

This item is the archived peer-reviewed author-version of:

Matching the phenology of Net Ecosystem Exchange and vegetation indices estimated with MODIS and FLUXNET in-situ observations

Reference:

Balzarolo Manuela, Vicca Sara, Nguy-Robertson A. L., Bonal D., Elbers J. A., Fu Yongshuo, Gruenwald T., Horemans Joanna, Papale D., Penuelas J.,- Matching the phenology of Net Ecosystem Exchange and vegetation indices estimated with MODIS and FLUXNET in-situ observations

Remote sensing of environment - ISSN 0034-4257 - 174(2016), p. 290-300

Full text (Publisher's DOI): <http://dx.doi.org/doi:10.1016/J.RSE.2015.12.017>

To cite this reference: <http://hdl.handle.net/10067/1303590151162165141>

1 **Matching the phenology of Net Ecosystem Exchange and**
2 **Vegetation Indices estimated with MODIS and FLUXNET *in-***
3 ***situ* observations**

4

5 **M. Balzarolo¹, S. Vicca¹, A. L. Guy-Robertson², D. Bonal³, J. A. Elbers⁴, Y. H. Fu^{1,5},**
6 **T. Grünwald⁶, J. A. Horemans¹, D. Papale⁷, J. Peñuelas^{8,9}, A. Suyker², F.**
7 **Veroustraete¹⁰**

8

9

10 [1] Center of Excellence PLECO (Plant and Vegetation Ecology), Biology Department,
11 University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium

12 [2] School of Natural Resources, University of Nebraska-Lincoln, Nebraska, U.S.A.

13 [3] INRA, UMR EEF, 54280 Champenoux, France

14 [4] Alterra, Wageningen UR, Droevedaalsesteeg 3, 6708PB, Wageningen, The Netherlands

15 [5] College of Urban and Environmental Sciences, Peking University, Yiheyuan Road 5, 100871
16 Beijing, China

17 [6] Technische Universität (TU) Dresden, Institute of Hydrology and Meteorology, Chair of
18 Meteorology, D-01062 Dresden, Germany

19 [7] Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University
20 of Tuscia, Via S.C. de Lellis snc, 01100 Viterbo, Italy

21 [8] CSIC, Global Ecology Unit CREAF-CSIC-UAB, 08913 Cerdanyola del Vallès, Catalonia,
22 Spain.

23 [9] CREAF, 08913 Cerdanyola del Vallès, Catalonia, Spain.

24 [10] University of Antwerp, Faculty of Sciences, Department of Bioscience Engineering,
25 Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

26

27

28

29

30 Correspondence to: M. Balzarolo (manuela.balzarolo@uantwerpen.be)

31

32

33

34

35 **Keywords:** Vegetation Indices (VI's), Net Ecosystem Exchange (NEE), Start of Net Carbon
36 Uptake (SGS_{NEE}), Start of Growing Season (SGS), Plant Functional Types (PFT's), Phenology.

37

38 **Abstract**

39 Shifts in ecosystem phenology play an important role in the definition of inter-annual variability
40 of net ecosystem carbon uptake. A good estimate at the global scale of ecosystem phenology,

41 mainly that of photosynthesis or gross primary productivity (GPP), may be provided by
42 vegetation indices derived from MODIS satellite image data.

43 However, the relationship between the start date of a growing (or greening) season (SGS) when
44 derived from different vegetation indices (VI's), and the starting day of carbon uptake is not well
45 elucidated. Additionally, the validation of existing phenology data with *in-situ* measurements is
46 largely missing. We have investigated the possibility to use different VI's to predict the starting
47 day of the growing season for 28 FLUXNET sites as well as MODIS data. This analysis included
48 main plant functional types (PFT's).

49 Of all VI's taken into account in this paper, the NDVI (Normalised Difference Vegetation Index)
50 shows the highest correlation coefficient for the relationship between the starting day of the
51 growing season as observed with MODIS and *in-situ* observations. However, MODIS
52 observations elicit a 20-21 days earlier SGS date compared to *in-situ* observations. The
53 prediction for the NEE start of the growing season diverges when using different VI's, and seems
54 to depend on the amplitude for carbon and VI and on PFT. The optimal VI for estimation of a
55 SGS date was PFT-specific - for example the WRDVI for cropland, but the MODIS NDVI
56 performed best when applied as an estimator for Net Ecosystem Exchange and when considering
57 all PFT's pooled.

58

59 **1 Introduction**

60 Ecosystem phenology shifts play an important role in describing the inter-annual variability of
61 NEE (Net Ecosystem Exchange) due to its impact on Gross Primary Productivity (GPP). A shift
62 in the start date of a growing season modulates annual GPP (Churkina et al. 2005; Keenan et al.

63 2014; Richardson et al. 2010). Multiple data sources - primarily carbon dioxide (CO_2) eddy
64 covariance flux data (NEE) as well as satellite imagery estimated vegetation indices (VI's) -
65 originating from different databases are used to estimate the start day of a growing season
66 (Garrity et al. 2011).

67 GPP and NEE seasonality is frequently defined as carbon-flux phenology. Both variables
68 describe the seasonality of ecosystem gross photosynthesis. Photosynthetic phenology is
69 represented by the starting day of GPP and NEE and more specifically when NEE becomes
70 positive. Explicitly the date when this occurs is by definition the day (SGS_{NEE}) when an
71 ecosystem transforms from a carbon source into a carbon sink. SGS_{NEE} can be estimated in
72 different ways. Eddy covariance data is on track to make the estimate (Baldocchi et al. 2005). On
73 the other hand, leaf phenology can also be observed and defined with remote sensing based
74 methods (Garrity et al., 2011). The exercise is to estimate the starting day of greening ($\text{SGS}_{\text{MODIS}}$
75 and $\text{SGS}_{\text{in-situ}}$) using an optical sensor (MODIS or *in-situ*). Intuitively, this is expected to
76 correspond to SGS_{NEE} , but this relationship, and hence the predictability of SGS_{NEE} from optical
77 sensors, has yet to be verified. It is assumed in this paper that a correspondence with SGS_{NEE}
78 exists. It is the objective of this paper to verify, even validate this correspondence and hence
79 whether SGS_{NEE} can be estimated from a space remote sensing platform (TERRA MODIS).

