y , $L'(p, m)$ be the climatological LAI in month m , and $\sum L'(p)$ be the climatological annual LAI. The

index $CVII(p, m, y)$ in month m and year y is then calculated as:

$$CVII(p, m, y) = 100 \times \frac{L(p, m, y) - L'(p, m)}{\sum L'(p)} \quad (1)$$

where

$$L'(p, m) = \frac{1}{N_y} \sum_y L(p, m, y) \quad (N_y, \text{the number of years})$$

$$\sum L'(p) = \sum_m L'(p, m).$$

Here, the CVII quantifies the percentage of the climatological annual grid-point production either gained or lost due to climatic variability in a given month. Preliminary results show that the CVII can successfully identify vegetation loss of up to 30–50% of the total annual production in impacted regions during historic drought events (Zhang et al., 2004). In the following, we examine the relationships between CVII and crop production at local, regional, and national scales.

3.1. CVII versus production at local scale

At the local scale, we used the 1-km resolution MODIS LAI data, from 2000 to 2004, to generate the Climate-Variability Impact Index. Data from 2000 to 2003 are examined in this section and the 2004 LAI data are saved for the evaluation of prediction schemes in Section 4. The MODIS IGBP land-cover map at 1-km resolution was used to select cropland pixels. We chose two states, Illinois and North Dakota, as representative states to examine the relationship between CVII and production at local scales. Illinois contains 102 counties, which are grouped into 9 crop-reporting districts (CRD). The principal crop in Illinois is corn, which is approximately 50% of all crops by area. Since we do not have a detailed land use map for each crop, corn is used as the representative crop in Illinois when we study the relationship between crop production and the LAI-based CVII. North Dakota contains 53 counties and 9 crop-reporting districts. The principal crop in North Dakota is spring wheat, which composes approximately 40% of all crops by area. For this reason, variability of spring wheat is studied using North Dakota data.

Crop production estimates are given at county- and CRD-levels by the United States Department of Agriculture. Accordingly, we aggregated LAI over the same regions by overlapping the LAI map with the county map and then calculated the Climate-Variability Impact Index for each county/CRD. Previous studies

have found that LAI integrated over the growing season is highly correlated with crop yield because both the magnitude and duration of photosynthetic activity is considered (e.g. Tucker et al., 1980). Thus, in this research, the LAI-based CVII is summed over the growing season. The usual growing period of corn in Illinois is from April (planting time) to early October (harvesting time). The corresponding times for spring wheat in North Dakota are April and late September. Accordingly, the CVII is summed over the growing season using the following equation:

$$CVII(p, y) = 100 \times \sum_m \left(\frac{L(p, m, y) - L'(p, m)}{\sum L'(p, m)} \right) \quad (2)$$

At the same time, survey-based crop production estimates (Y) are normalized by the 2000–2003 mean (\bar{Y}).

$$Y' = 100 \times \frac{Y}{\bar{Y}} \quad (3)$$

Our results demonstrate strong positive correlation between the crop production and the CVII (Fig. 1). At county- and CRD-levels, 50% of the variance in crop production can be explained by the CVII. In general, a positive CVII indicates more vegetative production than the climatological average and vice versa. In addition, significant decreases in crop production during 2002, associated with a large-scale drought (e.g. LeCompte and Cutlip, 2003) are successfully captured by the CVII. However, the relationship between the CVII and the production variation is not unique. Counties/CRDs with near-zero CVII can have a range of crop production. Overall, however, the CVII does capture a large fraction

of the variance in production, which indicates that the CVII can be used as a possible tool for monitoring climate impacts upon vegetation.

3.2. CVII versus production at regional scale

The United States was affected by a significant drought in 1988 and 1989. The drought in 1988 was a result of a dry winter followed by a dry spring, which resulted in yield declines of more than 30% in certain areas (Karrenbrock, 1989). A 30% reduction in winter wheat yield was noted in Kansas and Nebraska due to the drought occurred in the early growing season of 1989 (Kogan, 1995). Here, we chose three affected areas to investigate the relationship between CVII and production at regional scales (i.e. aggregated over counties) during these drought periods. In the first region (IL, MN, MI, IA, IN, and WI), more than 50% of the crop is corn (by area). In the second region (ND, MT, MN, and SD), spring wheat is one of the principal crops (25% by area), while in the third region (KS, OK, CO, and NE), winter wheat is the principal crop (35% by area).

Because MODIS was not yet launched during this drought period, GIMMS LAI data were used to generate the cumulative CVII in the middle of the growing season, which is from June to August for corn and spring wheat states, and from April to June for winter wheat states. Fig. 2 shows strong positive correlation between the CVII and production anomaly in all three regions. In general, more than 60% of the variance in crop production can be explained by the variation of CVII. In addition, the corn production decreased over

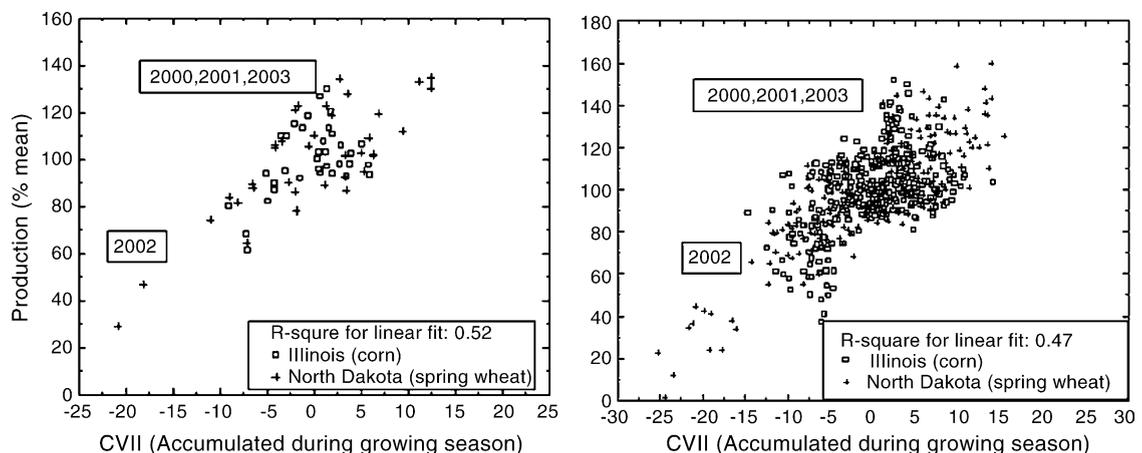


Fig. 1. Correlation between production (corn in Illinois and spring wheat in North Dakota) and Climate-Variability Impact Index (CVII) at the county-level (right panel) and crop-reporting districts (CRD)-level (left panel). The CVII is summed over the growing season and crop production is normalized as the percent of the 2000–2003 mean.

