The sensitivity of models of gross primary productivity to meteorological and leaf area forcing; a comparison between a Penman-Monteith ecophysiological approach and the MODIS light-use efficiency algorithm.

Paul B. Alton, Geography Dept., Swansea University, SA2 8PP, UK.

p.alton@swansea.ac.uk +44(0)1792 295069

November 10, 2015

1 Abstract

The current trend in land-surface and carbon modelling development is largely dichotomous: simple algorithms which minimise the number of biophysical parameters and meteorological drivers versus complex ecophysiologically based models which do not. Understanding the sensitivity of both types of approach to current uncertainties in Leaf Area Index (LAI) and meteorological forcing is an important step in producing accurate model predictions of land-atmosphere carbon exchange. We force two quite disparate models (the Moderate Resolution Imaging Spectroradiometer (MODIS) Light-Use Efficiency (LUE) algorithm and the ecophysiological model JULES-SF) with two LAI forcings (satellite and site-normalised) and two meteorologies (tower-based and reanalysis). Simulations are conducted for 67 sites and 10 vegetation classes. The sensitivity of modelled Gross Primary Productivity (GPP) to both LAI and meteorological forcing, thus derived, is compared with model bias against observed carbon fluxes. Our most novel findings are as follows: uncertainty in model formulation (LUE versus ecophysiological) is at least as important (20% change in simulated GPP) as that pertaining to LAI and meteorological forcing (10-20% change). However, all these uncertainties are modest compared to both model bias ($\leq$30%) and inconsistencies between observational datasets used for model calibration (45%). The ecophysiological model is more sensitive to meteorology (20% change in simulated GPP) than the LUE algorithm (10%) owing to the former's reliance on precipitation and shortwave radiation to calculate, respectively, the internal balances of water and energy.

Keywords

carbon cycle, process-based models, Moderate Resolution Imaging Spectroradiometer (MODIS), FLUXNET, Light-Use Efficiency (LUE), Leaf Area Index (LAI)

2 Introduction

Accurate model predictions of current ecosystem gross productivity are essential for our understanding of ecology, the carbon cycle and how environmental change is likely to have an impact on future photosynthesis (Running et al 1999; Friend et al 2007). Understanding the sensitivity of land-surface and carbon models to current uncertainties in meteorology and LAI forcing is an important step in producing accurate model predictions (IPCC 2007). Further, this sensitivity is likely to vary for models of fundamentally different structure and complexity. Although, in principle, more complex ecophysiological models confer greater flexibility to modelling vegetation under different and changing environmental conditions, they also impose greater demands in terms of process parameterisation and meteorological forcing compared to simple algorithms such as LUE (Abramowitz et al 2008; McCallum et al 2009; note that frequently used acronyms and algebraic quantities are listed in Tab. 1).

For the LUE algorithm used to calculate global daily GPP from MODIS reflectance observations, estimates vary by 15-25% according to the source of the reanalysis meteorology used to drive the model (Zhao et al 2006). Similarly, simulated GPP for 15 Ameriflux sites decreases by on average 28% when tower-based (ground) meteorological forcing replaces the standard reanalysis meteorology used to drive the MODIS LUE algorithm (Heinsch et al 2006). Heinsch et al conjecture that errors in meteorology are likely to be more important than those associated with landcover classification or LAI phenology.

Since Heinsch et al, the MODIS algorithm has been recalibrated. Furthermore, the fraction of Photosynthetically Active Radiation (fPAR) used to force the model, along with the corresponding LAI product, have been updated (Zhao & Running 2010). Using this new collection 5 release, Fang et al (2012) find, for 83 sites, a root mean square discrepancy between MODIS and ground-based LAI which is somewhat greater than that documented by Heinsch et al (1.1 m^2 m^{-2} versus 0.5-1.0 m^2 m^{-2}). This suggests that errors in LAI forcing might be more important than previously surmised. Indeed, Puma et al (2013) assert that LAI phenology rather than meteorology is the primary influence on simulated GPP, although this conclusion is based on interannual variability rather than driver uncertainty. Thus, an open question exists concerning the relative importance of LAI and meteorology. Moreover, it is possible that differences in model formulation and process complexity may be more significant than uncertainties arising from forcing (Knorr & Heimann 2001). For example, Cramer et al (2001) estimate a $\pm$20% uncertainty in simulated global Net Primary

Productivity (NPP) owing to model formulation. Similarly, for water fluxes, model complexity, rather than forcing, appears to contribute greatest uncertainty (50%) to predicted zonal evapotranspiration (Vinukolla et al 2011; see also Dirmeyer 2011).

The current trend in land-surface modelling is largely dichotomous (e.g. Fisher et al 2008; McCallum et al 2009). Relatively simple algorithms which minimise biophysical parameters and meteorological drivers vie with complex ecophysiological models (sometimes referred to as process-based or mechanistic) which require far more parameters and meteorological drivers. We acknowledge, though, the recent emergence of a third group of models which is essentially statistical, such as artificial neural networks (Beer et al 2010). Complex ecophysiological land-surface models typically incorporate explicit processes for light interception, photosynthesis, respiration and plant and soil hydrology (Alton 2013). They usually contain a Penman-Monteith energy balance (Monteith 1965) and require 7-9 meteorological variables rather than 1-3 variables used to force a LUE algorithm. Many of them contain internal (prognostic) calculations of LAI which, in contrast to LUE models, allow them to be used independently of satellite LAI forcing in simulations under future climate (Richardson et al 2012). Owing to its simplicity, LUE is well suited to current global satellite datasets such as the MODIS fPAR product, based on multi-spectral reflectance (McCallum et al 2009). In contrast, some of the ecophysiological models have been developed at specific sites where intensive field measurements allow parameterisation and testing of the model (e.g. Baldocchi & Wilson 2001; Williams et al 1996). However, both types of model are being calibrated at site level in order to scale (or with a view to scaling) to global level using grid-scale (1°) reanalysis meteorology (e.g. Yuan et al 2007; Friend et al 2007; Alton 2013).

In the current study we compare a LUE algorithm and an ecophysiological model in terms of validation against eddy covariance carbon fluxes and sensitivity to forcing. Several model comparisons already exist in the literature. However, typically they compare predicted regional and global fluxes (e.g. NPP by Cramer et al 2001 and evapotranspiration by Dirmeyer 2011). There has been little focus on the sensitivity at site level of different types of model to meteorological and LAI forcing. In particular, few studies attempt to contrast the extremes of model complexity i.e. the model dichotomy described above. The open-access availability of fluxes and tower-based meteorology within the expanding FLUXNET archive allow such a comparison to be made. The enterprise is aided by the recent compilation of ancillary variables such as field LAI for FLUXNET locations (Agarwal 2012). The continuing improvement in global reanalysis meteorology

(e.g. Princeton Reanalysis and Global Soil Wetness Project; Sheffield et al 2012; Dirmeyer 2011) and LAI satellite products (Collection 5 release from MODIS) justifies a reexamination of model sensitivity to forcing which can then be placed in the context of other uncertainties such as model formulation and bias in model calibration datasets.

The overarching aim of the current study is to quantify and to compare the impact on predicted GPP of uncertainties from LAI and meteorological forcing for two carbon models representing opposite extremes in model complexity (a simple LUE algorithm versus a complex ecophysiological land-surface model). To some extent we build on Heinsch et al (2006) but our sample is much larger (67 sites vs 15), determining sensitivity to *both* LAI forcing and meteorology for *two* quite different models using the latest datasets of LAI and reanalysis. Our specific objectives are:

1. to determine the sensitivity of simulated GPP to driving meteorology and LAI forcing for a large number of FLUXNET sites (67) which encompasses a diverse range (10) of Plant Functional Types (PFTs);

2. to compare simulated GPP to estimates based on observed eddy covariance carbon fluxes and thus place the sensitivity from objective (1) into the context of model accuracy and bias;

3. to compare the sensitivity from objective (1) with the impact of model formulation (LUE or ecophysiological), as well as previously documented errors (e.g. uncertainties in biophysical parameters);

4. to quantify the difference between satellite and field LAI for a large number of globally distributed FLUXNET sites by virtue of our LAI sensitivity test which makes use of both kinds of measurement.

3 Material and Methods

In summary, the methodology consists of driving two carbon models with two phenologies (satellite and site-normalised) and two meteorologies (tower-based and reanalysis) and comparing the simulated output in GPP (Fig. 1). First, we introduce the two models (§3.1). Then, in §3.2, we describe the datasets required both for forcing (LAI and meteorology) and for GPP validation (eddy covariance carbon fluxes). Finally, we set out the modelling protocol for the model simulations and sensitivity experiments (§3.3).

3.1 Models

3.1.1 MODIS-GPP

The MODIS-GPP algorithm is a simple model based on light-use efficiency i.e. the daily conversion of solar radiative energy to carbohydrate synthesis and carbon storage. Thus:

$$GPP \;=\; 0.45\ \Sigma SW\ \epsilon_{max}\ fPAR\ fVPD\ fT_{min} \tag{1}$$

where ΣSW denotes the daily (24 hr) total of downwelling shortwave solar radiation. Multiplication by 0.45 and fPAR yields that part of the solar spectrum which is being absorbed by the foliage for photosynthesis. Parameter ϵ_{max} is the maximum light-use efficiency. In Eq. 1, fVPD and fT_{min} are stress scalars determined on a daily timestep. They diminish the LUE by ramp functions which vary between 0 and 1 according to thresholds in average daytime vapour pressure deficit and minimum air temperature. Note that ϵ_{max} and the thresholds for the stress scalars are defined per PFT (Zhao & Running 2010).

In the standard MODIS product based on Eq. 1 (MOD17), fPAR is supplied by the MODIS product MOD15A2. This fPAR is inferred by matching observed multispectral reflectances over a 8-day period with reflectances simulated by a 3-D radiative transfer model, for a range of conditions, and stored in a look-up table. For MOD17, near real-time daily inputs of ΣSW, minimum air temperature and daytime averaged vapour pressure deficit are supplied from a Data Assimilation Office reanalysis meteorology. Precipitation is not input. Therefore, no water balance nor calculation of soil moisture stress on photosynthesis is undertaken. The neglect of seasonal drought is recognised as a limitation of the model (Zhao & Running 2010), although the stress scalar for Vapour Pressure Deficit (VPD) may to some extent act a proxy for diurnal drought stress. The model has no memory and does not require spin-up. Notably, photosynthesis within MODIS-GPP depends on air temperature (T_{air}) rather than on a canopy temperature which is derived from an energy balance such as that conducted in JULES-SF.

The main advantages of this LUE algorithm are: phenology depends on a readily available satellite dataset (MODIS product); the number of biophysical parameters is small; and the simplicity of the model allows rapid computation. The main disadvantage is an absence of an explicit canopy light interception which accounts for leaf-level light saturation, separate components of direct and diffuse sunlight and a photo-

synthetic capacity (maximum Rubisco-limited carboxylation rate) which declines with depth through the canopy (Meir et al (2002)). A further disadvantage is the neglect of soil moisture stress on photosynthesis.

3.1.2 JULES-SF

JULES-SF (Joint UK Land Environmental Simulator) is an enhanced version of the new UK Met.Office Surface Exchange Scheme (Cox et al 1999). Key equations for JULES-SF are given in the Appendix of Alton & Bodin (2010) with the exception of a subsequent reformulation of plant maintenance respiration which now consists of separate, additive terms for leaf, stem and root respiration according to Q_{10} relationships based, respectively, on canopy and soil temperature (Law et al 1999). In the following overview we focus on major differences with simple LUE models.

JULES-SF, like many process-based land-surface models, is forced by more meteorological variables compared to LUE models. In addition to SW, T_{air} and VPD (or equivalently specific humidity), JULES-SF requires downwelling longwave thermal radiation (for energy balance), precipitation (for water balance), wind speed (to determine boundary layer heat and water conductance), and pressure (for the CO_2 and water vapour gradient across the leaf stomata). The core energy calculation is the standard Penman-Monteith approach (Monteith 1965), ensuring the balance of ingoing and outgoing energy fluxes at the land-surface.

JULES-SF takes account of diffuse and direct sunlight at multiple heights within the canopy including sunfleck penetration (hence SF) and diffuse sky irradiance (from for example cloud and haze). It is one of most elaborate land-surface models which operates globally in terms of light interception (Alton et al 2007). Photosynthesis is calculated separately within each of 5 leaf layers according to a biochemical co-limitation model (Collatz et al 1991), before summing to produce a canopy total. In stark contrast to MODIS-GPP, the co-limitation implies a non-linear response to light, which saturates under high irradiance to a Rubisco-limited rate. The Rubisco limit is proportional to active leaf nitrogen, per unit area, and is fixed by a parameter for the top of the canopy, V^0_{cmax}, which is probably the most important determinant of GPP and NEE in an ecophysiological model of this kind. In accordance with field observation, active leaf nitrogen, and therefore the Rubisco limit, decline exponentially from the top of the canopy downwards (e.g. Lewis et al (2000); Meir et al (2002)). Leaf photosynthesis is linked to transpiration through a Ball-Berry stomatal model (Ball et al 1987) which is sensitive to the relative humidity and temperature of the canopy. Canopy

temperature is derived from the Penman-Monteith energy balance.