80 Several studies highlight a new application of remote sensing e.i., the integration of remote
81 sensing data as well as NEE and GPP data collected with the eddy covariance method, to predict
82 and map terrestrial carbon assimilation at the global and regional scales (Heinsch et al. 2006;
83 Verma et al. 2014). An important step in this research venture is to establish a correspondence
84 between phenological data - observed with remote sensing - versus *in-situ* optical and eddy
85 covariance flux data.

86 Remote sensing facilitates the global observation of the starting day of a growing season defined
87 as the starting day of gross photosynthesis. Several approaches are applied to monitor changes in
88 canopy development. These include changes in greening, acquired by digital camera imagery
89 (Betancourt et al. 2005; Richardson et al. 2009), spectral spaces, reflectance and reflectance
90 relationships (Nguy-Robertson et al. 2012) and vegetation indices (Wu 2014; Zhang et al. 2003).
91 The latter is a common approach and has been applied using proximal sensors, such as
92 radiometers (Huemmrich et al. 1999) or modified cameras (Petach et al. 2014; Sakamoto et al.
93 2010), and satellite sensor imagery (Walker et al. 2014).

94 Several VI's are considered as a useful estimator of bio-geophysical and biochemical parameters
95 regulating leaf and canopy phenology and hence, productivity. Typical bio-geophysical variables
96 derived from remote sensing platforms are leaf area index (LAI) and chlorophyll a and b
97 (Gitelson et al. 2006; Myneni et al. 2002). A great variety of VI's have been defined by remote
98 sensing scientists and all differ in their definition and in their sensitivity to changes in
99 photosynthesis as well. These so-called "Greenness indices" - such as the widely used
100 Normalized Difference Vegetation Index (NDVI) (Tucker 1979) - demonstrate to be a good
101 proxy for the fraction of absorbed PAR ($f\text{APAR}$) and PAR is Photosynthetically Active
102 Radiation and APAR is absorbed PAR. By definition, $f\text{APAR} = \text{APAR}/\text{PAR}$. Hence $f\text{APAR}$ and
103 the NDVI are related with green biomass and canopy structure. Furthermore, the NDVI has been
104 recognised to be a good proxy for the investigation of the impact of climate change on leaf and
105 ecosystem phenology (Peng et al. 2013; Piao et al. 2015).

106 In addition to the NDVI, many other vegetation indices have also been defined. Among many
107 others one can cite: the Enhanced Vegetation Index (EVI) (Huete et al. 1997). Both the NDVI
108 and EVI allow the observation of canopy greening based on their dependency on the RED and

109 near infrared (NIR) parts of the electromagnetic spectrum (Huete et al. 2002; Piao et al. 2006;
110 Reed et al. 1994). The EVI is generally less sensitive to soil background variations compared to
111 other VI's when vegetation cover fraction (f_{Cover}) is low (Huete et al. 2002). The EVI
112 incorporates an additional blue spectral band in addition to the commonly used RED and NIR
113 spectral bands. The use of a blue band is intended to reduce atmospheric scattering effects
114 typically due to the interaction of - most strongly, blue - light with aerosols and atmospheric
115 molecules. The EVI definition reduces noise, but its applicability is limited to those sensors
116 which dispose of a blue band, which puts a limit on the number of satellite sensors which can be
117 used for global studies.

118 Jiang et al. (2008) proposed an alternative definition for the EVI, e.g., the EVI2 in which the blue
119 spectral band is substituted by a red band. Though EVI2 does not make use of a blue band, EVI2
120 has been determined to be equivalent to EVI and seems helpful to observe canopy properties. A
121 benefit of EVI and EVI2 is that they remain more sensitive than the NDVI when canopies
122 become denser. However, even these vegetation indices do saturate at moderate LAI values
123 (Viña et al. 2011). Alternatively, the Wide Dynamic Range Vegetation Index (WDRVI) seems
124 more sensitive for the entire dynamic range of the LAI (Gitelson 2004). The Simple Ratio (SR)
125 however has been shown to be the most sensitive VI at high LAI values (Viña et al. 2011).

126 The Global Environmental Monitoring Index (GEMI) has been defined based on RED and NIR
127 band reflectances. GEMI minimizes atmospheric effects, similar to the EVI and minimizes
128 observational angular effects as well (e.g. BRDF effects) in the observed VI signal (Pinty and
129 Verstraete, 1992). Nevertheless GEMI is rarely used in canopy phenology observations.

130 The Soil Adjusted Vegetation index (SAVI) has been defined to minimize the influence of soil
131 brightness (Huete 1988). The SAVI involves the RED and NIR reflectance bands and a soil

132 brightness correction factor (L). L equals zero for a very high vegetation cover and unity for non-
133 vegetated land surfaces. Typically, L is assumed to be 0.5 for most vegetated areas. By definition
134 SAVI equals the NDVI when L equals zero.

135 A variety of *in-situ* optical sensors are commercially available for field, UAV and airborne
136 applications. They acquire NIR and RED band reflectances at top-of-the-canopy level (Balzarolo
137 et al. 2011). PAR sensors can be applied as broadband sensors for reflectances in the visible
138 spectral range. These data can then be used instead of RED band imagery, to calculate vegetation
139 indices. Likewise, pyranometers are sensitive in the global shortwave radiation band (GLR) and
140 they can be applied as a NIR sensitive reflectance band. GLR spans a broad spectral range,
141 including the visible, NIR, and mid-infrared spectral regions. The visible spectral region in the
142 GLR band can be brought to zero reflectance using the PAR sensor signal (Jenkins et al. 2007;
143 Wang et al. 2004). With this approach *in-situ* NDVI can be derived from measurements of the
144 PAR band (400-700 nm); and a visible corrected GLR band (700-2800 nm).

145 *In-situ* NDVI measurements provide distinct advantages. They are typically endowed with a high
146 temporal resolution since they acquire data at an hourly basis and can be programmed for data
147 collection at even higher frequencies. Important to mention is that *in-situ* NDVI measurements
148 offer the possibility for data acquisition under overcast conditions. Only low altitude remote
149 sensing systems like UAV's offer this capacity as well.

150 Finally, the objective of this paper is to explore the potential of six different VI's calculated from
151 *in-situ* radiation measurements, and obtained from MODIS RED and NIR reflectances. This
152 enables the estimation of the start of the carbon uptake season (i.e. SGS_{NEE}). Additionally the
153 approach should also enable the phenological monitoring at twenty-eight different FLUXNET
154 sites encompassing eight different PFT's (or ecosystems).

