Remote Sensing of Environment 112 (2008) 3520–3537

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US

Mutlu Ozdogan [a,*], Garik Gutman [b]

[a] Center for Sustainability and the Global Environment (SAGE), University of Wisconsin, 1710 University Avenue, Madison, WI 53726, USA
[b] NASA Headquarters, Earth Science Division, 300 E Street, SW, Washington, DC 20546, USA

ARTICLE INFO

Article history:
Received 2 January 2008
Received in revised form 1 April 2008
Accepted 12 April 2008

Keywords:
Irrigation
MODIS
Agriculture
Water
Climate

ABSTRACT

We present a dryland irrigation mapping methodology that relies on remotely sensed inputs from the MODerate Resolution Imaging Spectroradiometer (MODIS) instrument, globally extensive ancillary sources of gridded climate and agricultural data and on an advanced image classification algorithm. The methodology involves four steps. First, we use climate-based indices of surface moisture status and a map of cultivated areas to generate a potential irrigation index. Next, we identify remotely-sensed temporal and spectral signatures that are associated with presence of irrigation defined as full or partial artificial application of water to agricultural areas under dryland conditions excluding irrigated pastures, paddy rice fields, and other semiaquatic crops. Here, the temporal indices are based on the difference in annual evolution of greenness between irrigated and non-irrigated crops, while spectral indices are based on the reflectance in the green and are sensitive to vegetation chlorophyll content associated with moisture stress. Third, we combine the climate-based potential irrigation index, remotely sensed indices, and learning samples within a decision tree supervised classification tool to make a binary irrigated/non-irrigated map. Finally, we apply a tree-based regression algorithm to derive the fraction of irrigated area within each pixel that has been identified as irrigated. Application of the proposed procedure over the continental US in the year 2001 produces an objective and comprehensive map that exhibits expected patterns: there is a strong east-west divide where the majority of irrigated areas is concentrated in the arid west along dry lowland valleys. Qualitative assessment of the map across different climatic and agricultural zones reveals a high quality product with sufficient detail when compared to existing large area irrigation databases. Accuracy assessment indicates that the map is highly accurate in the western US but less accurate in the east. Comparison of area estimates made with the new procedure to those reported at the state and county levels shows a strong correlation with a small bias and an estimated RMSE of 2500 km^2, or little over 2% of the total irrigated area in the US. As a result, the future application of the new procedure at a global scale is promising but may require a better potential irrigation index, as well as the use of remotely sensed skin temperature measurements.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Accurate information on irrigation extent is fundamental to many aspects of the Earth systems science and global change research. This includes modeling of water exchange between the land surface and atmosphere (Boucher et al., 2004; Gordon et al., 2005; Ozdogan et al., 2006), analysis of the impact of climate change and variability on irrigation water requirements/supply (Alcamo et al., 2003; Rosenzweig et al., 2004; Vörösmarty et al., 2000), and management of water resources that affect global food security (Vörösmarty et al., 2005). However, the current extent of irrigated areas over continental to global scales is still uncertain and available maps are derived primarily from country-level statistics using maps that are often outdated or made with data from relatively coarse resolution sensors (Droogers, 2002). Even in countries such as the US, where the general extent of irrigated areas is known, irrigation-related information exists in disparate datasets and cannot be easily synthesized into a single continental scale database (Vörösmarty, 2002).

To overcome these limitations, we present a novel methodology to map irrigated agriculture globally at a 500-m spatial resolution with a combination of data from the MODerate resolution Imaging Spectroradiometer (MODIS) instrument and ancillary data on climate and agricultural area. For the purposes of the work presented here, we define *irrigated* as agricultural areas that receive full or partial application of water to the soil to offset periods of rainfall shortfalls under dryland conditions. More specifically, we focus on dryland cultivated areas, excluding irrigated pastures as well as paddy rice fields and other semiaquatic crops. The reasons for this are two folds.

* Corresponding author. Tel.: +1 608 890 0336.
E-mail address: ozdogan@wisc.edu (M. Ozdogan).

0034-4257/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2008.04.010

First, while paddy irrigated fields have a distinct remotely sensed signal associated with transplanting in water and maturity (Xiao et al., 2006), dryland irrigated conditions are much more ambiguous and require separate methods. Second, irrigated pastures make up a small portion of total irrigated lands and are not considered important for food production as well as environmental implications. We also selected a minimum mapping unit of 100 ha (roughly four MODIS 500-m pixels) to identify and map irrigated lands, in essence, not considering small, irrigated lawns and gardens.

Our irrigation mapping methodology is objective, uses contemporary data, is robust enough to handle complex forms of irrigation that occur around the globe, and can be repeated at any location and time. This irrigation mapping effort is part of our larger research program to understand anthropogenic effects, specifically that of irrigation on global water and energy cycles, climate, agricultural productivity, and agricultural water sustainability. In this paper we present the methodology and an example of its application to the continental US as well as assessment and validation of our algorithm.

2. Existing datasets on global irrigation

In recent years, there have been several attempts to determine the spatial extent of irrigation at global scale. One such attempt is the USGS Global Land Cover Map (Loveland et al., 2000), generated based on 1-km AVHRR observations between April 1992 and September 1993. Of several land-cover legends that the database contains, the Global Ecosystems Legend identifies four irrigated land classes: *irrigated grassland*, *rice paddy and field*, *hot irrigated cropland*, and *cool irrigated cropland*. When combined, these classes provide one of the few sources of remotely sensed information on spatial distribution of irrigation over the continental US (Fig. 1 — Panel A).

Another global irrigation mapping effort was undertaken by Siebert et al., 2005 (updated by Siebert et al., 2007) who combined heterogeneous information on the (approximate) location of irrigated areas with information on the total irrigated area from national and international sources to generate the first global "irrigated lands" map (Fig. 1 — Panel B). The map is a digital raster product with 5-min spatial resolution containing information for each cell on the percentage of area *equipped for irrigation* centered on the year 2000. For the US, this product was generated by assigning the maximum of the irrigated areas by county as reported in the census surveys of USGS and USDA for the years 1995, 1997, 2000 and 2002 to agricultural land mapped by USGS and US-EPA at 30 m resolution. As a result, the total area equipped for irrigation in the continental US is about 25% larger than the NASS-estimate for 2002 (Siebert et al., 2005). The product of Siebert et al. (2005) has become the *de facto* present-day information source for spatial distribution of global irrigated areas.

More recently, the Remote Sensing and GIS group at the International Water Management Institute (IWMI) released a beta version of the Global Map of Irrigated Areas (GMIA) *circa* 1999 (http://www.iwmigmia.org/info/main/index.asp). The dataset has been produced using twenty years of AVHRR data augmented with additional information from SPOT *Vegetation*, Japanese Earth Resources Satellite (JERS-1), and Landsat GeoCover 2000 data, and mapped into 10 km grid resolution (Fig. 1 — Panel C). The Beta release map has 53 irrigation classes, derived from the 628 classes in the master file. This approach follows that of Thenkabail et al. (2005).

While these data sets provide the best available source of information regarding the distribution of irrigation at continental scales, depending on the way these datasets are used, they may also have serious shortcomings. For example, the Siebert et al. (2005) map primarily represents areas *equipped to be irrigated circa* 2000.

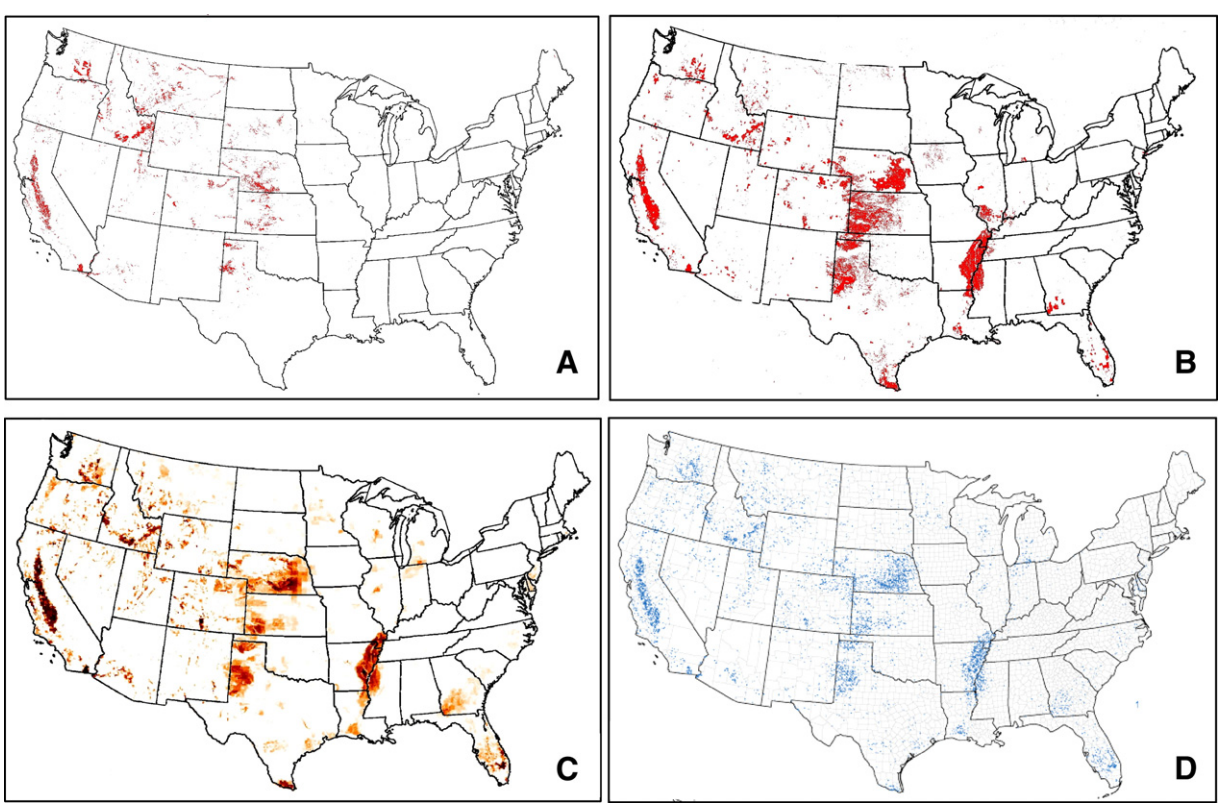

Fig. 1. Irrigation in the US represented by three different map products (A–C) and compared to the dot map made by the USDA (D). The map in the upper left (A) is the USGS map made from the AVHRR NDVI data. The upper right map (B) was made by the IWMI using 10-km AVHRR data and does not contain fractional irrigation information. The bottom left map (C) is the University of Frankfurt-FAO map showing the fraction of the total cell area equipped for irrigation ca. year 2000. The bottom right map (D) is the USDA dot map ca. year 2001 where each dot represents ~4000 ha.

However, actual area of irrigated lands can vary greatly driven by each year's water availability and the choice of crop. Therefore, this product will exhibit smaller variations in a study concerned with year-to-year variation in irrigated areas. The major shortcoming of the USGS map is that irrigated areas were determined as part of a broader classification scheme. Thus the emphasis was primarily placed on other land-cover types while irrigated classes received less attention and thus decreased classification accuracy. A recent comparison by Vörösmarty (2002) of irrigated lands depicted by the USGS map to the country-level reports of irrigated area (FAO, 1998) points to major uncertainties in classification and inventory irrigated lands due to the highly politicized nature of FAO data reports, as well as to the technical limitations of the more objective datasets made from remote sensing. The major drawback of the IWMI global irrigation map product is that the ground-truth data obtained only in India, SE Asia, Africa, and South America were used to adjust and refine global irrigation classes. This makes the IWMI product highly parameterized per region for which extensive ground data exists. However, in areas without such data, the irrigation classes may be less reliable.

The goal of our research is to overcome these shortcomings by relying on improved remotely sensed data sources and classification methodology and on ever-increasing continental and globally extensive ancillary data sources of gridded climate data and agricultural extent. The major differences between the new methodology (and the product) described here and existing large area irrigation datasets include improved spatial resolution, up-to-date information content, repeatability across time, and the use of machine learning algorithms that are intrinsically objective. The irrigation mapping effort is part of our larger research initiative to understand anthropogenic effects, specifically that of irrigation, on the global water, carbon, and energy cycles. This paper describes the development of a generic methodology for mapping irrigated lands over large areas. We apply this method to the US – a data-rich environment – to work out potential problems and understand the most important inputs to the algorithm and the development and refinement of the algorithm itself.

3. Irrigation mapping procedure

3.1. Definition of irrigation

We define *irrigation* as agricultural area that receives full or partial application of water to the soil to offset periods of rainfall shortfalls under dryland conditions. More specifically, we focus on *dryland* cultivated areas, excluding irrigated pastures, paddy rice fields, and other semiaquatic crops. The reasons for this are twofold. First, while paddy irrigated fields have a distinct remotely sensed signal associated with transplanting in water in maturity (Xiao et al., 2006), dryland irrigated conditions are much more ambiguous and require separate methods. Note that while irrigated paddy rice fields do not occupy large portions of the continental US, other countries, especially in southeast Asia, irrigated paddy rice is an important form of cultivation. Thus, as we move from the US effort to a global approach, we will have to pay more attention to irrigated paddy rice fields. Second, irrigated pastures make up a small portion of total irrigated lands and are not considered important for food production or environmental consideration such as water logging and soil salinity. We also selected a minimum mapping unit of 100 ha (roughly four MODIS 500 m pixels) to identify and map irrigated lands, in essence eliminating small irrigated lawns and gardens.

