Title
Reducing the impact of model scale on simulated, gridded switchgrass yields

Permalink
https://escholarship.org/uc/item/7dn2r4p4

Author
Vittorio, A.V. Di

Publication Date
2013-07-01

DOI

Peer reviewed
Reducing the impact of model scale on simulated, gridded switchgrass yields

Alan V. Di Vittorio\textsuperscript{a,b,*} and Norman L. Miller\textsuperscript{b,c}

*Corresponding author

\textsuperscript{a}Energy Biosciences Institute, University of California, Berkeley

\textsuperscript{b}Earth Sciences Division, Lawrence Berkeley National Laboratory

\textsuperscript{c}Department of Geography, University of California, Berkeley
Abstract

Results of gridded ecosystem simulations of bioenergy crops are used for estimating economic viability, environmental impacts, and potential land use change. Gridded model uncertainty propagates through these uses, thus we propose a simple method for estimating regional, spatial model error from sparse field data. We apply this method to the Agricultural-BioGeochemical Cycles (Agro-BGC) model to examine and reduce the model uncertainty associated with grid scale for simulated switchgrass yields in a 6° latitude × 5° longitude (~300,000 km²) region covering Illinois, United States of America. Based on three evaluation sites, changes in yield with scale result from complex intra-model interactions driven by a combination of meteorological rather than soil or terrain variables. Spatial bias of the regional mean significantly increases with increasing cell size for 11 of 15 measurement dates. This bias is primarily due to grid scale, thus bias correction of output yield reduces the model uncertainty associated with grid scale. The corresponding Root Mean Squared Error and Bias-Corrected RMSE ($\text{RMSE}_{\text{BC}}$) have effectively negligible trends with inconsistent signs. The range of $\text{RMSE}_{\text{BC}}$ for 2-year Average Mature August Yield (AMAY) is 267 - 285 g C m⁻² across 3- to 3600-arcsec resolution (~90 m to ~100 km) with biases from 9 - 61 g C m⁻². AMAY bias significantly increases with increasing cell size. Spatial bias of the regional mean is relatively consistent for resolutions ≤ 1200 arcsec (~33 km) (AMAY bias < 3%), and larger AMAY biases (4 - 13%) at coarse resolutions indicate poorly characterized spatial heterogeneity. Including the 68%
confidence interval around bias-corrected values, AMAY ranges from 0 to 1116 g C m$^{-2}$ across a 150-arsec grid, which is similar to the range reported for 24 eastern United States field sites. Spatial bias of the regional mean yield can vary across grid resolution by as much as 31% of the observed regional mean and can dramatically affect calculations dependent on the resolution of the estimate. We conclude that grid scale profoundly affects model accuracy such that regional studies must match evaluation and simulation scales and should utilize multi-scale analyses to determine robustness of results.

Keywords

Agro-BGC, bioenergy, Biome-BGC, scale, switchgrass, uncertainty
1. Introduction

Research on the potential productivity of bioenergy crops is expanding significantly in response to government mandates for annual increases in renewable energy production (U.S. Code of Federal Regulations (CFR), 40CFR80, 2008). This includes development of numerical models that utilize different approaches to simulate bioenergy crops. There are three primary reasons for the development of such models and subsequent scenario analysis studies: 1) there are a limited number of growth studies for lignocellulosic crops (Heaton et al., 2004; Lewandowski et al., 2003; McLaughlin and Kszos, 2005), 2) it is likely that projected climate change will affect future crop yields and distributions (Hatfield et al., 2008; Jarvis et al., 2010), and 3) there are concerns about the environmental impacts of bioenergy production (Lankoski and Ollikainen, 2011).

Mechanistic growth models (e.g. Di Vittorio et al., 2010; Miguez et al., 2009; VanLoocke et al., 2010) are of particular interest because their extrapolation across time and space is generally more robust than that of statistical models, which are often limited to the conditions under which they were developed (e.g. Miehle et al., 2009). Other bioenergy crop yield prediction approaches include light-use efficiency methods (Clifton-Brown et al., 2000; Corson et al., 2007; Hastings et al., 2009; Jain et al., 2010; Kiniry et al., 2005), environmentally regulated maximum productivity methods (Adler et al., 2007; Davis et al., 2010), and empirical climate-yield relationships (Jager et al., 2010). A few climate models have also incorporated bioenergy crop parameters to estimate shifts in land-atmosphere
exchange due to land use conversion (e.g. Georgescu et al., 2009). However, all of these
models have been developed and tested at a limited number of sites with insufficient data
for full accuracy assessment. While published model results generally agree with limited
field data, under-characterized model sensitivities and output uncertainties of models
upscaled from point sites to area-based grid cells raise questions about the reliability of
such gridded model results.

Nonetheless, gridded results are increasingly used to estimate environmental
impacts, economic viability, and land use change. Clifton-Brown et al. (2000) interpolated
point simulations to a 1 km grid to estimate miscanthus crop yields across Ireland. Other
models have been applied on regional grids from 2.5-arcmin to 0.5° resolution to make
claims about spatial distributions of productivity (Hastings et al., 2008; Hastings et al.,
2009; Jager et al., 2010; Somerville et al., 2010), and hydrologic impacts (Georgescu et al.,
2009; VanLoocke et al., 2010). Yield results have also been used as input to economic
models to determine viable feedstock prices (Jain et al., 2010). Additionally, efforts are
underway to couple economic and biophysical models to explore the potential effects of
human-environment feedbacks on land use and climate (e.g. Jones et al., 2013). A major
limitation of these studies is that single estimates of model outputs are presented without
corresponding uncertainties or confidence intervals, which gives the impression that these
single outputs are robust and reliable for decision-making.
Granted, sensitivities and output uncertainties of complex models are difficult to estimate, but several methods have been developed to address this challenge. For example, the climate model research community utilizes both single model (via multiple input data sets) and multiple model ensembles to estimate simulated means and uncertainties (Covey et al., 2003; Dai et al., 2004; Gates, 1992; Sanderson and Knutti, 2012). However, model ensembles are not always feasible due to high computational requirements and/or insufficient observational data (Crosetto et al., 2000; Heuvelink, 1998). A common recourse is to apply model evaluation methods using available observational data. Bennett et al. (2013) stress that the choice of evaluation method is context dependent, and have organized a wide range of such methods into five categories that are composed of summaries, point-by-point analyses, pattern preservation, parameter identification, and data transformations. These methods range from simple error statistics to sophisticated parameter reduction techniques and data decomposition. However, most sensitivity analysis methods focus on single parameters rather than groups of co-varying values, such as when entire input data sets are replaced (e.g. Pogson et al., 2012). More important than the methods used is the requirement that data independent of model development and calibration be used for evaluation and that statistics are appropriate to output variables of interest. For example, linear regression and its coefficient of determination are commonly used to evaluate simulated yield or productivity, but these data usually do not meet the
statistical assumption of independence because they are often from a single site and have temporal dependence (Di Vittorio et al., 2010; Marascuilo and Levin, 1983).

Grid-based models pose greater evaluation challenges than point-based models due to the effects of grid scale, point versus area data “support” mismatches, and spatial and temporal extrapolation (Chen and Knutson, 2008). Grid scale is defined by resolution and extent, and in this study we consider only fully tessellated model grids such that resolution refers both to grid cell size and the spacing between grid cell centers. One obvious evaluation challenge is that only remotely sensed data can provide measurements with coverage similar to such model grids, but these data are indirect measurements with associated errors and generally do not have the same resolution as the models being evaluated (e.g. Heinsch et al. 2006; Mao et al. 2012). “Support” is defined here as the size or aggregation of data entities and often does not match between site measurements and gridded model inputs/outputs (Heuvelink, 1998). It is distinct from resolution in its characterization of non-gridded and/or non-tessellated data sets. For example, site measurements such as soil cores and biomass samples have point data support, spatially gridded data support is characterized by grid cell size and shape, and thematic area data support is defined by particular spatial boundaries (e.g. watersheds, counties, nations).

