Available online at www.sciencedirect.com
SCIENCE DIRECT®
Remote Sensing of Environment 99 (2005) 341 – 356
Remote Sensing of Environment
www.elsevier.com/locate/rse

Image masking for crop yield forecasting using AVHRR NDVI time series imagery☆

Jude H. Kastens [a,*], Terry L. Kastens [b], Dietrich L.A. Kastens [c], Kevin P. Price [d,a], Edward A. Martinko [e,f], Re-Yang Lee [g]

[a] *Kansas Applied Remote Sensing Program, located at the University of Kansas in Lawrence, Kansas (USA)*
[b] *Department of Agricultural Economics at Kansas State University in Manhattan, Kansas*
[c] *Kastens Inc. Farms, located in Herndon, Kansas*
[d] *Department of Geography at the University of Kansas*
[e] *Kansas Biological Survey located at the University of Kansas*
[f] *Department of Ecology and Evolutionary Biology at the University of Kansas*
[g] *Department of Land Management at Feng Chia University in Taichung, Taiwan*

Received 16 February 2005; received in revised form 12 September 2005; accepted 17 September 2005

Abstract

One obstacle to successful modeling and prediction of crop yields using remotely sensed imagery is the identification of image masks. Image masking involves restricting an analysis to a subset of a region's pixels rather than using all of the pixels in the scene. Cropland masking, where all sufficiently cropped pixels are included in the mask regardless of crop type, has been shown to generally improve crop yield forecasting ability, but it requires the availability of a land cover map depicting the location of cropland. The authors present an alternative image masking technique, called yield-correlation masking, which can be used for the development and implementation of regional crop yield forecasting models and eliminates the need for a land cover map. The procedure requires an adequate time series of imagery and a corresponding record of the region's crop yields, and involves correlating historical, pixel-level imagery values with historical regional yield values. Imagery used for this study consisted of 1-km, biweekly AVHRR NDVI composites from 1989 to 2000. Using a rigorous evaluation framework involving five performance measures and three typical forecasting opportunities, yield-correlation masking is shown to have comparable performance to cropland masking across eight major U.S. region-crop forecasting scenarios in a 12-year cross-validation study. Our results also suggest that 11 years of time series AVHRR NDVI data may not be enough to estimate reliable linear crop yield models using more than one NDVI-based variable. A robust, but sub-optimal, all-subsets regression modeling procedure is described and used for testing, and historical United States Department of Agriculture crop yield estimates and linear trend estimates are used to gauge model performance.

Keywords: Image masking; Crop yield forecasting; AVHRR; NDVI; Time series

1. Introduction

Time series of the Advanced Very High-Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) have been used for crop yield forecasting since the 1980s. Image masking was a critical component of several of these yield forecasting efforts as researchers attempted to isolate subsets of a region's pixels that would improve their modeling results. Approaches generally sought to identify cropped pixels and when possible, pixels that corresponded to the particular crop type under investigation. In this paper, the former approach will be referred to as *cropland masking*, and the latter approach will be called *crop-specific masking*.

The research presented in this paper examines the underlying assumptions made when image masking for the purpose of regional crop yield forecasting. An alternative statistical image masking approach (called *yield-correlation masking*) is pro-

☆ Funding for this research was provided in part by NASA Earth Science Enterprise, Applications Program, grant numbers NAG5-4990, NAG5-9841, NAG5-11988, and NAG13-99009, and by USDA Small Business Innovative Research Program, grant agreement numbers 99-33610-7495 and 00-33610-9453.
* Corresponding author.
E-mail address: jkastens@ku.edu (J.H. Kastens).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2005.09.010

posed that is objective (i.e., can be automated) and has the flexibility to be applied to any region with several years of time series imagery and corresponding historical crop yield information. The primary appeal of yield-correlation masking is that, unlike cropland masking, no land cover map is required, yet we will show that NDVI models generated using the two methods demonstrate comparable predictive ability.

The goal of this research is very specific: We will establish the yield-correlation masking procedure as a viable image masking technique in the context of crop yield forecasting. To accomplish this objective, we empirically evaluate and compare cropland masking and yield-correlation masking for the purpose of crop yield forecasting. Cropland masking has been shown to benefit yield forecasting models, thus providing a practical benchmark. In the process, we present a robust statistical yield forecasting protocol that can be applied to any (region, crop)-pair possessing the requisite data, and this protocol is used to evaluate the two masking methods that are being compared.

This paper is developed as follows. A brief review of related research is presented. The two primary data sets, AVHRR NDVI time-series imagery and United States Department of Agriculture (USDA) regional crop yield data, are described. Details of the study regions, crop types, and time periods under investigation are specified. Three image masking procedures are discussed, two of which are evaluated in the research. Finally, the modeling approach and performance evaluation framework are described, along with a summary of results and conclusions drawn from the analysis. Details of the modeling strategy used are described in the Appendix.

2. Related research

Traditionally, yield estimations are made through agro-meteorological modeling or by compiling survey information provided throughout the growing season. Yield estimates derived from agro-meteorological models use soil properties and daily weather data as inputs to simulate various plant processes at a field level (Wiegand & Richardson, 1990; Wiegand et al., 1986). At this scale, agro-meteorological crop yield modeling provides useful results. However, at regional scales these models are of limited practical use because of spatial differences in soil characteristics and crop growth-determining factors such as nutrition levels, plant disease, herbicide and insecticide use, crop type, and crop variety, which would make informational and analytical costs excessive. Additionally, Rudorff and Batista (1991) indicated that, at a regional level, agro-meteorological models are unable to completely simulate the different crop growing conditions that result from differences in climate, local weather conditions, and land management practices. The scale of applicability of agro-meteorological models is getting larger, though, but presently only through the integration of remotely sensed imagery. For instance, Doraiswamy et al. (2003) developed a method using AVHRR NDVI data as proxy inputs to an agro-meteorological model in estimating spring wheat yields at county and sub-county scales in the U.S. state of North Dakota.

In the past 25 years, many scientists have utilized remote sensing techniques to assess agricultural yield, production, and crop condition. Wiegand et al. (1979) and Tucker et al. (1980) first identified a relationship between the NDVI and crop yield using experimental fields and ground-based spectral radiometer measurements. Final grain yields were found to be highly correlated with accumulated NDVI (a summation of NDVI between two dates) around the time of maximum greenness (Tucker et al., 1980). In another experimental study, Das et al. (1993) used remotely sensed data to predict wheat yield 85–110 days before harvest in India. These early experiments identified relationships between NDVI and crop response, paving the way for crop yield estimation using satellite imagery.

Rasmussen (1992) used 34 AVHRR images of Burkina Faso, Africa, for a single growing season to estimate millet yield. Using accumulated NDVI and statistical regression techniques, he found strong correlations between accumulated NDVI and yield, but only during the reproductive stages of crop growth. The lack of a strong correlation between accumulated NDVI and yield during other stages of growth was attributed to the limited temporal profile of imagery used in the study and the high variability of millet yield in his study area. Potdar (1993) estimated sorghum yield in India using 14 AVHRR images from the same growing season. He was able to forecast actual yield at an accuracy of ±15% up to 45 days before harvest. Rudorff and Batista (1991) used NDVI values as inputs into an agro-meteorological model to explain nearly 70% of the variation in 1986 wheat yields in Brazil. Hayes and Decker (1996) used AVHRR NDVI data to explain more than 50% of the variation in corn yields in the United States Corn Belt. Each of these studies found positive relationships between crop yield and NDVI, but the strength of the relationships depended upon the amount and quality of the imagery used.

Some studies have used large, multi-year AVHRR NDVI data sets. Maselli et al. (1992) found strong correlations between NDVI and final crop yields in the Sahel region of Niger using 3 years of AVHRR imagery (60 images). In India, Gupta et al. (1993) used 3 years of AVHRR data to estimate wheat yields within ±5% up to 75 days before harvest. The success of this study was dependent on the fact that over 80% of the study area was covered with wheat. In Greece, 2 years of AVHRR imagery were used to estimate crop yields (Quarmby et al., 1993). Actual harvested rice yields were predicted with an accuracy of ±10%, and wheat yields were predicted with an accuracy of ±12% at the time of maximum greenness. Groten (1993) was able to predict crop yield with a ±15% estimation error 60 days before harvest in Burkina Faso using regression techniques and 5 years of AVHRR NDVI data (41 images). Doraiswamy and Cook (1995) used 3 years of AVHRR NDVI imagery to assess spring wheat yields in North and South Dakota in the United States. They concluded that the most promising way to improve the use of AVHRR NDVI for estimating crop yields at regional scales would be to use larger temporal data sets, better crop masks, and climate data. Lee et al. (1999) used a 10-year, biweekly AVHRR data set to forecast

corn yields in the U.S. state of Iowa. They found that the most accurate forecasts of crop yield were made using a cropland mask and measurements of accumulated NDVI. Maselli and Rembold (2001) used multi-year series of annual crop yields and monthly NDVI to develop cropland masks for four Mediterranean African countries. They found that application of the derived cropland masks improved relationships between NDVI and final yield during optimal yield prediction periods. Ferencz et al. (2004) found yields of eight different crops in Hungary to be highly correlated with optimized, weighted seasonal NDVI sums using 1-km AVHRR NDVI from 1996 to 2000. They used non-forest vegetation masks and a novel time series interpolation approach and actually obtained their best results when using a greenness index equivalent to the numerator of the NDVI formula (NIR-RED; see Section 3).

Additionally, many researchers have found that crop condition and yield estimation are improved through the inclusion of metrics that characterize crop development stage (Badhwar & Henderson, 1981; Groten, 1993; Kastens, 1998; Lee et al., 1999; Quarmby et al., 1993; Rasmussen, 1992). Ancillary data have been found useful as well. For example, Rasmussen (1997) found that soil type information improved the explanation of millet and ground nut yield variation using 3 years of AVHRR NDVI from the Peanut Basin in Senegal. In a later study, Rasmussen (1998) found that the inclusion of tropical livestock unit density further improved the explanation of millet yield variation in intensively cultivated regions of the Peanut Basin.