As in most ecophysiological models, a water balance is conducted, accounting for input (precipitation) and output (evapotranspiration and above and below ground runoff). Therefore, in contrast to LUE models, JULES-SF possesses memory of past forcing. The water balance allows moisture content to be calculated in 4 soil layers of thickness (top downwards) of 0.1, 0.25, 0.65, 2.0 m. Plant water extraction depends on a fine root distribution with vertical exponential scale-depths of 0.1-0.3 m, depending on PFT (Jackson et al 1996). The soil moisture content enters the stomatal model as a stress factor. Thus, when soil moisture is limiting, extraction declines, the stomata close and photosynthesis decreases.

The main advantages of JULES-SF compared to a LUE approach are: explicit account for light interception and Rubisco photosynthetic capacity at different heights within the canopy; an ecophysiological approach which accounts for observed non-linear behaviour such as colimited photosynthesis and seasonal lag effects such as soil moisture stress; and a short timestep which accounts for the diurnal cycle. The main disadvantages are: a large number of parameters, many of which poorly known at least at PFT level; a large number of meteorological variables including precipitation which is difficult to reconstruct accurately in reanalysis; and a longer computational time owing to a shorter timestep and more complex calculations (e.g. energy balance and photosynthesis), though this is rarely inhibitive unless decadal or ensemble simulations are being run.

Tab.2 summarises the salient differences btw MODIS-GPP and JULES-SF.

3.2 Datasets

Datasets serve as input or validation and these two categories are discussed in turn below. As input, both models require biophysical parameter values (defined per PFT), a timeseries of meteorological forcing and a timeseries of LAI.

3.2.1 Input: Biophysical Parameters

As described above, MODIS-GPP requires maximum LUE (ϵ_{max}) and thresholds for the stress scalars moderating photosynthesis according to minimum air temperature and VPD stress. To permit a model

comparison, we define these for the same PFTs configured in JULES-SF. In JULES-SF there are typically 30-50 parameters per PFT although, in practice, less than 10 of them have a large influence on predicted carbon exchange. Many of the biophysical parameters are plant attributes which are either structural (e.g. rooting depth, canopy height), optical (e.g. leaf absorptance) or physiological (e.g. photosynthetic capacity, minimum stomatal conductance). They are assigned values from average collated field measurements (Alton & Bodin 2010). Probably the most influential parameter on modelled carbon fluxes is the maximum carboxylation rate, a measure of photosynthetic capacity (V^0_{cmax}), which is based on the average of leaf measurements compiled by Wright et al (2004) and Kattge et al (2009). The primary parameters for both models, ϵ_{max} and V^0_{cmax}, are given in Tab. 3 for each PFT.

To determine the soil hydraulic properties (e.g. conductivity at saturation, Clapp-Hornberger exponent) required by JULES-SF, we adopt the average soil composition measured at each site in the FLUXNET ancillary database (Agarwal 2012). The recorded clay and silt contents are related to the soil categorisation in Campbell & Norman (1998).

3.2.2 Input: Meteorological forcing

According to the experiment undertaken (discussed below), meteorological forcing either consists of that recorded *in situ*, typically 5-10 m above the vegetation (tower-based), or a reanalysis meteorology reconstructed globally by workers at Princeton University (Sheffield et al 1996) and continually improved and updated (Sheffield et al 2012). The tower-based meteorology, provided by FLUXNET (Falge et al 2002), is initially averaged over the JULES-SF 3hr timestep which is deemed of sufficient temporal resolution to simulate the diurnal cycle within the ecophysiological model. To run MODIS-GPP, we compress meteorological forcing to the one-day timestep conventionally used in Eq. 1 for the standard MODIS global GPP product (MOD17). For this model, we need only extract SW, specific humidity and T_{air}. Specific humidity is converted to VPD using T_{air} (e.g. Campbell & Norman 1998) and the daily minimum air temperature is extracted from T_{air}. The reanalysis already has a 3 hour timestep which we average to a daily interval for MODIS-GPP. The reanalysis timeseries is selected according to the 1° global grid cell containing the site being simulated and the corresponding year.

3.2.3 Input: LAI forcing

Both models depend either directly (JULES-SF) or indirectly (via fPAR for MODIS-GPP as discussed below) on LAI. To create a satellite LAI timeseries, we extract from the 8-day MOD15A2 Collection 5 LAI product a 7km×7km subset (49 pixels) centred on the site location. We mean average pixels of good quality (i.e. main algorithm, no significant cloud and >50% detectors working; Yang et al 2006). An ideal assessment of model sensitivity to uncertainty in LAI forcing requires a field-based timeseries of site LAI to compare against the satellite phenology. With the exceptions of a few sites, such a field-based phenology does not exist (Melaas et al 2013). Therefore, to assess sensitivity to LAI, two categories of simulation are conducted: an unnormalised satellite timeseries and one that is normalised to maximum site-recorded LAI.

Site-recorded values are available from the FLUXNET ancillary archive (Agarwal 2012; site LAI database hereafter) and represent measurements conducted using principally LiCOR, harvesting and leaf litter collection. We reject site measurements based on leaf litter as a detailed knowledge of senescence and leaf-out is required to reconstitute canopy LAI. However, the site LAI database only describes the method in about half the cases. Therefore, we must assume that some leaf litter measurements remain in our sample. Where possible, we match the siteyear of the simulation to the correct year from the site LAI database and extract the maximum site LAI in order to compare with the corresponding measurement (by interpolation across the 8-day interval if necessary) in the MOD15A2 timeseries (quality=1). If the siteyear precedes the available MODIS timeseries (<2002), a median MODIS timeseries is used, averaging years 2002-2010 (quality=2). In order to provide a sufficiently large sample to determine sensitivity over all PFTs, values are also adopted from the site LAI database where the year is incorrect/unknown (quality=2) and where the date is unknown (quality=3). For quality=3, we assume site LAI corresponds to the maximum annual value and we match it to the maximum value of the satellite timeseries. Note that there are no significant differences between the means of the 6 PFTs where there are sufficient site measurements to compare quality=1 against quality≤3. However, we check the impact of site LAI quality on our results.

For MODIS-GPP, an 8-day fPAR timeseries already exists in tandem with MOD15A2 LAI and this fPAR timeseries is generally adopted in Eq. 1 for the standard GPP product MOD17. However, the landcover adopted in the MODIS fPAR algorithm is not necessarily the same as that recorded at the FLUXNET site. Further, if we wish to perturb the LAI timeseries in our sensitivity experiments, fPAR would have to be

recalculated from the original radiances. Therefore, to render the sensitivity experiment feasible, we use the two-stream approximation (Sellers et al 1996) to derive fPAR from 8-day LAI for both the site-normalised and satellite (unnormalised) timeseries. This is carried out prior to the MODIS-GPP simulations in order to supply fPAR in Eq. 1. The two-stream approximation takes account of upwelling and downwelling direct and diffuse light in a uniform leaf distribution according to LAI, solar zenith angle and the fraction of diffuse sky irradiance over the course of the 8-day interval. We adopt the two-stream approximation for convenience because it is already used in JULES-SF to calculate surface albedo. Note that the timestep of both models (3 hr and 1-day for JULES-SF and MODIS-GPP, respectively) is smaller than that of the LAI and fPAR timeseries (8-day).

3.2.4 GPP Validation

To validate the GPP predicted by both models, we adopt Net Ecoystem Exchange (NEE) recorded in the main FLUXNET database. Siteyears that are available to the general modelling community lie between 1991-2010, though the bulk (93%) range 1997-2009 (Falge et al 2002; Yuan et al 2010). Sites are distributed worldwide but are biassed towards forest in North America and Europe (Fig. 2). To minimise the impact of incomplete energy closure (Foken 2008), we exclude fluxes recorded under low frictional velocity ($<$0.16 ms^{-1}; Goulden et al 1996; Reichstein et al 2003) or, if frictional velocity is unrecorded, where windspeed $<$2 ms^{-1} (Medlyn et al 2003). We recognise, however, that the effectiveness of these velocity filters may be site-dependent and that closure may depend on other factors such as storage terms which relate to the structure of the vegetation (Wilson et al 2002; Masseroni et al 2014).

To compare observations with model output, several steps are required, beginning by averaging good quality NEE measurements into 3 hr intervals. To convert NEE to GPP, we construct an ecosystem respiration model (R_e) for each sitemonth by best fitting a quadratic function of T_{air} against nocturnal 3 hr fluxes. GPP is then estimated at each 3 hr timestep using the corresponding T_{air} and sitemonth function for R_e. Thus;

$$GPP = R_e(sitemonth, T_{air}) - NEE \tag{2}$$

where negative values for NEE indicate carbon assimilation by the surface. The use of nighttime carbon fluxes to define ecosystem respiration has been adopted by many authors in the past (Valentini et al 2000;

Yuan et al 2007; Desai et al 2008) but some authors adopt an exponential function for R_e in Eq. 2 (e.g. Medlyn et al 2003). However, we find that a quadratic fit produces a lower root mean square error. Since the timestep of MODIS-GPP is daily (24 hr), both the flux-derived GPP inferred from Eq. 2 and GPP simulated by JULES-SF are averaged to a daily rate in $gCm^{-2}d^{-1}$.

3.3 Modelling Protocol and Experiments

Fig. 1 provides a schematic overview of the simulations and sensitivity experiments. Simulations are conducted for all 484 siteyears within the main FLUXNET database for which tower meteorology is recorded and available to the general ecological modelling community. For $\simeq$70% siteyears where site LAI is recorded, a sensitivity analysis is carried out by conducting 3 main simulations per model: (1)tower-based meteorology plus site-normalised LAI timeseries (default); (2) reanalysis meteorology plus site-normalised LAI timeseries (meteorology-perturbed); and (3) tower-based meteorology plus satellite (unnormalised) LAI timeseries (LAI-perturbed).

Sensitivity is defined as:

$$\Delta GPP = \frac{GPP(PERTURB) - GPP(DEF)}{GPP(DEF)} \tag{3}$$

where GPP(PERTURB) is either LAI-perturbed (GPP(LAI_PERTURB)) or meteorology-perturbed (GPP(MET_PERTURB)). GPP(DEF) derives from the default simulation. Both GPP(PERTURB) and GPP(DEF) are in $gCm^{-2}d^{-1}$. For the meteorology perturbation, we also carry out 7 auxillary simulations to ascertain the sensitivity to individual meteorological variables. We do this by replacing only one of the 7 tower-meteorology variables by its reanalysis counterpart.

For those siteyears without site LAI, only a default simulation is conducted with tower meteorology and satellite LAI timeseries. This allows these siteyears, where they contain valid NEE measurements, to be included in GPP validation. Thus, to make maximum use of the data, our sample sizes differ somewhat according to validation or sensitivity analysis (the number of sites and siteyears in Tab. 3 refer to sensitivity). In our results, we check the impact of mixing normalised and unnormalised LAI timeseries on our GPP validation.

To run the simulations every site must be attributed to one of the 10 PFTs defined in JULES-SF and given in Tab. 3. To run simulations for JULES-SF (a model containing soil water balance) a complete and continuous meteorology is required. Therefore, protracted gaps in the tower-meteorology (generally before/after the growing season) are filled with the reanalysis. Note, however, that the sensitivity analysis and the validation are carried out by only averaging over the period of the siteyear for which tower meteorology is available. Furthermore, for validation, we only average modelled and flux-derived GPP across timesteps where valid NEE is available under sufficient frictional velocity. For JULES-SF, the soil moisture content for each siteyear is spun-up by splicing the required meteorology and LAI timeseries back-to-back over a 5 yr period and pre-running the model over this period.

We recognise that there are large differences in the LAI and meteorological sampling size (footprint) between the perturbed and the default simulations. Field LAI has been upscaled to the satellite footprint for comparative purposes at individual sites by some authors (e.g.De Kauwe et al 2011) but such detailed ground sampling does not exist for the large number of sites in the present study. Further, we would argue that scaling mismatches of this kind constitute part of the typical error or uncertainty associated with running a global simulation of productivity that has been, for example, previously validated or calibrated at site-level. Note that the LAI and meteorological datasets differ not just in spatial scale but often in methodology. For example, satellite LAI derives from multispectral reflectance, which may saturate for dense canopies even at near-infrared wavelengths, whereas the field LAI is based on LiCor light-extinction profiles and harvesting measurements. Reanalysis meteorology comprises satellite and interpolated ground-based measurements with use of temporal disaggregation, whereas the tower-based meteorology is created from high frequency measurement with *in situ* instruments.

