



A data assimilation framework for constraining upscaled cropland carbon flux seasonality and biometry with MODIS

O. Sus^{1,*}, M. W. Heuer², T. P. Meyers², and M. Williams¹

¹The University of Edinburgh, School of GeoSciences and SAGES, West Mains Road, Edinburgh, EH9 3JN, UK

²Atmospheric Turbulence and Diffusion Division, NOAA/ARL, Oak Ridge, TN, USA

* now at: CREAM, Universitat Autònoma de Barcelona, Barcelona, Spain

Correspondence to: O. Sus (o.sus@creaf.uab.es)

Received: 31 July 2012 – Published in Biogeosciences Discuss.: 20 August 2012

Revised: 15 March 2013 – Accepted: 18 March 2013 – Published: 12 April 2013

Abstract. Agroecosystem models are strongly dependent on information on land management patterns for regional applications. Land management practices play a major role in determining global yield variability, and add an anthropogenic signal to the observed seasonality of atmospheric CO₂ concentrations. However, there is still little knowledge on spatial and temporal variability of important farmland activities such as crop sowing dates, and thus these remain rather crudely approximated within carbon cycle studies.

In this study, we present a framework allowing for spatio-temporally resolved simulation of cropland carbon fluxes under observational constraints on land management and canopy greenness. We apply data assimilation methodology in order to explicitly account for information on sowing dates and model leaf area index. MODIS 250 m vegetation index data were assimilated both in batch-calibration for sowing date estimation and sequentially for improved model state estimation, using the ensemble Kalman filter (EnKF), into a crop carbon mass balance model (SPAc). In doing so, we are able to quantify the multiannual (2000–2006) regional carbon flux and biometry seasonality of maize–soybean crop rotations surrounding the Bondville Ameriflux eddy covariance site, averaged over 104 pixel locations within the wider area.

(1) Validation at the Bondville site shows that growing season C cycling is simulated accurately with MODIS-derived sowing dates, and we expect that this framework allows for accurate simulations of C cycling at locations for which ground-truth data are not available. Thus, this framework enables modellers to simulate current (i.e. last 10 yr) carbon cycling of major agricultural regions. Averaged over

the 104 field patches analysed, relative spatial variability for biometry and net ecosystem exchange ranges from ~ 7 % to ~ 18 %. The annual sign of net biome productivity is not significantly different from carbon neutrality. (2) Moreover, observing carbon cycling at one single field with its individual sowing pattern is not sufficient to constrain large-scale agroecosystem carbon flux seasonality. Study area average growing season length is 20 days longer than observed at Bondville, primarily because of an earlier estimated start of season. (3) For carbon budgeting, additional information on cropland soil management and belowground carbon cycling has to be considered, as such constraints are not provided by MODIS.

1 Introduction

Agricultural ecosystems are of major importance to humankind. There are clear links between climate change, population growth (Zhang et al., 2011), and fluctuations in agricultural production (Lee et al., 2008). Global food demand is expected to double by 2050 (Tilman et al., 2001). Further agricultural intensification might have considerable detrimental effects on several crucial ecosystem services, including food production itself (Foley et al., 2005). Moreover, the biological dynamics of managed landscapes affect the fluctuations of atmospheric CO₂ levels on annual and inter-annual scales (Moureaux et al., 2008), and thus they need to be considered when quantifying, understanding, and regulating the global carbon (C) cycle (Sus et al., 2010).

Land management practices play a major role in determining global yield variability, of which only $\sim 30\%$ was found to be attributable to climate variables (Lobell and Field, 2007). However, still little is known about how important farmland activities such as crop sowing dates, cultivar selection, and fertilisation application vary in space and time (see Sacks et al., 2010; Siebert et al., 2005, as global examples). Uncertain crop phenology and sowing dates propagate into uncertain controls of food production (Sacks and Kucharik, 2011) and C cycling (McGuire et al., 2001) within large-scale agroecosystem models. This undermines progress in isolating natural from anthropogenic variability in atmospheric CO₂ concentrations, understanding the terrestrial C cycle, and predicting food security.

Measurements using eddy covariance (EC) to quantify net C fluxes are now being combined with simultaneous observations of crop biometry. These constraints on cropland C budgeting and climate controls are currently being used to improve biogeochemical models (BGCMS) on various scales (e.g. Bondeau et al., 2007; de Noblet-Ducoudré et al., 2004; Kucharik and Twine, 2007; Sus et al., 2010), which traditionally lacked a crop-specific plant functional type. Generally, BGCMS would benefit from a better representation of interannual phenological variability, which is poorly understood (Richardson et al., 2012; Stöckli et al., 2008).

As one attempt to overcome these shortcomings, remote sensing (RS) data have been used in crop growth modelling studies to estimate cropland management parameters (e.g. sowing or emergence date, Brown and de Beurs, 2008; Dente et al., 2008; Doraiswamy et al., 2004) and initial conditions (e.g. soil water content, Inoue and Oliosio, 2006), mostly applying batch-calibration data assimilation (DA) techniques. MODIS 250 m vegetation index (VI) data have been used for crop type classification (e.g. Wardlow and Egbert, 2008; Chang et al., 2007) and yield estimations (e.g. Doraiswamy et al., 2004). However, no study so far has attempted to assimilate RS time series of various field patches into a crop model simulating agroecosystem C exchange. In a recent study (West et al., 2010), cropland inventory data were combined with the MODIS land cover product to calculate spatially resolved estimates of C budgeting for the conterminous USA. However, as no mechanistic crop model has been applied and consequently no MODIS data were assimilated, the authors could not provide estimates of the seasonality of up-scaled C fluxes. To date, up-scaled model estimates of cropland net ecosystem exchange (NEE) fluxes have been forced with uncertain information on land management. In particular the variability of sowing dates in space and time remains rather crudely approximated within C cycle studies.

Here, we present a framework for spatio-temporally resolved simulation of cropland C fluxes under observational constraints on land management and canopy greenness. We apply DA methodology in order to explicitly account for information on sowing dates and model leaf area index (LAI). Sequential DA techniques such as the ensemble Kalman fil-

ter (EnKF, Evensen, 2003) have been developed and applied successfully within BGCMS (e.g. Quaife et al., 2008; Williams et al., 2005). We acquired 104 MODIS 250 m VI data time series, with pixel coordinates centred on field patches of sufficient size of 500 m \times 500 m surrounding the Ameriflux Bondville (Illinois, US) EC flux tower site. We then sequentially assimilated these data at observational time steps, using the EnKF, into a crop C mass balance model (SPAc, Sus et al., 2010) for improved C flux and biometry estimation, and compared model outputs against independent validation data sets.

In this study, we address the following research questions: (1) do MODIS 250 m VI time series contain appropriate information for improving crop C fluxes at the point scale? (2) Are the Bondville EC flux data representative of regional cropland C exchange, considering spatial heterogeneity in land use? (3) Are the region's maize–soybean crop rotations (MSCRs) a net sink or source of C?

2 Data and methods

2.1 Step 1: crop model calibration and application at the point scale

2.1.1 Study area and C flux data

We selected a study area of approximately 800 km² (32 km E–W, 25 km N–S), the centre pixel of which is close to the Ameriflux Bondville EC flux tower site (Ameriflux site ID: US-Bo1, latitude: 40.01° N, longitude: 88.29° W, Champaign, Illinois (USA), Fig. 1). The Bondville agricultural site has been used in no-till management of a soybean (*Glycine max*)–maize (*Zea mays*) crop rotation, with maize grown in odd years and the crop residues left on the field after harvest. The area surrounding the flux tower is characterised by predominantly well-drained silt loam soils with little surface slope (Wilson and Meyers, 2007). MSCRs have been analysed in various studies for their C sequestration potential under reduced or no-till management (Grant et al., 2007; Hollinger et al., 2004; Verma et al., 2005; Baker and Griffis, 2005), and can be considered as a model ecosystem for biogeochemical studies due to the flat terrain, homogeneous land-use pattern, and independent data sets available (Hollinger et al., 2004) for the wider Bondville region. The local climate is, according to the Köppen climate classification scheme, humid continental (class Dfa). Annual average precipitation over the period 2000 to 2006 is 754 mm.

At Bondville, half-hourly C exchange fluxes between the atmosphere and the biosphere have been measured continuously from 1996–present. Half hourly meteorological forcing data for radiation, temperature, wind speed, humidity, and precipitation have been recorded for diagnostics (see Meyers and Hollinger, 2004, for a detailed description). For several years, measurements have been taken of several aboveground

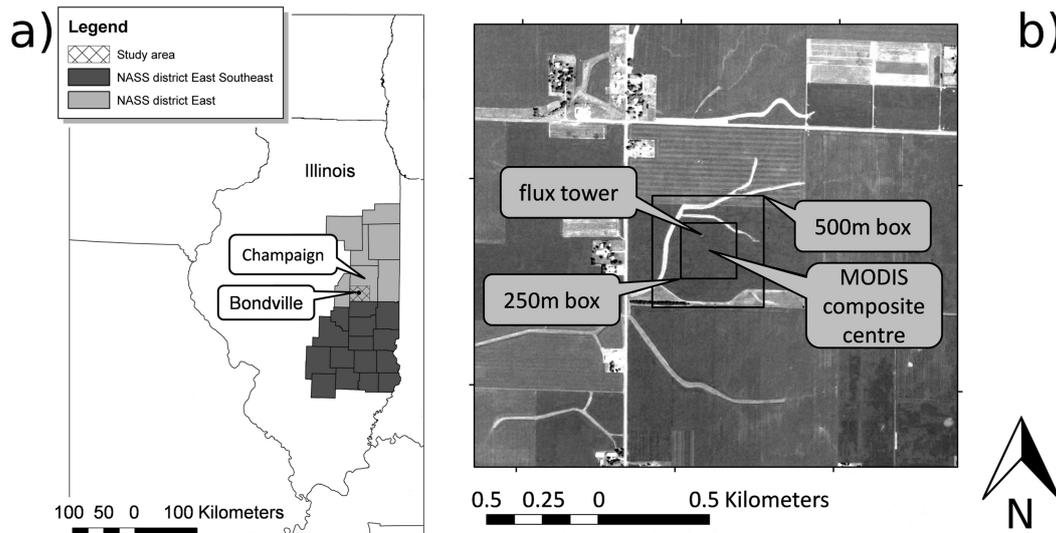


Fig. 1. (a) Location of the Bondville EC-flux tower site, study area, and NASS crop progress report districts. (b) The Bondville flux tower site. The selected MODIS composite centre location is indicated, together with approximated at-nadir extents of one 250 m MODIS pixel, and one 500 m MODIS pixel.

plant biometric variables, and farmer reported values on harvested biomass are available.