3.2. Proposed method

As part of our objective to map irrigated lands globally, we designed the mapping procedure to meet three criteria. First, the procedure must be automated and repeatable across space and time. Next, it must be robust enough to capture many different forms of irrigated lands across large geographic regions. Finally, it must rely on high quality and objective remotely sensed observations. To meet these criteria, we adapted an image classification approach to the irrigation-mapping problem, guided by a climate-based index that identifies locations that require irrigation. This index of irrigation potential was found necessary in earlier versions of our work to better identify real irrigation and reduce the rate of false positives when using remotely sensed data (Ozdogan and Gutman, 2007). Within the boundary conditions identified with the potential irrigation index, the classification of remotely sensed, multi-temporal, multi-spectral images was achieved in two steps: i) binary classification into *irrigated* and *non-irrigated* classes; and ii) fractional areal estimate of each pixel identified as irrigated in the first step. The subpixel estimation was necessary since often only a portion of individual pixels are often occupied by irrigated fields and this can significantly affect the estimates of agricultural/irrigated areas from remote sensing, especially in areas known to have small cultivated fields (Ozdogan and Woodcock, 2006).

Our classification-based irrigation mapping procedure is based on two key developments in the global remote sensing arena. First, remotely sensed inputs with improved radiometric and geometric quality, such as those provided by the MODIS instrument, have become available for continental to global scale studies. Second, a new generation of classification algorithms, such as classification and regression trees, initially developed within the machine learning community, have successfully found their way into large area remote sensing applications (DeFries et al., 1998; Friedl et al., 2002; Hansen et al., 2002).

Our irrigation mapping procedure has four parts that are schematically shown in Fig. 2. In part one, we calibrate a climatological moisture index along with existing agricultural maps to define irrigation potential. Second, we identify irrigation-related remotely sensed temporal and spectral indices. Third, we combine irrigation potential and remotely sensed indices within a supervised classification algorithm to locate irrigation. Finally, we estimate subpixel proportion of irrigation in each 500-m pixel identified as irrigated. We initially tested our procedure in the US to map irrigated lands across the entire country. In the sections that follow, we describe these steps in greater detail. In the last section, we show the initial examples from the US and present our results of validation.

3.3. Effective irrigation potential

Irrigation is practiced in many countries around the world at scales ranging from small subsistence farming to national enterprises. The location of irrigation is determined by a combination of factors including climate, resource availability, crop patterns, and technical expertise. Climate plays an important role in presence and distribution of irrigation as it determines natural moisture availability (precipitation), crop water demand (evaporation), and crop schedules. In this study, we delineated potentially irrigated areas using a climate-based index. A map of potentially irrigated areas provides the first approximation for areas that require irrigation from the climate perspective. As noted earlier, introduction of a climate-based potential irrigation index into the classification process greatly reduces ambiguity in identifying irrigation using remotely sensed inputs alone (Ozdogan and Gutman, 2007).

Over large areas, presence and distribution of irrigation is primarily controlled by natural moisture availability at the surface. For example, in arid and semi-arid parts of the world, dry atmosphere and the lack of rain-supplied moisture requires exclusive use of irrigation to grow crops. In more humid locations, on the other hand, irrigation is often supplemental and is used to meet the demand of crops whose growth cycle is out of sync with natural precipitation. Thus, climatic moisture availability provides the first level of information on potential presence of irrigation at a given location.

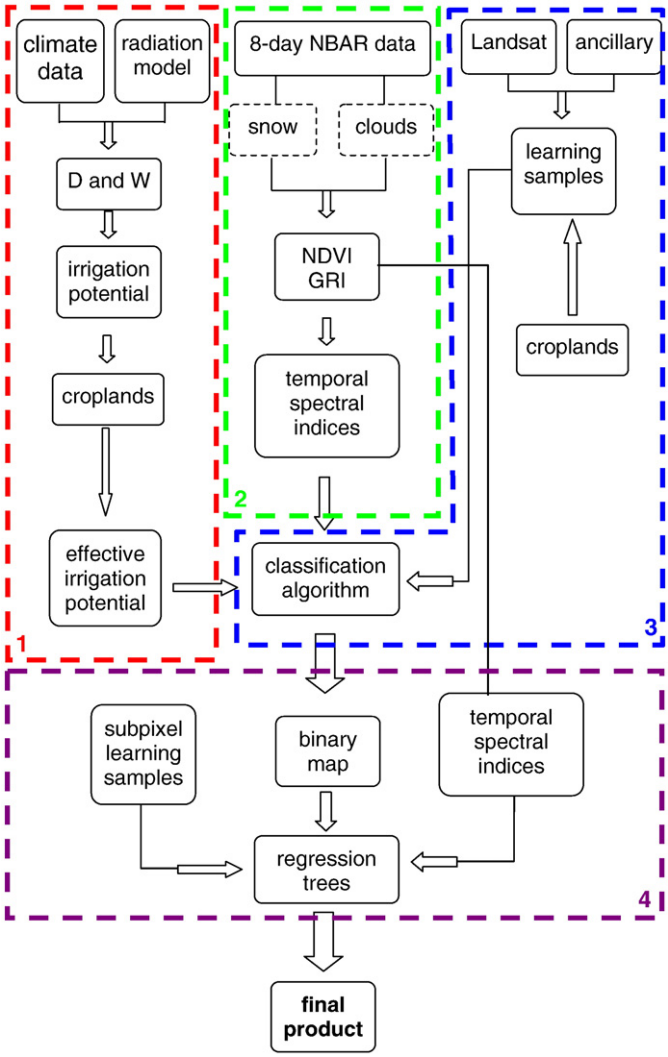

Fig. 2. Flow chart of the major steps in the proposed mapping algorithm. Each dashed box with a number refers to the processing step in the proposed irrigation mapping procedure. Please see text for details.

Budyko (1974) introduced the *Radiative Dryness Index* as:

$$D = \frac{R}{\lambda P} \tag{1}$$

where R is mean annual net radiation, which can be estimated from Earth–Sun geometry, observed mean air temperature, and observed humidity; P is mean measured annual precipitation; and λ is the latent heat of vaporization. In the present work, we calculated both R and P using gridded datasets of the WorldClim database (Hijmans et al., 2005). The WorldClim dataset is the only database that offers global coverage at around 1-km spatial resolution while providing a reliable and consistent source of climate information.

The dryness ratio has been widely used to classify climate regimes and the corresponding land cover types in simple climate models (e.g. Gutman et al., 1984). While D provides important information on climatic moisture availability, it is not directly related to irrigation. To relate D to irrigation, we plotted D against percent irrigation presence information from the GMIA product (Siebert et al., 2005). This relationship is shown in Fig. 3 on the left Y-axis as open circles (original aggregated data) with a fitted curve. The association between D and fractional irrigated area is non-linear. To linearize this relationship we follow Gutman et al. (1984) in using the empirical relationship suggested by Lettau (1969) for a water availability parameter:

$$W = \frac{\tanh D}{D}, \quad D \geq 0 \tag{2}$$

The association between W and fractional irrigated area is given in Fig. 3, right Y-axis as triangles and a linear fit of the original aggregated data.

Using this linear relationship between W and fractional irrigated area, we mapped climate-based irrigation potential, referring to it as the *effective irrigation potential*. As the final step, we used existing cultivated area masks (e.g. Ramankutty and Foley, 1998; Wood et al., 2000) to mask out those locations on the potential irrigation map that are not known to be cultivated, in essence limiting our investigation to major agricultural areas.

3.4. Remotely-sensed inputs

3.4.1. MODIS data

The MODIS instruments on board both Terra and Aqua platforms include seven spectral bands that are designed exclusively for monitoring Earth's land surfaces (Townshend and Justice, 2002). The

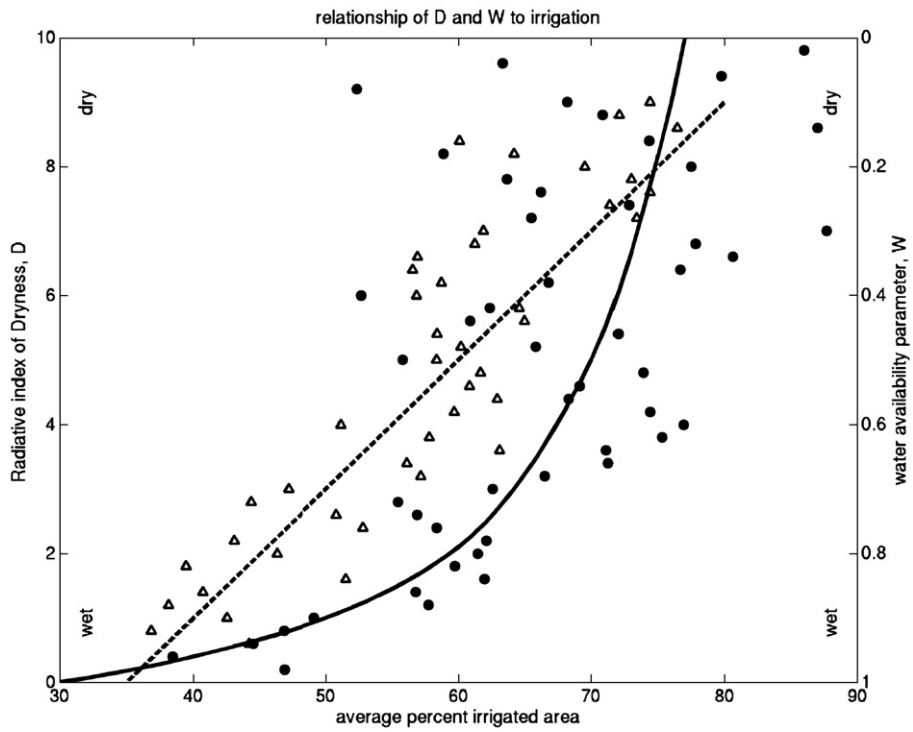

Fig. 3. The relationship between *D*, *W*, and global fractional irrigated area obtained from the GMIA product. *D* is plotted as filled circles with the fitted curve (solid), while *W* is plotted as open triangles with the straight fitted line (dashed). *W* linearizes the relationship between D and irrigated area and is used to produce effective irrigation potential. Note that the left *Y*-axis has been reversed to show consistent wet and dry directions (also indicated as text).

Terra and Aqua MODIS combined provide sub-daily global coverage at 250- and 500-m spatial resolutions and offer enhanced spectral, spatial, radiometric, and geometric quality for improved mapping and monitoring of vegetation activity. Hence, to date, MODIS land data have been an integral part of production of a variety of land cover maps, including irrigation (Friedl et al., 2002; Thenkabail et al., 2005; Wardlow and Egbert, 2008; Xiao et al., 2006).

A large array of standard MODIS data products are operationally produced by the MODIS Land Science Team and made available to the scientific community on a timely basis. One of these products is the Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance (NBAR) data (MOD34B4, Schaaf et al., 2002). This product provides cloud-screened and atmospherically corrected surface reflectances for all MODIS land bands that have been corrected for view- and illumination-angle effects. Currently, the NBAR data is produced at 500-m spatial resolution every 8 days with a total of 46 observations over the calendar year geographically organized in a MODIS tile system with the Sinusoidal Projection. In this study, we used one calendar year (2001) of NBAR data (total of 46 observations) for 14 tiles covering the continental US.

3.4.2. Irrigation-related indices

Remote sensing of irrigated lands over large geographic regions involves significant challenges both in selecting spectral bands or indices that contain the maximum amount of irrigation-related information and in relating this information to complex forms of irrigation presence. For instance, while satellite-derived indices are extremely useful in detecting vegetated areas in agricultural lands, the cause of spatial and temporal variability in biomass, whether as a result of rainfall or irrigation, is much less straightforward, hence more difficult to determine. A further complication may arise when only supplemental irrigation is practiced.

To overcome these challenges we first identified the characteristic attributes of irrigated lands, especially those characteristics that are observable with remotely sensed measurements. One such attribute is related to vegetation "greenness". There is an overwhelming consensus that the Normalized Difference Vegetation Index (NDVI) is an important vegetation monitoring tool (DeFries et al., 1998; Goward et al., 1991; Justice et al., 1985; Myneni et al., 1995; Tucker, 1979). NDVI is defined as:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{nir}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{nir}} + \rho_{\mathrm{red}}} \tag{3}$$

where ρ_{nir} and ρ_{red} respectively represent NIR and red reflectances. NDVI has been closely related to plant moisture availability (Nicholson et al., 1990), leaf area index (Xiao et al., 2002), primary production (Prince, 1991); and vegetation fraction (Gutman and Ignatov, 1998).