The few model evaluations performed with multiple types (e.g. point and area) or levels (e.g. county and country) of observational data support have shown that the magnitude and direction of changes in output uncertainty are model specific and depend on
the output variables of interest (Miehle et al., 2006; Ogle et al., 2010; Ogle et al., 2006). In most cases, however, evaluation of gridded models is performed with point data support, sometimes using only one site (e.g. Clifton-Brown et al., 2000; Hastings et al., 2008; VanLoocke et al., 2010). These gridded and site data are usually independent of each other, which would ensure random placement of sites within different grid orientations, positions, and resolutions (Di Vittorio and Miller, 2013). Spatial heterogeneity of input variables (e.g. precipitation, soil type/conditions, vegetation parameters) further complicates these mixed support comparisons because it can cause output errors to vary widely among calibration and evaluation sites. For example, Root Mean Squared Error (RMSE) of simulated switchgrass yields varied from 38 to 294 g C m$^{-2}$ among three calibration sites when all three sites were included in the calibration (Di Vittorio et al., 2010). However, assuming the same spatial density of field sites, progressively averaging site data to larger areas progressively decreases the error associated with point versus area data support mismatch, which means that averaged point data more accurately estimates area averages as resolution becomes coarser (Tustison et al., 2001). Additionally, temporal variability associated with historical climate/weather data, which is more easily estimated than spatial variability due to data availability, might be erroneously substituted for total output uncertainty when assessing differences between paired change experiments (VanLoocke et al., 2010).
To address these shortcomings in error and uncertainty analyses we present a relatively simple method to estimate and convey gridded model output errors using sparse point data (Section 2) and demonstrate its utility for quantifying and reducing regional uncertainty associated with grid scale (Section 3). Specifically, we quantify the effects of grid scale on spatial errors of switchgrass yield for three simulated sites (Sections 3.1 and 4.1) and also for the Illinois regional average (Sections 3.2 and 4.2). We determine which inputs drive these effects (Sections 3.1 and 4.1), produce confidence interval maps in addition to maps of bias-corrected estimated values (Sections 3.3 and 4.3), and discuss limitations and further implications of our method (Section 4.4).
2. Materials and methods

Our methodology utilizes a regional average to maintain consistent evaluation data support for separation and evaluation of spatial and temporal model uncertainties associated with grid scale. Furthermore, this method characterizes uncertainty due to grid scale independently of other sources of uncertainty, such as input, parameter, and measurement uncertainties. This study focuses on switchgrass to evaluate a recently developed capacity for multi-resolution ecosystem modeling of C₄ perennial grass as a bioenergy crop (Di Vittorio et al., 2010). After model calibration we perform site-level (i.e. baseline) simulations and 14 different input experiments at 13 grid resolutions (3- to 3600-arcsec; ~90 m to ~100 km) at three sites in Illinois, for a total of 549 simulations. Each simulation includes a spin-up to initialize soil conditions (up to 2800 years), a pre-switchgrass growing period from 1948 to 2001, and a switchgrass period from 2002 to Feb. 2007. The 14 input experiments comprise all possible combinations of the four available scale-dependent input sets having either baseline or scale-specific values for 1) soil depth and texture, 2) daily meteorology, 3) elevation, and 4) initial soil carbon, nitrogen, and water contents. This experimental design explores how model inputs interact at various scales to determine grid scale uncertainties and can provide insight on how spatial and temporal errors interact in a regional context. Due to limited data availability, the regional dependence of our approach, and high computational requirements of multi-
scale simulations, we focus on the state of Illinois (IL) in the United States of America (USA).

2.1. Model

We applied the Agricultural-BioGeochemical Cycles (Agro-BGC) ecosystem model (Di Vittorio et al., 2010) for switchgrass simulation. Agro-BGC is an extension of the Biome-BGC 4.2 model that mechanistically simulates ecosystem water, nitrogen, and carbon cycles (Running and Hunt, 1993; Thornton and Rosenbloom, 2005). White et al. (2000) performed an extensive sensitivity analysis on Biome-BGC Net Primary Productivity (NPP) with respect to vegetation parameters, and the model has been applied successfully to a variety of forest types (Migliavacca et al., 2009; Thornton et al., 2002; Ueyama et al., 2009), urban landscapes (Trusilova and Churkina, 2008), turf grasses (Milesi et al., 2005), agricultural fields (Wang et al., 2005), and heterogeneous landscapes at the regional, continental, and global scales (Jung et al., 2006; Running et al., 2004; Schimel et al., 1997; Turner et al., 2007). Agro-BGC adds agricultural practices, enzyme-driven C₄ photosynthesis, crop phenology, and standing dead matter to Biome-BGC in order to simulate C₄ perennial grasses, such as switchgrass, in managed and unmanaged contexts.

Most Agro-BGC outputs are on a per-area basis, and the grid resolution (determined mainly by data availability) of four spatially varying sets of input determines
the model output resolution. These input sets are 1) daily meteorology, 2) soil depth and
texture, 3) elevation, and 4) initial soil carbon, soil nitrogen, and soil water content
(Section 2.2), where the meteorological variables are daily maximum and minimum
temperature, precipitation, vapor pressure deficit, net downward shortwave radiation, and
daylight length. Furthermore, Agro-BGC requires 54 vegetation-specific parameters for
switchgrass, five of which are unavailable in the literature and must be determined through
calibration (see Appendix A for additional model description). Full details of the
development, optimization, and evaluation of Agro-BGC for switchgrass are presented in
Di Vittorio et al. (2010).

2.2. Sites, grids, and data

Our focus on regional uncertainties, combined with limited data availability and
high computational requirements for multi-scale analyses, restricts our study to a domain
encompassing IL, with coordinates 42°42’N 91°48’W x 36°42’N 86°48’W, and three field
sites (Shabbona, Urbana, and Simpson) with nearly identical management histories
spanning a 6° N-S gradient (Heaton et al. 2008) (Figure 1).

2.2.1. Site-specific data for baseline simulations

We compiled site-level data for these three evaluation sites (Table 1) to perform
site-specific simulations defined here as baseline simulations. Soil depth and texture for
these simulations were from the United States Department of Agriculture (USDA) Natural
Resources Conservation Service (NRCS) web soil survey (Soil Survey Staff accessed, 2009). We limited Agro-BGC’s soil depth to 1.5 m at Shabbona and Urbana because that is the approximate depth of the tile drainage systems. Heaton et al. (2008) provided 
switchgrass yield data for model evaluation at the three field sites that were established in 2002. Each March, starting in 2003, all field plots were mowed and cleared of cut grass. 
From June 2004 to Feb. 2007, a biomass sample consisting of four cuttings was taken from each site up to five times per growing season (June, Aug., Oct., Dec., Feb.). Five samples were also taken from the Urbana site from June 2003 to Feb. 2004. Samples were taken early in the growing season to evaluate the potential for multiple harvests and also taken during winter to evaluate the potential for maximum drying and retranslocation of nutrients. We used the reported sample means (sample size (n) = 4) to evaluate Agro-BGC and also estimate annual average mature switchgrass yields for each sample date by averaging the 2005 and 2006 growing seasons. These mature years were selected because fields take at least three years to mature (McLaughlin and Kszos, 2005). Estimates of wet nitrogen deposition were from the 2007 USA map available from the National Atmospheric Deposition Program (NADP 2009). Daily meteorology (1948 - 2006) and constant elevation input data were provided by our High Resolution Climate Downscaler (Di Vittorio and Miller, 2013).
Figure 1. Three Illinois evaluation sites within the 42°42′N 91°48′W x 36°42′N 86°48′W region bounding the gridded simulations.

2.2.2. Description of model grids

To facilitate multi-scale analyses we performed three sets of gridded simulations:

(1) a full-domain grid at 150-arcsec (~4 km) resolution covering IL, (2) a full-domain grid at 1500-arcsec (~42 km) resolution (minus the partial cells along the southern edge), and (3) 13 individual grid cells for each evaluation site, one each at 3-, 30-, 150-, 300-, 450-, 600-, 900-, 1200-, 1500-, 1800-, 2400-, 3000-, and 3600-arcsec resolutions. For reference, these resolutions have approximate north-south distances of 0.09, 1, 5, 9, 14, 19, 28, 37, 46, 56, 74, 93, and 111 km, respectively, and the shorter east-west distances decrease by 8% from south to north across the domain. Henceforth we present resolutions in arcsec to maintain N-S and E-W consistency and precision. The individual grid cells are subsets of their respective full-domain grids, which are aligned with the northwestern (upper left) corner of the full-domain. As is common with grid-based approaches, within-cell site position is determined by grid extent, orientation, and resolution (Figure 2).
Figure 2. Eight individual grid cells for the Simpson, Illinois evaluation site. These cells demonstrate the effect of grid position on the area represented by a cell. The box sizes are 3-arcsec (~90 m; appears as a dot), 30-, 150-, 300-, 450-, 600-, and 900-arcsec. The bounding box is the 1200-arcsec (~35 km) grid cell. The shading depicts the amount of soil clay (ranging from 15% to 33%) and is for a single layer aggregated from the 11-layer, 30 arcsec CONUS-SOIL data set (Miller & White 1998). White signifies that no data are available.

