Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 112 (2008) 1835–1845
Remote Sensing of Environment
www.elsevier.com/locate/rse

The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI

L.M. Montandon *, E.E. Small

Department of Geological Sciences, University of Colorado, Boulder, CO 80309, USA

Received 6 February 2007; received in revised form 7 September 2007; accepted 9 September 2007

Abstract

The green vegetation fraction (Fg) is an important climate and hydrologic model parameter. A common method to calculate Fg is to create a simple linear mixing model between two NDVI endmembers: bare soil NDVI ($NDVI_o$) and full vegetation NDVI ($NDVI_\infty$). Usually it is assumed that $NDVI_o$ is close to zero ($NDVI_o \sim 0.05$) and is generally chosen from the lowest observed NDVI values. However, the mean soil NDVI computed from 2906 samples is much larger (NDVI = 0.2) and is highly variable (standard deviation = 0.1). We show that the underestimation of $NDVI_o$ yields overestimations of Fg. The largest errors occur in grassland and shrubland areas. Using parameters for $NDVI_o$ and $NDVI_\infty$ derived from global scenes yields overestimations of Fg that are larger than 0.2 for the majority of U.S. land cover types when pixel NDVI values are $0.2 < NDVI_{pixel} < 0.4$. When using conterminous U.S. scenes to derive $NDVI_o$ and $NDVI_\infty$, the overestimation is less (0.10–0.17 for $0.2 < NDVI_{pixel} < 0.4$). As a result, parts of the conterminous U.S. are affected at different times of the year depending on the local seasonal NDVI cycle. We propose using global databases of $NDVI_o$ along with information on historical $NDVI_{pixel}$ values to compute a statistically most-likely estimate of Fg. Using *in situ* measurements made at the Sevilleta LTER, we show that this approach yields better estimates of Fg than using global invariant $NDVI_o$ values estimated from whole scenes. At the two studied sites, the Fg estimate was adjusted by 52% at the grassland and 86% at the shrubland. More significant advances will require information on spatial distribution of soil reflectance.

Keywords: Green fraction; Soil; Reflectance; NDVI; MODIS; AVHRR; Vegetation; Land cover

1. Introduction

Vegetation is important in climate studies due to its role in the hydrologic cycle. Remote measurements of land surface reflectance can be used to calculate the parameters, such as the Leaf Area Index (LAI) or the green vegetation fraction (Fg), needed to represent vegetation in climate and hydrologic models. These two parameters represent the vertical and the horizontal density of live vegetation, respectively (Gutman & Ignatov, 1997). This paper focuses on the impact of soil reflectance on Fg calculations. Fg is needed for many hydroclimatic applications due to its important contribution in climate models and surface fluxes (Matsui et al., 2005; Liang et al., 2004; Molders & Olson, 2004). The importance of accurately determining Fg can be illustrated by looking at its impact on the estimation of latent heat flux (LE). LE is a function of meteorological conditions and numerous surface properties, such as LAI, Fg, leaf stomatal resistance, soil texture, and water volumetric content at different depths. However, it is strongly dependent on Fg values, as illustrated by three sensitivity studies (Fig. 1). Keeping all other factors equal, a change of as much as 100 Wm^{-2} in LE can occur when varying Fg by only 0.2 (Gutmann & Small, 2007; Betts et al., 1997; Jacquemin & Noilhan, 1990). As a result, accurate estimation of Fg is crucial to better parameterize climate models.

Both Fg and LAI are normally inferred from Normalized Difference Vegetation Index (NDVI), an index calculated from reflectance measurements in the red and near-infrared wavelengths. These measurements are typically acquired by satellites over large areas (scenes) divided into sub-units (pixels) that represents the average reflectance over a smaller area. The methodologies used to infer LAI and Fg differ. Models using NDVI to derive LAI are based on empirical relationships (Qi et al., 2000; Price & Bausch, 1995, Price 1993). Models used to derive

* Corresponding author.
E-mail address: montando@colorado.edu (L.M. Montandon).

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2007.09.007

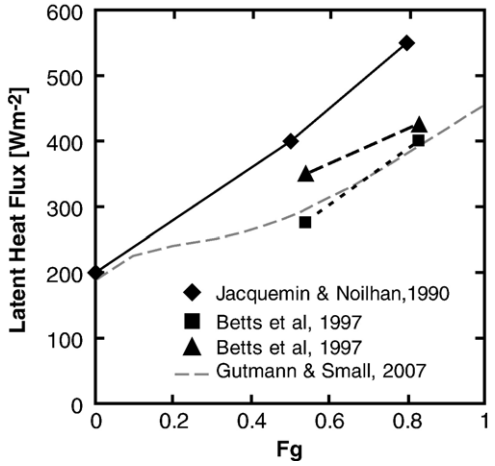

Fig. 1. Influence of vegetation cover on latent heat (all other factors kept equal) as computed from various land surface models.

Fg are generally simple linear or quadratic combinations of two endmembers: NDVI from dense (LAI > 3) live vegetation and soil. The difference in how LAI and Fg are modeled precludes studying the impact of soil reflectance on LAI and Fg using the same approach. Therefore, we focus only on Fg in this paper.

Common models to compute Fg include the linear Gutman and Ignatov (GI) model (Gutman & Ignatov, 1998) and the quadratic model (Carlson & Ripley, 1997). Both models are popular due to their ease of implementation. This ease stems partly from pre-selected values of NDVI for the soil and plant endmembers. However, the selection of these parameters is complicated by variations in spectral signals of vegetation due to differences in species, plant health, leaf water content, and other factors (Jensen, 2000). Likewise, the spectral signature of soil varies, depending upon mineralogy, moisture, grain size, etc (Baumgardner et al., 1985). In this paper, we primarily present our results within the framework of the linear GI model, although we include a discussion of the quadratic model. The linear model is still the one commonly used in products such as the community NOAH Land-Surface Model (Mitchell, 2001) and the NAM Eta model (Gallo et al., 2003).

Addressing the spatial variability of plant and soil reflectance requires extensive *in situ* measurements, which is unrealistic for most applications. As a result, both the linear and quadratic models are normally parameterized using single estimated soil NDVI ($NDVI_o$) and live vegetation NDVI ($NDVI_\infty$) values. The common technique to estimate the two endmembers is to infer them from the data themselves. $NDVI_\infty$ can be selected as the highest NDVI value within the scene (Li et al., 2003; Gallo et al., 2001; Gutman & Ignatov, 1998). Alternatively, the scene can be split into biomes and the maximum value can be selected from each (Matsui et al., 2005; Zeng et al., 2000; Oleson et al., 2000). $NDVI_o$ is commonly inferred from the historical lowest NDVI values within the scene (e.g. GI approach). However, many authors opt for the use of popular published $NDVI_o$ values of 0.05 or less (Table 1). The minimum NDVI approach is based on two important assumptions: (1) the pixels with the lowest NDVI are free of vegetation (bare soil assumption) and (2) the soil NDVI is the same everywhere in the scene (invariant assumption). The first assumption is likely true, particularly when studying large areas. However we show below that the second assumption is not often valid.

Here, we investigate how the observed values and variability of $NDVI_o$ can impact Fg calculations. First we examine several datasets of soil NDVI to constrain the variability of soil reflectance using both the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) bands spectral response functions. We evaluate the error on Fg introduced by underestimation of $NDVI_o$ for several land cover types by comparing values computed using a popular method to infer a single $NDVI_o$ from AVHRR and MODIS data to values from the soil dataset. Then we look at how this error varies throughout a typical year for the different land cover types in the conterminous U.S. using MODIS 16-day NDVI imagery. We illustrate the spatial distribution of the error throughout the U.S. with two examples: the error for April and June 2003 as derived from MODIS 16-day NDVI. These two images were selected to represent both a time of the year with overall low (April) and high (June) error. Finally we suggest a method to compute an adjusted Fg estimates and illustrate its impact on Fg time series at a grassland and shrubland site.