Based on the studies described, for the purpose of crop yield forecasting, longer time series of NDVI imagery are preferred to shorter ones. Also, few image masking techniques have been thoroughly and comparatively explored, likely due to the inherent complexities underlying this phase in any remote sensing-based yield forecasting methodology. Thus, to help achieve the goal of this project, an important objective of this research is to use historical yield information and historical time series AVHRR NDVI imagery to devise a thorough and robust statistical procedure for obtaining early to mid-season crop yield forecasts, with particular emphasis on image masking. The techniques described in this paper can be applied to any (region, crop)-pair that possesses sufficient historical yield information and corresponding time series NDVI imagery. Since few meaningful crop phenology metrics can be accurately derived at early points in the growing season, our research does not attempt to use this information. Also, no ancillary information is used, to prevent dependence on the availability of such data.

3. Description of data

The research presented in this paper relies on two data sets. The first is a time series of biweekly AVHRR NDVI composite imagery from 1989 to 2000, obtained from the U.S. Geological Survey Earth Resources Observation Systems (EROS) Data Center (EDC) in Sioux Falls, SD (Eidenshink, 1992). This data set was chosen because it is relatively inexpensive, reliable, and is updated in near real-time. NDVI is defined by the formula (NIR-RED)/(NIR+RED), where NIR is reflectance in the near-infrared spectrum (0.75–1.10 μm) and RED is reflectance in the red band of the visible spectrum (0.58–0.68 μm). Chlorophyll uses electromagnetic energy in the RED band for photosynthesis, and plant structure is reflective of energy in the NIR band. So, for vegetated surfaces, NDVI increases if plant biomass increases or if photosynthetic activity increases.

The NDVI data were received in unsigned 8-bit integer format, with the original NDVI range [−1,1] linearly scaled to the integer range 0–200. For analysis purposes, the integer data were rescaled to their native range of [−1,1]. As a consequence of the limited precision of the original 8-bit data, the precision of the rescaled data is 0.01, so there is an implicit expected numerical error of 0.005 in the pixel-level NDVI values.

The NDVI data set is not without uncertainty, both temporal and spatial. From 1989–2000, two polar orbiting National Oceanic and Atmospheric Administration (NOAA) satellites (NOAA-11 [1989–1994] and NOAA-14 [1995–2000]) carried the AVHRR sensors that collected the data comprising our data set. The U.S. annually invariant target curve is displayed in Fig. 1, with NDVI from periods 5–21 (February 26–October 21) shown for each year. This curve represents the average time series of nearly 3500 pixels selectively sampled from the 48 states in the conterminous U.S. to possess highly regular annual periodicity, thus exposing any artificial interannual NDVI value drift (Kastens et al., 2003). The NDVI data originating from NOAA-11 are fairly consistent over time. The data from NOAA-14 are less so, exhibiting a large artificial oscillation from 1997 to 2000. The range of the trend curve from 1989 to 2000 has width 0.0464. Comparing this width to the overall effective range of the AVHRR NDVI data being used (which is approximately [−0.05,0.95] for the full U.S. terrestrial range, but narrower in most practical situations), it follows that nearly 5% of the effective AVHRR NDVI data range can be attributed to artificial interannual NDVI value drift. In retrospect, we know that sensor orbit decay and sensor calibration degradation were the primary sources of the interannual NDVI value drift found in the NOAA-14 data.

Image resolution (1 km^2/pixel, or 100 ha/pixel) of the AVHRR NDVI data is also an issue because pixel size is more

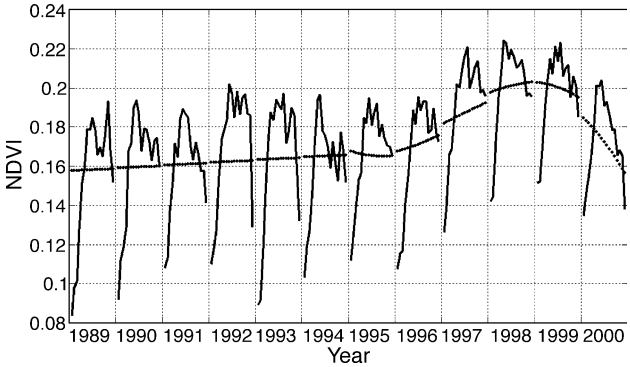

Fig. 1. Annually invariant target curve for U.S. AVHRR NDVI, 1989–2000. A linear trend line is depicted for 1989–1994, which corresponds to NOAA-11 data. A cubic trend curve is depicted for 1995–2000, which corresponds to NOAA-14 data. Only biweekly periods 5 to 21 are shown for each year, accounting for the yearly breaks seen in the curves.

than twice as large as the typical field size for soybeans and major grains in the U.S., which is roughly 40 ha (Kastens & Dhuyvetter, 2002). Furthermore, when considering spatial error of the image registration performed during the NDVI compositing process, the area of the region from which a single pixel's values can be obtained grows to more than 4 km^2, or more than 400 ha (Eidenshink, 1992). A combination of sensor factors (e.g., sensor stability, view angle, orbit integrity) and effects of image pre-processing and compositing induce this spatial variation.

The second data set is historical, final, state-level yield data, obtained from the USDA National Agricultural Statistics Service (NASS) through its publicly accessible website (http://www.usda.gov/nass). The database is updated annually for all crops, with each particular crop's final regional yield estimates released well after harvest completion. Updates to the final regional yield estimates can occur up to 3 years after their initial release, but generally these changes are not large. No historical or expected error statistics for these estimates are published below the national spatial scale, but they are nonetheless accepted by the industry as the best widely available record for average regional crop yield in the U.S.

4. Description of crops, regions, and time periods under investigation

The crops and regions under investigation in the present research are corn and soybeans in the U.S. states of Iowa (IA) and Illinois (IL), winter wheat and grain sorghum in the state of Kansas (KS), and spring wheat and barley in the state of North Dakota (ND). The locations of these states in the U.S. are shown in Fig. 2. Compared to other states, during 1989–2000, Iowa ranked first in corn production (100.2 million mt/year; "mt"=metric ton) and second in soybean production (26.9 million mt/year). Illinois ranked second in corn production (89.9 million mt/year) and first in soybean production (27.1 million mt/year). Kansas was the top-producing winter wheat state (25.5 million mt/year) and the top-producing grain sorghum state (14.2 million mt/year). North Dakota was the top-producer of both spring wheat (16.6 million mt/year) and barley (6.3 million mt/year).

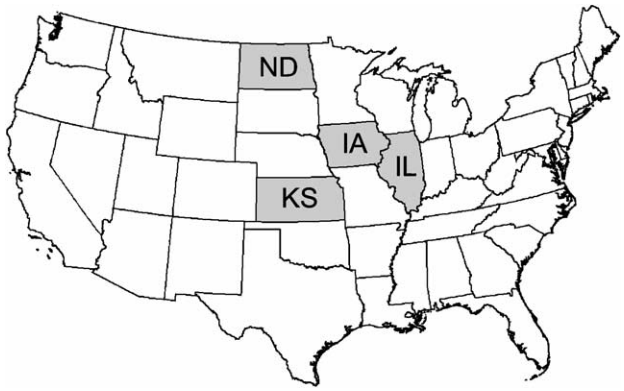

Fig. 2. Location of the study states in the conterminous United States. Iowa is designated by IA, Illinois by IL, Kansas by KS, and North Dakota by ND.

For each crop, a six-period window of early to mid-season NDVI imagery is considered. The source data for these six biweekly composites span nearly 3 months of raw AVHRR imagery, corresponding to Julian biweekly periods 5–10 (approximately February 26–May 20) for winter wheat, 9–14 (approximately April 23–July 15) for spring wheat and barley, and 11–16 (approximately May 21–August 12) for corn, soybeans, and sorghum. Labeling the six biweekly periods 1 to 6, yields are modeled using data from periods 1–4, 1–5, and 1–6, with each of these three ranges providing a unique yield forecasting opportunity corresponding to a different point in the growing season. To obtain the dates for the crop-specific ranges, the initial release dates of USDA NASS yield forecasts were considered. The first NDVI image generated after the release of the initial USDA state-level estimates for the season is assigned to period 6, which fixes periods 1–5 as well. Initial release dates for USDA state level estimates are approximately May 11 for winter wheat, July 11 for spring wheat and barley, and August 11 for corn, soybeans, and grain sorghum. Hereafter, winter wheat will be classified as an early-season crop, spring wheat and barley as mid-season crops, and corn, soybeans, and grain sorghum as late-season crops. With this timing, forecasts generated at periods 4 and 5 for each crop are produced before any state-level USDA yield estimates are released.

5. Approaches to image masking in crop yield forecasting

The purpose of image masking in the context of crop yield forecasting is to identify subsets of a region's pixels that lead to NDVI variable values that are optimal indicators of a particular crop's final yield. A good image mask should capture the essence (i.e., salient features) of the present year's growing season with respect to how the crop of interest is progressing. This growing season essence is a combination of climatic and terrestrial factors.

5.1. Crop-specific masking

In theory, the ideal approach to image masking for the purpose of crop yield forecasting would be to use crop-specific masking. This would allow one to consider only NDVI information pertaining to the crop of interest. However, when such masking is applied to multiple years of imagery, several difficulties are encountered. Principal among these is the widespread practice of crop rotation, which suggests that year-specific masks are needed rather than a single crop-specific mask that can be applied to all years. Regional trending in crop area (increase or decrease in the amount of a region's area planted to a particular crop over time), if severe enough, also may call for year-specific masking. Identifying a particular crop in the year to be forecasted presents even greater difficulties, as only incomplete growing season NDVI information is available. This is especially true early in the season when the crop has low biomass and does not produce a large NDVI response. In addition to hindering crop classification, this low NDVI response of a crop early in its development also

stifles crop yield modeling efforts, as AVHRR NDVI measurements from pixels corresponding to immature crops are not very sensitive and are thus minimally informative (see Wardlow et al., in press, for an example of such insensitivity occurring with 250-m Moderate Resolution Imaging Spectroradiometer [MODIS] NDVI data, and MODIS has better radiometric resolution than AVHRR).