155 The specific objectives pursued in this paper are:

156 (i) How well do SGS estimations derived from *in-situ* vegetation indices (referred to as
157 SGS_{*in-situ*}) correlate with SGS estimations derived from MODIS VI's (referred to as
158 SGS_{MODIS}) and secondly;

159 (ii) Which VI's as well as sensors are optimal for SGS_{NEE} detection based on *in-situ* NEE
160 flux data collected at FLUXNET sites.

161

162 **2 Materials and methods**

163 **2.1 FLUXNET data: site selection**

164 The study presented in this paper is based on VI's, determined with remote sensing and carbon
165 flux measurements acquired from the FLUXNET eddy covariance network (www.fluxdata.org,
166 “La Thuile” database, October 2010). The FLUXNET database contains half-hourly observations
167 of ecosystem CO₂, heat fluxes and meteorological data of more than 250 sites worldwide and for
168 a total of 960 site-years. The most representative sites used in this study have been selected
169 based on the following boundary conditions:

- 170 (i) The availability of continuous measurements of global incoming and outgoing
171 shortwave radiation (GLR_{in} and GLR_{out}) respectively, since both are required to calculate
172 *in-situ* VI's;
- 173 (ii) The availability of continuous measurements of global incoming and outgoing
174 PAR (PAR_{in}, PAR_{out}), since both are required to calculate *in-situ* VI's;
- 175 (iii) The availability of measured carbon mass fluxes (in particular NEE).

176 The application of these boundary conditions, leads to a subset of 28 FLUXNET sites (Table 1),
177 representing 72 site-years. They have a minimum of two years of both high quality flux
178 measurements and measured radiation data. The 28 sites have been selected to establish the basic
179 dataset used for the different procedures and analysis of which the results and conclusions are
180 reported in this paper.

181 The selected sites cover main global PFT's among which: CRO—Cropland; DBF—Deciduous
182 Broadleaf Forest; EBF—Evergreen Broadleaf Forest; ENF—Evergreen Needle-leaf Forest;
183 GRA—Grassland; OSH—Open Shrubland; WSA—Woody Savanna. The PFT's are defined as in
184 the International Geosphere–Biosphere Programme – IGBP (Loveland and Belward 1997).

185

186 **[Table 1]**

187

188

189

190

191

192 Table 1 – Description of FLUXNET sites, years of measurement and PFTs used in this study
193 (CRO—Cropland; DBF—Deciduous Broadleaf Forest; EBF—Evergreen Broadleaf Forest;
194 ENF—Evergreen Needle-leaf Forest; GRA—Grassland; OSH—Open Shrubland; WSA—
195 Woody Savanna).

Site ID	Site name	Country	Lat [decimal degrees]	Lon [decimal degrees]	Plant Functional Type	Measurement Interval (Years)
BR-Cax	Caxiuna Forest-Almeirim	Brazil	-1.719720	-51.459000	EBF	1999-2002
BR-Sa2	Santarem-Km77-Pasture	Brazil	-3.011900	-54.536499	CRO	2001-2002
BR-Sa3	Santarem-Km83-Logged Forest	Brazil	-3.018030	-54.971401	EBF	2002-2003

BR-Sp1	Sao Paulo Cerrado	Brazil	-21.619499	-47.649899	WSA	2001-2002
CA-NS5	UCI-1981 burn site	Canada	55.863098	-98.485001	ENF	2004-2005
CA-NS6	UCI-1989 burn site	Canada	55.916698	-98.964401	OSH	2002-2005
CA-NS7	UCI-1998 burn site	Canada	56.635799	-99.948303	OSH	2003-2005
DE-Geb	Gebesee	Germany	51.100101	10.914300	CRO	2004-2006
DE-Hai	Hainich	Germany	51.079300	10.452000	DBF	2004-2006
DE-Kli	Klingenberg – cropland	Germany	50.892899	13.522500	CRO	2004-2006
DE-Meh	Mehrstedt 1	Germany	51.275299	10.655500	GRA	2004-2006
DE-Tha	Anchor Station Tharandt - old spruce	Germany	50.963600	13.566900	ENF	2005-2006
DE-Wet	Wetzstein	Germany	50.453499	11.457500	ENF	2004-2006
FI-Hyy	Hyytiala	Finland	61.847401	24.294800	ENF	2004-2006
GF-Guy	Guyaflux	French Guiana	5.277700	-52.928799	EBF	2004-2006
JP-Tak	Takayama	Japan	36.146198	137.423004	DBF	1999-2004
JP-Tom	Tomakomai National Forest	Japan	42.739498	141.514893	DBF	2001-2003
NL-Loo	Loobos	Netherlands	52.167900	5.743960	ENF	2005-2006
US-ARM	ARM Southern Great Plains site- Lamont	USA	36.605801	-97.488800	CRO	2005-2006
US-Bar	Bartlett Experimental Forest	USA	44.064602	-71.288078	DBF	2004-2005
US-Bo1	Bondville	USA	40.006199	-88.290398	CRO	2003-2006
US-CaV	Canaan Valley	USA	39.063301	-79.420799	GRA	2004-2005
US-FPe	Fort Peck	USA	48.307701	-105.101898	GRA	2004-2006
US-Goo	Goodwin Creek	USA	34.254700	-89.873497	GRA	2002-2006
US-MOz	Missouri Ozark Site	USA	38.744099	-92.199997	DBF	2005-2006
US-Ne2	Mead - irrigated maize-soybean rotation site	USA	41.164902	-96.470100	CRO	2001-2004
US-Ne3	Mead - rainfed maize-soybean rotation site	USA	41.179699	-96.439697	CRO	2002-2004
US-SRM	Santa Rita Mesquite	USA	31.821400	-110.865997	WSA	2004-2006

196

197 Furthermore, nine additional sites have been selected from the FLUXNET “La Thuile” database,
 198 representing 20 site-years, (cited in Table S1 in the Supplementary data; validation sites). They
 199 were used as an independent evaluation (validation) of NEE phenology compared with the more
 200 common descriptors of phenology (i.e. fAPAR and EVI). These validation sites have been
 201 selected because they have acquisitions of all radiation components required to derive fAPAR:
 202 incident PAR at the top of the canopy (i.e. PAR_{in}) and below canopy PAR (PAR_{bc}).

203 More details on NEE phenology evolution are given in section 2.4.1. Two of the FLUXNET
204 validation sites (i.e. DE-Tha and FI-Hyy) have also been used as well, in the main analysis.