Although NDVI has been widely used to monitor vegetation greenness in agricultural settings under a variety of climatic conditions, overwhelmingly it is the temporal NDVI signal that has often been most related to irrigation (Kamthonkiat et al., 2005; Martinez-Beltran and Calera-Belmonte, 2001; Pax-Lenney et al., 1996; Thenkabail et al., 2005; Tucker and Gatlin, 1984). Consider an irrigated landscape. Under ideal conditions, irrigation would be practiced when soil moisture deficit occurs, leading to a strong temporal mismatch between the greenness cycle of rain-fed crops and that of irrigated crops, particularly in arid and semi-arid locations. Greenness associated with non-irrigated crops in arid/semi-arid landscapes is often a direct result of rainfall events while greenness associated with irrigated sites is generally independent of rainfall and would show a development cycle different than that of rainfed crops. This differential temporal behavior of irrigated and non-irrigated cultivation is illustrated in Fig. 4 for two relatively arid locations in the US. Here, the seasonal dynamics of vegetation greenness for irrigated (solid) and non-irrigated (dashed) croplands are illustrated in the form of mean smoothed NDVI signature (left *Y*-axis) for the year 2001. Also plotted in each panel is the monthly mean precipitation for the same year (right *Y*-axis). The top panel (A) is an example from semi-arid eastern Washington state (Northwest USA), where the non-irrigated crops

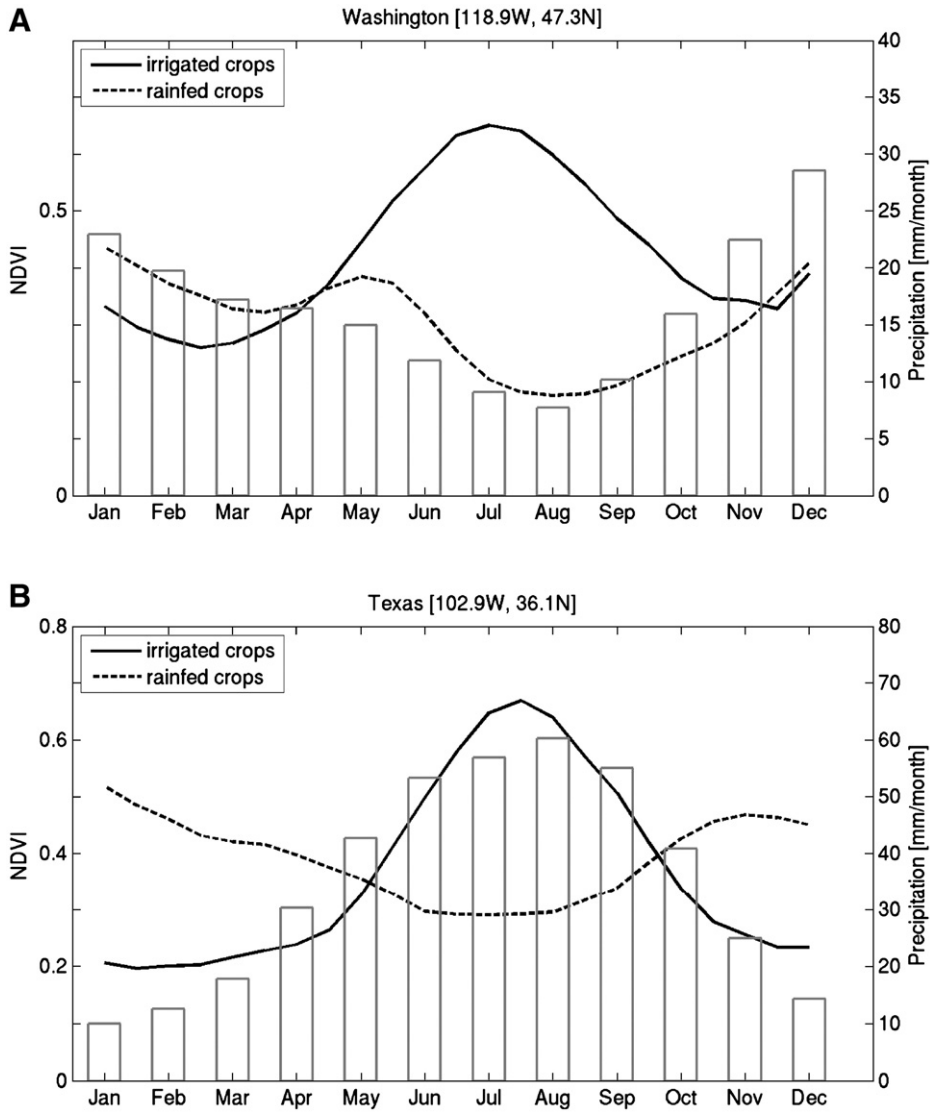

Fig. 4. Temporal NDVI and precipitation profiles of irrigated and non-irrigated fields in two different locations in the US. In each location, the lines represent average response from a sample of approximately five sites in each category.

(mostly cereals) exhibit two peaks, first following planting in the fall and second before harvest in late spring/early summer, closely following the moisture availability. In contrast, irrigated crop greenness peaks during mid-summer when moisture availability is smallest and greenness value of non-irrigated crops drops to its lowest value. Note that the lack of precipitation in the summer time at this location causes a large moisture deficit and makes irrigation absolutely necessary. The bottom panel (B) reflects a similar mismatch of greenness dynamics between irrigated and non-irrigated crops for a site in northern Texas (South Central US). However, the greenness dynamics at this location do not necessarily reflect moisture availability and irrigated crops that exhibit greenness peak about the same time when maximum precipitation occurs. Of course, the available precipitation may not be enough to meet the crop demands in this region and thus irrigation may be necessary. In fact, crop-distribution maps provided by the US Department of Agriculture's National Agricultural Statistics Service (NASS) for this region show a large presence of cotton, a crop with heavy water demand. Nevertheless, the irrigated and non-irrigated crops exhibit clearly distinct temporal greenness profiles and by monitoring the seasonal greenness behavior of crops via NDVI it is possible to distinguish irrigated crops from non-irrigated ones.

While such temporal analysis of the NDVI signal clearly reveals the differences between irrigated and non-irrigated crops, it also reveals that the timing difference of peak greenness in each location is related to the crop type. For example, both in Washington and Texas, the non-irrigated crops are predominantly winter cereal varieties (e.g. winter wheat) and the irrigated crops are primarily summer row crops (e.g. corn, soybean, or cotton). Nevertheless, to detect this crop type difference through temporal analysis, especially if related to moisture availability, also reveals information about the irrigation status of crops. For example, cereals (specifically winter varieties) are predominantly not irrigated while most summer crops are irrigated. Thus, detecting crop type through temporal analysis is used here as a proxy for detecting irrigation.

A more difficult case for distinguishing irrigated crops from non-irrigated counterparts occurs in locations where the same crop type is grown with and without irrigation in the same growing season. A prime example of this situation is central Nebraska where, depending on a particular year's soil moisture availability and installed irrigation infrastructure, irrigated and non-irrigated corn fields are often

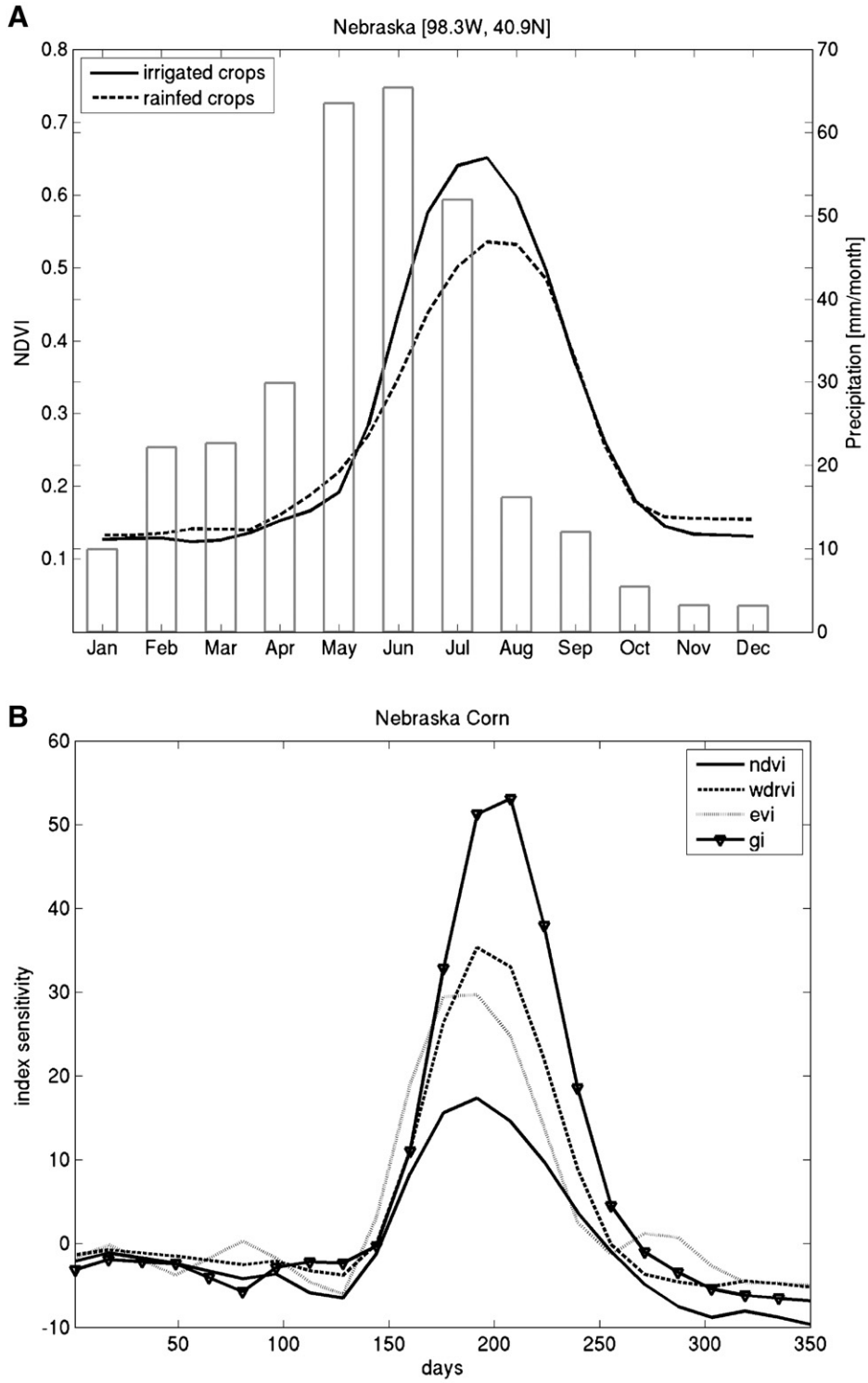

Fig. 5. (A) Seasonal dynamics of NDVI for irrigated and non-irrigated corn in Nebraska. Each line represents an average response from a sample of approximately five individual sites for each category. Also plotted is the long-term average monthly total precipitation acquired from a nearby meteorological station. (B). Sensitivity of four vegetation indices, namely NDVI, EVI, GI, and WDRVI, to irrigation presence as measured by Relative Sensitivity Index (RSI) in Nebraska for 2002. Each index was generated from an average response using a sample of approximately five individual sites. Please see the text for details on the RSI.

simultaneously present. The temporal NDVI profile of irrigated and non-irrigated corn from Central Nebraska in Fig. 5A exhibits an identical pattern with a sharp peak in greenness around late June/early July followed by a rapid decline. While irrigated corn fields exhibit slightly larger NDVI, possibly due to constant availability of moisture, the difference is small and probably useless in distinguishing irrigated fields. Thus, a more sensitive index is required to make this distinction.

A large body of research of spectral remote sensing of vegetation canopies indicates that moisture stress in vegetation is strongly manifested in spectral indices related to chlorophyll content (Gitelson et al.,

2003; Zarco-Tejada et al., 2002). One such index, suggested by Gitelson et al. (2005) to be used with the MODIS sensor, is the Green Index (*GI*) defined as:

$$\mathrm{GI} = \rho_{\mathrm{nir}}/\rho_{\mathrm{green}} \tag{4}$$

where ρ_{green} is the reflectance in the green spectral region. The theoretical foundations of the *GI* are given in Gitelson et al. (2003). In short, it is based on the evidence that in the green spectrum (centered around 510 nm) the specific absorption coefficient of chlorophylls is very low while green leaves absorb more than 80% (e.g., Gitelson and Merzlyak, 1994). In contrast, depth of light penetration into leaves in the blue and red spectral ranges is four to six times lower (e.g., Merzlyak and Gitelson, 1995). Therefore, in the green spectrum, absorption of light is high enough to provide high sensitivity of *GI* to chlorophyll content but much lower than in the blue and red to avoid saturation (Gitelson et al., 2003).