2.2.3. Input data for gridded simulations

To obtain area-based, gridded Agro-BGC output we first compiled fine-resolution, area-based, gridded input data sets and then aggregated to coarser resolutions using area-weighted averaging. The finest resolutions were 3-arcsec for the meteorological and topographic data and 30-arcsec for the soil depth and texture data. The statistical downscaling algorithm, HRCD, provided daily meteorological and constant elevation data (1948 - 2006) for the gridded simulations at 3-, 30-, and 150-arcsec resolutions and coarser resolution data were obtained from area-weighted averages of these 150-arcsec data (Di Vittorio and Miller, 2013). Gridded soil depth and texture (% sand, % silt, % clay) data were obtained from the 30-arcsec CONterminous United States multilayer Soil characteristics data set (CONUS-SOIL) (Miller and White, 1998). CONUS-SOIL defines 11 layers between the surface and 2.5 m depth, with layer thicknesses ranging from 5 to 50
cm. We calculated Agro-BGC’s single soil layer texture as a thickness-weighted average of layer values and an associated soil depth as the sum of the rock-free thicknesses of each layer. The 3-arcsec soil data were selected via nearest-neighbor interpolation. Soil data for resolutions coarser than 30-arcsec were obtained from area-weighted averages of the 30-arcsec data. The initial soil carbon, nitrogen, and water contents of the gridded and baseline simulations were calculated via model spin-up and subsequent pre-switchgrass growth at each grid resolution and at each site (Section 2.3). All gridded input data set cell boundaries were aligned with the northwestern (upper left) corner of the full-domain (Figure 1).

Some input data, however, were the same for point and gridded model runs. For all simulations we adapted a constant albedo (0.22) from a crop and prairie study (Song, 1999) and a constant annual rate of nitrogen fixation (9x10^-4 kg N m^-2 yr^-1) for North American grassland (Cleveland et al., 1999). The 1948 to 1979 annual atmospheric CO2 concentration data were included with the Biome-BGC 4.2 distribution and global mean annual 1980 to 2006 CO2 data were obtained from the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory website (Conway, 2008).

2.3. Simulation experiments

All simulations followed the same procedure where C3 grassland initialized the soil prior to six years of switchgrass production (see Appendix B for details). Fourteen
simulation experiments for each grid resolution tested the output response of switchgrass yield to grid scale changes using four available scale-dependent input data sets (Table 2). Each input data set had either baseline or grid-specific values, and each experiment implemented one possible combination of baseline and grid-specific input data sets. This experimental design enabled us to isolate and determine the relative influence of each individual data set and the interactions among multiple data sets on output sensitivities to grid scale. We define model uncertainty associated with grid scale as the output yield sensitivity to scale changes of all four input sets at once, as determined by experiment 14 (Table 2). Here grid-specific initial soil conditions refer to the results of C$_3$ grass initialization and growth for a particular grid resolution.

2.4. Multi-scale analysis

Using the results of our experiments we analyze the effects of changing grid resolution on model error to quantify model uncertainty associated with grid scale. First we estimate site-specific and gridded model output errors from sparse point data using common residual statistics. We use residual analysis because it is simple and also the most common form of model evaluation (Bennett et al., 2013), which makes it easier to replicate and to apply the results for further analyses. For a single simulation, errors include the effects of input, parametric, and structural uncertainty, but do not include measurement uncertainty because we assumed the measurements are correct. We use these error analyses
to reduce output uncertainty through bias correction, and to generate maps showing confidence intervals around single value estimates.

We quantify the model uncertainty associated with grid scale as changes in error statistics with changes in input data resolution. We isolate the effects of grid resolution on model error by using fine-resolution gridded input data as a basis for changing only input and output resolution. This allows for evaluation of model output sensitivity to changes in grid resolution of the four scale-dependent inputs. To facilitate our analyses we do not explicitly incorporate parametric, input, or measurement uncertainty. One could, however, incorporate these other uncertainties by applying our method stochastically to distributions of parameters, inputs, and measurements. We report only the spatial error and uncertainty results because we found that separate temporal and spatio-temporal analyses did not add to our understanding of model uncertainty associated with grid scale, likely due to the combination of a short, two-year data set and high inter-annual variability of meteorological inputs (Di Vittorio and Miller, 2013). It is important to note, however, that the temporal sampling error of our regional, mature, two-year yield averages is comparable to our estimated spatial root mean squared error. This suggests that our spatial confidence intervals underestimate the total confidence intervals.
2.4.1. Relationships with scale at individual evaluation sites

We first compared differences at each site between simulated gridded yields and observed yields across input experiments and grid scales. We repeated these comparisons with baseline (site-specific) yields in place of observed yields. These comparisons were performed for each measurement month, individually and for two-year mature averages, during the March 2003 to February 2006 growth cycles. The percent differences in simulated yield \( Y_{\text{mod}} \) (g C m\(^{-2}\)) with respect to observed \( Y_{\text{obs}} \) (g C m\(^{-2}\)) or baseline \( Y_{\text{base}} \) (g C m\(^{-2}\)) yields were calculated to normalize these comparisons as follows:

\[
\text{percent difference}_{\text{obs}(m,e,r)} = \frac{1}{Y_{\text{obs}}(m,e,r)} \left( Y_{\text{mod}(m,e,r)} - Y_{\text{obs}(m)} \right), \tag{1a}
\]

\[
\text{percent difference}_{\text{base}(m,e,r)} = \frac{1}{Y_{\text{base}}(m,e,r)} \left( Y_{\text{mod}(m,e,r)} - Y_{\text{base}(m)} \right), \tag{1b}
\]

where \( m \) is the measurement month (Section 2.2.1), \( e \) is the input experiment (Table 2), and \( r \) is the grid resolution (Section 2.2.2). For each measurement month we also calculated the Standard Error of the mean (SE), Root Mean Squared Error (RMSE) and Bias-Corrected RMSE (RMSE\(_{\text{BC}}\)) of the 2005-2006 simulated yields used to calculate the mature averages (\( n = 2 \) years per month; see section 2.2.1) and compared these statistics across grid resolution. Since RMSE represents the total error deviation of biased model estimates, RMSE\(_{\text{BC}}\) (g C m\(^{-2}\)) represents the error deviation of bias-corrected model estimates and was calculated as:
\( RMSE_{BC} = \sqrt{MSE - bias^2} \), \hspace{1cm} (2) 

where MSE \( (\text{g C m}^{-2})^2 \) is the Mean Squared Error of yield estimates, and bias \( (\text{g C m}^{-2}) \) is the simulated mean minus the observed mean of yield estimates. Linear trends with scale were evaluated by regressing the percent differences and statistics against grid resolution (probability of Type I error \( (\alpha) = 0.05 \)).

We performed several tests on individual input variables across grid scale to determine which ones contributed most to changes in model output. Two different tests were applied to the meteorological data. First, at each grid resolution, we linearly regressed each individual meteorological input variable (see section 2.1) against its baseline values. We then performed F-tests of significant difference between these regression models for all possible pairs of grid resolutions \( (\alpha = 0.05) \) (Marascuilo and Levin, 1983). A significantly different model slope indicates a consistent difference in daily values of one input variable between two resolutions that could be related to changes in output. Second, we calculated five-year means (2002 – 2006, \( V_i \)) of daily, monthly, and yearly meteorological input variables for the baseline and each grid-specific data set to test for differences in average conditions among grid scales:

\[
\bar{V}_{(\text{time, base/r})} = \frac{1}{5} \sum_{y=2002}^{2006} V_{i,y(\text{time, base/r})}, 
\]

(3) 

where \( V_i \) is a particular meteorological variable (Section 2.1), \( y \) is the simulation year, \( time \) is either daily (1-365), monthly (1-12), or yearly (1), and \( base/r \) represents either the
baseline simulation or grid resolution (Section 2.2). The monthly and yearly minimum and maximum temperatures, vapor pressure deficit, and daylight length values were averages of the daily values:

$$V_{i,(y,\text{time},\text{base} / r)} = \frac{1}{nd} \sum_{j=1}^{nd} V_{i,(j,y,\text{base} / r)},$$  \hspace{1cm} (4)$$

while the downward radiation and precipitation were sums of daily values:

$$V_{i,(y,\text{time},\text{base} / r)} = \sum_{j=1}^{nd} V_{i,(j,y,\text{base} / r)},$$  \hspace{1cm} (5)$$

where $y$ is the simulation year and $\text{time}$ is a particular month or year as in eqn. (3), and $nd$ is the number of days in $\text{time}$. We then calculated the minimum, mean, and maximum differences between grid-specific and baseline five-year means:

$$\bar{V}_{i,\text{min}}, \bar{V}_{i,\text{mean}}, \bar{V}_{i,\text{max}} \left( \bar{V}_{i,(\text{time},r)} - \bar{V}_{i,(\text{time},\text{base})} \right).$$  \hspace{1cm} (6)$$

These minimum, mean, and maximum values were linearly regressed against grid resolution at each site ($\alpha = 0.05$). The mean difference in eqn. (6) is the bias of the gridded data with respect to the baseline data (except for the monthly variables because the mean is not weighted by the number of days per month). A large bias indicates that a particular variable mean is different between two resolutions and could be related to changes in output.

Elevation, soil depth, and each soil texture component of individual grid cells were linearly regressed against grid resolution at each site ($\alpha = 0.05$). We did not statistically
analyze relationships between grid resolution and separate initial soil carbon, nitrogen, and water pools because our initial analyses showed that their values were dependent on the other scale-dependent inputs and that as a group they had the least impact on simulated yields across grid scale.