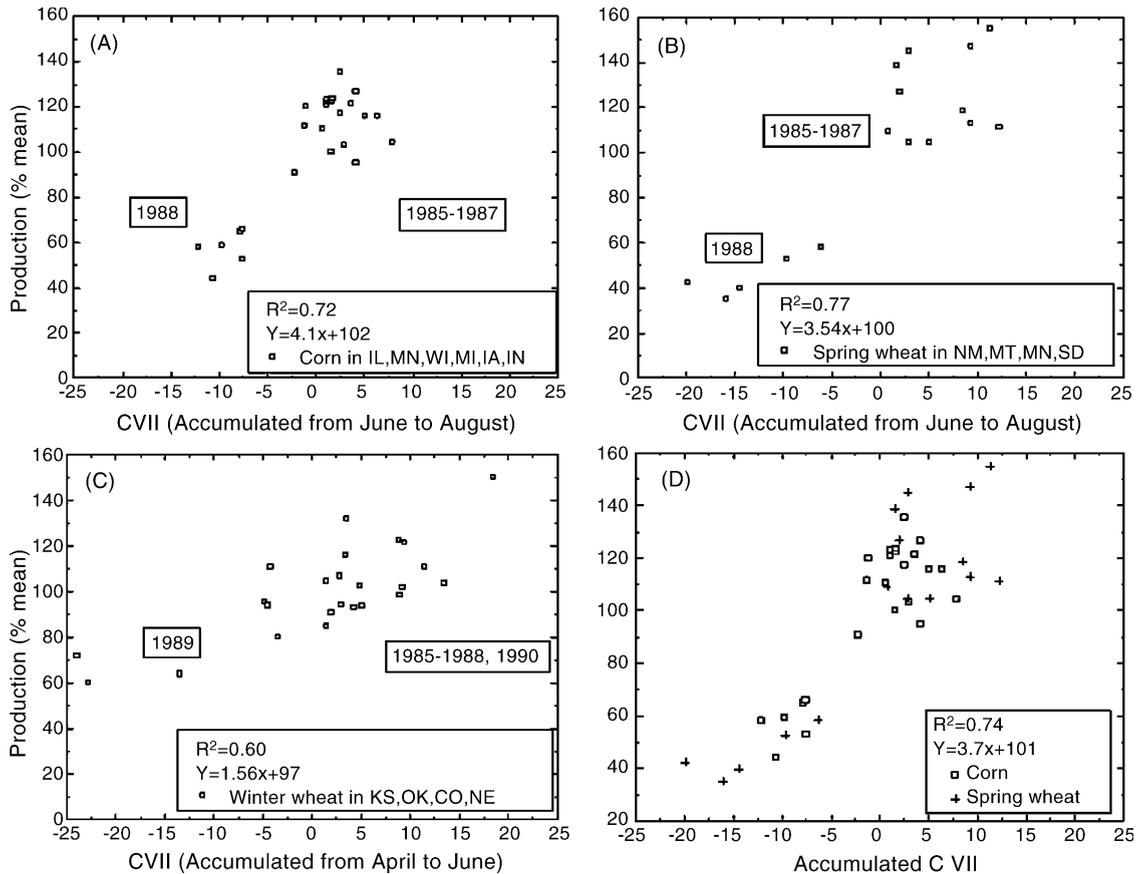


Fig. 2. Correlation between production and Climate-Variability Impact Index (CVII) at state-levels for corn (A), spring wheat (B), winter wheat (C), and corn and spring wheat together (D). GIMMS LAI is used as a substitute to generate the CVII. The CVII is accumulated over the growing season and crop production is normalized as the percent of the mean.

30% and the spring wheat decreased over 40% in 1988, and the winter wheat decreased up to 30% in 1989. As before, these results suggest that CVII may be potentially useful in a large-scale monitoring capacity as well as at finer scales. It is also interesting to note that the relationship in corn states is similar to that in spring wheat states. If the data from these two crops are combined, the explained variance as well as the coefficients change only slightly (Fig. 2D), suggesting that the CVII–production relationship of corn and spring wheat are almost identical at regional scales. However, this is not universal as the winter wheat in fact does have a different CVII–production relationship. Nevertheless, the result is encouraging in that it suggests that for certain crop types this relationship may be more homogenous than would have been expected. In Section 4, we test whether there is also a geographic and/or temporal dependence of the relationship by performing independent tests with data outside of the training set.

3.3. CVII versus production at national scale

At the national scale, we chose four European countries (England, Ireland, Germany, and France) to study the relationship between the CVII and crop production. One of the principal crops in these countries is wheat, which has the largest harvest area. The normal growing period of wheat in these countries is from October to August of the following year. We calculated the integrated AVHRR-based CVII and production anomaly for each country from 1982 to 2000. Fig. 3 demonstrates that the LAI-based CVII is positively related to crop production at the national scale. Although the coefficients are significant at the 99% confidence level, the correlation is not as high as at the local- and state-levels, which suggests a weakening of the relationship between CVII and crop production with decreasing resolution. At the country-level, it is likely not accurate to use one crop type to represent all vegetation activity because countries contain a more