The flux data analysed in this study are half-hourly observations of NEE which were gap-filled based on a light use efficiency model for daytime values and on a respiration function for nighttime values (Bernacchi et al., 2005; Hollinger et al., 2004) to derive daily sums. The higher sink strength of the maize crop growing periods largely reflects the difference between C_3 (soybean) and C_4 (maize) net photosynthesis rates (Hollinger et al., 2004; Baldocchi, 1994).

2.1.2 Crop model: structure, parameterization, and initial conditions

The Soil Plant Atmosphere (SPA) model (Williams et al., 1996, 2001) is a process-based model that simulates ecosystem photosynthesis and water balance at fine temporal and spatial scales (30 min time step, up to 10 canopy and 20 soil layers). SPA employs some well-tested theoretical representations of ecophysiological processes, such as for the calculation of photosynthesis (the Farquhar model, Farquhar and von Caemmerer, 1982) and leaf-level transpiration (Penman–Monteith equation, Jones, 1992). These two processes are linked by a model of stomatal conductance, which optimises the daily gain of C per unit of leaf nitrogen within the limits of canopy water storage and soil to canopy water transport (Williams et al., 1996). A C mass balance model as described in Williams et al. (2005) has been added to SPA, and a C_4 photosynthesis model based on Collatz et al. (1992) was integrated.

Moreover, a crop C partitioning scheme and a developmental model have been added (SPA version 2 – Crop, Sus

et al. (2010), hereafter referred to as SPAc). The C partitioning scheme is based on empirical values of field crop growth analyses (Penning de Vries et al., 1989), and is a function of crop developmental stage (DS). The model representation of DS is introduced into SPAc by a new state variable, varying between 0 at emergence, 1 at flowering, and 2 at maturity. The duration of the phase between sowing and emergence is calculated through growing degree days, and lasts typically around 1 week. The progression of DS is based on non-linear functions for temperature $f(T)$ and photoperiod $f(P)$, with $0 < f(T, P) < 1$ (Streck et al., 2008; Sus et al., 2010; Setiyono et al., 2007). DS is calculated as the cumulative sum of daily maximum developmental rate (DR_{max}) multiplied by $f(T)$ and $f(P)$. For maize, $f(T)$ is the only control on developmental rate throughout the crop's life cycle within SPAc (Streck et al., 2008). For soybean, crop development is affected by $f(T)$ from sowing to maturity, and by $f(P)$ until $DS = 1$ (as in Setiyono et al. (2007), but simplified). Based on this development-linked C allocation pattern, SPAc simulates the allocation of C to roots, foliage, stems, and storage organs. Around emergence, C gained through photosynthesis is mainly allocated to leaves and roots at approximately equal amounts. However, C allocation begins to favour growth of stems as the crop matures throughout spring ($0.3 < DS < 1$), until before all assimilated C is allocated to storage organs in the reproductive phase ($DS > 1$). Leaf senescence for both crops is calculated as the bigger value of leaf senescence rate due to mutual shading (SR_{sh} , if $LAI > 5$) and leaf senescence rate as a function of physiological maturity or age (SR_{age} , if $DS > 1$, van Laar et al., 1997). We set SR_{age} to increase exponentially with DS, and 50% of senesced C mass is reallocated to storage organs.

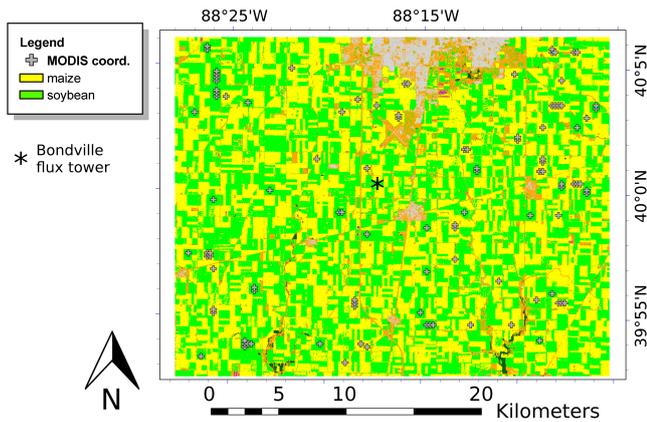


Fig. 2. The crop data layer of the study area. Selected MODIS composite coordinates are indicated (grey crosses). The study area is dominated by maize (yellow) and soybean (green) cultivations.

SPAc sensitivity to parameters controlling crop establishment and development has been analysed for cereal crops in Sus et al. (2010). Particularly small changes in cumulative NEE were found for a 25 % change in model parameters controlling timing of emergence (< 0.1 % change in cumulative NEE), critical LAI beyond which self-shading senescence is triggered (< 5 %), and maximum senescence rate due to self-shading (< 1 %). SPAc is particularly sensitive towards changes in DR_{\max} (< 20 %), and temperature (< 29 %) and photoperiod (< 30 %) developmental parameters.

SPAc runs were conducted for three different plant functional types: maize, soybean, and C_3 weed grasses growing in fallow periods between harvest and sowing. The initial conditions for the SPAc modelling runs were a soil organic matter (SOM) C content of 1300 g C m^{-2} for which modelled SOM C is in equilibrium, an adjusted litter C content of 400 g C m^{-2} to reflect annual variability in litter C as observed by Verma et al. (2005), and a labile C content of 10 g C m^{-2} at sowing for soybean and maize (i.e. the seed C content, approximated from Aubinet et al., 2009) and of 1 g C m^{-2} after harvest for the fallow period C_3 weed crop (approximated value).

2.2 Step 2: selection of field patches

For the identification of MODIS VI time series coordinates that are centred over single field MSCRs, USDA-NASS (US Department of Agriculture–National Agricultural Statistics Service) Cropland Data Layers (CDLs) were used as a classification basis (30 m ground resolution, total crop mapping accuracies range from 85 % to 95 % for major crop categories, Boryan et al., 2011). The seven NASS CDL raster images (one for each year in 2000–2006) served to extract time series of those MODIS pixels for which we find minimum requirements of crop type coverage to be satisfied (i.e. > 95 % coverage of crop type within a $500 \text{ m} \times 500 \text{ m}$ search

window). We found 104 pixels that met the defined requirements, 59 with maize sown in year 2000 (and with alternating crop types in following years), and 45 with soybean (Fig. 2).

2.3 Step 3: extracting single-field crop VI time series

We downloaded MODIS Terra (MOD13Q1) and Aqua (MYD13Q1) collection 5 (C5) 250 m data subsets from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL, 2010). Apart from vegetation indices, these MODIS data products also contain the red and NIR reflectance values themselves. We decided to remove all MODIS observations with a viewing angle > 40° , which eliminates all pixel distortions of factor > 1.6 (Wolfe et al., 1998). Thus, we expect only minor neighbouring effects in the retrieved Bondville MODIS data time series. Note that at the Bondville site our criterion for minimum crop type coverage as defined earlier is not strictly met (Fig. 1b).

We extracted MODIS time series information for all 104 selected pixels and all study years. We applied a filter to screen for pixel values affected by thick clouds by excluding all composite data with a blue reflectance value of > 10 % (Sakamoto et al., 2005), and removed all composite data with a reliability of > 1 (i.e. pixels most probably cloudy, ORNL, 2010). To improve temporal accuracy of the retrieved VI time series, all MODIS data were associated with their true observation date using the “composite day of year” information (Solano et al., 2010).

To scale from modelled LAI to modelled VI, we applied an empirical relationship based on the renormalized difference vegetation index (RDVI, developed by Haboudane et al., 2004, for maize and soybeans grown in Ottawa, Canada). The RDVI has been developed in order to attain a more linearised relationship with vegetation biophysical variables compared to the NDVI (Roujean and Breon, 1995). The RDVI is defined as

$$RDVI = \frac{\rho_{NIR} - \rho_{red}}{\sqrt{\rho_{NIR} + \rho_{red}}}, \quad (1)$$

and the empirical relationship as

$$LAI = 0.0918 \times \exp^{6.0002 \times RDVI}. \quad (2)$$

When comparing RDVI-derived LAI with ground truth measurements, Haboudane et al. (2004) found an R^2 of 0.90–0.95 and an overestimation of LAI values > $5 \text{ m}^2 \text{ m}^{-2}$. The coupling of a canopy transfer model to the crop C mass balance model for the provision of a modelled VI output is beyond the scope of this study, but has been tested by Quaife et al. (2008), who assimilated MODIS spectral reflectance rather than LAI-product data into an ecosystem model.

2.4 Step 4: determination of individual sowing dates for each pixel

We ran SPAc 80 times in forward mode (i.e. no data assimilated, forced with Bondville observed meteorology) for all

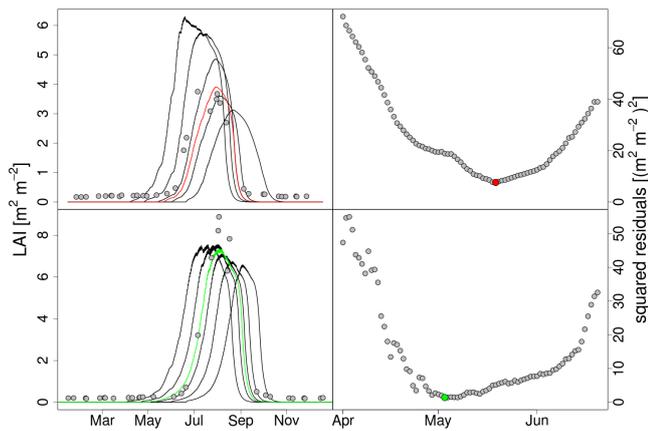


Fig. 3. Left panels: model LAI curves are shown for maize (top) and soybean (bottom) and 5 different sowing dates (DOY 90, 110, 130, 150, 170, black lines) for one study area example field patch. MODIS LAI values are shown as grey dots. Model LAI with lowest squared residuals is shown as coloured line. Right panels: squared residuals between model LAI and MODIS LAI for each of the 80 sowing dates and maize (top) and soybean (bottom). The coloured dot depicts the lowest squared residuals value.