2.4.2. Regional evaluation and uncertainty reduction

We performed spatial output yield error analysis for the fully grid-based experiment (EXP14; Table 2) at all grid resolutions to determine and reduce model uncertainty associated with grid scale. This spatial analysis operates on the three IL site yields for each measurement month to analyze annual values or two-year mature averages (n = 3 sites per month), through calculations of the mean, SE, model bias (simulated mean - observed mean), RMSE, and RMSE$_{BC}$ of simulated yields. Trends in error across scale were evaluated by linearly regressing these statistics against grid resolution ($\alpha = 0.05$).

We applied the resulting spatial error estimates to the full-domain simulations to reduce uncertainty and estimate confidence intervals of switchgrass yields in addition to single values. To reduce output uncertainty associated with grid scale we subtracted the model biases (g C m$^{-2}$) uniformly from the simulated yield grid values ($Y_{mod}$, g C m$^{-2}$) to obtain bias-corrected values ($Y_{BC}$, g C m$^{-2}$):

$$Y_{BC(p,r)} = Y_{mod(p,r)} - bias,$$

(10)
where \( p \) is a single pixel in the grid with resolution \( r \). The bias-corrected 68% confidence interval (68% CI) for each pixel was determined by subtracting (adding) the confidence and sample size adjusted RMSE_{BC} uniformly from (to) the bias-corrected grids:

\[
Y_{68\% CI(p,r)} = Y_{BC(p,r)} \pm 1.3161 \cdot RMSE_{BC} .
\]
3. Results

3.1. Results at individual evaluation sites

Scale-dependent input/output analysis shows that changes in meteorological inputs drive the effects of grid scale on simulated yield, but spatial and temporal variations in soil texture and initial conditions can sometimes (e.g. Simpson) counter the meteorological grid scale effects. Table 3 shows percent changes in simulated yield from baseline values due to grid scale-induced changes in meteorology, interactions between soil depth and texture and soil initial conditions, and soil depth and texture alone. The ranges of percent change include comparisons at all grid scales of selected inputs and for all individual and two-year mature average values at each measurement month. The only site with a significantly increasing trend in yield change with coarsening resolution is Shabbona (Figure 3a), and it is coincident with high output sensitivity to the resolution of meteorological inputs (Table 3).

Although no consistent pattern with respect to grid scale emerged from two different evaluations of individual meteorological inputs, only Shabbona meteorology exhibited significant cross-scale trends in bias between gridded and baseline inputs for all variables except precipitation (Table 4, eqn. 6). The Shabbona meteorological biases at grid cell sizes ≥ 1500-arcsec were dramatically larger than those at finer resolutions, which corresponds well with cross-scale patterns in simulated yield (Figure 3a). Simpson precipitation exhibited the only other significant trend in input bias with respect to grid
scale (slope = 6.01E-06, p-value < 0.05), with increased variability in bias at grid cell sizes \[ \geq 900\text{-arcsec} \], but this did not translate to a trend in simulated yield (Figure 3c). Other metrics did not distinguish additional variables or sites, so their values are not presented in the following summary. All daily and monthly meteorological input variables had significant trends in minimum and maximum differences between gridded and baseline five-year means at all sites (eqn. 6). Almost all cross-scale comparisons of regression model slopes for individual meteorological inputs showed significant differences \( p < 0.05 \), even if slopes differed by only 0.001.

Although gridded soil depth and texture differed from baseline soil depth and texture at all three sites, only a few Urbana and Simpson soil texture inputs had significant trends with grid scale (Table 5). These cross-scale soil trends affected yields only in later years (Table 3), but, in contrast to Shabbona meteorology, did not translate into significant cross-scale trends in yield changes (Figure 3). The trend in soil texture at Simpson was due to the relatively high spatial heterogeneity in soil clay and sand amounts (Figure 2, Table 5). Significant, albeit small, changes in elevation (Table 5) did not have any discernible influence on output yields. Changes in simulated yield with respect to scale-dependent changes in inputs are generally less than simulated regional RMSE and observed regional standard deviation (Figure 3).

Changes in meteorological inputs across scale drive changes in Shabbona simulated yield (Tables 3 and 4, Figure 3a). Differences in simulated yields from baseline and
observation, for all individual and two-year average observations and across all grid
resolutions, remain relatively constant across scale for cell sizes ≤ 900-arcsec. At larger
cell sizes these differences become more variable and from June 2004 forward, for all eight
experiments with grid-specific meteorology, result in significantly increasing differences
from baseline (18 of 23 dates and 2-year averages) and observation (16 of 21 dates and 2-
year averages) with increasing grid cell size (p < 0.05). For example, figure 3a shows the
fully grid-based (EXP14, Table 2) Average Mature August Yield (AMAY) percent
differences at Shabbona compared to regional RMSE_{BC} of simulated AMAY and regional
standard deviation of observed AMAY.

The Urbana simulated yields are influenced nearly exclusively by meteorological
inputs (Table 3). Differences in simulated yields from baseline and observation are less
consistent across scale than for Shabbona, but generally are larger at cell sizes ≥ 1500-
arcsec than at smaller cell sizes. No significant trends in yield differences with scale are
present in the Urbana results, and changes in yield differences across scale are generally
less than the simulated regional RMSE and observed regional standard deviation. For
example, figure 3b shows EXP14 AMAY percent differences at Urbana.

The Simpson simulated yields are influenced primarily by meteorological inputs
(Table 3) from June 2003 to August 2005, and from September 2005 until Feb. 2007 soil
depth and texture and their interactions with soil initial conditions become more important.
Increases in yields due to changing soil parameters tend to compensate for decreases in
yields due to changing meteorology (Table 5). There are no discernible trends in yield

differences with scale for any inputs. In general, the differences in simulated yields from
baseline and observation are relatively consistent across scale, although some EXP14
individual dates have variations across scale as high as 11%. For example, figure 3c shows
no trends with scale in EXP14 AMAY at Simpson.
Figure 3. Cross-scale comparison of simulated two-year average August switchgrass yields (2005 - 2006) and estimated and observed regional errors of mature yields (n = 6) for three Illinois sites (Figure 1). Simulated yields are presented as percent differences from respective baseline or observed site yields and RMSE and observed standard deviation are presented as the percent of observed site yield. a) Shabbona: baseline = 317 g C m\(^{-2}\), % difference from baseline slope = 0.008 (p < 0.05), observed = 358 g C m\(^{-2}\), % difference from observed slope = 0.007 (p < 0.05); b) Urbana: baseline = 480 g C m\(^{-2}\), observed = 786 g C m\(^{-2}\), non-significant slopes; c) Simpson: baseline = 616 g C m\(^{-2}\), observed = 301 g C m\(^{-2}\), non-significant slopes. RMSE = Root Mean Square Error.

3.2. Results for regional evaluation and uncertainty reduction

Results from the fully grid-based input configuration (EXP14, Table 2) provide regional yield error estimates for the full-domain simulations. Regional spatial bias, RMSE, and RMSE\(_{BC}\) of simulated yields for Feb. 2005 to Feb. 2007 (including the mature two-year averages) generally follow the same scale pattern as for AMAY (Figure 4). Biases increase significantly with increasing cell size (p < 0.05) for Feb. 2005 and from Aug. 2005 to Feb. 2007 (including the mature two-year averages). For all significant trends, the bias shows marked increases at 1500-arcsec, and June, Oct., and Dec. 2004 have large negative biases for all grid resolutions. RMSE decreases significantly with increasing cell size (p < 0.05) for AMAY, but increases significantly (p < 0.05) for Dec.
2006, Feb. 2007, and for the Dec. and Feb. two-year averages. RMSE$_{BC}$ decreases significantly with increasing cell size ($p < 0.05$) for AMAY, for Oct. and Dec. 2005, and for the Dec. and Feb. two-year averages, but increases significantly ($p < 0.05$) for Feb. and Dec. 2006. While the bias consistently increases with increasing cell size, alternating signs for changes in RMSE and RMSE$_{BC}$ indicate that these two statistics effectively do not have trends with scale. RMSE and RMSE$_{BC}$ are closely related, and henceforth we will refer to them together as RMSE$_{(BC)}$ where appropriate.