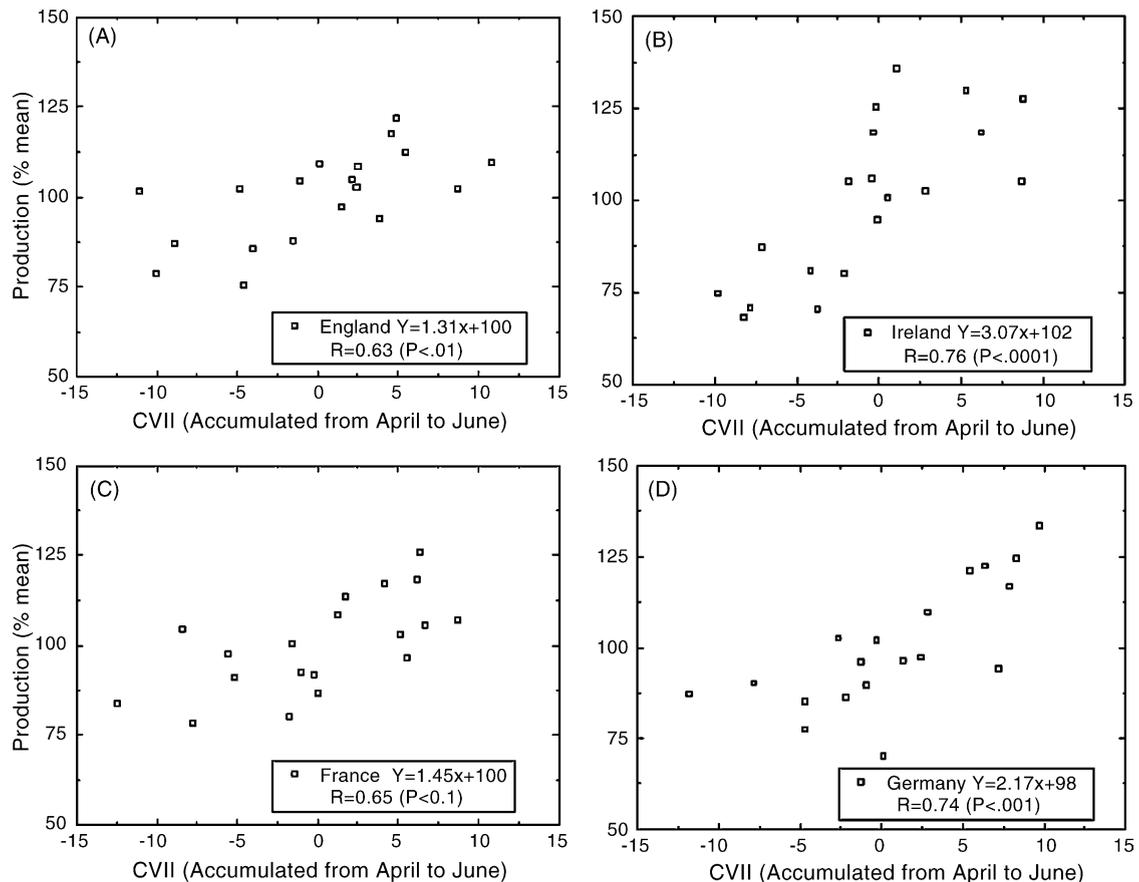


Fig. 3. Correlation between wheat production and Climate-Variability Impact Index (CVII) in England (A), Ireland (B), France (C), and Germany (D). GIMMS LAI (1982–2000) is used as a substitute to generate the CVII. The CVII is accumulated over the growing season and wheat production is normalized as the percent of the mean.

diverse mix of crop types with a large range in the rates of physiological development and growth phases by date. As a result, the satellite signals contain large variations in the composite of crop anomalies, which is not comparable with yield information for a single crop. Furthermore, when we calculate the CVII and production anomaly over a large area, the fluctuations tend to be reduced because the positive variations in a particular sub-region will cancel negative variations in another sub-region, which will further weaken the relationship. Yet, another factor may be the quality of the AVHRR data series itself. The 1982–2000 data series was assembled from data from four different satellites: NOAA-7 (1982–1984), NOAA-9 (1985–1988), NOAA-11 (1989–1994), and NOAA-14 (1995–2000). The NASA GIMMS group implemented corrections for residual sensor degradation and sensor intercalibration differences (Vermote and Kaufman, 1995; Los, 1998) and stratospheric aerosols (Vermote et al., 1997) in developing this data set. Nevertheless, some residual

noise persisted, and as a result CVII variability during this period may be partially related to non-vegetation factors.

Overall, however, the results in this section highlight that the Climate-Variability Impact Index, derived from both the MODIS and AVHRR LAI, can provide both fine-scale and aggregated spatial information on vegetation productivity and may serve as a diagnostic tool for yield estimation and monitoring.

4. Prediction of production with CVII

The previous section detailed the positive correlation between the CVII and crop production anomaly at different locations. This section examines how the relationship between these two variables can provide additional information to generate predictions of crop production from the LAI-based CVII.

Using the monthly 1-km MODIS LAI and production estimates from USDA, we fit linear models for the

Table 1
Linear model between crop production (dependent) and Climate-Variability Impact Index (independent) at district-level

	Unstandardized coefficients		Standardized coefficients	<i>t</i>	Significance
	<i>B</i>	S.E.			
Model 1					
Constant	1.011	0.016		63.238	<0.001
CVII	0.023	0.003	0.723	8.747	<0.001
Model 2					
Constant	1.005	0.022		46.622	<0.001
CVII	0.017	0.005	0.467	3.077	<0.001

The first model is generated from 72 samples from Illinois and North Dakota using both corn and spring wheat. The second model is generated from 36 corn samples from Illinois.

CVII and production anomaly from 2000 to 2003. The first model uses 72 samples calculated at the CRD-level from Illinois (corn) and North Dakota (spring wheat). The second model uses 36 samples calculated from Illinois CRDs only. The production anomaly is the dependent variable and the accumulated CVII over the growing season is the independent variable. We only model corn production here because we only have access to the 2004 production of corn from Illinois and not the 2004 production of spring wheat from North Dakota. At the same time in one of the models we include spring wheat data from North Dakota to see whether predictor coefficients generated from different crops significantly affects the estimation ability of the model. From Table 1, we note that the two models are similar and the coefficients are significantly different

than zero ($p < 0.001$). Using these models, the corn production of 2004 is predicted for the nine crop-reporting districts in Illinois (Fig. 4). Using the following equation, we calculated the correlation to test how closely the model predictions are related with the USDA estimates:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (4)$$

Here, x_i is the model prediction of CRD i , y_i the USDA estimates of CRD i , and \bar{x} and \bar{y} are the mean of model prediction and USDA estimates, respectively. Due to the data limitation, only nine CRDs are used in this prediction. This sample is biased as all the observations have positive anomalies. Since the expected sample mean is the population mean, we use the population mean (which by definition is one) as a substitute of the sample mean. The result represents how well the predictions and the estimates are fitted to the unity line.

