Validation of 3D-CMCC Forest Ecosystem Model (v.5.1) against eddy covariance data for ten European forest sites

A. Collalti[1,2], S. Marconi[1,2], A. Ibrom[3], C. Trotta[2], A. Anav[4], E. D'Andrea[5], G. Matteucci[5,6], L. Montagnani[7], B. Gielen[8], I. Mammarella[9], T. Grünwald[10], A. Knohl[11], F. Berninger[12], Y. Zhao[13], R. Valentini[1,2], M. Santini[1]

[1]{EuroMediterranean Center on Climate Change – Division on Impacts on Agriculture, Forest and Ecosystem Services (IAFES), 01100 Viterbo, Italy}

[2]{University of Tuscia – Department for Innovation in Biological, Agro-Food and Forest Systems (DIBAF), 01100 Viterbo, Italy}

[3]{Centre for Ecosystems and Environmental Sustainability, Dept. Chem. Engineering, Technical University of Denmark (DTU), Roskilde, Denmark}

[4]{College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, United Kingdom}

[5]{CNR-IBAF – National Research Council of Italy, Institute of Agroenvironmental and Forest Biology, 00015 Monterotondo Scalo, RM, Italy}

[6]{CNR-ISAFOM – National Research Council of Italy, Institute for Agriculture and Forestry Systems in the Mediterranean, 87036 Rende, CS, Italy}

[7]{Forest Services, Autonomous Province of Bolzano, 39100, Bolzano, Italy; and Faculty of Science and Technology, Free University of Bolzano, Piazza Università 5, 39100 Bolzano, Italy}

[8]{Research Group of Plant and Vegetation Ecology, Department of Biology, University of Antwerp, Belgium}

[9]{University of Helsinki – Department of Physics, Division of Atmospheric Sciences, 00014 Helsinki, Finland}

[10]{Technische Universität (TU) Dresden, Institute of Hydrology and Meteorology, Chair of Meteorology, D-01062 Dresden, Germany}

[11]{Bioclimatology, Faculty of Forest Sciences and Forest Ecology, Georg-August University of Göttingen, Büsgenweg 2, 37077, Göttingen, Germany}

29   [12]{Department of Forest Sciences, University of Helsinki, POBox 27, 00014, Helsinki}

30   [13]{LOCEAN/IPSL, UMR 7159, Sorbonne Universitès, Unitè mixte UPMC-CNRS-IRD-

31   MNHN, 4 place Jussieu, 75005, Paris, France}

32   Correspondence to: A. Collalti (alessio.collalti@cmcc.it)

Abstract

34   This study evaluates the performances of the new version (v.5.1) of 3D-CMCC Forest

35   Ecosystem Model (FEM) in simulating gross primary productivity (GPP), against eddy

36   covariance GPP data for ten FLUXNET forest sites across Europe. A new carbon allocation

37   module, coupled with new both phenological and autotrophic respiration schemes, was

38   implemented in this new daily version. Model ability in reproducing timing and magnitude of

39   daily and monthly GPP fluctuations is validated at intra-annual and inter-annual scale,

40   including extreme anomalous seasons. With the purpose to test the 3D-CMCC FEM

41   applicability over Europe without a site-related calibration, the model has been deliberately

42   parameterized with a single set of species-specific parameterizations for each forest

43   ecosystem. The model consistently reproduces both in timing and in magnitude daily and

44   monthly GPP variability across all sites, with the exception of the two Mediterranean sites.

45   We find that 3D-CMCC FEM tends to better simulate the timing of inter-annual anomalies

46   than their magnitude within measurements' uncertainty. In six of eight sites where data are

47   available the model well reproduces the 2003 summer drought event. Finally, for three sites

48   we evaluate if a more accurate representation of forest structural characteristics (i.e. cohorts,

49   forest layers) and species composition can improve model results. In two of the three sites

50   results reveal that model slightly increases its performances although, statistically speaking,

51   not in a relevant way.

52

1 Introduction

Terrestrial ecosystems have a relevant role in the global carbon cycle, acting also as climate regulators (Peters et al., 2007; Bonan, 2008; Huntingford et al., 2009). In fact terrestrial ecosystems store large carbon stocks and cause most of the variance of carbon exchange between the atmosphere and land surface (Batlle Bayer et al., 2012). Among terrestrial ecosystems, forests are an essential component in the global carbon cycle because of their high capacity to store carbon in the vegetation and soil pools (Kramer et al., 2002). Through Gross Primary Production (GPP) plants fix atmospheric carbon dioxide (CO_2) as organic compounds, enabling terrestrial ecosystems to offset part of the anthropogenic CO_2 emissions (Janssens et al., 2003; Cox & Jones, 2008; Battin et al., 2009). Consequently, changes in GPP could have relevant impacts on atmospheric CO_2 concentration. Thus, accurately simulating terrestrial GPP is key to quantify the global carbon cycle and predict the future trajectories of the atmospheric CO_2 concentration (Wu et al., 2015), and taking into account the various spatial and temporal scales of the processes is a major challenge (Yuan et al., 2007). Terrestrial ecosystem models, used to simulate carbon, water and energy fluxes, are valuable tools for advancing the knowledge of the role of ecosystems in maintaining a multitude of their fundamental services, like the provision of products and the regulation of climate (Ibrom et al., 2006). Such numerical models are also useful to: 1) predict the impacts of climate variability on terrestrial biosphere and related carbon fluxes (Ciais et al., 2005; Brèda et al., 2006; Richardson et al., 2007), ranging from long term anomalies (Santini et al., 2014) up to extreme events (Zscheischler et al., 2014); and 2) reproduce biophysical and biogeochemical feedbacks of vegetation cover and change on climate, especially when coupled to atmosphere-ocean climate models through land surface schemes (Bonan, 2008; Arneth et al., 2012; Taylor et al., 2012; Fisher et al., 2015).

At European level, terrestrial ecosystems have been reported to be a significant sink of CO_2 (Luyssaert et al., 2012), with forests playing a relevant role in absorbing anthropogenic emissions for about 10% (Nabuurs et al., 2003; UNECE and FAO, 2011).

In the last decade some studies have identified systematic errors when modelling terrestrial ecosystem sensitivity to climate variability at multiple time scales (Friedlingstein et al., 2006; Piao et al., 2013; Dalmonech et al., 2015) while sometimes differences in model predictions are very large (Wang et al., 2014a).

84   To improve the models capability in reproducing relevant processes related to the land carbon
85   cycle, detailed representation of missing processes should be increasingly developed (Sykes et
86   al., 2001; Campioli et al., 2013; Nolè et al., 2013; Ciais et al., 2013; Prentice et al., 2014). For
87   instance, spatial and temporal environmental hetereogeneity is known to play an important
88   role in the dynamics of populations and communities (Kobe, 1996; Chesson, 2000; Clark et
89   al., 2010; 2011). However, the implications of this hetereogeneity for developing and testing
90   regional to global scale forest dynamics models that are also able to take into account forest
91   management have still largely to be explored (Zhang et al., 2014). As reported by Wramneby
92   et al. (2008), incorporating increased mechanistic details is expected to improve the
93   explanatory power of a model. Many models for example calculate leaf photosynthesis
94   through the Farquhar model (Farquhar et al., 1980; Farquhar & Sharkey, 1982), while few
95   models take in proper consideration the canopy vertical stratification. Increasing model
96   complexity can sometimes mask a lack of understanding, although models including a larger
97   subset of important processes should be more realistic than a simpler model. However,
98   complex models are tuned to perform well at standard tests but produce widely divergent
99   results when projected beyond the domain of calibration (Prentice et al., 2014). Since
100   European forests are mostly managed and not homogeneous in terms of structure,
101   composition and cohorts, only a few models are able to represent this particular ecosystem
102   complexity and heterogeneity (Grote et al., 2011; Morales et al., 2005; Seidl et al., 2012; Yin
103   et al., 2014). For simulating the impact of forest management on the carbon cycle, it is
104   important to consider the vertical structure of forests and the age-related changes in structure
105   and physiology.

106   In this study we investigate the performance of the new version of the 3D-CMCC Forest
107   Ecosystem Model (3D-CMCC FEM, Collalti et al. 2014) in quantifying GPP across different
108   forest types and climate conditions in Europe. In contrast to Dynamic Global Vegetation
109   Models (DGVMs), 3D-CMCC FEM incorporates accurate process description focusing on the
110   effects of hierarchy in vertical forest structure and ages on productivity and growth at species
111   level. The model has been designed to maintain computational efficiency, as postulated for
112   the Light Use Efficiency (LUE) Models (Monteith, 1977), coupled to the accuracy of the
113   Process-Based Models (PBMs) (Makela et al., 2000). As described by Wang et al. (2014a; b),
114   a model with both high accuracy and computation efficiency is highly desirable for the
115   purpose of simulating long time series of GPP at high spatial resolution.

Thanks to FLUXNET, a global network of flux tower sites, half hourly net CO_2, water and energy eddy covariance (EC) flux measurements (Baldocchi, 2003) are now available for a wide range of forest ecosystems. The network provides a continuously increasing set of annual series of half-hourly data (Balzarolo et al., 2014). These data provide valuable information to investigate seasonal phasing and amplitudes of carbon fluxes (Aubinet et al., 2000; Falge et al., 2002; Gielen et al., 2013; Slevin et al., 2015) and to test terrestrial models at the ecosystem scale (e.g. Richardson et al., 2010; Blyth et al., 2011; Chang et al., 2013; Wißkirchen et al., 2013; Bagnara et al., 2014; Balzarolo et al., 2014; Liu et al., 2014; Wang et al., 2014a; Wu et al., 2015). In the present paper daily meteorological and GPP data are provided by FLUXNET. GPP data are exploited as an independent dataset to compare, over different time-scales, 3D-CMCC FEM simulations for ten European forest stands varying in species composition, forest structure, cohorts and climates.

The objective of this work is to answer the following questions:

1   Does the model reproduce the magnitude and the timing of ~~seasonal fluctuations in~~ GPP ~~and their effects~~ across different forest types and forest canopy structures?

2   Does the model reproduce the observed inter-annual GPP variability?

3   Is the model generic enough so that a single set of species-specific parameterizations (i.e. without a site-related calibration) allows reproducing GPP behaviour across different biomes?

4   Do the model outputs improve when considering a complex heterogeneous three-dimensional canopy structure compared to a simple "big leaf" model canopy representation?

To investigate these issues, we introduced a 3D canopy representation into the 3D-CMCC FEM, while however maintaining its flexibility and the generic features to be applied to different forest ecosystems. The new model can now run on a daily time step and includes as main changes an improved allocation-phenology scheme (with new carbon pools including the non-structural carbon pool, NSC), an implemented water cycle (including snow processes) and the computation of autotrophic respiration.

2 Materials and Methods

2.1 Model description

The three-dimensional Forest Ecosystem Model, 3D-CMCC FEM (Collalti, 2011; Collalti et al., 2014) (source code and executables are available upon request to the corresponding authors and downloadable at http://dev.cmcc.it/git/3D-CMCC-FEM-git) is hybrid between an empirical and a process-based model relying on the concepts of the LUE approach at canopy level for carbon fixation. The 3D-CMCC FEM is designed to simulate forest ecosystems at flexible scale (from hectare to 1km per 1km) and on a daily time step. The model simulates tree growth as well as carbon and water fluxes, at species level, representing ecophysiological processes in hetereogeneous forest ecosystems including complex canopy structures. The 3D-CMCC FEM v.5.1 uses daily meteorological data, site-specific data and ecophysiological data (e.g. maximum canopy conductance, specific leaf area, etc.; see Table S3 and Collalti et al., 2014) to simulate forest processes. The model code architecture allows aggregating trees into representative classes, each characterized with its variables (e.g. carbon pools, leaf area index, tree height) based on their ages, species-specific and structural traits. These variables are identified by the model through four indexes: i.e. species (x index), diameter class (Diameter at Breast Height, DBH) (y index), height class (z index), and age cohort (k index); such indexes represent the main state variables considered by the model in distinguishing ecosystems across sites. To deal with forest hetereogenity within and across different ecosystems, 3D-CMCC FEM v.x.x (all model versions follow the same architecture) uses a species-specific parameterization for each species simulated. Moreover, based on the assumption made by Magnani et al. (2007) that the above-ground net primary production decreases with the ageing of a forest, the model explicitly takes into account all ages within the stand, reproducing a year by year reduction due to senescence (Landsberg & Waring, 1997; Waring & McDowel, 2002). Height classes and the tree position within the forest vertical profile are explicitly treated by the model to estimate the light availability (version 5.1 includes also the albedo effects) using the Monsi-Saeki formulation of exponential attenuation coupled with the "Big-leaf" approach developed for a multi-layered model without considering canopy depth (Collalti et al., 2014; Medlyn et al., 2003). DBH together with stand density control grid cell horizontal canopy coverage (and gaps) through the computation of the single tree crown coverage and then upscale to grid-cell level (Collalti et

176   al., 2014). In this way, the model is able to reproduce different combinations of uneven-aged,
177   multi-layered and multi-species forests, by optional simulation of e.g. light competition, age
178   related decline and different species-specific traits. This aspect makes the model flexible to be
179   theoretically used for a wide range of applications in forests and allows quantifying the effects
180   of a particular simulation of forest structure on model performance.

181   ## 2.2   Model implementations

182   In this study, the 3D-CMCC FEM described in Collalti et al. (2014) has been advanced to
183   version 5.1 to improve the representation of forest processes, like phenology, canopy
184   photosynthesis, including autotrophic respiration, tree carbon allocation and water dynamics.
185   The improved phenology routine is based on a new C allocation scheme, that includes new
186   carbon pools among which the Non-Structural-Carbon (NSC) pool, related to five
187   phenological transitions for deciduous species, and three phenological transitions for
188   evergreen species, both updated once per day. Autotrophic respiration is explicitly simulated
189   and separated into maintenance and growth respiration. Maintenance respiration is the
190   function of the nitrogen content (a new added pool) in the living pools, while growth
191   respiration is computed proportionally to the carbon allocated to the different tree
192   compartments.

193

194   **Photosynthesis and net primary production**

195   As in the Collalti et al. (2014) the carbon flux is still estimated in 3D-CMCC FEM through
196   the Light Use Efficiency approach multiplying, for a particular species x, the absorbed
197   photosynthetic active radiation (APAR, i.e. the radiation intercepted by the canopy) with the
198   leaf area index (LAI, m^2m^{-2}) with either the prognostic potential radiation use efficiency (ε_x,
199   grams of dry matter MJ^{-1}) or the maximum canopy quantum use efficiency (α_x, µmol CO_2
200   µmol^{-1} PAR) (for a full list of model parameters see Table S3). Parameters ε_x or α_x are
201   controlled by the product of several environmental factors (modifiers) indicated as $mod_{x,k}$
202   (dimensionless values varying between 0 and 1 and differing for each species x and age class
203   k) depending on: vapour pressure deficit, daily maximum and minimum air temperatures, soil
204   water content and site nutrient status (for a full modifiers description see Landsberg &
205   Waring, 1997). Gross Primary Production (GPP; $gCm^{-2}day^{-1}$) is thus calculated using the
206   following equation:

207 $$GPP_{x,y,z,k} = (\varepsilon_x \, or \, \propto_x) * APAR_z * mod_{x,k} \qquad (1)$$

208 where $APAR_z$ is the absorbed radiation by the trees at the zth layer (where z represents the

209 layer of representative height for each height class).

210 Conversely from the previous version where Autotrophic Respiration (AR) was set as a

211 constant fraction of GPP (Waring & Landsberg, 1998), in this version AR is explicitly

212 simulated. AR is treated distinguishing into Maintenance Respiration (MR), governed by a

213 Q_{10} type response function (Ryan, 1991; Bond-Lamberty et al., 2005) and Growth Respiration

214 (GR) assumed to be a constant proportion (30%) of all new tissues produced (Larcher, 2003).

215 Net Primary Production (NPP), is then calculated as follows:

216 $$NPP_{x,y,z,k} = GPP_{x,y,z,k} - AR_{x,y,z,k} \qquad (2)$$

217 NPP is then partitioned into biomass compartments and litter production following dynamic

218 allocation patterns that reflect environmental constraints (i.e. light and water competition) and

219 age.

220

221 **Daily meteorological forcing and snow dynamics**

222 The model implements a daily time step (previous version was at monthly time step) thanks to

223 the temporal frequency of meteorological forcing input data: daily maximum (T_{max}; °C) and

224 minimum air temperature (T_{min}; °C), soil temperature (T_{soil}; °C), vapour pressure deficit (hPa),

225 global solar radiation (MJ m^{-2}day^{-1}) and precipitation amount (mm day^{-1}). In addition, the

226 model uses the day-time (T_{day}; °C) and night-time (T_{night}; C°) average temperature computed

227 as follows (Running & Coughlan, 1988):

228 $$T_{day} = 0.45 * (T_{max} - T_{avg}) + T_{avg} \qquad (3)$$

229 $$T_{night} = (T_{day} + T_{min})/2 \qquad (4)$$

230 Where T_{avg} is the daily average air temperature (°C). When the soil temperature is missing

231 among in situ observed data, the model estimates it for the upper 10 cm of the soil layer

232 through an 11-day running weighted average of T_{avg} and further corrected by the presence of a

233 snowpack as in Thornton (2010), Kimball et al. (1997) and Zeng et al. (1993). The variable

234 related to the snowpack thickness was included as a water cycle component by reproducing

235 the daily amount (mm day^{-1}) of snow melt driven by average air temperature (T_{avg}) and

236 incident net global radiation (Rad_{soil}; W m^{-2}), while snow sublimation is only driven by T_{avg}.

237 In case of snow presence, if T_{avg} is higher than 0°C, considered the melting point as in
238 Running & Coughlan (1988) and Marks et al. (1992), the rate of daily snowmelt is estimated
239 by:

240 $$Snow_{melt} = \left(t_{coeff} * T_{avg}\right) + \left(\frac{Rad_{soil} * \varepsilon_{snow}}{H_{fus}}\right) \qquad (5)$$

241 where t_{coeff} is the snowmelt coefficient (0.65 Kg m^{-2} °C^{-1}day^{-1}), ε_{snow} is the absorptivity of
242 snow (0.6), H_{fus} is the latent heat of fusion (335 kJ kg^{-1}), Rad_{soil} is the incident net global
243 radiation at the soil surface (kJ m^{-2} day^{-1}).

244 Otherwise, if T_{avg} is lower than 0°C snow sublimation is computed by:

245 $$Snow_{subl} = \left(\frac{Rad_{soil} * \varepsilon_{snow}}{H_{sub}}\right) \qquad (6)$$

246 where H_{sub} is the latent heat of sublimation (2845 kJ kg^{-1}).