Our hypothesis for using the *GI* for irrigation assessment is that irrigated crops with very little or no soil moisture stress will exhibit larger Chlorophyll content than non-irrigated crops with potential moisture stress. To test this hypothesis and the sensitivity of the *GI* to irrigation (moisture) presence, we compared four vegetation indices, NDVI, Enhanced Vegetation Index (EVI) (Huete et al., 1999), Wide Dynamic Range Vegetation Index (WDRVI) (Gitelson, 2004), and *GI*, using the following Relative Sensitivity Index (*RSI*):

$$\mathrm{RSI} = \frac{I_{\mathrm{irr}} - I_{\mathrm{non}}}{I_{\mathrm{non(max)}} - I_{\mathrm{non(min)}}} x\,100 \tag{5}$$

In Eq. (5), I_{irr} and I_{non} are the irrigated and non-irrigated values for each index, *I*, at each time period, normalized by the seasonal amplitude (maximum − minimum) of non-irrigated values of each index. *RSI* represents the difference between irrigated and non-irrigated index value compared to the seasonal maximum change in the same index's non-irrigated value. Comparison of relative sensitivity of all four indices reveals that the *GI* shows the largest sensitivity to irrigation presence during peak crop growth (Fig. 5B). Similar results were obtained by Gitelson et al. (2006) over other maize canopies in Nebraska. Recognizing the increased sensitivity to irrigation, we developed a *GI* dataset from MODIS *NIR* and *green* NBAR bands for each 8-day period.

The temporal analysis of NDVI and *GI* data presented above reveals that the timing of phenological transition dates such as the timing of greenup, maturity, and senescence are nearly identical between the two indices while the absolute magnitude of greenness (or maximum greenness) as one indication of irrigation presence is better captured by the *GI*. In essence, the *GI* by itself captures all of the salient temporal features of irrigation in addition to the spectral signature associated with the chlorophyll content. Therefore, we developed a full year (46 observations) *GI* only dataset and incorporated it into our classification algorithm. The complete list of input features into the irrigation classification algorithm is given in Table 1. Note that the methods involving the *GI* are only applicable to sensors such as MODIS that are capable of sensing in the green part of the electromagnetic spectrum. Without this capability, NDVI may be the only data source to monitor irrigation remotely and may require new methods.

Table 1
List of input features in the classification algorithm

Input feature	Source	Frequency	Period	Description
GI	MODIS reflectance	8-days	12 months	Ch2 & Ch4 NBAR reflectance
Irr. Potential	WorldClim	static	climatology	Precip. and temp. climatology
croplands	Ramankutty & Foley	static	*circa* 1995	LC classification and country

3.5. Classification algorithms

Our goal is to separate irrigated crops from non-irrigated croplands. Remote sensing based greenness indices suggest that mapping is possible through the use of certain rules based on input data. While there are several different methods to formulate these rules (e.g. Thenkabail et al., 2005), in our research we use a supervised classification methodology derived from the tree-based models. Popularized by Brieman et al. (1984), tree-based models have been successfully used to determine distributions of a variety of biogeophysical fields using remote sensing data (DeFries et al., 1998; Friedl and Broadley, 1997; Friedl et al., 2002; Hansen et al., 2002). Tree-based methods have substantial advantages for remote sensing classification problems because of their flexibility, intuitive simplicity, and computational efficiency. Because of their ability to handle noisy and missing data as well as eliminating requirements for distributional assumptions these methods are useful for remote sensing applications, particularly at continental to global scales (Friedl et al., 2002).

For the research presented here, we consequently used two tree-based models: i) decision trees to classify the input data into irrigated and non-irrigated classes; and ii) regression trees to estimate the fraction of irrigation within each pixel that has been identified as irrigated. At the heart of our decision tree phase is a supervised decision-tree classification algorithm (C4.5) (Quinlan, 1993). Widely used in the machine learning community, the C4.5 algorithm we use has been augmented with the Boosting method (Friedman et al., 1998) to optimize classification accuracies and to provide spatially explicit class confidence estimates (McIver and Friedl, 2001). In this application, we used 10 boosting iterations as suggested in McIver and Friedl (2001). We also used C4.5's internal pruning algorithm to limit overfitting.

As in classification trees, the regression-tree algorithms produce rule-based models for prediction based on training data but allow prediction of continuous variables such as subpixel distribution of irrigation. Each rule set defines the conditions under which a multivariate linear regression model is established. We use a specific regression-tree algorithm called *Cubist* (Rulequest, 2001). The Cubist algorithm consists of a set of linear models and a set of inequality "cuts" on the variables to select among the individual linear models, yielding a piecewise linear model.

3.6. Development of a continental-scale training database

Our approach to irrigation mapping is a supervised learning methodology that requires training data that well characterizes the desired output. Thus, the training data that provide exemplars of all different types and amount of irrigation across large regions are critical. A globally representative, consistent, and accurate training database is required to establish a relationship between remotely sensed signal, climate constraints and irrigation and plays a key role in land-cover classification based on remote sensing. In addition to the quantity, the quality of the training site database strongly influences the quality of classification results. Because irrigation is highly diverse, a key requirement of the database is that it be geographically comprehensive and include variations in irrigation practices manifested across different landscapes.

In our approach, we derive two separate sets of training data, one to train the decision-tree algorithm and the other for the regression-tree algorithm. The method to derive both training datasets involves several dozen high spatial resolution satellite images acquired by the Landsat 7 ETM+ sensor *circa* 2000. The location of the training sites is chosen to represent major irrigated land areas of the US. In the case of

decision trees, each training site is derived from a corresponding Landsat scene in the form of a polygon that represents examples of both irrigated and non-irrigated lands and served as a basis for classification and quality assessment. Each site is chosen to represent a minimum of 0.25 km^2 uniform land cover (irrigated vs. non-irrigated). The non-irrigated class included all other forms of land cover including agriculture. In most cases, each training site is interpreted based on characteristic shapes of irrigated fields (e.g. center pivot irrigation systems), as well as with the help of ancillary data sources such as county-based irrigated area maps, climate data, state reports, and the internet. For example, in the Pacific Northwest, agricultural fields having healthy green vegetation cover on the mid-summer Landsat images are interpreted to be irrigated as very little precipitation is available to sustain crops during summer. Similar logic applies to most arid western states. Note that, in the more humid east, the selection and interpretation of the training data is more subjective. Here we additionally rely on county-level irrigation infrastructure data, state reports and production estimates. However, the reliability and the quality of these learning samples may be less than those obtained in the west and therefore lead to poor classification accuracies. At the same time, the eastern part also represents a smaller portion of the total irrigated area of the United States and hence potential errors would be expected to affect a smaller area where irrigation is mostly supplementary in nature.

For training sites to be used in the regression tree, we first classify each Landsat image (wall-to-wall) into irrigated and non-irrigated classes at native resolution (30 m) and then aggregate these maps up to 500 m using a square-wave filter. The outcome is a small map that shows the fraction of irrigated area in each 500-m pixel as a continuous variable. The quality of the classified maps at 30-m resolution ranged from 68% in the east to 95% in the west, based on a 10-fold cross-validation exercise. While it is hard to predict the impact of lower accuracy samples on the training of the Cubist model, aggregation of the high-resolution samples to MODIS scale would be expected to reduce labeling errors. Moreover, by filtering the fractional irrigated area samples by pixels that have been identified as irrigated at the decision tree phase, we reduced the occurrence of these labeling errors.

Note that we rely heavily on the recently available Landsat datasets (either free or at cost of reproduction) such as those from the University of Maryland Global Land Cover Facility (http://glcf.umiacs.umd.edu/data/landcover/index.shtml). The existence of these databases significantly reduces the cost of our research. Moreover, these datasets are global in nature and thus the same procedure can be used to extract a global training database.

4. Implementation of the proposed method in the continental US

The first step in implementing the irrigation mapping algorithm for the continental US is to screen the MODIS time series data for clouds and snow cover as well as for non-agricultural land cover. For cloud cover, we use the quality control cloud flags included in the NBAR data files in the form of cloud masks for each 8-day time period for a total of 46 files per year. Pixels corresponding to these cloud masks are then excluded from the analysis. To minimize the potential impact of snow cover, especially in the spring months, we use MODIS snow masks (Hall et al., 1995) for each of the 8-day periods; pixels identified as snow are removed from further analysis.

The second step in the implementation of our algorithm is to train the decision tree model (tree growing stage) with the example datasets derived from the training sites described above to automatically generate the rules and thresholds that identify irrigation. We then apply the trained decision tree to all of the individual MODIS tiles that cover the continental US. Each tile consists of cloud- and snow-screened time series of GI values representing a total of 46 per-pixel observations. The end result of the classification algorithm is a two-class map showing irrigated and non-irrigated areas over the continental US.

In the last step, we train and apply the regression-tree algorithm to estimate the subpixel distribution of irrigation presence as a continuous variable within each 500-m pixel identified earlier as irrigated. Our final product is a continental scale map of irrigation distribution *circa* 2001 (Fig. 7).

5. Validation

Accuracy assessment of moderate resolution (500-m to 1-km) land-cover products is a challenging task as these maps cover large areas and can over- or under-estimate areas of land-cover types due to the fragmentation and subpixel proportion of individual types (Wulder et al., 2006). Because of budget constraints and resource limitations, we are not able to conduct extensive field surveys for collection of site-specific data. Instead, we rely on freely available high-resolution data (Ikonos and Landsat), as well as assembled sub-national (county- and state-level) irrigated agriculture statistical datasets. We evaluate the MODIS-derived irrigation map in three ways: (1) statistical estimate of the binary irrigation map accuracy (resulting from the decision-tree phase) using an independent set of ground-truth observations selected from a probability-based sample design; (2) comparison of area estimates made from the continuous irrigation map (resulting from the regression tree phase) to sub-national level (state and county level); and (3) comparison to other irrigation map products such as those reported in Section 2.

As part of the first method, we select a two-tier stratified random sampling approach where accuracy assessment samples are stratified on both geographic distribution and class type. To accomplish this, we first divide the country right from the middle and generate 100 randomly-located accuracy assessment sites (points) in each half with a total of 200 samples. This first-order geographic stratification is necessary as the western and eastern halves of the continental US contain very distinct irrigation presence. In the arid west, irrigation is often absolutely necessary to grow crops during very dry summer growing seasons. Indeed, roughly 75% of total irrigation applied in the US occurs in the arid west (Gollehon and Quinby, 2000). On the other hand, the eastern US primarily supports supplemental irrigation. This stark difference of moisture availability and irrigation practices between the eastern and western portions of the country requires careful consideration of how the MODIS-based irrigation map should be evaluated. Our qualitative assessment of the binary irrigation map (resulting from the decision tree stage), based on a visual comparison to the county-level irrigation database, suggests that the map is more accurate in the west than it is in the east.

Within each geographic area, we also stratify our test samples on the class label (irrigated vs. non-irrigated) and select approximately 70% of the sites from the non-irrigated class and 30% of the sites from the irrigated class. The increase in the number of samples from the non-irrigated class is directly related to its larger area. We identify the irrigation status of each site using available high-resolution datasets from Landsat, as well as from those available through Google Earth (http://earth.google.com). We use these high-resolution datasets as a substitute for ground truth, although these images may be snapshots of cropping systems and irrigation status taken at different times. The high-resolution datasets have the advantage of being comprehensive, providing information on much larger areas, and, therefore, more representative. When labeling these test sites, we rely on shapes of cultivated fields such as center pivot irrigation, ancillary information at the county level, and other sources of information extracted from reports and the internet. Once identified, we compare the class label of each test sample to the class label of the pixel extracted from the binary irrigation map at the same location within a confusion matrix — a standard tool for assessing accuracy of land-cover classifications (Table 2).

Table 2
Confusion matrices for the eastern (top) and western (bottom) US

		Predicted		
		non-irr	irr	total
Observed	non-irr	70	19	89
	irr	2	9	11
	total	72	28	100
Observed agreement = 0.79				
Chance agreement = 0.67				
Kappa = 0.36 (±0.2)				
Observed	non-irr	69	1	70
	irr	12	18	30
	total	81	19	100
Observed agreement = 0.87				
Chance agreement = 0.62				
Kappa = 0.65 (±0.16)				

Note that, when compared to the western US, the quality and accuracy of the ground truth data in the eastern US may be less than ideal. This is partially due to inherent difficulty of correctly identifying what is and isn't irrigated in this region as irrigation is not very extensive and often supplementary in nature. Thus, it is possible that the accuracy assessment for the eastern US provided in the following sections maybe slightly inflated. However, the eastern US also represents a smaller portion of the total irrigated area in the continental US and hence errors would affect less area.

We also evaluate the MODIS irrigation map using sub-national statistics (both at state and county level) obtained by the NASS as part of the 2002 Census of Agriculture (http://www.nass.usda.gov/Census_of_Agriculture/). This dataset provides the total irrigated cropland area sown to various crops based on a sample of farms and contains data from a year (2002) similar to our remote sensing analysis. Results of this comparison are provided in the next section.

The last method of evaluation of our methodology and its results is to compare the final MODIS-based irrigation intensity map to other irrigated area maps available from a variety of sources. One such map is for the state of California where a digital database of distribution of agricultural resources, including irrigation, exists and is available for each county. While the dataset spans a few years around the target year of 2001 due to a rolling census approach, it represents the best available digital map of irrigation in the state (California DWR).