In addition, we calculated the Heidke Skill Score (HSS; Wilks, 1995) to evaluate the relative accuracy of the model forecasts with respect to a random model:

$$HSS = \frac{C - E}{N - E} \quad (5)$$

Here, C is the number of correct forecasts, N is the total number of forecasts, and E is the number of forecasts expected to be correct by chance (which is $N/2$ in this case). For this case, we choose one (i.e. climatology) as the threshold to calculate the HSS. By

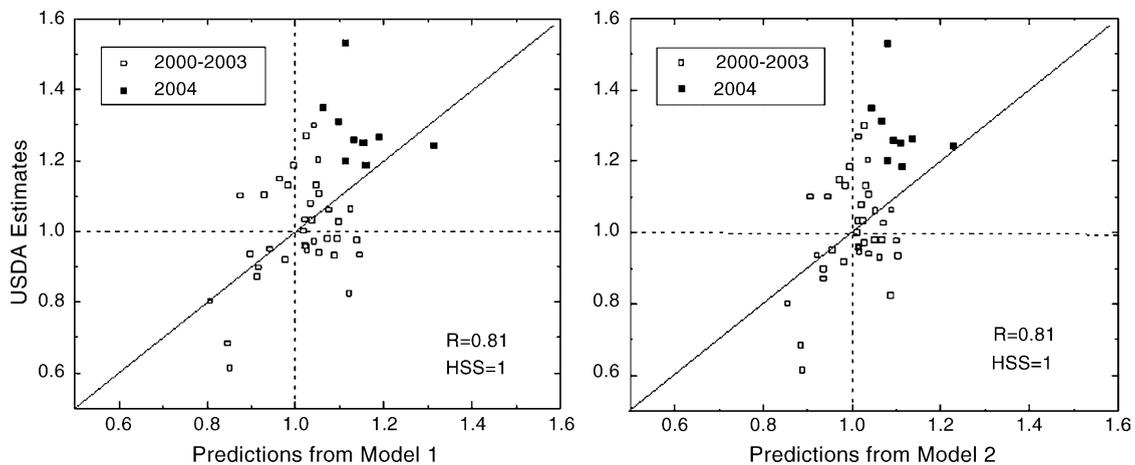


Fig. 4. Correlation between corn production from model predictions and USDA estimates at nine Illinois crop-reporting districts in 2004 (filled squares). The empty squares are model predictions vs. USDA estimates in 2000–2003 (the training period for the prediction model). Correlation (R) and Heidke Skill Score (HSS) are calculated using the equations in Section 4. Left panel is the result of linear model one, which is generated from 72 samples in both Illinois (corn) and North Dakota (spring wheat). Right panel is the result of linear model two, which is generated from 36 samples in Illinois (corn only).

definition, HSS represents the percentage improvement over a random forecast (Wilks, 1995). For a perfect forecast, HSS equals 1; for a totally random forecast, HSS equals 0. For this evaluation, a random forecast would be correct 50% of the time. Hence, HSS = 0.5 represents a hit rate of 75%.

In general, both models predict positive corn production anomalies at all CRDs, which is consistent with the USDA estimates (hence the HSS equals 1). In addition, both models and USDA estimates show that the 2004 production is the largest anomaly for each district (not shown). However, the model predictions still tend to underestimate the USDA values. Overall though, the correlation values demonstrate that the model predictions and the USDA estimates are well fitted to the unity line. The similar predictions from the two models again demonstrate that the CVII relationships with corn and spring wheat production are very similar at regional scales.

While only nine Illinois districts could be used in this study due to limited availability of USDA estimates, our predictions are consistent with the USDA estimates and suggest that the LAI-based CVII is a good predictor for the crop production. At the same time, some of the data limitations will be eliminated once more crop estimates become available in the future; at that point more robust tests of the predictive capabilities of the CVII will be possible.

We have demonstrated that the linear model generated from one study area can be used to predict the future crop production at the same location. Now we discuss whether a model generated from one location can be applied to predict crop production in another

Table 2

Linear model between winter wheat production (dependent) and Climate-Variability Impact Index (independent)

Model	Unstandardized coefficients		Standardized coefficients Beta	<i>t</i>	Significance
	<i>B</i>	S.E.			
Constant	0.977	0.027		35.676	<0.001
CVII	0.016	0.003	0.775	5.751	<0.001

location. Using the linear model generated from the winter wheat states in the United States, we calculate winter wheat production in the European countries. According to the results in Section 3.2, the dependent variable for the training data is the production anomaly for each winter wheat state from 1985 to 1990, and the independent variable is the cumulative CVII from April to June derived from AVHRR LAI. From Table 2, the coefficient is significantly different than zero ($p < 0.001$). Using this model, the wheat production from 1985 to 1990 is predicted for the four European countries (France, Germany, England, and Ireland). We also use the same model to predict the winter wheat production for a longer time period, from 1982 to 2000. To evaluate the model predictions, we use the previous equations to calculate the correlation and HSS. In general, the heterogeneous predictions are consistent with the FAO estimates (Fig. 5). The model predictions and FAO estimates are closely related to the unity line for both the short- and long-time periods. In addition, the HSS shows the model performance represents a 50–60% improvement over a random model.

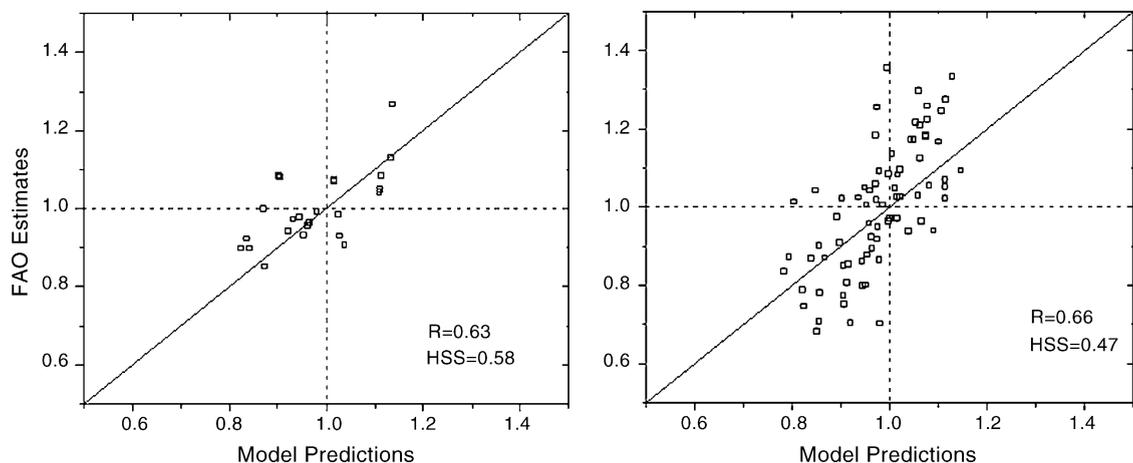


Fig. 5. Correlation between winter wheat production from model predictions and FAO estimates at four European countries (England, Ireland, France, and Germany). The linear model generated from US winter wheat states from 1985 to 1990 is applied to calculate the wheat production in European countries. Correlation (R) and Heidke Skill Score (HSS) are calculated using the equations in Section 4. Left panel is the prediction from 1985 to 1990. Right panel is the prediction from 1982 to 2000.