Moreover, with the coarse-resolution (about 100 hectares/pixel) AVHRR NDVI imagery used in this study, identifying monocropped pixels becomes an improbable task. This is particularly true in low-producing regions and in regions with sparse crop distribution. As noted, a single pixel covers an area well over twice the average field size, and when error of the image registration is considered, a pixel's effective ground coverage can become more than 400 hectares/pixel, or roughly ten times the typical field size.

5.2. Cropland masking

A more feasible alternative to crop-specific masking is cropland masking, which refers to using pixels dominated by land in general agricultural crop production. Kastens (1998, 2000) and Lee et al. (1999) obtained some of their best yield modeling results using this approach. Rasmussen (1998) used a percent-cropland map to improve his yield modeling by splitting the data into two categories based on cropland density and building different models for the two classes. Maselli and Rembold (2001) used correlations between 13 years of monthly NDVI composites and 13-year series of national crop yields to estimate pixel-level cropland fraction for four Mediterranean African countries. Upon application of these derived cropland masks, the authors found improved relationships between NDVI and final estimated yield.

Cropland masks usually are derived from existing land use/land cover maps (one exception being Maselli and Rembold (2001)). If relatively small amounts of land in a study area have been taken out of or put into agricultural crop production during a study period, a single mask can be obtained and applied to all years of data. Considering that all traditional agricultural crops are now lumped in the general class of "cropland," heavily cropped pixels are more prevalent in heavily cropped regions, which allows for the construction of well-populated masks dominated by cropland. But the generation of such masks becomes difficult when low-producing regions are encountered, as well as in regions where cropland is widely interspersed with non-cropland.

As with crop-specific masking, cropland masking also can suffer the effects of minimally informative NDVI response early in a crop's growing season (e.g., March for early-season crops, May for mid-season crops, and late May and early June for late-season crops). Many important agricultural regions are almost completely dominated by single-season crop types (e.g., Iowa produces predominately late-season crops). In such cases, cropland AVHRR NDVI from time periods early in the particular growing season may not be very useful for predicting final yield.

By late May and June, some of the year's terrestrial and weather-based growth-limiting factors may have already been established for some crops or regions. For instance, in the U.S., soil moisture has largely been set by this time, and soil moisture is an important determinant of crop yields in the four states comprising our study area. Such moisture information is not readily detectable in immature crops because it is usually not a limiting factor until the plant's water needs become significant and its roots penetrate deeper into the soil. On the other hand, available soil moisture can noticeably affect other regional vegetation that is already well developed, such as grasslands, shrublands, and wooded areas, and in some cases, early- and mid-season crops.

5.3. Yield-correlation masking

For the reasons noted above, we propose a new masking technique, which we call yield-correlation masking. All vegetation in a region integrates the season's cumulative growing conditions in some fashion and may be more indicative of a crop's potential than the crop itself. Thus, all pixels are considered for use in crop yield prediction. This premise is most sound early in a crop growing season (especially for mid- and late-season crops), when the NDVI response of the immature crop is not yet strong enough to be a useful indicator of final yield. Also, as noted, when the crop is in early growth stages, problems such as lack of subsoil moisture may not yet have impacted the immature crop while having already affected more mature nearby vegetation.

Each NDVI-based variable captures a different aspect of the current growing season. This aspect manifests itself in different ways within the region's vegetation, suggesting that optimal masks for the different NDVI-based variables likely will not be identical. Thus, for each (region, crop)-pair, yield-correlation masking generates a unique mask for each NDVI variable. The technique is initiated by correlating each of the historical, pixel-level NDVI variable values with the region's final yield history, a strategy similar to the initial step of the cropland classification strategy presented in Maselli and Rembold (2001). The highest correlating pixels, thresholded so that some pre-specified number of pixels is included in the mask (this issue is addressed later), are retained for further processing and evaluation of the variable at hand. Fig. 3 shows a diagram outlining this process for a single variable.

Though much more computationally intensive, the yield-correlation masking technique overcomes the major problems afflicting crop-specific masking and cropland masking. Unlike these approaches, yield-correlation masking readily can be applied to low-producing regions and regions possessing sparse crop distribution. Also, since yield correlation masks are not constrained to include pixels dominated by cropland, they are not necessarily hindered by the weak and insensitive NDVI responses exhibited by crops early in their respective growing seasons. Furthermore, once the issue of identifying optimal mask size (i.e., determining how many pixels should be

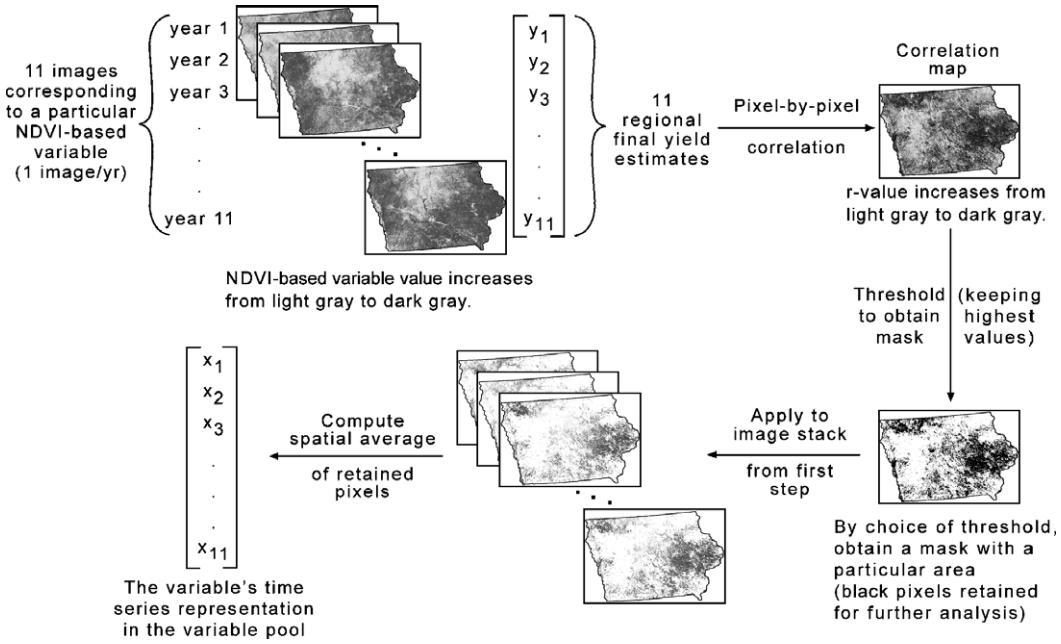

Fig. 3. Flowchart for a single-variable application of the yield-correlation masking technique for a single choice of mask size. Example shown is for Iowa corn using the NDVI variable obtained by accumulating values across periods 3 and 4 during the 11-year span 1989–1999 (and thus would have been used for the prediction of Iowa corn yields when 2000 was the 'out year'). A different variable vector will be obtained for each unique mask size choice.

included in the masks) is addressed, the entire masking/modeling procedure becomes completely objective.

6. Description of cropland masks

For Iowa, Illinois, and North Dakota, the cropland masks used in this study were derived from the United States Geological Survey (USGS) National Land Cover Database (NLCD) (Vogelmann et al., 2001). The original 30-m resolution land cover maps can be obtained from the website http://landcover.usgs.gov/natllandcover.html. After generalizing the classes to cropland and non-cropland, the data were aggregated to a 1-km grid corresponding to the NDVI imagery used in this study. All annual crops, as well as alfalfa, were assigned to the cropland category, and all other cover types were classified as non-cropland. Pixel values in the resulting map corresponded to percent cropland within the 1-km^2 footprint of the pixel. Fig. 4 shows the percent-cropland map that was used for Iowa. For Kansas, a 30-m land cover map produced by the Kansas Applied Remote Sensing (KARS) Program (Egbert et al., 2001) was used. As with the other state land cover maps, the classes were reassigned to the categories of cropland and non-cropland, and the data were then aggregated to a 1-km grid. Fig. 5 shows the percent-cropland map that was used for Kansas.

7. Description of framework for statistical evaluation of performance

Twelve years (1989–2000) of NDVI imagery form the backbone of this study. Thirty-one years (1970–2000) of final yield data were acquired so that any persistent linear trends in

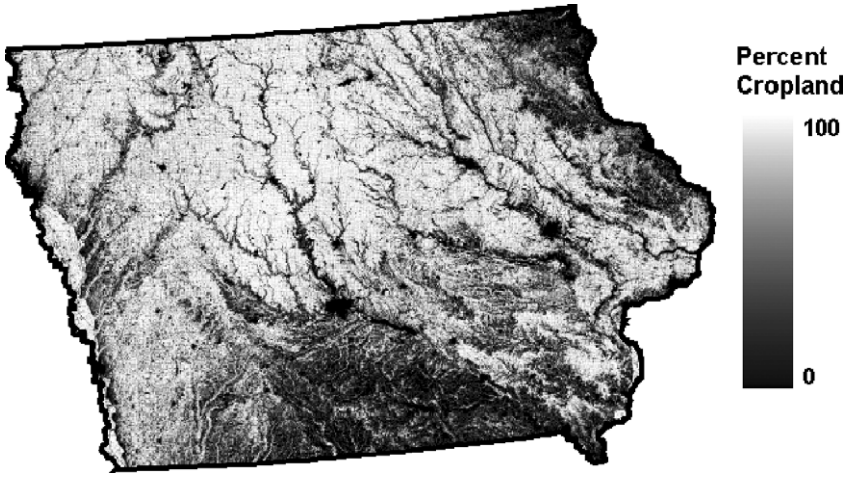

Fig. 4. Percent cropland map for Iowa.