6. Results

6.1. Decision rules

The first type of result is concerned with the decision rules generated at the decision tree training phase presented in graphic form in Fig. 6. This representation of the tree structure is informative as it provides physically-based meaningful interpretation. Note that the tree structure in Fig. 6 was generated from data but represents a simplified version of the decision process made using only a third of

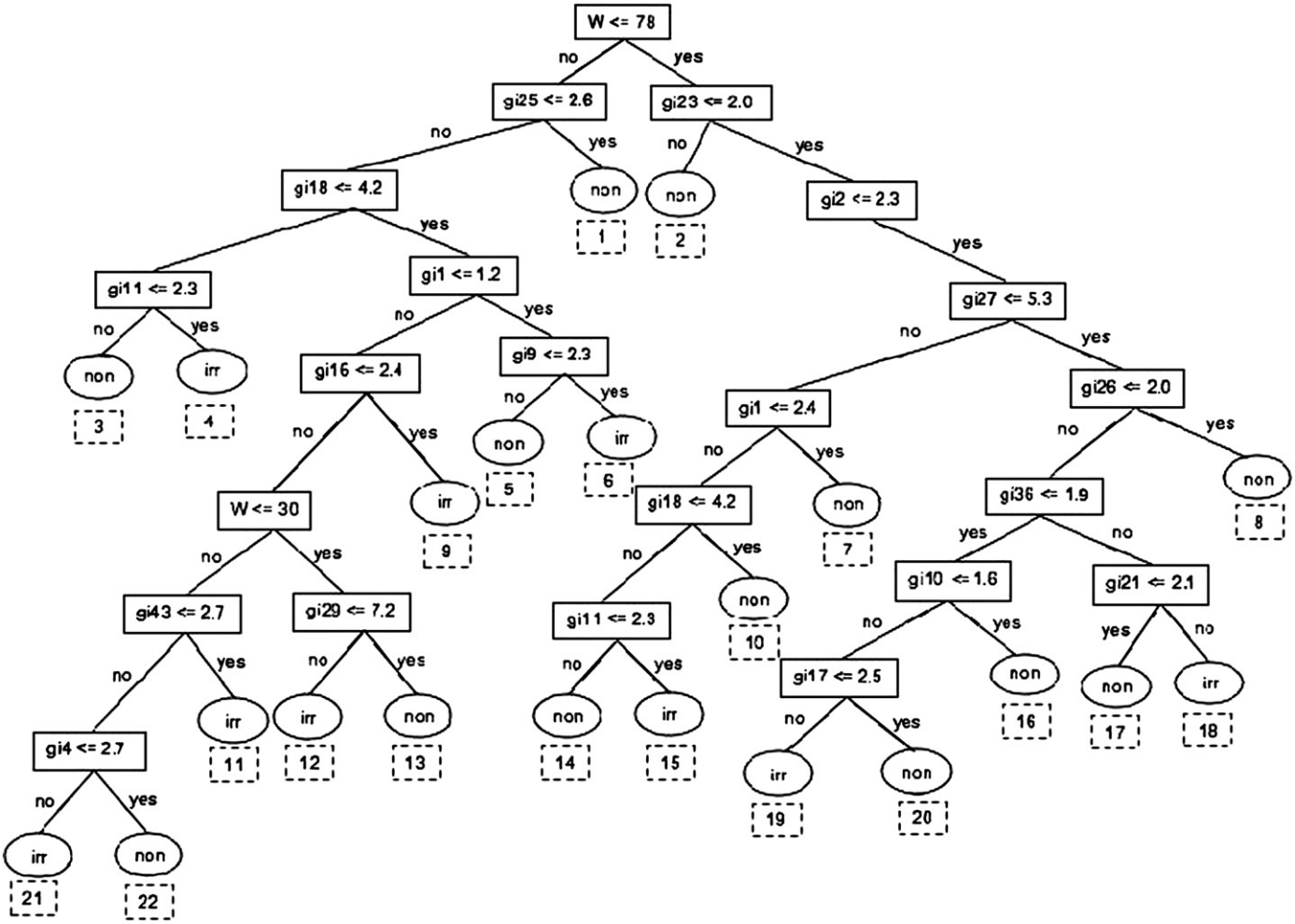

Fig. 6. Simplified decision tree that displays decision rules for mapping irrigation. Each decision node (rectangle) shows the green index observation time (e.g. gi11 means Green Index observed on observation period 11 – Julian day 88) and leaf nodes (ellipse) shows prediction results.

the training data, selected randomly. The input features and decision boundaries are given in rectangles and the terminal nodes (leaves) are shown in ellipses. Several important observations can be made by examining this tree structure. First, the tree is balanced, meaning that it uses roughly an equal number of input features and splits throughout the decision process following the initial node at the top. Next, the first split in the tree uses the Effective Irrigation Potential, represented by the Water Availability Parameter W, indicating the importance of this variable in irrigation mapping. This initial split represents moisture availability that is so important in discerning irrigation and roughly coincides with the break in the east-west moisture gradient across the continental US, hence slightly different decision rules are applied to each portion. This is also intuitive, as our climate-based effective irrigation potential shows very little irrigation probability for locations with small moisture availability (i.e. eastern US).

Following the splitting rules down the tree path further provides evidence for the decision process to map irrigation. Take the non-irrigated decision made at node 2, for instance. A pixel is labeled "non-irrigated" when climate-derived irrigation potential is lower (<78%) and there is enough mid-season greenness (gi23>2.0) typical of conditions in humid mid-western states. A more complicated irrigation labeling is exemplified by node 6 where a pixel with high irrigation potential (W>78%) and high mid-season greenness (gi25>2.6) *and* low early- to mid-season greenness (gi18<4.2, gi1<1.2, and gi9<2.3) would be labeled as irrigated, suggesting conditions for drier mid-western states such as Nebraska. Contrast this with node 3, where a pixel with similar high irrigation potential (W>78%) and high early-to mid-season greenness (gi25>2.6 and gi18>4.2) but low early-season greenness (gi11<2.3) would be labeled as "irrigated." These rules exemplify typical conditions in southern California where a long growing season and water availability allows irrigated forage crops such as hay and grass.

These results suggest a set of physical/logical rules that use concurrent information from multiple variables (temporal greenness and irrigation potential in this case) to discern irrigated agriculture. Note that an additional benefit of the tree-based analysis used here would be to project rule/node information onto a geographic space since each pixel's predicted label has a known decision path through the tree structure. Our preliminary analysis (not shown) indicates that there is a strong relationship between geographic distribution of attributes (different time slices of MODIS GI) and irrigation, in essence revealing how irrigation/non-irrigation class labels are assigned on the basis of both the value *and* geographic position of each input feature. This sort of information is useful, especially for expanding the irrigation mapping work to other parts of the globe because it allows geographic discovery of the relationship between irrigation presence and the remotely sensed signal (Andrienko and Andrienko, 1999).

6.2. Map accuracy

The purpose of assessing the accuracy of a MODIS irrigation map is to provide objective information on classification results and its error structure. As described above, we used traditional confusion matrix approach at the decision tree classification phase and then compared the area estimates obtained from the subpixel mapping from the regression tree phase to those reported at county and state levels. The confusion matrices provide information on the accuracy of the classification process as applied to an independent set of observations. In this section, we report the results of the confusion matrices for the eastern and western portions of the continental US separately (Table 2). The eastern portion has an index of agreement of 0.79 while the western, and more heavily irrigated portion of the country has 0.87 index of agreement. A better index to determine map accuracy from confusion matrices is the Kappa Index, which compares the agreement against that which might be expected by chance. Kappa can be thought of as the chance-corrected proportional agreement, and possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to −1 (complete disagreement). For the irrigation map that represents the eastern half of the country the Kappa statistic is around 0.36 (0.16–0.56 at 95% CI), which is considered to represent a fair agreement between independent and mapped irrigation (Landis and Koch, 1977). For the western half, on the other hand, the Kappa statistic increases to 0.65 (0.49–0.82 at 95% CI), which is considered to be a substantial agreement. These accuracy measures suggest that many irrigated lands in the arid western

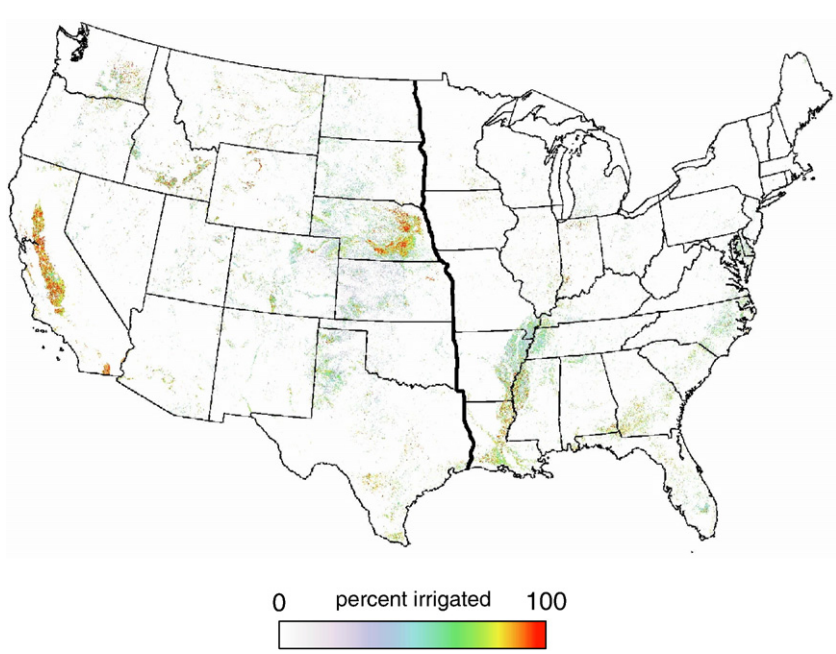

Fig. 7. Spatial distribution of irrigation in the US *circa* 2001 mapped form MODIS and ancillary data using the proposed procedure. The thick vertical line separates the east and west portions of the country individually selected for accuracy assessment. See text for detailed description of the procedure.

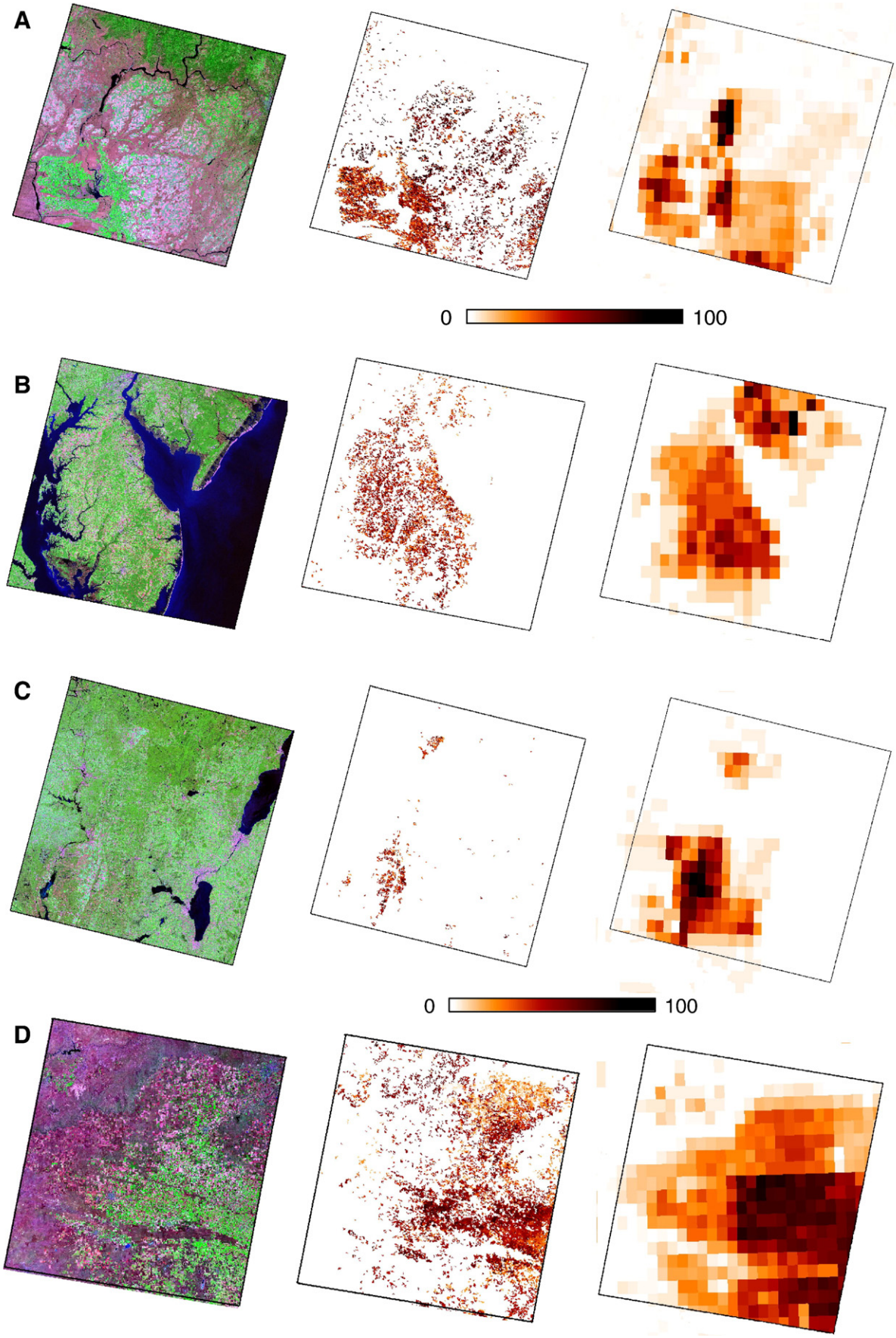

Fig. 8. Regional views of irrigation depicted by Landsat ETM+ data (left), MODIS-based procedure (center), and by the FAO map (right) in the Columbia River Plateau (A), eastern US (B), midwestern US (C), and south-central US (D).