247

248 **Phenology and Carbon allocation**

249 Phenology plays a fundamental role in regulating photosynthesis and other ecosystem
250 processes (e.g. carbon and nitrogen dynamics), as well as inter-individual and inter-species
251 competitive relations and feedbacks to the climate system (Richardson et al., 2012a). In the
252 updated model version phenology and carbon allocation depend on six different carbon and
253 nitrogen pools (while three carbon pools where considered in the previous versions). Five
254 pools represent the main tree organs: foliage, (fine and coarse) roots, stem, branch and bark
255 fraction. One new pool corresponds to non-structural carbon (NSC; starch and sugar) stored in
256 the whole tree. The inclusion of this new pool was necessary to represent NSC mobilization
257 and consequently leaf phenology (e.g. leaf production during spring for deciduous trees) and
258 carbon allocation. Woody pools are furthermore distinguished between live and dead wood.
259 In the new version of 3D-CMCC FEM, LAI values are predicted for sun and shaded leaves
260 (De Pury & Farquhar, 1997; Thornton & Zimmermann, 2007; Wu et al., 2015), minimizing
261 the effects of the "Big-leaf" approach (Monteith, 1965; Sellers et al., 1997), as a function of
262 the amount of carbon allocated to the leaf pool. It is noteworthy that each pool and each
263 structural state variables is daily updated according to the meteorological data, forest structure
264 and simulated fluxes.

Following Arora & Boer (2005), for deciduous species the model considers five phenological transitions (being just four in the previous versions: bud burst, peak LAI, leaf fall period and dormancy) that drive the seasonal progression of vegetation through phases of dormancy/quiescence, budburst, maximum growth, active growth, and senescence as in the following:

1. Leaf onset starts from quiescence when thermic sum (the sum of the T_{day} air temperatures exceeding the threshold T_{base} value of 5°C) exceeds a species- and site-specific temperature threshold value (Rötzer et al., 2004; Dufrene et al., 2005) and when the LAI value reaches LAI = max(LAI) * 0.5. The costs of expanding buds during this period of high carbon demand are supported by NSC (Landhausser, 2010; Dickmann & Kozlowski, 1970).

2. During the budburst phase, carbon and NSC are allocated to the foliage pool, as long as the balance between GPP and AR is positive (Barbaroux & Bréda, 2002; Campioli et al., 2013; Scartazza et al., 2013).

3. During the succeeding maximum growth phase and lasting up to peak LAI, carbon is allocated into foliage and fine root pools (Sabatè et al., 2002), based on the pipe model theory (Shinozaki et al., 1964 a; b), to optimize photosynthesis; otherwise, no growth occurs and NSC is used.

4. Successively, the full growing phase lasts up to the day when day length (in hours) is shorter than a species-specific threshold value. In this phase carbon is allocated into stem, fine and coarse roots, branch and bark, and into non-structural carbon pools in order to refill the reserves for the next years.

5. Finally, during the leaf fall (i.e. yellowing or senescence) phase, lasting until the leaf fall (assumed linear) is complete, the total positive carbon balance is allocated to the NSC pool.

Outside the growing season (dormancy) trees consume NSC for fuelling maintenance respiration (Ogren, 2000).

For evergreen species the model follows a similar but simplified approach simulating a first maximum growth phase, when the model allocates NSC to foliage and fine roots up to reach peak LAI, and a second full growing phase, when the model allocates to the other pools. As in

295   Lawrence et al. (2011) for litterfall we assume and simplify that there are no distinct periods,

296   but rather a continuous shedding of foliage and fine roots of the previous years.

297   All tree pools are updated at a daily time step depending on NPP balance. Nitrogen

298   concentration for each pool is considered as a C/N ratio following Dufrene et al. (2005) and

299   Thornton (2010). The C/N stoichiometry is constant and depends on species; unfortunately,

300   the model still lacks of an interactive C-N cycle. Forest stand structural attributes, e.g. DBH,

301   tree height, and crown competition are also updated at a daily time step based on species-

302   specific biometric relationships.

303

304   **Autotrophic respiration**

305   Based on the approach of BIOME-BGC model (Thornton, 2010) 3D-CMCC FEM computes

306   the daily AR of all living tissues. MR is a modified Van't Hoff function (Davidson et al.,

307   2006; Mahecha et al., 2010) of temperature with the temperature sensitivity parameter Q_{10}

308   (see below) and a linear function of the nitrogen content ($N_{content}$ = 0.218 kgC kgN^{-1} day^{-1};

309   Ryan, 1991) in the living compartments. The Q_{10} function is an exponential function for

310   which a 10°C increase in temperature relates to a Q_{10} factor change in the rate of respiration.

311   MR is partitioned into day time and night time respiration using, in place of *temp* in Eq.(7):

312   t_{day} and t_{night} for foliage, t_{soil} for fine and live coarse roots, and t_{avg} for live stem and branch.

313    $$MR_{x,y,z,k} = 0.218 * Ncontent_{x,y,z,k} * Q_{10}^{(temp-20)/10} \tag{7}$$

314    $GR_{x,y,z,k}$ is considered as a fixed ratio (30%) of all newly grown (i.e. living) tissues as

315   proposed by Larcher (2003).

316

317   2.3   Data description

318   Model validation has been performed for ten different forest sites (Table 1) included in the

319   European EC fluxes database cluster (URL: http://www.europe-fluxdata.eu). For each site,

320   3D-CMCC FEM v.5.1 simulations were performed averagely for 10 years, forced with gap-

321   filled daily meteorological data, according to the available time series. The selected sites

322   cover a wide range of European forest ecosystems across different latitudes, landscapes and

323   three climatic zones: temperate, Mediterranean and subalpine.

For all sites, daily time series of meteorological variables (maximum and minimum air temperature, precipitation, vapour pressure deficit and incoming solar global radiation) were used as drivers, while GPP was used for model output validation. The GPP derives from Net Ecosystem Exchange (NEE) measurements that have been previously quality checked and processed including storage correction, spike detection, and low turbulence condition (u*) filtering according to the method in Papale et al. (2006) and gap filled using the Marginal Distribution Sampling method (MDS; Reichstein et al., 2005). The GPP is not directly measured by the eddy covariance technique but it is estimated using a partitioning technique as described in Reichstein et al. (2005). In the rest of the paper, we will refer to these data as "measured" or "observed" GPP for simplicity but it is important to highlight that they are obtained using a modeling approach (although strongly based on direct measurements).

2.4 Model and experimental set-up

Site data needed for model initialization concerned information on forest structure (DBH, tree height, age and density), its species composition, and soil characteristics (e.g. soil depth, texture and bulk density). These data were used for each site to initialize the model, i.e. to describe soil characteristics and the initial forest conditions at which the model starts to simulate forest processes. Initialization data were taken from the BADM (Biological, Ancillary, Disturbance, Metadata) files, available at http://www.europe-fluxdata.eu, for each of the selected sites, and complemented by a literature review and personal contacts with the sites' Principal Investigators. Length of model simulations, basic sites description and forest attributes used for model initialization are shown in Table 1. As a whole, for all sites, the species-specific ecophysiology has been parameterized generically (i.e. not related to the simulated site) using only data from the literature (e.g. Breuer et al., 2003; Mollicone et al., 2003; Pietsch et al., 2005; White et al., 2000) independently from site-related measurements (for a list of model ecophysiological and structural species-related parameters see Table S3). As in White et al. (2000), Reich et al. (2007) and in Naudts et al. (2014) in case of multiple values available for a single parameter, the mean values were used. Using the mean parameter estimates avoided hidden model-tuning (i.e. the use of unrealistic value to obtain the best fit) and largely reduces the likelihood that simulation results are biased by hidden calibration.

In addition, several studies (Bolstad et al., 1999; Griffin et al., 2001; Ibrom et al., 2006; Misson et al., 2007; Cescatti et al., 2012; Guidolotti et al., 2013; Migliavacca et al., 2015)

claim that beside environmental variables, spatial heterogeneity (horizontal and vertical) of the stand structure and composition (age, species) also plays an important role at the ecosystem level. To evaluate if a more detailed simulation of forest heterogeneity improves model performances, a number of replicated simulations were performed for three heterogeneous sites (BE-Bra, IT-Ren and DE-Tha), based on different model initializations in terms of forest layers, species composition and/or ages (Table 1). These replicates start from a forest representation very close to reality (e.g. cohorts, mixed species composition and different canopy layers) to a more generalized one. For reasons of comparability, in these test sites the model has been forced with the same meteorological input data, and eco-physiological species-related parameterizations, i.e. only model initializations data, related to stand attributes, differ. These data are based on different sources: site measurements and/or literature data and/or experimental settings.

In the case of BE-Bra we initialized the model with near all possible combinations of initialization datasets. The first simulation (BE-Bra P_Q-3L) has explicitly taken into account the site heterogeneity (vertical and horizontal) (following Gielen et al., 2013, and ancillary data sources) consisting in mixed species composition at a different canopy coverage rate of *Quercus robur* (Q) and *Pinus sylvestris* (P) (20 and 80%, respectively), with two cohorts (oaks and pines, 65 and 72 years old, respectively) and three forest layers. In the second simulation (BE-Bra P), only a single-layer of Scots pines was considered (following Janssens et al., 2002 and Verbeeck et al., 2007). In the third, fourth and fifth simulations (BE-Bra Q_3L, BE-Bra Q_2L, BE-Bra Q_1L, respectively) only three, two and one layer of pedunculate oaks (following Curiel Yuste et al., 2005 and experimental set up) were assumed. Additionally, two more experimental set-ups combined two layers of oak and one layer of pine (BE-Bra P_Q-2L) and one layer of oak and pine (BE-Bra P_Q-1L).

For IT-Ren, in the first simulation, two layers and two cohorts were considered (IT-Ren 2L_2C) following Montagnani et al. (2009). In the second case, stand heterogeneity has been grouped into one layer, i.e. minimizing forest structure, and one single averaged cohort (IT-Ren 1L_1C; experimental set up).

For DE-Tha, two species (DE-Tha 2S) (spruce 80% and pine 20%, respectively) were modelled in the first simulation (following Grünwald & Bernhofer, 2007), while in the second experiment only the dominant species (spruce; DE-Tha 1S) was considered (BADM source).

2.5 Validation approach

In order to analyse model performance, we used daily (X_{daily}), monthly ($X_{monthly}$) and annual (X_{annual}) time series for modelled and observed GPP values, which were compared at the different time scales. At first, we conducted a comparison via appropriate performance indices on long-term annual average (i.e. over the full series of all the available years), then we evaluated how the model performed in the different seasons aggregating values for months of the same season.

In addition, to avoid misleading results in the daily and monthly signal comparisons due to the strong seasonality for both daily and monthly signals, we followed the decomposition technique proposed by Zhao et al. (2012). To partially remove the seasonal cycle signal, we build a new daily (Y_{daily}) and a new monthly ($Y_{monthly}$) dataset for both observed and modelled data, respectively. The Y_{daily} is created by subtracting the daily time series from the daily mean of the month, and the $Y_{monthly}$ by substracting the monthly time series from the annual mean (see Table S1-b).

For both X and Y datasets we firstly adopted the Pearson coefficient of correlation (r).Then, we calculated the Normalized Root Mean Square Error (NRMSE) (Anav et al., 2010; Keenan et al., 2012) as a standardized index of error. The NRMSE reports the mean difference between observed and modelled GPP values (GPP_{EC} and GPP_{MD}, respectively) normalized on the variability in the GPP_{EC}, in order to have an indication of the average distance between GPP_{MD} and GPP_{EC}, comparable among the different sites. NRMSE was quantified as:

$$NRMSE_{GPP} = \frac{\sqrt{\sum_{i=1}^{N}\left(GPP_{EC_i} - GPP_{MD_i}\right)^2}}{\sum_{i=i}^{N} \sigma\left(GPP_{EC_i}\right)} \tag{8}$$

where i represents the day (or month), and $\sigma(GPP_{EC})$ is the standard deviation of the full daily (or monthly) series of observed GPP consisting of N records.

In addition, model performances were measured for the same series through the 'Model Efficiency' index (MEF) following Reichstein et al. (2002) and Migliavacca et al. (2015):

$$MEF = 1 - \frac{\sum_{i=1}^{N}\left(GPP_{EC_i} - GPP_{MD_i}\right)^2}{\sum_{i=1}^{N}\left(GPP_{EC_i} - avg(GPP)_{EC}\right)^2} \tag{9}$$

In contrast to correlation coefficient r, the MEF index (Bowman & Azzalini, 1997) measures not only the correlation between modelled and observed data (in other words, how well they

414 reproduce the phase of observations), but also their 'coincidence', i.e. the deviation from the
415 1:1 line, and it is sensitive to systematic deviations between model and observations
416 (Reichstein et al. 2002).

417 Another index used in model evaluation is the standardized Mean Absolute Bias (MABstd)
418 (Li et al., 2010) instead of classical Bias index to avoid compensations for errors of opposite
419 signs and standardized (as for NRMSE) to allow comparison across sites:

420 $$MABstd = \sum_{i=1}^{N} \left(\frac{\left| GPP_{MD_i} - GPP_{EC_i} \right|}{\sigma \left(GPP_{EC_i} \right)} \right) \frac{1}{N} \qquad (10)$$

421 To evaluate the model performances in terms of variability patterns, we adopted a procedure
422 to compare each GPP_{MD} value to both its correspondent GPP_{EC} value and the GPP_{EC}-GPP_{MD}
423 difference, at daily and monthly levels. Since the different sites have different ranges of GPP,
424 we arranged in ascending order GPP_{EC} time series, then divided the whole series in 18 classes,
425 each one containing values of a 5 percentile class. For each group of GPP_{EC} we calculated the
426 median and reported the range. We calculated the same statistics for the values of GPP_{MD}
427 arranged so that dates of GPP_{MD} and GPP_{EC} matched. We chose the median rather than the
428 average because it is less influenced by outliers. We decided to use the range rather than the
429 variance as a measure of variability, because giving information on asymmetry.
430 In order to assess the Inter-Monthly and Inter-Annual Variability (IMV and IAV
431 respectively), individual GPP values for each month and year considered were normalized
432 following Vetter et al. (2008) and Keenan et al. (2012). Shortly, we subtracted the respective
433 observed or modelled average from individual (monthly and yearly) observed and modelled
434 value as follows:

435 $$IMV_{(EC \ or \ MD)i} \ or \ IAV_{(EC \ or \ MD)i} = GPP_{(EC \ or \ MD)i} - avg(GPP)_{(EC \ or \ MD)} \qquad (11)$$

436 where avg(GPP) is the long-term (full series of all the available years) average of monthly
437 (for IMV) or yearly (for IAV) GPP from observations (EC) and modeling (MD), respectively.
438 A kernel density estimation (*kde*) was performed to qualitatively observe probability
439 distribuition functions (PDFs) respectively of the IMV and IAV values (Bowman & Azzalini,
440 1997).

441 To evaluate 3D-CMCC FEM ability in reproducing the observed IMV and IAV, we
442 calculated the NRMSE based on monthly and annual time series of IMV and IAV values,
443 respectively. The NRMSE, adopted as a normalized index of error allowing comparability

444 among different sites, was thus calculated as in Eq. 8 but using IMV and IAV instead of GPP
445 individual values, following the approach of Keenan et al. (2012).

3 Results

3.1 GPP evaluation over long-term annual and seasonal scale

448 Both monthly and daily simulated (MD) GPP show high correlations with EC data and these
449 results are consistent with MEF values as well as with NRMSE and MABstd (Table S1a, and
450 Figure 1-a and 1-b). On average, deciduous forests reveal better correlation between MD and
451 EC data than evergreen forests, with a mean r of 0.86, while evergreen and mixed stands
452 show average *r* of 0.81 and 0.77, respectively. For all stations p<0.0001. These results are
453 confirmed by Taylor diagrams (Taylor, 2001) (Figure 2a) which show that the model
454 performs satisfactorily for daily fluxes, in four (i.e. DE-Hai, DK-Sor, DE-Tha, FI-Hyy) of ten
455 sites falling within ±0.5 normalized standard deviations from the reference point (REF;
456 representing observed data) and having correlation around 0.9. For six sites (all the evergreen
457 needleleaf plus deciduous except FR-Hes), the normalized standard deviation of simulated
458 data is close to that of observed data (represented by reference line with normalized standard
459 deviation, i.e. radial distance from the axis origin equal to 1). Simulated data for IT-Cpz, FR-
460 Hes and FR-Pue have, respectively, a normalized standard deviation of approximately +0.2,
461 +0.3 and +0.4 (as difference from that of observations) consistently with the lower correlation
462 values; BE-Bra shows the highest negative difference, in terms of standard deviation, of
463 around -0.3. On average, the least performing result is for IT-Cpz that shows a correlation
464 below 0.60 and falls outside ±1 normalized standard deviation from the reference point. The
465 Taylor diagram in Figure 2b shows the model's capability to better simulate GPP at monthly
466 scale. For seven sites (all deciduous and evergreen needleleaf), the normalized standard
467 deviations of modelled data are close to that of observations (reference line), the correlation is
468 well above 0.90 and within ±0.5 normalized standard deviation from the reference point. IT-
469 Cpz and BE-Bra show improved results with respect to daily data: respectively, their
470 correlation increases by more than 0.1 units, they fall within the +0.2 and -0.2 units of
471 normalized standard deviation differences with respect to that of observations, and they enter
472 in the field of ±1 and ±0.5 normalized standard deviation from the reference point,
473 respectively, although for IT-Cpz the values for all statistical indexes are consistently the
474 lowest. Although less strongly, FR-Pue monthly data also have better performances than daily

475   data results in terms of higher correlation (0.89) and closer position in terms of normalized

476   standard deviations units from the reference point even if the other indexes are a little bit far

477   from the average values of the other sites.

478   To reduce the effects of seasonality, we also examine model performance using

479   decomposition method (section 2.5). In the daily time-step, the overall model performance is

480   much lower in Y dataset (Figure 1-c and Table S1b) than in X dataset, that is, r= 0.51, MEF =

481   -0.43, NRMSE = 1.18 and MABstd = 0.8 in Y dataset vs r= 0.82, MEF = 0.~~63~~57, NRMSE =

482   0.~~57~~ 63 and MABstd = 0.44 in X dataset. The large model error at synoptic scale have been

483   well recognized by previous studies (Dietze et al., 2011; Zhao et al., 2012). The model shows

484   to be a good predictor for DE-Tha and FR-Pue and to be less predictive for DK-Sor and FR-

485   Hes with respect to the X dataset. Accordingly, for FR-Pue comparisons between X and Y

486   datasets show that this site is less affected by seasonality while DK-Sor is the most affected

487   one. As expected, in the monthly time step, the decomposition technique returns more similar

488   results between X and Y datasets. Worst results are for IT-Cpz while best results are for DE-

489   Hai, DK-Sor, DE-Tha and IT-Ren (see Table S1b and Figure 1-d). Overall, after smoothing

490   the seasonality the model shows to be slightly better predictive with average values among

491   sites consistent with observed data (r = 0.94, MEF = 0.85, NRMSE = 0.36 and MABstd =

492   0.27). Comparison between X and Y datasets shows that DE-Hai is less affected by

493   seasonality and IT-Cpz is the most affected one. In brief, comparison between X and Y

494   datasets shows similar reconstruction ability in the monthly time step, but very different in the

495   daily time step because X dataset contains the feature of large seasonality. Given that one of

496   the objects of this study focuses on seasonality fluctuation, we mainly show the results based

497   on X dataset hereafter without specification.