It should be noted that the estimated production calculated from this model is the anomaly from the mean, not the absolute production. The climatological mean is needed to obtain the actual production. However, even without this information, this model can potentially provide near real-time, global coverage of the percentage of the climatological crop production either gained or lost due to climate variations during the growing season.

5. Operational tool for agriculture monitoring

The previous sections demonstrated the correlation between LAI-based CVII and crop production for various regions. The following examines how the satellite data may be able to provide an operational tool for crop growth monitoring. To study the relation between monthly CVII and growing-season production, we examine the timing of the relationship between CVII and crop production. Here, we construct a statistical model for production: $Y = \alpha + \beta_i CVII_i + \mu$, in which $CVII_i$ is the Climate-Variability Impact Index in month i , Y the production anomaly during the study year, α and β_i are the regression coefficients, and μ is a regression residual. This linear equation is examined four times with different regression selection procedures. In the *ideal model*, the monthly CVII time-series are iteratively added into the model following a forward selection procedure; in the *cumulative model*, the CVII is progressively accumulated into one predictor for each month; in the *chronological model*, the CVII series for each month is added to the model

chronologically and the regressions coefficients can change with the addition of each new predictor; in the *fixed-coefficient model*, the CVII time-series are added in the model chronologically however the regression coefficients calculated from the pre-existing predictor variables are fixed.

5.1. Ideal model

To construct this model, the monthly CVII value most highly correlated with the dependent variable is the first variable to enter the model. If this regression is significant, we start to add the next variable with the largest partial F statistic unless the largest partial F is not statistically significant (Kleinbaum et al., 1998). Once we identify all the important predictors, we can determine which months of the growing season contribute the most *independent* information concerning the overall variation of crop production. As such, if 1 month's CVII value is well correlated with overall production anomalies but is also well correlated with other values of the CVII, it may not serve as an important predictor in that it provides redundant information already contained in previous predictors. In addition, the important predictors might be different for various crop types and study areas. For example, the most predictive months in Illinois and North Dakota are June, August, and September. As a result, although this model does not contain any redundant predictor, it is not a feasible monitoring tool for cross-crop or cross-region comparisons. However, the ideal model explains 60% of the variance in crop production (Fig. 6), which

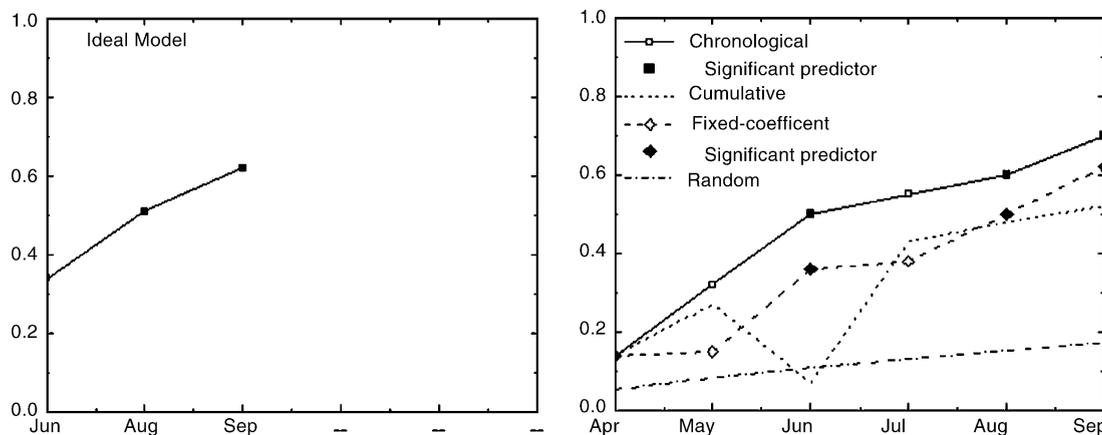


Fig. 6. Evolution of the R -square of four models as a function of the number of predictors. Left panel shows the result from the ideal model in Illinois and North Dakota at crop-reporting district-level. Only significant predictor months are shown—see text for details. Right panel shows the results from the chronological (solid line), fixed-coefficient (dash line), and cumulative (dot line) models. Filled markers indicate the monthly coefficients are significantly different from zero. The random R -square increase (dash-dot line) is used as a reference to determine whether the chronological/fixed-coefficient models are statistically significant (at 5% level).

provides the best correlation between the CVII and subsequent production using the fewest number of predictors.

5.2. Cumulative model

In the cumulative model, the CVII is progressively accumulated into one predictor month by month over the growing season. At any point in the growing season, the cumulative CVII represents the variations in the total growth up to that point.

$$Y = \alpha + \beta(n) \sum_1^n \text{CVII}_n + \mu \quad (6)$$

Note that the regression coefficient, β , is a function of the number of months, n , and hence will change over the course of the growing season. Previous studies reported a closer relation when cumulative rather than instantaneous vegetation indices are used because the accumulation represents the duration of photosynthetic activity (e.g. Tucker et al., 1980; Pinter et al., 1981). In this research, we used monthly MODIS LAI products to fit the linear model between cumulative CVII and crop production in Illinois and North Dakota. The cumulative model explains 50% of the variance in the crop production at the end of the growing season (Fig. 6). As more LAI data are added to the cumulative CVII, the explained variance does not monotonically increase until after June. This agrees with previous works that indicate the performance of the integrated metric is only optimal over a specific integration period (Atzberger, 1997).