We limit this study to the problems resulting from the bare soil and invariant assumptions when using NDVI in linear models such as the linear GI model and how it impacts Fg estimations. Previous work addressing the issue of the interaction of soil reflectance with overlying vegetation yielded alternatives to NDVI such as the Soil Adjusted Vegetation Index (SAVI, Huete, 1988), Generalized SAVI (GESAVI, Gilabert et al., 2002), and Transformed SAVI (TSAVI, (Baret & Guyot, 1991). It is important to note that these indices offer low sensitivity to soil background noise and are possible alternatives to the use of NDVI.

2. Methods

NDVI is computed from reflectance measurements in the red (R) and near-infrared (NIR) wavelengths:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - R}{\mathrm{NIR} + R} \tag{1}$$

The GI model is used to relate Fg to NDVI values with a linear mixing model:

$$\mathrm{Fg} = \frac{\mathrm{NDVI}_{\mathrm{pixel}} - \mathrm{NDVI}_o}{\mathrm{NDVI}_\infty - \mathrm{NDVI}_o} \tag{2}$$

where $NDVI_{pixel}$ is the pixel NDVI value, $NDVI_o$ is the bare soil NDVI, and $NDVI_\infty$ is the live vegetation NDVI. This linear mixing model is based on the assumption that vegetation is dense where it exists, with a LAI ≥ 3. Alternatively, other authors (Carlson & Ripley, 1997; Choudhury et al., 1994) have suggested a quadratic relationship between Fg and NDVI:

$$\mathrm{Fg} = \left(\frac{\mathrm{NDVI}_{\mathrm{pixel}} - \mathrm{NDVI}_o}{\mathrm{NDVI}_\infty - \mathrm{NDVI}_o}\right)^2 \tag{3}$$

Table 1
Examples of $NDVI_o$ and $NDVI_\infty$ used in recent publications

Paper and study area	$NDVI_o$	$NDVI_\infty$	Dataset used in the study	Parameter determination
Gutman and Ignatov (1998) Global	0.04	0.52	Global AVHRR NDVI	Desert annual minimum ($NDVI_o$) and evergreen annual maximum ($NDVI_\infty$) NDVI
Yang and Yang (2006) China	0.05	0.49 grassland 0.60 open shrubland 0.68 mixed forest 0.70 broadleaf and deciduous forests	AVHRR 10-day NDVI composite over China	Zeng et al. (2000) values
Gan and Burges (2006) Eastern US	0.04	0.52	Global AVHRR NDVI	GI values
Gebremichael and Barros (2006) Mexico and Nepal	0.05	0.52 open shrubland 0.95 mixed forests	MODIS 16-day NDVI over study area	Zeng et al. (2000) values for open shrubland
Matsui et al. (2005) North America	0.03 arid 0.04 seasonal 0.05 evergreen	0.52 arid 0.74 seasonal 0.67 evergreen	Global monthly AVHRR NDVI	Historical minimum and maximum over 20 years for different biomes. $NDVI_o$=3rd percentile $NDVI_\infty$=97th percentile
Li et al. (2003) North China	0.04	0.61	AVHRR reflectance over North China	Steppe minimum ($NDVI_o$) and maximum ($NDVI_\infty$) NDVI values in 2001
Ek et al. (2003) North America	0.04	0.52	–	GI values
Sridhar et al. (2003) Conterminous U.S.	0.04	0.52	Conterminous U.S. AVHRR 14-day NDVI composite	GI values
Gallo et al. (2003) Conterminous U.S.	0.09	0.69	Conterminous U.S. AVHRR 14-day NDVI composite	5-year average NDVI in January for the North-central U.S. $NDVI_o$=50th percentile of grassland and cropland And yearly maximum NDVI for different biomes $NDVI_\infty$=98th percentile of areas with dense vegetation only
Oleson et al. (2000) Upper Mississippi Basin	0.048	0.752 crop, grass, desert, shrub 0.816 mixed woodland, forest 0.824 broadleaf deciduous	Conterminous U.S. AVHRR 10-day NDVI composite	Monthly maximum NDVI over 5 years for different biomes. $NDVI_o$=2nd percentile of desert and semi-desert $NDVI_\infty$=98th percentile of groups of biomes
Zeng et al. (2000) Conterminous U.S.	0.05	0.49 grassland 0.60 open shrubland 0.68 mixed forest 0.70 broadleaf and deciduous forests	Global AVHRR 10-day NDVI composite	Maximum NDVI over 1 year time series for different biomes: $NDVI_o$=5th percentile of barren and sparsely vegetated land areas $NDVI_\infty$=90th percentile (shrubland, barren, sparsely vegetated), 75th percentile (other)

In order to constrain $NDVI_o$ and its variability, we compiled 2906 reflectance spectra of soils collected worldwide from the following datasets: (1) LARS at Purdue University available at www.lars.purdue.edu, (2) ICRAF Soil Laboratory at www.worldagroforestry.org, and (3) ASTER spectral library available at speclib.jpl.nasa.gov. A majority of these spectra were measured in the lab and might differ from field spectra values. However a study by Stoner et al. (1980) suggests that lab wet soil reflectances in the visible and NIR are similar to spectral field values. For each soil, we computed two NDVI values using the following band paths convolved by their corresponding spectral response function: 1) AVHRR NOAA-16 band 1 (580–680 nm) and 2 (725–1100 nm) and 2) MODIS Terra band 1 (620–670 nm) and 2 (841–876 nm). While the LARS soil spectra were measured on wet soils, the distribution of NDVI values of these soils is normal and similar to that of dry soils (Fig. 2). We used this 2906 soils composite dataset to compute statistically most-likely Fg estimate that, unlike current methods, does not use a single soil NDVI value but takes into account the observed soil NDVI variability to produce an adjusted Fg estimate.

We calculated Fg by using pairs of $NDVI_o$ and $NDVI_\infty$ for each land cover type defined by the International Geosphere–Biosphere Program (IGBP). $NDVI_o$ was defined using two approaches: 1) using an invariant $NDVI_o$ computed from a time series of AVHRR and MODIS imagery for the conterminous U.S. and 2) using a set of $NDVI_o$ from the soil dataset to compute a number of possible Fg values whose mean is the most-likely estimate. $NDVI_\infty$ values were computed using the imagery time series of the first approach.

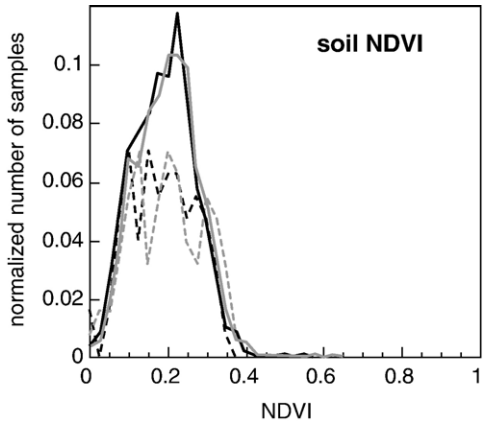

Fig. 2. Histograms of NDVI values of the two soil databases (dashed line—80 wet soils, solid line—2826 dry soils) computed using corresponding AVHRR (grey) and MODIS (black) spectral response functions.