205 **2.1.1 *In-situ* radiation measurements**

206 The most commonly used instrument for the measurement of PAR_{in} and PAR_{out} at the flux tower
207 sites, is the quantum sensor. In a typical set-up at a FLUXNET site, an upward facing quantum
208 sensor is used to measure PAR_{in} while concomitantly a downward facing sensor measures
209 outgoing PAR_{out} . Measurements of respectively GLR_{in} and GLR_{out} , in the optical spectral range
210 (305 to 2800 nm) have been performed with two pyranometers, of which one faces upward to
211 measure GLR_{in} , and the other faces downward to measure GLR_{out} . More details on radiation
212 sensor set-ups at the FLUXNET sites are given by Balzalrolo et al. (2011).

213 **2.1.2 Carbon flux measurements**

214 Eddy covariance (EC) measurements of ecosystem CO_2 mass fluxes have been acquired from the
215 FLUXNET database (Baldocchi et al. 2001). EC data are collected by the site manager according
216 to a standard procedure and provide to the FLUXNET database. Typically data are collected at
217 high sampling frequencies (at least at 10 Hz) and subsequently converted into mass fluxes
218 integrated over a thirty minute time interval. Here, we used gap-filled NEE data from FLUXENT
219 “La Thuile” database (www.fluxdata.org, October 2010) where half hourly data are processed
220 following the standardized methodology described in Papale et al. (2006) and Reichstein et al.
221 (2005). In particular, the NEE data are storage corrected, spike filtered, u^* -filtered, and
222 subsequently gap-filled. The datasets thus obtained typically correspond with a source area
223 footprint of hundreds of meters in the vicinity of the EC tower, depending on tower and
224 vegetation height (Schmid 2002).

225 2.2 Computation of *in-situ* VI's from *in-situ* radiation measurements

226 *In-situ* VI's are calculated from half-hourly *in-situ* acquisitions of PAR_{in} , PAR_{out} , GLR_{in} and
 227 GLR_{out} according to the method proposed by Huemmrich et al. (1999). PAR reflectance (ρ_{PAR} ,
 228 400-700 nm) is derived from PAR_{in} and PAR_{out} measurements. NIR Reflectance (700-2800 nm)
 229 is derived from GLR_{in} and PAR_{in} and GLR_{out} and PAR_{out} measurements. Summarizing, ρ_{PAR} and
 230 ρ_{NIR} are calculated according to Eq. 1 and 2:

$$231 \quad \rho_{PAR} = \frac{PAR_{out}}{PAR_{in}} \quad (1)$$

$$\rho_{NIR} = \frac{GLR_{out} - PAR_{out}}{GLR_{in} - PAR_{in}} \quad (2)$$

The physical units of both incoming and outgoing PAR are obtained by a physical unit conversion $\mu\text{mol.photons.m}^{-2.\text{s}^{-1}}$ to $\text{J.m}^{-2.\text{s}^{-1}}$ using a conversion factor of $4.55 \mu\text{mol.J}^{-1}$ as proposed by Goudriaan and Van Laar (1994).

236 *In-situ* data are calculated as an average of five observations per hour before and after solar noon
237 (i.e. between 11h00 and 13h00 local solar time (LST)) for each of the 28 main sites. *In-situ* VI's
238 (Table 2) are derived using ρ_{PAR} and ρ_{NIR} reflectances and calculated according to eq.1 and eq. 2.
239 The acquisition dates of MODIS 8-day composite NDVI data are used to obtain representative
240 *in-situ* VI data.

241

242 [Table 2]

243

244

245 Table 2 - Definitions of MODIS and *in-situ* VIs. See running text for a definition of ρ_{NIR} and ρ_{PAR} .

246

VI	Definition	Literature reference
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\rho_{NIR} - \rho_{PAR}}{\rho_{NIR} + \rho_{PAR}}$	Rouse et al., 1974
Simple Ratio (SR)	$SR = \frac{\rho_{NIR}}{\rho_{PAR}}$	Rouse et al., 1974
Wide Range Dynamic Vegetation Index (WRDVI)	$WRDVI = \frac{a\rho_{NIR} - \rho_{PAR}}{a\rho_{NIR} + \rho_{PAR}}$ $a = 0.1$	Gitelson et al., 2004
Enhanced Vegetation Index 2 (EVI2)	$EVI2 = \frac{2.5 * (\rho_{NIR} - \rho_{PAR})}{\rho_{NIR} + 2.4 * \rho_{PAR} + 1}$	Jiang et al., 2008
Global Environmental Monitoring Index (GEMI)	$GEMI = \eta(1 - 0.25\eta) - \frac{\rho_{PAR} - 0.125}{1 - \rho_{PAR}}$ $\eta = \frac{2(\rho_{NIR}^2 - \rho_{PAR}^2) + 1.5\rho_{NIR} + 0.5\rho_{PAR}}{\rho_{NIR} + \rho_{PAR} + 0.5}$	Pinty and Verstraete, 1992
Soil-Adjusted vegetation index (SAVI)	$SAVI = \frac{\rho_{NIR} - \rho_{PAR}}{\rho_{NIR} + \rho_{PAR} + L} * (1 + L)$ $L = 0.5$	Huete, 1988

L is the soil brightness correction factor

247

248

249

250 The quality assessment and control (QA/QC) of half-hourly radiometric measurements is
 251 performed applying various physical limit tests. Typically, GLR_{in} has to be less than the
 252 corresponding extraterrestrial radiation (R_{ext}) at the same point in time (hence $GLR_{in} < R_{ext}$). An
 253 analysis of the statistical variability of the data (quantified by the standard deviation, σ) has been
 254 performed as well in the QA/QC procedure. As a matter of fact GLR_{in} can only be larger than

255 R_{ext} for high latitude regions (hence, above the 65° Northern or 65° Southern latitudes) and only
256 under the condition that convective clouds are present (Yang et al. 2010). Physically, at the
257 Earth's surface, GLR_{in} when interacting with clouds and aerosols always drops to values lower
258 than R_{ext} for all optical wavelengths. Henceforth, the Atmospheric Impact Ratio (AIR) or $AIR =$
259 GLR_{in} / R_{ext} must always be smaller or equal to unity, the last case only under exo-atmospheric
260 conditions.