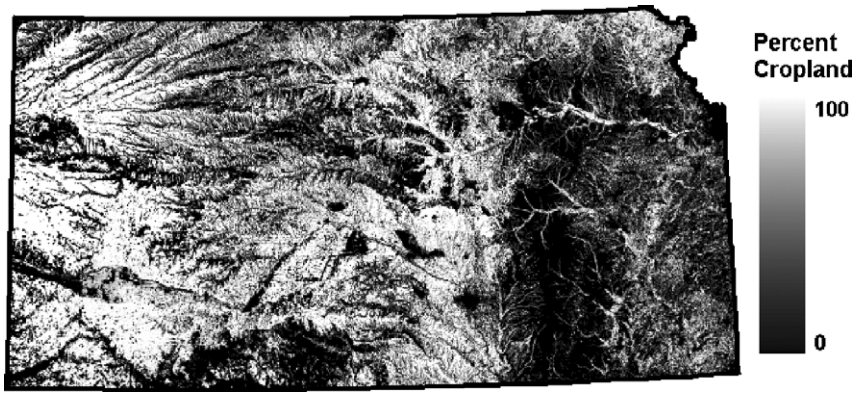

Fig. 5. Percent cropland map for Kansas.

yields could be identified and accounted for, providing a robust performance benchmark. 1970 is approximately the time point by which "modern" farming practices had begun across the entire study area, in particular, the use of herbicides, artificial fertilizers, and hybrid seeds. Thereafter, increased adoption by farmers, coupled with continual improvements in the technologies, induced temporally increasing crop yields. Table 1 shows summary statistics for the yield time series that were used in this study. Statistics from three time periods are given: the extended yield period (1970–2000), the pre-study period (1970–1988), and the study period (1989–2000). Increases in yields over the time spans considered are likely the direct result of technological advances in crop varieties, crop inputs, and farming practices.

Table 1
Crop statistics for year spans 1970–2000, 1970–1988, and 1989–2000

Region and crop	Years	Average yield	Standard deviation	Mean squared deviation from 1970–2000 trend line
IA Corn	1970–2000	7.29	1.395	1.040
	1970–1988	6.70	1.131	1.005
	1989–2000	8.24	1.274	1.092
IA Soybeans	1970–2000	2.58	0.381	0.255
	1970–1988	2.38	0.265	0.234
	1989–2000	2.89	0.329	0.286
IL Corn	1970–2000	7.41	1.385	1.056
	1970–1988	6.80	1.282	1.186
	1989–2000	8.37	0.939	0.806
IL Soybeans	1970–2000	2.50	0.358	0.250
	1970–1988	2.31	0.312	0.299
	1989–2000	2.80	0.165	0.141
KS Winter Wheat	1970–2000	2.33	0.410	0.371
	1970–1988	2.26	0.292	0.270
	1989–2000	2.45	0.542	0.491
KS Sorghum	1970–2000	3.82	0.795	0.603
	1970–1988	3.52	0.734	0.612
	1989–2000	4.29	0.664	0.589
ND Spring Wheat	1970–2000	1.94	0.377	0.341
	1970–1988	1.82	0.359	0.347
	1989–2000	2.12	0.339	0.330
ND Barley	1970–2000	2.44	0.503	0.418
	1970–1988	2.26	0.489	0.451
	1989–2000	2.74	0.380	0.359

All statistics are in metric tons per hectare yield units.

All results presented were derived from out-of-sample evaluation using a delete-1 cross-validation, or CV(1), framework. Twelve years of NDVI and final yield information were available for each (region, crop)-pair under study, and each year was treated at some point in the analysis as an out-of-sample year. Yields were linearly detrended (also in a CV(1) framework) in an attempt to reduce the impact of temporal order on the analysis, which supports the use of the CV(1) framework.

A single step of the CV(1) is described as follows. From the 12-year NDVI and final yield data sets, 1 year (called the *out year*) was set aside, and the remaining 11 years of information were used for both mask and model determination. No information from the out year entered into the masking or modeling process so that true out-of-sample forecasts were obtained for the out year. The sole exception to this claim regards the original land use/land cover maps that were used in the cropland masking phase of the analysis. These maps were generated during the study period. This slightly favors their use in the analysis, but the effect should be quite small given that the landscapes of the study states did not vary substantially from 1989 to 2000.

The linear detrending of yields was done separately in each step of the CV(1) exercise by leaving a hole in the yield time series at the location of the out year. The resulting trend line was then evaluated at the out year to serve as a benchmark forecast (referred to as the linear trend forecast). Models targeted deviation from this trend line, and model output was added to the CV(1) trend estimate to generate an actual yield estimate. Given that yield information from 1970 to 2000 was used, at each step, there were 30 years of data from which a trend could be extracted. Because this longer time series was used for the yields, the resulting linear trends varied little across all steps of the CV(1) exercise. Consequently, the CV(1) root mean squared error realized using the linear trend was not much greater than the standard deviation of the data from 1989 to 2000 after removing the 31-year trend line.

7.1. Mask application and evaluation

Once maps are available that can be thresholded to create an image mask, the arbitrary decision remains regarding how

many pixels to include in the mask. In crop-specific masking, this quantity is typically guided by historical areal coverage of the crop of interest, but this is not necessarily justified once one moves away from this masking method. To explore the relationship between model performance and mask area, we developed and evaluated models using several different mask sizes.

All percent-cropland and yield-correlation maps were evaluated for yield-prediction ability using many different thresholds determining mask inclusion/exclusion sets. For the percent-cropland maps, each integer percent from 0 to 100 was used as an inclusion/exclusion threshold and determined a mask for evaluation. Thus, the first mask evaluated contained all pixels that were $\geq$0% cropland, the second contained all pixels that were $\geq$1% cropland, and so on. This provided a dense but irregularly spaced sampling of mask sizes considered as a percent of the study region's total area. The yield-correlation maps, on the other hand, were thresholded such that a complete set of masks ranging in integer sizes from 0% to 100% of total region area was examined. For simplicity of analysis of the yield-correlation masking approach, at each percent of total area evaluation step, all of the variable-specific yield-correlation maps were thresholded to include the same number of pixels in the tested masks. Note that the 0% cropland threshold mask and the 100% of total area yield-correlation mask both corresponded to applying a mask with no pixels excluded, which is the same as applying no mask at all. We call this the "no mask" case.

7.2. Modeling approach

First- and second-order NDVI variables were used in this study. First-order NDVI variables are defined to be running sums of NDVI values during the six periods under study for each (region, crop)-pair, and second-order variables consist of squares and pair-wise products of first-order variables. See the Appendix for a complete description of the considered variables. The generated masks were used to reduce the spatial stacks of NDVI variables to single time series (see Fig. 3). To do this, the masks were applied (in a variable-specific fashion when yield-correlation masks were considered) to the variable stacks, and one-dimensional annual time series were subsequently produced for all the variables by averaging the values of the retained pixels. In the context of the CV(1) exercise, each of the variable arrays that comprised the NDVI variable pool thus consisted of 11 points. It should be emphasized that, unlike cropland masking, the maps generated during yield-correlation masking are time (year) dependent by construction. Consequently, the yield-correlation masking procedure was repeated anew (throwing out the out year) during each step of the CV(1) so as to not bias the results. In no way did information from the respective out year enter into the masking/modeling procedures used during the 12 iterations of the CV(1) exercise.

Two elementary model families were selected for evaluation. Family-1 consists of two-parameter linear models (one NDVI-based variable and an intercept), and Family-2 consists of three-parameter linear models (two NDVI-based variables and an intercept). See the Appendix for a more detailed description of the two model families. All model parameters were estimated using ordinary least-squares regression, and all estimated models targeted deviation of yield from linear trend. At a particular forecasting time period, the one-variable models (Family-1) were constrained to use a variable containing information from that time period, and the two-variable models (Family-2) were constrained to include at least one such variable. This ensured uniqueness of the individual models evaluated at each of the three forecasting time periods.

Table 2
Variable number and model number information, by time period

	Time period					
	1	2	3	4	5	6
Number of period-specific first-order variables	1	2	3	4	5	6
Number of period-specific second-order variables	1	5	15	34	65	111
Total number of period-specific pool variables	2	7	18	38	70	117
Available first-order variables	1	3	6	10	15	21
Available second-order variables	1	6	21	55	120	231
Available pool variables	2	9	27	65	135	252
Number of one-variable models evaluated	–	–	–	38	70	117
Number of two-variable models evaluated	–	–	–	1729	6965	22581

One-variable (two-parameter) linear models correspond to Family-1, and two-variable (three-parameter) linear models correspond to Family-2.

Due to small sample size, all of the models within each family had a few degrees of freedom, leading to low confidence in model output. To mitigate this problem, we used the statistical result that forecast error variance typically is reduced when competing forecasts obtained from different models are averaged together (Granger, 1989). Furthermore, when predicting agricultural series, combined forecasts are frequently more accurate predictors than individual, uncombined forecasts (Allen, 1994). Thus, the out year forecasts were obtained by averaging across all individual model forecasts from a model family, and these values served as the actual forecasts for evaluation at a particular (region, crop, time period)-triple. To be clear, all presented results involving NDVI-based models pertain to estimates obtained using family-specific "all-subsets" average models. This output averaging increased the robustness of the estimates at the expense of optimality, but such a trade-off is necessary in the small-sample context of this research to enhance reliability of results. Table 2 shows details regarding the numbers of models from the different families evaluated at the three forecasted time periods. An explanation of how the values in Table 2 are obtained is provided in the Appendix. Details explaining a single application of the modeling method are also included in the Appendix, along with some discussion regarding assumptions made leading to our choice of method.