In our first approach, we used the method described in Zeng et al. (2000) to compute the pairs of $NDVI_o$ and $NDVI_\infty$ for each land cover types. For each pixel over the conterminous U.S., we selected the highest NDVI of two 2003 time series: AVHRR 14-day NDVI and MODIS 16-day NDVI. Additional information on the imagery used is summarized in Table 2. We then produced histograms of maximum NDVI values for each IGBP land cover type. The 5th percentile of barren and sparsely vegetated land was used as $NDVI_o$. $NDVI_\infty$ was selected as the 90th percentile for barren, sparsely vegetated and shrubland and the 75th percentile for all other land cover types. Table 3 summarizes the different pairs of parameters obtained. We included the values obtained by Zeng et al. (2000) from a time series of global (i.e. covering most of the Earth) AVHRR 10-day NDVI (Table 3A, and Table 2). Note how the values derived from global scenes are generally lower than the ones derived from conterminous U.S. imagery and how MODIS-derived values are higher than the one derived from coarser resolution AVHRR. We then used these parameters along with Eqs. (2) and (3) to compute Fg. Here $NDVI_o$ is invariant at 0.05 (global AVHRR), 0.10 (U.S. AVHRR), and 0.07 (U.S. MODIS).

Our second approach was used to derive an adjusted estimate, Fg*, that was computed using the same $NDVI_\infty$ as in Table 3. However we varied $NDVI_o$ for each pixel using the n soil NDVI values from our dataset that respect the linear mixing model condition: $NDVI_o \leq NDVI_{pixel}$ (Fig. 3A). This yielded n Fg

Table 2
Description of the remote sensing datasets used for this study

	AVHRR 14-day NDVI	Global AVHRR 10-day NDVI	MODIS 16-day NDVI
Product	AVHRR NDVI biweekly composites for the conterminous U.S.	AVHRR 10-day global NDVI composite	MODIS NDVI composite for the conterminous U.S.
Source	USGS EarthExplorer http://earthexplorer.usgs.gov/	USGS Earth Resources Observation and Science http://edc.usgs.gov/	University of Maryland, Global Land Cover Facility http://glcf.umiacs.umd.edu/
Spatial resolution	1 km	1 km	250 m
Input data	NOAA-16 calibrated raw data	NOAA-12 calibrated raw data	MODIS daily Surface Reflectance product (MOD09)
Composite	14-day	10-day	16-day
Atmospheric corrections	The NOAA-16 radiance data were corrected for ozone and Rayleigh scattering only	Radiance data were corrected for ozone and Rayleigh scattering only	The MOD09 data include correction for atmospheric gazes (ozone, water vapor, etc.), aerosols (e.g. Rayleigh scattering), and thin cirrus clouds

Table 3
$NDVI_o$, $NDVI_\infty$ values derived using Zeng et al. (2000) method on three time series: A) global AVHRR 10-day NDVI imagery between April 1992 and March 1993 (Zeng et al., 2000 values), B) conterminous U.S. AVHRR 14-day NDVI imagery for 2003, and C) conterminous U.S. MODIS 16-day NDVI imagery for 2003

	IGBP land cover type	*A* global AVHRR	*B* cont. U.S. AVHRR	*C* cont. U.S. MODIS	Mean Fg	% cont. U.S. cover
$NDVI_\infty$	Bare and shrubland	0.6	0.83	0.87	0.60	13.9
	Grassland	0.49	0.61	0.67	0.49	11.1
	Cropland and national vegetation mosaic	0.65	0.79	0.86	0.65	10.1
	Cropland	0.61	0.81	0.86	0.61	9.3
	Evergreen needleleaf forest	0.63	0.81	0.89	0.63	8.3
	Deciduous broadleaf forest	0.7	0.85	0.89	0.70	7.3
	Woody savanna	0.62	0.69	0.76	0.62	4.6
	Mixed forest	0.68	0.85	0.89	0.68	3.4
	Savanna	0.58	0.70	0.80	0.58	0.3
	Evergreen broadleaf forest	0.69	0.79	0.85	0.69	0.1
	Permanent wetland	0.56	0.57	0.84	0.56	0.1
$NDVI_o$	All land cover types	0.05	0.10	0.07	0	

Also shown are the mean fractional vegetation cover as defined by Zeng et al. (2000), and the area of the U.S. covered by each IGBP land cover class.

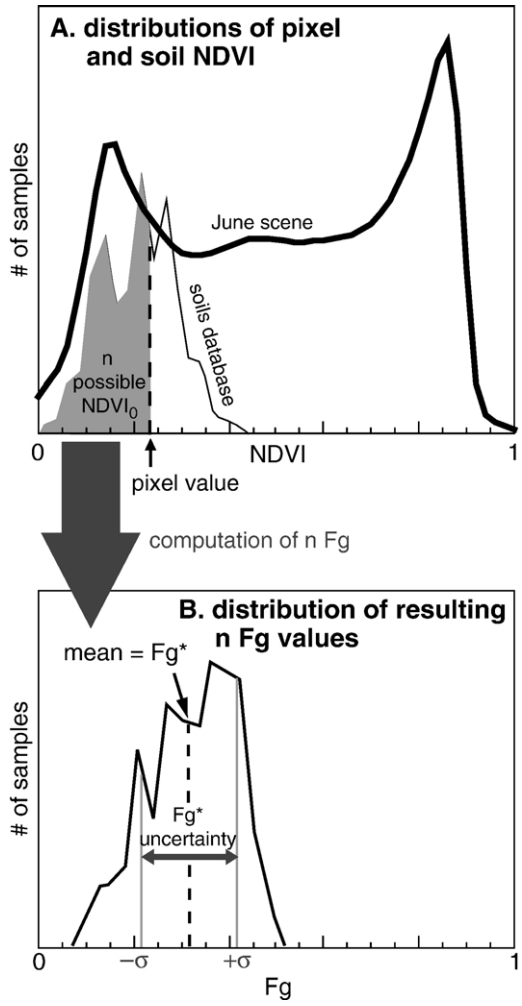

Fig. 3. Sketch showing how Fg* and its uncertainty σ are computed using all soils from the database that respect the linear mixing model condition: $NDVI_o \leq NDVI_{pixel}$. A) For each pixel, the n soils whose NDVI are equal or smaller than the $NDVI_{pixel}$ are identified and then used to computed n possible Fg values using the linear or quadratic model (Eqs. (2) and (3)). B) The distribution of the resulting n Fg values is then used to compute the most-likely estimate Fg* (Eq. (4)) and its uncertainty (Eq. (6)).

values, which approximate a normal distribution (Fig. 3B). The mean of the n values is statistically the most-likely Fg value and therefore an adjusted estimate Fg*:

$$Fg^* = \frac{\sum_{i=1}^{n} \left(\frac{NDVI_{pixel} - NDVI_{o,i}}{NDVI_{\infty} - NDVI_{o,i}} \right)^a}{n} \quad (4)$$

where a is equal to 1 for the linear model, and 2 for the quadratic model.

We then quantified the error on Fg estimation due to the incorrect assignment of $NDVI_o$, hereafter referred to as ΔFg^*. It is the difference between the Fg values computed from the two approaches for each pixel:

$$\Delta Fg^* = Fg^* - Fg \quad (5)$$

For pixels with $NDVI_{pixel} > NDVI_{\infty}$, the Fg models saturate and the Fg values were set to 1 indicating full vegetation cover exists. Hence, $\Delta Fg^* = 0$ for these pixels.

The uncertainty on Fg* values due to the variability of soil NDVI was then estimated by computing the standard deviation (σ) of the resulting n Fg values for each pixel (Fig. 3B):

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} Fg_i - Fg^*} \quad (6)$$

The distribution of the n Fg values approximates a normal distribution, regardless of the pixel value (see example in Fig. 3B). Therefore, the adjusted Fg estimate has about a 68% chance of being within the range defined by $Fg^* \pm \sigma$, Fg* remaining the most-likely estimate.