261 The variation of AIR (σ_{AIR}) between two successive 30 minutes measurement intervals cannot
262 exceed 0.75. A smaller σ_{AIR} value indicates a pyranometer failure. For example σ_{AIR} will be
263 equal to zero when a pyranometer ceases to operate for a considerable period. Conversely, when
264 a pyranometer works intermittently, the variability of σ_{AIR} becomes unrealistically high. Hence,
265 half-hourly radiometric acquisitions are flagged out for further use, when the following boundary
266 condition is met:

$$267 \frac{1}{8}\mu\left(\frac{GLR_{in}}{R_{ext}}\right) < \sigma\left(\frac{GLR_{in}}{R_{ext}}\right) \leq 0.35.$$

268 This QA/QC statistic is computed using half-hourly radiation measurements acquired between
269 sunrise and sunset. *In-situ* VI's as defined in Table 2, are computed at half hour time intervals
270 from the radiometric data when these are not rejected by the QA/QC procedure criteria as
271 described above.

272 **2.3 MODIS products**

273 We used the 8-day 500 m surface reflectance product (MOD09G1, collection 5) from the
274 MODIS/TERRA satellite sensor/platform as provided by ORNL DAAC (see
275 http://daac.ornl.gov/cgi-bin/MODIS/GR_col5_1/mod_viz.html). MOD09G1 pixels matching the
276 coordinates of a FLUXNET site have been extracted. The VIs, as reported in Table 2, are

277 calculated from these surface reflectance values using band 1 (red: 620-670 nm) and band 2
278 (NIR: 841-876 nm). 8-day MOD09G1 pixel values represent the optimal reflectance values for
279 8-day compositing windows, selecting pixels with optimal viewing angles and minimal cloud or
280 cloud shadow impacts. Extracted time series Quality Assurance / Quality Control (QA/QC) flags
281 have been used, ensuring the quality of the MOD09G1 product. Specifically the MOD35 QA/QC
282 flags have been used to identify the presence of snow (i.e. “MOD35 snow/ice” flags equal to
283 “no”), clouds and cloud shadows (i.e. “MOD35 cloud” flags equal to “clear”). The MOD09G1
284 reflectance bands at 500 m were flagged as having the optimal quality for all bands (i.e.
285 “MODLAND QA bits” flags equal to “corrected product produced at ideal quality all bands”).
286 Only the pixels with the highest quality (e.g. clear conditions without snow) have been selected
287 and retained for further use.

288 **2.4 Canopy phenological variable derivation from NEE and *in-situ* and MODIS
289 VI's**

290 Canopy phenological variables are derived using MODIS as well as *in-situ* VI's as well as NEE
291 time-series data. TIMESAT v.3.1 software has been selected for VI time-series processing.
292 TIMESAT is available at the following URL: <http://www.nateko.lu.se/TIMESAT/timesat.asp>
293 (Jonsson and Eklundh 2002; 2004).

294 With respect to TIMESAT options for use, the adaptive Savitzky-Golay method for time-series
295 smoothing and the double sigmoid method to extract seasonally dependent variables from a time-
296 series have been selected for application. By definition, the adaptive Savitzky-Golay method
297 smooths a time-series with a total of N points i , which comply with: (t_i, V_i) with taking values of i
298 = 1, 2, ..., N. Each point i is fitted with a quadratic polynomial function as defined by eq. 3:

$$f(t) = c_1 + c_2 t + c_3 t^2 \quad (\text{eq. 3})$$

300 For all $2k + 1$ points within a time window ranging from $n = i - k$ to $m = i + k$, a linear combination
 301 of nearby values is solved according to eq. 4:

$$302 \quad \sum_j^n = c_j V_{i+j} \quad (\text{eq. 4})$$

303 In the simplest case, coefficients c_j are defined as $c_j = 1/(2n + 1)$ while the data value V_i is
304 replaced by the average of the data values in the time window as defined earlier. Time window
305 extent determines the amplitude of the degree of VI time-series smoothing.

306 A fitting time window of N=4 points i has been used to represent the temporal variability of NEE
307 fluxes as well as both VI types, i.e. the MODIS and *in-situ* VI's. A double sigmoid is applied to
308 be fitted through each smoothed time-series of beforementioned data types:

$$309 \quad V(t) = \frac{1}{\frac{vi_1-t}{1+e^{\frac{vi_2}{v_i}}}} - \frac{1}{\frac{vi_3-t}{1+e^{\frac{vi_4}{v_i}}}} \quad (\text{eq. 5})$$

310 In eq. 5:

- vi_1 is the position of the V curve part before an inflection point;

312 - v_i_2 is the rate of change of the variable curve before an inflection point;

- vi_3 is the position of the V curve part after an inflection point and;

314 - vi_4 is the rate of curve change after an inflection point.

315 The main phenological variables - the start and end of a growing season amplitude - are
316 determined by a threshold method as implemented in TIMESAT. A seasonal starting point (i.e.
317 SGS_{NEE} for NEE; SGS_{MODIS} for the MODIS VI's and $SGS_{in-situ}$ for *in-situ* VI's) is defined using
318 the double sigmoid function (see eq. 5) to determine the time point (in days) corresponding with

319 50% of the V time-series amplitude height and defined as the distance between the time-series
320 left side minimal and maximal levels. The end of the a growing season is defined similarly,
321 starting however from the time-series right side minimum.

322 **2.4.1 Evaluation of NEE phenology**

323 For the nine sites listed in Table S1 in the Supplementary data, NEE fluxes are continuosly
324 measured, and the resulting SGS_{NEE} date estimates are evaluated against a remote sensing
325 variable and index respectively, commonly used to estimate flux phenology, e.i., $fAPAR$ and
326 EVI. $fAPAR$ is derived from both components of radiation (i.e. PAR_{in} and PAR_{bc}) by using the
327 formula given by Monteith (1993) as:

328
$$fAPAR = 1 - \frac{PAR_{bc}}{PAR_{in}} e^{1.35} \quad (\text{eq. 6})$$

329 where PAR_{bc} is below canopy PAR and PAR_{in} is incident PAR. The exponent with value 1.35,
330 accounts for the mean effect of the different absorptivities in the PAR and global solar radiation
331 spectral bands. QA/QC of the half-hourly PAR_{in} and PAR_{bc} data is performed applying the same
332 tests as used to checking the quality of all other radiation measurements used to calculated VI's
333 (see section 2.2).

334 EVI has been derived from the 8-day 500 m surface reflectance MODIS product (MOD09G1,
335 collection 5, being the same dataset used for the main analysis in this study) and is calculated
336 according to eq. 7:

337
$$EVI = 2.5 * \frac{(\rho_{NIR} - \rho_{PAR})}{1 + \rho_{NIR} + 6 * \rho_{PAR} - 7.5 * \rho_{blue}} \quad (\text{eq. 7})$$

338 where ρ_{PAR} is band 1 (620-670 nm), ρ_{NIR} is band 2 (841-876 nm) and ρ_{blue} is band 3 (459-479
339 nm). Furthermore, the same QA/QC flags applied for the VI's calculated for MODIS (see section
340 2.3) have been used to check the QA/QC of the EVI time-series.