Although there little consensus on what constitutes current uncertainties in model forcing, the present study follows Heinsch et al (2006) in comparing simulations based on *in situ* (relatively accurate) forcing with those based on satellite observations in order to determine model sensitivity to forcing uncertainties that are typical of spatial upscaling. The justification for this approach is that many models are calibrated or validated at site level, using *in situ* meteorology and possibly some estimate(s) of field LAI, to be run globally using satellite-based data (e.g. Yuan et al 2007; 2010).

4 Results & Discussion

First we validate the models. Then we analyse model sensitivity to LAI and meteorological forcing. Validation also includes a comparison of satellite LAI against field estimates since this defines the uncertainty in LAI forcing used within the sensitivity experiment. Furthermore, it provides a check on a new release of a widely used MODIS product.

4.1 Validation: LAI

MOD15A2 LAI exhibits a saturating (exponential) relationship against field-based estimates from the site LAI database (Fig. 3). This relationship is largely independent of the quality of field LAI used (quality=1 or quality$\leq$3). The bias of satellite LAI with respect to ground-based measurements is -12% for site LAI $<$3.9 m^2 m^{-2} (median site LAI) and -25% for site LAI $\geq$3.9 m^2 m^{-2}. Removing C3 crops from the sample, reduces the bias to -11% and to -8%, respectively. Thus, this Collection 5 MOD15A2 release appears to remove positive bias at low LAI which characterises preceding releases (Abuelgasim et al 2006; Aragao et al 2005; Heinsch et al 2006) and reveals an underestimation of site LAI by MODIS especially for non-woody PFTs. Severe underestimation of C3 crops may arise from the inclusion of surrounding vegetation with less vigorous growth in the satellite footprint. We recognise that field measurements sample a much smaller area ($\sim$100 m) than the satellite footprint ($\sim$1 km) but, in general, we would expect this mismatch to generate dispersion in Fig. 3 rather than a systematic offset or bias.

Our field averages are somewhat smaller than those of a much larger field database (Asner et al 2003), but this only confirms a general tendency for field estimates to exceed remote sensing measurements (Tab. 4). This accords with Fang et al (2012) who find a tendency for both MODIS and SPOT to underestimate field LAI when field estimates exceed 3 m^2 m^{-2}. Note that, in Tab. 4 and many subsequent results, we show the median value of the PFT means (or, for sensitivity, the median value of the absolute PFT means) owing to the small number of PFTs being evaluated. The reader should bear in mind, however, that some ecosystems (e.g. North American broadleaf and needleleaf forests) are better represented numerically than others (Tab. 3 and Fig. 2).

4.2 Validation: GPP

Despite representing extremes in process complexity, both models exhibit a similar saturating (exponential) response against observational (eddy covariance flux) estimates of GPP (Fig. 4). Thus, for <5.7 $gCm^{-2}d^{-1}$ (median observed), MODIS-GPP and JULES-SF both overestimate observational estimates by +28% and +37%, respectively. Above the median, the respective bias is -26% and -13%. The highest values of observation-derived GPP (for tropical broadleaf forest and C3 crops) are underestimated by 50-100% by both models. Note that our median observed (and modelled) daily GPP is quite high for a sample dominated by temperate ecosystems (equivalent to 2.1 kg $m^{-2}yr^{-1}$; c.f. Luyssaert et al 2007). This is because model and observation can only be compared where the tower-based meteorology is available (bias towards the growing season) and frictional velocity is moderately high (bias towards daytime; see §3.2.4 and §3.3).

The tendency for models, regardless of their formulation, complexity and calibration, to underestimate the highest productivity rates inferred from eddy covariance fluxes (either GPP or maximum daytime assimilation rates, |NEE|) is evident in previous studies. For example, the ecophysiological CSIRO Biosphere Model underestimates peak |NEE| recorded at two FLUXNET sites (one needleleaf and one broadleaf) at both half-hourly and monthly (25% underestimation) timescales (Wang et al 2007). Even an ecophysiological model with an elaborate light canopy interception, accounting for leaf-clumping, underestimates peak daytime assimilation by at least 50% for a broadleaf forest (Baldocchi & Harley 1995). A purely empirical carbon model, regressed against multiple satellite drivers (land-surface temperature and enhanced vegetation index) for 42 Ameriflux sites, underestimates the highest observed 8-day assimilation rates by 50% (Xiao et al 2011). A calibrated LUE model also underestimates 8-day GPP at high productivity FLUXNET sites (Yuan et al 2007). A novel machine-learning technique (neural network model) appears to underestimate GPP at the most productive sites by 25% (Jung et al 2011).

Given the tendency for diverse models to exhibit a similar bias against eddy covariance fluxes, we should consider whether observational values are systematically in error. The observational GPP values are not measured directly but inferred from measured NEE (Eq. 2). Several methods have been applied to separate respiration from photosynthesis but most of them yield estimates that vary by 5-10% (Desai et al 2008). Nevertheless, we check our observational GPP for 19 sites which overlap with the sample of Yuan et al (2007) who adopt a slightly different respiration model. The difference in mean observational GPP aver-

aged, respectively, across non-tropical broadleaf forest, non-mediterranean needleleaf forest and the whole Yuan et al sample is 6%, 10% and 1%. Thus uncertainty in partitioning of respiration and GPP cannot account for the observation-model discrepancies in Fig.4 which are up to 50-100%. Eddy covariance measurements at the original 30 second timestep are very noisy but averaging over the siteyear, as we do in the present study, reduces the random error considerably (Hollinger & Richardson 2005). Systematic errors are more problematic, with incomplete energy closure strongly suggesting sizeable bias in detected carbon fluxes (Wilson et al 2002). Growing season closure averaged across all FLUXNET siteyears is 0.77, with closure somewhat higher during the day compared to night. Incomplete daytime closure implies assimilation by the canopy is actually higher. Underestimation of ecosystem respiration, owing to incomplete closure at night, will also lead to underestimation of observed GPP via Eq. 2. Correction for incomplete closure, therefore, would likely exacerbate the pronounced model underestimation at high GPP. Furthermore, we find that observation-model discrepancies in Fig. 4 do not correlate with siteyear closure values.

Exploring further the possibility of observational bias, we note that both JULES-SF and MODIS-GPP appear to overestimate (rather than underestimate) annual GPP when converting NPP measured at Ecosystem Model-Data Intercomparison (EMDI) class A sites (Olson et al 2008) to GPP (Fig. 5). A similar qualitative response also characterizes the ecophysiological model LPJ (Hickler et al 2006). Thus, large systematic differences are apparent between observations of productivity based on eddy covariance fluxes and those inferred from allometric estimates of biomass change (NPP) for the same PFT. One possibility is that some of the most productive FLUXNET sites are recovering from disturbance, a process unaccounted for in carbon models (Friend et al 2007), and that the associated high carbon assimilation is, therefore, atypical of the PFTs that the FLUXNET sites represent. Further, the distribution of EMDI NPP sites is more evenly distributed globally compared to our FLUXNET sample (Olson et al 2008). This geographical disparity appears to invalidate the comparison against EMDI in Fig. 5, even if it is being conducted for the same PFTs as those used to simulate FLUXNET sites. However, both observational datasets (EMDI NPP and FLUXNET) are currently being used for validation purposes (e.g. Zaehle et al 2005; Yuan et al 2007), and are therefore liable to introduce bias into the model calibration. Thus, systematic observational differences, owing to both sampling (geographical and successional) and measurement bias, impose a significant limitation on the accuracy of carbon models and constitute a major uncertainty in the modelling process.

Owing to its overall lower bias against eddy covariance fluxes, the ecophysiological model has a slightly higher modelling efficiency than the LUE algorithm (0.2 versus 0.1; Tab. 5). Notably, however, JULES-SF exhibits more scatter (Fig.4). The median root mean square error of both models is quite similar (Tab. 5). The lack of scatter for MODIS-GPP is striking given the simplicity of its parameterisation. For example, maximum LUE is defined per PFT (Tab. 3) but biophysical parameters such as Rubisco-limited photosynthetic capacity are known to vary by an order of magnitude for plants within the same PFT (Wright et al 2004). Despite the pronounced bias with the LUE algorithm, the lack of scatter suggests that there may be advantages to simplifying complex processes which are modelled explicitly by ecophysiological models. Within the latter, a greater number of biophysical parameters potentially confers more flexibility in simulating fluxes at a diverse range of sites. In practice, however, many of the parameters used in ecophysiological models have poorly constrained values which may increase the scatter. A similar conclusion is drawn when comparing ecophysiological land-surface models against statistical models (Abramowitz et al 2008).

4.3 Sensitivity

The sensitivity of both models to LAI and meteorological forcing is comparable when examining the median response across all PFTs (10-20% change in simulated GPP; Tab. 6). However, the ecophysiological model is more sensitivity to meteorology ($\simeq$20% change) than the LUE algorithm ($\simeq$10%). Notably, model formulation i.e. ecophysiological versus LUE is at least as important (sensitivity $\simeq$20%) as LAI forcing and meteorology. The large uncertainty in carbon fluxes owing to model formulation is already noted in previous studies, for example $\sim$20% for global NPP (Cramer et al 2001; Knorr & Heimann 2001). Simulating annual GPP for Europe using 3 ecophysiological models, Jung et al (2007) infer a greater average uncertainty owing to the selected model (15-35%) compared to that owing to meteorological forcing (5-20%). Similarly for energy/water fluxes, Vinukolla et al (2011) conclude that model formulation, rather than forcing, contributes the greatest uncertainty to zonal latent heat exchange ($\simeq$50%; see also Dirmeyer 2011).

Using the interannual variability of LAI, Puma et al (2013) determine a sensitivity to LAI phenology (10% change in GPP for ΔLAI=0.3-0.6 m^2 m^{-2}) which is close to that found in the current study. These authors identify sensitivity to LAI as more important than sensitivity to meteorology. However, their interannual variability in meteorology is unquantified making it difficult to assess that statement against the current results. Tab. 6 demonstrates that sensitivity to LAI forcing varies greatly according to PFT. In general,

low-LAI systems are more sensitive, when expressing per unit LAI change, owing to the tendency for GPP to saturate at high LAI (Fig. 6).

Sensitivity to meteorological forcing also varies greatly between PFTs. In part, this is attributable to higher uncertainty in climate for certain regions. For tropical broadleaf trees for example, average SW and VPD are, respectively, 12% higher and 0.2 kPa lower in the reanalysis meteorology compared to the tower-based meteorology, perhaps owing to localised cloud or weather systems (Fig. 7). As a median average across all PFTs, the greatest sensitivity to individual meteorological drivers is SW(12%), specific humidity (9%) and precipitation (8%) for JULES-SF and SW(8%), T_{air}(8%) and specific humidity (4%) for MODIS-GPP (Tab.7). In percentage terms, the greatest root mean square error between tower and reanalysis drivers is for precipitation (42%) and VPD (33%), explaining the sensitivity of JULES-SF to the former and the sensitivity of both models to specific humidity. Of all the meteorological variables, precipitation is one of the most difficult to predict in reanalysis and is highly variable both spatially and temporally (Sheffield et al 1996). In contrast, SW exhibits a relative low bias (3%) and a root mean square error of 14 W m^{-2} (9%), although regional exceptions exist such as those mentioned above for tropical broadleaf trees (12% bias; Fig. 7).

The sensitivity of both models to relatively small average percentage uncertainties in SW is striking. Further, it is somewhat surprising that the ecophysiological model, which calculates photosynthesis as a co-limitation of light and Rubsico-related capacity (therefore GPP saturating at high light levels), is more sensitive to SW than a LUE algorithm which is directly proportional to SW. The stronger SW dependence in JULES-SF arises from the energy balance conducted at each timestep which is largely determined by SW. This Penman-Monteith balance also determines the temperature used in the calculation of photosynthesis and therefore accounts, at the same time, for the relatively low dependence of JULES-SF on T_{air}. The Penman-Monteith balance is missing from the LUE algorithm which explains this model's strong dependence on T_{air}. Overall, JULES-SF is more sensitive to climate than MODIS-GPP (Tab.7), perhaps owing to the additional dependence of the ecophysiological model on precipitation which determine soil moisture stress on photosynthesis. Diverse methods in calculating drought stress within models, e.g. via water balance (as in JULES-SF) or via VPD (as in both MODIS-GPP and JULES-SF), have already been cited as contributing to the large range in modelled global NPP (Cramer et al 2001).

For 15 Ameriflux sites, Heinsch et al (2006) claim a greater sensitivity of MODIS-GPP to meteorology than that estimated here (20-25% vs 10%). Similarly, Zhao et al (2006) find that simulated global GPP varies by 15-25% according to the reanalysis meteorology (Data Assimilation Office, ECMWF or NCEP) employed to drive the MODIS LUE algorithm. However, the Princeton reanalysis that we adopt in the current study is a hybrid product which improves on the Data Assimilation Office reanalysis used in the standard MODIS global GPP product (and used in the comparisons of Heinsch et al and Zhao et al). The improvement in reanalysis stems from a more extensive calibration against ground measurements (Sheffield et al 2006; see also Dirmeyer 2011 for Global Soil Wetness Project). Thus, our reanalysis bias against tower-based meteorology is only 3% and 0.5 K for annual SW and daily minimum air temperature, respectively (Fig. 7), whereas the corresponding bias for the Data Assimilation Office reanalysis is 20% and several K. The improvement in reanalysis SW is particularly important given its influential role in both MODIS-GPP and JULES-SF. As in Heinsch et al, we find that implementation of reanalysis generally increases simulated GPP compared to tower-based simulations, due to either a regional overestimation of SW (forecast models used to reconstruct meteorology are generally too transparent) and a general underestimation of VPD (lower humidity stress).