In order to understand how the error (ΔFg^*) varies throughout a typical year, we computed for each land cover type the mean NDVI value for each 2003 scene of the MODIS 16-day NDVI time series. Using these NDVI values, we calculated the corresponding temporal Fg* and ΔFg^*. Two NDVI 16-day MODIS images over the conterminous U.S. were also selected to represent the spatial distribution of the error. The two scenes were chosen by studying the overall NDVI values over a four-year time series (2001–2004). The MODIS June 10–26 2003 and April 7–23 2003 scenes were the two images with the highest and lowest number of pixels with large errors respectively (defined by the number of pixels with $\Delta Fg^* > 0.15$).

Finally, in order to test our adjusted estimate Fg*, we compared the standard Fg and Fg* values with the one obtained using *in situ* $NDVI_o$ measurements as ground truth. These measurements were made in August 2006 at two locations within the Sevilleta LTER in New Mexico: a black grama grassland and a creosote shrubland site. These two sites were selected as grassland and shrubland are the two most common land cover types in the conterminous U.S. and represent areas where the potential error is maximized. The soil at the grassland site is a Turney sandy clay loam and the soil at the shrubland site is a Bluepoint fine sand (United States Department of Agriculture, 1988). Their respective NDVI are 0.12 and 0.09 and were computed from field spectra convolved by the MODIS spectral response functions.

3. Results

3.1. Observed soil NDVI

Fig. 2 shows the distribution of the NDVI of the 2906 soil samples computed using AVHRR and MODIS spectral response functions. The mean soil NDVI values are the same between the wet and dry soil datasets, and are 0.20 for AVHRR band paths and 0.21 for MODIS band paths with a standard deviation of 0.08 (wet) and 0.09 (dry). This result is consistent with the computations made by Price and Bausch (1995) on a smaller dataset.

The data summarized in Fig. 2 have two implications for the calculation of Fg using GI or similar approaches. First, the mean

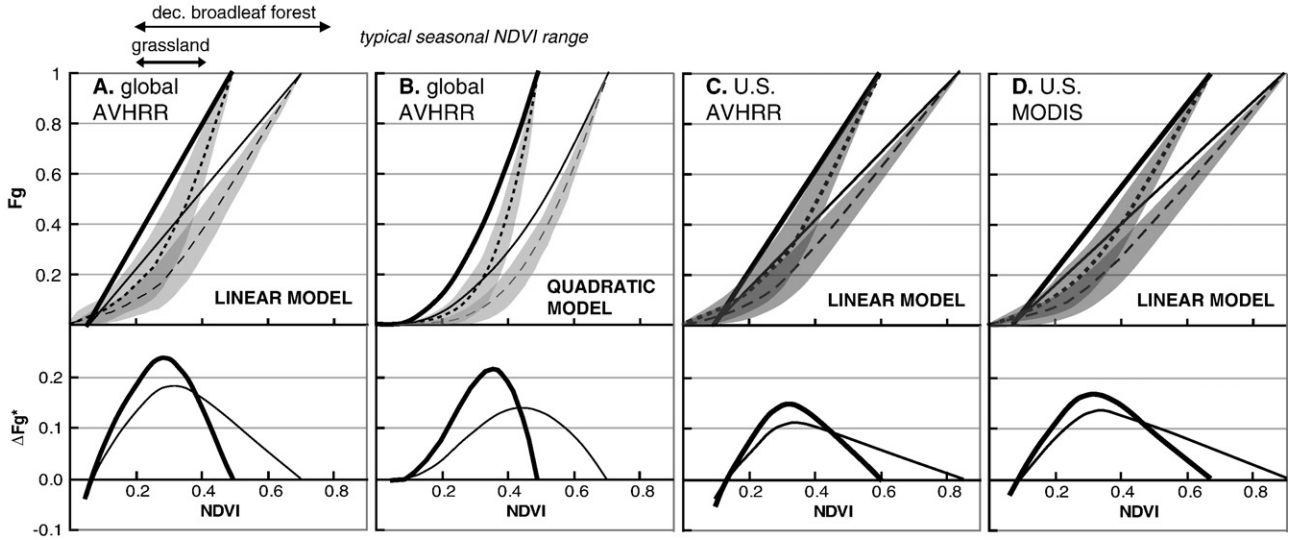

Fig. 4. Upper quadrants: Fg (solid lines) and corresponding Fg* (dashed lines) computed for grassland (thick lines) and deciduous broadleaf forest (thin lines) using three different pairs of $NDVI_o$ and $NDVI_\infty$: A) and B) Zeng et al. (2000) values derived from global AVHRR 10-day NDVI imagery, C) values derived from conterminous U.S. AVHRR 14-day NDVI imagery, and D) values derived from conterminous U.S. MODIS 16-day NDVI imagery. The grey-shaded areas show the uncertainty on Fg* from Eq. (6). Lower quadrants: corresponding Fg overestimation, ΔFg* (Eq. 5).

soil NDVI value is about five times larger than the $NDVI_o$ derived from global scenes, i.e. 0.21 versus 0.03–0.05 and two to five times larger than the values derived from local scenes, i.e. 0.21 versus 0.04–0.10 (Table 1 and 3). These commonly used $NDVI_o$ values are also equal to the lowest values of the entire distribution of soil NDVI. This correspondence shows consistency between the soil NDVI measured *in situ* and via remote sensing. Second, given the variability of soil NDVI, using a single $NDVI_o$ value for a whole scene is not appropriate, particularly for continental scale studies.

3.2. Influence of $NDVI_o$ on Fg calculations

First, we show how Fg values differ when a $NDVI_o$ that is too low is used instead of the statistically most-likely $NDVI_o$, as estimated from the soil database. The solid lines in Fig. 4 illustrate the linear Fg models for grassland and deciduous broadleaf forest, determined using three pairs of parameters: Zeng et al. (2000) values derived from global AVHRR (Fig. 4A and B), values derived from conterminous U.S. AVHRR NDVI imagery (Fig. 4C), and values derived from conterminous U.S. MODIS NDVI imagery (Fig. 4D). These two land cover types represent the two extremes for $NDVI_\infty$, which yields the maximum and minimum Fg errors respectively. All other land cover types exist between these two endmembers (see $NDVI_\infty$ values in Table 3). The dotted lines in Fig. 4 show the adjusted estimate, Fg*, computed using the distribution of soil NDVI from the soil database (Eq. (4)). Fig. 4 shows that the underestimation of $NDVI_o$ yields an overestimation of Fg for most $NDVI_{pixel}$ values measured. This is illustrated by the positive ΔFg* values (Eq. (5)) in the lower quadrants of Fig. 4.

For grassland, the land cover with the largest error, ΔFg* exceeds 0.15 for pixels with $0.17 \le NDVI_{pixel} \le 0.39$ and reaches a maximum of 0.24 when using the parameters derived from global scenes along with the linear model (Fig. 4A). This error is reduced when using parameters derived from corresponding local scenes (here the conterminous U.S.) and reaches 0.15 for AVHRR and 0.17 for MODIS. For deciduous broadleaf forest, the land cover type associated with the smallest error, ΔFg* exceeds 0.15 for pixels with $0.22 \le NDVI_{pixel} \le 0.43$ and reaches a maximum of 0.18 when using globally derived parameters. This sensitive

Table 4
Summary of the amount of overestimation on Fg due to underestimation of $NDVI_o$, and the uncertainty on the adjusted Fg* estimate