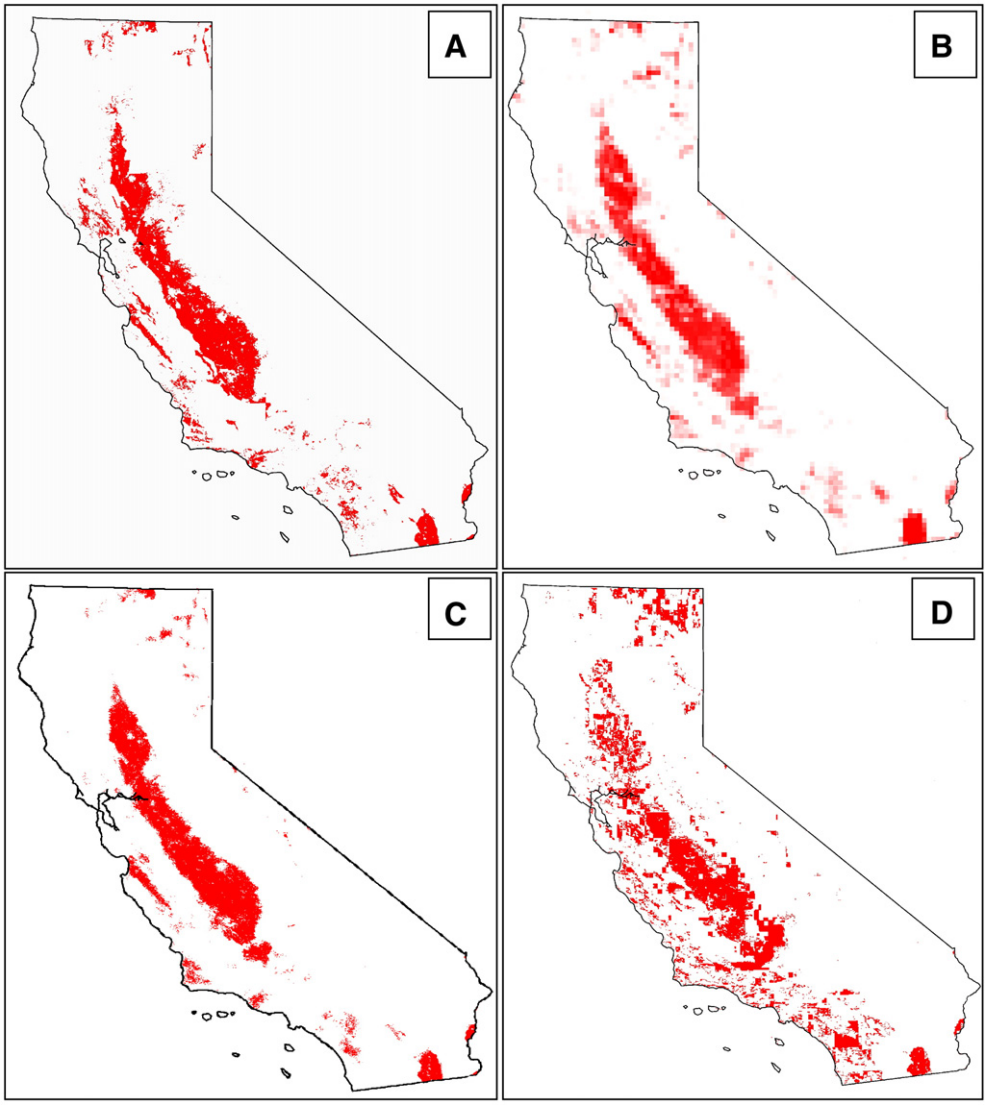

Fig. 9. Comparison of MODIS irrigation map (C) to other irrigation map products in California. The other data sets are from California DWR (A), the FAO map (B), and the GIAM map (IWMI-Sri Lanka) (C).

portion of the continental US are more accurately represented by the MODIS irrigation map than the irrigated lands of the eastern portions of the country.

6.3. Spatial distribution of irrigated agriculture from MODIS

The spatial distribution of irrigation across the continental US is given in Fig. 7 and shows some expected patterns. First, irrigation has a strong east-west divide. The MODIS map indicates that over 70% (by area) of all irrigated lands in the US are located in the western part of the country primarily due to its relatively arid nature, a result also supported by Gollehon and Quinby (2000). The eastern part of the country, on the other hand supports less irrigated land area, primarily in the form of supplemental irrigation, although not distinguished here.

At the continental scale, major irrigated areas are distributed along dry lowland valleys in the west such as the Central Valley of California, the Snake River Basin of Idaho, and in the Columbia basin of the Pacific Northwest. In the central parts of the country, irrigation occurs in small, dry depressions of Colorado, in the Texas Pan Handle, in western dry portions of Kansas and Oklahoma, and in major corn growing regions of Nebraska. In the Eastern parts of the country, irrigation is found in locations where mostly sand soils and high water-demand crops such as cotton occur including the Mississippi valley, southwestern Georgia, and along agricultural areas of the eastern coastal plain.

More detailed, regional representations of irrigation mapping results in four locations with very different climatic and irrigation patterns across the continental US are given in Fig. 8. In each row in Fig. 8, the left panel shows the Landsat ETM+ view of the landscape (spectral bands 7, 4, and 2 as RGB) during the peak growing season where bright green color indicates irrigated agriculture *circa* 2000. The center panel shows the MODIS continuous irrigation map and the right panel shows the same area mapped as *equipped to be irrigated* circa year 2000 by Siebert et al. (2005) (hereafter the FAO map). The first row (8a) depicts the irrigated landscape of the arid Columbia River Plateau in the Pacific Northwest. The plateau is comprised mostly of agricultural fields, punctuated by river valleys and surrounded by lush, green forests. While the plateau is primarily devoted to agriculture, only a part of this agricultural landscape is irrigated during summer. In the Landsat image, the irrigated fields are depicted by green vegetation in the southwestern corner of the image while non-irrigated, summer fallow fields show up as bright purple color. Both the MODIS and FAO irrigation maps accurately identify this

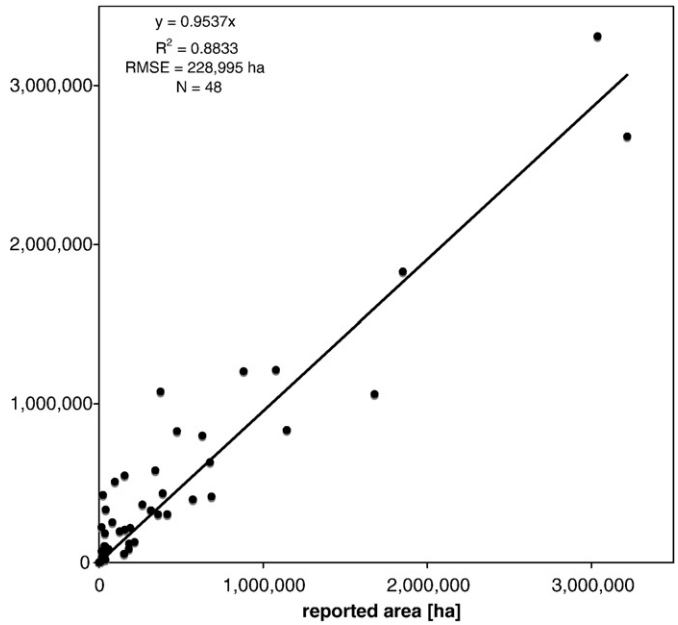

Fig. 10. Correlation between NASS-reported (2002) and MODIS-estimated (2001) irrigated area at the state-level.

pattern of irrigation presence but the MODIS map does so in greater detail.

In contrast to the arid Columbia plateau, the second row (8b) shows an example from the more humid eastern US where the practice of irrigation is more supplementary in nature (Solley et al., 1998). Here, agricultural fields are interspersed with forested areas and irrigation mostly occurs over sandy soils that, in recent years, support large acreages of soybean crop. Again the MODIS irrigation map captures this pattern of irrigation as accurately as the FAO map but in greater detail.

The third regional view shows an area in north-central US in the state of Wisconsin (8c). This example is particularly interesting because in this area, which receives ample natural precipitation during the growing season (~700 mm annual average of which roughly 60% falls during the growing season), large scale irrigation is only practiced in what is known as the "Central Sand Plain", a diamond-shaped area in the lower left portion of the Landsat ETM+ image. Here a large expanse of sandy soils allows rapid infiltration of precipitation and depresses the amount of plant available soil moisture, making irrigation necessary, especially in below normal precipitation years. As portrayed in the middle panel, the Central Sands area is nicely depicted by the MODIS irrigation map. What is interesting about this example from the irrigation mapping perspective is that the majority of the landscape shown in the Landsat image is agriculture where crop phenology is very similar to that observed in the Sand Plain area. Yet the irrigation mapping procedure accurately identifies only the irrigated sites.

The example shown in the last row (8d) is from the south-central US in the state of Texas, where large scale groundwater-fed center-pivot irrigation is very common. The summer irrigation crops in the western half of the Landsat image are accurately depicted in the MODIS irrigation map (center) and provide detail that is missing from the FAO map. In an area where accurate information on irrigated area and related water use from the groundwater supply by irrigated crops is critical for irrigation sustainability, the detail provided by the MODIS approach as well as its potential for operational status is of critical importance. Note that our primary goal with these comparisons is to highlight the differences in spatial detail between the MODIS and FAO maps. The larger irrigated area depicted by the FAO map is expected given its maximal nature.

Note that this qualitative evaluation of regional map views is not the same as statistically based accuracy assessment using high resolution data. Large area land-use/land-cover studies that use medium- to coarse-resolution (e.g. MODIS) data often use Landsat type datasets for quantitative accuracy evaluation (Friedl et al., 2002; Loveland et al., 2000; Xiao et al., 2005). While we also use quantitative analysis to assess the binary map accuracy, the purpose of map comparison presented in Fig. 8 is to suggest to the reader that both the location and intensity of irrigation predicted by the new method exhibit accurate patterns. Ideally, independent estimates, especially of irrigation intensity, would be used to evaluate how well areas are predicted by the new procedure. However, because of budget limitations, we were not able to conduct extensive field surveys across the US. Nevertheless, the visual interpretation of Landsat ETM+ images in Fig. 8 suggest that our irrigation-mapping algorithm accurately identifies both location and density of irrigated fields across a variety of moisture and cultivation conditions.

Another regional view, this time over the state of California, (which is the largest irrigating state in the US), compares the irrigation map generated as a result of the work presented here to three other sources of irrigation data (Fig. 9). The upper left panel is the irrigation data from the California Department of Water Resources (hereafter the

Table 3
Irrigated area of US states in hectares derived from MODIS (right column) and reported by USDA-NASS for the year 2002 (left column)

State	NASS [ha]	MODIS [ha]
Alabama	41,264	334,539
Arizona	359,347	304,985
Arkansas	1,677,263	1,061,009
California	3,216,867	2,680,779
Colorado	881,708	1,204,303
Delaware	39,264	59,981
Florida	686,452	415,996
Georgia	343,183	580,214
Idaho	1,145,253	834,801
Illinois	158,022	207,223
Indiana	126,247	197,329
Iowa	57,238	85,107
Kansas	1,078,791	1,213,093
Kentucky	14,859	223,411
Louisiana	375,569	1,076,736
Maine	7929	9160
Maryland	32,317	104,677
Massachusetts	9514	5336
Michigan	183,714	120,836
Minnesota	182,001	85,439
Mississippi	475,491	827,926
Missouri	415,699	304,661
Montana	629,956	800,138
Nebraska	3,037,935	3,311,778
Nevada	216,366	130,591
New Jersey	38,766	19,248
New Mexico	264,734	366,539
New York	27,889	16,400
North Carolina	96,455	509,130
North Dakota	81,350	253,570
Ohio	16,139	72,107
Oklahoma	190,584	219,676
Oregon	572,964	397,380
Pennsylvania	16,544	32,989
South Carolina	37,148	184,403
South Dakota	155,468	548,427
Tennessee	23,616	426,727
Texas	1,849,955	1,831,209
Utah	315,750	330,027
Virginia	38,564	101,828
Washington	675,796	632,405
Wisconsin	153,936	55,839
Wyoming	388,672	437,005
United States	20,369,456	22,623,781

States with less than 5000 ha are omitted.

DWR) and is considered the best available map of irrigation over the entire state, although it is updated only every few years. The upper right panel shows the FAO map, which closely matches the patterns shown in the DWR but at a coarser resolution. The lower left panel is the MODIS-based irrigation map presented here and, with a few exceptions, shows a strong correlation with the FAO and the DWR map. Finally, the lower right panel portrays the IWMI GIAM dataset at 8-km resolution and this dataset shows the least correlation to the DWR dataset.