341 **2.5 Statistical analysis**

342 A correlation analysis is performed to investigate the relationship between $\text{SGS}_{\text{MODIS}}$ and $\text{SGS}_{\text{in-}}
343 \text{situ}$. The relationship was characterised using the following statistics:

- 344 - the coefficient of determination (R^2);
345 - the root mean square error (RMSE) and;
346 - the normalized mean bias (NMB).

347 Differences between PFT's have been assessed for each VI applying a statistical analysis to
348 quantify the correlation between $\text{SGS}_{\text{MODIS}}$ and $\text{SGS}_{\text{in-situ}}$ by binning FLUXNET sites according
349 to PFT type. The robustness of the statistical analysis has been tested by a leave-one-out cross-
350 validation technique. The predictive performance is expressed as a cross-validated root mean
351 square error (RMSE_{CV}).

352 To test the impact of VI on the relation between $\text{SGS}_{\text{MODIS}}$ and $\text{SGS}_{\text{in-situ}}$, we performed a
353 covariance analysis (ANCOVA) with $\text{SGS}_{\text{MODIS}}$ as response variable, $\text{SGS}_{\text{in-situ}}$ as the
354 explanatory variable of primary importance and VI as covariate using the PROC GLM routine
355 implemented in SAS (SAS 9.4; ©SAS Institute Inc., Cary, NC, USA). The relationship between
356 a VI and $\text{SGS}_{\text{in-situ}}$ is also included in the analysis. A second ANCOVA analysis where VI's were
357 replaced by PFT's, has been performed to test the PFT impact on the $\text{SGS}_{\text{MODIS}}$ vs. $\text{SGS}_{\text{in-situ}}$
358 relationship.

359 A two by two comparison of the slope of the regression relationship between SGS_{MODIS} and
360 $SGS_{in-situ}$ has been performed for the VI selected according to a best fit criterion (i.e. highest R^2
361 and minimal RMSE value) compared to the other VI's investigated. A two by two comparison
362 has been conducted to test the significance of differences in regression slopes between VI's
363 eliciting the highest correlation between SGS_{MODIS} and $SGS_{in-situ}$ compared to all other VI's.

364 To better characterise the impact of each PFT on the SGS_{MODIS} - $SGS_{in-situ}$ correlation on a best
365 fit, a third ANCOVA analysis with as response variable SGS_{MODIS} and explanatory variables
366 $SGS_{in-situ}$, PFT and their interaction, has been performed.

367 The performance of SGS_{MODIS} and $SGS_{in-situ}$ estimates derived from different MODIS and *in-situ*
368 VI's intended to predict SGS_{NEE} , has been investigated similarly to the procedures described
369 earlier in this chapter. However, the response variable selected for this case is SGS_{NEE} .

370 To confirm the hypothesis that differences between SGS_{NEE} and respectively the SGS values
371 derived from MODIS and *in-situ* VI's (i.e. SGS_{NEE} - SGS_{MODIS} and SGS_{NEE} - $SGS_{in-situ}$,
372 respectively) are related to VI seasonality, VI type and PFT properties, we applied a general
373 linear mixed effects model (GLMM). In this respect seasonality is represented by the amplitude
374 of a VI time-series evaluated with the TIMESAT software. Using the GLMM, a boundary
375 condition is that FLUXNET sites and measurement years are considered as random variables.
376 Time as a variable (i.e. measurement year) is spatially nested (i.e. FLUXNET sites are spatially
377 nested). The GLMM analysis is performed using the PROC MIXED routine implemented in
378 SAS (SAS 9.4; ©SAS Institute Inc., Cary, NC, USA).

379 **3. Results**

380 **3.1 Comparison of SGS_{MODIS} and SGS_{in-situ}**

381 For all PFT's, MODIS VI's predict the date of the start of season earlier in time than for the *in-*
382 *situ* VI's (Table 3), SGS_{in-situ} vs. SGS_{MODIS} correlations differ according to VI type ($F=8.17$; p
383 <0.0001 , not shown in Table 3 nor Fig 1).

384 Clearly, the SGS estimated using the NDVI elicits the highest correlation coefficient for MODIS
385 as well as *in-situ* observations (Fig. 1a; Table 3; $R^2 = 0.68$; $p < 0.05$). For the NDVI, the
386 SGS_{MODIS-NDVI} occurs roughly 20-21 days before the SGS_{in-situ-NDVI} (RMSE = 20.89 days). The
387 VI's, SR and WDRVI show quite satisfactory correlation coefficient values as well ($R^2 = 0.43$
388 and $R^2 = 0.46$, respectively). But SGS_{MODIS-SR} occurs more than 27 days before the SGS_{in-situ-SR}
389 (see Table 3 - RMSE of SR). For all other VI's, the values of the correlation coefficients (R^2)
390 drop below acceptable values for SGS_{MODIS} and SGS_{in-situ}.

391 A two by two comparison of the regression slopes of the VI relationships SGS_{MODIS} vs. SGS_{in-situ}
392 for the VI showing the highest R^2 values (NDVI) versus each of the other VI's reveals that the
393 SGS dates slopes derived from the SR and WDRVI VI's are not significantly different from the
394 NDVI slope ($F = 1.47$ and $p = 0.14$; $F = -0.99$ and $p = 0.32$, respectively, not shown in Table 3
395 nor Fig 1).