498   To summarize, although with similar inter-sites variability, monthly correlations across

499   different sites are higher than daily ones, with average correlations of 0.94 for deciduous, 0.90

500   for evergreen and 0.92 for mixed stand (Figure 1 and Table S1a).

501   Daily and monthly NRMSE are usually less than 1.00. Monthly NRMSE is less than daily

502   NRMSE, 0.41 vs. 0.63 on average, respectively (Table S1a). These results confirm that the

503   model performs better at a monthly than at a daily time scale (Figure 1), likely because of

504   averaging effects of daily variability in GPP estimation.

505　The same consistency is shown for MEF index that is on average 0.79 (monthly) and 0.57

506　(daily), with largely lower values for the two Mediterranean forests (IT-Cpz and FR-Pue) at

507　both the daily and monthly time scale (Table S1a and Figure 1).

508　Considering the annual mean in deciduous forests (Table S1a), the model slightly

509　underestimates the GPP by -2.4% (average among DE-Hai, DK-Sor), while in FR-Hes and

510　IT-Col it shows an overestimation of 5.2% on average. Concerning evergreen forests, we find

511　an overall model underestimation of 2.1%, with higher variability compared to deciduous

512　forests, and more divergent in the case of the two Mediterranean ecosystems, ranging from

513　underestimation of 18.4% (318 gC m^{-2} year^{-1}; IT-Cpz) to overestimation of 12.1% (158 gC m$^-$

514　^{2}year^{-1}; FR-Pue). Results for the mixed forest site of BE-Bra are reasonable, with an

515　underestimation of about 4.4%.

516　In terms of inter-annual variability of the yearly mean, GPP$_{MD}$ falls well within the range of

517　GPP$_{EC}$ standard deviations for all sites except at IT-Cpz (Figure 3). Deciduous broadleaved

518　and the evergreen needleleaf are the best reproduced.

519　Performance indices from daily and monthly observed and modelled GPP series analysed at

520　seasonal level are shown in Table S2 and Figures 4 and 5. Winter (DJF) and summer (JJA)

521　correlations were generally lower than those in autumn (SON) and spring (MAM).

522　Specifically, DJF and JJA showed a correlation of 0.45 and 0.46 respectively on a daily scale

523　and a value of 0.59 and 0.50 on a monthly scale; MAM and SON showed on a daily scale an

524　average correlation of 0.72 and 0.77 respectively, while on monthly scale a correlation of 0.82

525　and 0.86 with two low values of 0.05 and 0.06 for monthly DJF and MAM for IT-Cpz was

526　shown. Winter and summer monthly average NRMSE of 1.13 and 1.00, respectively, were not

527　significantly different to the 0.66 and 0.57 of spring and fall. MEF and MABstd indexes

528　values suggest similar findings than NRMSE.

529　Figure 6 shows overall modelled vs. observed fluxes over daily and monthly scales, and the

530　absolute difference (ΔGPP MD, i.e. GPP$_{MD}$ minus GPP$_{EC}$) vs. observed fluxes (GPP$_{EC}$) as

531　calculated by the difference matrix described in section 2.5. Overall, the aggregated data

532　reveal high correlation also due to a progressively reduced range of data, and then variability,

533　at higher GPP values (Figures 6; top plots). Figures 6 (bottom plots) show patterns of ΔGPP

534　MD with increasing GPP$_{EC}$. These differences result in strong reduction of discrepancies for

535　GPP$_{EC}$ greater than 8.5 gC m^{-2} d^{-1} for daily, or 7.3 gC m^{-2} d^{-1} for monthly time series.

536   The average intra-annual GPP variations are analysed by calculating the long-term average

537   and standard deviation values for each month of the year (Figure 7). In spring, the modelling

538   results from deciduous forests present a larger variability than the observed data, especially

539   during budburst and in late spring. The model generally matches the observed phenology

540   timing (budburst, peak LAI, leaf senescence and their fall, i.e. length of growing season, data

541   not shown). Consistent biases were observed in late summer.

3.2   Inter-monthly and inter-annual variability

543   The distribution of the IMV for the analysed sites reveals in general lower variance for

544   modelled than observed data (Figure 8 and Table 2). Regarding deciduous forests, both DK-

545   Sor and FR-Hes show IMV_{MD} distributions with a larger interquartile range in comparison

546   with IMV_{EC} (p-value < 0.05). Conversely, for DE-Hai and IT-Col the IMV_{MD} variance is

547   statistically representative for the IMV_{EC}; however IT-Col shows a significantly biased

548   median (p-value < 0.05). Less variability than IMV_{EC} is generally observed for IMV_{MD} of

549   conifers. While DE-Tha shows significant agreement for both variance and central tendency

550   (median) (p-value > 0.05), at FI-Hyy the IMV_{MD} appears statistically in disagreement with

551   IMV_{EC} for both variance and central mean tendency (p-value < 0.05). We find a small

552   difference between IMV_{MD} and IMV_{EC} probability density modal values in IT-Ren (Table 2).

553   Concerning broadleaved evergreen vegetation, we observe very good agreement between

554   observed and modelled IMV central tendency measures in FR-Pue with most of the

555   frequencies between ± 2 $gCm^{-2}d^{-1}$. In FR-Pue, however, we notice that the distributions are

556   slightly shifted, especially around the median, with resulted variance from modelled data in

557   disagreement with observed data. We detect high IMV distributions disagreement in IT-Cpz,

558   where the PDF from observed IMV is normally distributed and the one from modelled IMV is

559   not (as resulted by a χ^2 goodness of fit test). IMV_{MD} series in BE-Bra (mixed forest) are in

560   low agreement with those from EC. Modelled variance is low, and positive IMV values are

561   especially scarcely represented. Table 2 also shows the NRMSE for IAV and IMV series.

562   There is no apparent correlation either between sites species and average error, or between

563   distributions uniformity and NRMSE. In fact, the lowest NRMSE for IMV was found in BE-

564   Bra and IT-Col and the highest in DE-Hai and DK-Sor. On average the model has a NRMSE

565   for IMVs of about 1.2.

566 Figure 9 shows the modelled and measured individual IAV values for each studied site. The
567 magnitude of IAV_{MD} was on average of the same order as IAV_{EC}, showing the model's ability
568 to reproduce the inter-annual variability range, and capturing about 62% of the anomalies
569 signs (i.e. timing) for the total set of years. The model generally better captured conifers' IAV
570 sign (i.e. DE-Tha, FI-Hyy, and IT-Ren), with 66% of the times against about 59% for the
571 deciduous forests (i.e. DE-Hai, DK-Sor, FR-Hes, IT-Col) and 55% for the Mediterrenean
572 ones (i.e. FR-Pue and IT-Cpz). However, the IAV difference in magnitude was better
573 represented for deciduous forests rather than conifers, as inferred by the average NRMSE of
574 respectively 1.45 and 1.67 (calculated by averaging values reported in Table 2). Although the
575 model reproduced the timing of anomalies satisfactory in more than half of cases (a little bit
576 more than in a random selection), the correlations had a wide spread across sites.
577 Quantitatively, modelled anomalies suggest better results for FR-Pue (r = 0.76) and worse
578 results for IT-Ren (r = -0.54).

579 In the case of the year 2003 with its summer heat and drought extreme (Ciais et al., 2005;
580 Vetter et al., 2008), the anomaly sign has been well captured by the model for six of the eight
581 sites analyzed for that year (no observations were available for BE-Bra and IT-Ren) (Figure
582 9). At IT-Cpz and DK-Sor, average IAV_{MD} has the opposite sign to IAV_{EC}, while 2003 was
583 recognized as not remarkably anomalous at IT-Col. Similarly, the model results match with
584 that found by Delpierre et al. (2009) about the anomalous carbon uptake during the warm
585 spring of 2007 compared with the decadal mean for FR-Pue, FR-Hes, and DE-Tha.

3.3 Comparison within different forest structure simulations

587 Considering the presence of only one species (either pines or oaks) strongly limits the model
588 to simulate the daily and monthly GPP patterns in BE-Bra (Table 3). This site represents a
589 mixed stand of deciduous and evergreen tree species that assimilates CO_2 all year round,
590 although low temperatures in winter and spring reduce photosynthesis for pines also. The
591 observed GPP fluxes are then caused by the 'mixture', at a varying degree, of both oak and
592 pine trees. Considering BE-Bra as a pure oak forest with a variable number of layers
593 (simulation codes: BE-Bra Q_3L, BE-Bra Q_2L, BE-Bra Q_1L) the model results for annual
594 GPP deviate from -1.2 up to -7.4%; considering a pure pine forest (BE-Bra P) or a
595 combination of pines and one layer of oak (BE-Bra P_Q-1L) the model underestimates
596 annually from -15.9% to -11.5%, respectively. It is noteworthy that the daily GPP values

show a markedly different seasonal distribution on fluxes (data not shown). Conversely, there is no clear evidence that in simulating pines coupled with one, two or three oak layers (BE-Bra P_Q-1L, BE-Bra P_Q-3L BE-Bra P_Q-2L) model results largely benefitted of this differentiation both on a daily, monthly and annual scale. Similar results are obtained for DE-Tha site when simulating one single species (DE-Tha 1S) or two (DE-Tha 2S), since the similar phenology behaviour of modelled species does not cause a marked difference in the seasonal GPP cycle. On the other hand, IT-Ren initialized as a single layer and with one single cohort (IT-Ren 1L_1C) instead of two layers and two cohorts (IT-Ren 2L_2C) differs strongly from observed GPP values overestimating the annual cumulated GPP by 43.2%. However, for this site, the analysis of performance indices based on daily and monthly series shows no evidence of improved model results, slightly better "r" values for 1L_1C in respect to 2L_2C are counterbalanced by the other indexes (i.e. NRMSE, MEF and MABstd) where the 2L_2C representation (for both daily and monthly time step, respectively) shows to slightly improve model results; no difference is found for MABstd.

4 Discussion

In this paper, we have analyzed the capability of the latest version of the 3D-CMCC FEM (v.5.1) to simulate intra-annual to inter-annual GPP variability over ten heterogeneous European forest sites, representative of different ecosystems and bioclimatic regions, by comparing model results with observations based on the EC technique. Although the model provides a reasonable reproduction of the observed values, we may evince some critical issues. First, the observed GPP data are affected by high uncertainties (Kenan et al., 2002; Papale et al., 2006; Richardson et al., 2012a, b). According to Luyssaert et al. (2007) these uncertainties in the ten case studies considered here, although at the biome level, have a very high spread, varying from ±557.9 (for FI-Hyy) to ±700 gC m^{-2}yr^{-1} (for IT-Cpz). Besides uncertainty in the EC technique, model assumptions and parameterizations can increase discrepancies compared to observed GPP data.

A potential further source of error in the model runs that may need to be considered or accounted for is related to our choice of not making a site-specific parameterization. Since we used general parameterizations, large uncertainties could be detected especially in the variables that determine, for example, the length of the growing season (Richardson et al., 2010), and the latitudinal differences (acclimation) of the maximum, minimum and optimum temperatures for photosynthesis. Improvement could be achieved with a site-specific

629 parameterization, but this falls beyond our goal to make the model generally applicable. In
630 addition, to avoid a misleading model evaluation coming from strong seasonality (especially
631 for deciduous sites) we followed the decomposition technique proposed by Zhao et al. (2012).

632 On average, 10 years of simulations have been conducted for each site. In addition, in three
633 sites different model initializations (i.e. considering different forest structure, composition and
634 cohorts) were used to quantify improvements in model results when a more detailed
635 heterogeneity forest structure representation and processes are simulated. Modelled GPP
636 results were compared against those from EC observations collected for these sites
637 encompassing three mono-specific (pure) stands of Beech, Holm oak and Scots pine, and
638 three uneven-aged, multi-layered and mixed stands.

639 Based on results, we can now provide answers to the four initial questions:

640 ***1.   Does the model reproduce the magnitude and timing of ~~seasonal fluctuations in~~***
641 ***GPP ~~and their effects~~ across different forest types, structures and compositions?***

642 Overall, as desirable, the model is skilful in reproducing the annual GPP and its intra-annual
643 (seasonal) cycle, calculated as both daily and monthly value averages, with the monthly scale
644 performing better across all statistical indexes considered. These results can be however
645 considered as a "false positive" due to the strong seasonality of GPP patterns that influences
646 and causes higher values of correlation than the model's capabilities to reproduce GPP fluxes
647 (Zhao et al., 2012). This is clearly related to the tendency to linearize the relationship among
648 CO_2 flux and PAR and/or temperature, as also reported by Ruimy et al. (1995) and Wu et al.
649 (2015). Overall, statistical indexes of daily and monthly modelled values for both X and Y
650 datasets were highly consistent with EC data, except for the Mediterranean sites (where
651 seasonality is less pronounced) (see Table S1a and b). Summer drought stress appeared to be
652 the most limiting factor on photosynthesis at FR-Pue (Falge et al., 2002; Reichstein et al.,
653 2002; Sabatè et al., 2002) while the presence of shallow groundwater table at IT-Cpz seemed
654 to reduce the severity of summer drought. This reduction cause a smoothing of seasonality
655 well highlighted in the Y dataset (see Table S1-b) where IT-Cpz showed to be unanimously
656 one of the worst simulated site at both daily and monthly timescale while FR-Pue and DE-
657 Tha, both evergreens, the less affected by seasonal patterns. This behaviour is confirmed by
658 the daily values of DK-Sor and IT-Col for monthly data, both deciduous, that showed to be
659 the most affected, in other words if we smooth over the seasonal trends results get worse
660 while the model indicated to be less sensitive for those evergreen sites where seasonality is

661 not marked with high values of correlation for DE-Tha, FI-Hyy and Fr-Pue. These results
662 confirm that seasonality has a remarkable effects on a model evaluation.

663 Concerning the seasonality, all statistical indexes divided by seasons in Table S2 are
664 consistent in showing a non-negligible uncertainties in representing GPP patterns, as well as
665 inferred by temporal mismatches in variance. The overall agreement despite temporal
666 mismatches suggested that errors compensated over the year, but are cumulated in specific
667 time windows (e.g. seasons). As reported for other models (Morales et al., 2005 and Naudts et
668 al., 2014), the model's performances are generally worse in winter (DJF) and summer (JJA).
669 Biases and differences in winter GPP variance may be related, among other things, to the
670 model algorithms used to simulate LAI and to the algorithm used to calculate GPP from EC
671 data (Reichstein et al., 2005), since GPP variability should be low during DJF or absent for
672 deciduous forests. However, mismatches are also related to the way in which 3D-CMCC
673 FEM represents winter and early spring ecosystem processes. The model in fact does not
674 consider the influence of ground vegetation that appears to be not negligible in some cases
675 (Kolari et al., 2006). High GPP variance for evergreen species could be strongly related to
676 low temperatures during winter (Del Pierre et al., 2009). Systematic overestimation in winter
677 and spring GPP could then be associated with a lack in representing conifers acclimation or to
678 soil and atmosphere thermal constraints. At high latitudes and altitudes, another source of
679 uncertainty may be related to freezing and thawing dynamics in soil water (Beer et al., 2007)
680 which are not considered by the model, as with snow sublimation and melting, which are still
681 simplistically represented.

682 GPP of deciduous forests in summer and autumn are also affected by uncertainties for
683 surface, which is represented by LAI in the model. In addition, GPP is linear with respect to
684 PAR (Monteith, 1977) over monthly or annual time scales, while the relation is strongly
685 nonlinear at the daily scale (Leuning et al., 1995; Gu et al., 2002; Turner et al., 2003; Wu et
686 al., 2015). The linear response of GPP to PAR led to the underestimation/overestimation of
687 GPP under conditions of low/high incident PAR (Propastin et al., 2012; He et al., 2013). In
688 the case of stress or photoinhibition, leaves reduce or stop photosynthesis at too high levels of
689 radiation, while in normal conditions, photosynthesis is light-saturated at high PAR (Mäkelä
690 et al., 2008) which lets canopy photosynthesis saturate at relatively low PAR even in dense
691 tropical forests with high LAI (Ibrom et al., 2008). The model overestimation of summer GPP

692 may thus be partially related to the lack of representation of the canopy photosynthesis
693 saturation processes.

694 Although adopting a more complex phenology scheme in the comparison between decidous
695 and evergreen forests, our model showed better performances for deciduous compared to
696 evergreen forests. This behaviour is due to the strong seasonality patterns that the deciduous
697 species show, which is consistent with the findings of Zhao et al. (2012) at the two french
698 sites, but contrasts to the results of Morales et al. (2005) who showed that it is generally easier
699 for models to simulate evergreen forests due to the simpler phenology. The present results for
700 evergreen forests are, however, highly affected by the low model performances for the two
701 evergreen Mediterranean forests. As previously stated, overestimation during summer at FR-
702 Pue, and during winter and spring for IT-Cpz, are mostly related to neglecting species-
703 specific drought stress response functions. As in Landsberg & Waring (1997), the water
704 modifier is only based on soil physical characteristics and no consideration is given to the
705 stress tolerance or strategy of the species (Larcher, 2003), suggesting that further model
706 developments should focus on this aspect.

707 Other discrepancies affecting other sites could probably be reduced with a site-specific
708 parameterization.

709 *2.   Does the model reproduce the observed inter-monthly and inter-annual GPP*
710 *variability?*

711 Overall, the distribution of the modelled inter-monthly variability was sufficiently consistent
712 with the observed one. The model, however, showed reduced variability in the distribution for
713 both conifers and deciduous species. The model's ability in better representing higher rather
714 than lower anomalies suggests that it may still be less sensitive to some drivers of variability.
715 In this context, the phenological cycle may have an important role, since it influences canopy
716 cover and is controlled by environmental drivers (Richardson et al., 2010). According to Suni
717 et al. (2003) and Jeong et al. (2013), spring phenology largely affects the summertime carbon
718 budget. Hence, uncertainties in the growing season start date may affect 3D-CMCC FEM's
719 ability to reproduce IMV. In summer and autumn, petioles loss of turgor, cavitation in xylem
720 vessels and leaf yellowing may have an important role in the GPP variability of temperate
721 forests (Reichstein et al., 2007).

722 Even though evergreen forests do not experience complete dormancy in winter, changes in
723 'greenness' can be attributed to seasonal variation in canopy biochemistry, the production of

new foliage by canopy species and, particularly where the overstorey is sparse, the phenology of understory vegetation (Richardson et al., 2010). Leaves of different ages have different efficiency, sensitivity to solar radiation, temperature and water related stresses (Chabot & Hicks, 1982). All these elements may have an important role in affecting GPP dynamics, but are still scarcely or not represented by mechanistic ecosystem or forest models. As a confirmation of these suspects, slight modifications in representing phenology and leaf turnover resulted in general improvement of model consistency with EC data (Marconi, 2014).