The changes in error statistics across scale quantify the model uncertainty associated with grid scale. Spatial biases of simulated yield generally increase with increasing cell size, with significant trends driven by a combination of the Shabbona trends and Urbana patterns (Figures 3 and 4). Variations in bias across grid resolution are smaller than simulation RMSE and observed regional standard deviation, but range from 11% (Aug.) to 14% (Feb.) of the observed regional mean yield for the two-year averages, and from 4% (June 2004) to 31% (Feb. 2007) for individual measurement dates. For 9 of 15 individual dates and two-year averages, the regional bias is larger than the standard error of the observed regional mean at all grid scales. This standard error ranges from 14% to 32% and from 14% to 80% of the observed regional mean for two-year averages and individual dates, respectively. The standard error of the simulated regional mean can be larger, smaller, or comparable to that of the observed regional mean, and in most cases decreases slightly, but significantly, with increasing cell size ($p < 0.05$). Variations across grid
resolution in RMSE (2% to 45% of observed standard deviation) and RMSE\textsubscript{BC} (1% to 38% of observed standard deviation) can be high even though they do not have consistent patterns across resolution. For example, at cell sizes \( \leq 1200\text{-arcsec} \) the biases are relatively low and the maximum variations across these cell sizes in RMSE (~0% to 6% of observed standard deviation) and RMSE\textsubscript{BC} (~1% to 10% of observed standard deviation) are much lower than when all resolutions are included.
Figure 4. Cross-scale comparison of estimated spatial errors and variability of observed and simulated two-year (2005–2006) Average Mature August Yield (AMAY) of Illinois switchgrass (n = 3). The bias slope is 0.0102 (p < 0.05), the Root Mean Square Error (RMSE) slope is -0.00183 (p < 0.05), and the Bias-Corrected RMSE (RMSE$_{BC}$) slope is -0.00291 (p < 0.05). The observed regional mean yield is 481.66 g C m$^{-2}$.

3.3. Confidence intervals for gridded switchgrass yield estimates

The model adequately captures the spatial variability of potential yield in this region (Table 6). Simulated, bias-corrected, gridded yields are consistent with minimum and mean values from published field trials, but are lower than maximum values from these same field trials (McLaughlin and Kszos, 2005). However, including the 68% confidence interval allows the simulated regional maximum to match well with field trials. Figures 5 and 6 show maps of estimated AMAY ranges for the full-domain 150-arcsec and 1500-arcsec resolution simulations, respectively. The bias-corrected, regional AMAY based on all whole grid cells is 10% lower for the 1500-arcsec simulation (490 g C m$^{-2}$) than for the 150-arcsec simulation (513 g C m$^{-2}$). The 68% confidence intervals are 370 g C m$^{-2}$ and 362 g C m$^{-2}$ for 150- and 1500-arcsec, respectively, which are 77% and 75% of the observed regional mean based on three sites (482 g C m$^{-2}$).
Figure 5. Bias-corrected 68% confidence interval (CI) of simulated, two-year (2005-2006) Average Mature August Yield (AMAY) of Illinois switchgrass at 150-arcsec resolution (~4 km), corrected for spatial bias to reduce uncertainty (bias = 14 g C m$^{-2}$, bias corrected 68% CI = ±370 g C m$^{-2}$). a) Lower envelope of bias-corrected 68% CI (image range: 0 - 376 g C m$^{-2}$); b) Bias-corrected values are based on averages of the direct output (image range: 218 - 746 g C m$^{-2}$); c) Upper envelope of bias-corrected 68% CI (image range: 588 - 1116 g C m$^{-2}$).
Figure 6. Bias-corrected 68% Confidence Interval (CI) of simulated, two-year (2005 - 2006) Average Mature August Yield (AMAY) of Illinois switchgrass at 1500-arcsec resolution (~42 km), corrected for spatial bias to reduce uncertainty (bias = 37 g C m\(^{-2}\), bias-corrected 68% CI = ±362 g C m\(^{-2}\)). a) Lower envelope of bias-corrected 68% CI (image range: 0 - 318 g C m\(^{-2}\)); b) Bias corrected values are based on averages of the direct output (image range: 252 - 680 g C m\(^{-2}\)); c) Upper envelope of bias-corrected 68% CI (image range: 614 - 1042 g C m\(^{-2}\)). The black area along the bottom represents un-calculated partial cells.
4. Discussion

4.1. Multi-scale analysis at individual evaluation sites

The sensitivity of output yield to changes in input resolution demonstrates that uncertainty associated with scale is driven primarily by concurrent changes in several meteorological inputs (Tables 3 and 4, Figure 3). In a similar type of study, Pogson et al. (2012) drove a miscanthus growth model with different meteorological and soil data sets, and also found that changes in several input variables, rather than individual ones, caused variations in simulated yields.

The only significant trends in yield difference with scale occur at Shabbona, where the effects of meteorology are most pronounced and the effects of soil are negligible. At Urbana the changes in yield due to changes in meteorology are considerable, and generally increase with increasing cell size, but not with significant trends. At Simpson the changes in yield are consistent across scale, with differences from baseline generally being much smaller than differences from observation. These three sites span nearly the entire possible range of results from significant scale effects to no change at all due to changes only in input data grid resolution. Thus, assuming that any individual grid cell will randomly exhibit scale effects within this possible range, using these sites for regional analyses will give robust results.

The dramatic shift from meteorological to soil sensitivity of simulated Simpson yields (Table 5) was driven by large differences in soil depth (62%) and soil sand amount
There were corresponding differences in average gridded soil silt (-15%) and clay (-18%) amounts. Urbana also showed slight sensitivity to soil input data at later dates (Table 5) with an average gridded soil sand amount 55% higher than baseline, but silt and clay differences were less than 10% and the difference in depth was negligible. In contrast, Shabbona yields showed no sensitivity to soil input data even with a 157% increase with respect to baseline in average gridded soil sand amount, 15% decreases in corresponding soil silt and clay amounts, and a negligible difference in depth. Soil depth and texture determines soil water conductivity and maximum availability, and likely has more of an influence when more vegetation is present in later years because of interactions between higher plant water demand, precipitation, and infiltration of water into the soil. The inter-site variability in yield sensitivity to baseline versus gridded soil input data, combined with consistent yield sensitivities to changes in meteorological inputs, support the hypothesis that complex intra-model interactions dictate the scale-dependent meteorological influences on output yield.

4.2. Multi-scale analysis of regional evaluation and uncertainty reduction

Figures 3 and 4 show that variations across scale in regional bias and RMSE$_{(BC)}$ are driven primarily by the interaction of model accuracy among sites for any given time and resolution. Thus, we can infer from the site analyses that the regional bias trends with scale are driven by changes in meteorology with changes in scale. These bias trends represent
uncertainty associated grid scale that can be reduced via bias correction for a particular
grid. Significant trends occur at later dates most likely because the model parameters were
optimized using mature vegetation data (Di Vittorio et al., 2010), but this is difficult to
confirm with such short temporal data sets. Nonetheless, scale-dependent bias estimates
can be used to reduce output uncertainty, regardless of trend significance, because bias is a
systematic error. The significance of bias trends mainly indicates that estimating these
biases is important for producing reliable outputs. Based on our analyses, Illinois
switchgrass yield estimates and associated regional spatial errors are relatively stable at
cell sizes $\leq 1200$-arcsec, indicating that coarser resolution grids are not appropriate for this
region and its degree of heterogeneity. These results highlight the importance of multi-
scale analyses in determining appropriate grid scales in addition to reinforcing the need to
match model evaluation and model experiment grid scales.

The result that significant trends with grid scale exist for bias but not for RMSE$_{(BC)}$
suggests that coarser resolution grids do not adequately represent the area average of
individual grid cells (Figure 4). If the changes in uncertainty with scale were due primarily
to an increasing mismatch between cell area and point observations, one might expect a
systematic increase in regional RMSE$_{(BC)}$ with increasing cell size and a more consistent
bias across scale. The expectation of increasing error comes from the decreasing likelihood
that an area average represents a point randomly located in that area. The expectation of
consistent bias comes from the assumption that the regional mean should remain relatively
constant whether calculated from point data or area averaged data, unless high spatial
heterogeneity is not adequately represented. In this case, RMSE_{BC} changes inconsistently
with scale and output yield has different patterns with scale among the three sites (Figure
3). This variation in yield pattern among sites occurs because coarse resolution inputs
poorly represent the spatial heterogeneity of meteorology surrounding Shabbona in
comparison with relatively less heterogeneity surrounding the other sites. Thus, the
resulting variations across scale in regional model bias are also due to inadequate
representation of high spatial heterogeneity by coarse inputs rather than a mismatched
comparison.

4.3. Robustness of gridded switchgrass yield estimates

Using bias-corrected outputs and corresponding 68% confidence intervals for
gridded yield estimates gives results with low bias at resolutions finer than or equal to
1200-arcsec that are consistent with field trials (Table 6, Figures 3 and 4). The estimated
regional minimum and mean AMAY are comparable to the minimum and mean multi-year
average site yields reported by McLaughlin and Kszos (2005) for 24 sites across fourteen
US states east of the Rocky Mountains. The estimated regional maximum AMAY is
expectedly lower than the reported maximum because 71% of the reported sites are south
of IL and several studies include higher-productivity lowland varieties in contrast to the
model parameterization of upland switchgrass. However, including the 68% confidence
interval provides a maximum estimate very similar to the maximum reported site average
The lowest estimates occur in the southeast corner of the full-domain and cannot be explained by any of the individual scale-independent inputs, while low estimates in the northern part of the domain appear to be related to relatively high soil sand and silt concentrations (data not shown).