396

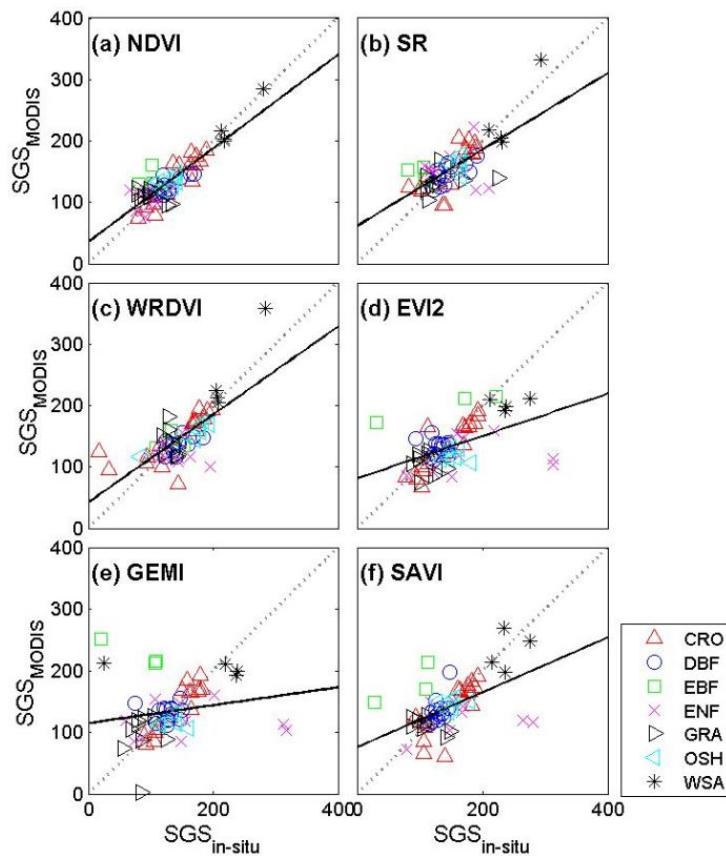
397

398

399

400

401 [Figure 1]



402

403 Figure 1 – Relationship between the start day of a growing season ($SGS_{in-situ}$) as derived from *in-*
 404 *situ* and MODIS (SGS_{MODIS}) VI's (see VI definitions in Table 2) for different PFT's (CRO—
 405 Cropland; DBF—Deciduous Broadleaf Forest; EBF—Evergreen Broadleaf Forest; ENF—
 406 Evergreen Needle-leaf Forest; GRA—Grassland; OSH—Open Shrubland; WSA—Woody
 407 Savanna). Black lines represent linear interpolation functions (for all PFT's pooled), dotted lines
 408 1:1 relationships.

409

410

411 [Table 3]

412

413 Table 3 - Statistics of correlation analysis between the starting day of the growing season derived
414 from MODIS (SGS_{MODIS}) and *in-situ* ($SGS_{in-situ}$) VI's (see VI definitions in Table 2), for all
415 PFT's, pooled (see Fig. 1). N. obs—number of available sites and years; R^2 —coefficient of
416 determination; RMSE—root mean square error; NMB—normalized mean bias; Y-int—y-
417 intercept of the linear model; Slope—slope of the linear model; R_{cv}^2 —cross-validated coefficient
418 of determination; and RMSE_{cv}—cross-validated root mean square error. Bold letters indicate the
419 model with the highest value of R^2 .

420

	N. obs	R^2 (-)	RMSE (day)	NMB (day)	Y-int (day)	Slope (-)	R_{cv}^2 (-)	RMSE _{cv} (day)
SGS_{MODIS} vs. $SGS_{in-situ}$								
NDVI	73	0.68*	20.89	0.0007	38.53	0.75	0.68	20.93
SR	73	0.43*	27.22	0.0002	70.40	0.56	0.42	27.24
WRDVI	71	0.46*	30.56	0.0000	55.62	0.63	0.46	30.48
EVI2	74	0.26*	32.73	-0.0012	83.25	0.35	0.26	32.80
GEMI	70	0.03	40.65	0.0004	117.29	0.13	0.03	40.67
SAVI	71	0.24*	34.23	-0.0002	84.80	0.39	0.25	34.26

421 *: p-value < 0.05

422

423 The relationship between $SGS_{MODIS-NDVI}$ and $SGS_{in-situ-NDVI}$ differs in magnitude according to the
424 type of PFT (Table 4; $F = 6.89$; $p < 0.0001$). For SGS dates derived from the NDVI, woody
425 savanna (SWA) elicits the highest correlation coefficient value (see Table 4). $SGS_{MODISNDVI}$ is
426 only 11 days earlier than $SGS_{in-situ-NDVI}$. Cropland (CRO) SGS dates, derived with the NDVI, is of
427 all PFT's considered, the one with the highest correlation coefficient ($R^2 = 0.81$). The RMSE

428 value of 17.20 days is quite high though.. $SGS_{MODIS-NDVI}$ and $SGS_{in-situ-NDVI}$ dates are
 429 significantly correlated as well for deciduous forest (See Table 4: DBF, $R^2 = 0.51$ and RMSE =
 430 8.70). The $SGS_{MODIS-NDVI}$ date occurs only 8 to 9 days before the $SGS_{in-situ-NDVI}$ date. For the
 431 remaining PFT's, non-significant relationships were found (see Table 4).

432

433 **[Table 4]**

434

435 Table 4 - Statistics of the correlation between the start day of the growing season (SGS) derived
 436 from MODIS (SGS_{MODIS}) and from *in-situ* observations ($SGS_{in-situ}$), with the SGS derived from
 437 the NDVI for each PFT's considered in this paper. R^2 —coefficient of correlation; RMSE—root
 438 mean square error; Y-int—y-intercept of the linear model; Slope—slope of the linear model;
 439 R_{cv}^2 —cross-validated coefficient of determination; and $RMSE_{cv}$ —cross-validated root mean
 440 square error. Bold letters indicate the model with the highest value of R^2 .

	N.obs	R^2	RMSE	NMB	Y-int	Slope	R^2_{cv}	$RMSE_{cv}$
	(-)	(day)	(day)	(day)	(-)	(-)	(-)	(day)
SGS_{MOSIS} vs. $SGS_{in-situ}$								
CRO	19	0.81*	17.20	0.000	0.05	1.00	0.81	17.14
DBF	15	0.51*	8.70	0.001	73.76	0.42	0.50	8.88
EBF	5	0.68	24.50	0.073	105.20	0.49	0.76	24.08
ENF	10	0.04	16.84	0.012	89.18	0.17	0.05	17.28
GRA	13	0.14	13.97	0.007	141.22	-0.28	0.15	14.35
OSH	7	0.35	9.95	0.003	66.12	0.51	0.30	9.86
WSA	4	0.94*	11.93	-0.007	-48.12	1.18	0.96	11.11

441 *: p-value < 0.05

442

443 A two by two comparison of the regression slope of the relationship SGS_{MODIS} vs. $SGS_{in-situ}$
444 between woody savanna (WSA), the PFT with the best highest R^2 , and all other PFT's reveals
445 that the correlation between WSA differs significantly from all other PFT's except for cropland
446 (CRO, $F = -0.63$; $p = 0.53$) and open shrubland (OSH, $F = -1.27$; $p = 0.21$).