Distribution of IMV values showed specific patterns attributable to the dominant species. Beech forests IMV PDFs were concentrated around the average value and strongly influenced by high biases. This pattern was probably due to the fact that half of the months in one year have no or little photosynthesis (i.e. early spring, fall and winter) and most of the photosynthetic activity occurs in late spring and summer, when carbon assimilation is influenced by temperatures and solar radiation (Mercado, et al.; 2009). Conifers PDFs were usually smoother, non-skewed, with reduced variability and fitted by a statistical normal curve. The model showed an average NRMSE for IMV of 1.22 but still captured about two thirds of the annual anomalies sign (a little bit more than the fifty percent that represents a simple causality).

The results for IAV (see Figure 9) are quite contrasting and largely depend on the site and the number of annual-by-annual comparisons. The recent modelling studies, that we are aware, show unanimously the difficulties of models to explain the large interannual variability in cases where no obvious triggers like management or climatic extreme are at work (e.g. Keenan et al., 2012; Wu et al., 2013). In 3D-CMCC FEM better results have been obtained for FI-Hyy and FR-Pue, so there is no apparent correlation with latitudes and forest species. Interestingly, the performance of a DGVM for IAV in FR-Pue is also higher than other sites (Zhao et al., 2012), indicating the main determinant factor for GPP simulation in this Mediterranean site may not come from the treatment of canopy representation. However, the advantage of a 3D canopy representation needs to be revalued in the future. Similarly, worse results are reported for IT-Ren, IT-Cpz and BE-Bra where the number of annual correlations are lower than the other sites. The magnitude of differences in the standard deviation generally follows the same tendency, particularly for BE-Bra, IT-Ren and IT-Cpz. These

755 results confirm the model's limited ability to represents the inter-annual variability in these
756 specific sites rather than in these ecosystems.

757 The comparison between modelled and observed data at the inter-annual time scale shows the
758 model to be sufficiently able to reproduce the sign of variability through the years including
759 the extreme events (heat wave combined to drought) during the summer 2003 (Ciais et al.,
760 2005; Vetter et al., 2008) and, for some sites, the anomalous carbon uptake during the warm
761 spring of 2007 described by Del Pierre et al. (2009). Potentially negative effects from the
762 anomalous 2003 were modelled into negative GPP anomaly at DK-Sor and IT-Cpz due to
763 model simulation of summer drought stress, while such anomalies are not evident from
764 measurements for DK-Sor (Pilegaard et al., 2011) and IT-Cpz. This could be due to the more
765 maritime climate for DK-Sor and the presence of shallow groundwater for IT-Cpz that
766 weakened the effects in the first part of the summer. In both sites, and included DE-Tha, the
767 effects during July to September were captured by the model (data not shown). As reported by
768 Ciais et al. (2005), Mediterranean sites showed a smaller degree in carbon fluxes, largely
769 dominated by less respiration. It is noteworthy that IT-Col, differently from other european
770 beech stands, does not seems to have suffered from this anomalous heat wave in 2003 (G.
771 Matteucci, personal communication). Both simulated and observed data showed a positive
772 GPP anomaly, demonstrating that this beech forest benefited by moderate higher temperature
773 values and consequently had "extra" days for assimilation and growth (see also Churkina et
774 al., 2002; Richardson et al., 2010). A similar behaviour was reported also by Jolly et al.
775 (2005) for the Swiss Alps, especially between March and July. This pattern seems to be
776 mostly related to an untimely beginning of the growing season (see Piao et al., 2006), to a
777 reduction in plant transpiration that causes an increase in plant water use efficiency throught
778 the partial closure of stomata (Warren et al., 2011) and to high fluxes related to forest floor
779 vegetation.

780 It is also noticeable that in FR-Hes during the summer of 2004 a negative anomaly occurred,
781 larger than in 2003; and while its sign was captured by the model, its magnitude was not. This
782 can be explained by the modelled postponed effects of a low NSC allocation during the year
783 2003 to the subsequent periods (Granier et al., 2007; Gough et al., 2009). These results
784 highlight that model has a sort of "memory" linked to short-term events (e.g. drought stress)
785 and that these events affect the long-term processes.

786   Quantitatively, modelled inter-annual anomalies show a very large spread across the sites.
787   Correlations vary widely, without any apparent relation with latitude and/or species. If
788   modelled anomaly signs are potentially agreeing with the observed ones most of the time,
789   their magnitude was not. This behaviour seems to be related to several aspects, mainly to an
790   over/under estimation of the causes that reproduce anomalies, e.g. processes simulated linked
791   to the type of climate anomaly, mismatches in phenology or to a missed representation of
792   other processes (e.g. mast years, disturbances, shallow water). Keenan et al. (2012) asserts
793   that a lack in phenological variability and in canopy and soil dynamics are the main culprits of
794   these mismatches but also that flux measurements are affected by random errors especially
795   when fluxes are higher. Poulter et al. (2009) found a similar magnitude of errors with models
796   that were driven by remote-sensing data. Open questions remain as to the proportion of
797   interannual variability in land-atmosphere carbon exchange that is directly explainable by
798   variability in climate (Hui et al., 2003; Richardson et al., 2007)

799   *3.   Is the model generic enough that a single set of species-specific parameterization*
800   *allows reproducing GPP behaviour across different ecosystems without further need of a*
801   *site-related calibration?*

802   Overall, the model showed good flexibility although the sites showed a pronounced spatial
803   and temporal heterogeneity (i.e. a variable number of forest layers, different cohorts and
804   species). The model was able to reliably represent the ecophysiology of beech and spruce
805   species at different latitudes, without modifying or tuning the parameterization sets. However,
806   annual and seasonal performance indices, calculated exploiting daily and monthly series,
807   evidenced different performances between the two northern beech sites and the two southern
808   ones. Tables S1 and S2 show a systematic difference in all the statistics used, suggesting the
809   presence of a latitudinal gradient in 3D-CMCC FEM's ability to represent beech forest
810   processes. This gradient could be explained by how the model represents the different limiting
811   factors and their impacts on GPP. For example, we expect low temperatures to be the most
812   important limiting factor at higher latitudes, compared to soil water availability at lower
813   latitudes (Chapin et al., 2002).

814   We obtained similar results for the two spruce sites. The model showed better performance at
815   higher latitudes. While phenotypic plasticity, and thus the parameter set, may influence the
816   model results, it is noteworthy that the IT-Ren site has different topographic and climatic
817   conditions. Lower average temperatures, higher slopes, and non-negligible encroachment of

818   different species in a more complex canopy, may negatively influence the model performance

819   in IT-Ren with respect to DE-Tha. Since the model showed unrealistic results for the two

820   Mediterranean forests, we think it is not easy to determine if and how differences in

821   performances are related to the generality of the model rather than to bad assumptions behind

822   the simulated processes. From our findings, we conclude that for non-water limited conditions

823   it is possible to yield satisfying results with general parameter sets.

824   ***4.   Do the model's results improve when considering a complex 3D canopy structure?***

825   We evaluated possible improvements that could be made if a more accurate model

826   representation at a higher rate of heterogeneity of: forest structure, differences in ages and

827   species composition and their linked structural-ecophysiological processes, are assumed.

828   These analyses helped us to understand the importance of each process within the represented

829   combination (i.e. light competition, age related decline and the specific differences in

830   ecophysiology) on modelled GPP. Doubtless, a direct comparison between modelled and

831   observed GPP data is not possible due to the lack of partitioned measurements of GPP across

832   different layers, cohorts and species. However, in situations where the different

833   ecophysiological behaviours express themselves in the species specific canopy responses

834   during certain periods of the seasonal cycle, the test of a mixed forest tree model with flux

835   measurements is possible, as the results by Oltchev et al. (2002) showed using the model

836   MixFor-SVAT.

837   This preliminary analysis can be considered as a sensitivity analysis in terms of processes

838   explicitly simulated instead of lumped parameterisation. As a whole, model results using

839   different initialization data are within the observed GPP uncertainties but a quantitative

840   assessment for two sites, BE-Bra and IT-Ren showed the potential to increase the model's

841   ability in simulating fluxes, while for DE-Tha there is no evidence that model performances

842   could benefit of these efforts. For BE-Bra, taking into account two species (that differ

843   especially for their phenological traits) was beneficial in terms of model performances, the

844   same occurred for different layers (with the exception of BE-Bra P_Q-3L vs. BE-Bra P_Q-2L

845   whose results were similar) and different cohorts. Better performances, in terms of seasonal

846   GPP representations, were obtained when each of the above mentioned characteristics was

847   accounted for by the model. For IT-Ren, similar results were obtained, although no

848   differences were found in the simulation of phenological patterns in daily and monthly results.

849   Differently, for DE-Tha a differentiation between the two evergreen coniferous species did

850 not cause marked differences in model results, due to low differences in species
851 ecophysiological traits, justifying in these cases the use of a Plant Functional Type (PFT)
852 level of parameterization instead species level (Poulter et al., 2015).

5 Conclusions

854 This study aimed at evaluating the performances of the updated version of 3D-CMCC FEM
855 compared to nearly 10x10 sites x years GPP data across eddy-covariance European forest
856 sites. Although the sites showed high spatial and temporal environmental heterogeneity the
857 model appears able to reproduce trends in all of the ten sites. Different performance indexes
858 showed that daily and monthly level model results match well, both for the annual and
859 seasonal scale, against observed data, with some exceptions. Mediterranean sites (IT-Cpz and
860 FR-Pue) showed to be the most problematic in reproducing carbon fluxes. This is likely due
861 to their specific ecosystem peculiarity, e.g. shallow groundwater for IT-Cpz and, for both
862 sites, a low pronounced seasonality. In these two sites, the model showed less generalisation
863 unless additional processes were included. Differently from other models, 3D-CMCC FEM,
864 both for daily and monthly simulations and for both X and Y datasets, performs better for
865 deciduous species rather than for evergreen, although deciduous species have a more complex
866 phenology and a more pronounced seasonality. Some mismatches in the simulation over the
867 seasons and over the sites still remain, especially during winter and summer. The first reason
868 for these low agreements in winter can be also attributable to errors during the estimation of
869 GPP from NEE and Ecosystem Respiration values from measurements data. The second can
870 be related to the model's lack or simplicity in representation of snow pack dynamics as
871 reported by Krishnan et al. (2008; 2009), especially for evergreen sites (Keenan et al., 2012).
872 Disagreements in summer could be related to model simplicity in simulating soil drought and,
873 using the Monteith approach (Monteith, 1977), to the strong nonlinearity at the daily scale of
874 GPP and PAR, and to the lack of representation of the light saturation processes. In addition,
875 as reported by Keenan et al. (2012), the apparent high variability in the data during the
876 summer season could therefore be due to random errors in the flux measurements, generating
877 larger variability and then lower correlations against modelled data.

878 No marked differences were found in simulations across different latitudes, so model
879 parameterizations for the different tree species could be useful over Europe with quite a high
880 rate of confidence, with the exception of specific cases in Mediterranean forests.

881 As for other models, 3D-CMCC FEM showed to have the potential to correctly reproduce the
882 signs of interannual variability, like the 2003 heat wave and drought extreme and the
883 anomalous carbon uptake during the warm spring of 2007 and their instantaneous biological
884 response to these events. Significant disagreements were, however, found in reproducing the
885 magnitude of these anomalies.

886 The consideration of stand hetereogeneity, when possible or existing (i.e. layers, cohorts and
887 mixed composition), led the model to improve its results in two of the three sites compared to
888 generalized simulations of forest attributes. This plasticity makes the model able to be used in
889 a wider range of forest ecosystems.

890

891 **Author contribution**

892 A.C. conceived the paper, designed the experiments, co-developed the model code, performed
893 the simulations, and wrote the manuscript with contribution from all co-authors

894 S.M. co-developed the model code, performed the simulations, and contributed to data
895 analysis

896 A.I. contributed to manuscript improvement

897 C.T. contributed to data analysis

898 A.A. contributed to data analysis and to the manuscript improvement

899 E.A. contributed to data analysis

900 G.M. contributed to manuscript improvement and data analysis

901 L.M. contributed to manuscript improvement and data analysis

902 B.G. contributed to manuscript improvement and data analysis

903 I.M. contributed to manuscript improvement and data analysis

904 T.G. contributed to manuscript improvement and data analysis

905 A.K. contributed to manuscript improvement and data analysis

906 F.B contributed to manuscript improvement and data analysis

907 Y.Z. contributed to data analysis and to the manuscript improvement

908 R.V. contributed to manuscript improvements.

909   M.S. contributed to manuscript conceiving and data analysis improvements

910

917

918 **Table captions**

919 **Table 1** Main characteristics of the study sites. Sites were classified according to the IGBP

920 (International Geosphere Biosphere Program) legend as in the FLUXNET database:

921 MF=mixed forest; DBF=deciduous broadleaf forest; EBF=evergreen broadleaf forest;

922 ENF=evergreen needle leaf forest. Year of simulation starting and ending depend on available

923 time series of observed data.

924 **Table 2** IMV and IAV NRMSE for the analyzed sites. Each specific IMV distribution was

925 tested for normality goodness of fit (N = normal distribution, P = non normal distribution). A

926 test for equivalence of central tendency was performed between IMV_{MD} and IMV_{EC} values.

927 (na) refers to the case of sites with inconsistent distributions (one normally, one not normally

928 distributed). (*) marks refer to the acceptance of the null hypothesis that the two distributions

929 are equivalent for the specific statistic ($\alpha=0.05$). ECT stands for "Equivalence for Central

930 Tendency"; EV for "Equivalence for Variance".

931 **Table 3** Performance statistics (r, NRMSE, MEF, MABstd) are reported as derived from daily

932 and monthly series of GPP_{EC} and GPP_{MD} values over long-term annual scale, for the different

933 forest structure simulations. The (*) refers to p-value < 0.0001 in correlation between GPP_{EC}

934 and GPP_{MD} data. In addition, long term average of annual GPP_{MD} and GPP_{EC} values ($gC\ m^{-2}$

935 yr^{-1}) for the different forest structures are shown.

936

937

938

939

940

941

942

943

944

945

946

Figure captions

Figure 1 3D-CMCC FEM performance indices at different time scales; daily (Figure 1-a) and daily aggregated to month (Figure 1-b) for X dataset. Figure 1-c and 1-d refer to Y daily and Y monthly dataset following decomposition technique proposed in Zhao et al. (2012). DE-Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

Figure 2 Taylor diagrams for daily (a) and daily aggregated to month (b) GPP evaluated representing: the deviation of model results from observations in terms of normalized standard deviation of observations, represented by the distance from the site point to the point on the x-axis identified as reference (REF); the difference of model normalized standard deviation from that of observations, represented by the distance of the site point with respect to the quarter arc crossing REF; and the correlation, given by the azimuthal position of the site point to the x-axis. The sites are numbered in ascending order as follows: (1) DE-Hai, (2) DK-Sor, (3) FR-Hes, (4) IT-Col, (5) FR-Pue, (6) IT-Cpz, (7) DE-Tha, (8) FI-Hyy, (9) IT-Ren, (10) BE-Bra. Colors refer to different IGBP vegetation classes: DBF (yellow), EBF (orange), ENF (light-blue), MF (green).

Figure 3 Distributions of annual GPP ($gC\ m^{-2}\ yr^{-1}$). MD (red) are model results, EC (blue) measured by eddy covariance. The vertical bars represent $\pm$ 1 standard deviation. DE-Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

Figure 4 3D-CMCC FEM performances indices of daily GPP at different seasons. DE-Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

Figure 5 3D-CMCC FEM performances indices of daily GPP aggregated to months at different seasons. DE-Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

Figure 6 Comparison between GPP_{MD} and GPP_{EC} data. The top plots show the average $GPP_{EC}:GPP_{MD}$ correlation for (left) monthly ($gC\ m^{-2}\ month^{-1}$) and (right) daily ($gC\ m^{-2}d^{-1}$) data. The bottom plots show absolute difference range between GPP_{MD} and GPP_{EC} while increasing GPP_{EC} values. Negative values are excluded because of model assumptions. DE-

978   Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L

979   simulation (see text).

980   **Figure 7** Seasonal (monthly) cycle of GPP across the ten sites. The grey line and margins of

981   the grey area represent long-term average of monthly GPP_{EC} ($gCm^{-2}month^{-1}$) and its ±1

982   standard deviation, respectively. The green and red dashed lines represent the long-term

983   average of monthly GPP_{MD} ($gCm^{-2}month^{-1}$) and its ±1 standard deviation, respectively. DE-

984   Tha refers to the 1S simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L

985   simulation (see text).

986   **Figure 8** Distribution of the magnitude for the inter-monthly variability values (IMVs, $gC\ m^{-}$

987     $^{2}d^{-1}$) for each specific site, resulted by standard kernel density estimation. The vertical red line

988   is the media, the box plot limit the 25th and 75th percentiles, the dashed black bars represent

989   the rest of the distribution range excluding outliers (red crosses) DE-Tha refers to the 1S

990   simulation, IT-Ren to the 2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

991   **Figure 9** Inter-Annual Variability (IAV) based on Keenan et al. (2012). Red and blue bars

992   indicate the observed and modelled IAV values, respectively; *r* values refer to correlation

993   between observed and modelled variations. DE-Tha refers to the 1S simulation, IT-Ren to the

994   2L_2C simulation, BE-Bra to the P_Q-3L simulation (see text).