6.4. Comparison of area estimates

An important use of large area irrigation datasets concerns area estimates. To better understand how well the MODIS based irrigation map compares with the estimates reported by the NASS, we aggregate the 500-m MODIS product to two levels of sub-national polygon maps, state and county. At the state level, we calculate the area of irrigated land in each state and compare it to the NASS reported estimate both in tabular and graphic forms. As shown in Fig. 10, MODIS-based and reported estimates at the state level agree reasonably well with an r-squared value of 0.88 and a root mean squared error (RMSE) of 2290 km^2 based on a simple linear regression model. The MODIS map estimates the total irrigated area of the US to be a little over 226,000 km^2, which is similar to the NASS estimate of 203,694 km^2 (Table 3). Estimated and reported irrigated areas for individual states are also given in this table as well. Note that there are also areas with significant disagreement between two datasets, especially in the eastern US where high annual rainfall and the mosaic nature of agricultural and natural vegetation make it difficult to distinguish irrigated from rainfed agriculture. Potential irrigation index derived from average annual inputs further complicates this issue. Further discussion on this is provided in the next section.

At the county level, we calculate the percentage of irrigated area estimated with the MODIS map as well as irrigated area reported by NASS over the total land area for each county, normalizing comparisons between large and small counties. We compare these results in Fig. 11. Except for a few significant differences, the spatial pattern of the MODIS irrigation map strongly resembles that of NASS.

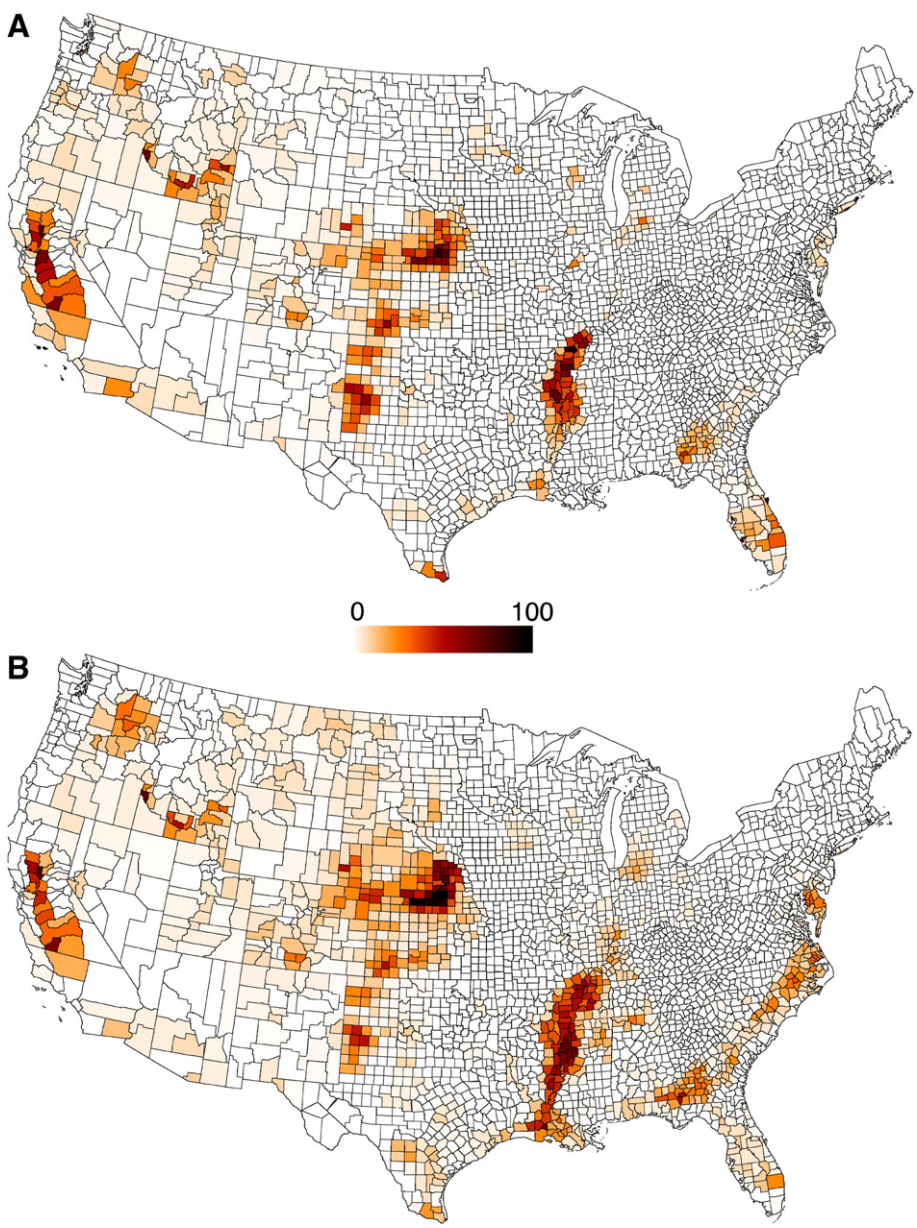

Fig. 11. County-level irrigated area presented as a percentage of total county area in 2002 in NASS dataset (A) and MODIS-irrigation map in 2001 (B).

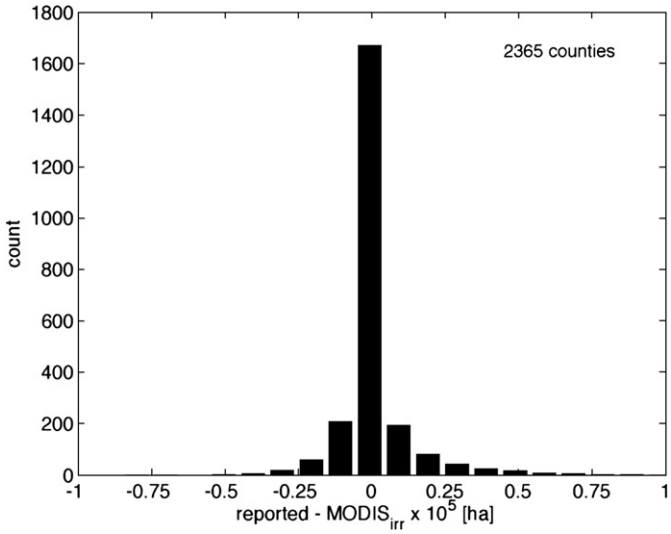

Fig. 12. Histogram distribution of the difference between NASS-reported and MODIS-estimated irrigated area at the county-level.

For instance, the pattern and fraction of irrigated area in most of the counties in large irrigating states such as California, Nebraska, and Texas show remarkably good agreement with those reported in the NASS dataset, indicating the strength of our irrigation mapping methodology. Moreover, the large concentration of bias between the MODIS and NASS maps near zero (Fig. 12) further lends credibility to remote sensing of irrigation through the method we propose.

7. Discussion

The results of the new irrigation mapping procedure presented here suggest that the MODIS algorithm is capable of identifying both the location and intensity of the majority of irrigated lands across the continental US, but there remain areas of significant disagreement between area estimates from the MODIS map and those reported by NASS, even at the county scale. These areas are mostly concentrated along the eastern US where high annual rainfall and vigorous natural vegetation make it difficult to distinguish irrigation from rainfed agriculture and natural vegetation. This problem is further exacerbated by the choice of a climate-based potential irrigation index that makes use of average annual inputs. The need for irrigation over agricultural fields in relatively humid eastern US area is directly related to sporadic, unreliable water supplies through precipitation that often comes in the form of fast moving fronts. This, combined with the sandy nature of the soils in the region significantly diminishes the plant available water content, thus creating the need for irrigation, even though annual total precipitation is always relatively high. However, because of its annual nature, our current potential irrigation index does not capture this phenomenon well and, as a consequence, assigns lower probability to irrigation in this region. While not tested here, an ideal potential irrigation index would be able to capture both the frequency and duration of natural water availability from precipitation and combine this information with soil types to identify locations where irrigation maybe required. This approach may become especially important for areas with small field size distributions outside the US where an irrigation-related remotely sensed signal would be weakened by mixed pixel effects.

The second reason for the observed difference between the NASS reported irrigated area and the MODIS estimates at the state level is related to the timing of each dataset, especially in more arid western states. The NASS irrigated area statistics are extracted from the 2002 Census of Agriculture (NASS, 2002) and are based on a sample over the census year. On the other hand, the MODIS irrigated area is a direct observation, pixel-by-pixel, of conditions in 2001. In any particular year, the irrigated area maybe reduced as a response to infrastructure or water scarcity issues. In fact, 2001 was one of the dry years with below normal precipitation levels in California (NOAA-NCDC, 2001). Reduced irrigated area is one response to this limited water availability and thus less irrigated area prediction (less than the NASS reports) by MODIS is highly likely.

While not a significant source of error in the US, the presence of small patch size for agricultural areas around the world would be one of the challenges when the methods developed here are applied globally. A long line of research suggests that when landscape patches (cultivated fields) cover a small fraction of a pixel, accurate estimation of the land cover of interest and its fraction remains challenging. This could be further exacerbated by the presence of topographic relief and persistent cloud cover over the growing period that is most important for dicerning irrigation (Thenkabail et al., 2005).

The purpose of the new mapping procedure presented here is to develop an operational irrigation-monitoring tool applicable over large areas using remotely sensed inputs. The algorithm involves a two-step mapping procedure: first, a climate-based potential irrigation index and remotely sensed temporal and spectral indices are combined within a decision tree-based supervised image classification tool to map the presence of irrigation; in the second step, continuous subpixel distribution of irrigation is predicted with a regression-tree tool for those pixels identified as irrigated in the first step. This two-step approach is considerably different than other well-publicized continuous field mapping projects such as the Vegetation Continuous Fields (Hansen et al., 2002) or impervious surfaces (Yang et al., 2003), which directly estimate subpixel proportions of the geophysical field of interest. There are several reasons behind the two-step choice for irrigation mapping in this study. First, irrigated lands have non-uniform distribution across large landscapes. They depend on moisture availability, crop type, infrastructure presence, and cultural traditions. For example, in any given year, irrigated and non-irrigated fields could simultaneously occur on the same landscape. As a result, irrigation exhibits complex patterns of spectral-temporal remotely sensed signatures that make it difficult to map their location, let alone their intensity. Hence, a two-step approach that first identifies the location and then the intensity of irrigated fields improves the quality of final predictions. Second, identification of irrigated agricultural fields with remote sensing is complicated by the fact that irrigated landscapes are a subclass of croplands that have traditionally been difficult to map with remote sensing, especially over large areas (Friedl et al., 2002; Loveland et al., 2000). Agricultural fields (and therefore irrigated agricultural fields) are highly dynamic because each cultivated field may be at a different stage of development and thus easily confused with natural land cover classes. As the desired accuracy of a land-cover map is often inversely related to its categorical detail, mapping irrigated agriculture as a subclass of agriculture – already an inherently difficult category to identify and map – becomes even more difficult.

While it is impossible to predict the exact uses of the MODIS-derived irrigation dataset presented in this paper, it is highly likely that one major use (at least by the authors for their irrigation/climate relationship study) will be to assess irrigation water requirements and use by major irrigated cropland areas. As modeling of irrigation water requirements is (among other things) directly related to the area of irrigated crops, it is likely that the errors in the MODIS map will propagate through irrigation water use estimates. This is especially true for those states where the largest discrepancy between the NASS and MODIS datasets occurs (e.g. Alabama, Kentucky, North Carolina, North Dakota, Ohio, South Carolina, South Dakota, Tennessee) (Table 3) and users should exercise caution when using the MODIS-derived dataset in those locations. However, it is also important to note that these states are not necessarily the most arid states with major irrigation needs such as Texas, California, and Nebraska. Thus, calculated irrigation water requirements for these areas

will be much less than the major irrigation states and errors related to area estimates may be less important.

Despite these challenges, the application of the MODIS irrigation algorithm to the continental US produces a reasonably accurate map with several advantages over existing products. First, the new procedure is automated in a sense that the decision- and regression-tree tools automatically identify irrigation based only on the relevant information using a physically based set of objective rules. This objectivity by itself leads to improved prediction accuracies that are free of user interpretation of irrigation-related spectro-temporal indices (e.g. Thenkabail et al., 2005). The new algorithm may also be generalized across both space and time. In terms of space, signatures generated from the US can be directly applicable to other areas with similar irrigation conditions without having to retrain/relearn patterns of irrigation presence. In terms of time, the same signatures can be applied across years to map irrigation over time without the necessity of recreating training samples for each year for which the map is intended. This becomes particularly important when we wish to quantify change and variability of irrigated areas across time under population and climate pressure with the least amount of error. This type of change analysis is considered superior to an approach that uses two moderate-resolution maps generated from algorithms based on spatial pattern recognition to infer land-cover and land-use changes. Given the possibility that irrigated lands will expand globally to feed growing population, as well as to increasingly accommodate bioenergy production, our methodology is well positioned to understand changes in irrigated areas. The successful application of our algorithm in the US is a testament of the capability of the new procedure to map irrigation over large, continental areas as an input to a variety of hydrologic, agronomic, economic, and climate studies.

Experience shows that the quality and quantity of the learning dataset is of paramount importance for accurate predictions of the location and intensity of irrigation with tree-based models. Currently, the irrigation learning dataset contains examples only from the US where they are eventually applied. As the irrigation mapping procedure is expanded worldwide, more training data, representing an even greater variety of irrigation-related signatures, will be added to our database. This global learning sample will then be capable of locating irrigation in many different environments, especially when combined with a better climate-based potential irrigation index.