104 selected pixels and each year of the study period, but each time with a different sowing date. These sowing dates span a range of day of year (DOY) from 90 to 170, thus encompassing reported usual values for maize and soybean in Illinois (USDA, 1997). We expect that the application of uniform meteorology over the study area has negligible effects on model performance regarding spatially continuous climate variables such as temperature and vapour pressure deficit. However, the spatial distribution of precipitation is not accounted for. Modelled field patches are located less than ~ 16 km away from the flux tower, and so their water balance is subjected to manageable uncertainty. Out of the resulting 80 modelled LAI curves, we solved for the sowing date with the minimum sum of squared residuals between its corresponding modelled LAI and the MODIS-derived LAI data over the entire calendar year. In order to compare our estimation of modelled sowing dates with general patterns of land management within the wider area, we referred to NASS crop progress reports for two neighbouring census districts, but note the study area's limitation in spatial representivity with respect to the crop census districts (Fig. 1a).

2.5 Step 5: model upscaling through MODIS VI DA

DA can be considered as a set of techniques that aims at finding an optimal combination of observations and models, referred to as the “analysis” (Mathieu and O’Neill, 2008). For this study, we selected – apart from batch-calibration of sowing dates – the EnKF approach for model state estimation as such a DA technique (Evensen, 2003; Williams et al., 2005). We expected considerable changes in model perfor-

mance after DA updates only of LAI and root C mass, but for the purpose of model diagnosis and validation, the state vector contained all above- and belowground biometric variables and the RDVI. We selected a model uncertainty of 1 % for all biometric variables, and a higher uncertainty of 10 % for the RDVI in order to account for additional uncertainties in scaling from LAI to RDVI. We used the temporal separation approach (Hollinger and Richardson, 2005) to estimate RDVI uncertainty, defining any differences in MODIS values that are < 4 days apart as data uncertainty originating from various sources such as sensor calibration and atmospheric conditions. This approach produced an uncertainty value of ± 0.017 s.d. for the RDVI, which is comparable to the MODIS Land Discipline Team (MODLAND, <http://modis-land.gsfc.nasa.gov/>) values for the NDVI (± 0.025) and EVI (± 0.015). We chose an ensemble size of 50 members, for which we found stabilising RMSE between SPAC modelled and observed NEE in a previous study (Sus, 2011, Sect. 4). Similarly, de Wit and van Diepen (2007) found that the soil moisture ensemble mean of a crop model can be well estimated with an ensemble size of 50, and improvements are small when this metric is increased to 100.

3 Results

3.1 Simulated sowing dates

Sowing date is a key control of LAI magnitude and seasonality. A model run initialised at the earliest sowing date (DOY = 90) of maize produces a maximum LAI value about twice as large compared to a model run initiated by the end of the plausible range of values (DOY = 170, Fig. 3). The timing of these maximum LAI, dependent on sowing date, can be up to 2 months apart. For soybean, maximum LAI is less sensitive to sowing date, but a strong control on seasonality is obvious as well. Due to its sensitivity to day length and its indeterminate character (meaning that vegetative and reproductive phases overlap), soybean is, phenologically speaking, more stable than maize. The duration of leaf growth and thus maximum LAI are less strongly affected by delays in sowing date. There is a clear minimum of squared residuals as a function of sowing date for the example maize field patch, whereas a range of sowing dates spanning about 1 week appears plausible for soybean (Fig. 3). The example illustrates both the model's sensitivity to sowing dates, and that an optimal sowing date can be determined within an uncertainty range of ± 3 days. It remains to be determined whether a sowing date optimisation for seasonality rather than magnitude would further improve the phase agreement of observed and modelled LAI.

Study-area-modelled sowing dates broadly reflect observed trends of planting progress as reported by NASS for the two crop census districts, but discrepancies exist (Fig. 4). Except for 2003, modelled soybean sowing dates are at least

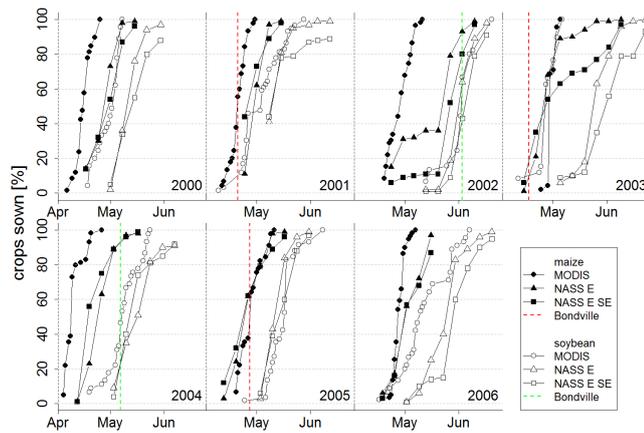


Fig. 4. Crop sowing progress for the study area and the two surrounding NASS census districts (east and east-south east) for the seven study years (2000–2006). MODIS-derived model estimates (circles) and NASS reported values (triangles and squares) are shown as percentage of crops sown within the respective regions. Reported sowing dates for the Bondville EC flux site are indicated by dashed vertical lines (red: maize; green: soybean).

1–2 weeks later than maize (as also reported by NASS and USDA, 1997). NASS observations and MODIS-derived sowing dates are closest in 2005 for both crops (< 1 week difference). Soybean sowing dates appear better reproduced by MODIS than those of maize, which are generally premature. 100 % sowing progress of maize is often reached by end of April, whereas reported sowing activity often lasts well into May. Model–observations differences are most obvious in 2002 and 2003.

In contrast to the above, a comparison of farmer reported (i.e. directly observed, not NASS-derived) and simulated sowing dates for the Bondville site shows that modelled maize sowing DOYs, except for 2001, are overestimates (reported + modelled DOY (year): 109 + 108 (2001), 106 + 118 (2003), 116 + 127 (2005)), whilst soybean sowing DOYs are only slightly underestimated (153 + 152 (2002), 127 + 123 (2004)). The Bondville data also show that reported sowing of the 2002 soybean crop is clearly delayed due to abnormal precipitation in April–June, which is well reproduced by the MODIS-based model value. Except for 2003, Bondville observations are within the 40–80 % range of modelled sowing progress (vertical coloured lines in Fig. 4).

3.2 Proof of concept – sequential assimilation of MODIS RDVI time series at Bondville

Our assessment in this section is twofold: firstly, to analyse whether sequential MODIS DA improves C flux and biomass estimation when SPAC is driven with reported sowing dates. Secondly, we conduct the same analysis for SPAC outputs under satellite-derived (i.e. modelled) sowing dates. We conducted the following 4 model experiments: (1) model for-

ward (no sequential DA) run forced with farmer reported sowing dates (FWrep), (2) as FWrep but with sequential MODIS DA (DArep), (3) model forward run forced with modelled sowing dates (FWmod), and (4) as FWmod but with sequential MODIS DA (DAmo).

Ground-observed and MODIS-derived LAI generally agree well with each other at Bondville ($R^2 = 0.74$), but MODIS data are negatively biased (mean error (ME, i.e. mean of MODIS LAI minus Bondville ground-truth) = $-0.49 \text{ m}^2 \text{ m}^{-2}$, RMSE = $1.08 \text{ m}^2 \text{ m}^{-2}$, Fig. 5). The agreement is particularly strong with regards to the observed seasonality of LAI evolution of maize and soybean, which is in turn largely controlled by site-specific sowing dates. However, maximum ground-observed LAI values are often underestimated by MODIS, e.g. by $\sim 0.5\text{--}1 \text{ m}^2 \text{ m}^{-2}$ for maize 2003 and soybean 2005, but note overestimation in 2004. MODIS-derived LAI is mostly $< 0.3 \text{ m}^2 \text{ m}^{-2}$ during fallow periods.

The sequential assimilation of MODIS RDVI data generally improves the simulation of LAI by reducing a negative bias (ME = $-0.45 \text{ m}^2 \text{ m}^{-2}$ (FWrep), -0.41 (DArep), -0.57 (FWmod), -0.40 (DAmo)) and constraining seasonality (Fig. 5), which is generally better captured than overall magnitude. With the assimilation of MODIS RDVI (DAmo), modelled LAI is now generally closer to ground-truth data (R^2 increased by 0.10, RMSE reduced by $\sim 20\%$). Growth of C_3 grass is reduced due to the assimilation of relatively low RDVI during fallow periods. Sequential DA successfully informed about rapid growth following late sowing of the soybean crop in 2002 (which was underestimated by FWmod, Fig. 5), and thus appears suitable for correcting model deficiencies during anomalous years.

Growing season observed NEE data are generally well reproduced in terms of magnitude and seasonality, with relatively similar model fits for all experiments: $R^2 = 0.55$ (FWrep), 0.53 (DArep), 0.56 (FWmod), 0.55 (DAmo). Model bias is reduced through sequential DA (ME (mod. – obs.) = $0.99 \text{ g C m}^{-2} \text{ d}^{-1}$ (FWrep), 0.65 (DArep), 1.15 (FWmod), 0.76 (DAmo)). However, there are still discrepancies in terms of a delayed start of the crop C assimilation phase (2002), timing/rate of senescence (data not shown), and of an overall underestimation of the growing season sink strength as indicated by biases. Sequential MODIS DA appears generally useful to correct for model deficiencies during growing seasons, but the informational content about fallow season C_3 grass growth and C assimilation needs further validation.