8. Results and discussion

There are many options for evaluating model performance. Following Mathews and Diamantopoulos (1994), five of these

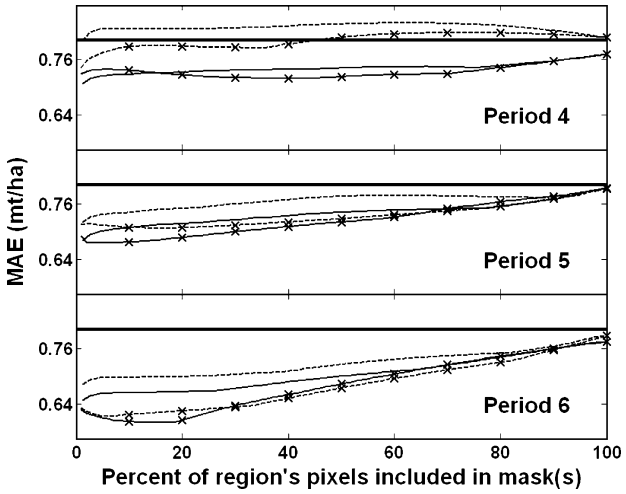

Fig. 6. MAE for IA corn yield forecasts. The horizontal black line is the linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

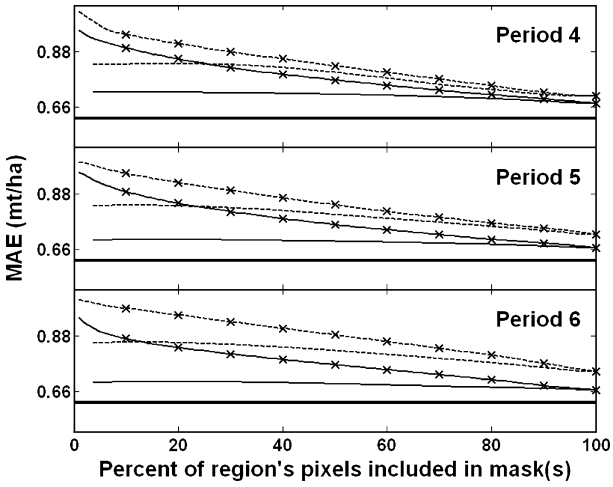

Fig. 8. MAE for IL corn yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

were selected for use in this study. All error values were calculated by comparing actual yields to their corresponding out year estimates obtained while stepping through the overarching CV(1) framework used in this research. Recall that the out year estimate is equal to the out year linear trend estimate plus (when NDVI-based models are involved) the out year estimate of deviation from linear trend, which is obtained from an all-subsets average model. Mean error (ME) indicates the bias of the forecasts over the 12 years, but does not generally reflect forecast accuracy. Mean absolute error (MAE) tells us the average magnitude of error and is reflective of model accuracy. Root mean squared error (RMSE) also gives an error magnitude, with bad misses amplified through the squaring operation. Mean absolute percent error (MAPE) gives a dimensionless error measure by relating absolute error magnitudes to the magnitudes of the target yields. Finally, the correlation coefficient (CORR) indicates the degree of collinearity that exists between the forecasted yields and the actual yields and is also a dimensionless measure.

8.1. Mask-size error curves

Figs. 6–13 show MAE obtained using cropland masking and yield-correlation masking while letting the mask area vary. MAE is displayed because it played no direct role in the masking/modeling procedures and thus might provide a more appropriate accuracy measure than CORR (the correlation operation was used during yield-correlation masking) or RMSE (implicitly minimized in the regression modeling). Fig. 7 shows results for Iowa soybeans, which was the (region, crop)-pair giving the best models relative to linear trend performance.

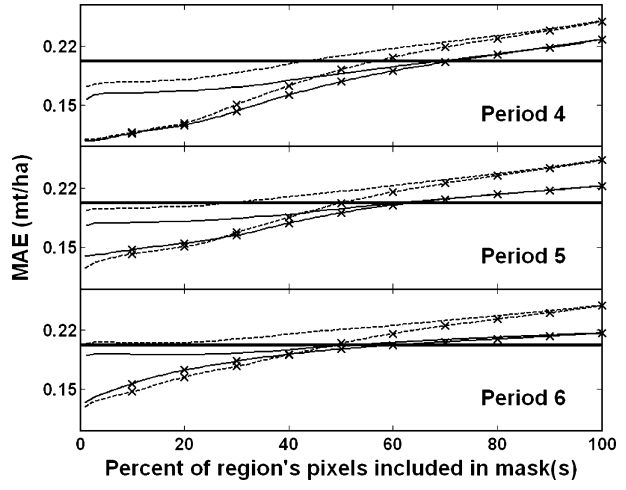

Fig. 7. MAE for IA soybean yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

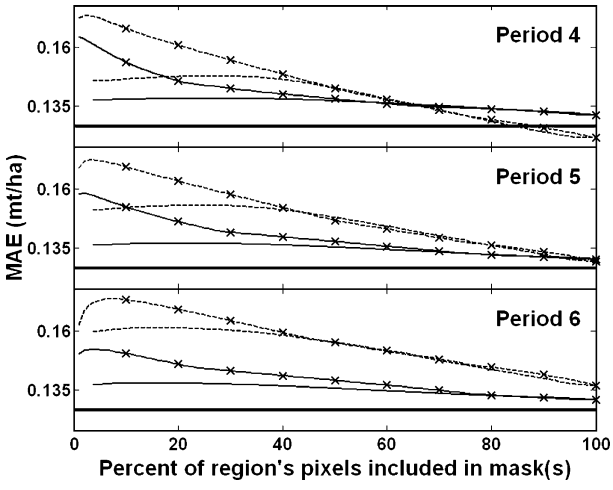

Fig. 9. MAE for IL soybean yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

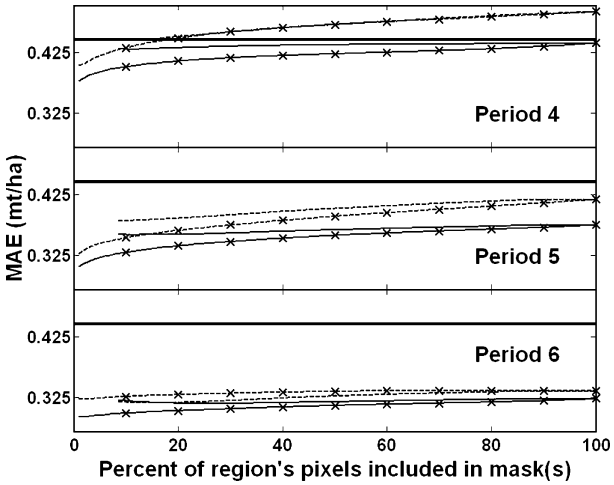

Fig. 10. MAE for KS winter wheat yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

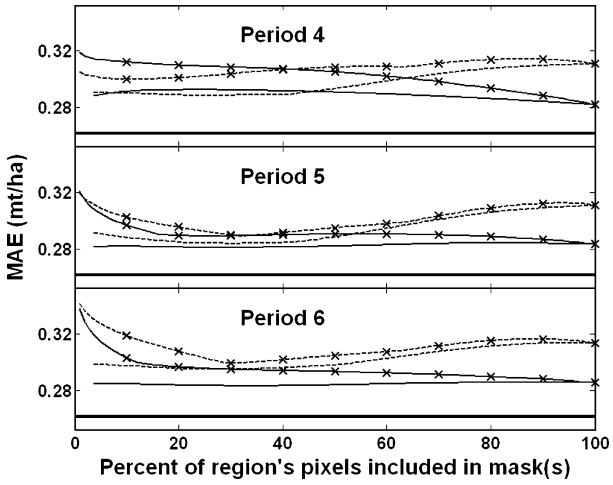

Fig. 12. MAE for ND spring wheat yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

Fig. 8 shows results for Illinois corn, which gave the worst models relative to linear trend performance. Note that in Figs. 6–13, the error curves associated with cropland masking do not extend to the smallest percent-inclusion values along the *X*-axis. This is because of the large number of 100%-cropped pixels that were present in the percent-cropland maps for the study states. The size of the set of 100%-cropped pixels defined the smallest cropland masks that were obtainable for these regions using the described thresholding techniques.

Looking at Figs. 6–13, the all-subsets average models from Family-1 (composed of one-variable linear models, depicted by the solid lines in the figures) out-performed the all-subsets average models from Family-2 (composed of two-variable linear models, depicted by dashed lines in the figures) in the majority of the cases evaluated in the CV(1) modeling exercise. Had the analysis been performed using only in-sample results, this outcome could not have occurred in any of the cases, suggesting that to some extent the CV(1) exercise served its purpose of exposing false prophets. This result indicates that in this small sample exercise, the Family-2 all-subsets average models (and thus their component models) had a tendency to overfit the data rather than uncover systematic relationships between NDVI and crop yields. In future work, extending the temporal reach of the data to include more years of information will help expose more complex relationships, if they exist.