4.4 Caveats and Limitations of Current Study

1. Our validation of the default simulation includes $\simeq$30% siteyears where site normalisation of LAI phenology was not possible. Removal of these unnormalised siteyears changes the model bias against observed carbon fluxes in Fig. 4 by only a modest amount (3%).

2. Using only the highest quality (=1) field LAI measurements in our sensitivity analysis (only possible for 6 PFTs), rather than quality$\leq$3, produces the same median sensitivities to LAI forcing and meteorology as Tab. 6. However, sensitivity to model formulation increases moderately from 17% to 29%. A similar conclusion is drawn if the mean, rather than the median, is adopted when averaging the model sensitivity across PFTs in Tab. 6 (with the original quality$\leq$3 field LAI measurements).

3. Our inferences for sensitivity depend on how we define the current uncertainties in LAI and meteorological forcing. For meteorology, as discussed above, the reanalysis bias against ground-based observations has decreased with the release of improved datasets. For LAI sensitivity, our test entails a small change in LAI, at least for some PFTs e.g. ΔLAI = -0.2 m^2 m^{-2} for non-tropical broadleaf forest (Tab. 6). Moreover, our perturbation only changes the amplitude and not the phase of the

timeseries. Some models adopt a constant (field-based) LAI across the growing season (e.g. Medlyn et al 2005). By adopting this approach as our perturbed LAI phenology, ΔLAI increases from 0.6 to 0.9 in Tab. 6 (median of absolute PFT means). However, the corresponding average increase in simulated GPP is 10% for both models i.e. close to the change in the original sensitivity experiment (8-12%). The increase in GPP is quite modest for both models owing to physiological limitations on photosynthesis outside the growing season (e.g. low temperature). Some dynamic vegetation global models generate LAI internally. The intramodel variability in GPP (10-20%; Richardson et al 2012), owing to uncertainties in this prognostic LAI, is comparable to the sensitivity to LAI forcing derived in the current study.

4. The uncertainties owing to forcing and model formulation are significant compared to the impact of anthropogenic greenhouse gases ($\sim$5% global GPP; Houghton 2007). However, they are only moderate compared to some other sources of modelling uncertainty e.g. bias of observational datasets used to calibrate carbon models (Tab. 8).

5 Summary and Conclusions

We have driven two carbon models (a simple LUE algorithm, MODIS-GPP, and a complex ecophysiological land-surface model, JULES-SF) with two LAI forcings (satellite and site-normalised) and two meteorologies (tower-based and reanalysis) and compared the simulated output in GPP. These experiments have been conducted for 67 sites and 10 PFTs in order to determine model sensitivity to LAI and meteorological forcing. Output from the default simulation, using site-normalised LAI forcing and tower-based meteorology, was also compared with GPP inferred from observed eddy covariance carbon fluxes. Our sensitivity experiment for LAI forcing allowed us to compare satellite (MODIS) LAI with field-based measurements.

Our conclusions are as follows:

1. For both models, the sensitivity to LAI and meteorological forcing is 10-20% (change in simulated GPP), which is comparable to the sensitivity owing to model formulation i.e. LUE versus ecophysiological (20% change in GPP).

2. Compared to the LUE algorithm, the ecophysiological model is more sensitive to meteorology (20% versus 10% change in GPP) owing to the reliance of JULES-SF on precipitation and SW to calculate,

respectively, the internal water and energy balances (important for drought stress and leaf temperature).

3. For MODIS-GPP, the average uncertainty owing to meteorology (10% GPP) is less than that previously found (15-25% GPP; Heinsch et al 2006; Zhao et al 2006) owing to the improving accuracy of reanalysis meteorology compared to ground-based observations. For some regions, however, we find uncertainties are larger (e.g. 30% GPP for tropical broadleaf trees) owing to substantial errors in SW and VPD reanalysis.

4. Despite their disparity in complexity, both models underestimate flux-derived observational GPP for the more productive FLUXNET sites (by 10-30% for the 50% most productive sites and by 50-100% for tropical broadleaf trees and C3 crops). This bias, which cannot be attributed to forcing uncertainties, is shared with a large range of other models, possibly indicating a general inadequacy in land-surface and carbon modelling. However, there are also inconsistencies of a comparable magnitude between the observational datasets used to validate, and potentially calibrate, the models (e.g. FLUXNET versus Ecosystem Model-Data Intercomparison sites).

5. Although MODIS-GPP possesses a greater bias than JULES-SF against flux-derived observational GPP, it produces less scatter. This suggests that, once adequately calibrated, the LUE approach may allow acceptable simplification of the complex process of canopy photosynthesis.

6. Satellite measurements of growing season LAI, based on the latest (Collection 5) MODIS product (MOD15A2), underestimate field-based estimates by 10-25%. Underestimation is more pronounced for grasses and crops.

Acknowledgements

We are grateful to the PIs and Co-Is of FLUXNET who make their data freely available to the ecological modelling community through the FLUXNET archive(http://fluxnet.ornl.gov/). These contributors include, but are not limited to, Falge, E., M. Aubinet, P. Bakwin, D. Baldocchi, P. Berbigier, C. Bernhofer, A. Black, R. Ceulemans, A. Dolman, A. Goldstein, M. Goulden, A. Granier, D. Hollinger, P. Jarvis, N. Jensen, K. Pilegaard, G. Katul, P. Kyaw Tha Paw, B. Law, A. Lindroth, D. Loustau, Y. Mahli, R. Monson, P.Moncrieff, E. Moors, W. Munger, T. Meyers, W. Oechel, E. Schulze, H. Thorgeirsson, J. Tenhunen, R.

Valentini, S. Verma, T. Vesala, and S. Wofsy.

References

Abramowitz, G., Leuning, R., Clark, M., Pitman, A. (2008) Evaluating the performance of land surface models *Journal of Climate*, 21, 5468-5481

Abuelgasim, A., Fernandes, R., Leblanc, S. (2006) Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1872-1884

Agarwal, D., (ed.) 2012 Biological and Ancillary Data for FLUXNET sites. Data set. Available on-line [http://www.fluxdata.org]

Alton (2013) From site-level to global simulation: reconciling carbon, water and energy fluxes over different spatial scales using a process-based ecophysiological land-surface model *Agricultural and Forest Meteorology*, 176, 111-124

Alton, P. and Bodin, P. (2010) A comparative study of a multilayer and a productivity (light-use) efficiency land-surface model over different temporal scales *Agricultural and Forest Meteorology*, 150, 182-195

Alton, P, Ellis, R., Los, S., North, P. (2007) Improved global simulations of Gross Primary Product based on a separate and explicit treatment of diffuse and direct sunlight. *JGR*, 112, D07203

Aragao, L, Shimabukuro, Y., Espirito-Santo, F., Williams, M. (2005) Spatial validation of the Collection 4 MODIS LAI product in Eastern Amazonia *IEEE Transactions in Geoscience and Remote Sensing*, 43, 2526-2534

Asner, G., Scurlock, J., Hicke, J. (2003) Global synthesis of leaf area index observations: implications for ecological and remote sensing studies *Global Ecology & Biogeography* 12, 191-205

Baldocchi, D. & Harley, P. (1995) Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. II Model testing and application *Plant, Cell and Environment* 18, 1157-1173

598

599 Baldocchi, D., Wilson, K. (2001) Modelling CO_2 and water vapour exchange of a temperate broadleaved
600 forest across hourly to decadal time scales *Ecological Modelling*, 142, 155-184

601

602 Ball, J., Woodrow, E., Berry, J. (1987) A model predicting stomatal conductance and its contribution to the
603 control of photosynthesis under different environmental conditions *In* Progress in Photosynthesis Research
604 Ed. Biggins, J., Nijhoff, M., 221-224, *Dordrecht, Netherlands*

605

606 Beer, C, Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Roedenbeck, C., Arain, M.,
607 Baldocchi, D., Bonon, G., Bondeau,A., Cescatti., A., Lasslop, G., Lindroth, A., Lomas., M., Luyssaert, S.,
608 Margolis, H., Oleson, K., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F., Papale,
609 D. (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate *Science*,
610 329, 834-838

611

612 Campbell, B. and Norman, J. (1998) Environmental Biophysics *Ed. Springer-Verlag, New York*

613

614 Collatz, G., Ball, J., Grivet, C., Berry, J. (1991) Physiological and environmental regulation of stomatal con-
615 ductance, photosynthesis and transpiration: a model that includes laminar boundary layer *Agric.For.Meteorol.*
616 54, 107-136

617

618 Cox, P., Betts, R., Bunton, C., Essery, R., Rowntree, P., Smith, J. (1999) The impact of new land surface
619 physics on the GCM simulation of climate and climate sensitivity *J.Climate Dynamics* 15, 183-203

620

621 Cramer, W., Bondeau, A., Woodward, F. et al (2001) Global response of terrestrial ecosystem structure
622 and function to CO_2 and climate change: results from six dynamic global vegetation models *Global Change*
623 *Biology* 7, 357-373

624

625 DeLucia, E., Drake, J., Thomas, R., Gonzalez-Meler, M., (2007) Forest carbon use efficiency: is respiration
626 a constant fraction of gross primary production ? *Global Change Biology*, 13, 1157-1167

627

De Kauwe, M., Disney, M., Quaife, T., Lewis, P., Williams, M., (2011) An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest *Remote Sensing of Environment*, 115, 767-780

Desai, A.R., Richardson, A.D., Moffat, A.M., Kattge, J., Hollinger, D.Y., Barr, A., Falge, E., Noormets, A., Papale, D., Reichstein, M., Stauch, V.J., 2008. Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. Agricultural and Forest Meteorology 148, 821-838

Dirmeyer, P. (2011) A history and review of the Global Soil Wetness Project (GSWP) *J.Hydrometeor.*, 12, 729-749

Falge, E., Tenhunen, J., Baldocchi, D., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, C., Bonnefond, J-M., Burba, G., Clement, R., Davis, K., Elbers, J., Falk, M., Goldstein, A., Grelle, A., Granier, A., Gruenwald, T., Guomundsson, J., Hollinger, D., Janssens, I., Keronen, P., Kowalski., A., Katul, G., Law, B., Malhi, Y., Meyers, T., Monson, R., Moors, E., Munger, J., Oechel, W., Kyaw Tha Paw, U., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Thorgeirsson, H., Tirone, G., Turnipseed, A., Wilson, K., Wofsy, S. (2002) Phase and amplitude of ecosystem carbon release and uptake potential as derived from FLUXNET measurements *Agricultural and Forest Meteorology*, 113, 75-95

Fang, H., Wei, S., Liang, S. (2012) Validation of MODIS and CYCLOPES LAI products using global field measurement data *Remote Sensing of Environment*, 119, 43-54

Fisher, J., Tu, K., Baldocchi, D. (2008) Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites *Remote Sensing of Environment*, 112, 901-919

Foken, T. (2008) The energy balance closure problem – an overview *Ecological Applications*, 18, 1351-1367

Friend, A., Arneth, A., Kiang, N., Lomas, M., Ogee, J., Roedenbeck, C., Running, S., Santaren, J., Sitch, S., Viovy, N., Woodward, F., Zaehle, S., (2007) FLUXNET and modelling the global carbon cycle *Global Change Biology*, 13, 610-633

Goulden M., Munger, J., Fan, S., Daube, B., Wofsy, S. (1996) Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy *Global Change Biology* 2, 169-182

Heinsch, F., Zhao, M., Running, S. (2006) Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1908-1925

Hickler, T., I.C Prentice, B. Smith, M.Sykes, S.Zaehle (2006) Implementing plant hydraulic architecture within the LPJ Dynamic Global Vegetation Model *Global ecology and biogeography*, 15, 567-577

Hollinger, D., Richardson, A. (2005) Uncertainty in eddy covariance measurements and its application to physiological models *Tree Physiology*, 25, 873-885

Houghton, R. (2007) Balancing the global carbon budget *Annu.Rev.Earth Planet.Sci.*, 35, 313-347

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jackson, R., Canadell, J., Ehleringer, J., Mooney, H., Sala, O., Schulze, E. (1996) A global analysis of root distributions for terrestrial biomes *Oecologia*, 108, 389-411

Jung, M., Vetter, M., Herold, M., Churkina, G., Reichstein, M., Zaehle, S., Ciais, P., Viovy, N., Bondeau, A., Chen, Y., Trusilova, K., Feser, F., Heimann, M., (2007) Uncertainties of modelling gross primary productivity over Europe: a systematic study on the effects of using different drivers and terrestrial biosphere models *Global Biogeochemical Cycles*, 21, GB4021

Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bern-

688 hofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E.,
689 Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams,
690 C. (2011) Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat de-
691 rived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research -*
692 *Biogeosciences*, 116, doi:10.1029/2010JG001566

693

694 Kattge, J., Knorr, W., Raddatz, T., Wirth, C. (2009) Quantifying photosynthetic capacity and its relation-
695 ship to leaf nitrogen content for global-scale terrestrial biosphere models *Global Change Biology*, 15, 976-991

696

697 Knorr., W. and Heimann, M. (2001) Uncertainties in global terrestrial biosphere modeling, Part I: A com-
698 prehensive sensitivity analysis with a new photosynthesis and energy balance scheme *Global Biogeochemical*
699 *Cycles, 15*, 207-225

700

701 Law, B., Ryan, M., Anthoni, P. (1999) Seasonal and annual respiration of a ponderosa pine ecosystem *Global*
702 *Change Biology*, 5, 169-182

703

704 Lewis, J., McKane, R., Tingey, D., Beedlow, P. (2000) Vertical gradients in photosynthetic light response
705 within an old-growth douglas-fir and western hemlock canopy *Tree Physiology* 20, 447

706

707 Luyssaert, S., Inglima, I., Jung, M., A. D. Richardson, M. Reichstein, D. Papale, S. L. Piao, E.-D. Schulze,
708 L. Wingate, G. Matteucci, L. Aragao,M. Aubinet,C. Beers,C. Bernhoffer,K. G. Black,D. Bonal,J. -M. Bonne-
709 fond,J. Chambers,P. Ciais,B. Cook, K. J. Davis,A. J. Dolman,B. Gielen,M. Goulden,J. Grace,A. Granier,A.
710 Grelle,T. Griffis,T. Gruenwald,G. Guidolotti, P. J. Hanson,R. Harding, D. Y. Hollinger,L. R. Hutyra,P. Ko-
711 lar,B. Kruijt,W. Kutsch,F.Lagergren,T. Laurila,B. E. Law, G. Le Maire,A.Lindroth,D. Loustau,Y. Malhi,J.
712 Mateus,M. Migliavacca,L. Misson,L. Montagnani,J. Moncrieff, E. Moors,J. W. Munger,E. Nikinmaa,S. V.
713 Ollinger,G. Pita,C. Rebmann,O. Roupsard,N. Saigusa,M. J. Sanz,G. Seufert, C. Sierra,M. -L. Smith,J. Tang,
714 R. Valentini,T. Vesala, I. A. Janssens (2007) CO_2 balance of boreal, temperate and tropical forests derived
715 from a global database *Global Change Biology*, 13, 2509-2537

716

717 Masseroni, D., Corbari, C., Mancini, M., (2014) Limitations and improvements of the energy balance closure

with reference to experimental data measured over a maize field *Atmósfera*, 27(4), 335-352

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obsteiner, M., Fritz, S., Nilsson, S. (2009) Satellite-based terrestrial production efficiency modelling *Carbon Balance and Management*, 4, 1-14

Medlyn, B., Barrett, D., Landsberg, J., Sands, P., Clement, R. (2003) Conversion of canopy-intercepted radiation to photosynthate: review of modelling approaches for regional scales *Functional Plant Biology* 30, 153-169

Medlyn, B., Robinson, A., Clement, R., McMurtie, E. (2005) On the validation of models of forest CO_2 exchange using eddy covariance data: some perils and pitfalls *Tree Physiology* 25, 839-857

Meir, P., Kruijt, B., Broadmeadow, M. et al (2002) Acclimation of photosynthetic capacity to irradiance in tree canopies in relation to leaf N concentration and leaf mass per unit area *Plant, Cell and Environment* 25, 343-357

Melaas, E., Richardson, A., Friedl, M., Dragonic, D., Gough, C., Herbst, M., Montagnanif, L., Moors, E (2013) Using FLUXNET data to improve models of springtime vegetation activity onset in forest ecosystems *Agricultural and Forest Meteorology* 171-172, 46-56

Monteith, J.L. (1965) Evaporation and environment. *Symp. Soc. Exptl. Biol.* 19, 205-234.

Müller, C., and W. Lucht (2007), Robustness of terrestrial carbon and water cycle simulations against variations in spatial resolution, *J. Geophys. Res., [Atmos.]*, 112, 6105

Olson, R. J., J. M. O. Scurlock, S. D. Prince, D. L. Zheng, and K. R. Johnson (eds.). 2008. NPP Multi-Biome: NPP and Driver Data for Ecosystem Model-Data Intercomparison, R[evision]1. Data set. Available on-line [http://www.daac.ornl.gov] from the Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

748 Puma, M., Koster, R., Cook, B. (2013) Phenological versus meteorological controls on land-atmosphere
749 water and carbon fluxes *JGR*, 118, 14-29

750

751 Quaife, T., Quegan, S., Disney, M., Lewis, P., Lomas, M., Woodward, F. (2008) Impact of land cover un-
752 certainties on estimates of biospheric carbon fluxes *Global Biogeochemical Cycles*, 22, GB4016

753

754 Reichstein, M., Tenhunen, J, Roupsard, O., Ourcival J.M., Rambal S., Miglietta F., Peressotti A., Pecchiari
755 M., Tirone G., Valentini R.(2003) Inverse Modelling of seasonal drought effects on canopy CO_2/H_2O ex-
756 change in three Mediterranean ecosystems *JGR* 108, D23, 4726

757

758 Richardson, A.D., R.S. Anderson, M.A. Arain, A.G. Barr, G. Bohrer, G. Chen, J.M. Chen, P. Ciais, K.J.
759 Davis, A.R. Desai, M.C. Dietze, D. Dragoni, S.R. Garrity, C.M. Gough, R. Grant, D.Y. Hollinger, H.A.
760 Margolis, H. McCaughey, M. Migliavacca, R.K. Monson, J.W. Munger, B. Poulter, B.M. Raczka, D.M.
761 Ricciuto, A.K. Sahoo, K. Schaefer, H. Tian, R. Vargas, H. Verbeeck, J. Xiao, and Y. Xue. 2012. Terrestrial
762 biosphere models need better representation of vegetation phenology: Results from the North American
763 Carbon Program site synthesis. Global Change Biology, 18: 566-584.

764

765 Running, S., Baldocchi, D., Turner, D., Gower, S., Bakwin, P., Hibbard, K., (1999) A global terrestrial
766 monitoring network integrating tower fluxes, flask sampling, ecosystem modelling and EOS satellite data
767 *Remote Sensing of Environment*, 70, 108-127

768

769 Sellers, P., Randall, D., Collatz, G., et al (1996) A revised land surface parameterization (SiB2) for atmo-
770 spheric GCMs. Part I: Model formulation *Journal of Climate* 9, 676-705

771

772 Sheffield, J., G. Goteti, and E. Wood (2006), Development of a 50-year high-resolution global dataset of
773 meteorological forcings for land surface modeling, *J. Clim.*, 19, 3088–3111.

774

775 Sheffield, J., Wood,E., Roderick, M. (2012) Little change in global drought over the past 60 years *Nature*,
776 491, 435-438

777

778 Valentini, R., Matteucci G., Dolman A. J., Schulze E. - D., Rebmann C., Moors E. J., Granier A., Gross P.,
779 Jensen N. O., Pilegaard K., Lindroth A., Grelle A., Bernhofer C., Grnwald T., Aubinet M., Ceulemans R.,
780 Kowalski A. S., Vesala T., Rannik ., Berbigier P., Loustau D., Gumundsson J., Thorgeirsson H., Ibrom A.,
781 Morgenstern K., Clement R., Moncrieff J., Montagnani L., Minerbi S., & Jarvis P. G. (2000) Respiration as
782 the main determinant of carbon balance in European forests *Nature*, 404, 861-865

783

784 Vinukolla, R., Meynadier, R., Sheffield, J., Wood, E. (2011) Multi-model, multi-sensor estimates of global
785 evapotranspiration: climatology, uncertainties and trends *Hydrological Processes*, 25, 3993-4010

786

787 Wang, Y.P., Baldocchi, D., Leuning, R., Falge, E., Vesala, T., (2007) Estimating parameters in a land-
788 surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET
789 sites *Global Change Biology*, 13, 652-670

790

791 Williams, M., Rastetter, E., Fernandes, D. et al (1996) Modelling the soil-plant-atmosphere continuum in
792 a *Quercus-Acer* stand at Harvard Forest; the regulation of stomatal conductance by light, nitrogen and
793 soil/plant hydraulic properties *Plant, Cell and Environment* 19, 911-927

794

795 Wilson, K., Goldstein, A., Falge, E., Aubinet M., Baldocchi D., Berbigier P., Bernhofer C., Ceulemans R.,
796 Dolman H., Field C., Grelle A., Ibrom A., Law B.E., Kowalski A., Meyers T., Moncrieff J., Monson R.,
797 Oechel W., Tenhunen J., Valentini R., Verma S. (2002) Energy balance closure at FLUXNET sites *Agricul-*
798 *tural & Forest Meteorology*, 113, 223-243

799

800 Wright, I., Reich, P., Westoby, M., Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cor-
801 nelissen JH, Diemer M, Flexas J, Garnier E, Groom PK, Gulias J, Hikosaka K, Lamont BB, Lee T, Lee W,
802 Lusk C, Midgley JJ, Navas ML, Niinemets U, Oleksyn J, Osada N, Poorter H, Poot P, Prior L, Pyankov
803 VI, Roumet C, Thomas SC, Tjoelker MG, Veneklaas EJ, Villar R. (2004) The worldwide leaf economics
804 spectrum *Nature*, 428, 821-827

805

806 Xiao,J.F., Q.L. Zhuang, B.E. Law, D.D. Baldocchi, J.Q. Chen, A.D. Richardson, J.M. Melillo, K.J. Davis,
807 D.Y. Hollinger, S. Wharton, R. Oren, A. Noormets, M.L. Fischer, S.B. Verma, D.R. Cook, G. Sun, S.

McNulty, S.C. Wofsy, P.V. Bolstad, S.P. Burns, P.S. Curtis, B.G. Drake, M. Falk, D.R. Foster, L.H. Gu, J.L. Hadley, G.G. Katulk, M. Litvak, S.Y. Ma, T.A. Martinz, R. Matamala, T.P. Meyers, R.K. Monson, J.W. Munger, W.C. Oechel, K.T. Paw U, H.P. Schmid, R.L. Scott, G. Starr, A.E. Suyker, M.S. Torn (2011) Assessing net ecosystem carbon exchange of U.S. terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations *Agric. For. Meteorol.*, 151 (2011), 60-69

Yang, W, Shabanov, NV, Huang, D, Wang, W, Dickinson, RE, Nemani, RR, Knyazikhin, Y, Myneni, RB (2006). Analysis of leaf area index products from combination of MODIS Terra and Aqua data. *Remote Sensing of Environment*, 104(3), 297-312

Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A., Goulden, M., Hollinger, D., Hu, Y., Law, B., Stoy, P., Vesala, T., Wofsy, S. (2007) Deriving a light-use efficiency model from eddy covariance covariance flux data for predicting daily gross primary production across biomes *Agricultural and Forest Meteorology* 143, 189-207

Yuan, W., Liu, S., Yu, G., Bonneford, J-M., Chen, J., Davis, K., Desai, A., Goldstein, A., Gianelle, D., Rossi, F., Suyker, A., Verma, S. (2010) Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data *Remote Sensing of Environment*, 114, 1416-1431

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F. (2005) Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochem. Cycles, 19, GB3020

Zhao, M., Running, S., Nemani, R. (2006) Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses *JGR*, 111, G01002

Zhao, M., Running, S., (2010) Drought-induced reduction in global terrestrial net primary production from 2000 through 2009 *Science*, 329, 940-943

Table 1: An alphabetical list of acronyms and abbreviations used in the main text. Units are given where appropriate.

	Definition
fPAR	fraction of Photosynthetically Active Radiation
GPP	Gross Primary Productivity ($gCm^{-2}d^{-1}$)
JULES-SF	Joint UK land environmental simulator
LAI	Leaf Area Index (m^2 m^{-2})
LUE	Light-Use Efficiency
MODIS	Moderate Resolution Imaging Spectroradiometer
NEE	Net Ecosystem Exchange
NPP	Net Primary Productivity
PFT	Plant Functional Type
SW	downwelling Short Wave radiation (W m^{-2})
T_{air}	Air temperature (K)
VPD	Vapour Pressure Deficit (kPa)

Table 2: Comparison of the salient features of each model. Driving meteorology is denoted as follows: shortwave radiation (SW), Vapour Pressure Deficit (VPD), air temperature (T_{air}), longwave radiation (LW), precipitation (PPT), wind speed (WS), pressure (P) and specific humidity (Q). Driving phenology is denoted as Leaf Area Index (LAI) and fraction of photosynthetically active radiation (fPAR). Note that fPAR is derived from multispectral reflectance in the standard MODIS GPP product but in the present study it is derived from LAI. The primary model parameters are top-of-canopy maximum carboxylation rate or photosynthetic capacity (V^0_{cmax}) and the maximum light-use efficiency (ϵ_{max}). N refers to leaf nitrogen.