Cumulative all-year NEE is generally overestimated by model experiments (by $\sim 600 \text{ g C m}^{-2}$ after 5 yr), with this overestimation being $\sim 100 \text{ g C m}^{-2}$ larger for experiments with modelled sowing dates (data not shown). In general, there is little difference in final estimates between FW and DA experiments ($< 30 \text{ g C m}^{-2}$). Observations and experimental data are rather close until before weed grass growth in spring 2003. A careful analysis of these data should bear in mind that “observed” cumulative NEE is estimated on the

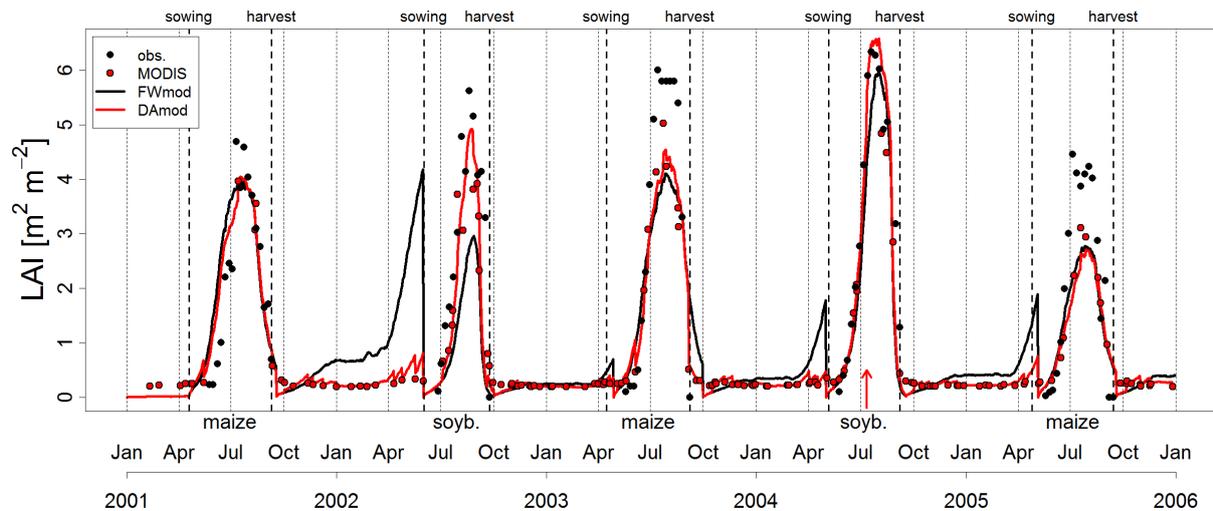


Fig. 5. Observed (black circles), MODIS-derived (red circles), and modelled LAI (experiments FWmod and DAMod, black and red lines) for the Bondville EC flux tower site, 2001–2006. The vertical dashed lines denote farmer reported sowing and harvest dates for the flux site. The red arrow in July 2004 denotes the location of maximum MODIS-derived LAI ($\sim 9.1 \text{ m}^2 \text{ m}^{-2}$, not shown in figure).

basis of both measured and gap-filled half-hourly NEE. Of the data analysed here, $\sim 51\%$ of half-hourly NEE is gap-filled.

Whereas Bondville measurements suggest that this no-till agroecosystem is still a C sink after accounting for harvested C mass ($\text{NBP} = 209 \text{ g C m}^{-2}$, with NBP defined as net biome productivity, $= -\text{NEE} - \text{C}_{\text{stor}}$), model data suggest that the MSCRs are about C neutral at best ($\text{NBP} = -17$ (FWrep), -58 (DArep), -64 (FWmod), -256 (DAmo) g C m^{-2}). The gap between FW and DA model data primarily results from the failure of sequential MODIS DA to inform about fallow season C uptake and upward corrections of simulated yield (see also Fig. 6, but note that model data are study area averages, and Bondville-only model data are not shown).

3.3 Upscaling – cropland C cycling within the study area after sequential MODIS DA

3.3.1 Statistical assessment of study area model averages and observations

Averaged over 104 pixels, the annual cumulative NBP of the study area is positive for soybean ($13 \text{ g C m}^{-2} \text{ yr}^{-1}$) and negative for maize ($-58 \text{ g C m}^{-2} \text{ yr}^{-1}$, Table 1). Through DA, we are able to quantify a considerable spatial variability of cumulative NEE (mean standard deviation: $62 \text{ g C m}^{-2} \text{ yr}^{-1}$ (maize), $38 \text{ g C m}^{-2} \text{ yr}^{-1}$ (soybean)) and NBP ($72 \text{ g C m}^{-2} \text{ yr}^{-1}$ (maize), $47 \text{ g C m}^{-2} \text{ yr}^{-1}$ (soybean)) within the study area. This shows that the NBP of both crop rotations is not significantly different from 0, which is also true for most individual growing seasons including fallow periods.

Observed NBP is negative for four years, but strongly positive in 2003. The strong sink in 2003 alone is largely re-

sponsible for overall positive observed NBP after five years, as the ecosystem loses on average $\sim 59 \text{ g C m}^{-2} \text{ yr}^{-1}$ during other years. Observed NEE is clearly lower than modelled for maize, but equal to or larger during soybean years. In terms of NEE, simulated sink size is larger for maize than soybean (by $\sim 54 \text{ g C m}^{-2} \text{ yr}^{-1}$), but maize is a stronger source of C (by $\sim 71 \text{ g C m}^{-2} \text{ yr}^{-1}$) in terms of NBP (Table 1).

Model yield is on average $\sim 130 \text{ g C m}^{-2}$ higher for maize than soybean. Yield variability is $\sim 9\%$ of the study area mean for both crops. Compared to Bondville observations, model C_{stor} is on average $\sim 130 \text{ g C m}^{-2}$ lower for maize, and about equal to larger for soybean (by $\sim 24 \text{ g C m}^{-2}$ on average, but note differences to farmer reported yield). Maximum LAI of soybean is on average $\sim 2.6 \text{ m}^2 \text{ m}^{-2}$ greater than that of maize, and study area variabilities of both crops are comparable ($\sim 0.5 \text{ m}^2 \text{ m}^{-2}$). Differences between model values and observations are, except for 2004, within $0.5 \text{ m}^2 \text{ m}^{-2}$ (Table 1).

In general, assimilation of RDVI data through the EnKF resulted in considerable updates in state variables LAI, leaf, and root C mass. Whereas LAI (but not leaf C mass) changes have significant consequences on gross primary productivity (GPP) and thus C cycling, changes in root C mass slightly altered model sensitivity to drought. Stem and storage organ C mass showed only small to moderate updates, and are not physiologically integrated into the model. Biometric updates had no immediate consequences on autotrophic respiration, which within SPA is simulated through the gradual turnover of a respiration C pool and thus is not directly biomass dependent.

Table 1. Observed (at Bondville) and study area model averages and standard deviations for maximum LAI ($\text{m}^2 \text{m}^{-2}$), storage organ C at maturity (StOrg (g C m^{-2})), and all-year cumulative NEE and NBP ($\text{g C m}^{-2} \text{yr}^{-1}$). Observed storage organ C is shown for field dry matter samples, and farmer reported values (in parentheses). Bold model values are same crop types as at Bondville. Model values (experiment DAmo) are averages of all soybean (SB) or maize (MZ) field patches in a given year. Data are only shown for years when Bondville dry matter data are available for comparison (2001–2005).

	2001	2002	2003	2004	2005	mean	Σ
LAI							
obs	4.7	5.6	6.0	6.3	4.5	–	–
SB	7.3 ± 0.3	6.1 ± 0.6	7.4 ± 0.4	7.9 ± 0.5	7.7 ± 0.5	7.4 ± 0.5	–
MZ	5.2 ± 0.4	3.8 ± 0.7	5.5 ± 0.3	4.9 ± 0.3	4.1 ± 0.4	4.8 ± 0.5	–
StOrg							
obs	487 (474)	217 (174)	545 (542)	267 (169)	635 (463)	–	–
SB	278 ± 24	204 ± 24	297 ± 22	329 ± 34	312 ± 29	290 ± 27	–
MZ	441 ± 38	239 ± 45	472 ± 36	509 ± 33	383 ± 38	420 ± 38	–
NEE							
obs	–405	–143	–992	–252	–570	–316	–2362
SB	-287 ± 42	-139 ± 27	-342 ± 40	-358 ± 48	-357 ± 34	-297 ± 38	-1483 ± 87
MZ	-333 ± 72	-204 ± 73	-492 ± 48	-430 ± 49	-296 ± 66	-351 ± 62	-1755 ± 140
NBP							
obs	–82 (–69)	–74 (–31)	447 (450)	–15 (83)	–65 (107)	29 (95)	211 (540)
SB	9 ± 48	-65 ± 36	45 ± 46	29 ± 59	45 ± 45	13 ± 47	63 ± 106
MZ	-108 ± 81	-35 ± 86	20 ± 60	-79 ± 59	-87 ± 76	-58 ± 72	-210 ± 164

3.3.2 Time series analysis of modelled and observed NEE and NBP

Simulated study area average values are 240 to 300 g C m^{-2} lower than the Bondville cumulative NEE flux data value ($\sim -2970 \text{ g C m}^{-2}$) after the seven study years and when only considering the growing season C balance (sowing–harvest, Fig. 6a). For all season NEE data however (Fig. 6b), the gap between modelled and observed NEE by end of 2006 is somewhat smaller and its sign has changed (~ 108 to 224 g C m^{-2} higher than Bondville observations). The study area-wide modelled NEE variability has a standard deviation of $\sim \pm 87$ to 140 g C m^{-2} after 7 yr. Observed cumulative NBP data indicate a net sink of $\sim 211 \text{ g C m}^{-2}$ after five years, however with changing sign so that observations are a mix of source and sink years. Modelled crop rotations indicate a net source of ~ 12 to 208 g C m^{-2} (Fig. 6c). Note that these values are similar to values shown for specific crop types (i.e. not crop rotations) in Table 1. Model data are close to Bondville observations until after autumn 2001 (dashed line in Fig. 6c), but subsequently a gap continuously builds up during fallow periods (especially spring 2003).