Finally, the spatial detail provided by the new MODIS irrigation map directly serves the needs of hydroclimatological studies where information on landscape heterogeneity is important. It is becoming increasingly clear that changes in land surface properties involving croplands, especially irrigated croplands, influence local and regional climates and hydrology by modifying the partitioning of water between the surface and the atmosphere (Bonan, 2001; Lobell et al., 2006; Moore and Rojstaczer, 2002; Otterman et al., 1990). However, at continental and global scales, these climatic and hydrologic effects of irrigation are still poorly understood, primarily because reliable datasets that can resolve the presence of dry (non-irrigated) and wet (irrigated) patches have not been available. The MODIS-based irrigation map is one way to fill this gap.

8. Conclusions

In this paper, we present a novel classification-based irrigation mapping procedure that utilizes MODIS time series data coupled with ancillary data on climate and agricultural extent. The procedure is built on four parts: i) calibration of a climatological moisture index along with existing agricultural maps to define irrigation potential; ii) identification of irrigation-related remotely sensed temporal and spectral indices; iii) supervised decision tree-based classification of remotely sensed input and irrigation potential; and iv) estimation of subpixel distributions of irrigation within each pixel identified as irrigated. Application of these steps to the continental US produces a relatively accurate map of irrigated agriculture under dryland irrigated conditions (excluding irrigated pastures and paddy rice) *circa* 2001. As an objective and repeatable methodology that is robust enough to handle complex forms of irrigation that occur around the earth, the successful application of our procedure in the US warrants pursuing a global application. The result of this application is a first step in a larger research agenda to understand anthropogenic effects of cultivation, and specifically that of irrigation, on worldwide water and energy cycles, climate, agricultural productivity, and sustainability where accurate, objective, up-to-date maps of irrigation at a global scale are essential.

Acknowledgements

This research was funded by the US National Research Council and NASA Postdoctoral Program Fellowships awarded to Mutlu Ozdogan. Dr. Ozdogan acknowledges generous computation and logistical support from the Hydrological Sciences Branch at NASA-GSFC. Suggestions by Drs. Anatoly Gitelson and Mark Friedl greatly improved this research. Discussion with Stefan Siebert of Frankfurt University and Prasad Thenkabail of IWMI greatly enhanced the scope of this research. Finally, three anonymous reviewers are thanked for greatly improving the quality of this contribution.

References

Alcamo, J., Doll, P., Henrichs, T., Kaspar, F., Lehner, B., Rosch, T., et al. (2003). Global estimates of water withdrawals and availability under current and future "business-as-usual" conditions. *Hydrologic Sciences Journal*, *48*(3), 339–348.

Andrienko, G. L., & Andrienko, N. V. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, *13*(4), 355–374.

Bonan, G. B. (2001). Observational evidence for reduction of daily maximum temperature by croplands in the Midwest United States. *Journal of Climate*., *14*, 2430–2442.

Boucher, O., Myhre, G., & Myhre, A. (2004). Direct human influence of irrigation on atmospheric water vapor and climate. *Climate Dynamics*, *22*, 597–603.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees, Wadsworth*.

Budyko, M. I. (1974). *Climate and life, international geophysics series.* Orlando, FL: Academic Press. 508 pp.

California DWR, http://www.landwateruse.water.ca.gov/basicdata/landuse/landusesurvey.cfm

DeFries, R. S., Hansen, M., Townshend, J. R. G., & Sohlberg, R. (1998). Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, *19*, 3141–3168.

Droogers, P. (2002). Global irrigated area mapping: Overview and recommendations. *Working Paper*, *vol. 36*. Colombo, Sri Lanka: International Water Management Institute.

FAO (Food and Agriculture Organization of the United Nations) (1998). *FAOSTAT-PC, FAO Statistics on diskette.* Rome: Food and Agriculture Organization of the United Nations.

Friedl, M., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*, 399–409.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*, 287–302.

Friedman, J. H., Hastie, T., & Tibshirani, R., (1998). Additive logistic regression: A statistical view of boosting, Dept. of Statistics, Stanford University Technical Report, 45 pp.

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, *161*, 165–173.

Gitelson, A. A., Gritz, U., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*, 271–282.

Gitelson, A. A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll *a* using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology (B)*, *22*, 247–252.

Gitelson, A. A., Vina, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, *32*, L08403. doi:10.1029/2005GL022688

Gitelson, A. A., Vina, A., Verma, S. B., Rundquist, D. C., Arkebauer, T. J., Keydan, G., et al. (2006). Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research*, *111*, D08S11. doi:10.1029/2005JD006017

Gollehon, N., & Quinby, W. (2000). Irrigation in the American West: Area, water and economic activity. *Water Resources Development*, *16*, 187–195.

Gordon, L. J., Steffen, W., Jonsson, B. F., Folke, C., Falkenmark, M., & Johansen, A. (2005). Human modification of global water vapor flows from the land surface. *Proceedings of the National Academy of Sciences*, *102*(21), 7612–7617.

Goward, S. N., Markham, B., Dye, D., Dulaney, W., & Yang, J. (1991). Normalized difference vegetation index measurements from the advances very high resolution radiometer. *Remote Sensing of Environment*, *35*, 257–277.

Gutman, G., & Ignatov, A. (1998). Derivation of green vegetation fraction from NOAA/AVHRR for use in weather prediction models. *International Journal of Remote Sensing*, *19*, 1533–1543.

Gutman, G., Ohring, G., & Joseph, J. H. (1984). Interaction between the geobotanic state and climate: A suggested approach and a test with a zonal model. *Journal of the Atmospheric Sciences*, *41*, 2663–2678.

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, *54*, 127–140.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., & Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, *83*, 303–319.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Huete, A.R., Justice, C.O., & van Leeuwen, W.J.D., (1999). MODIS Vegetation Index, Algorithm Theoretical Basis Document, http://modarch.gsfc.nasa.gov/MODIS/ATBD/atbd_mod13.pdf

Justice, C., Townshend, J. R. G., Holben, B., & Tucker, C. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, *6*, 1271–1318.

Kamthonkiat, D., Honda, K., Turral, H., Tripathi, N. K., & Wuwongse, V. (2005). Discrimination of irrigated and rainfed rice in a tropical agricultural system using SPOT VEGETATION NDVI and rainfall data. *International Journal of Remote Sensing*, *26*, 2527–2547.

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

Lettau, H. (1969). Evapotranspiration climatology: A new approach to numerical prediction on monthly evapotranspiration, runoff and soil moisture storage. *Monthly Weather Review*, *97*, 691–699.

Lobell, D. B., Bala, G., & Duffy, P. D. (2006). Biogeophysical impacts of cropland management changes on climate. *Geophysical Research Letters*, *33*, L06708. doi:10.1029/2005GL025492

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Yang, L., et al. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. *International Journal of Remote Sensing*, *21*, 1,303–1,330.

Martinez-Beltran, C., & Calera-Belmonte, A. (2001). Irrigated crop estimation using Landsat TM imagery in La Mancha, Spain, Photogramm. *Eng. Rem. Sen.*, *67*, 1177–1184.

Mciver, D. K., & Friedl, M. A. (2001). Estimating pixel-scale land cover classification confidence using non-parametric machine learning methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 1959–1968.

Merzlyak, M. N., & Gitelson, A. A. (1995). Why and what for the leaves are yellow in autumn? On the interpretation of optical spectra of senescing leaves (*Acer platanoides* L.). *Journal of Plant Physiology*, *145*, 315–320.

Moore, N., & Rojstaczer, S. (2002). Irrigation's influence on precipitation: Texas High Plains, U.S.A. *Geophysical Research Letters*, *29*(16), 1755. doi:10.1029/2002GL014940

Myneni, R. B., Hall, F. G., Sellers, P. S., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 481–486.

Nicholson, S. E., Davenport, M. L., & Malo, A. R. (1990). A comparison of vegetation response to rainfall in the Sahel and East Africa using normalized difference vegetation index from NOAA-AVHRR. *Climatic Change*, *17*, 209–241.

NASS, 2002. United States Department of Agriculture, National Agricultural Statistics Service, 2002 Census of Agriculture – Volume 1 – Geographic Area Series – Farm&Ranch Irrigation Survey. http://www.nass.usda.gov/Census/Create_Census_FRIS.jsp

NOAA–NCDC, 2001. US National Oceanic and Atmospheric Administration (NOAA) and US National Climatic Data Center (NCDC). http://www.ncdc.noaa.gov/oa/climate/research/2001/aug/us_national.html

Otterman, J., Manes, A., Rubin, S., Alpert, P., & Starr, D. O.'C. (1990). An increase of early rains in southern Israel following land use change? *Bound.-Layer Meteor*, *53*, 333–351.

Ozdogan, M., & Gutman, G. (2007). Towards global mapping of irrigated agriculture. *Geophysical Research Abstracts*, *9*, 00329 2007 SRef-ID: 1607-7962/gra/EGU2007-A-00329.

Ozdogan, M., Salvucci, G. D., & Anderson, B. C. (2006). Examination of the Bouchet–Morton complementary relationship using a mesoscale climate model and observations under a progressive irrigation scenario. *Journal of Hydrometeorology*, *7*(2), 235–251.

Ozdogan, M., & Woodcock, C. E. (2006). Resolution dependent errors in remote sensing of cultivated areas. *Remote Sensing of Environment*, *103*, 203–217.

Pax Lenney, M., Woodcock, C. E., Collins, J. C., & Hamdi, H. (1996). The status of agricultural lands in Egypt: The use of multitemporal NDVI features derived from Landsat TM. *Remote Sensing of Environment*, *56*(1), 8–20.

Prince, S. D. (1991). Satellite remote sensing of primary production: comparison of results for Sahelian grasslands. *International Journal of Remote Sensing*, *12*, 1133–1421.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* Los Altos: Morgan Kaufmann.

Ramankutty, N., & Foley, J. A. (1998). Characterizing patterns of global land use: An analysis of global croplands data. *Global Biogeochemical Cycles*, *12*(4), 667–685.

Rosenzweig, C., Strzepek, K. M., Major, D. C., Iglesias, A., Yates, D. N., McCluskey, A., & Hillel, D. (2004). Water resources for agriculture in a changing climate: International case studies. *Global Environmental Change A*, *14*, 345–360. doi:10.1016/j.gloenvcha.2004.09.003

Rulequest, 2001. Rulequest Research data mining tools. htpp://www.rulequest.com

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, Albedo and Nadir reflectance products from MODIS. *Remote Sensing of Environment*, *83*, 135–148.

Siebert, S., Döll, P., Feick, S., Frenken, K., & Hoogeveen, J. (2007). *Global map of irrigated areas version 4.0.1.* Rome, Italy: University of Frankfurt (Main), Germany / Food and Agriculture Organization of the United Nations.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J. -M., Frenken, K., & Feick, S. (2005). Development and validation of the global map of irrigation areas. *Hydrology and Earth System Sciences*, *9*, 535–547.

Solley, W.B., Pierce, R.R., and Perlman, H.A., 1998. Estimated use of water in the United States in 1995. Circular 1200, U.S. Geological Survey, U.S. Department of Interior.

Thenkabail, P. S., Schull, M., & Turral, H. (2005). Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sensing of Environment*, *95*, 317–341.

Townshend, J. R. G., & Justice, C. (2002). Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. *Remote Sensing of Environment*, *83*, 351–359.

Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*, 127–150.

Tucker, C. J., & Gatlin, T. A. (1984). Monitoring vegetation in the Nile Delta with NOAA 6 and NOAA 7 AVHRR imagery. *Photogrammetric Engineering and Remote Sensing*, *50*(1), 53 61.

Vörösmarty, C. J. (2002). Global water assessment and potential contributions from Earth System Science. *Aquatic Science*, *64*, 328–351.

Vörösmarty, C. J., Douglas, E. M., Green, P. A., & Revenga, C. (2005). Geospatial indicators of emerging water stress: An application to Africa. *Ambio*, *34*, 230–236.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *Science*, *289*, 284–288.

Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment*, *112*, 1096–1116.

Wood, S., Sebastian, K., & Scherr, S. J. (2000). *Pilot analysis of global ecosystems: agroecosystems.* Washington D.C.: World Resources Institute/International Food Policy Research Institute.

Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., & Magnussen, S. (2006). An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing*, *27*(4), 663–683.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J. Y., Salas, W., et al. (2006). Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sensing of Environment*, *100*, 95–113.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., Moore, B., et al. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, *95*, 480–492.

Xiao, X., He, L., Salas, W., Li, C., Moore, B., Zhao, R., et al. (2002). Quantitative relationships between field-measured leaf area index and vegetation index derived from VEGETATION images for paddy rice fields. *International Journal of Remote Sensing*, *23*, 3595–3604.

Zarco-Tejada, P. J., Miller, J. R., Mohammed, G. H., Noland, T. L., & Sampson, P. H. (2002). Vegetation stress detection through chlorophyll a+b estimation and fluorescence effects on hyperspectral imagery. *Journal of Environmental Quality*, *31*, 1433–1441.

Yang, L., Huang, C., Homer, C. G., Wylie, B. K., & Coan, M. J. (2003). An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, *29*(2), 230–240.