	MODIS-GPP	JULES-SF
Meteorology	SW, daytime VPD, minimum T_{air}	SW, LW, PPT, T_{air}, WS, P, Q
Phenology	fPAR (via reflectances or LAI)	LAI
Primary Parameter	ϵ_{max}	V^0_{cmax}
Photosynthesis	light-use efficiency with ramp stress functions for stress owing to T_{air} and VPD	co-limited by SW, Rubisco (N) concentration, enzyme kinematics (sensitive to canopy temperature), water availability in root zone and stomatal conductance (sensitive to relative humidity)
Canopy Structure	none	1-D but accounting for sunlit-shade foliage and declining Rubisco with decreasing height in canopy
Energy Balance	none	Penman-Monteith to determine surface fluxes and canopy temp
Water Balance	none	full account of PPT, evapotranspiration, runoff and changes in soil moisture
Output	daily GPP	3hr carbon,water and energy fluxes including GPP

Table 3: Key parameters adopted for each model: top-of-canopy maximum carboxylation rate or photosynthetic capacity (V^0_{cmax}; μmol m^{-2} s^{-1}) for JULES-SF and maximum light-use efficiency (ϵ_{max}; gCm^{-2} d^{-1} MJ^{-1}) for MODIS-GPP. Parameters are assigned according to Plant Functional Types (PFT) as defined in JULES-SF. The corresponding abbreviation for PFT (Desig.) is adopted in subsequent figures and tables. Design.(UMD) defines the corresponding University of Maryland land cover classification which is conventionally adopted with MODIS-GPP (Zhao & Running 2010). The number of sites and siteyears in the present study is given by n$_{site}$ and n$_{siteyr}$, respectively.

Plant Functional Type	Desig.	Desig.(UMD)	n$_{siteyr}$	n$_{site}$	V^0_{cmax}	ϵ_{max}
Non-tropical Broadleaf Forest	BL	DBF	92	15	52	1.165
Non-Mediterranean Needleleaf Forest	NL	ENF	96	18	59	0.962
C3 crop	Cr3	Crop	59	4	95	1.044
C4 crop	Cr4	Crop	2	1	28	1.044
Tundra Shrub	Tu	OShrub	19	4	45	0.841
Tropical Broadleaf Forest	TBL	EBF	15	5	41	1.268
C3 Grass	C3	Grass	54	10	76	0.860
C4 Grass	C4	Grass	19	4	28	0.860
Non-Tundra Shrub	SH	CShrub	12	2	51	1.281
Mediterranean Needleleaf Forest	MNL	ENF	36	4	61	0.962

Table 4: Comparison of satellite MOD15A2 Leaf Area Index (LAI) against measurements from the site LAI database (Agarwal 2012). For each Plant Functional Type (PFT), the mean and Standard Deviation (SD) are given. ΔLAI(RMS) is the Root Mean Square (RMS) difference between satellite and field measurements. The number of measurements, n, is generally less than the number of siteyears used in the sensitivity experiments owing to the removal of duplicates where the same site LAI is adopted across multiple siteyears. To validate LAI, we also include values for savanna (SAV) and mixed forest (MX). The median of the PFT means is indicated in the bottom row (median($\bar{x}$)). Where possible, comparison is made with the field measurements compiled by Asner et al (2003).

PFT	n	Site LAI mean(SD) (m^2 m^{-2})	Satellite LAI mean(SD) (m^2 m^{-2})	ΔLAI(RMS) (m^2 m^{-2})	Asner et al mean(SD) (m^2 m^{-2})
BL	45	4.4(1.1)	4.3(1.2)	1.3	5.1(1.6)
NL	31	4.7(2.3)	3.1(1.0)	2.7	5.7(3.0)
Cr3	34	4.7(1.1)	2.1(0.6)	2.8	3.6(2.1)
Cr4	1	5.2(–)	3.4(–)	–	3.6(2.1)
Tu	5	1.4(0.2)	0.7(0.4)	0.8	1.9(1.5)
MX	12	2.7(1.3)	4.8(0.9)	2.5	–(–)
TBL	6	5.2(0.3)	6.3(0.4)	1.2	4.8(1.7)
C3	20	2.3(0.6)	1.6(0.9)	1.3	1.7(1.2)
C4	8	2.3(0.8)	1.5(0.7)	0.8	1.7(1.2)
SH	4	1.8(1.6)	2.5(2.2)	1.1	2.1(1.6)
SAV	12	1.3(0.5)	1.6(0.4)	0.6	–(–)
MNL	10	3.6(1.5)	3.6(1.6)	1.0	5.5(3.4)
median($\bar{x}$)	11	3.1(1.1)	2.8(0.9)	1.2	3.6(1.6)

Table 5: Validation of simulated GPP from MODIS-GPP and JULES-SF against observation-based estimates from eddy covariance fluxes (GPP(obs)). RMSE, SD and MEF are, respectively, the Root Mean Square Error, Standard Deviation and Modelling Efficiency. MEF is similar to the coefficient of determination (r^2) but takes account of model bias (Medlyn et al 2003). For C4 crops and tundra the sample size is too small to derive the RMSE and MEF. The median of the PFT means is indicated in the bottom row (median($\bar{x}$)).

PFT	GPP(obs)	MODIS-GPP			JULES-SF		
	mean(SD) ($gCm^{-2}d^{-1}$)	mean(SD) ($gCm^{-2}d^{-1}$)	RMSE ($gCm^{-2}d^{-1}$)	MEF (–)	mean(SD) ($gCm^{-2}d^{-1}$)	RMSE ($gCm^{-2}d^{-1}$)	MEF (–)
BL	6.5(2.0)	5.1(1.8)	1.9	0.09	6.5(2.0)	1.1	0.69
NL	4.7(2.7)	4.5(1.8)	2.1	0.43	5.1(1.9)	2.2	0.34
Cr3	9.0(4.9)	6.0(1.8)	4.6	0.10	7.5(3.4)	4.6	0.11
Cr4	8.1(–)	5.5(–)	–	–	10.7(–)	–	–
Tu	1.5(–)	2.8(–)	–	–	3.0(–)	–	–
TBL	13.1(3.6)	9.9(3.0)	3.9	-0.17	8.5(1.2)	5.4	-1.18
C3	2.9(2.3)	4.3(2.3)	2.2	0.06	3.6(3.1)	2.0	0.22
C4	3.0(2.7)	3.5(1.5)	1.6	0.63	6.5(5.0)	4.3	-1.60
SH	4.5(3.2)	7.8(2.8)	3.5	-0.24	5.0(2.4)	3.0	0.18
MNL	6.4(4.2)	5.9(2.4)	3.0	0.49	6.6(3.7)	2.5	0.64
median($\bar{x}$)	5.5(3.0)	5.3(2.0)	2.6	0.10	6.5(2.8)	2.8	0.20

Table 6: Sensitivity of simulated GPP to uncertainty in LAI and meteorological forcing. Column 1 is the Plant Funct
maining columns contain mean values averaged over all siteyears comprising any given PFT, except the last row which
absolute PFT means (median($|\bar{x}|$)). ΔLAI is the average change in Leaf Area Index (LAI). MOD and JUL refer to resu
JULES-SF, respectively. ΔGPP(LAI-DEF) is the difference between the LAI-perturbed simulation (GPP(LAI_PERT
ulation (GPP(DEF)). Similarly, ΔGPP(MET-DEF) is the difference between the meteorology-perturbed simulation
and GPP(DEF). The right most column shows the sensitivity of simulated GPP to model. Thus, GPP(MOD) is GI
and ΔGPP(MOD-JUL) is the difference between GPP(MOD) and the default simulation for JULES-SF.

PFT	ΔLAI (m^2m^{-2})	GPP(DEF) ($gCm^{-2}d^{-1}$)		$\frac{\Delta GPP(LAI-DEF)}{\Delta LAI}$ ($gCm^{-2}d^{-1}$ $[m^2\ m^{-2}]^{-1}$)		$\frac{\Delta GPP(LAI-DEF)}{GPP(DEF)}$ (%)		$\frac{\Delta GPP(MET-DEF)}{GPP(DEF)}$ (%)		$\frac{\Delta GPP(M}{GPP}$ (
		MOD	JUL	MOD	JUL	MOD	JUL	MOD	JUL	
BL	-0.2	4.3	5.7	0.2	0.3	-1.0	-1.2	22.1	16.6	3
NL	-1.3	3.4	3.9	0.1	0.1	-2.1	-4.4	11.2	35.5	1
Cr3	-1.3	4.2	5.7	0.8	1.1	-22.4	-23.3	2.9	8.6	3
Cr4	-0.7	4.1	8.0	0.8	1.5	-13.0	-12.6	7.3	20.7	9
Tu	-0.5	1.4	1.4	1.0	1.3	-34.7	-42.2	29.5	32.4	
TBL	1.2	7.3	6.7	-0.0	0.0	-0.1	0.2	21.2	26.4	-
C3	-0.4	3.4	3.6	1.3	0.5	-15.1	-5.6	5.7	26.4	
C4	-0.2	2.3	4.2	1.6	1.9	-16.5	-10.5	-11.1	-2.8	8
SH	1.0	4.3	3.5	0.5	0.4	10.7	10.6	6.5	25.1	-
MNL	-0.5	4.4	3.7	0.4	0.4	-5.2	-5.2	-5.1	2.0	-
median($\lvert\bar{x}\rvert$)	0.6	4.1	4.0	0.6	0.5	11.8	8.0	9.2	22.9	1

Table 7: Sensitivity of MODIS-GPP (MOD) and JULES-SF (JUL) to individual meteorological variables mean averaged over all relevant siteyears for each Plant Functional Type (PFT). GPP(DEF) represents Gross Primary Productivity (GPP) from the default simulation. ΔGPP(MET-DEF) is the difference in GPP between the meteorology-perturbed simulation and the default simulation. Meteorological forcing variables are perturbed in turn and are denoted as follows: downwelling shortwave radiation (SW), downwelling longwave radiation (LW), precipitation (PPT), air temperature (T_{air}), windspeed (WS), pressure (P), specific air humidity (Q). Although MODIS-GPP is forced by minimum air temperature and daytime vapour pressure deficit, we show the sensitivity of both models to T_{air} and Q in order to permit a comparison. The last row contains the median of the absolute PFT means (median($|\bar{x}|$)).

PFT	$\frac{\Delta GPP(MET-DEF)}{GPP(DEF)}$ [%]													
	SW		LW		PPT		T_{air}		WS		P		Q	
	MOD	JUL	MOD	JUL	MOD	JUL	MOD	JUL	MOD	JUL	MOD	JUL	MOD	JUL
BL	8	11	0	1	0	2	-17	0	0	0	0	0	18	7
NL	5	12	0	2	0	5	0	3	0	-2	0	1	4	13
Cr3	2	13	0	1	0	-8	-6	-1	0	0	0	0	4	9
Cr4	8	11	0	1	0	-1	-8	1	0	2	0	0	3	10
Tu	25	15	0	7	0	-6	3	11	0	1	0	0	-1	1
TBL	13	15	0	-1	0	8	-16	-4	0	2	0	0	7	15
C3	1	12	0	3	0	7	-2	1	0	-1	0	0	6	9
C4	8	15	0	1	0	-11	-18	-5	0	-1	1	3	-1	-4
SH	6	12	0	3	0	13	-8	-3	0	1	0	0	4	3
MNL	13	8	0	1	0	-10	-30	-8	0	1	1	6	9	13
median($\|\bar{x}\|$)	8	12	0	1	0	8	8	3	0	1	0	0	4	9

Table 8: Categories of uncertainty, to the nearest 5%, for carbon fluxes (GPP and NPP) at site, regional and global level. Categories are approximately ordered with greatest uncertainties at the top. For each category, a range of uncertainty is given according to the cited studies. The bias introduced into the model by calibrating against FLUXNET, rather than against Ecosystem Model-Data Intercomparison observations, is estimated by comparing the model bias in Fig. 4 with that in Fig. 5. PFT is Plant Functional Type and LAI is Leaf Area Index.

Category	Uncertainty (%)	Studies
Bias owing to calibration observations	45	current
Model formulation (process complexity)	20-25	current; Cramer et al (2001) Knorr & Heimann (2001)
LAI and meteorological drivers	10-20	current
Biophysical parameterisation	15	Zaehle et al (2005)
PFT classification and land cover	0-10	Quaife et al (2008); Jung et al (2007)
Spatial resolution of global simulation	5	Mueller & Lucht (2007)

Figure Captions:

Fig.1: A schematic overview of model input/output for the simulations. Sensitivity experiments are conducted separately for each siteyear within a Plant Functional Type (PFT) and are denoted by DEF (default), LAI_PERTURB (LAI perturbed) and MET_PERTURB (meteorology perturbed). Gross Primary Productivity (GPP; $gCm^{-2}d^{-1}$) for the default experiment (GPP(DEF)) is validated against observed eddy covariance (EC) fluxes. ΔGPP is the difference between the perturbed simulation and DEF.