8.2. Modeling comparison

To compare modeling results, we must select a mask size. When comparing with respect to MAE, we will choose the

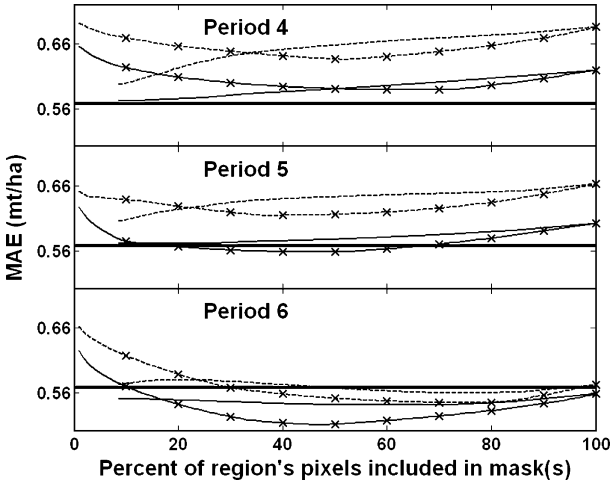

Fig. 11. MAE for KS sorghum yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

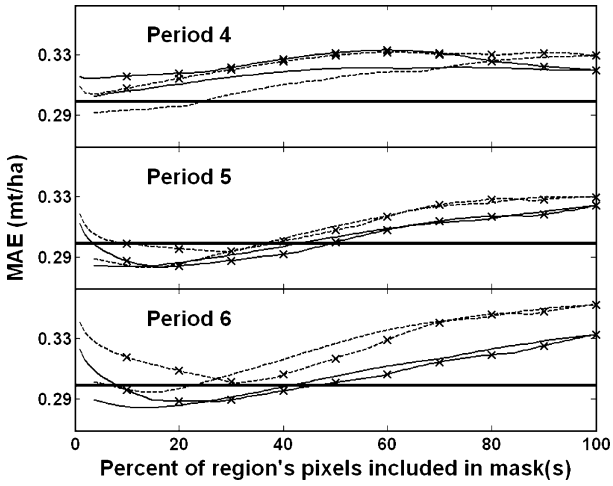

Fig. 13. MAE for ND barley yield forecasts. The horizontal black line is linear trend error. The unmarked solid (Family-1) and dashed (Family-2) curves represent error obtained using cropland masks. The curves marked with '×' show corresponding errors obtained using yield-correlation masks. Yield units are given in metric tons per hectare.

Table 3
Optimal model counts

	Linear trend	Cropland masking	Yield-correlation masking	No masking
ME Optimal Model Count	6	5	34	3
MAE Optimal Model Count	21	6	20	1
RMSE Optimal Model Count	22	8	16	2
MAPE Optimal Model Count	20	5	22	1
CORR Optimal Model Count	24	0	23	1

There are 48 observations [8 (region, crop)-pairs × 3 time periods × 2 model families] of best mask/model approach in the 12-year CV(1) exercise. CV(1) linear trend is considered among the models to serve as a benchmark.

location of the minimum of each MAE error curve depicted in each Figs. 6–13. Likewise, a "best mask size" will be selected for each performance measure in the same fashion. This is a biased selection since it is being made after the CV(1) exercise, one that will ensure any performance values reported will appear better than they likely would in real-time use. However, the same selection bias is imparted to the two masking methods (cropland masking and yield-correlation masking) under study, which keeps the field level between them. Also, Maselli and Rembold (2001) found little variation between the cropland masks they estimated using yield-correlation maps in a similar, 13-year CV(1) framework, which suggests that adding or removing a year from our analysis may not greatly alter the shape of the mask-size error curves. If this is true, then the presented results, though certainly overly optimistic, will not be that far from their "true" values in the study context.

The ME performance measure warrants further discussion so that it can be interpreted properly. This is the only measure of the five where an outcome can occur on either side of its optimal value, which is zero. In a smooth, multiple-scenario study such as the present one (which is smooth with respect to the effects of mask size variation on yield prediction), it is not uncommon to see ME cross over from positive to negative values (or vice versa), sometimes even more than once. Thus, movements in the ME curve due to data sampling effects will essentially relocate the zero-crossings, whereas with the other four measures, their best cases will generally maintain a non-vanishing amount of separation from their theoretically optimal values due to incompleteness of the information used in the models. Not that the ME values are totally misleading, as the mask sizes with smaller ME values are probably close to the "correct" ones, but the ME values themselves near the theoretical minimum will not reflect reality in the operational setting. Further, density of ME occurrences near zero will still be reflective of the relative, general bias of the cropland masking and yield-correlation masking methods.

In this study, there are eight (region, crop)-pairs, each with three forecast periods, and forecast performance from two model families at each forecast period. That gives $8 \times 3 \times 2 = 48$ opportunities to determine which of four tested methods (linear trend, cropland masking with regression modeling, yield-correlation masking with regression modeling, or regression modeling with no masking) performed best in a particular situation. As making a linear trend estimate is a once-per-year event (i.e., it does not depend on forecast period), each CV(1) linear trend estimate was replicated across the three forecast periods. Table 3 shows counts, with respect to choice of error measure, reflecting the number of instances (out of 48) when each of the techniques performed the best. In light of the above discussion regarding ME, it is not surprising that scenarios were encountered where NDVI-based models outperformed the linear trend with respect to ME. However, the fact that yield-correlation masking accounted for the vast majority of these cases rather than cropland masking is worth noting, and will be revisited later.

Looking at the other four accuracy measures (MAE, RMSE, MAPE, and CORR) in Table 3, we see that the linear trend outperformed the other techniques in the most cases (87 out of 192). Yield-correlation masking picked up the bulk of the remainder (81 out of 192), followed by cropland masking (19 out of 192), and finally no masking (5 out of 192). Due to the aforementioned mask-size selection bias afflicting results from cropland masking and yield-correlation masking, the counts for linear trend and the "no mask" case may be higher in an operational setting using the same sub-optimal, all-subsets average modeling strategy presented here. But, this does not derail the comparison between cropland masking and yield-correlation masking, each of which has been subjected to the same treatment throughout the study.

Table 4 shows "best of the rest" counts, which look at how many times the three non-trivial forecasting techniques came in second place when linear trend forecasting was optimal. Looking at MAE, RMSE, and MAPE, we see that in these cases cropland masking clearly outperformed yield-correlation masking, and no masking made a strong presence as well. Note the clear dominance of yield-correlation masking in the "best of the rest" scenarios with respect to CORR. This is likely a consequence of the correlation operation's integral part in yield-correlation masking.

One benchmark that can be used to gauge the level of model accuracy is comparison to accuracies of corresponding USDA initial season yield forecasts. Historical estimates corresponding to the case studies in question were retrieved from the USDA NASS website given earlier. Relevant accuracy measures are summarized in Table 5, where results presented for the "NDVI" method were obtained using yield-correlation masking with measure-specific, best-identified mask sizes and Family-1 all-subsets average models. NDVI-driven forecasts from period 4 and period 5 theoretically would have been generated prior to the release of initial USDA estimates. In most cases, the USDA

Table 4
"Best of the rest" optimal model counts

	Cropland masking	Yield-correlation masking	No masking
ME Optimal Model Count	0	6	0
MAE Optimal Model Count	10	0	11
RMSE Optimal Model Count	11	1	10
MAPE Optimal Model Count	8	1	11
CORR Optimal Model Count	5	17	2

This table shows first runner-up counts for the mask/model approaches in the cases when the CV(1) linear trend performed the best.

Table 5
Observed errors

Forecast period	Method description	IA Corn	IA Soy	IL Corn	IL Soy	KS Wwt*	KS Sorg*	ND Spwt	ND Barl
MAE									
–	Trend	0.80	0.20	0.62	0.13	0.45	0.57	0.26	0.30
p5–p6	USDA	0.58	0.20	0.50	0.11	0.36	0.43	0.25	0.28
p6	NDVI	*0.60*	**0.13**	0.67	0.13	**0.29**	*0.51*	0.29	*0.29*
p5	NDVI	*0.68*	**0.14**	0.67	0.13	**0.31**	*0.56*	0.28	*0.28*
p4	NDVI	*0.72*	**0.11**	0.67	0.13	*0.38*	0.59	0.28	0.31
MAPE									
–	trend	11.5	7.5	7.5	4.5	19.8	14.0	12.4	11.2
p5–p6	USDA	8.3	7.0	5.9	4.1	14.4	10.4	11.1	10.2
p6	NDVI	*8.4*	**4.8**	8.2	4.7	**12.5**	*12.4*	13.8	*10.6*
p5	NDVI	*9.3*	**5.0**	8.2	4.7	**13.2**	*13.6*	13.6	*10.5*
p4	NDVI	*10.0*	**4.1**	8.4	4.7	*16.6*	14.3	13.5	11.7
RMSE									
–	trend	1.16	0.31	0.86	0.15	0.53	0.64	0.35	0.38
p5–p6	USDA	0.81	0.23	0.70	0.14	0.48	0.53	0.31	0.39
p6	NDVI	**0.78**	**0.19**	0.91	0.16	**0.40**	*0.63*	0.36	0.39
p5	NDVI	*0.84*	**0.18**	0.92	0.16	**0.41**	0.67	0.35	0.39
p4	NDVI	*0.88*	**0.15**	0.96	0.16	*0.51*	0.69	0.37	0.40
CORR									
–	Trend	0.49	0.44	0.45	0.52	0.54	0.40	0.02	0.18
p5–p6	USDA	0.75	0.74	0.66	0.87	0.51	0.48	0.39	0.20
p6	NDVI	**0.77**	**0.81**	0.25	0.36	**0.67**	0.39	*0.08*	*0.20*
p5	NDVI	*0.73*	**0.83**	0.15	0.35	**0.66**	0.29	*0.13*	0.17
p4	NDVI	*0.70*	**0.90**	0.07	0.32	0.47	0.21	*0.10*	0.10

Results shown for "NDVI" method were obtained using yield-correlation masking with best-identified mask sizes (specific to each performance measure) and Family-1 all-subsets average models. NDVI model accuracies surpassing USDA accuracies are in bold face. NDVI model accuracies surpassing trend accuracies (but not USDA) are in italics. *USDA forecasts for winter wheat and sorghum in Kansas were unavailable for 1989. Units for MAE and RMSE are metric tons per hectare.

estimates outperformed both trend models and the NDVI-based models, not an unexpected result given the tremendous effort and resources expended to generate the USDA estimates. To help put the results shown in Table 5 in proper perspective, we reiterate that the all-subsets average models used in this research are not expected to be optimal.