Comparison of 150- and 1500-arcsec AMAY gridded estimates shows discrepancies large enough to significantly affect accuracies of further calculations (Figures 5 and 6, Table 6), such as conversion to ethanol or electricity (e.g. Adler et al., 2007; Campbell et al., 2009) and break-even farm-gate prices (e.g. Jain et al., 2010).

While uncorrected regional mean yield estimates are identical, the bias-corrected, 1500-arcsec estimate is 4% lower than the bias-corrected, 150 arcsec estimate, due to the effects of scale on spatial bias. The 1500-arcsec and 150-arcsec biases are 8% and 3% of the observed regional AMAY, respectively, and the 1500-arcsec and 150-arcsec RMSE<sub>BC</sub> are 4% and 6% greater than the observed regional standard deviation, respectively. An 8% bias would pass directly through to an 8% difference in estimated ethanol or electricity output (Adler et al., 2007), and to an approximately 6% difference in estimated break-even farm gate price (Jain et al., 2010). In this case, the 150- and 1500-arcsec regional mean estimates are not drastically different from each other, but the biases are, and swapping these biases would not provide reliable gridded yield estimates. Furthermore, the estimated harvest date will also influence the impact of grid scale on error. The biases for the February mature average in this study vary across grid scale from 21% to 35% of the

...
observed regional mean and those for the February 2007 measurement range from 7% to
38% (Section 3.2). Thus, using bias-corrected output can considerably improve accuracy,
but only if the bias calculation is characteristic for the region, output of interest, and grid
scale.

Our results suggest that the 150-arcsec resolution is more appropriate than 1500-
arсsec for simulating IL switchgrass AMAY with Agro-BGC, considering its relatively
low bias (Table 6) and the consistency of error among cell sizes ≤ 1200-arcsec (Figure 4).
However, the bias-corrected, 150-arcsec gridded regional mean estimate is 6% higher than
the observed three-site regional mean, while the 1500-arcsec estimate is only 2% higher
(Table 6). This evidence seems to counter the argument for 150-arcsec as a more
appropriate resolution, but it results from the high positive model bias at 1500-arcsec that
forces a large decrease to the model output. This large bias adjustment could cause equally
large overcompensation for individual grid cells, thus, the low 150-arcsec bias is more
desirable. Furthermore, three sites do not adequately represent the regional mean, and
additional sites (and longer time periods for the averages), if available, will improve
estimates of regional error and more clearly demonstrate the effects of grid scale on model
outputs.

4.4. Caveats and implications for gridded simulations
We have presented a simple method for estimating regional uncertainties of gridded
simulations, but have not explored all the possible sources of error in Agro-BGC.
Sensitivities to uncertainties in meteorological inputs are difficult to isolate, so we reduce the effects of these uncertainties by estimating multi-year average yields and assuming that these input uncertainties are randomly distributed in time at any particular location. Furthermore, Agro-BGC is sensitive to vegetation parameters in addition to the four scale-dependent input sets evaluated in this study, which are 1) daily meteorology, 2) soil depth and texture, 3) elevation, and 4) initial soil carbon, nitrogen, and water contents. Vegetation parameters could also be scale dependent because they are highly heterogeneous in space and time with dependencies on environmental conditions as well as genetics and morphology. But this heterogeneity and the associated limits required for sensitivity analysis are relatively unknown due to the limited number of studies that have measured model-relevant parameters (Di Vittorio et al., 2010).

The focus of this study is the effect of grid scale on output yield, and as such we use self-consistent multi-scale inputs, hold vegetation parameters constant, and do not include estimations of measurement uncertainty. Nonetheless, using multiple sites to estimate regional uncertainty incorporates some of the error associated with spatially and temporally heterogeneous inputs such as meteorology, vegetation parameters, soil, and terrain. This input error manifests as variability in multi-site regional error, which enables a more comprehensive uncertainty estimate than one determined at a single site. The representativeness of a regional uncertainty estimate increases with an increasing number of sites because the “representativeness error” of using points to estimate a spatial value
decreases (Tustison et al., 2001). We hold this representativeness error constant by holding the number of evaluation sites constant as we change scale. With respect to our inputs, Tustison et al. (2001) also show that increasing the aggregation area also decreases this representativeness error (assuming the same density of point data is available), which means that each aggregated input represents area averages better as resolution becomes coarser. All of our scale-dependent inputs at cell sizes $\geq 300$-arcsec were averaged from finer resolution data to take advantage of this error reduction and also to minimize influences of particular data sets on results. We have thus isolated the effects of scale on results by making comparisons using the same input data. HRCD did, however, generate separate inputs at 3-, 30-, and 150-arcsec, but HRCD values at these resolutions were not significantly different for 17 weather stations across the United States. Furthermore, HRCD outputs do not degrade in accuracy when averaged from 150-arcsec up to 1 degree (Di Vittorio and Miller, 2013).

While gridded data increase in accuracy through aggregation, we have shown that yield outputs gain bias with aggregation of meteorological inputs. This supports the hypothesis that the coarse resolution model response to a combination of average inputs is not the same as an average of fine resolution model outputs because of non-linear interactions among model components. This hypothesis is analogous to reported differences between non-linear model outputs generated by two cases: interpolate or
average first then calculate, or calculate first then interpolate or average (e.g. Addiscott and Tuck, 2001; Leterme et al., 2007).

Thus, our regional error estimates are driven by model response to changes in input data grid scale, not by a changing relationship with observational data, and show bias increasing as resolution becomes coarser. This is in contrast to a modeling study that estimated total United States (U.S.) soil carbon and reported relative uncertainties driven by increasing cumulative carbon values with increasing area, resulting in higher relative uncertainty regionally than nationally (Ogle et al., 2010). On the other hand, our results show the limitations of having only three available evaluation sites, which is consistent with the general conclusion of Ogle et al. (2010) that additional sites are needed to improve spatial modeling results.

This discussion highlights the complexity of scale and the importance of consistency among input, parameterization, evaluation, and output data when addressing scale questions. More consistent with our results, a study estimating annual changes in U.S. soil carbon reported high national-scale and low regional-scale absolute differences between estimates based on national and regional parametric sensitivity analyses (Ogle et al., 2006). However, the same study reported narrower confidence intervals at the national-scale than at the regional-scale, while we found few significant differences in RMSE.(BC) across grid resolution. This discrepancy is likely due to the dramatic differences between the two studies’ analyses (regional error and uncertainty vs. parametric sensitivity
analysis), scale and support (finer than 1° grids vs. regional and national boundaries), model types (mechanistic vs. accounting), scaled inputs (meteorology and soil vs. land practices), and outputs analyzed (per area vs. change in aggregate total). Nonetheless, Ogle et al. (2006) reached a conclusion that scales of model parameterization and interpretation must match, which is consistent with our conclusion that scales of error and uncertainty analyses and estimation must match. In contrast to all three of these studies discussed here, other methodological (e.g. data sources) or geographic (e.g. size of study area) differences contribute to a lack of scale-dependency for model estimates of tree plantations in Australia (Miehle et al., 2006).

5. Conclusion

Considerable variations in model bias and RMSE$_{(BC)}$ with grid scale indicate that grid scales of model evaluation and experiment must match. Through careful attention to scale, data grouping, inputs, and outputs we have shown that Agro-BGC errors in simulated IL switchgrass yields vary across sites and generally increase with increasing grid cell size. Significant trends with grid scale for yield differences and spatial biases are driven by complex interactions among meteorological inputs. Our results show that simultaneous increases of multi-year, gridded-versus-baseline biases of several meteorological variables (all except precipitation, in this case) with increases in grid cell size have the largest impact on simulated yields. Significant trends with scale for regional
spatial RMSE_{BC} are inconsistent with respect to sign and are effectively negligible. Soil texture and initial conditions can also influence gridded output yield when gridded inputs are considerably different from site-based inputs. In this study, regional spatial error is relatively consistent for grid cell sizes ≤ 1200 arcsec. Changes in bias and estimated yield across scale are less than estimated regional RMSE_{BC}, but cross-scale changes in regional spatial bias for a particular yield estimate can be as high as 31% of the observed regional mean yield. This cross-scale variation in bias quantifies model uncertainty associated with grid scale, which can be reduced via bias correction. Increases in bias without increases in RMSE_{BC} indicate that coarser resolutions do not adequately capture spatial heterogeneity. Riley et al. (2009) reached a similar conclusion that coarse-resolution model estimates of regional land-atmosphere CO₂ exchange incurred biases with respect to higher-resolution estimates because of inadequate representation of high spatial heterogeneity. Regional spatial bias is greater than the standard error of the observed regional mean yield for most measurement dates.