Most importantly, model data suggest clear differences in the seasonality of MSCR C fluxes within the study area compared to observations. These differences are mainly attributable to earlier estimated start of season (SoS, i.e. timing of sink-crossover after sowing using a 10 day moving average of the time series analysed) values. Study area aver-

ages of modelled soybean–maize rotations generally reflect interannual patterns in SoS and end of season (EoS, i.e. timing of source-crossover after SoS) as observed at Bondville (Table 2), but discrepancies exist. Model-derived SoS are generally earlier than Bondville values (by ~ 2 weeks on average, soybean–maize (SM) rotations in Table 2), and DOYs are closer to Bondville values for soybean than maize years. This prematurity in SoS is even larger for maize–soybean (MS) rotations (~ 25 days) and the study area flux average (AV, ~ 23 days). In 2002, model SoS values are > 1.5 months premature (MS and AV). These discrepancies are smaller for EoS estimates, which are again generally earlier than “observed” by ~ 5 days (SM), ~ 11 days (MS), and ~ 4 days (AV). Consequently, modelled average growing season length ($\text{GSL} = \text{EoS} - \text{SoS}$) is about 9 to 20 days longer than derived from Bondville measurements.

Moreover, the Bondville flux data appear representative of SM rotations within the study area. The magnitude of modelled growing season C fluxes reflects observed patterns well (green line in Fig. 7a). Fallow season weed growth (C uptake) and decay (C emission) are not reproduced, but post-harvest NEE fluxes (dominated by crop residue decomposition) are close to measurements. As expected, modelled MS rotations (red line in Fig. 7a) clearly differ in growing season C uptake magnitude and seasonality during all study years.

The pattern of study area mean model fluxes reveals the year-to-year variability in regional ecosystem sink strength (black line in Fig. 7b, which is the average of all maize and

Table 2. Observed (obs., at Bondville) and modelled (mod.) start of season (SoS), end of season (EoS) DOYs, and growing season length (GSL) in days. Model data (experiment DAmoD) are shown for pixels with soybean grown in even years (soybean–maize (SM)), as Bondville site rotation), with soybean grown in odd years (maize–soybean (MS)), and averaged over all pixels (AV). ME = mean error (mod. – obs.).

	2000	2001	2002	2003	2004	2005	2006	mean	ME
Bondville crop	soy	maize	soy	maize	soy	maize	soy		
SoS									
obs.	166	139	184	155	152	150	155	157	–
mod. (SM)	160	123	182	127	145	130	132	143	–14
mod. (MS)	126	139	130	132	110	156	128	132	–25
mod. (AV)	145	131	131	130	131	135	130	134	–23
EoS									
obs.	238	242	254	236	231	242	259	243	–
mod. (SM)	244	216	256	229	249	215	256	238	–5
mod. (MS)	216	246	212	252	219	252	227	232	–11
mod. (AV)	225	239	241	236	243	244	245	239	–4
GSL									
obs.	72	103	70	81	79	92	104	86	–
mod. (SM)	84	93	74	102	104	85	124	100	14
mod. (MS)	90	107	82	120	109	96	99	95	9
mod. (AV)	80	108	110	106	112	109	115	106	20

soybean pixels for each year), rather than sink strength variability due to C₃/C₄ crop type rotation. The magnitude of C sink strength in 2002, when sowing dates have been reported to be particularly late, is considerably lower compared to other years. As expected, model data show larger (lower) C uptake during Bondville soybean (maize) growing periods than observed. Clear seasonality shifts towards earlier DOYs are most obvious during Bondville soybean years (especially 2002).

4 Discussion

4.1 Does MODIS DA improve model reproduction of Bondville ground-truth data?

4.1.1 Model improvement through simulated sowing dates

Modelled sowing dates allow for a realistic simulation of NEE seasonality ($R^2 = 0.56$ (FWmod) at Bondville site). However, C sink underestimation increases simultaneously by $\sim 100 \text{ g C m}^{-2}$ (data not shown). MODIS RDVI was used as an independent constraint on crop LAI for finding the model sowing date which best reproduces observed seasonality in green aboveground biomass. This procedure clearly constrains the seasonality of cropland NEE. Most probably, we could have produced a better agreement between observed and modelled sowing dates by using only (and temporally more highly resolved) MODIS data of the vegetative phase. However, this approach would probably be detri-

mental for the merits of our sequential DA procedure during senescence, the simulated timing and rate of which still need improvement (Kucharik and Twine, 2007; Sus et al., 2010).

Upscaled differences between MODIS-derived sowing dates and NASS-observed values are indicative of deficiencies in model LAI and photosynthetic potential under ambient climatic conditions (Fig. 4). As modelled maize sowing dates are mostly premature, we conclude that forward mode green biomass is probably underestimated. In other words, a negative model bias in maize LAI is compensated by an earlier start of the growing season. Even though discrepancies with observations exist (2000 and 2003), MODIS-derived soybean sowing dates appear more realistic. The reported delay in sowing progress in 2002 is well captured for soybeans, but is not reflected for maize. A more detailed analysis of the quality of modelled sowing dates is not possible, as scale differences between sub-county model data and census district-level observations (Fig. 1a) are prohibitively large. More steps need to be taken to move from a qualitative to a more quantitative validation of satellite-derived sowing dates (Wardlow et al., 2006). Progress in this respect is probably most hindered by the lack of reliable, spatially resolved validation data.

The general applicability of MODIS data for sowing date assessment has been successfully demonstrated in other studies. Using MODIS NDVI and rainfall data, Brown and de Beurs (2008) were able to produce high relationships between derived sowing dates and observations for West Africa ($R^2 = 0.89$). MODIS-derived sowing dates were consistent

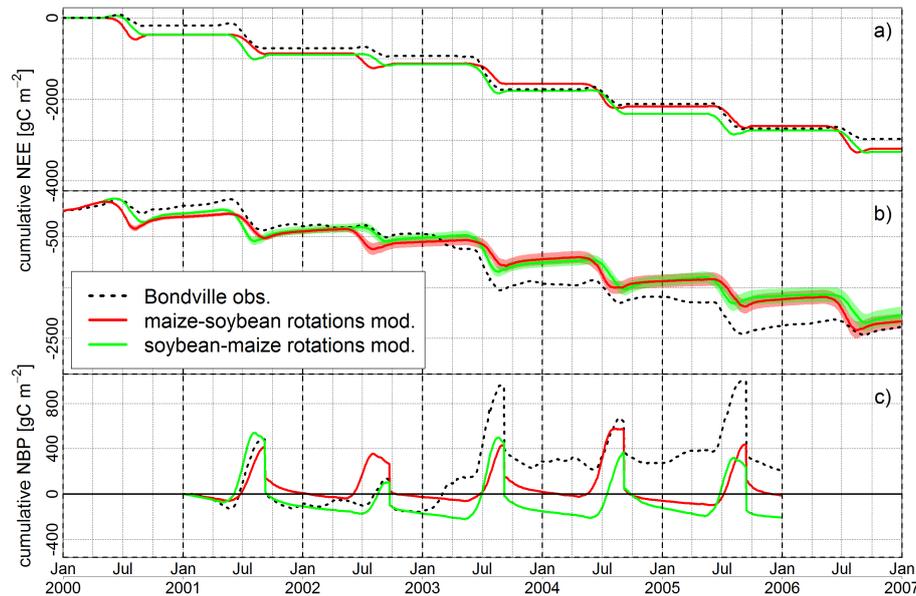


Fig. 6. Study area model averages (experiment DAMod) and Bondville EC data of (a) cumulative NEE for growing periods only, (b) cumulative NEE ($\pm 1 \sigma$ study area model variability, shown as shaded areas), (c) cumulative NBP. Maize–soybean rotations: maize sown in even years, and vice versa for soybean–maize rotations.

with the relative order of sowing of major crops in Kansas (US, Wardlow et al., 2006).

Even though we expect mixed-pixel effects on retrieved MODIS time series for Bondville (Fig. 1b), there is clear evidence that forward model runs with modelled sowing dates (FWmod) reproduce the agreement of model runs with reported sowing dates (FWrep) with independent validation data (NEE). These findings suggest that modelled sowing dates themselves are suitable constraints for agroecosystem C cycling and biomass growth at the point scale, and also partly compensate for model shortcomings in simulating young crop establishment as indicated by bias towards premature sowing in Fig. 4. Despite this compensation, the qualitative agreement of the temporal development of modelled sowing progress with NASS observations suggests that a realistic spatial pattern of study area sowing dates is captured nonetheless.

4.1.2 Model improvement through sequential MODIS DA

Sequential MODIS DA reduces a bias in estimating growing season C fluxes at Bondville (data not shown). This bias is partly a consequence of running SPAC with modelled instead of reported sowing dates: in contrast to overall seasonality, the magnitude of observed LAI and NEE is then less well reproduced. Using simulated sowing dates in model runs is necessary for capturing the overall observed seasonality, whereas sequential DA is necessary to compensate generic model deficiencies and additional biases introduced. However, understanding to what extent this improvement is

caused by an increase in GPP, brought about by an increase in LAI, or a decrease in weed litter mineralization is complicated by the lack of observational constraints. We found clear indications that forward model GPP is an underestimate, and sequential DA brought GPP closer to EC-derived observations through increasing LAI (data not shown). Moreover, as published field data suggest (Buyanovsky et al., 1987), litter mineralization during the growing period is rather small ($\sim 15 \text{ g C m}^{-2} \text{ growing period}^{-1}$ assuming a litter C input of $\sim 200 \text{ g C m}^{-2}$ before sowing) compared to GPP and autotrophic respiration, and thus of second-order importance.