A. Global AVHRR		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.24	0.18
	$\Delta Fg^* \ge 0.15$	$0.17 \le NDVI_{pixel} \le 0.39$	$0.22 \le NDVI_{pixel} \le 0.43$
*Uncertainty on Fg**	Max σ	0.15	0.11
	$\sigma = max$	$0.30 \le NDVI_{pixel} \le 0.35$	$0.32 \le NDVI_{pixel} \le 0.37$
	$\sigma \ge 0.10$	$0.20 \le NDVI_{pixel} \le 0.42$	$0.28 \le NDVI_{pixel} \le 0.4$
B. AVHRR U.S.		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.15	0.11
	$\Delta Fg^* \ge 0.15$	$0.32 \le NDVI_{pixel} \le 0.33$	None
*Uncertainty on Fg**	Max σ	0.12	0.09
	$\sigma = max$	$0.31 \le NDVI_{pixel} \le 0.37$	$0.33 \le NDVI_{pixel} \le 0.40$
	$\sigma \ge 0.10$	$0.24 \le NDVI_{pixel} \le 0.43$	None
C. MODIS U.S.		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.17	0.13
	$\Delta Fg^* \ge 0.15$	$0.25 \le NDVI_{pixel} \le 0.39$	None
*Uncertainty on Fg**	Max σ	0.11	0.08
	$\sigma = max$	$0.32 \le NDVI_{pixel} \le 0.39$	$0.30 \le NDVI_{pixel} \le 0.44$
	$\sigma \ge 0.10$	$0.27 \le NDVI_{pixel} \le 0.42$	None

Results are shown for the linear model as computed using 3 different pairs of $NDVI_0$ and $NDVI_\infty$ (see Table 3) for both grassland and deciduous broadleaf forest.

Table 5
Same as Table 4, but using the quadratic model

A. Global AVHRR		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.21	0.14
	ΔFg* ≥ 0.15	$0.26 \le NDVI_{pixel} \le 0.43$	None
*Uncertainty on Fg**	Max σ	0.13	0.08
	σ = max	$0.36 \le NDVI_{pixel} \le 0.41$	$0.43 \le NDVI_{pixel} \le 0.52$
	σ ≥ 0.10	$0.31 \le NDVI_{pixel} \le 0.44$	None

B. AVHRR U.S.		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.12	0.08
	ΔFg* ≥ 0.15	None	None
*Uncertainty on Fg**	Max σ	0.09	0.06
	σ = max	$0.38 \le NDVI_{pixel} \le 0.48$	$0.46 \le NDVI_{pixel} \le 0.63$
	σ ≥ 0.10	None	None

C. MODIS U.S.		Grassland	Deciduous broadleaf forest
Over-estimation	Max ΔFg*	0.13	0.09
	ΔFg* ≥ 0.15	None	None
*Uncertainty on Fg**	Max σ	0.08	0.06
	σ = max	$0.40 \le NDVI_{pixel} \le 0.52$	$0.52 \le NDVI_{pixel} \le 0.61$
	σ ≥ 0.10	None	None

range, for which ΔFg* ≥ 0.15, varies depending on the land cover, moving towards larger NDVI values as the $NDVI_\infty$ increases. However, for the parameters derived here, the sensitive range remains around $0.2 \le NDVI_{pixel} \le 0.4$. These results are summarized in Table 4.

Fig. 4B show the results for the quadratic model (Eq. (3)) applied to Zeng et al. (2000) parameters. Here the maximum ΔFg* error is smaller and occurs at higher NDVI values than for the corresponding linear model. ΔFg* then decreases for both lower and higher NDVI values, similar to the linear model. For example, the quadratic model yields a maximum ΔFg* of 0.21 for grassland versus 0.24 using the linear model, and 0.14 (quadratic model) versus 0.18 (linear model) for deciduous broadleaf forest when using globally derived parameters. The sensitive range of the quadratic model is also narrower, meaning that the range of NDVI values at which the overestimation is the largest is less than for the linear model. For grassland, the ΔFg* error exceeds 0.15 for $0.26 \le NDVI_{pixel} \le 0.43$ with the quadratic model against $0.17 \le NDVI_{pixel} \le 0.39$ for the linear model. The quadratic model results are summarized in Table 5.

3.3. Uncertainty on the adjusted estimate

Because Fg* is a statistical estimate, it is accompanied by an uncertainty, σ, defined in Eq. (6). As discussed in Section 2, the range defined by Fg* ± σ (grey-shaded areas in Fig. 4) includes 68% of the possible adjusted Fg estimates. When using parameters derived from global scenes and the linear model, σ is the highest (±0.15) for $0.30 \le NDVI_{pixel} \le 0.35$, and is larger than ±0.1 for $0.20 \le NDVI_{pixel} \le 0.42$ for grassland. For deciduous broadleaf forest, σ is the highest (±0.11) for $0.32 \le NDVI_{pixel} \le 0.37$, and is larger than ±0.1 for $0.28 \le NDVI_{pixel} \le 0.42$. The maximum σ is smaller when using parameters derived from local scenes, i.e. σ = 0.12 for AVHRR and σ = 0.11 for MODIS. Table 4 summarizes the uncertainties on Fg* for both land cover types and include the corresponding values for the local AVHRR and MODIS scenes.

It is important to consider this uncertainty because, although Fg* and ΔFg* have been defined as the most-likely Fg value and associated error, the uncertainty in Fg* implies that there can still be instances of larger overestimation. For example, consider a grassland pixel with NDVI = 0.3. The most-likely Fg (Fg*) for that pixel is 0.33 when using the global AVHRR parameters (Table 3A) along with the linear model. But in reality the adjusted Fg estimate has a 68% chance of being anywhere between 0.18 and 0.48 (0.33 ± 0.15). Comparing this range to Fg = 0.57, the value obtained using Zeng et al. (2000) parameters, the potential overestimation of Fg can be as much as 0.39. This overestimation is large compared to some natural cycles such as seasonal variability in semi-arid environments, and underlines the importance of better defining $NDVI_o$.

3.4. Seasonal variability of the overestimation relative to the Fg signal

The time of the year at which the ΔFg* error is maximum varies due to differences in land cover NDVI seasonal cycles. This can be illustrated by mean NDVI time series for different locations and land cover computed from the mean MODIS 16-day NDVI values of each land cover type over the year 2003 (Fig. 5). In areas with strong seasonality (e.g. croplands and deciduous broadleaf forests), the error is maximum (within the $0.2 < NDVI_{pixel} < 0.4$ sensitive range) in the spring and in the winter, when photosynthetic activity is small and NDVI values are within the sensitive range. In the case of grasslands, the seasonality of its NDVI is less than for deciduous forests and croplands. Grassland NDVI peaks at around 0.4, which

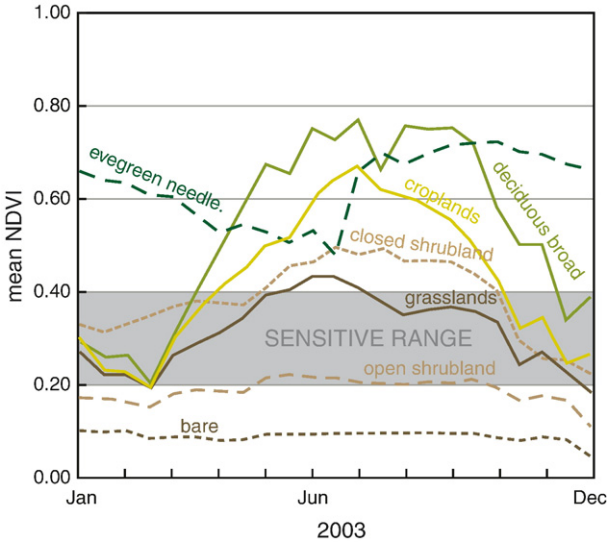

Fig. 5. Example of typical NDVI seasonal cycle for seven different land cover derived from conterminous U.S. MODIS 16-day NDVI imagery. The curves were derived using the mean NDVI value for each land cover type.

yields large overestimation throughout a typical year. The same can be said for shrublands, where vegetation remains fairly stable throughout a typical year. In areas covered by evergreen forest, the NDVI values are large throughout the year so that the overestimation in these areas remains small.