Fig.2: Site locations used in the current study, based on open-access FLUXNET data. For clarity, sites are coarsely categorised (tree, grass/crop and shrub) although 10 PFTs are used in the simulations.

Fig.3: Field-based maximum Leaf Area Index (site LAI), measured for a given siteyear, compared against the corresponding satellite (MOD15A2) measurement. The dashed curve shows a least-squares exponential fit a - c exp(-x/b) excluding the 3 outliers at x<2, y>5 (a=4.32, b=2.59 and c=4.53). To validate LAI, we also include siteyears for savanna (SAV) and mixed forest (MX) which are not simulated in the sensitivity experiment.

Fig.4: Daily Gross Primary Productivity (GPP) derived from the MODIS-GPP algorithm (bottom) and from JULES-SF (top) compared against observationally based GPP from eddy covariance fluxes. For this validation exercise, both models use the default meteorological and LAI forcing. Each point represents an average for the siteyear. However, GPP is expressed as a daily average to reduce the impact of data gaps across the annual cycle. In each case, the dashed curve shows a least-squares exponential fit a - c exp(-x/b), with a=17.50, b=24.88 and c=15.41 for MODIS-GPP and a=10.94, b=7.35 and c=9.95 for JULES-SF.

Fig.5: Modelled annual Gross Primary Productivity (GPP) for FLUXNET sites, used in the current study, compared against values inferred from observed NPP at Ecosystem Model-Data Intercomparison (EMDI) sites. Markers denote median averages for each PFT. Panels (a) and (b) refer to JULES-SF and MODIS-GPP, respectively, for the modelled values. EMDI sites are class A, meaning that NPP is measured both above and below ground (Olson et al 2008). For EMDI, we assume a net-to-gross primary productivity ratio of 0.45 (DeLucia et al 2007). Vertical error bars correspond to the standard error. Horizontal error bars

assume a range ±0.15 in the NPP-to-GPP ratio, with symbols moving to the right for a low ratio of 0.3 corresponding to old, undisturbed sites (DeLucia et al 2007). Note that annual EMDI values can only be compared against modelled, rather than observation-based, values at FLUXNET sites because FLUXNET observations often contain gaps outside the growing season.

Fig.6: Change in Gross Primary Productivity (ΔGPP), simulated by MODIS-GPP, per unit change in Leaf Area Index (ΔLAI). The change is plotted against mean LAI of PFT. The solid line represents a least-squares exponential fit a - c exp(-x/b), where a=-0.770, b=3.54 and c=-2.60. The response for JULES-SF is very similar (best fit a=-0.607, 3.18, -2.60).

Fig.7: A comparison of Princeton reanalysis meteorology against tower-based meteorology. The upper panels represent a primary meteorological used to force JULES-SF (annual total precipitation, PPT) and to force both JULES-SF and MODIS-GPP (average annual shortwave radiation, SW). The bottom panels depict two primary meteorological variables used to force MODIS-GPP (daily minimum air temperature (T_{air}) and daytime Vapour Pressure Deficit (VPD), both annually averaged). Each marker represents one siteyear and is categorised according to the PFT given in the key. The solid and dashed lines represent, respectively, the best linear fit and y=x.

Figure 1: A schematic overview of model input/output for the simulations. Sensitivity experiments are conducted separately for each siteyear within a Plant Functional Type (PFT) and are denoted by DEF (default), LAI_PERTURB (LAI perturbed) and MET_PERTURB (meteorology perturbed). Gross Primary Productivity (GPP; $gCm^{-2}d^{-1}$) for the default experiment (GPP(DEF)) is validated against observed eddy covariance (EC) fluxes. ΔGPP is the difference between the perturbed simulation and DEF.

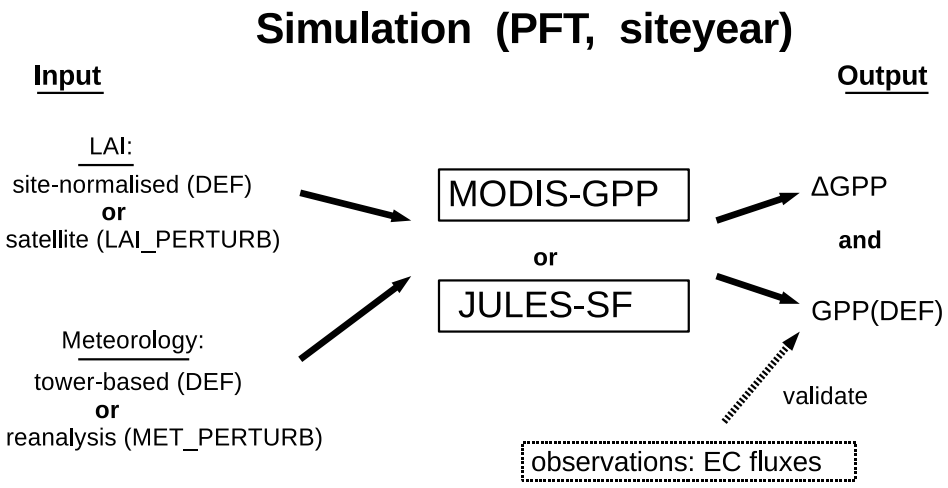

Figure 2: Site locations used in the current study, based on open-access FLUXNET data. For clarity, sites are coarsely categorised (tree, grass/crop and shrub) although 10 PFTs are used in the simulations.

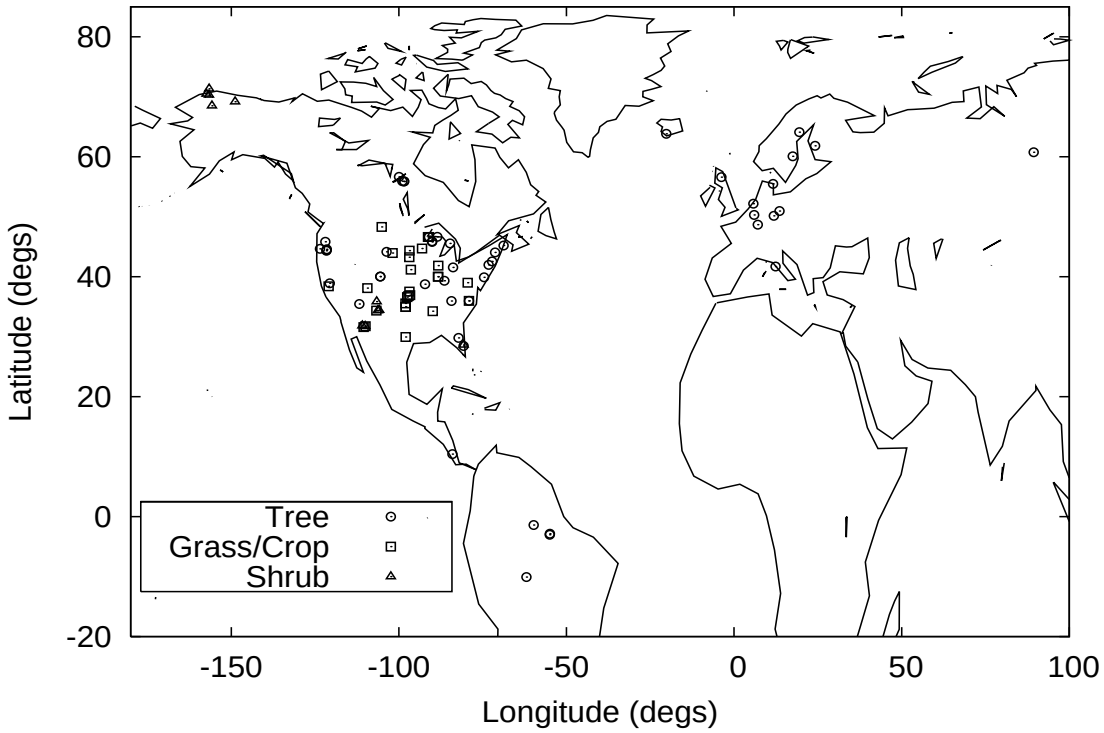

Figure 3: Field-based maximum Leaf Area Index (site LAI), measured for a given siteyear, compared against the corresponding satellite (MOD15A2) measurement. The dashed curve shows a least-squares exponential fit a - c exp(-x/b) excluding the 3 outliers at x<2, y>5 (a=4.32, b=2.59 and c=4.53). To validate LAI, we also include siteyears for savanna (SAV) and mixed forest (MX) which are not simulated in the sensitivity experiment.

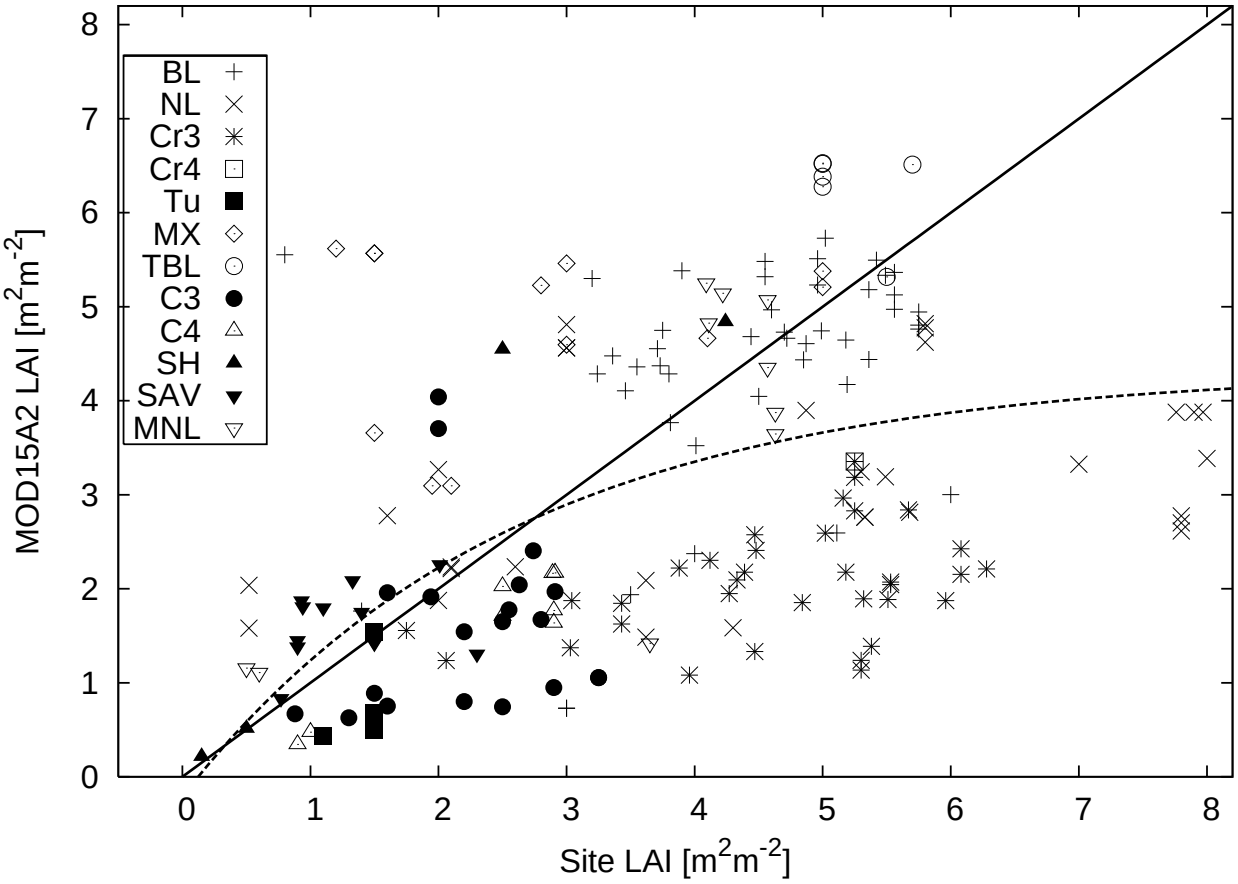

Figure 4: Daily Gross Primary Productivity (GPP) derived from the MODIS-GPP algorithm (bottom) and from JULES-SF (top) compared against observationally based GPP from eddy covariance fluxes. For this validation exercise, both models use the default meteorological and LAI forcing. Each point represents an average for the siteyear. However, GPP is expressed as a daily average to reduce the impact of data gaps across the annual cycle. In each case, the dashed curve shows a least-squares exponential fit a - c exp(-x/b), with a=17.50, b=24.88 and c=15.41 for MODIS-GPP and a=10.94, b=7.35 and c=9.95 for JULES-SF.