5.3. Chronological model

In the chronological model, the CVII series for each month of the growing season is added to the model chronologically. As more independent variables are added, the coefficients of the pre-existing independent variables will change. In general, the larger the number of independent variables in the model, the more the variance of the dependent variable is explained by chance. In this manuscript, we used a random model to check the statistical significance of the chronological models. The random R -square tells us how much of the variance in the dependent variable can be explained by the same number of random independents. At the same time, while the overall model may explain more variance than a random one with the same number of predictors, some of the coefficients for the predictor variables may not be statistically significant, indicating

they do not provide any additional information. To identify the important predictors, a partial F statistic is calculated to test whether the addition of one particular predictor variable adds significantly to the prediction of Y achieved using the pre-existing predictor variables.

Fig. 6 indicates that the chronological model generated from monthly MODIS products can explain about 70% of the variance of crop production in Illinois and North Dakota at the CRD-level. In particular, CVIIs for August and September are statistically significant predictors, indicating the ideal prediction comes at the end of the growing season. However, the explained variance rises above 0.5 three months into the season, indicating significant predictability even during the course of the growing season.

5.4. Fixed-coefficient model

Similar to the third model, for this model the CVIIs are added in the model chronologically. However, the coefficients of the pre-existing independent variables are fixed when more variables are added. To do this, the additional predictor variable is regressed with the residual of Y instead of Y . Similarly, we used a partial F test to find the important predictors.

From Fig. 6, the fixed-coefficient models explain less variance in the crop production compared to the chronological one. Models derived from GIMMS LAI data for other regions provide similar results (see Fig. 7). However, as more remotely sensed data become available in the growing season, it is easier to add the latest CVII into the fixed model because the coefficients of the pre-existing CVIIs do not have to be recalculated as in the chronological model. Thus, the fixed model may serve as a better operational tool for crop monitoring compared with the chronological model while still explaining about as much variance during the latter part of the season.

5.5. Application of models for yield prediction

To evaluate how these models work on crop predictions of out-of-sample production, we used two sets of remotely sensed data, LAI and NPP, to generate the chronological and fixed-coefficient models based on corn production in five states, Minnesota, Michigan, Iowa, Indiana, and Wisconsin. Then, we apply the models to estimate the corn production in Illinois from 1982 to 1999. To get a robust evaluation, longer time period GIMMS products are used as a substitute because a 5-year cover of MODIS products cannot provide enough statistical information.

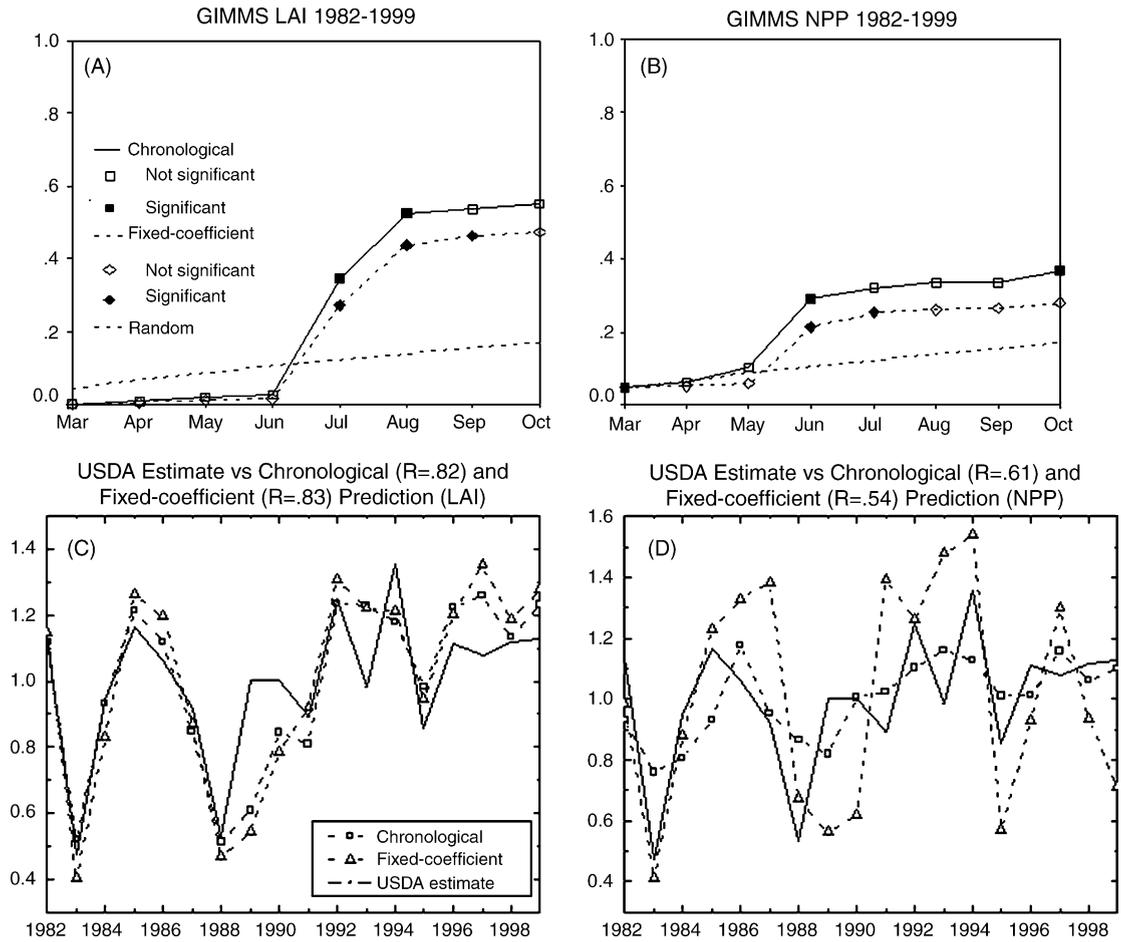


Fig. 7. The evolution of R -square of the chronological (squares) and fixed-coefficient (diamonds) models using GIMMS LAI (A) and NPP (B). The models are generated from five corn states, MN, MI, IA, IN, and WI, using the data from 1982 to 1999. Filled markers indicate the monthly coefficients are significantly different from zero. The USDA estimates of corn production in Illinois from 1982 to 1999 are then compared with the LAI model predictions (C) and NPP model predictions (D).