We now compare the amplitude of the overestimation to the one of the Fg signal. Fig. 6 summarizes the ΔFg* and Fg* for the six most common land cover types in the conterminous U.S. (Table 3). Each point shows the Fg* value computed using the mean NDVI values illustrated on Fig. 5 and the parameters in Table 3C. In the case of both evergreen needleleaf forests and open shrublands, the overestimation remains below 0.1, however the relative overestimation for both land cover types is different. Because the Fg signal for evergreen needleleaf forests is large throughout the year, the relative overestimation remains under 25%. In the case of open shrublands however, Fg is low (<0.2) throughout the year. As a result, the relative overestimation is large (25–100%). In the case of closed shrublands and grasslands, the overestimation is equal to or larger than 25% of the Fg value throughout most of the year. As for croplands and deciduous forests, the relative overestimation strongly oscillates and is negligible in the summer and fall, but can be large in the spring and winter (up to 100% the Fg signal).

As a result of the dependence of error on seasonality, the areas in the U.S. where the Fg overestimation is the largest migrate across the landscape as NDVI values fluctuate through time (Fig. 7). The overestimation affects the greatest areas in the spring and winter, as illustrated by the April 2003 data (Fig. 7D) for which 44% of the pixels have ΔFg*>0.10, compared to 28% in June 2003 (Fig. 7E). Although the overestimation is larger overall in April 2003, the maximum errors (ΔFg*>0.15) occupy about the same portion of the U.S. (7% of the pixels) than for the June data (5%). This is explained by the fact that in June, most U.S. grassland NVDI values reach the sensitive

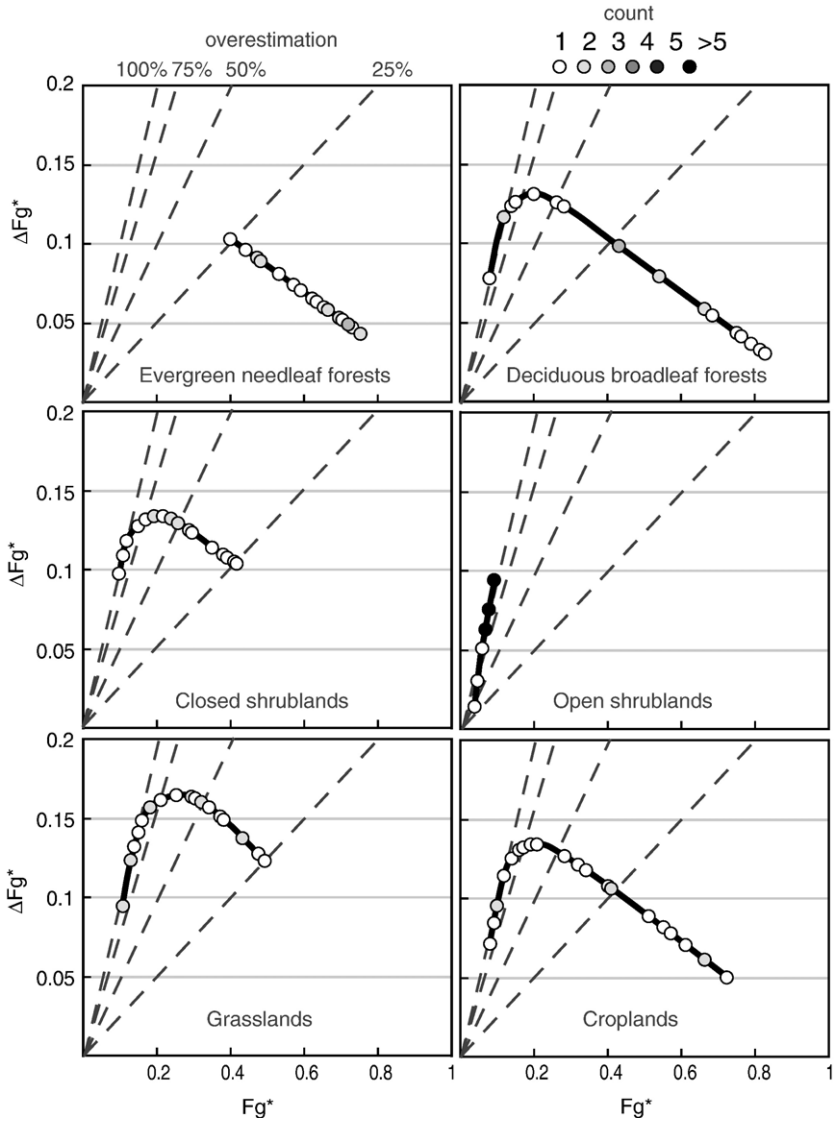

Fig. 6. ΔFg* and Fg* values for the six most common land cover types (bare excepted) in the conterminous U.S. as computed from the 2003 time series of MODIS 16-day NDVI data. Each curve shows the range of values derived from the mean NDVI value of each image (dots colors indicate the number of observations). The dashed isolines show the amount of relative Fg overestimation linked to the underestimation of $NDVI_o$.