Since we are making a post facto determination of optimal mask size, this USDA-based accuracy assessment is not entirely fair, as our model accuracies will appear more accurate than they probably should be. However, had we imposed the logical mask size restriction that all masks used include an area equivalent in size to historical harvested area (avoiding mask size selection bias in this regard), the Family-1 all-subsets average models would have still outperformed USDA estimates for soybeans in Iowa and winter wheat in Kansas (at least when using MAE as an accuracy measure; see Figs. 7 and 10 and Table 5). For reference, soybeans covered approximately 25% of Iowa annually from 1989 to 2000, and approximately 20% of Kansas was annually covered by winter wheat during this same period.

8.3. The influence of data sampling

Regarding the major discrepancies in forecast performance (relative to linear trend) among the various (region, crop) pairs, we posit that the variation in sample statistical behavior of crop yields among the time spans considered might be responsible for this effect. For example, NDVI-based models performed relatively well (compared to linear trend performance) when applied to IA soybeans and KS winter wheat, but performed relatively poorly when applied to IL corn and soybeans. The final column in Table 1 shows "Mean Squared Deviation from 1970 to 2000 Trend Line," which helps explain this result. The deviation values for IA soybeans and KS winter wheat from 1989 to 2000 (the study period) are substantially greater than their counterparts from the other two time spans considered, thus presenting opportunity for improvement over using the linear trend model during the study period. Conversely, the deviation values for IL corn and soybeans from 1989 to 2000 are substantially lower than their counterparts, suggesting that the linear trend model ought to fare well during the study period, leaving less room for NDVI models to improve upon. Note that these are hindsight observations, not something that can be ascertained in real time.

Continuing with this logic, in situations similar to that in IL, where the linear trend performs better than expected, this leaves a "deviation from linear trend" series that has a smaller than expected magnitude of variation (i.e., a lower signal-to-noise ratio). This increases the chances of encountering spuriously high correlation values between pixel-level NDVI values and final regional yield, leading to sub-optimal yield-correlation maps and diminished performance in forecasting

Table 6
Head-to-head comparison between yield-correlation masking and cropland masking

	Yield-correlation masking	Cropland masking
ME	40	8
MAE	20	28
RMSE	17	31
MAPE	23	25
CORR	40	8
Total	140	100

There are 48 evaluation scenarios for each performance measure. Values indicate number of occurrences where the method was superior to the other method.

situations. Recall that in situations where linear trend performed the best, cropland masking starkly dominated yield-correlation masking, perhaps due to the effects just described. The argument can be reversed in favor of yield-correlation masking in situations where the linear trend performs worse than expected, which may help account for the better performance of yield-correlation masking than cropland masking in these situations.

These observations illustrate the increased dependence of the performance of the yield-correlation masking procedure (relative to the performance of cropland masking) on the sample yields. Like in many statistical modeling situations, we have traded a decrease in bias for an increase in variance, in that yield-correlation masks may result in less biased yield forecasts on average when compared to cropland masking, but the yield forecasts may exhibit increased error variance (or equivalently, increased error standard deviation). That yield-correlation masking out-performed cropland masking 40 to 8 with respect to the ME measure of forecast bias (Table 6) lends support to this line of reasoning. Likewise, that cropland masking out-performed yield-correlation masking 31 to 17 with respect to RMSE (Table 6), which is analogous to forecast error standard deviation, also supports this statistical explanation, but the lesser rate also suggests that the bias-for-variance trade-off observed when using yield-correlation masking instead of cropland masking may be worthwhile. However, the evidence we have presented is inconclusive in this last regard.

9. Conclusions

We have presented evidence that yield-correlation masking is a viable alternative to cropland masking in the context of regional crop yield forecasting. The appeal of yield-correlation masking is that no land cover map is required to implement the procedure, but the procedure results in forecasts of comparable accuracy to those obtained when using cropland masking. As cropland masking, when compared to using a "no masking" method, has been shown in previous work (and in this study) to generally improve regional crop yield models, this result is significant.

The all-subsets average model design used in this research was chosen because of its simplicity and robustness and certainly could be improved upon if the goal is to minimize forecast error, which is desirable when developing operational yield prediction models. For example, regression model subset selection and/or variable or model weighting procedures could be used in an attempt to improve prediction efficiency. However, as our primary goal was to establish the validity of the yield-correlation masking procedure as an alternative to cropland masking for the purpose of crop yield forecasting, given our small sample size, we favored model generality and statistical robustness over prediction accuracy. We sought to minimize the dependence of our results on the act of variable or model selection, so that individual linear parameter estimates were of as little consequence as possible. As a result, we have exposed some general, probabilistic tendencies of performance of the different masking methods and model families.

From a modeling efficacy perspective, our results indicate that 11 years (or less, no doubt) of time series AVHRR NDVI data may not be enough to build reliable, two-NDVI variable linear crop yield models. This does not imply that more data points will or will not remedy the situation, nor does it preclude that using a second variable that is not derived from NDVI (such as a climate data series or a crop condition index) could prove to be useful.

It must be noted that this research constitutes a cross-validation exercise, not a true test of forecasting ability. No matter how much care was taken to ensure that legitimate "out-of-sample" prediction was used for model evaluation, some unknown amount of selection bias (Miller, 2002, p. 6) has occurred because the same data were used to both construct the models and evaluate their aptitude for prediction (not at each step of the CV(1), but in the CV(1) exercise as a whole). In the present study, this effect will be the strongest with respect to optimal mask-size determination, which was performed after the CV(1) masking/modeling exercise had been completed.

Data from certain states were easier to model than others. Relatively speaking, Iowa corn and soybeans and Kansas winter wheat permitted models that showed improvement over using the linear trend as a forecast, whereas Illinois corn and soybeans and North Dakota spring wheat gave the masking/modeling procedures the most difficulty. That Kansas winter wheat models performed well may be due to the distinct early growing season of the crop, which undergoes a rapid emergence from winter dormancy that precedes emergence of surrounding natural vegetation. This allows the crop to have a strong influence on early season NDVI values, favoring the extraction of information pertaining to yield potential. Iowa's agricultural landscape is greatly dominated by corn and soybeans, so spatial homogeneity may have been a factor in the good model performance seen in this state. Ultimately, though, sample statistical behavior of the final regional yields during the study period was reasoned to be largely responsible for the major performance discrepancies observed among the various (region, crop)-pairs examined in this study.

In an operational setting, assuming the availability of adequate computer processing power, one could perform an

analysis just like the one presented here in order to obtain an optimal mask size, and then use this information when extracting NDVI data to use as inputs into operational models. Alternatively, one could even determine variable-specific optimal mask sizes if these are desired, though this design deviates from that presented in this paper, and would introduce more variability into subsequent forecasts. Ultimately, optimal masks could be identified by simultaneously evaluating both cropland masks and yield-correlation masks. In fact, any numerical map for the region, such as vegetation phenology maps (e.g., average date of green-up onset, average length of growing season, average maximum NDVI, etc.) and other mathematically derived maps (e.g., NDVI mean or variance, Fourier amplitude and phase, etc.), can be thresholded and evaluated for masking in the context of yield forecasting.

To solidify or dismiss results uncovered in this research, the analysis could be extended to additional (region, crop)-pairs. The analysis need not be restricted to states, as other spatial scales such as agricultural statistics districts, counties, or the conterminous U.S. as a whole can also serve as regions of study. As a final note, crop yields are not the only target that can be pursued using the correlation masking procedure described in this paper. Rather, any quantity dependent on seasonal vegetation conditions (and possessing an adequate historical record) can serve as the dependent variable.

Acknowledgements

The authors would like to thank John Lomas for assistance with database development and Kevin Dobbs for valuable proofreading and editorial recommendations. We also wish to thank the reviewers for their thorough and constructive commentary, which led to a much-improved paper.

Appendix A

A single year of input data for each (region, crop)-pair consists of a six-point, biweekly NDVI time series, roughly spanning the beginning of a specific crop growing season to the peak of that growing season. Let $\{X_1,\ldots,X_6\}$ be the set of vectors corresponding to these time series points, with one entry per vector per year, so that each vector has length 11 (the number of years used for masking and modeling in each iteration of the CV(1) exercise). The only prior assumption we will make is that as a growing season progresses, the more complete the information (with respect to predicting final crop yield) contained in the NDVI time series becomes. Constraining the models to always include information from the last available period is justifiable in this regard, and it also ensures no redundancy between individual models considered during the three forecasting periods.

For a given prediction period $t=4$, 5, or 6, expand the time series to a variable pool consisting of 65, 135, or 252 variables, respectively (see Table 2). First generate all possible NDVI-1 variables, which are running sums of NDVI within the t-point span (the "Available First Order Variables" in Table 2). Thus, there is one length–t span, two length–$(t-1)$ spans, and so on, resulting in $(1+2+\ldots+t)=t(t+1)/2$ NDVI-1 variables. Next, multiply all of the NDVI-1 variables together pair-wise, including squaring, to obtain NDVI-2 variables (the "Available Second-Order Variables" in Table 2). As there are $(t(t+1)/2)$-choose-2 ways to select two distinct variables from a list of length $t(t+1)/2$, as well as $t(t+1)/2$ unique squares, there are $\binom{t(t+1)/2}{2}+t(t+1)$ NDVI-2 variables. We then define the variable pool to consist of all NDVI-1 and NDVI-2 variables (the "Available Pool Variables" in Table 2), each of which is a vector of length 11. If we let X denote the variable pool, we have

$$X=\left\{\left(\sum_{j=l}^{u}x_j\right)\,|\,1\le l\le u\le t\right\}\cup\left\{\left(\sum_{j=l_1}^{u_1}x_j\right)\left(\sum_{j=l_2}^{u_2}x_j\right)\,|\,1\le l_1\le u_1\le t \text{ and } 1\le l_2\le u_2\le t\right\}.$$

NDVI-1 variables are readily justified, given that linear relationships have been documented in the remote sensing literature between such variables and crop yields. NDVI-2 variables are included to partially account for curvature that might exist in the relationship between yields and NDVI. There are also logical reasons for including NDVI-2 variables. For instance, an NDVI-1 variable from early in the season can be informative for establishing a yield potential, whereas an NDVI-1 variable with information from later periods can be indicative of how much of that potential is being realized. Consequently, the product of these two variables (which is an NDVI-2 variable) can reflect such an interaction in the form of a single predictor.