As expected, the fixed-coefficient models explain less variance in the corn production compared to the chronological one (Fig. 7A and B). In addition, models generated from NPP explain less variance in corn production than those from LAI data. The model predictions in Illinois are highly correlated with the USDA estimates ($R > 0.8$) from 1982 to 1999 for both the chronological model and the fixed-coefficient model. Although the Net Primary Production contains information about both vegetation properties and climate variations, the models generated from the NPP data are not as good as the models from the LAI data, at least for this case of corn yield prediction. We then compared the relationship between the monthly NPP and LAI from 1982 to 1999 in Illinois and Minnesota. Fig. 8 shows that NPP saturates as the LAI becomes larger. Thus, the NPP tends to plateau during

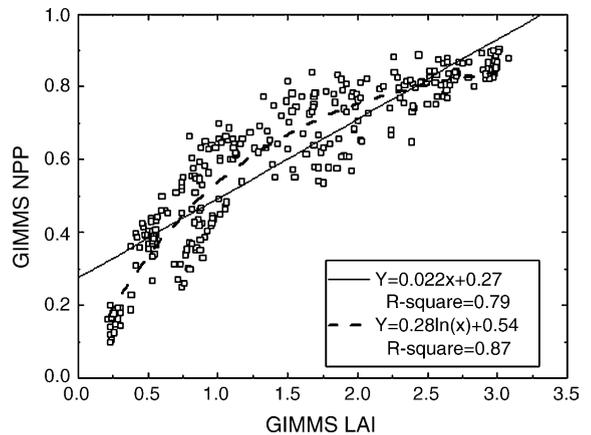


Fig. 8. Relationship between monthly GIMMS LAI and GIMMS NPP from 1982 to 1999 in Illinois and Minnesota.

the summer and contributes less to the variance in yields, while the LAI itself shows larger variability and can provide more specific information regarding the variations in vegetation productivity.

6. Conclusions

In this paper, the Climate-Variability Impact Index, defined as the monthly contribution to anomalies in annual growth, quantifies the percentage of climatological production either gained or lost due to climatic variability in a given month, with positive CVII values indicating production greater than the climatological average. Both fine-scale and aggregated information on vegetation productivity for various crop types can be obtained by examining the integrated CVII over the growing season. In general, about 60% of the variance in crop production could be explained by variations in CVII. The CVII is best correlated with crop production at regional scales when the variation of the crop production is large enough, e.g. there is homogenous behavior over a relatively large study area. At local (county-wide and smaller) scales, individual fluctuations in the remotely sensed data introduce excessive noise into the aggregate field; at much coarser resolution (state-wide and larger) the linear relationship again weakens due to regional differences in seasonal climate variations between disparate locations.

Once the relationship between the CVII and crop production is developed based on the historical record, the model can be applied to produce homogeneous yield forecasts (in which the model is trained and tested for a particular region), and heterogeneous yield forecasts (in which the model is trained in a particular region and applied to a different region). We used correlation and Heidke Skill Scores to evaluate how well the model performs on out-of-sample predictions. Results suggest that the CVII-based empirical model provides significant predictability for both the sign and magnitude of production variations over the training regions. Furthermore, the model derived using data from the United States also provides predictability for crop production in several European counties, which suggests that the single-crop CVII–production relationship may be quasi-independent of location. In addition, for certain crop types such as corn and spring wheat, the CVII–production relationship appears to be crop-independent as well (although this does not hold for winter wheat, for instance). Both of these results are encouraging because they suggest that the model may be applicable for regions even when historical production data are not available. As a result, the CVII-based

model can provide near real-time, global coverage of the percent change in the climatological crop yield (either gained or lost) in both types of forecast scenarios.

Lastly, the predictive value of the CVII is assessed by comparing the estimated production as a function of growing-season months. Several models were developed towards this goal. The ideal model, which contains no redundant predictors and provides the best correlation between CVII and production, is not a good monitoring tool for cross-crop or cross-region comparisons because the important predictors differ with crop types and study areas. The cumulative and chronological models could be potential tools for crop monitoring/forecast at various locations; here, we show that the chronological model provides satisfactory predictions before the end of the growing season, however both models require a matrix of coefficients that change with each subsequent CVII value added during the course of the growing season. A fixed-coefficient model, in which the coefficients from previous months remain constant as additional data is included in the model, serves well for operational crop monitoring. The results presented here also indicate that NPP is not as good a predictor as LAI, at least for corn production prediction in Illinois. One possible reason of this might be the saturation of the NPP during the high growth period. Overall, the high temporal and spatial resolution as well as the availability of the timely access to the needed MODIS products makes CVII a useful tool in near real-time crop growth monitoring.

Acknowledgements

We thank Dr. Tucker for providing the AVHRR GIMMS NDVI data and Dr. Nemani for providing the GIMMS NPP data. We also acknowledge X. Song and W. Yang for the AVHRR and MODIS LAI data sets and discussions. This work was funded by the NASA Earth Science Enterprise.

References

- Asrar, G., Myneni, R.B., Choudhury, B.J., 1992. Spatial heterogeneity in vegetation canopies and remote sensing of absorbed photosynthetically active radiation: a modeling study. *Remote Sens. Environ.* 41, 85–103.
- Atzberger, C., 1997. Estimates of winter wheat production through remote sensing and crop growth modelling. A case study on the Camargue region In: *Akademische Abhandlungen zu den Geowissenschaften*, Verlag für Wissenschaft und Forschung, Berlin, Germany, p. 267, ISBN 3-89700-013-X.