We further acknowledge that our current model–DA scheme is not exploiting the full potential of MODIS VI data for constraining agroecosystem C cycling. One deficiency in our model scheme is the representation of crop senescence. Adopting a simple modelling approach (van Laar et al., 1997), simulated crop senescence appears premature for soybean (Fig. 5), and thus reduces sequential MODIS DA potential during this developmental phase. Consequently, an improved senescence model is necessary that provides a better forward model agreement with observations. We expect further considerable benefits from the assimilation of MODIS data with higher temporal resolution (e.g. MOD09GQ). Moreover, scaling from MODIS RDVI to LAI using an empirical relationship is a considerable source of uncertainty in this study. However, we expect the influence on derived seasonality metrics to be comparably small, as timing of Bondville observed biomass growth and decay are clearly reproduced. Nonetheless, the application of a canopy

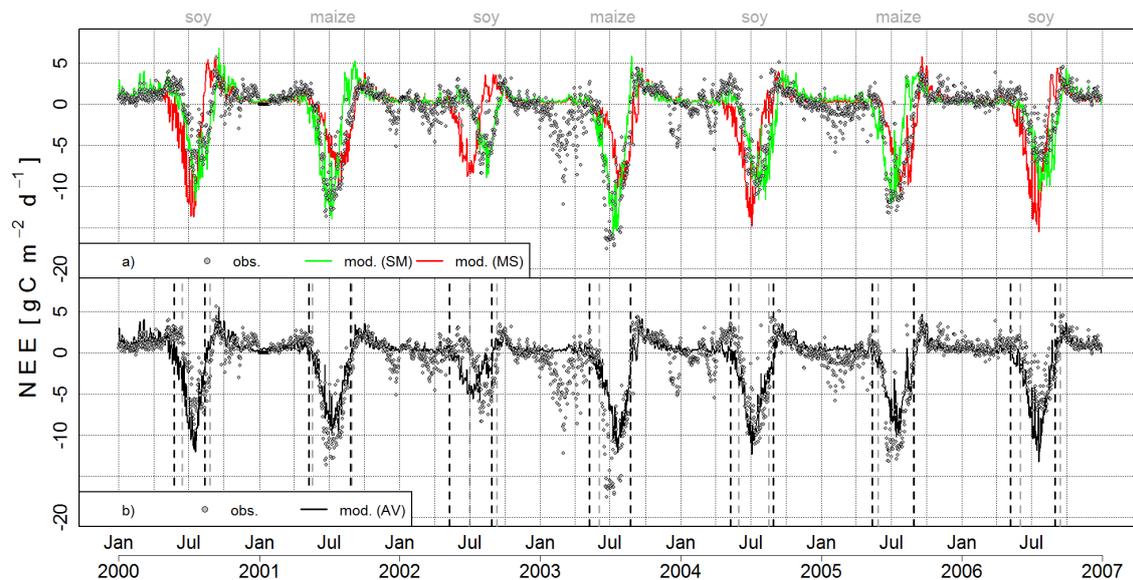


Fig. 7. Study area averages of modelled NEE fluxes (lines) and as observed at Bondville EC site (grey circles). Model data shown are (a) crop rotations with soybean sown in even years (green line, soybean–maize (SM), as Bondville site rotation), soybean sown in odd years (red line, MS), and (b) means over all 104 MODIS pixel locations (black line, average (AV), i.e. not crop type specific). Vertical dashed lines in (b) denote SoS and EoS dates as estimated from model data (black) and observations (grey).

reflectance model could provide improved estimates of C_{stor} (Doraiswamy et al., 2004).

MODIS data allow for spatio-temporal applications of SPAC without a priori knowledge on sowing dates. Certain model deficiencies remain unresolved after DA. SPAC simulations are less reliable for yield than NEE and LAI, and improvements in representing yield formation are warranted for further study. As Dente et al. (2008) show, considerable improvements can be expected through DA. Nonetheless, we are confident that our methodology provides a representative upscaled estimate of agroecosystem C cycling. Our results confirm previous findings that MODIS data contain enough useful information for correcting some deficiencies in global BGCMs, such as a $\sim 40\%$ reduction of RMSE in modelled agricultural C fluxes and improved estimation of GSL (Demarty et al., 2007). The benefits of MODIS DA are particularly considerable during years of abnormal sowing patterns (as observed here in 2002, Fig. 5), and their influence on crop establishment and growth are still rather poorly represented within SPAC. RS DA is also a suitable tool for mitigating uncertainties due to model parameters and weak understanding of phenological processes (Stöckli et al., 2008). There is clear value in making use of MODIS' full spatiotemporal richness when addressing current key uncertainties of upscaled crop modelling. The applicability of our DA approach is limited by the availability of crop type classification data outside of the US. However, MODIS seasonality has been used to provide such classification data in the US (Wardlow and Egbert, 2010) and Ukraine (Becker-Reshef et al., 2010), and is ap-

plicable to all agricultural regions with sufficient ground element (i.e. field patch) sizes given appropriate training data.

4.2 Are point-scale cropland C flux observations spatially representative?

Based on regionally resolved MODIS data, we explored the spatial variation in cropland C fluxes. Our results show that, next to local meteorology, regional patterns of land management are important drivers of agricultural C cycling and major sources of uncertainty if not appropriately accounted for (Tables 1 and 2). Average relative spatial variability for C_{stor} , LAI, and NEE for both crops ranges from $\sim 7\%$ to $\sim 18\%$ per year, which propagates into an NBP variability that is larger than NBP magnitude. Consequently, observing C cycling at one single field with its individual sowing pattern is not sufficient to constrain large-scale agroecosystem behaviour in its total, land management-driven variability.

Bondville data considerably deviate from the pattern of C flux seasonality produced by spatial heterogeneity in cropland management (Fig. 7, Table 2). However, they are comparably representative of the large-scale growing season C budget (Fig. 6). The EC data underestimate upscaled GSL by 20 days. Mean study area NEE fluxes contain crop-rotation “signals” not observed at Bondville. It is plausible that mean GSL increases when cropland NEE is averaged over several field locations: few “premature” sites will decrease mean NEE considerably, as their quickly increasing sink strength is soon larger than the comparably small C losses of fallow sites (and vice versa around maturity). That is partly

why modelled SM rotations show little NEE phase shifts compared to observations (Fig. 7a), even though growing season metrics differ (Table 2). The accurate quantification of GSL is important, as GSL influences microclimatic variables through the longevity of vegetation cover (de Noblet-Ducoudré et al., 2004; Kucharik and Twine, 2007), indicates farmers adopting longer season cultivars in order to maximise yield (Sacks and Kucharik, 2011), and affects seasonal atmospheric CO₂ amplitudes (Keeling et al., 1996) and possibly the terrestrial C balance (Piao et al., 2007).

Differences between Bondville and upscaled NEE are especially large in years with non-optimal climatic sowing conditions (here in 2002, Fig. 7b). When intense spring precipitation delays field work, NASS reports suggest that farmers use at times relatively short time windows with drier weather conditions for crop sowing. This can lead to potentially large differences in maize and soybean sowing progress, as their usual time windows for sowing do not overlap. Cool and wet conditions are favourable for pest development, and potentially affect the timing of major phenological events of rain-fed crops. DAMod data suggest that this was the case in 2002, when maize sowing and establishment was about normal, but soybean progress was negatively affected by strong precipitation. If this modelled difference in NEE seasonality between the two crops were merely an artefact due to weaknesses in MODIS-derived sowing dates for maize, early maize growth and C uptake would have been restricted through the sequential assimilation of low MODIS RDVI data. Instead, model results suggest that the 2002 NEE seasonality shift (Fig. 7) is realistic and supported by empirical evidence (NASS).

As the seasonality and regional variability of agroecosystem C cycling is considerably sensitive to sowing dates, the timing of this land management action needs particular attention in large-scale simulation runs. Models driven with sowing dates that are either static (e.g. soybean within LPJml, Bondeau et al., 2007) or estimated through temperature thresholds (e.g. 10 °C for maize in ORCHIDEE-STICS, de Noblet-Ducoudré et al., 2004) would not be able to reproduce the observed NEE phase shift and its consequence on C budgeting and biometry in 2002. Large-scale applications of cropland BGCs certainly necessitate these simplifications, but associated uncertainties are large and need to be quantified. In contrast to natural ecosystems whose C cycling primarily responds to climatic constraints and disturbances, croplands carry an additional “disturbance” signal provided by human management. Sowing progress is clearly linked to atmospheric variables, but their relationship is poorly quantified and difficult to predict from time series analysis of climatic data alone.

Future large-scale applications such as for the conterminous US could provide a thorough assessment of the current state of agroecosystem C cycling and an improved quantification of the relationship between climate variables and sowing dates. Further, our DA scheme could allow for a detailed assessment of the “yield gap” (i.e. climatic potential yield

– actual yield, Licker et al., 2010), or possibly the “C sequestration gap” (modelled GSL might correlate with terrestrial C uptake, Piao et al., 2007) of major agricultural areas. DA-derived “observed” GSL metrics could be compared with maximum potential GSL to see to what extent farmers exploit the full time period allowing for crop cultivation.

4.3 Are study area croplands a sink or source of carbon?

Our results are a considerable improvement compared to a previous study, where SPAC reproduced daily cropland C fluxes observed at six different European flux tower sites with high accuracy, but growing season cumulative NEE was clearly overestimated (by $\sim 123 \text{ g C m}^{-2} \text{ yr}^{-1}$ on average, Sus et al., 2010; Wattenbach et al., 2010). Here, assuming measured Bondville C budget is spatially representative, observed growing season cumulative NEE is underestimated by ~ 240 to 300 g C m^{-2} after 7 yr (Fig. 6), which corresponds to $\sim 10\%$ of the observed value or an annual underestimation of just 34 to $43 \text{ g C m}^{-2} \text{ yr}^{-1}$. All-season cumulative NEE on the other hand is overestimated and closer to the observed value, with a deviation of now ~ 110 to 220 g C m^{-2} . This shows that fallow season weed grass C uptake is underestimated by SPAC by ~ 350 to 520 g C m^{-2} , as model soil C is in equilibrium. A large fraction of this discrepancy is explained by model–observations differences in 2003, when pre-sowing weed growth was particularly intense. Consequently, the underestimation of cumulative NBP falls within that range ($\sim 420 \text{ g C m}^{-2}$; see green line in Fig. 6c, which is for the same crop rotation as at Bondville).