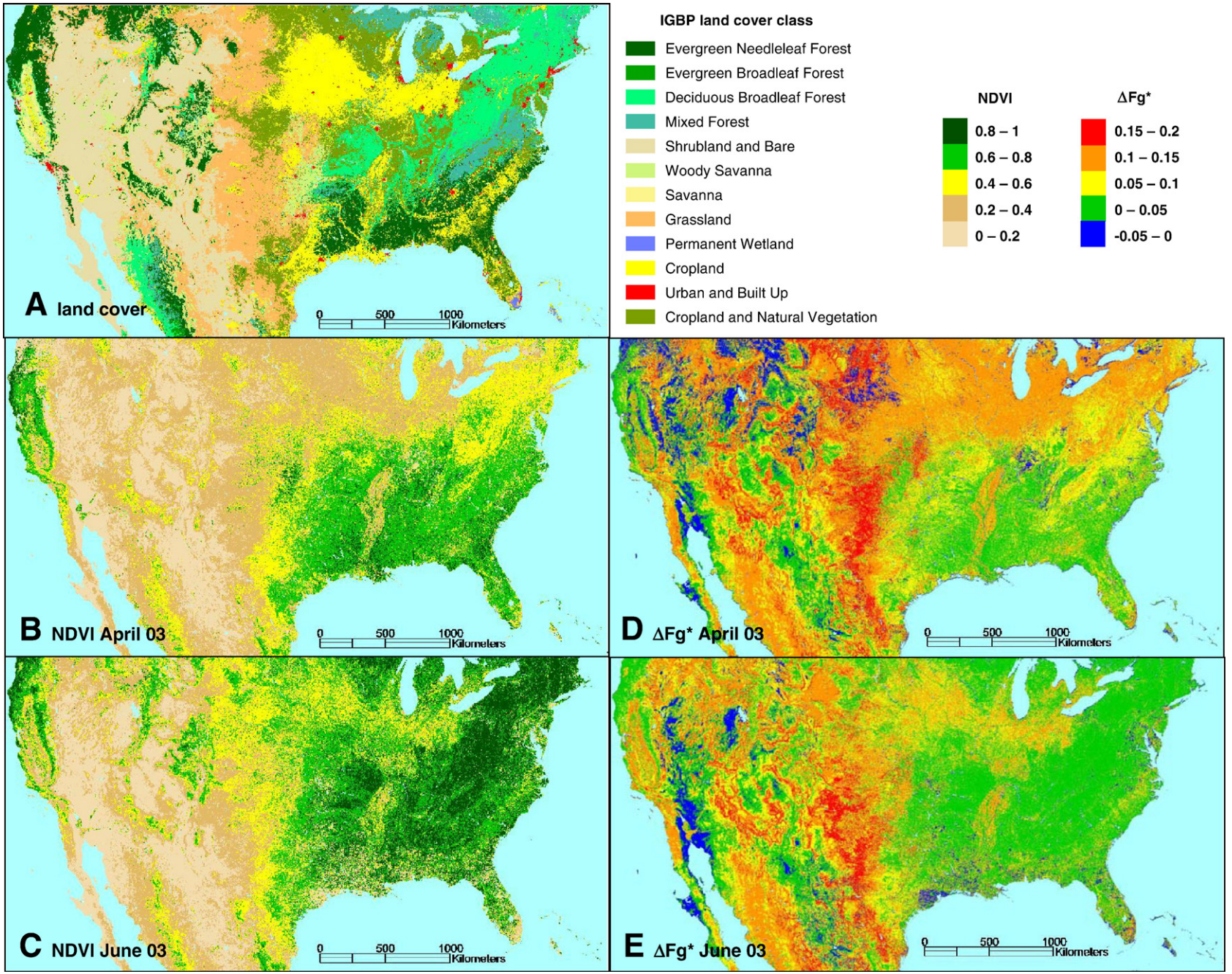

Fig. 7. Maps of the conterminous U.S. showing A) IGBP-based land cover classification, B) MODIS 16-day NDVI data for April 7 2003, C) MODIS 16-day NDVI data for June 10 2003, and ΔFg* errors computed using $NDVI_o$ and $NDVI_\infty$ in Table 3C from the D) April, and E) June NDVI values.

range, yielding maximum overestimation in a large part of the U.S. (most red areas in Fig. 7E are grassland).

4. Validation

We showed that underestimating $NDVI_o$ yields an overestimation of the green vegetation fraction. We also showed that this overestimation can be large relative to the Fg of some land cover types (in particular grassland and shrubland), and can vary with seasons. Because $NDVI_o$ can vary significantly between soils (Fig. 2) it is important to find a way to better constrain $NDVI_o$.

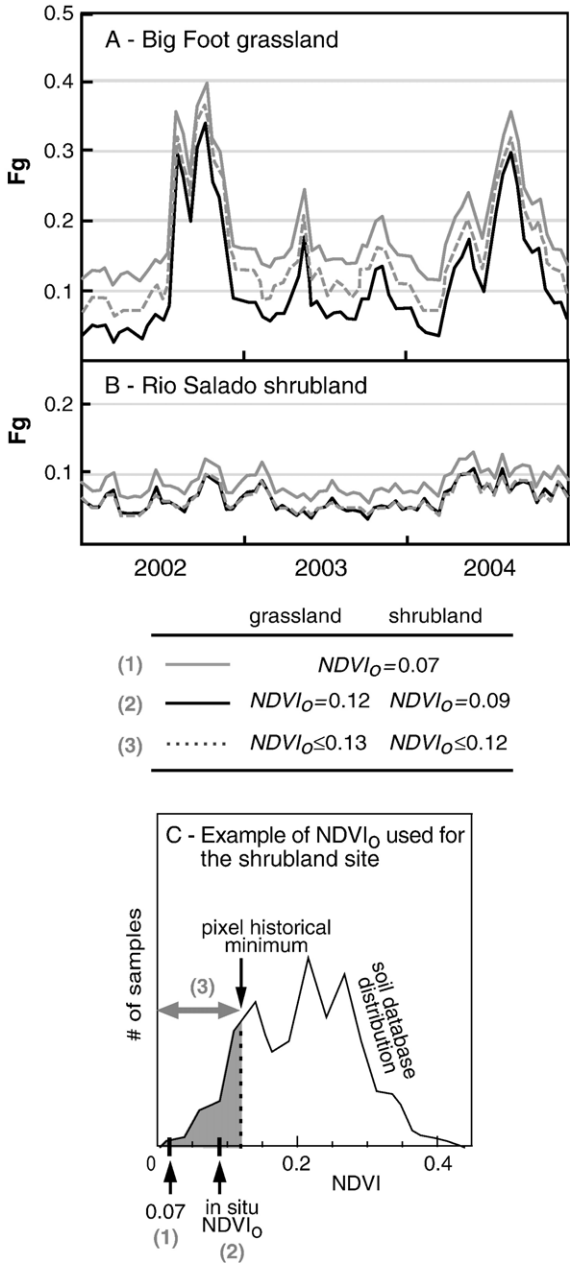

Fig. 8. Comparison of Fg values computed with different $NDVI_o$ at two locations within the Sevilleta LTER: A) a black grama grassland on Turney sandy clay loam with soil NDVI=0.12, and B) a creosote shrubland on Bluepoint fine sand with soil NDVI=0.09. The Fg values were computed using the corresponding $NDVI_\infty$ (Table 3C) along with 1) *in situ* $NDVI_o$ measurements (black solid line), 2) $NDVI_o=0.07$ (Table 3C, grey solid line), and 3) a range of $NDVI_o$ from the soil database that respect the condition $NDVI_o \leq$ lowest $NDVI_{pixel}$ (grey dashed line) as illustrated in C) for the shrubland site.

When no information on soils is available, the most popular method to estimate $NDVI_o$ is to use the lowest NDVI values of the studied scene. This method does not take into account the spatial variability of soils and tends to underestimate $NDVI_o$. Instead, we suggest using soil NDVI databases to compute a statistically most-likely Fg value as described by Eq. (4). With this statistical method, each pixel Fg is computed from all $NDVI_o$ samples that satisfy the $NDVI_o \leq NDVI_{pixel}$ condition. When dealing with temporal NDVI data, the range of possible $NDVI_o$ values can be further constrained by using the historical lowest NDVI of each pixel so that $NDVI_o \leq$ lowest $NDVI_{pixel}$.

Fig. 8 shows how this new approach affects Fg estimates. It compares the Fg values computed using 3 different methods to estimate $NDVI_o$: 1) using the *in situ* $NDVI_o$ measurements made at the grassland and shrubland Sevilleta LTER sites described earlier, 2) using Zeng et al. (2000) method as described in section 2, here $NDVI_o=0.07$ as derived from local MODIS time series, and 3) using for each pixel all samples from the soil database with $NDVI_o \leq$ lowest $NDVI_{pixel}$. The minimum NDVI used to constrain the possible $NDVI_o$ were inferred from a series of 2002 to 2004 MODIS 16-day NDVI data and are 0.13 for the grassland and 0.12 for the shrubland. As Fig. 8 shows, in both examples, the Fg computed using the historical minimum NDVI at each pixel along with the soil database yields much better estimates than the standard method using a single invariant $NDVI_o$ for each land cover type. In these two examples, the Fg estimate is adjusted by 52%, on average, for the grassland and 86% for the shrubland.

An alternative to the method we suggest here could be to use only the historical lowest NDVI value of each pixel as $NDVI_o$. However, this method is only valid for pixels that are over bare soil areas or where plants photosynthesis activity stops during part of the year (Fg=0). These conditions are hard to meet for most MODIS pixels, particularly because one pixel averages the reflectance signal over large (250×250 m) areas.

5. Conclusions

The soil reflectance data available from several datasets show that soils have a highly variable NDVI and that the mean value (0.20–0.21) is much larger than the $NDVI_o$ commonly used in Fg models ($\leq$0.05). As NDVI values for soils are not always available, $NDVI_o$ is generally inferred from the lowest NDVI value within a local remote sensing scene. This approach requires two assumptions to be true: the pixel with the lowest soil NDVI is free of vegetation (bare soil) and the soil spectral properties are invariant within the studied area. However, the soil dataset suggests that the invariant assumption is likely to be false. As a result, $NDVI_o$ is generally underestimated which yields overestimations of Fg. This problem is most severe in areas with sparse vegetation cover (e.g. grassland and shrubland) where typical seasonal NDVI values are in the sensitive range ($0.2<NDVI_{pixel}<0.4$) which results in the largest Fg overestimation.