A 6° x 5°, 150-arcsec resolution simulation covering IL adequately captures spatial variation in yield across the region and is consistent with a wide range of field estimates. Based on consistency of error and relatively low bias, cell sizes ≤ 1200-arcsec are more appropriate than those ≥ 1500-arcsec for simulating switchgrass in IL with Agro-BGC. Single-scale studies (e.g. Georgescu et al., 2008; Hastings et al., 2009; VanLoocke et al., 2010) can perform our scale-appropriate error analysis to reduce output uncertainty, but
they lack the ability to evaluate the representativeness of their respective grids. Applying our simple analysis method to simulations at multiple resolutions has successfully characterized and reduced model uncertainty associated with grid scale and demonstrated that uncorrected estimates can have potentially significant effects on further calculations of conversion to ethanol or electricity (e.g. Adler et al., 2007; Campbell et al., 2009) and break-even farm-gate prices (e.g. Jain et al., 2010). Our results indicate that scale effects are driven by complex intra-model interactions and show that strictly site-based evaluations are not appropriate for gridded results. This suggests that, at the very least, single-scale studies must match model evaluation and estimation scales, and multi-scale analyses are crucial for determining the robustness of results.
Acknowledgments

The Energy Biosciences Institute funded model development and data analysis under grant EBI07-J120. Support for project guidance and manuscript completion was provided by the Director, Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under contract No. DE-AC02-05CH11231 as part of their Integrated Assessments Program. The authors are grateful to Emily Heaton and Frank Dohleman for providing switchgrass data from the Illinois sites. We also thank Jeffrey Chambers and our reviewers for their very helpful suggestions and insights.
Appendix A. Agro-BGC model description

Agro-BGC is a mechanistic model that adds enzyme-driven C₄ photosynthesis, agricultural practices, crop phenology, and standing dead matter to Biome-BGC. These additions facilitate simulation of C₄ perennial grass, such as switchgrass, in managed and unmanaged contexts. All model photosynthesis is adapted from Farquhar et al. (1980) The C₄ algorithm uses a phosphoenolpyruvate carboxylase model adapted from Chen et al. (1994) to calculate bundle sheath CO₂ and O₂ concentrations for use in adapted Farquhar equations. The agricultural practices are prescribed at specific dates and include harvest, irrigation, and fertilization. Harvest is simulated by removing all above ground standing biomass from the system. Irrigation is simulated by saturating the soil to field capacity, but the model is not sensitive to one-time additions (as implemented in this study) because there is very little foliage and growing potential at the time of irrigation. Fertilization is simulated as the addition of mineral nitrogen to the soil, but the model is not sensitive to one-time additions (as implemented in this study) because the denitrification parameterization rapidly transfers unused mineral nitrogen from the soil to the atmosphere. Crop phenology is based on growing degree days and triggers perennial grass stages including stem elongation, senescence, flowering, and retranslocation. Standing dead matter exists in all natural grasslands, and is important in the context of bioenergy crop modeling because in practice dead grasses are often left standing in the field to dry before harvest. The canopy has a two-big-leaf structure representing sun and shade.
Evapotranspiration is calculated by a Penman-Monteith formulation and stomatal conductance is determined by reducing maximum values by water, temperature, radiation, and humidity factors. The soil is a single layer with uniform properties that does not uptake water beyond its saturation capacity; the excess water is runoff out of the system.

Following Di Vittorio et al. (2010), five switchgrass vegetation parameters were re-optimized using site-specific input data and available mature yield data from Rockspring, Pennsylvania (Adler et al., 2006) because the Magnus-Tetens formula constants (for relating water vapor pressure to temperature) in Agro-BGC and HRCD have been corrected (Buck, 1981; Campbell and Norman, 1998). These five parameters—annual coarse root turnover rate (4.08x10^{-1} root fraction yr^{-1}), fine root carbon to leaf carbon allocation ratio (3.36x10^{-1}), stem carbon to leaf carbon allocation ratio (3.14x10^{-1}), coarse root carbon to stem carbon allocation ratio (1.49), and proportion of new carbon allocated to current growth (2.22x10^{-1})—were originally optimized because values were not available in the literature. We used the values from Di Vittorio et al. (2010) supplementary material for the remaining parameters.
Appendix B. Agro-BGC simulations

All simulations followed the same basic procedure in which C$_3$ grassland initialized the soil prior to six years of switchgrass production (Di Vittorio et al., 2010). We used the C$_3$ grass vegetation parameters of White et al. (2000) with an annual, year-long phenology cycle. The 1948 to 2006 daily HRCD outputs were set as repeating meteorological forcings over the long initialization runs, and were applied chronologically from 1948 to 2006 for the C$_3$ grass growth and switchgrass simulations with the 1948 data recurring in 2007 to complete the final litter fall season. The initialization stage used a constant, 1948, global average atmospheric CO$_2$ concentration (310.344 ppm), and each input configuration and grid resolution was initialized independently. Due to differences in inputs among these initializations, the length of time needed for the long-term average of soil carbon to reach a quasi-steady state varied from about 900 to 2800 years. The grassland growth stage ran from 1948 through 2001 and the switchgrass stages from 2002 through 2007, all with annually increasing CO$_2$ concentration (1.23 ppm average annual increase). At the beginning of 2002 the C$_3$ grassland was cleared of all roots, litter, and above ground vegetation (live and dead), leaving the soil carbon, nitrogen, and water pools unchanged. These soil carbon, nitrogen, and water concentrations at the beginning of 2002 comprised the fourth set of scale-dependent inputs in this study. Upon clearing, a very small switchgrass leaf was added to the ecosystem to emulate seeding. The leaf remained dormant until the onset of plant emergence. Switchgrass phenology was based on critical
soil temperature for emergence (12 °C), growing degree days for delineating growth and senescence stages, and a meteorologically determined killing frost date for end of growing season. The baseline and individual grid cell simulations incorporated site-specific management practices as described by Heaton et al. (2008) including nitrogen fertilization in July 2003, annual harvests from 2003 to 2007 (implemented on 1 March), and two irrigations at Urbana in 2002. However, the model is not sensitive to these one-time additions of water and nitrogen and so the full-domain simulations included annual March harvests, but no fertilization or irrigation. The model is not sensitive to nitrogen deposition, so the full-domain simulations used a constant annual nitrogen deposition value ($6 \times 10^{-4} \text{ kg N m}^{-2} \text{ yr}^{-1}$) for all areas, but the baseline and individual grid cell simulations used the IL site estimates for consistency in multi-scale analyses.
859 References


Heinsch, F.A., Maosheng, Z., Running, S.W., et al., 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network


simulation by the ALMANAC model at diverse sites in the southern US. Biomass & Bioenergy 29, 419-425.


statistical regression model to predict growth of Eucalyptus globulus plantations.

Ecological Modelling 220, 734-746.


related to the sensitivity of heterotrophic respiration. Agricultural and Forest Meteorology 149, 582-602.


Figure Captions

Figure 1. Three Illinois evaluation sites within the 42°42’N 91°48’W x 36°42’N 86°48’W region bounding the gridded simulations.

Figure 2. Eight individual grid cells for the Simpson, Illinois evaluation site. These cells demonstrate the effect of grid position on the area represented by a cell. The box sizes are 3-arcsec (~90 m; appears as a dot), 30-, 150-, 300-, 450-, 600-, and 900-arcsec. The bounding box is the 1200-arcsec (~35 km) grid cell. The shading depicts the amount of soil clay (ranging from 15% to 33%) and is for a single layer aggregated from the 11-layer, 30 arcsec CONUS-SOIL data set (Miller & White 1998). White signifies that no data are available.

Figure 3. Cross-scale comparison of simulated two-year average August switchgrass yields (2005 - 2006) and estimated and observed regional errors of mature yields (n = 6) for three Illinois sites (Figure 1). Simulated yields are presented as percent differences from respective baseline or observed site yields and RMSE and observed standard deviation are presented as the percent of observed site yield. a) Shabbona: baseline = 317 g C m⁻², % difference from baseline slope = 0.008 (p < 0.05), observed = 358 g C m⁻², % difference from observed slope = 0.007 (p < 0.05); b) Urbana: baseline = 480 g C m⁻², observed = 786.
g C m$^{-2}$, non-significant slopes; c) Simpson: baseline = 616 g C m$^{-2}$, observed = 301 g C m$^{-2}$, non-significant slopes. RMSE = Root Mean Square Error.

Figure 4. Cross-scale comparison of estimated spatial uncertainties of observed and simulated two-year (2005 – 2006) Average Mature August Yield (AMAY) of Illinois switchgrass (n = 3). The bias slope is 0.0102 (p < 0.05), the Root Mean Square Error (RMSE) slope is -0.00183 (p < 0.05), and the Bias-Corrected RMSE (RMSE$_{BC}$) slope is -0.00291 (p < 0.05). The observed standard error of the regional mean is 153.09 g C m$^{-2}$.

The observed regional mean yield is 481.66 gC m$^{-2}$.