Hollinger et al. (2004) (but see also Hollinger et al., 2005) assessed the regional C sequestration potential of no-till MSCR based on the Bondville flux data (1997–2002). Their analysis showed that the annual C sink strength is rather small (NBP $\sim 30 \text{ g C m}^{-2} \text{ yr}^{-1}$). Cumulated over the 5 study years for which C_{stor} observations are available (Table 1), this estimate is comparable to the observed NBP data presented here ($\sim 211 \text{ g C m}^{-2}$ after 5 yr). However, their published estimate of cumulative NEE for years 2000–2002 is considerably lower (by $\sim 266 \text{ g C m}^{-2}$) than what we derived from the FLUXNET database. EC studies of other maize and/or soybean agroecosystems indicate C neutrality at best: a maize–fennel crop rotation in Italy was found to lose $\sim 417 \text{ g C m}^{-2} \text{ yr}^{-1}$ under organic manure fertilisation (Kutsch et al., 2010), and rainfed no-till MSCRs were found to be approximately C neutral (Verma et al., 2005) or net sources of C (40 to $80 \text{ g C m}^{-2} \text{ yr}^{-1}$, Grant et al., 2007; Baker and Griffis, 2005).

The observations analysed provide no clear constraints on the sign of NBP, and generic uncertainties in EC data have been documented (Anthoni et al., 2004; Falge et al., 2001). The role of fallow season C sequestration necessitates more detailed analysis. Model data indicate a source of C, but the magnitude of NBP is smaller than its spatial

variability. However, wind-velocity corrections of Bondville EC data were published recently (Kochendorfer et al., 2012), involving an overall increase in flux magnitude of $\sim 11\%$. We expect only minor effects of this correction on our results about observed flux seasonality, but the gap between modelled and observed all-year cumulative NEE will further increase by $\sim 230 \text{ g C m}^{-2}$ (2000–2006). Regarding these uncertainties, we are as yet not able to firmly establish the magnitude and sign of NBP for the MSCR agroecosystems of the study area. However, under the assumption of soil C equilibrium, our model data suggest that the region's no-till croplands ($\sim 40\%$ of US croplands are no-till, Kucharik and Twine, 2007) are unlikely to be sinks of C. Fallow weed grass C uptake might have been responsible for converting the Bondville-observed C balance into a sink.

5 Conclusions

In this study, we developed and tested a model–DA framework for the simulation of spatio-temporally resolved cropland C fluxes for maize–soybean rotations close to the Bondville (IL, US) FLUXNET site. MODIS RDVI data were assimilated both for the batch-calibration of sowing dates and for improved model state estimation using the EnKF. A comparison with NASS land management data showed that MODIS-derived sowing dates appear realistic and thus will probably allow for more accurate simulations of agroecosystem C fluxes at locations for which ground-truth sowing dates are not available. The EnKF analysis provides further information on seasonality of C fluxes and biometry through MODIS constraints on crop establishment and senescence, which is especially important in anomalous years when the forward model performs poorly. This framework is readily applicable for simulating current C cycling of major agricultural regions as long as crop type classification data are available and/or minimum requirements on field patch size are satisfied. However, model results do not allow for firm conclusions on whether the agroecosystems of the study area are net sources or sinks of C. Additional data on land management and belowground C cycling remain mandatory, as such information is not provided by MODIS.

The relative spatial variability of yield, LAI, and NEE for both crops ranges from $\sim 7\%$ to $\sim 18\%$, and is considerably larger for NBP, which is negative but close to C neutrality. These results show that regional patterns of land management are important drivers of agricultural C cycling and major sources of uncertainty if not appropriately accounted for. Observing C cycling at one single field with its individual sowing pattern is not sufficient to constrain large-scale agroecosystem behaviour. Whereas EC data appear representative for a particular crop type rotation, upscaled phenology shows considerable differences due to land-use heterogeneity. Study-area GSL is 20 days longer than observed, primarily because of an earlier estimated SoS. Anomalies in sowing

dates as observed here in 2002 add a particularly strong anthropogenic signal to large-scale C exchange fluxes. The DA scheme developed and tested here appears suitable for accounting for human intervention and its knock-on effects on ecosystem services such as fluxes of C, water, and energy. Our approach is a step forward in improving large-scale applications of BGCMs.

Acknowledgements. We acknowledge support from the Scottish Alliance of Geoscience, Environment and Society (SAGES) and the University of Edinburgh School of Geosciences (Torrance Bequest studentship). Observational data were provided by NOAA/ATDD's SEBN program, USDA-NASS (crop progress reports and data layers), and ORNL DAAC (MODIS subsets).

Edited by: J. Pongratz

References

- Anthoni, P. M., Freibauer, A., Kolle, O., and Schulze, E.-D.: Winter wheat carbon exchange in Thuringia, Germany, *Agr. Forest. Meteorol.*, 121, 55–67, 2004.
- Aubinet, M., Moureaux, C., Bodson, B., Dufranne, D., Suleau, B. H. M., Vancutsem, F., and Vilret, A.: Carbon sequestration by a crop over a 4-year sugar beet/winter wheat/seedpotato/winter wheat rotation cycle, *Agr. Forest. Meteorol.*, 149, 407–418, 2009.
- Baker, J. and Griffis, T.: Examining strategies to improve the carbon balance of corn/soybean agriculture using eddy covariance and mass balance techniques, *Agr. Forest. Meteorol.*, 128, 163–177, 2005.
- Baldocchi, D.: A comparative study of mass and energy exchange rates over a closed C_3 (wheat) and an open C_4 (corn) crop: II. CO_2 exchange and water use efficiency, *Agr. Forest. Meteorol.*, 67, 291–321, 1994.
- Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C.: A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data, *Remote Sens. Environ.*, 114, 1312–1323, 2010.
- Bernacchi, C. J., Hollinger, S. E., and Meyers, T.: The conversion of the corn/soybean ecosystem to no-till agriculture may result in a carbon sink, *Glob. Change Biol.*, 11, 1867–1872, 2005.
- Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Muller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Glob. Change Biol.*, 13, 679–706, 2007.
- Boryan, C., Yang, Z., Mueller, R., and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, *Geocarto Int.*, 26, 341–358, 2011.
- Brown, M. E. and de Beurs, K. M.: Evaluation of multi-sensor semi-arid crop season parameters based on NDVI and rainfall, *Remote Sens. Environ.*, 112, 2261–2271, 2008.
- Buyanovsky, G. A., Kucera, C. L., and Wagner, G. H.: Comparative Analyses of Carbon Dynamics in Native and Cultivated Ecosystems, *Ecology*, 68, 2023–2031, 1987.

- Chang, J., Hansen, M. C., Pittman, K., Carroll, M., DiMiceli, C.: Corn and Soybean Mapping in the United States Using MODIS Time-Series Data Sets, *Agronomy J.*, 99, 1654–1664.
- Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C₄ Plants, *Aust. J. Plant Physiol.*, 19, 519–538, 1992.
- de Noblet-Ducoudré, N., Gervois, S., Ciais, P., Viovy, N., Brisson, N., Seguin, B., and Perrier, A.: Coupling the soil-vegetation-atmosphere-transfer scheme ORCHIDEE to the agronomy model STICS to study the influence of croplands on the European carbon and water budgets, *Agronomie*, 24, 397–407, 2004.
- de Wit, A. and van Diepen, C.: Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts, *Agr. Forest. Meteorol.*, 146, 38–56, 2007.
- Demarty, J., Chevallier, F., Friend, A. D., and Viovy, N.: Assimilation of Global MODIS Leaf Area Index Retrievals Within a Terrestrial Biosphere Model, *Geophys. Res. Lett.*, 34, 1–6, 2007.
- Dente, L., Satalino, G., Mattia, F., and Rinaldi, M.: Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield, *Remote Sens. Environ.*, 112, 1395–1407, 2008.
- Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., and Stern, A.: Crop condition and yield simulations using Landsat and MODIS, *Remote Sens. Environ.*, 92, 548–559, 2004.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., and Clement, R.: Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agr. Forest. Meteorol.*, 107, 43–69, 2001.
- Farquhar, G. D. and von Caemmerer, S.: Modelling of photosynthetic response to the environment, in: *Physiological Plant Ecology II. Encyclopedia of plant physiology, New Series*, edited by Lange, O. L., Nobel, P. S., Osmond, C. B., and Ziegler, H., vol. 12B, 549–587, Springer, Berlin, 1982.
- Foley, J. A., DeFries, R., Asner, G. P., and Barford, C.: Global Consequences of Land Use, *Science*, 309, 570–574, 2005.
- Grant, R. F., Arkebauer, T. J., Dobermann, A., Hubbard, K. G., Schimelfenig, T. T., Suyker, A. E., Verma, S. B., and Walters, D. T.: Net Biome Productivity of Irrigated and Rainfed Maize–Soybean Rotations: Modeling vs. Measurements, *Agron. J.*, 99, 1404–1423, 2007.
- Haboudane, D., Millera, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B.: Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture, *Remote Sens. Environ.*, 90, 337–352, 2004.
- Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiol.*, 25, 873–885, 2005.
- Hollinger, S. E., Bernacchi, C. J., and Meyers, T. P.: Carbon Budget of Mature No-till Ecosystems in North Central Region of the United States, *Agr. Forest. Meteorol.*, 130, 59–69, 2004.
- Hollinger, S. E., Bernacchi, C. J., and Meyers, T. P.: Corrigendum to “Carbon budget of Mature No-till Ecosystem in North Central Region of the United States”, *Agr. Forest. Meteorol.*, 136, 88–89, 2005.
- Inoue, Y. and Oliosio, A.: Estimating the dynamics of ecosystem CO₂ flux and biomass production in agricultural fields on the basis of synergy between process models and remotely sensed signatures, *J. Geophys. Res.-Atmos.*, 111, D24S91, doi:10.1029/2006JD007469, 2006.
- Jones, H. G.: *Plants and Microclimate*, Cambridge University Press, Cambridge, 1992.
- Keeling, C. D., Chin, J. F. S., and Whorf, T. P.: Increased activity of northern vegetation inferred from atmospheric CO₂ measurements, *Nature*, 382, 146–149, doi:10.1038/382146a0, 1996.
- Kochendorfer, J., Meyers, T., Frank, J., Massman, W., and Heuer, M.: How Well Can We Measure the Vertical Wind Speed? Implications for Fluxes of Energy and Mass, *Bound.-Lay. Meteorol.*, 1–16, doi:10.1007/s10546-012-9738-1, 2012.
- Kucharik, C. J. and Twine, T. E.: Residue, Respiration, and Residuals: Evaluation of a Dynamic Agroecosystem Model Using Eddy Flux Measurements and Biometric Data, *Agr. Forest. Meteorol.*, 146, 134–158, 2007.
- Kutsch, W., Aubinet, M., Buchmann, N., Smith, P., Osborne, B., Eugster, W., Wattenbach, M., Schrumpf, M., and Schulze, E.: The net biome production of full crop rotations in Europe, *Agriculture, Ecosyst. Environ.*, 139, 336–345, 2010.
- Lee, H., Fok, L., and Zhang, D.: Climatic change and Chinese population growth dynamics over the last millennium, *Climatic Change*, 88, 131–156, 2008.
- Licker, R., Johnston, M., Foley, J. A., Barford, C., Kucharik, C. J., Monfreda, C., and Ramankutty, N.: Mind the gap: how do climate and agricultural management explain the yield gap of croplands around the world?, *Global Ecol. Biogeogr.*, 19, 769–782, 2010.
- Lobell, D. B. and Field, C. B.: Global scale climate–crop yield relationships and the impacts of recent warming, *Environ. Res. Lett.*, 2, 1, doi:10.1088/1748-9326/2/1/014002, 2007.
- Mathieu, P.-P. and O’Neill, A.: Data assimilation: From photon counts to Earth System forecasts, *Remote Sens. Environ.*, 112, 1258–1267, 2008.
- McGuire, A. D., Sitch, S., Clein, J. S., Dargaville, R., Esser, G., and Foley, J.: Carbon Balance of the Terrestrial Biosphere in the Twentieth Century: Analyses of CO₂, Climate and Land Use Effects with Four Process-based Ecosystem Models, *Global Biogeochem. Cy.*, 15, 183–206, 2001.
- Meyers, T. P. and Hollinger, S. E.: An assessment of storage terms in the surface energy balance of maize and soybean, *Agr. Forest. Meteorol.*, 125, 105–115, 2004.
- Moureaux, C., Debacq, A., Hoyaux, J., Suleau, M., Tourneur, D., Vancutsem, F., Bodson, B., and Aubinet, M.: Carbon Balance Assessment of a Belgian Winter Wheat Crop (*Triticum aestivum* L.), *Glob. Change Biol.*, 14, 1353–1366, 2008.
- ORNL: MODIS subsetted land products, Collection 5. available at ORNL DAAC, Oak Ridge, Tennessee, USA <http://daac.ornl.gov/MODIS/modis.html>, last access 26 November 2010, 2010.
- Penning de Vries, F. W. T., Jansen, D. M., ten Berge, H. F. M., and Bakema, A.: Simulation of Ecophysiological Processes of Growth in Several Annual Crops, Vol. 29 of *Simulation Monographs*, Pudoc, Wageningen, 1989.
- Piao, S., Friedlingstein, P., Ciais, P., Viovy, N., and Demarty, J.: Growing season extension and its impact on terrestrial carbon cycle in the Northern Hemisphere over the past 2 decades, *Global Biogeochem. Cy.*, 21, 1–11, 2007.