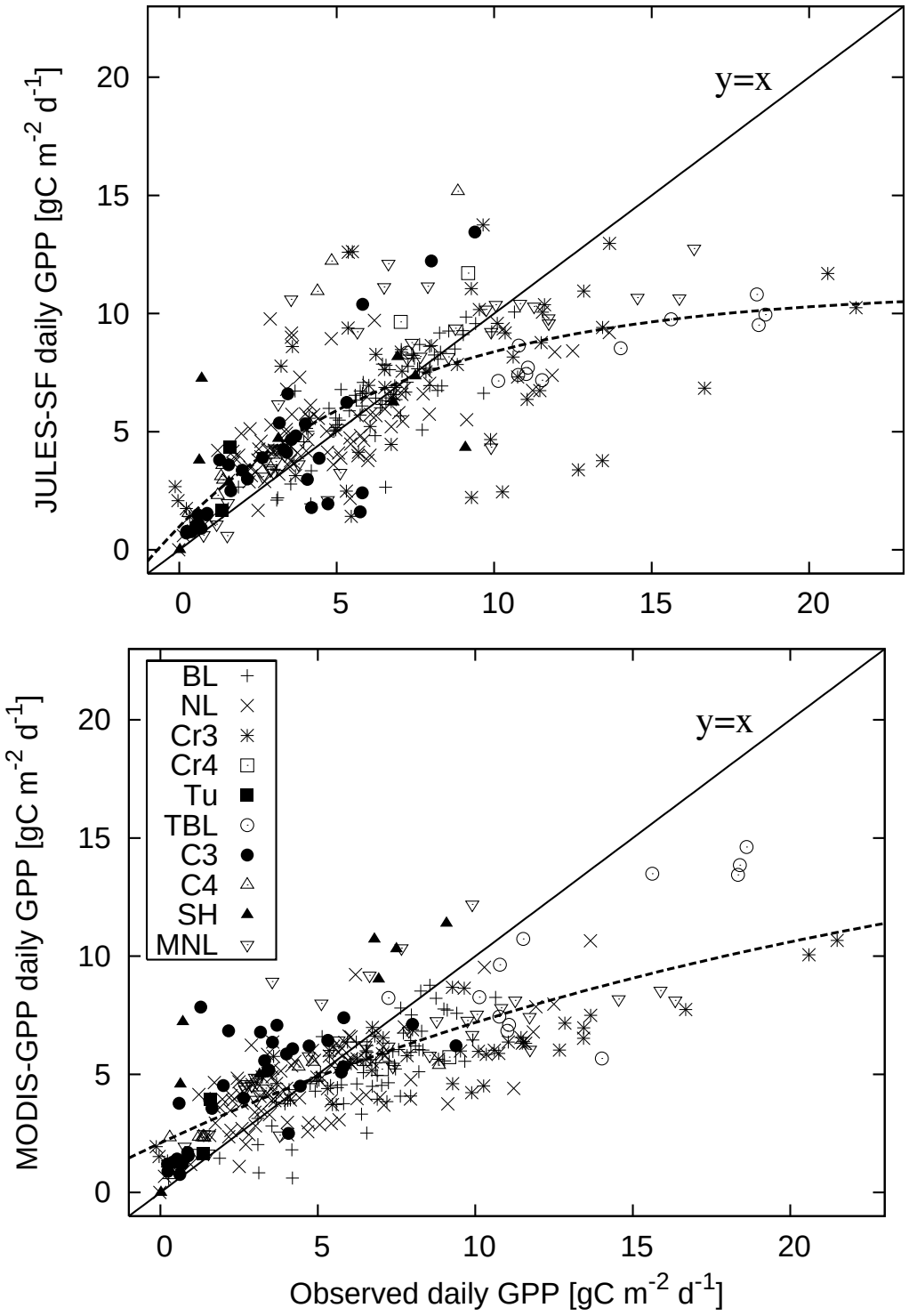

Figure 5: Modelled annual Gross Primary Productivity (GPP) for FLUXNET sites, used in the current study, compared against values inferred from observed NPP at Ecosystem Model-Data Intercomparison (EMDI) sites. Markers denote median averages for each PFT. Panels (a) and (b) refer to JULES-SF and MODIS-GPP, respectively, for the modelled values. EMDI sites are class A, meaning that NPP is measured both above and below ground (Olson et al 2008). For EMDI, we assume a net-to-gross primary productivity ratio of 0.45 (DeLucia et al 2007). Vertical error bars correspond to the standard error. Horizontal error bars assume a range ±0.15 in the NPP-to-GPP ratio, with symbols moving to the right for a low ratio of 0.3 corresponding to old, undisturbed sites (DeLucia et al 2007). Note that annual EMDI values can only be compared against modelled, rather than observation-based, values at FLUXNET sites because FLUXNET observations often contain gaps outside the growing season.

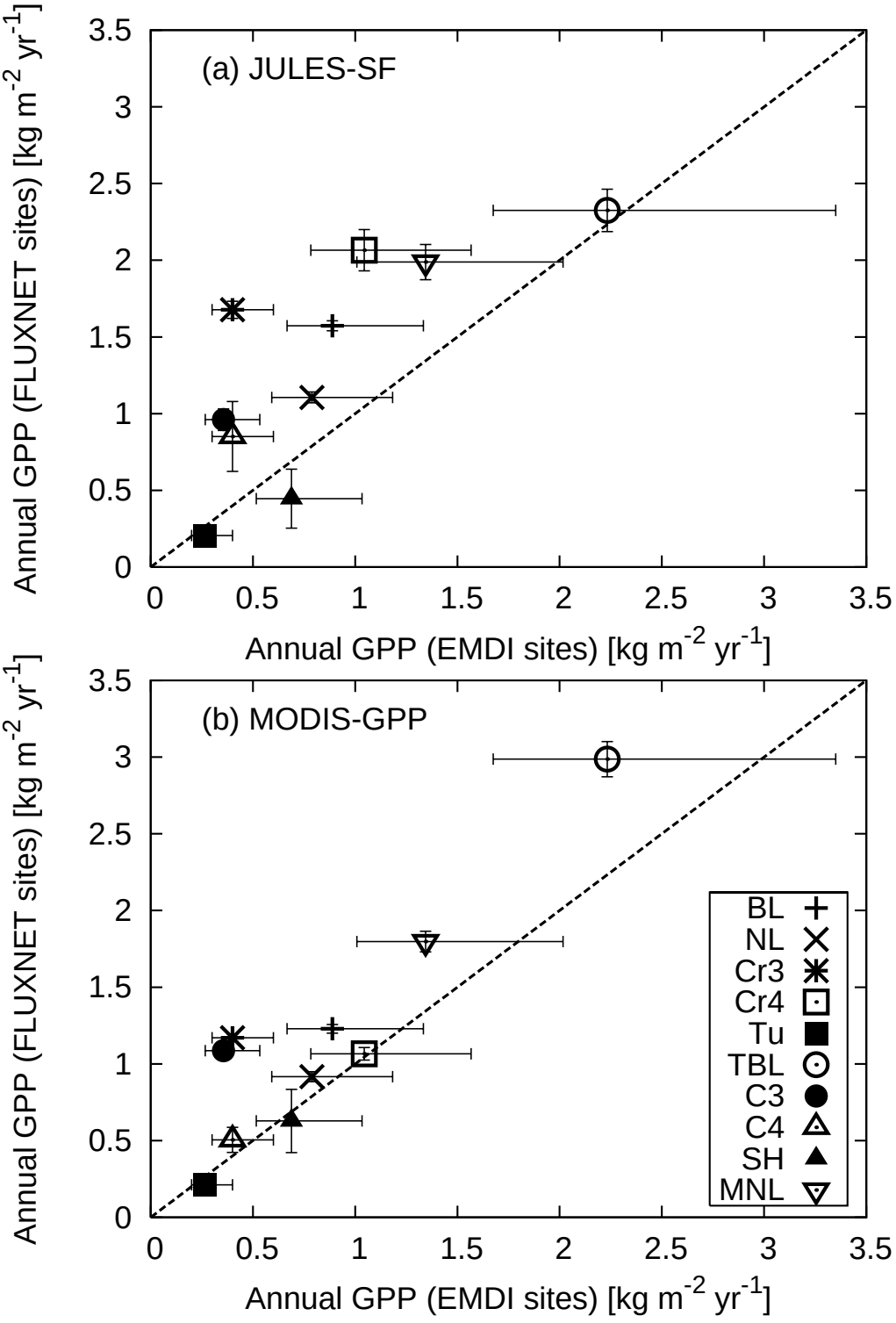

Figure 6: Change in Gross Primary Productivity (ΔGPP), simulated by MODIS-GPP, per unit change in Leaf Area Index (ΔLAI). The change is plotted against mean LAI of PFT. The solid line represents a least-squares exponential fit a - c exp(-x/b), where a=-0.770, b=3.54 and c=-2.60. The response for JULES-SF is very similar (best fit a=-0.607, 3.18, -2.60).

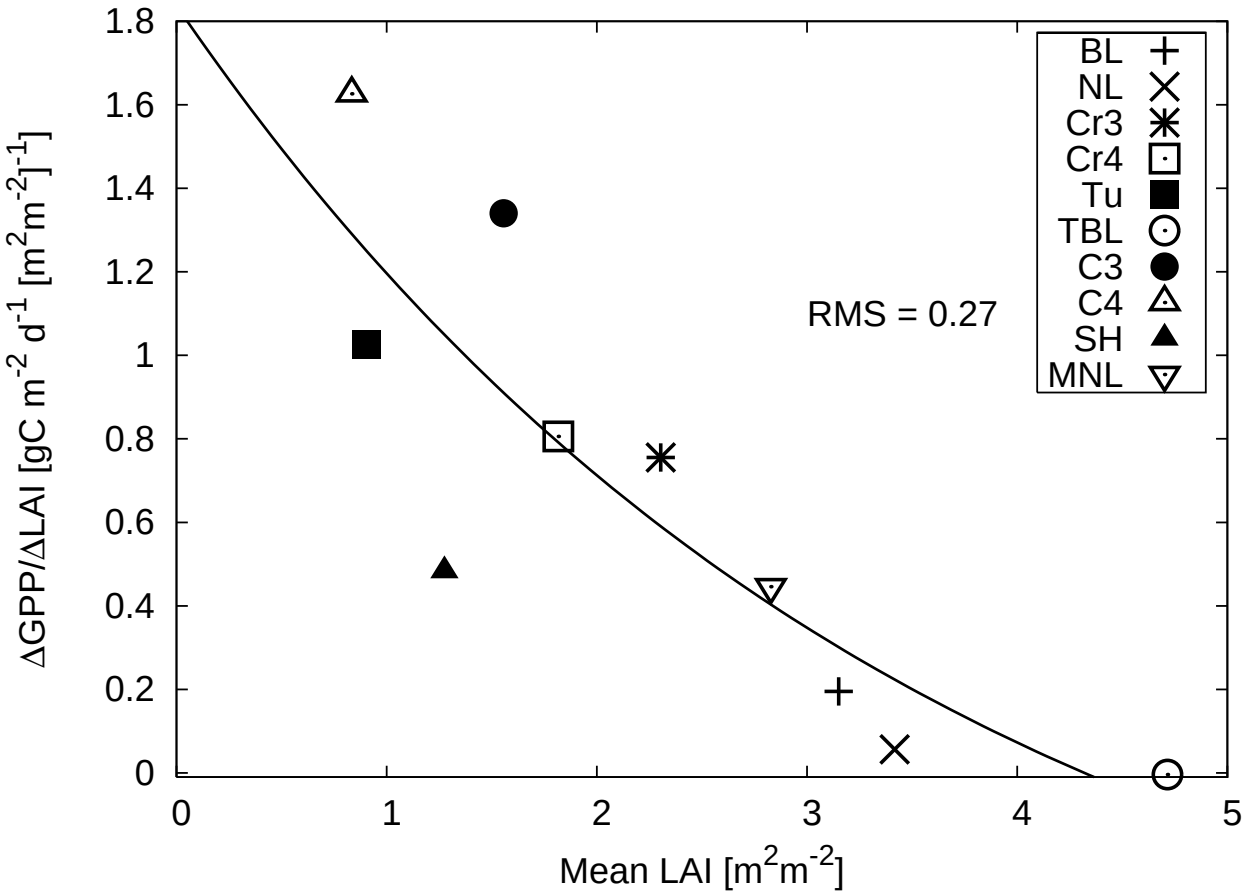

Figure 7: A comparison of Princeton reanalysis meteorology against tower-based meteorology. The upper panels represent a primary meteorological used to force JULES-SF (annual total precipitation, PPT) and to force both JULES-SF and MODIS-GPP (average annual shortwave radiation, SW). The bottom panels depict two primary meteorological variables used to force MODIS-GPP (daily minimum air temperature (T_{air}) and daytime Vapour Pressure Deficit (VPD), both annually averaged). Each marker represents one siteyear and is categorised according to the PFT given in the key. The solid and dashed lines represent, respectively, the best linear fit and y=x.