1   Table 1.

Site Name (Site code)	Lat (°) / Lon (°)	IGBP	Simulation year (Starting - Ending)	Mean Annual Temperature (°C)	Annual Precipitation Amount (mm yr^{-1})	Elevation (m a.s.l.)	Main species and forest description	Main references
Hainich (DE-Hai)	51.08 / 10.45	DBF	2000 – 2007	8.3	720	445	Uneven-aged, unmanaged multi-layered forest of beech (*Fagus sylvatica*, 250 years, mean DBH 30.8 cm, mean tree height 23.1 m, stand density 334 trees/ha)	Knohl et al., 2003 + BADM files
Sorø (DK-Sor)	55.49 / 11.64	DBF	2001 – 2009	8.2	660	40	Beech (*Fagus sylvatica*, averagely 80 years, mean DBH 36.13 cm, mean tree height 25 m, stand density 283 trees/ha)	Pilegaard et al., 2003 + BADM files
Hesse (FR-Hes)	48.67 / 7.07	DBF	2001 – 2007	9.2	820	300	Beech (*Fagus sylvatica*, averagely 35 years, mean DBH 8.19 cm, mean tree height 13 m, stand density 3384 trees/ha)	Granier et al., 2000 + BADM files
Collelongo (IT-Col)	41.85 /13.59	DBF	1997 – 2012	6.3	1180	1550	Beech (*Fagus sylvatica*, averagely 100 years, mean DBH 20.2 cm, mean tree height 19.8 m, stand density 900 trees/ha)	Scartazza et al., 2013 + BADM files
Puechabon (FR-Pue)	43.74 /3.60	EBF	2000 – 2011	13.5	883	270	Holm oak (*Quercus ilex*, averagely 59 years, mean DBH 7 cm, mean tree height 6 m, stand density 8500 trees/ha)	Loustau et al., 2005 + BADM files
Castelporziano (IT-Cpz)	41.71 /12.38	EBF	2000 – 2008	15.6	780	3	Holm oak (*Quercus ilex*, averagely 45 years, mean DBH 16 cm, average tree height 12.5 m, stand density 458 trees/ha)	Vitale et al., 2003 + BADM files
Tharandt (DE-Tha)	50.96/ 13.57	ENF	2000 – 2010	7.7	820	380	Mixed Norway spruce (*Picea abies*, averagely 113 years, mean DBH 33 cm, tree height 26, density 396 trees/ha) and Scots Pine (*Pinus sylvestris*, averagely 113 years, mean DBH 33.1 cm, tree height 26.1 m, density 81 trees/ha)	Grünwald & Bernhofer, 2007 + BADM files
Hyytiälä (FI-Hyy)	61.85 / 24.29	ENF	2001 – 2011	3.8	709	170	Scots pine (*Pinus sylvestris*, 39 years, mean DBH 30.8 cm, mean tree height 23.1 m, stand density 334 trees/ha)	Suni, et al., 2003 + BADM files

Renon (IT-Ren)	46.59 / 11.43	ENF	2006 - 2010	4.7	809	1735	Uneven-aged multi-layered forest of Norway spruce (*Picea abies* averagely, 190 and 30 years, average DBH 30.8 cm, average tree height 23.1m, stand density 334 trees/ha)	Montagnani et al.(2009) + BADM files
Brasschaat (BE-Bra)	51.30/4.52	MF	2001 - 2010	9.8	750	16	Mixed, uneven-aged multi-layered forest of Scots pine (*Pinus sylvestris*, averagely 72 years) and Pedunculate oak (*Quercus robur*, averagely 65 years)	Gielen et al., 2013 + BADM files

1

1 *Table 2*

		DE-Hai	DK-Sor	FR-Hes	IT-Col	FR-Pue	IT-Cpz	DE-Tha (1S)	FI-Hyy	IT-Ren (2L-2C)	BE-Bra (P_Q-3L)
NRMSE	IAVs	2.4	1.8	1.3	0.3	0.6	1.1	1.0	2.7	1.3	0.9
NRMSE	IMVs	1.7	2.7	1.1	0.6	1.1	1.2	1.1	1.2	1.0	0.5
ECT	p-value	1.00*[N]	0.12*[N]	0.54*[N]	0.00 [N]	0.15*[N]	1.00*[na]	1.00*[P]	0.04 [N]	0.88*[P]	0.85*[N]
EV	p-value	0.53*[N]	0.00 [N]	0.00 [N]	0.46*[N]	0.00 [N]	0.02 [na]	0.78*[P]	0.00 [N]	0.27*[P]	0.01 [N]

2

3

1   *Table 3.*

2

Site	Model set-up code	Daily				Monthly				Yearly	
		r	NRMSE	MEF	MABstd	r	NRMSE	MEF	MABstd	GPP$_{MD}$ gC m^{-2} yr^{-1}	GPP$_{EC}$ gC m^{-2} yr^{-1}
BE-Bra	P	0.72*	0.73	0.47	0.51	0.86*	0.55	0.70	0.39	1003	
	Q_3L	0.76*	0.91	0.18	0.67	0.84*	0.71	0.49	0.52	1105	
	Q_2L	0.74*	0.89	0.21	0.66	0.86*	0.74	0.45	0.55	1179	
	Q_1L	0.75*	0.95	0.01	0.70	0.86*	0.68	0.53	0.50	1147	
	P_Q-3L	0.77*	0.65	0.57	0.46	0.93*	0.39	0.84	0.28	1141	1193
	P_Q-2L	0.75*	0.67	0.55	0.46	0.91*	0.44	0.81	0.30	1037	
	P_Q-1L	0.75*	0.66	0.56	0.46	0.91*	0.68	0.53	0.50	1056	
IT-Ren	**2L_2C**	0.82*	0.62	0.61	0.44	0.95*	0.30	0.91	0.23	1349	1362
	1L_1C	0.83*	0.85	0.27	0.61	0.96*	0.61	0.62	0.45	1950	
DE-Tha	**1S**	0.90*	0.46	0.79	0.31	0.96*	0.27	0.93	0.19	1840	1869
	2S	0.89*	0.48	0.80	0.31	0.95*	0.29	0.91	0.19	1898	

3

a   X dataset daily

b   X dataset monthly

c   Y dataset daily

d   Y dataset monthly
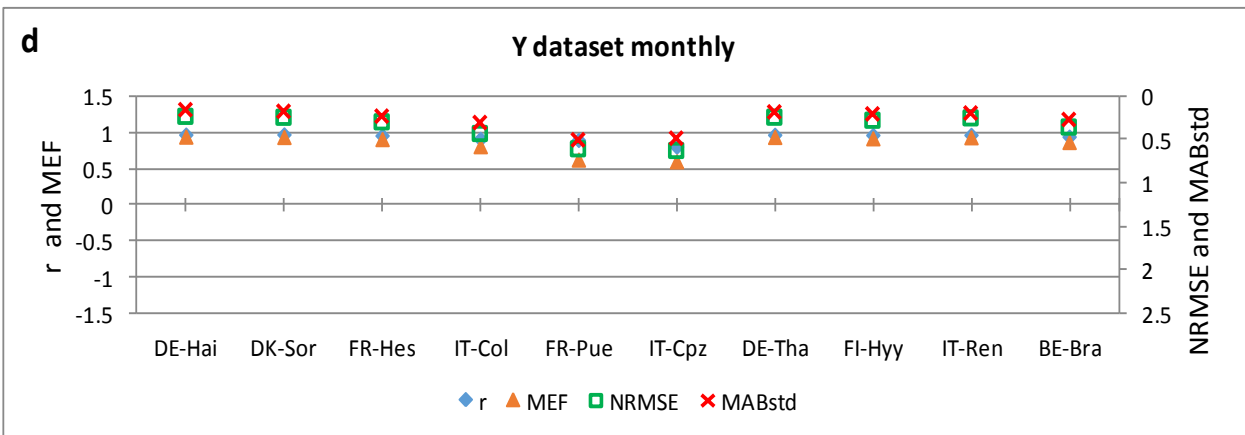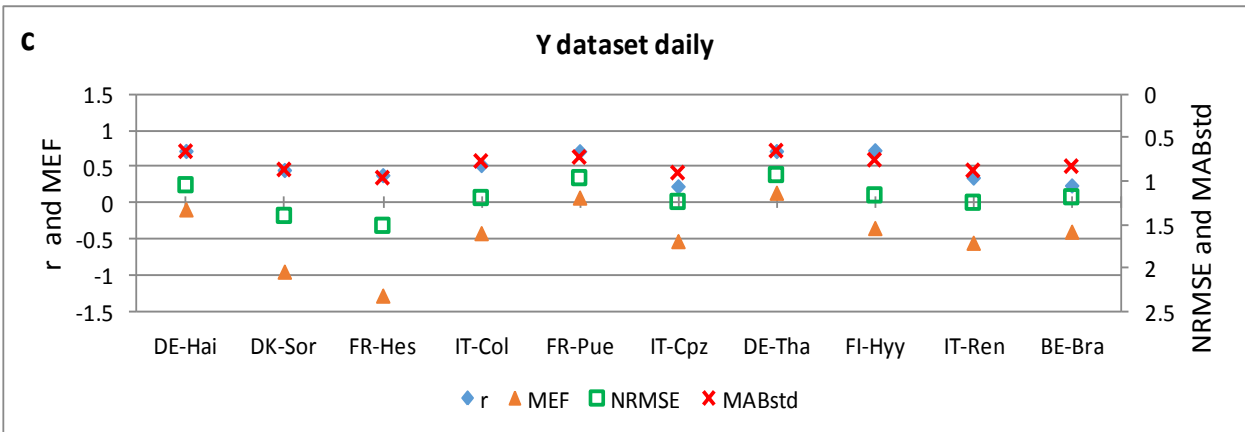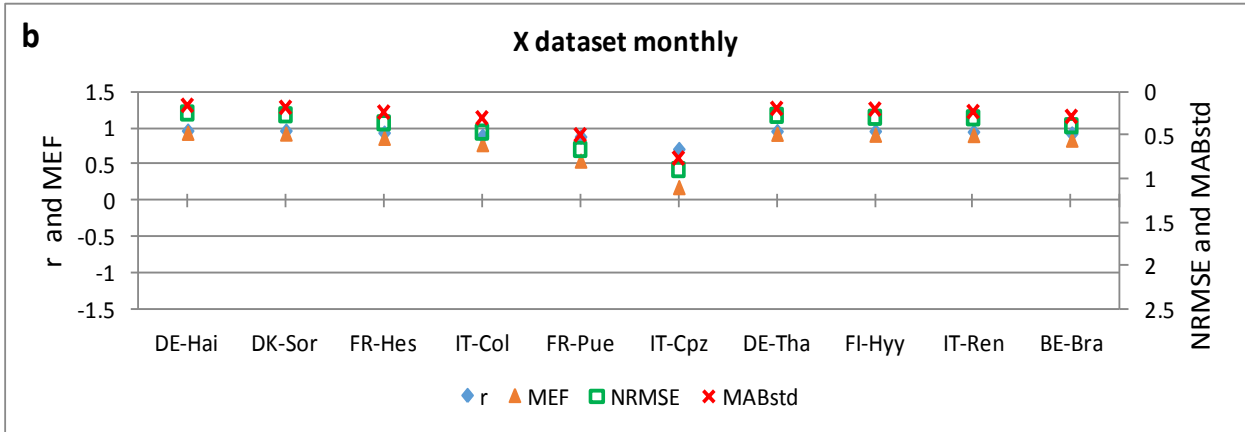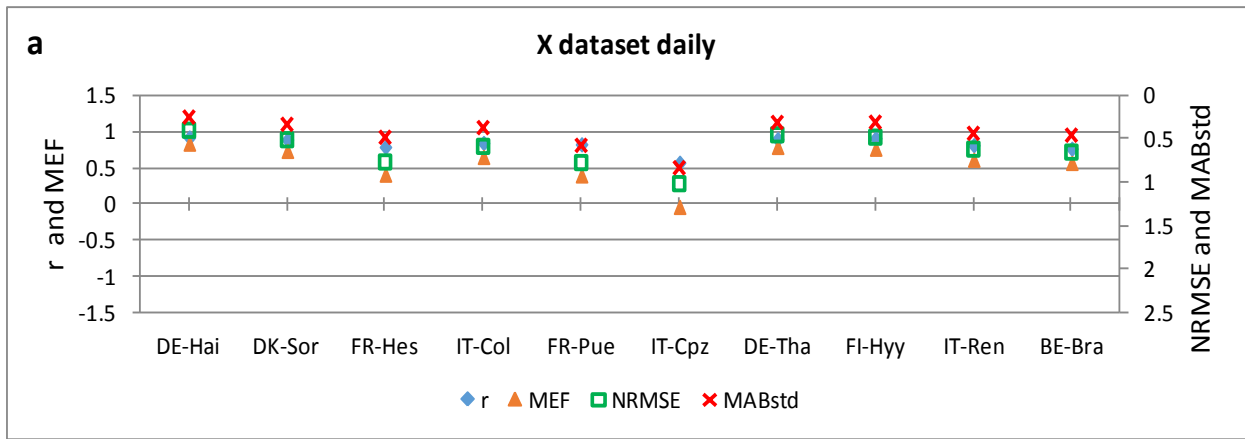
Figure 1

4

MEAN DAILY CYCLE
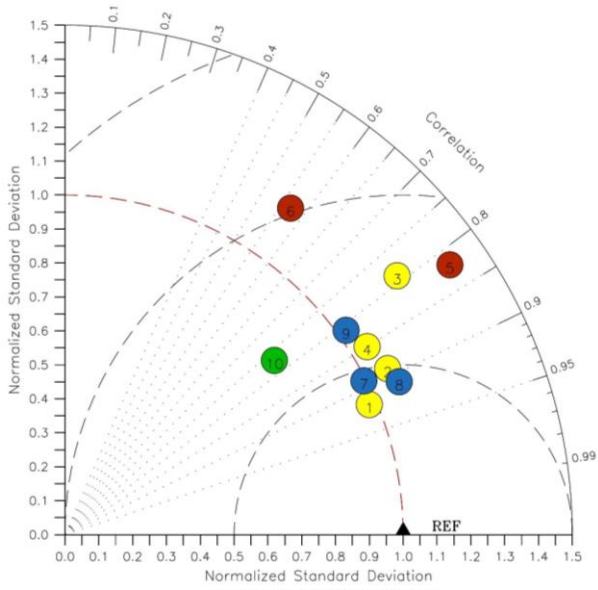
5

MEAN MONTHLY CYCLE
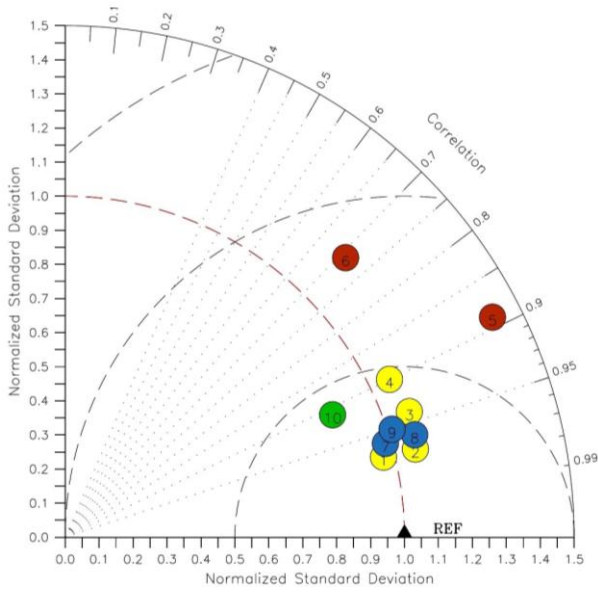
6

7   *Figure 2*
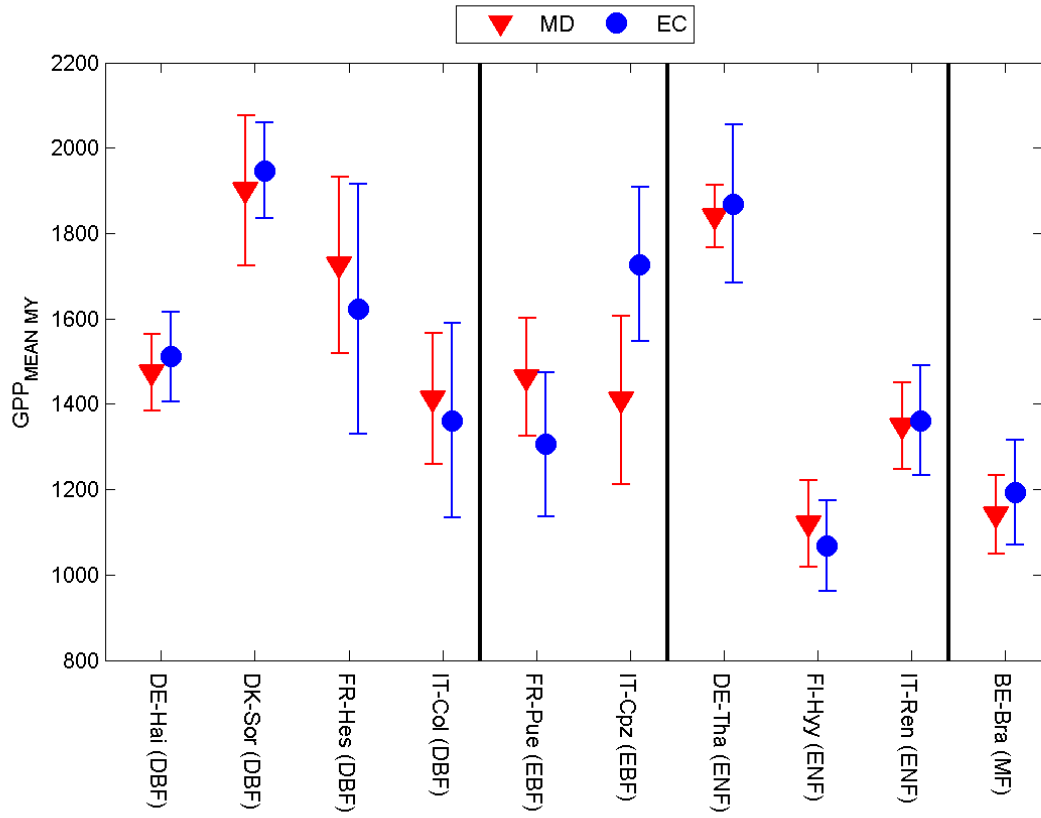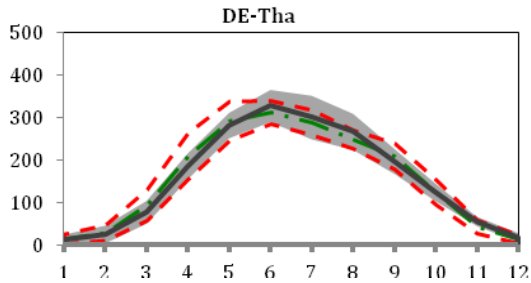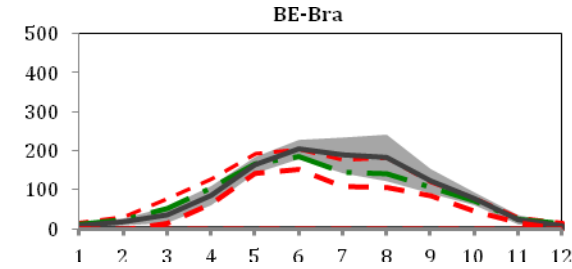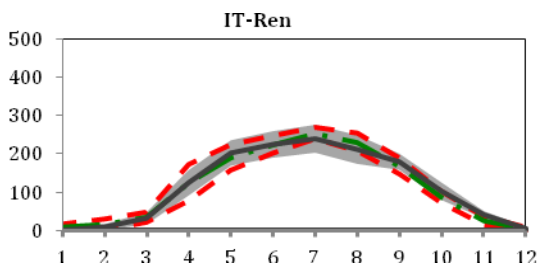
Figure 3