- Quaife, T., Lewis, P., Kauwe, M. D., Williams, M., Law, B. E., Disney, M., and Bowyer, P.: Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter, *Remote Sens. Environ.*, 112, 1347–1364, 2008.
- Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., and Xue, Y.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, *Global Change Biology*, 18, 556–584, 2012.
- Roujean, J.-L. and Breon, F.-M.: Estimating PAR absorbed by vegetation from bidirectional reflectance measurements, *Remote Sens. Environ.*, 51, 375–384, 1995.
- Sacks, W. J. and Kucharik, C. J.: Crop management and phenology trends in the US Corn Belt: Impacts on yields, evapotranspiration and energy balance, *Agr. Forest. Meteorol.*, 151, 882–894, 2011.
- Sacks, W. J., Deryng, D., Foley, J. A., and Ramankutty, N.: Crop planting dates: an analysis of global patterns, *Global Ecol. Biogeogr.*, 19, 607–620, 2010.
- Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuaka, N., and Ohno, H.: A Crop Phenology Detection Method Using Time-Series MODIS Data, *Remote Sens. Environ.*, 96, 366–374, 2005.
- Setiyono, T., Weiss, A., Specht, J., Bastidas, A., and Dobermann, K. C. A.: Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions, *Field Crop. Res.*, 100, 257–271, 2007.
- Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., and Feick, S.: Development and validation of the global map of irrigation areas, *Hydrol. Earth Syst. Sci.*, 9, 535–547, doi:10.5194/hess-9-535-2005, 2005.
- Solano, R., Didan, K., Jacobson, A., and Huete, A.: MODIS Vegetation Indices (MOD13) C5 User's Guide, Tech. rep., Terrestrial Biophysics and Remote Sensing Lab, The University of Arizona, Arizona, 2010.
- Stöckli, R., Rutishauser, T., Dragoni, D., O'Keefe, J., Thornton, P. E., Jolly, M., Lu, L., and Denning, A. S.: Remote sensing data assimilation for a prognostic phenology model, *J. Geophys. Res.*, 113, G4, doi:10.1029/2008JG000781, 2008.
- Streck, N. A., Lago, I., Gabriel, L. F., and Samboranza, F. K.: Simulating maize phenology as a function of air temperature with a linear and a nonlinear model, *Pesqui. Agropecu. Bras.*, 43, 449–455, 2008.
- Sus, O.: Simulating the carbon cycling of croplands — model development, diagnosis, and regional application through data assimilation, Ph.D. thesis, School of Geosciences, University of Edinburgh, Edinburgh, UK, 2011.
- Sus, O., Williams, M., Bernhofer, C., Beziat, P., Buchmann, N., Ceschia, E., Doherty, R., Eugster, W., Gruenwald, T., Kutsch, W., Smith, P., and Wattenbach, M.: A linked carbon cycle and crop developmental model: Description and evaluation against measurements of carbon fluxes and carbon stocks at several European agricultural sites, *Agriculture, Ecosyst. Environ.*, 139, 402–418, 2010.
- Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W. H., Simberloff, D., and Swackhamer, D.: Forecasting Agriculturally Driven Global Environmental Change, *Science*, 292, 281–284, 2001.
- USDA: Usual Planting and Harvesting Dates for US Field Crops, Agricultural Handbook No. 628, US Department of Agriculture, Washington, DC, 1997.
- van Laar, H. H., Goudriaan, J., and van Keulen, H.: SUCROS97: Simulation of crop growth for potential and water-limited production situations, in: Quantitative Approaches in Systems Analysis, 14, DLO Research Institute for Agrobiology and Soil Fertility, The C.T. de Wit Graduate School for Production Ecology, Wageningen, 1997.
- Verma, S. B., Dobermann, A., Cassman, K. G., Walters, D. T., Knops, J. M., and Arkebauer, T. J.: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, *Agr. Forest. Meteorol.*, 131, 77–96, 2005.
- Wardlow, B. D. and Egbert, S. L.: Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains, *Remote Sens. Environ.*, 112, 1096–1116, 2008.
- Wardlow, B. D. and Egbert, S. L.: A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for Southwest Kansas, *Int. J. Remote Sens.*, 31, 805–830, 2010.
- Wardlow, B. D., Kastens, J. H., and Egbert, S. L.: Using USDA Crop Progress Data for the Evaluation of Greenup Onset Date Calculated from MODIS 250-Meter Data, *Photogramm. Eng. Rem. Sens.*, 72, 1225–1234, 2006.
- Wattenbach, M., Sus, O., Vuichard, N., Lehuger, S., Gottschalk, P., Li, L., Leip, A., Williams, M., Tomelleri, E., Kutsch, W. L., Buchmann, N., Eugster, W., Dietiker, D., Aubinet, M., Ceschia, E., Béziat, P., Grünwald, T., Hastings, A., Osborne, B., Ciais, P., Cellier, P., and Smith, P.: The carbon balance of European croplands: A cross-site comparison of simulation models, *Agriculture, Ecosyst. Environ.*, 139, 419–453, 2010.
- West, T. O., Brandt, C. C., Baskaran, L. M., Hellwinckel, C. M., Mueller, R., Bernacchi, C. J., Bandaru, V., Yang, B., Wilson, B. S., Marland, G., Nelson, R. G., De La Torre Ugarte, D. G., and Post, W. M.: Cropland carbon fluxes in the United States: increasing geospatial resolution of inventory-based carbon accounting, *Ecological Applications*, 20, 1074–1086, 2010.
- Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L. and Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the Soil-Plant-Atmosphere Continuum in a Quercus-Acer Stand at Harvard Forest: the Regulation of Stomatal Conductance by Light, Nitrogen and Soil/Plant Hydraulic Properties, *Plant, Cell Environ.*, 19, 911–927, 1996.
- Williams, M., Rastetter, E. B., Shaver, G. R., Hobbie, J. E., Carpino, E., and Kwiatkowski, B. L.: Primary production of an Arctic watershed: an uncertainty analysis, *Ecol. Appl.*, 11, 1800–1816, 2001.
- Williams, M., Schwarz, P. A., Law, B. E., and Irvine, J.: An Improved Analysis of Forest Carbon Dynamics Using Data Assimilation, *Glob. Change Biol.*, 11, 89–105, 2005.
- Wilson, T. and Meyers, T.: Determining vegetation indices from solar and photosynthetically active radiation fluxes, *Agr. Forest. Meteorol.*, 144, 160–179, 2007.

Wolfe, R. E., Roy, D. P., and Vermote, E.: MODIS Land Data Storage, Gridding, and Compositing Methodology: Level 2 Grid, *IEEE T. Geosci. Remote*, 36, 1324–1338, 1998.

Zhang, D. D., Lee, H. F., Wang, C., Li, B., Zhang, J., Pei, Q., and Chen, J.: Climate change and large-scale human population collapses in the pre-industrial era, *Global Ecol. Biogeogr.*, 20, 520–531, 2011.