- Barnett, T.L., Thompson, D.R., 1983. Large-area relation of Landsat MSS and NOAA-6 AVHRR spectral data to wheat yields. *Remote Sens. Environ.* 13, 277–290.
- Colwell, J.E., Rice, D.P., Nalepka, R.F., 1977. Wheat yield forecast using Landsat data. In: *Proceedings of the 11th International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, pp. 1245–1254.
- Domenikiotis, C., Spiliotopoulos, M., Tsiros, E., Dalezios, N.R., 2004. Early cotton yield assessment by the use of the NOAA/AVHRR derived Vegetation Condition Index (VCI) in Greece. *Int. J. Remote Sens.* 25 (14), 2807–2819.
- Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* 83, 287–302.
- Goward, S.N., Markham, B., Dye, D.G., Dulaney, W., Yang, A.J., 1991. Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sens. Environ.* 35, 257–277.
- Gutman, G.G., 1991. Vegetation indices from AVHRR data: an update and future prospects. *Remote Sens. Environ.* 35, 121–136.
- Hatfield, J.L., 1983. Remote sensing estimators of potential and actual crop yield. *Remote Sens. Environ.* 13, 301–311.
- Hochheim, K.P., Barber, D.G., 1998. Spring wheat yield estimation for western Canada using NOAA NDVI data. *Can. J. Remote Sens.* 24 (1), 17–27.
- Holben, B.N., Tucker, C.J., Fan, C.-J., 1980. Spectral assessment of soybean leaf area and leaf biomass. *Photogram. Eng. Remote Sens.* 46, 651–656.
- Johnson, G.E., Achutuni, V.R., Thiruvengadachari, S., Kogan, F., 1993. The role of NOAA satellite data in drought early warning and monitoring: selected case studies. In: Wilhite, D.A. (Ed.), *Drought Assessment, Management, and Planning: Theory and Case Studies*. Kluwer Academic, pp. 31–49.
- Justice, D.H., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, R., Myneni, R., Knjazihhin, Y., Running, S., Nemani, R., Vermote, E., Townshend, J., Defries, R., Roy, D., Wan, Z., Huete, A., van Leeuwen, W., Wolfe, R., Giglio, L., Muller, J., Lewis, P., Barnsley, M., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* 36, 1228–1249.
- Karrenbrock, J.D., May/June 1989. The 1988 Drought: Its Impact on District Agriculture. The Federal Reserve Bank of St. Louis Review, pp. 3–13.
- Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A., 1998. *Applied Regression Analysis and Other Multivariable Methods*. Duxbury Press, Belmont, Calif, pp. 386–422.
- Kogan, F.N., 1990. Remote sensing of weather impacts on vegetation. *Int. J. Remote Sens.* 6, 1417–1434.
- Kogan, F.N., 1995. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bull. Am. Meteorol. Soc.* 76, 655–668.
- Kogan, F.N., 1997. Global drought watch from space. *Bull. Am. Meteorol. Soc.* 78, 621–636.
- LeComte, D., Cutlip, K., 2003. The weather of 2002. *Weatherwise* 56 (2), 14–22.
- Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agric. Ecosyst. Environ.* 94, 205–220.
- Los, S.O., 1998. Estimation of the ratio of sensor degradation between NOAA-AVHRR channels 1 and 2 from monthly NDVI composites. *IEEE Trans. Geosci. Remote Sens.* 36, 206–213.
- Monteith, J.L., 1977. Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. Lond. Ser. B* 281, 277–294.
- Moulin, S., Biondeau, A., Delecalle, R., 1998. Combining agricultural crop models and satellite observations: from field to regional scales. *Int. J. Remote Sens.* 19 (6), 1021–1036.
- Myneni, R.B., Nemani, R.R., Running, S.W., 1997. Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35, 1380–1393.
- Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B., Running, S.W., 2003. Climate driven increases in global net primary production from 1982 to 1999. *Science* 300, 1560–1563.
- Patel, N.K., Singh, T.P., Sahai, B., 1985. Spectral response of rice crop and its relation to yield and yield attributes. *Int. J. Remote Sens.* 6, 657–664.
- Pinter, P.J., Jackson, R.D., Idso, S.B., Reginato, R.J., 1981. Multidate spectral reflectance as predictors of yield in water stressed wheat and barley. *Int. J. Remote Sens.* 2 (1), 43–48.
- Price, J.C., 1993. Estimating leaf area index from satellite data. *IEEE Trans. Geosci. Remote Sens.* 31, 727–734.
- Rasmussen, M.S., 1992. Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from AVHRR. *Int. J. Remote Sens.* 13 (18), 3431–3442.
- Rudorff, B.F.T., Batista, G.T., 1990. Yield estimation of sugarcane based on agrometeorological-spectral models. *Remote Sens. Environ.* 33, 183–192.
- Running, S.W., Justice, C.O., Salomonson, V., Hall, D., Barker, J., Kaufmann, Y.J., Strahler, A.H., Huete, A.R., Muller, J.-P., Vanderbilt, V., Wan, Z.M., Teillet, P., Carneggie, D., 1994. Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *Int. J. Remote Sens.* 15 (17), 3587–3620.
- Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J., Randall, D.A., 1994. A global 1° by 1° NDVI data set for climate studies. Part 2: the generation of global fields of terrestrial biophysical parameters from the NDVI. *Int. J. Remote Sens.* 15, 3519–3545.
- Sharman, M., Boissezon, H., de Gonzales, G., Pous, B., 1992. Rapid estimates of acreages and potential yield at a European scale. In: *Proceedings of the 43rd Congress of the International Astronautical Federation*, Washington, DC, August 28–September 5.
- Tucker, C.T., Holben, B.N., Elgin, J.H., McMurtery, J.E., 1980. Relationship of spectral data to grain yield variation. *Photogram. Eng. Remote Sens.* 46 (5), 657–666.
- Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D., Pak, E.W., Mahoney, R., Vermote, E., El Saleous, N. An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.*, in press.
- Vermote, E.F., Kaufman, Y.J., 1995. Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views. *Int. J. Remote Sens.* 16, 2317–2340.
- Vermote, E., El Saleous, N., Kaufman, Y.J., Dutton, E., 1997. Data pre-processing: stratospheric aerosol perturbing effect on the remote sensing of vegetation: correction method for the composite NDVI after the Pinatubo Eruption. *Remote Sens. Rev.* 15, 7–21.
- White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to inter-annual climatic variability. *Global Biogeochem. Cycles* 11, 217–234.

- Wilks, D.S., 1995. *Statistical Methods in Atmospheric Sciences*, International Geophysics Series, vol. 59 Academic Press, pp. 248–250.
- Yang, W., Huang, D., Tan, B., Stroeve, J.C., Shabanov, N.V., Knyazikhin, Y., Nemani, R.R., Myneni, R., 2005. Analysis of leaf area index and fraction of PAR absorbed by vegetation products from terra MODIS sensor: 2000–2005. *IEEE Trans. Geosci. Remote Sens.* (in second review).
- Zhang, P., Anderson, B., Barlow, M., Tan, B., Myneni, R., 2004. Climate related vegetation characteristics derived from MODIS LAI and NDVI. *J. Geophys. Res.* 109, D20105 doi:10.1029/2004JD004720.
- Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* 84, 471–475.