Our results suggest that the overestimation is not as great when using the quadratic model (Eq. (3)) instead of the linear GI model (Eq. (2)). Also, $NDVI_o$ and $NDVI_\infty$ parameters should be derived

from local scenes. Using global values, such as those suggested by GI, yields larger overestimations (Tables 4 and 5). Most land cover types are affected by the overestimation as, apart from evergreen needleaf forests, they all reach the sensitive range at some time of the year. Shrubland and grassland areas are affected most strongly. Their NDVI tends to remain $0.2 < NDVI_{pixel} < 0.4$ throughout a typical year and the error, often equal to the actual Fg, reaches up to 0.24 when using globally derived $NDVI_o$ and $NDVI_\infty$. The overestimation in deciduous broadleaf forests and croplands is large in the spring and winter (up to 0.18 when using globally derived $NDVI_o$ and $NDVI_\infty$) but is negligible in the summer.

Because using an underestimated $NDVI_o$ can have a large impact on Fg computations, we recommend using $NDVI_o$ databases to derive adjusted estimates of Fg. Using *in situ* data at a grassland and a shrubland site at the Sevilleta LTER, we showed that combining this database with temporal NDVI information for each pixel can yield better estimations of Fg than using global invariant $NDVI_o$ values estimated from whole scenes. At the two sites the Fg estimation was adjusted by 52% and 86% for the grassland and shrubland respectively. Taking into account soil NDVI variability when using Fg models such as GI shows significant potential in improving Fg estimates. However more research is needed to quantify the degree of improvement this technique might provide. More significant advances will require information on the spatial distribution of soil reflectance. Additional spatial information, such as soil maps, might be useful to better constrain soil reflectance if a correlation exists between soil classifications and soil reflectances.

Acknowledgments

This research was supported by grant #:NNG04GO83G (CU Boulder) from the NASA Earth Science Enterprise (program manager J. Entin). We also would like to thank K. Shepperd from the World Agroforestry Center (ICRAF) Soil Laboratory for providing soil reflectance data.

References

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and PAR assessment. *Remote Sensing of Environment*, *35*, 161−173.

Baumgardner, M. F., Silva, L. F., Biehl, L. L., & Stoner, E. R. (1985). Reflectance properties of soils. *Advances in Agronomy*, *Vol. 38* (pp. 1−44). Orlando, Florida: Academic Press Inc.

Betts, A. K., Chen, F., Mitchell, K. E., & Janjić, Z. I. (1997). Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta Model using FIFE data. *Monthly Weather Review*, *125*, 2896−2916.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*, 241−252.

Choudhury, B. J., Ahmed, N. U., Idso, S. B., Reginato, R. J., & Daughtry, C. S. T. (1994). Relations between evaporation coefficients and vegetation indices studies by model simulations. *Remote sensing of Environment*, *50*, 1−17.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of NOAH land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, *108*, 12−11 to 12−16.

Gallo, K., Owen, T., & Reed, B. (2003). Land use and seasonal green vegetation cover of the Conterminous USA for use in numerical weather models. Paper presented at 17th Conference on Hydrology, Annual Meeting of the American Meteorological Society, Long Beach, Calif., 9–13 February.

Gallo, K., Tarpley, D., Mitchell, K., Csiszar, I., Owen, T., & Reed, B. (2001). Monthly fractional green vegetation cover associated with land cover classes of the conterminous USA. *Geophysical Research Letters*, *28*, 2089−2092.

Gan, T. Y., & Burges, S. J. (2006). Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. *Journal of Hydrology*, *320*, 117−131.

Gebremichael, M., & Barros, A. P. (2006). Evaluation of MODIS Gross Primary Productivity (GPP) in tropical monsoon regions. *Remote Sensing of Environment*, *100*, 150−166.

Gilabert, M. A., González-Piqueras, J., García-Haro, F. J., & Meliá, J. (2002). A generalized soil-adjusted vegetation index. *Remote Sensing of Environment*, *82*, 303−310.

Gutman, G., & Ignatov, A. (1997). Satellite-derived green vegetation fraction for the use in numerical weather prediction models. *Advanced Space Research*, *19*, 477−480.

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

Gutmann, E. D., & Small, E. E. (2007). A comparison of land surface model soil hydraulic properties estimated by inverse modeling and pedotransfer functions. *Water Resources Research*, *43*, W05418.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*, 295−309.

Jacquemin, B., & Noilhan, J. (1990). Sensitivity study and validation of a land surface parameterization using the Hapex-Mobilhy data set. *Boundary-Layer Meteorology*, *52*, 93−134.

Jensen, J. R. (2000). Remote sensing of the environment: An Earth resource perspective. *Series in geographic information science* (pp. 333−377). Upper Saddle River, New Jersey: Prentice Hall.

Li, X. -B., Chen, Y. -H., Shi, P. -J., & Chen, J. (2003). Detecting vegetation fractional coverage of typical steppe in northern China based on multi-scale remotely sensed data. *Acta Botanica Sinica*, *45*, 1146−1156.

Liang, X. Z., Li, L., Kunkel, K. E., Ting, M. F., & Wang, J. X. L. (2004). Regional climate model simulation of US precipitation during 1982–2002. Part I: Annual cycle. *Journal of Climate*, *17*, 3510−3529.

Matsui, T., Lakshmi, V., & Small, E. E. (2005). The effects of satellite-derived vegetation cover variability on simulated land–atmosphere interactions in the NAMS. *Journal of Climate*, *18*, 21−40.

Mitchell, K. (2001). *The community NOAH Land-surface Model (LSM) user's guide.* http://www.emc.ncep.noaa.gov/mmb/gcp/noahlsm/README_2.2

Molders, N., & Olson, M. A. (2004). Impact of urban effects on precipitation in high latitudes. *Journal of Hydrometeorology*, *5*, 409−429.

Oleson, K. W., Emery, W. J., & Maslanik, J. A. (2000). Evaluating land surface parameters in the biosphere–atmosphere transfer scheme using remotely sensed data sets. *Journal of Geophysical Research*, *105*, 7275−7293.

Price, J. C. (1993). Estimating leaf area index from satellite data. *IEEE Transactions on Geosciences and Remote Sensing*, *31*, 727−733.

Price, J. C., & Bausch, W. C. (1995). Leaf area index estimation from visible and near-infrared reflectance data. *Remote Sensing of Environment*, *52*, 55−65.

Qi, J., Kerr, Y. H., Moran, M. S., Weltz, M., Huete, A. R., Sorooshian, S., et al. (2000). Leaf area index estimates using remotely sensed data and BRDF models in a semiarid region. *Remote Sensing of Environment*, *73*, 18−30.

Sridhar, V., Elliott, R. L., & Chen, F. (2003). Scaling effects on modeled surface energy-balance components using the NOAH-OSU land surface model. *Journal of Hydrology*, *280*, 105−123.

Stoner, E. R., Baumgardner, M. F., Weismiller, R. A., Biehl, L. L., & Robinson, B. F. (1980). Extension of laboratory-measured soil spectra to field conditions. *Soil Science Society of America Journal*, *44*, 572−574.

United States Department of Agriculture. (1988). *Soil survey of Socorro County area, New Mexico.* Washington D.C., USA: Soil Conservation Service.

Yang, H., & Yang, Z. (2006). A modified land surface temperature split window retrieval algorithm and its applications over China. *Global and Planetary Change*, *52*, 207−215.

Zeng, X., Dickinson, R. E., Walker, A., Shaikh, M., DeFries, R. S., & Qi, J. (2000). Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. *Journal of Applied Meteorology*, *39*, 826−839.