8

9

10

Figure 4

Figure 5

12
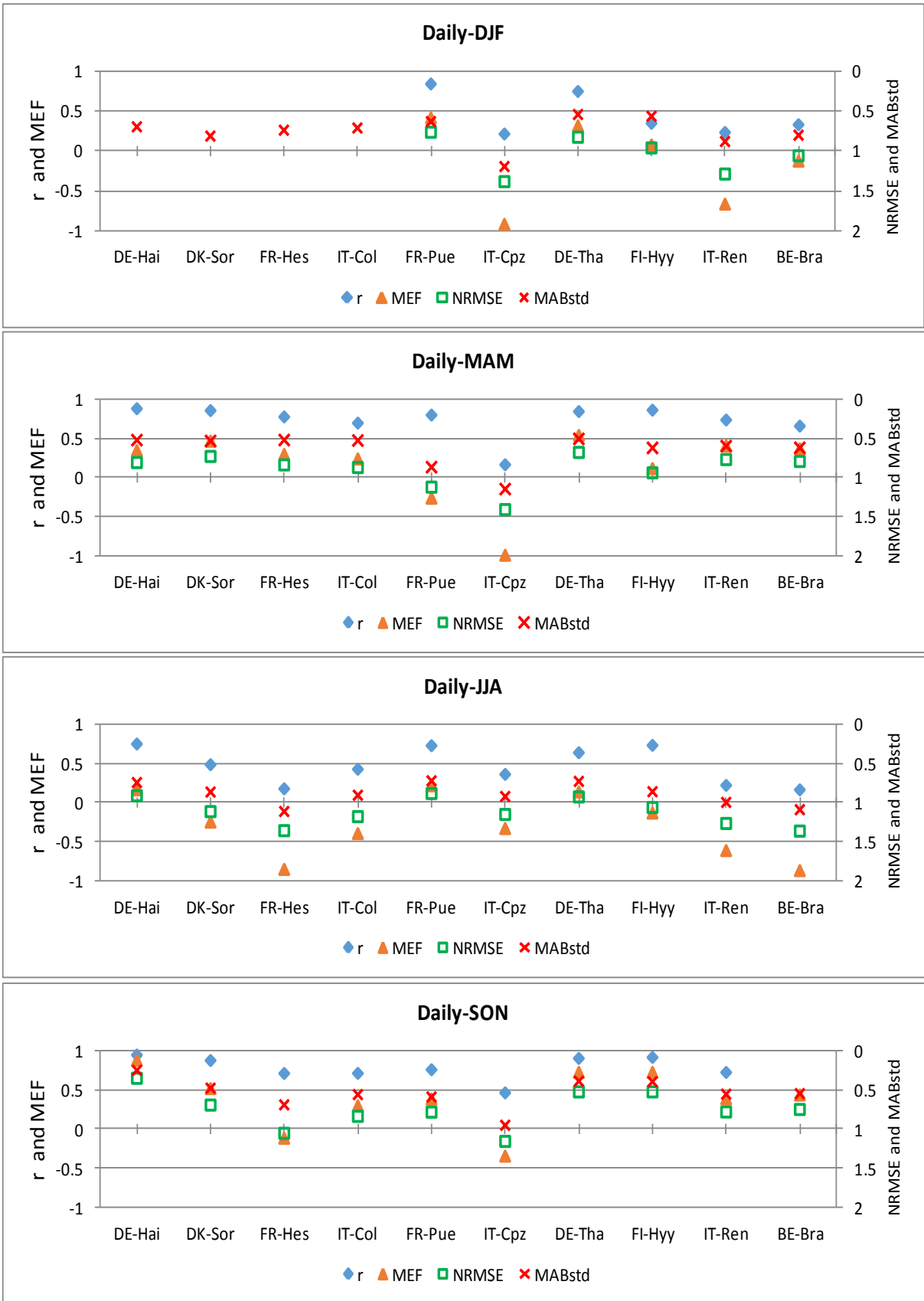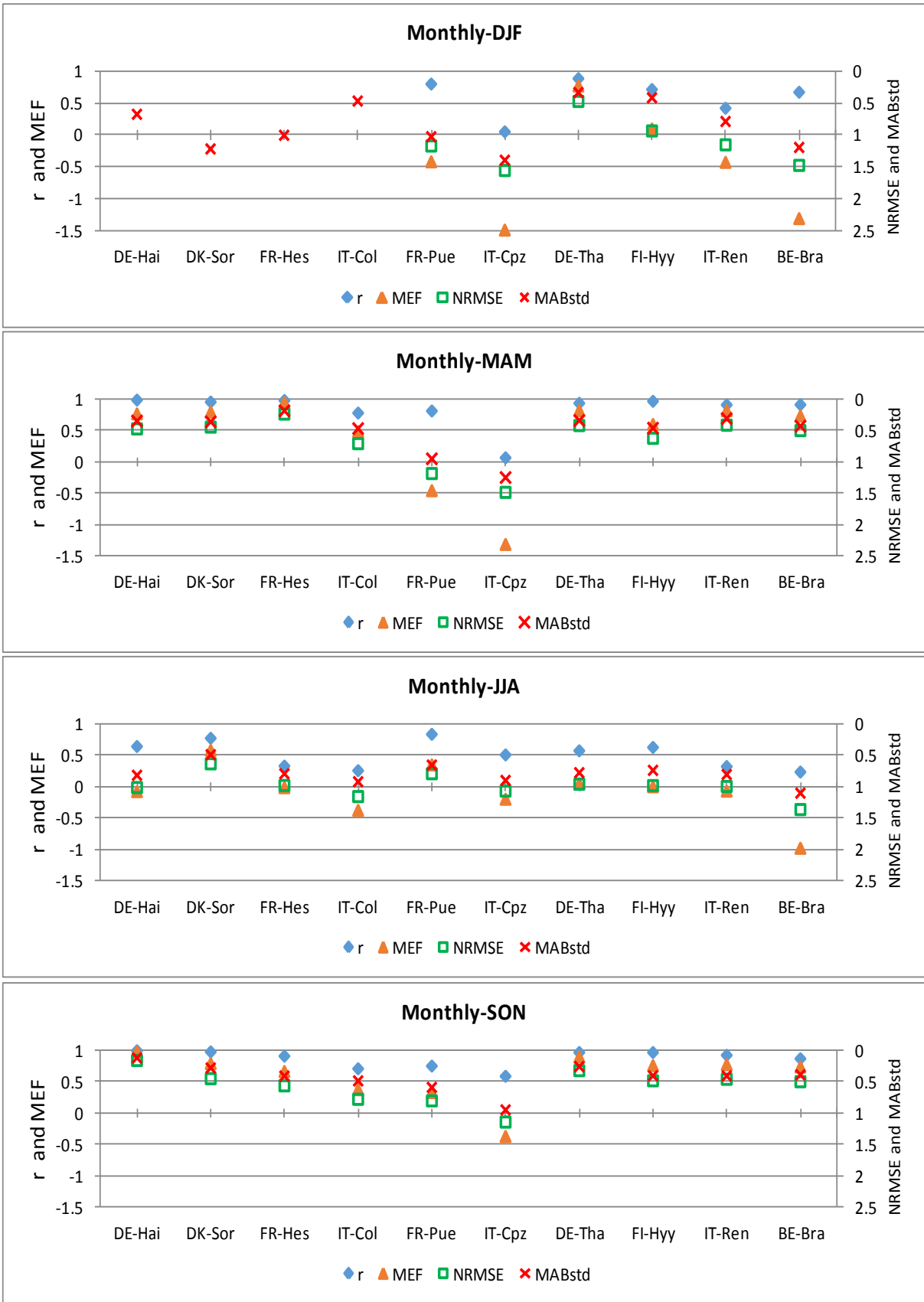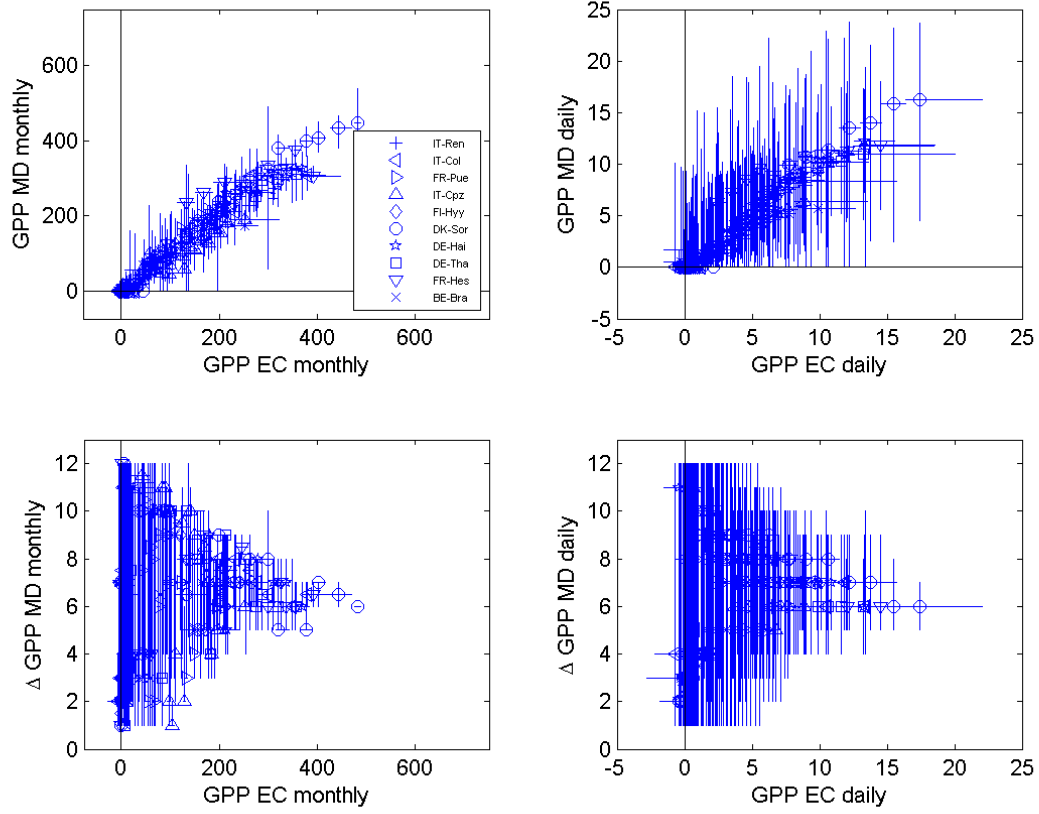
13

14   *Figure 6*

15

16
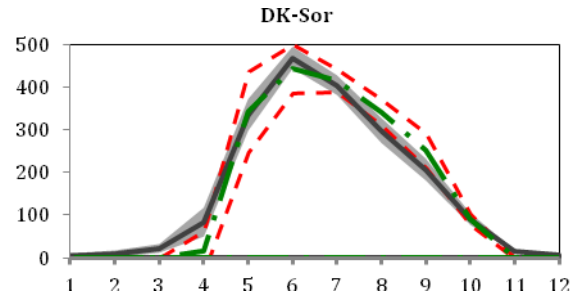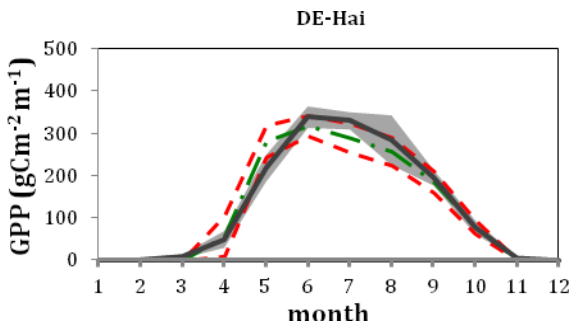
17
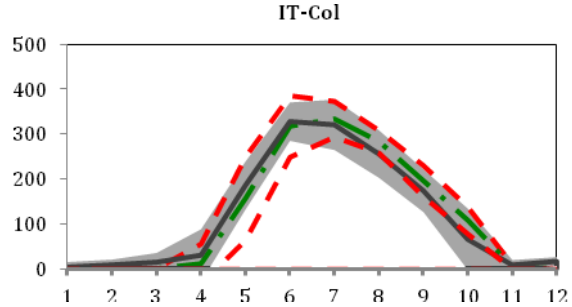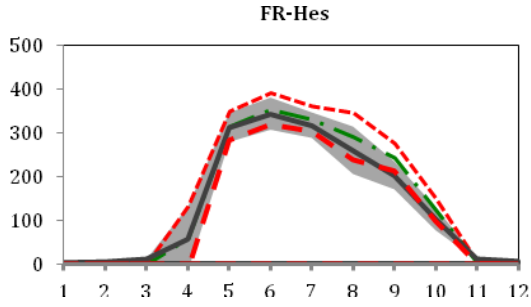
18
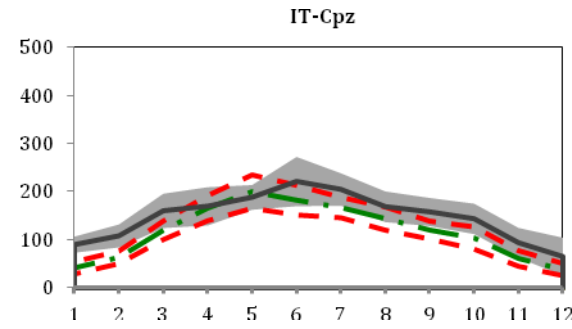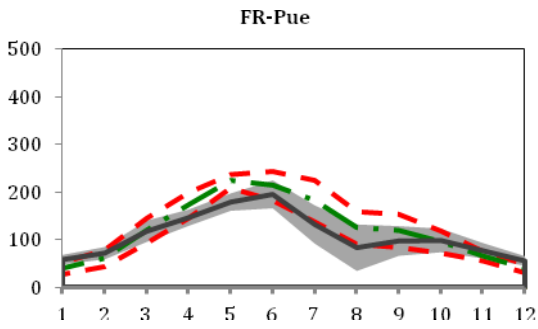
19
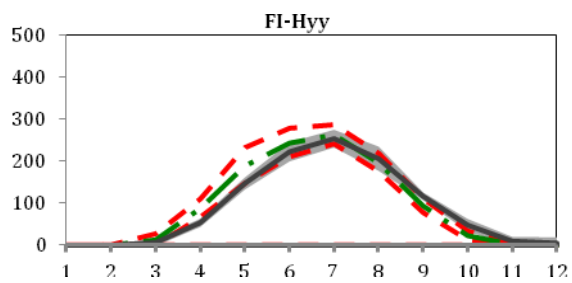
20

21   *Figure 7*

22
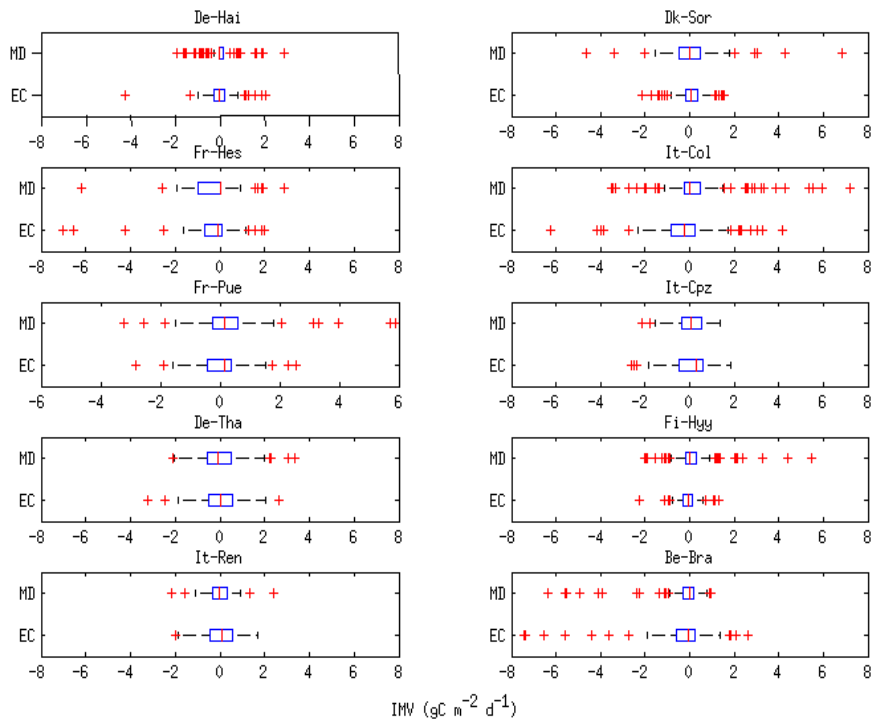
23

24   *Figure 8*

25

26
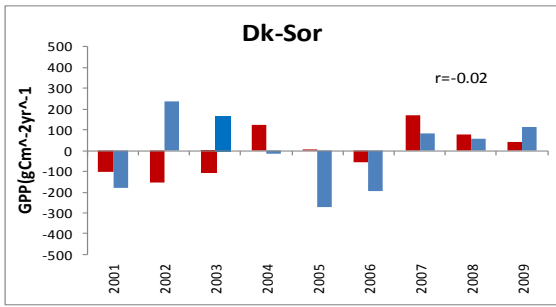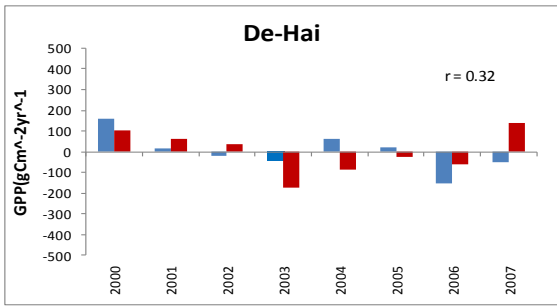
27
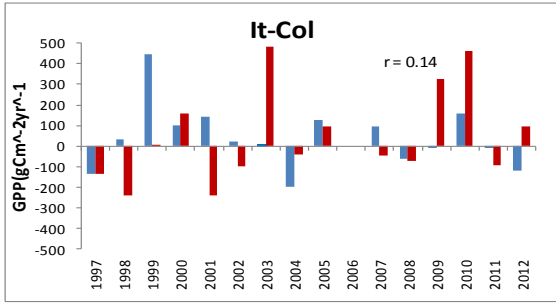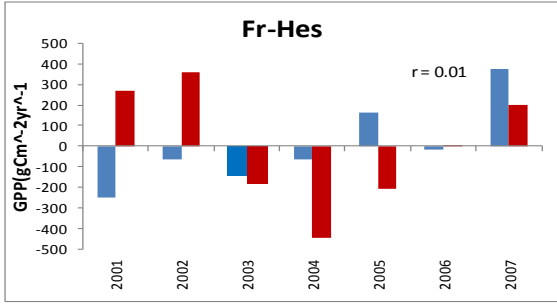
28
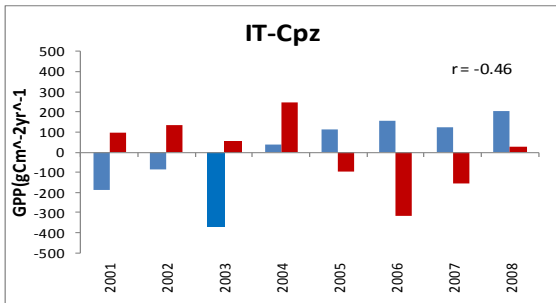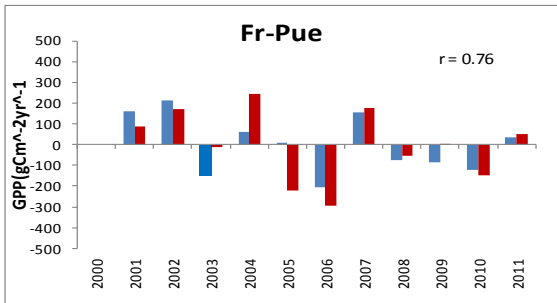
29

30

31   *Figure 9*

32

33

References

Anav, A., D'Andrea, F., Viovy, N., & Vuichard, N. (2010). A validation of heat and carbon fluxes from high-resolution land surface and regional models. *Journal of Geophysical Research* , 115: 1-20.