Figure 5. Bias-corrected 68% confidence interval (CI) of simulated, two-year (2005 - 2006) Average Mature August Yield (AMAY) of Illinois switchgrass at 150-arcsec resolution (~4 km), corrected for spatial bias to reduce uncertainty (bias = 14 g C m$^{-2}$, bias corrected 68% CI = ±370 g C m$^{-2}$). a) Lower envelope of bias-corrected 68% CI (image range: 0 - 376 g C m$^{-2}$); b) Bias-corrected values are based on averages of the direct output (image range: 218 - 746 g C m$^{-2}$); c) Upper envelope of bias-corrected 68% CI (image range: 588 - 1116 g C m$^{-2}$).

Figure 6. Bias-corrected 68% Confidence Interval (CI) of simulated, two-year (2005 - 2006) Average Mature August Yield (AMAY) of Illinois switchgrass at 1500-arcsec
resolution (~42 km), corrected for spatial bias to reduce uncertainty (bias = 37 g C m$^{-2}$, bias-corrected 68% CI = ±362 g C m$^{-2}$). a) Lower envelope of bias-corrected 68% CI (image range: 0 - 318 g C m$^{-2}$); b) Bias corrected values are based on averages of the direct output (image range: 252 - 680 g C m$^{-2}$); c) Upper envelope of bias-corrected 68% CI (image range: 614 - 1042 g C m$^{-2}$). The black area along the bottom represents un-calculated partial cells.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Shabbona</th>
<th>Urbana</th>
<th>Simpson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude (dd N)</td>
<td>41.844156</td>
<td>40.042431</td>
<td>37.454856</td>
</tr>
<tr>
<td>Longitude (dd W)</td>
<td>88.852311</td>
<td>88.237928</td>
<td>88.722975</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>265</td>
<td>218</td>
<td>143</td>
</tr>
<tr>
<td>Soil depth (m)</td>
<td>1.50</td>
<td>1.50</td>
<td>0.97</td>
</tr>
<tr>
<td>Soil sand (%)</td>
<td>8.60</td>
<td>12.7</td>
<td>3.00</td>
</tr>
<tr>
<td>Soil silt (%)</td>
<td>60.8</td>
<td>58.7</td>
<td>71.9</td>
</tr>
<tr>
<td>Soil clay (%)</td>
<td>30.6</td>
<td>28.6</td>
<td>25.1</td>
</tr>
<tr>
<td>Nitrogen deposition (kg N m(^{-2}) yr(^{-1}))(^a)</td>
<td>7.5x10(^{-4})</td>
<td>4.0x10(^{-4})</td>
<td>6.0x10(^{-4})</td>
</tr>
</tbody>
</table>

\(^a\)N = nitrogen.
Table 2. Input configuration experiments for each grid resolution.

<table>
<thead>
<tr>
<th>Experiment (EXP)</th>
<th>Baseline values</th>
<th>Grid-specific values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>soil depth and texture</td>
<td>daily meteorology</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

\(^a\)C = carbon, N = Nitrogen
Table 3. Sensitivity of output yield to scale-dependent inputs for 3 Illinois (IL) sites. Values are percent differences from baseline. The data cover 13 grid resolutions ranging from 3- to 3600-arcsec and all individual dates and 2-year mature averages from 2003 to 2007.

<table>
<thead>
<tr>
<th>Input</th>
<th>Shabbona, IL</th>
<th>Urbana, IL</th>
<th>Simpson, IL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meteorology (Aug. 2005 forward)</td>
<td>-6 to 47%</td>
<td>-9 to 18%</td>
<td>-13 to 7% (7 to 2%)</td>
</tr>
<tr>
<td>Soil depth/texture and Soil initial conditions</td>
<td>~0%</td>
<td>-1 to 0%&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5 to 18%&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Soil depth/texture</td>
<td>~0%</td>
<td>-1 to 0%&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1 to 6%&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Only for dates and averages from August 2006 forward; otherwise ~0%.

<sup>b</sup>Only for dates and averages from August 2005 forward, otherwise ~0%.
Table 4. Significant cross-scale trends in biases of gridded meteorological input variables with respect to baseline values, for Shabbona, IL. The biases are calculated for a 5-year period from 2002 through 2006 (see eqns. 3-6) and for 13 grid resolutions ranging from 3- to 3600-arcsec.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Slope</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daylight length</td>
<td>7.79E-06</td>
<td>2.89E-02</td>
</tr>
<tr>
<td>Downward short wave radiation</td>
<td>-1.38E-03</td>
<td>1.10E-03</td>
</tr>
<tr>
<td>Minimum temperature</td>
<td>3.29E-04</td>
<td>3.25E-04</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>2.55E-04</td>
<td>2.32E-03</td>
</tr>
<tr>
<td>Vapor pressure deficit</td>
<td>5.79E-03</td>
<td>4.48E-05</td>
</tr>
</tbody>
</table>

*aThe precipitation trend was not significant at the 5% type I error rate, and thus not included here.
Table 5. Significant relationships between soil/elevation inputs and scale based on 13 grid resolutions ranging from 3- to 3600-arcsec. The data cover 3 Illinois (IL) evaluation sites.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Elevation (m)</th>
<th>Silt (%)</th>
<th>Sand (%)</th>
<th>Clay (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shabbona, IL</td>
<td>Urbana, IL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>209</td>
<td>52</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Maximum</td>
<td>272</td>
<td>54</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Slope</td>
<td>-0.013</td>
<td>-3.3x10⁻⁴</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P-value</td>
<td>0.012</td>
<td>0.048</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<p>| Simpson, IL |           |          |          |          |
| Minimum     | 135        | -        | 17       | 19       |
| Maximum     | 173        | -        | 20       | 22       |
| Slope       | -6.5x10⁻³  | -        | 4.8x10⁻⁴ | -5.2x10⁻⁴|
| P-value     | 0.015      | -        | 0.002    | 0.008    |</p>
<table>
<thead>
<tr>
<th>Statistic</th>
<th>Eastern United States field trials</th>
<th>150-arcsec Average Mature August Yield (AMAY)</th>
<th>1500-arcsec Average Mature August Yield (AMAY)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Mg ha⁻¹)</td>
<td>(Mg ha⁻¹) a</td>
<td>(g C m⁻²) a</td>
</tr>
<tr>
<td>Regional minimum</td>
<td>5.5 b</td>
<td>4.6 c</td>
<td>218 c</td>
</tr>
<tr>
<td>Regional Bias-Corrected mean</td>
<td>14 b</td>
<td>11 c</td>
<td>513 c</td>
</tr>
<tr>
<td>Regional non-bias-corrected mean</td>
<td>10 d</td>
<td>11 e</td>
<td>527 e</td>
</tr>
<tr>
<td>Regional maximum</td>
<td>23 b</td>
<td>16 c</td>
<td>746 c</td>
</tr>
<tr>
<td>Regional Minimum (At lower limit of 68%CI) f</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Regional Maximum (At upper limit of 68%CI) f</td>
<td>-</td>
<td>24</td>
<td>1116</td>
</tr>
<tr>
<td>Spatial bias</td>
<td>-</td>
<td>0.3</td>
<td>14</td>
</tr>
<tr>
<td>Spatial RMSE_{BC} g</td>
<td>5.6 h</td>
<td>6</td>
<td>282</td>
</tr>
</tbody>
</table>

a Conversion factor: 0.47 g C per 1.0 g biomass.
b Based on site mean harvest values from 24 sites across 14 states (McLaughlin and Kszos 2005).
c Based on Bias-corrected values of all whole cells in the full 6x5 degree grid.
d Based on Bias-corrected values of all whole cells in the full 6x5 degree grid.
e Based on all whole cells in the full 6x5 degree grid.
f Based on Bias-corrected values ± spatial 68% confidence interval (68%CI) of all whole cells in the full 6x5 degree grid.
g RMSE_{BC} = Bias-corrected Root Mean Square Error.
h Based on all whole cells in the full 6x5 degree grid.
Figure 2

Grid cell boundaries and counties within the Simpson area.
Figure 3

a) Shabbona

b) Urbana

c) Simpson
Figure 4

- Bias
- RMSE
- Bias–corrected RMSE
- Observed Std Dev

Grid cell size (arcsec):

Yield units (g C m\(^{-2}\))

% of observed regional mean yield
a) Lower 68% CI of AMAY

b) bias-corrected AMAY

c) Upper 68% CI of AMAY

Figure 5

- Lower 68% CI of AMAY
- Bias-corrected AMAY
- Upper 68% CI of AMAY

Legend:
- states
- lakes
- sites

Color bar:
- 1000 g C m\(^{-2}\)
- 800
- 600
- 400
- 200
- 0
Figure 6

a) Lower 68% CI of AMAY

b) bias-corrected AMAY

c) Upper 68% CI of AMAY

Legend:
- states
- lakes
- sites

Color scale:
- 0 g C m⁻²
- 1000 g C m⁻²

N

0 200 400 600 800 1000
DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.