Given this variable pool, the modeling situation can be characterized as follows:

1) Small sample of observations ($n=11$ for each model evaluation)
2) Many more predictors than observations
3) Massive collinearity among the predictors
4) Dynamic and unpredictable predictor extraction methods (predictors are extracted summarily from spatial data) preclude the presumption of priors relating particular predictors with the quantity being predicted.

In the light of these observations, we need a modeling method that is robust, making maximal use of the few degrees of freedom that are available in the data. In particular, we are most concerned with robustness and generality (rather than optimality) so that results we obtain are as unassailable as possible. All-subsets regression with output averaging provides just such a method.

In the all-subsets regression application used here, design matrices for individual regression models consist of two or three columns, an intercept and one or two variables from X. Models with two NDVI variables are included to allow for simple multilinear relationships between NDVI variables and crop yields and can also be logically justified for reasons similar to those supporting the inclusion of NDVI-2 variables. All models to be evaluated are shaped as follows.

Let $X_t \subset X$ denote the subset of X containing all variables that include NDVI information from prediction time period t (see Table 2, row 3). Define $F_1(t) = \{M | M = [1V], V \in X_t\}$ and $F_2(t) = \{M | M = [1V_1V_2], V_1 \in X_t, V_2 \in X, V_1 \neq V_2\}$, where the "1" in brackets is a column vector of ones (i.e., an intercept variable). F_1 and F_2 contain the design matrices associated with model families Family-1 and Family-2, respectively, described in the main text. Let n_1 and n_2 denote the cardinalities of F_1 and F_2, respectively (see Table 2, rows 7 and 8). These values can be obtained using basic counting procedures, subject to the constraint of the one prior we have assumed.

Let Y denote the dependent variable (i.e., the vector of in-sample deviations of yield from linear trend). Index the members of F_1 and F_2, and then define functions f_1 and f_2 such that, for $k=1$, 2, we have $f_k(t) = \frac{1}{n_k}\sum_{j=1}^{n_k} M_j\beta_j$, where $M_j \in F_k(t)$ and $\beta_j = \left(M_j^T M_j\right)^{-1} M_j^T Y$.

Under standard regression assumptions, f_k will be statistically unbiased (i.e., it will have a long-term average error of 0), as it is the average of unbiased OLS regression models, each of which is unbiased because each possesses an intercept term. Also, if we index members of the sets X_t and X, after determining regression coefficients, we can rearrange f_k into the form $f_k(t) = a_0 + \sum_{j=1}^{m_t} a_j V_j$, where m_k is the cardinality of X_t and $V_j \in X_t$ if $k=1$, and m_k is the cardinality of X and $V_j \in X$ if $k=2$. Presuming more than 11 a_j's are different from 0, it becomes clear that we are actually using a supersaturated model (i.e., a model with more parameters to estimate than observations to use for this purpose), but one that has been derived deterministically. Also, the output of the model will be robust to small perturbations in NDVI values due to the linear regression model averaging that is used and the symmetric entry of the variables into the model. The latter factor aids robustness because most of the variables used involve discrete integrations of NDVI, which is a smoothing operation that in the present situation will serve to diminish the effects of signal noise.

Once the β_j's have been estimated using the 11 available years of input data, deviation from trend yield is predicted at time t for the out year that was set aside as part of the CV(1) procedure using the models $f_1(t)$ and $f_2(t)$, which are precisely the all-subsets regression average models described in the main text. All results obtained in this study are based on out-of-sample estimates derived in this fashion.

References

Allen, P. G. (1994). Economic forecasting in agriculture. *International Journal of Forecasting*, *10*(1), 81–135.

Badhwar, G. D., & Henderson, K. E. (1981). Estimating development stages of corn from spectral data—an initial model. *Agronomy Journal*, *73*, 748–755.

Das, D. K., Mishra, K. K., & Kalra, N. (1993). Assessing growth and yield of wheat using remotely sensed canopy temperature and spectral indices. *International Journal of Remote Sensing*, *14*, 3081–3092.

Doraiswamy, P. D., & Cook, P. W. (1995). Spring wheat yield assessment using NOAA AVHRR data. *Canadian Journal of Remote Sensing*, *21*, 43–51.

Doraiswamy, P. D., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote Sensing*, *69*(6), 665–674.

Egbert, S. L., Peterson, D. L., Stewart, A. M., Lauver, C. L., Blodgett, C. F., Price, K. P., et al. (2001). The Kansas gap land cover map: Final report. *Kansas biological survey report*, *vol. 98*. Kansas: Lawrence.

Eidenshink, J. C. (1992). The 1990 conterminous U.S. AVHRR data set. *Photogrammetric Engineering and Remote Sensing*, *58*(6), 809–813.

Ferencz, Cs., Bognar, P., Lichtenberger, J., Hamar, D., Tarcsai, Gy., Timar, G., et al. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, *20*, 4113–4149.

Granger, C. W. J. (1989). Combining forecasts—twenty years later invited review. *Journal of Forecasting*, *8*, 167–173.

Groten, S. M. E. (1993). NDVI-crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing*, *14*, 1495–1515.

Gupta, R. K., Prasad, S., Rao, G. H., & Nadham, T. S. V. (1993). District level wheat yield estimation using NOAA/AVHRR NDVI temporal profile. *Advanced Space Research*, *13*, 253–256.

Hayes, M. J., & Decker, W. L. (1996). Using NOAA AVHRR data to estimate maize production in the United States corn belt. *International Journal of Remote Sensing*, *17*, 3189–3200.

Kastens, D. L. (1998). Estimating wheat yields from time series analysis of remotely sensed data. *United States Department of Agriculture small business innovative research phase I grant (Grant agreement number 99-33610-7495)*.

Kastens, D. L. (2000). Forecasting pre-harvest winter wheat yields in the great plains using remotely sensed data. *United States Department of Agriculture small business innovative research phase II grant (Grant agreement number 00-33610-9453)*.

Kastens, T. L., & Dhuyvetter, K. C. (2002). *2002 harvest year report for USCHI's Custom Harvester Analysis and Management Program (CHAMP)* online at. http://www.agmanager.info/farmmgt/machinery/default.asp, last accessed October 14, 2005.

Kastens, J. H., Jakubauskas, M. E., & Lerner, D. E. (2003). Using temporal averaging to decouple annual and non-annual information in AVHRR NDVI time series. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(11), 2590–2594.

Lee, R., Kastens, D. L., Price, K. P., & Martinko, E. A. (1999). Forecasting corn yield in Iowa using remotely sensed data and vegetation phenology information. *Proceedings, PECORA 14, land satellite information regional conference, American society of photogrammetric Engineering and remote sensing, Denver, CO, December 6–10*.

Maselli, F., Conese, C., Petkov, L., & Gilabert, M. A. (1992). Use of NOAA-AVHRR NDVI data for environmental monitoring and crop forecasting in the Sahel. Preliminary results. *International Journal of Remote Sensing*, *13*, 2743–2749.

Maselli, F., & Rembold, F. (2001). Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogrammetric Engineering and Remote Sensing*, *67*(5), 593–602.

Mathews, B. P., & Diamantopoulos, A. (1994). Towards a taxonomy of forecast error measures. *Journal of Forecasting*, *13*, 409–416.

Miller, A. (2002). Subset selection in regression (2nd Edition). Boca Raton, FL: Chapman and Hall/CRC Press LLC.

Potdar, M. B. (1993). Sorghum yield modeling based on crop growth parameters determined from visible and near-IR channel NOAA AVHRR data. *International Journal of Remote Sensing*, *14*, 895–905.

Quarmby, N. A., Milnes, M., Hindle, T. L., & Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing*, *14*, 199–210.

Rasmussen, M. S. (1992). Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR. *International Journal of Remote Sensing*, *13*, 3431–3442.

Rasmussen, M. S. (1997). Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability. *International Journal of Remote Sensing*, *18*, 1059–1077.

Rasmussen, M. S. (1998). Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information. Part II: Crop yield assessment. *International Journal of Remote Sensing*, *19*, 119–139.

Rudorff, B. F. T., & Batista, G. T. (1991). Wheat yield estimation at the farm level using TM Landsat and agro meteorological data. *International Journal of Remote Sensing*, *12*, 2477–2484.

Tucker, C. J., Holben, B. N., Elgin, J.H, Jr., & McMurtrey, J. E. (1980). Relationship of spectral data to grain yield variation. *Photogrammetric Engineering and Remote Sensing*, *45*, 657–666.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, N. (2001). Completion of the 1990s national land cover data set for the conterminous United States from landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*, 650–652.

Wardlow, B.D. Kastens, J.H., Egbert, S.L. (in press). Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data. *Photogrammetric Engineering and Remote Sensing*.

Wiegand, C. L., & Richardson, A. J. (1990). Use of spectral vegetation indices to infer leaf area, evapotranspiration and yield. *Agronomy Journal*, *82*, 623–636.

Wiegand, C. L., Richardson, A. J., & Kanemasu, E. T. (1979). Leaf area index estimates for wheat from Landsat and their implications for evapotranspiration and crop modeling. *Agronomy Journal*, *71*, 336–342.

Wiegand, C. L., Richardson, A. J., Jackson, R. D., Pinter Jr., P. J., Aase, J. K., Smika, D. E., et al. (1986). Development of agrometeorological crop model inputs from remotely sensed information. *IEEE Transactions on Geoscience and Remote Sensing*, *24*(1), 83–89.