Arneth, A., Sitch, S., Bondeau, A., Butterbach-Bahl, K., Forestr, P., Gedney, N., et al. (2010). From biota to chemistry and climate: towards a comprehensive description of trace gas exchange between the biosphere and atmosphere. *BioGeosciences* , 7, 121-149.
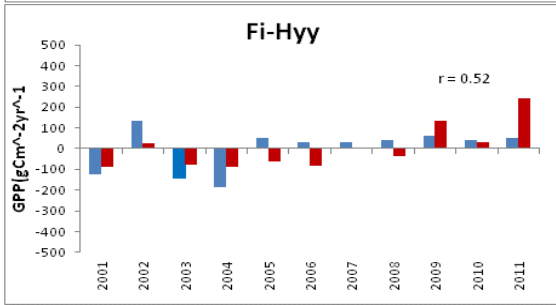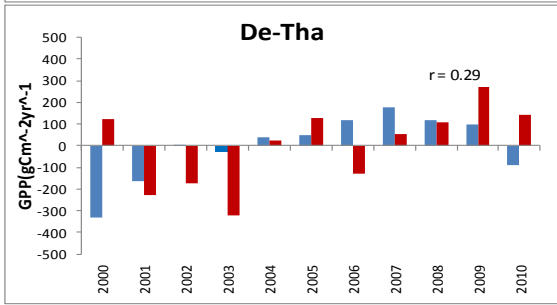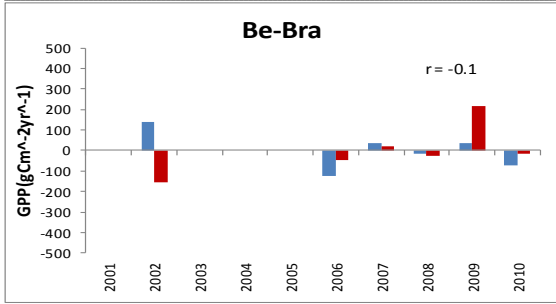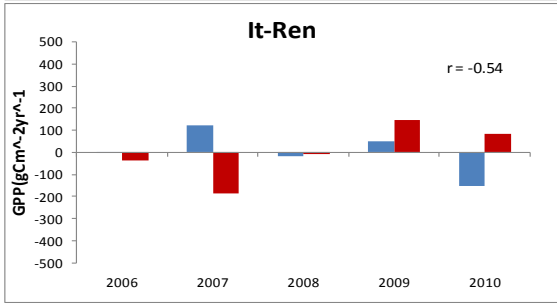
Arora, V. K., & Boer, G. J. (2005). A parameterization of leaf phenology for the terrestrial ecosystem component of climate models. *Global Change Biology* , 11: 39-59.

Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., et al. (2000). Estimates of the annual net carbon and water exchange of European forests: the EUROFLUX methodology. *Advances in Ecological Research* , 30, 113-175.

Bagnara, M., Van Oijen, M., Cameron, D., Gianelle, D., Magnani, F., & Sottocornola, M. (2014). A user-friendly forest model with a multiplicative mathematical structure: a Bayesian approach to calibration. *Geosci. Model Dev. Discuss.* , 7, 6997-7031.

Baldocchi, D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology* , 9, 479-492.

Balzarolo, M., Boussetta, S., Balsamo, G., Beljaars, A., Miagnan, F., Calvet, J.-C., et al. (2014). Evaluating the potential of large-scale simulations to predict carbon fluxes of terrestrial ecosystems over a European Eddy Covariance network. *Biogeosciences* , 11-2661,2678.

Barbaroux, B., & Bréda, N. (2002). Contrasting distribution and seasonal dynamics of carbohydrate reserve in stem wood of adult ring-porous sessile oak and diffuse-porous beech trees. *Tree Physiology* , 1201-1210.

Batlle Bayer, L., van den Hurk, B., Strengers, B., & van Minnen, J. (2012). Regional feedbacks under changing climate and land-use conditions. *Earth Syst. Dynam. Discuss.* , 3, 201-234.

Battin, T., Luyssaert, S., Kaplan, L., Aufdenkampe, A., Richter, A., & Tranvik, L. (2009). The boundless carbon cycle . *Nature Geoscience* , 2, 598-600.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2012). Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science* , 329, 834-838.

Blyth, E., Clark, D., Ellis, R., Huntingford, C., Los, S., Pryor, M., et al. (2011). A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.* , 4, 255-269.

Bolstad, P., Mitchell, K., & Vose, J. (1999). Foliar response functions for broad-leaved tree species in the southern Appalachians. *Tree Physiology* , 19, 871-878.

Bonan, G. B. (2008). Forest and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* , 320, 1444-1449.

Bond-Lamberty, B., Gower, S., Ahl, D., & Thornton, P. E. (2005). Reimplementation of the Biome-BGC model to simulate successional change. *Tree Physiology* , 25, 413-424.

Bowman, A., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis.* New York: Oxford University Press Inc.

Brèda, N., Huc, R., Granier, A., & Dreyer, E. (2006). Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. *Ann. For. Sci.* , 625-644.

Breuer, L., Eckhardt, K., & Frede, H.-G. (2003). Plant parameter values for models in temperate climates. *Ecological Modelling* , 169:237-293.

Campioli, M., Verbeeck, H., Van den Bossche, J., Wu, J., Ibrom, A., D'Andrea, E., et al. (2013). Can decision rules simulate carbon allocation for years with contrasting and extreme weather conditions? A case study for three temperate beech forests. *Ecological Modelling* , 263, 42-55.

Cescatti, A., Marcolla, B., Santhana, S., & al., e. (2012). Remote sensing of environment intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network. *Remote Sensing of Environment* , 121, 323-334.

Chabot, B. F., & Hicks, D. (1982). The ecology of leaf life spans. . *Annual Review of Ecology and Systematics* , 229-259.

Chai, T., & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* , 7, 1247-1250.

Chang, J., Viovy, N., Vuichard, N., Ciais, P., Wang, T., Cozic, A., et al. (2013). Incorporating grassland management in ORCHIDEE: model description and evaluation at 11 eddy-covariance sites in Europe. *Geosci. Model Dev.* , 6, 2165-2181.

Chapin, F., Matson, P., & Vitousek, P. (2002). *Principles of Terrestrial Ecosystem Ecology.* New York, NY, USA: Springer.

Chesson, P. (2000). Mechanism of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* , 31, 343-366.

Churkina, G., Schimel, D., Braswell, B., & Xiao, X. (2002). Spatial analysis of growing season length control over net ecosystem exchange. *Gobal Change Biology* , 11, 1777-1787.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., et al. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* , 437: 529-533.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., et al. (2013). Carbon and Other Biogeochemical Cycles. In T. Stocker, D. Qin, G. -K. Plattner, M. Tignor, S. Allen, J. Boschung, et al., *Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description - Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.* , 4, 701-722.

Clark, J., Bell, D., Chu, C., Courbaud, B., Dietze, M., Hersh, M., et al. (2010). High-dimensional coexistence based on individual variation: a synthesis of evidence. *Ecol. Monogr.* , 80, 569-608.

Collalti, A. (2011). *Sviluppo di un modello ecologico-forestale per foreste a struttura complessa.* Viterbo, University of Tuscia: PhD. Thesis.

Collalti, A., Perugini, L., Santini, M., Chiti, T., Nolè, A., Matteucci, G., et al. (2014). A process-based model to simulate growth in forests with complex structure:Evaluation and use

of 3D-CMCC Forest Ecosystem Model in a deciduous forest in Central Italy. *Ecological Modelling* , 362-378.

Cox, P., & Jones, C. (2008). Illuminating the MOdern Dance of Climate and CO2. *Science* , 321, 1642-1644.

Curiel Yuste, J., Janssens, I., & Cuelemans, R. (2005). Calibration and validation of an empirical approach to model soil CO2 efflux in a deciduous forest. *Biogeochemistry* , 73: 209-230.

Dalmonech, D., Zaehle, S., Schurmann, G., Brovkin, V., Reick, C., & Schnur, R. (2015). Separation of the Effects of Land and Climate Model Errors on Simulated Contemporary Land Carbon Cycle Trends in the MPI Earth System Model version 1. *J. Climate* , 28, 272-291.

Davidson, E., Janssens, I., & Luo, Y. (2006). On the variability of respiration in terrestrial ecosystems : moving beyond Q(10). *Global Change Biology* , 12: 154-164.

De Pury, D., & Farquhar, G. (1997). Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Pant Cell Environ.* , 20, 537-557.

Delpierre, N., Soudani, K., Francois, C., Kostner, B., Pontailler, J.-Y., Nikinmaa, E., et al. (2009). Exceptional carbon uptake in European forests during the warm spring of 2007: a data-model analysis. *Global Change Biology* , 15, 1455-1474.

Dickmann, D., & Kozlowski, T. (1970). Photosynthesis by rapidly expanding green strobili of Pinus resinosa. *Life science* , 9, 549-552.

Dietze, M. C., Vargas, R., Richardson, A. D., Stoy, P. C., Barr, A., & Anderson, e. a. (2011). Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis,. *Journal of Geophysical Research* .

Dufrene, E., Davi, H., Fracois, C., le Maire, G., Le Dantec, V., & Granier, A. (2005). Modelling carbon and water cycles in beech forest. Part I: Model description and uncertainty analysis on modelled NEE. *Ecological Modelling* , 185: 407-436.

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., et al. (2002). Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agrocoltural and Forest Meteorology* , 113, 53-74.

151 Farquhar, G., & Sharkey, T. (1982). Stomatal condictance and photosynthesis. *Annual Review*
152 *of Plant Physiology* , 33, 317-345.

153 Farquhar, G., von Caemmerer, S., & Berry, J. (1980). A biogeochemical model of
154 photosynthetic CO2 assimilation in leaves of C3 species. *Planta* , 149, 78-90.

155 Feikema, P., Morris, J., Beverly, C. C., Baker, T., & Lane, P. (2010). Validation of plantation
156 transpiration in south-eastern Australia estimated using the 3PG+ forest growth model. *Forest*
157 *Ecology and Management* , 260, 663-678.

158 Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., et al. (2006).
159 Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J.*
160 *Climate* , 19, 3337-3353.

161 Gielen, B., De Vos, B., Campioli, M., Neirynck, J., Papale, D., Verstraen, A., et al. (2013).
162 Biometric and eddy covariance-based assessment of decadal carbon sequestration of a
163 temperate Scots pine forest. *Agricoltural and Forest Meteorology* , 174-175; 135-143.

164 Gough, C., Flower, C., Vogel, C., Dragoni, D., & Curtis, P. (2009). Whole-ecosystem labile
165 carbon production in a north temperate deciduous forest. *Agricultural and Forest*
166 *Meteorology* , 149, 1531-1540.

167 Granier, A., Ceschia, E., Damesin, C., Dufrene, E., Epron, D., Gross, P., et al. (2000). The
168 carbon balance of a youg beech forest. *Functional Ecology* , 14, 312-325.

169 Granier, A., Reichstein, M., Brèda, N., Janssens, I., Falge, E., Ciais, P., et al. (2007).
170 Evidence for soil water control on carbon and water dynamics in European forests during the
171 extremely dry year: 2003. *Agricultural and Forest Meteorology* , 143, 123-145.

172 Griffin, K., Tissue, D., Turnbull, M., Schuster, W., & Whitehead, D. (2001). Leaf dark
173 respiration as a function of canopy position in Nothofagus fusca trees grown at ambient and
174 elevated CO2 partial pressure for 5 years. *Functional Ecology* , 15, 497-505.

175 Grote, R., Korhonen, J., & Mammarella, I. (2011). Challenges for evaluating process-based
176 models of gas exchange at forest sites with fetches of various species. *Forest systems* , 20(3),
177 389-406.

178 Grünwald, T., & Bernhofer, C. (2007). A decade of carbon, water and energy flux
179 measurements of an old spruce forest at the Anchor Station Tharandt. *Tellus* , 59B: 387-396.

180 Gu, L., Baldocchi, D., Verma, D., Black, T., Vesala, T., Falge, E., et al. (2002). Advantages
181 of diffuse radiation for terrestrial ecosystem productivity . *J. Geophys. Res.* , 107.

182 Guidolotti, G., Rey, A., D'Andrea, E., Matteucci, G., & De Angelis, P. (2013). Effect of
183 environmental variables and stand structure on ecosystem respiration components in a
184 Mediterranean beech forest. *Tree Physiology* , 00, 1-13.

185 He, M. J., Zhou, Y., J.M., C., He, H., Wang, S., Wang, H., et al. (2013). Development of a
186 two-leaf light use efficiency model for improving the calculation of terrestrial gross primary
187 production. *Agr. Forest Meteorol.* , 173, 28-39.

188 Hui, D., Luo, Y., & Katul, G. (2003). Partitioning interannual variability in net ecosystem
189 exchange between climatic variability and functional change. *Tree Physiology* , 23, 433-442.

190 Huntingford, C., Lowe, J., Booth, B., Jones, C., Harris, G., Gohar, L., et al. (2009).
191 Contributions of carbon cycle uncertainty to future climate projection spread. *Tellus B* , 61,
192 355-360.

193 Ibrom, A., Jarvis, P., Clement, R., Morgenstern, K., Oltchev, A., Medlyn, B., et al. (2006). A
194 comparative analysis of simulated and observed photosynthetic CO2 uptake in two coniferous
195 canopies. *Tree Physiology* , 26, 845-864.

196 Ibrom, A., Oltchev, A., June, T., Kreilein, H., Rakkibu, G., Ross, T., et al. (2008). Variation
197 in photosynthetic light-use efficiency in a mountains tropical rain forest in Indonesia. *Tree*
198 *Physiol.* , 28, 499-508.

199 Janssens, I., Freibauer, A., Ciais, P., Smith, P., Nabuurs, G.-J., Folberth, G., et al. (2003).
200 Europe's Terrestrial Biosphere Absorbs 7 to 12% of European Anthropogenic CO2 emissions.
201 *Science* , 300, 1538-1542.

202 Janssens, I., Sampson, D., Curiel Yuste, J., Carrara, A., & Ceulemans, R. (2002). The carbon
203 cost of fine root turnover in a Scots pine forest. *Forest Ecology and Management* , 168: 231-
204 240.

205 Jeong, S., Medvigy, D., Shevliankova, E., & Malyshev, S. (2013). Predicting changes in
206 temperate forest budburst using continental-scale observations and models. *Geophys. Res.*
207 *Lett.* , 40, 1-6.

208 Jolly, W., Dobberlin, M., Zimmerman, N., & Reichstein, M. (2005). Divergent vegetation
209 growth responses to the 2003 heat wave in th Swiss Alps. *Geophys. Res. Lett.* ,
210 doi:10,1029/2005GL023027.

211 Keenan, T., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., et al. (2012). Terrestrial
212 biosphere model performances for inter-annual variability of land-atmosphere CO2 exchange.
213 *Global Change Biology* , 18(6): 1971-1987.

214 Kimball, J., Thornton, P., White, M., & Running, S. (1997). Simulating forest productivity
215 and surface-atmosphere carbon exchange in the BOREAS study region. *Tree Physiology* , 17,
216 589-599.

217 Knohl, A., Schulze, E.-D., Kolle, O., & Buchmann, N. (2003). Large carbon uptake by an
218 unmanaged 250-year-old deciduous forest in Central Germany. *Agricultural and Forest*
219 *Meteorology* , 118, 151-167.

220 Kobe, R. (1996). Intraspecific variation in sapling mortality and growth predicts geographic
221 variation in forest composition. *Ecol. Monogr.* , 66, 181-201.

222 Kolari, P., Pumpanen, J., Kumala, L., Ilvesniemi, H., Nikinmaa, E., Gronholm, T., et al.
223 (2006). Forest floor vegetation plays an important role in photosynthetic production of boreal
224 forests. *Forest Ecol. Manag.* , 221, 241-248.

225 Kramer, K., Leinonen, I., Bartelink, H., Berbigier, P., Borghetti, M., Bernhofer, C., et al.
226 (2002). Evaluation of six process-based forest growth models using eddy-covariance
227 measurements of CO2 and H2O fluxes at six forest sites in Europe. *Global Change Biology* ,
228 8, 212-230.

229 Krishnan, P., Black, T., Barr, A., Grant, N., Gaumont-Guay, D., & Nesic, Z. (2008). Factors
230 controlling the interannual variability in the carbon balance of southern boreal black spruce
231 forest. *Journal of Geophysical Reasearch* , 113, 1-16.

232 Krishnan, P., Black, T., Jassal, R., Chen, B., & Nesic, Z. (2009). Interannual variability of the
233 carbon balance of a southern boreal black spruce forest. *Journal of Geophysical Research* ,
234 114, 1-18.

235 Landhausser, S. (2010). Aspen shoots are carbon autonomus during budbreak. *Tree* , 25- 531-
236 536.

237 Landsberg, J. (1996). *Physiological Ecology of Forest Production.* Sydney, Australia:
238 Academic Press.

239 Landsberg, J., & Waring, R. (1997). A generalised model of forest productivity using
240 simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest*
241 *Ecology and Management* , 172, 199–214.

242 Larcher, W. (2003). *Physiological plant ecology: ecophysiology and stress physiology of*
243 *functional groups.* New York: Springer.

244 Lawrence, D., Oleson, K., Flanner, M., Thornton, P., Swenson, S., Lawrence, P., et al. (2011).
245 Parameterization Improvements and Fuctional and Structural Advances in Version 4 of the
246 Community Land Model. *J. Adv. Model. Earth Syst* , 3, 1-27.

247 Leuning, R., Kelliher, F., DePury, D., & Schulze, E. (1995). Leaf nitrogen, photosynthesis,
248 conductance and transpiration: Scaling from leaves to canopy . *Plant Cell Environ.* , 18, 1183-
249 1200.

250 Li, H., Sheffield, J., & Wood, E. (2010). Bias correction of monthly precipitation and
251 temperature fields from Intergovernmental Panel on Climate Change AR4 models using
252 equidistant quantile matching. *Journal of Geophysical Research* , DOI:
253 10.1029/2009JD012882.

254 Liu, Z., Wang, L., & Wang, S. (2014). Comparison of Different GPP Models in China Using
255 MODIS Image and ChinaFLUX Data. *Remote Sensing* , 6, 10215-10231.

256 Loustau, D., Bosc, A., Colin, A., Ogèe, J., Davi, H., Francois, C., et al. (2005). Modelling
257 climate change effects on the potential production of French plains forests at the sub-regional
258 level. *Tree Physiology* , 25, 813-823.

259 Luyssaert, S., Abril, G., Andres, R., Bastviken, D., Bellasen, V., Bergamaschi, P., et al.
260 (2012). The European CO2, CO, CH4, and N2O balance between 2001 and 2005 .
261 *Biogeosciences Discussion* , doi:10.5194/bgd-9-2005-2012.

262 Luyssaert, S., Inglima, I., Jung, M., Richardson, D., Reichstein, M., Papale, D., et al. (2007).
263 CO2 balance of boreal, temperate, and tropical forests derived from a global database. *Global*
264 *Change Biology* , 13, 2509-2537.

265 Magnani, F., Mencuccini, M., Borghetti, M., Berbigier, P., Delzon, S., & al., e. (2007). The
266 human footprint in the carbon cycle of temperate and boreal forests. *Nature* , 447, 848-850.

267 Mahecha, M., Reichstein, M., Carvalhais, N., Lasslop, G., Lange, H., Seneviratne, S., et al.
268 (2010). Global Convergence in the Temperature Sensitivity of Respiration at Ecosystem
269 Level. *Science* , 329, 838-840.

270 Makela, A., Landsberg, J., Ek, A., Burk, T., Ter-Mikaelian, M., Agren, G., et al. (2000).
271 Process-based models for forest ecosystem management: current state of the art and
272 challenges for practical implementation. *Tree Physiology* , 20: 289,298.

273 Makela, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., et al. (2008).
274 Developing an empirical model of stand GPP with the LUE approach: analysis of eddy
275 covariance data at five contrastinf conifer sites in Europe. *Global Change Biology* , 14, 92-
276 108.

277 Marconi, S. (2014). *Assessing NEE and Carbon Dynamics among 5 European Forest types:*
278 *Development and Validation of a new Phenology and Soil Carbon routines within the process*
279 *oriented 3D-CMCC-Forest Ecosystem Model.* Viterbo: Master thesis, University of Tuscia.

280 Marks, D., Dozier, J., & Davies, R. (1992). Climate and energy exchange at the snow surface
281 in the alpine region of the Sierra Nevada. I. Meteorology measurments and monitoring. *Water*
282 *Resouces Research* , 4, 719-739.

283 Medlyn, B., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P., Kirschbaum, M., et al.
284 (2003). Temperature response of parameters of a biogeochemically based model of
285 photosynthesis. II. A review of experimental data. *Plant, Cell & Environment* , 25, 1167-
286 1179.

287 Mercado, L., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., et al. (2009).
288 Impact of changes in diffuse radiation on the global land carbon sink. *Nature* , 458, 1014-
289 1017.

290 Migliavacca, M., Reichstein, M., Richardson, A., Mahecha, D., Cremonese, E., DelPierre, N.,
291 et al. (2015). Influence of physiological phenology on the seasonal patterns of ecosystem
292 respiration in deciduous forests. *Global Change Biology* , 21, 363-376.

293 Misson, L., Baldocchi, D., Black, T., Blanken, P., Brunet, Y., Curiel Yuste, J., et al. (2007).
294 Partitioning forest carbon fluxes with overstory and understory eddy-covariance
295 measurements: A synthesis based on FLUXNET data. *Agricoltural and Forest Meteorology* ,
296 144, 14-31.

297 Mollicone, D., Matteucci, G., Koble, R., Masci, A., Chiesi, M., & Smits, P. (2003). A Model-
298 Based Approach for the Estimation of Carbon Sinks in European Forests. In R. Valentini,
299 *Fluxes in Carbon, Water and Energy of European Forests* (pp. 164, 179-205). Heidelberg:
300 Springer-Verlag .

301 Montagnani, L., Manca, G., Canepa, E., Georgieva, E., Acosta, M., Feigenwinter, C., et al.
302 (2009). A new mass conservation approach to the study of CO_2 advection in an alpine forest.
303 *Journal of Geophysical Research* , 114: D07306.

304 Monteith, J. (1977). *Climate and the efficiency of crop production in Britain.* London: Phil.
305 Trans. Roy. Soc.

306 Monteith, J. (1965). Evaporation and environment. In J. Monteith, *Symposium, Society of*
307 *Experimental Botany* (pp. 19, 205-234). Cambridge: Cambridge University Press.

308 Morales, P., Sykes, M., Prentice, I., Smith, P., Smith, B., Bugmann, H., et al. (2005).
309 Comparing and evaluating process-based ecosystem model predictions of carbon and water
310 fluxes in major European forest biomes. *Global Change Biology* , 11, 2211-2233.

311 Nabuurs, G., Schelhaas, M., Mohren, G., & Field, C. (2003). Temporal evolution of the
312 European forest sector carbon sink from 1950 to 1999. *Global Change Biology* , 9, 152-160.

313 Naudts, K., Ryder, J., McGrath, M., Otto, J., Chen, Y., Valade, A., et al. (2014). A vertically
314 discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the
315 energy, water and carbon fluxes. *Geosci. Model Dev. Discuss.* , 7, 8565-8647.

316 Nolè, A., Collalti, A., Magnani, F., Duce, P., Ferrara, A., Mancino, G., et al. (2013).
317 Assessing temporal variation of primary and ecosystem production in two Mediterranean
318 forests using a modified 3-PG model. *Annals of Forest Science* , DOI 10.1007/s13595-013-
319 0315-7.

320 Ogren, E. (2000). Maintenance respiration correlates with sugar but not nitrogen
321 concentration in dormant plants. *Physiologia Plantarum* , 108:295-299.

322 Oltchev, A., Cermak, J., Nadezhdina, N., Tatarinov, F., Tishenko, A., Ibrom, A., et al. (2002).
323 Transpiration of a mixed forest stand: field measurements and simulation using SVAT
324 models. *Boreal Environment Research* , 7, 389-397.

325 Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., et al. (2006).
326 Towards a standardized processing of Net Ecosystem Exchange measured with eddy
327 covariance technique: algorithms and uncertainty estimation. *Biogeosciences* , 3, 571-583.

328 Peters, W., Jacobson, A., Sweeney, C., & al., e. (2007). An atmospheric perspective on North
329 American carbon dioxide exchange: carbontracker. *Proceedings of the National Academy of*
330 *Sciences of the United States of America* , 104, 18925-18930.

331 Piao, S., Friedlingstein, P., Ciais, P., Zhou, L., & Chen, A. (2006). Effects of climate and CO2
332 changes on the greening of the Northern Hemisphere over the past two decades. *Geophys.*
333 *Res. Lett.* , 33, L23402, doi:10.1029/2006GLO28205.

334 Piao, S., Sitch, S., Ciais, P., Friendlingstein, P., Peylin, P., Wang, X., et al. (2013). Evaluation
335 of terrestrial carbon cycle models for their response to climate variability and to CO2 trends.
336 *Global Change Biology* , 19, 2117-2132.

337 Pietsch, S., Hasenauer, H., & Thornton, P. (2005). BGC-model parameters for tree species
338 growing in central European forests. *Forest Ecology and Management* , 211, 264-295.

339 Pilegaard, K., Ibrom, A., Courtney, M., Hummelshoj, P., & Jensen, N. (2011). Increasing net
340 CO2 uptake by a Danish beech forest during the period from 1996 to 2009. *Agrocoltural and*
341 *Forest Meteorology* , 151: 934-946.

342 Poulter, B., Heyder, U., & Cramer, W. (2009). Modelling the Sensitivity of the Seasonal
343 Cycle of GPP to Dynamic LAI and Sil Depths in Tropical Rainforest. *Ecosystems* , 12, 517-
344 533.

345 Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., et al. (2015). Plant
346 functional type classification for Eart System Models: results from the European Space
347 Agency's Land Cover Climate Change Initiative. *Geosci. Model Dev. Discuss.* , 8, 429-462.

348 Prentice, I., Liang, B., B.E., M., & Wang, Y.-P. (2014). Reliable, robust and realistic: the
349 three R's of next-generation land surface modelling. *Atmos. Chem. Phys. Discuss.* , 14, 24811-
350 24861.

351 Propastin, P., Ibrom, A., Knohl, A., & Erasmi, S. (2012). Effects of canopy photosynthesis
352 saturation on the estimation of gross primary productivity from modis datain a tropical forest.
353 *Remote Sens. Environ.* , 121, 252-260.

354 Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Valentini, R., Aubinet, M., et al. (2005).
355 On the separation of net ecosystem exchange into assimilation and ecosystem respiration:
356 review and improved algorithm. *Global Change Biology* , 11, 1424-1439.

357 Reichstein, M., Tenhunen, J., Roupsard, O., Ourcival, J.-M., Rambal, S., Dore, S., et al.
358 (2002). Ecosystem respiration in two Mediterranean evergreen Holm Oak forests: drought
359 effects and decomposition dynamics. *Functional Ecology* , 16, 27-39.

360 Richardson, A., Anderson, R., Altafarain, M., Barr, A., Bohrer, G., Chen, G., et al. (2012a).
361 Terrestrial biosphere models need better representation of vegetation phenology: results from
362 the North American Carbon Program Site Synthesis. *Global Change Biology* , 18, 566-584.

363 Richardson, A., Aubinet, M., Barr, A., Hollinger, D., Ibrom, A., Lasslop, G., et al. (2012b).
364 Uncertainty quantification. In M. Aubinet, T. Vesala, & D. Papale, *Eddy Covariance.*
365 Heidelberg London New York: Springer Dordbrecht.

366 Richardson, A., Black, T., Ciais, P., Delbart, N., Friedl, M., Gobron, N., et al. (2010).
367 Influence of spring and autumn phenological transitions on forest ecosystem productivity.
368 *Phyl. Trans. R. Soc. B.* , 365, 3227-3246.

369 Richardson, A., Hollinger, D., Aber, J., Ollinger, S., & Braswell, B. (2007). Environmental
370 variation is directly responsible for short- but not long-term variation in forest-atmosphere
371 carbon exchange. *Global Change Biology* , 13, 788-803.

372 Rotzer, T., Grote, R., & H., P. (2004). The timinig of bud burst and its effect on tree growth.
373 *Internation Journal of Biometeorology* , 48, 109-118.

374 Ruimy, A., Jarvis, P., Baldocchi, D., & Sugier, B. (1995). CO2 fluxes over plant canopies and
375 solar radiation: A review. *Adv. Ecol. Res.* , 26, 1-51.

376 Running, S., & Coughlan, J. (1988). A general model of forest ecosystem processes for
377 regional applications I. Hydrologic balance, Canopy Gas Exchange and Primary Production
378 Processes. *Ecological Modelling* , 42:125-154.

379 Ryan, M. (1991). Effects of climate change on plant respiration. *Ecological Applications* ,
380 1:157-167.

381 Sabatè, S., Gracia, C., & Sànchez, A. (2002). Likely effects of climate chamge on growth of
382 Quercus ilex, Pinus halepensis, Pinus pinaster, Pinus sylvestris and Fagus sylvatica forests in
383 the Mediterranean region. *Forest Ecology and Management* , 162, 23.37.

384 Santini, M., Collalti, A., & Valentini, R. (2014). Climate change impacts on vegetation and
385 water cycle in the Euro-Mediterranean region, studied by a likelihood approach. *Reg.*
386 *Environ. Change* , DOI 10.1007/s10113-013-0582-8.

387 Scartazza, A., Moscatello, S., Matteucci, G., Battistelli, A., & Brugnoli, E. (2013). Seasonal
388 and inter-annual dynamics of growth, non-structural carbohydrates and C stable isotopes in
389 Mediterranean beech forest. *Tree Physiology* , 730-742.

390 Seidl, R., Rammer, W., Scheller, R., & Spies, T. (2012). An individual-based process model
391 to simulate landscape-scale forest ecosystem dynamics. *Ecological Modelling* , 231, 87-100.

392 Sellers, P., Dickinson, R., Randall, D., & al., e. (1997). Modelling the exchanges of energy,
393 water, and carbon between continents abd the atmosphere. *Science* , 275, 502-509.

394 Shinozaki, K., Yoda, K., Hozumi, K., & Kira, T. (1964a). A quantitative analysis of plant
395 form-the pipe model theory. I Basic analyses. *Japanese Journal of Ecology* , 4, 97-105.

396 Shinozaki, K., Yoda, K., Hozumi, K., & Kira, T. (1964b). A quantitative analysis of plant
397 form-the pipe model theory. II. Further evidence of the theory and its application in forest
398 ecology. *Japanese Journal of Ecology* , 14, 133-139.

399 Slevin, D., Tett, S., & Williams, M. (2015). Multi-site evaluation of the JULES land surface
400 model using global and local data. *Geosci. Model Dev.* , 8, 295-316.

401 Suni, T., Beninger, F., Vesala, T., Markkanen, T., Hari, P., Makela, A., et al. (2003). Air
402 temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. *Global*
403 *Change Biology* , 9, 1410-1426.

404 Suni, T., Berninger, F., Vesala, T., Markkanen, T., Hari, P., Makela, A., et al. (2003). Air
405 temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. *Global*
406 *Change Biology* , 9, 1410-1426.

407 Sykes, M., Prentice, I., Smith, B., Cramer, W., & Venevsky, S. (2001). An introduction to the
408 European terrestrial ecosystem modelling activity. *Global Ecol. Biogeogr.* , 10, 581-594.

409 Taylor, K. (2001). Summarizing multiple aspects of model performance in a single diagram.
410 *Journal of Geophysical Research* , 106, 7183-7192.

411 Taylor, K., Stouffer, R., & Meehl, G. (2012). An overview of CMIP5 and the experiment
412 design. *Bull. Am. Meteorol. Soc.* , 90, 485-498.

413   Thornton, P. E. (2010). *Biome BGC version 4.2: Theoretical Framework of Biome-BGC*.

414   Thornton, P., & Zimmermann, N. (2007). An Improved Canopy Integration Scheme for a
415   Land Surface Model with Prognostic Canopy Structure. *Journal of Climate* , 20, 3902-3923.

416   Turner, D. U., Bremer, D., Wofsy, S., Meyers, T., Gower, S., & Gregory, M. (2003). A cross-
417   biome comparison of daily light use efficiency for gross primary production. *Global Change*
418   *Biol.* , 9, 383-395.

419   UNECE, & FAO. (2011). *State of Europe's Forest.* Status and Trends in Suistainable Forest
420   Management in Europe.

421   Verbeeck, H., Steppe, K., Nadezhdina, N., Op De Beeck, M., Meiresonne, L., Lemeur, R., et
422   al. (2007). Model analysis of the effects of atmospheric drivers on storage water use in Scots
423   pine. *Biogeosciences* , 4, 657-671.

424   Vetter, M., Churkina, G., Jung, M., Reichstein, M., Zaehle, S., Bondeau, A., et al. (2008).
425   Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using
426   seven models. *Biogeosciences* , 5, 561-583.

427   Vitale, M., Scimone, M., Feoli, E., & Manes, F. (2003). Modelling leaf gas exchanges to
428   predict functional trends in Mediterranean Quercus ilex forest under climatic changes in
429   temperature. *Ecological Modelling* , 166, 123-134.

430   Wang, F., Chen, J., Gonsamo, A., Zhou, B., Cao, F., & Yi, Q. (2014a). A two-leaf rectangular
431   hyperbolic model for estimating GPP across vegetation types and climate conditions. *J.*
432   *Geophys. Res. Biogeosci.* , 119, doi: 10.1002/2013JG002596.

433   Wang, H., Prentice, I., & Davis, T. (2014b). Biophysical constraints on gross primary
434   production by the terrestrail biosphere. *Biogeosciences* , 11, 5987-6001.

435   Waring, R., & Landsberg, J. (1998). Net primary production of forests: a constant fraction of
436   gross primary production? *Tree Physiology* , 18, 129-134.

437   Waring, R., & McDowel, N. (2002). Use of a physiological process model with forestry yield
438   tables to set limits on annual carbon balances. *Tree Physioloogy* , 22, 179-188.

439   Warren, J., Potzelsberger, E., Wullschleger, S., Thornton, P., Hasenauer, H., & Norby, R.
440   (2011). Ecohydrologic impact of reduced stomatal conductance in forests exposed to elevated
441   CO2. *Ecohydrology* , 4, 196-210.

442 White, M., Thornton, P., & Running, S. (2000). Parameterization and sensitivity analysis of
443 the BIOME-BGC terrestrial ecosystem model: net primary production controls. *Earth*
444 *interactions* , 4 (3), 1.

445 Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error error (MAE)
446 over the root mean square error (RMSE) in assessing average model performance. *CLIMATE*
447 *RESEARCH* , 30, 79-82.

448 Wißkirchen, K., Tum, M., Günther, K., Niklaus, B., Eisfelder, C., & Knorr, W. (2013).
449 Quantifying the carbon uptake by vegetation for Europe on a 1km2 resolution using a remote
450 sensing driven vegetation model. *Geoscientific Model Development* , 6, 1623-1640.

451 Wramneby, A., Smith, B., Zaehle, S., & Sykes, M. (2008). Parameter uncertainties in the
452 modelling of vegetation dynamics-Effects on tree community structure and ecosystem
453 functioning in European forest biomes. *Ecological Modelling* , 216, 277-290.

454 Wu, J., Jansson, P., van der Linden, L., Pilegaard, K., Beier, C., & Ibrom, A. (2013).
455 Modelling the decadal trend of ecosystem carbon fluxes demonstrates the important role of
456 functional changes in a temperate deciduous forest. *Ecological Modelling* , 260, 50-61.

457 Wu, X., Ju, W., Zhou, Y., He, M., Law, B., Black, T., et al. (2015). Performance of Linear
458 and Nonlinear Two-leaf Light Use Efficiency Models at Different Temporal Scales. *Remote*
459 *sensing* , xxxx.

460 Yin, Z., Dekker, S., van den Hurk, B., & Dijkstra, H. (2014). Effects of vegetation structure
461 on biomass accumulation in a Balanced Optimally Structure Vegetation Model (BOSVM
462 v1.0). *Geosci. Model Dev.* , 7, 821-845.

463 Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L., Baldocchi, D., et al. (2007). Deriving a
464 light use efficiency model from eddy covariance flux data for predicting daily gross primary
465 production across biomes. *Agricoltural and Forest Meteorology* , 143, 189-207.

466 Zeng, D., Hunt, E., & Running, S. (1993). A Daily Soil Temperature Model Based on Air
467 Temperature and Precipitation for Continental Apllications. *Climate Research* , 2, 183-191.

468 Zhang, T., Lichstein, J., & R.A., B. (2014). Spatial and temporal hetereogeneity in the
469 dynamics of eastern U.S. forests: Implications for developing braod-scale forest dynamics
470 models. *Ecological Modelling* , 279, 89-99.

471   Zhao, Y., Ciais, P., Peylin, P., Viovy, N., Longdoz, B., Bonnefond, J., et al. (2012). How

472   errors on meteorological variables impact simulated ecosystem fluxes: a case study for six

473   French sites. *Biogeosciences* , 9, 2537-2664.

474   Zscheischler, J., Michalak, A., Schwalm, C., Mahecha, M., Huntzinger, D., Reichstein, M., et

475   al. (2014). Impact of large-scale climate extremes on biospheric carbon fluxes: An

476   Intercomparison based on MsTMIP data. *Global Biogeochemical Cycles* , 28, 585-600.

477

478