447 **3.2 Performance of SGS_{MODIS} and $SGS_{in-situ}$ to predict SGS_{NEE}**

448 SGS estimates derived from MODIS VI's correlate better with SGS_{NEE} than those derived from
449 *in-situ* VI's (Fig. 2 and 3; Table 5). $SGS_{MODIS-NDVI}$ dates show the highest correlation coefficient
450 ($R^2 = 0.77$ and $R^2 = 0.65$). The $SGS_{MODIS-NDVI}$ prediction occurs at a point in time, 21-22 days
451 earlier than that of NEE (i.e. SGS_{NEE}). In contrast, the $SGS_{in-situ-NDVI}$ date occurs 25-26 days
452 earlier than that of SGS_{NEE} (see Table 5).

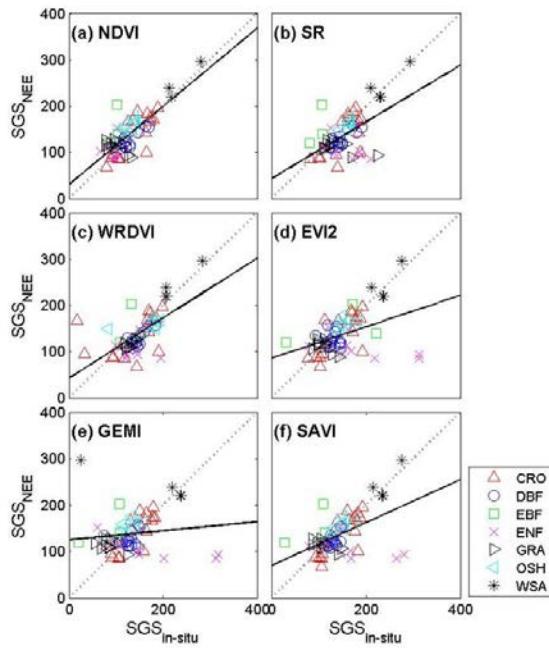
453 Non-significant correlations are found for the other *in-situ* VI's (except for the NDVI). Though
454 the SGS dates derived from the MODIS WRDVI performs satisfactory as well (Table 5; $R^2 =$
455 0.70). It predicts SGS dates 23-24 days earlier than the SGS_{NEE} dates. GEMI shows the poorest
456 correlations (see Table 5).

457 For many PFT's, SGS_{NEE} shows a higher correlation coefficient value with SGS_{MODIS} than with
458 $SGS_{in-situ}$, both SGS data derived from the NDVI (see Table 6). Woody savanna (WSA) elicits a
459 very good correlation between SGS_{NEE} and $SGS_{MODIS-NDVI}$ and $SGS_{in-situ-NDVI}$ (see Table 6).
460 Nevertheless, only one site has been used for this PFT.

461 Note however, that $SGS_{MODIS-NDVI}$ based estimates for deciduous forest (Table 6: DBF; $R^2 =$
462 0.74) elicits a high correlation coefficient. Moreover, the $SGS_{MODIS-NDVI}$ date is only only 8-9
463 days earlier than SGS_{NEE} .

464

465 [Figure 2]



466

467 Figure 2 – Relationships between the start day of the growing season, derived from *in-situ*
468 (SGS_{in-situ}) VI's (see VI definitions in Table 2) and NEE (SGS_{NEE}) for the different PFT's
469 considered in this paper. Black lines represent linear interpolation functions (for all PFT's
470 pooled), dotted lines 1:1 relationships.

471

472

473

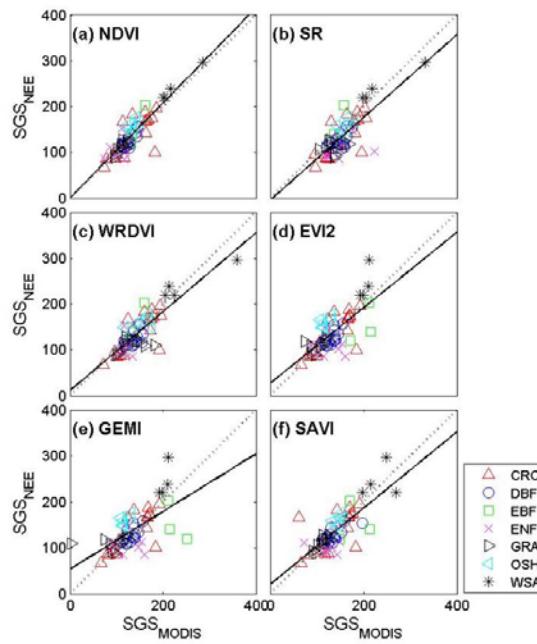
474

475

476

477

478 [Figure 3]



479

480 Figure 3 – Relationships between the day of the start of a growing season as derived from
481 MODIS (SGS_{MODIS}) vegetation indices and net carbon uptake (SGS_{NEE}) for different plant
482 functional types as in Fig. 2, except that the start of the growing season day is derived from
483 MODIS vegetation indices (SGS_{MODIS}) instead of *in-situ* vegetation indices.

484

485

486

487

488

489

490

491 [Table 5]

492

493 Table 5 - Statistics of correlation analysis between the day of the start of the growing season
494 (SGS) derived from NEE (SGS_{NEE}) and SGS derived from MODIS (SGS_{MODIS}) and *in-situ*
495 ($SGS_{in-situ}$) VI's (see definition in Tab. 2), and net carbon uptake (SGS_{NEE}), for all PFT's, pooled.
496 N. obs—number of available sites and years; R^2 —coefficient of determination; RMSE—root
497 mean square error; NMB—normalized mean bias; Y-int—y-intercept of the linear model;
498 Slope—slope of the linear model; R_{cv}^2 —cross-validate coefficient of determination; and
499 RMSE_{cv}—cross-validated root mean square error. Bold letters indicate the model with highest R^2
500 value.

	N. obs	R^2 (-)	RMSE (day)	NMB (day)	Y-int (day)	Slope (-)	R_{cv}^2 (-)	RMSE _{cv} (day)
SGS_{NEE} vs. SGS_{MODIS}								
NDVI	64	0.77*	20.50	0.0006	0.38	1.04	0.77	20.47
SR	64	0.59*	27.43	-0.0020	-4.50	0.90	0.59	27.44
WRDVI	64	0.70*	23.58	-0.0010	11.06	0.86	0.70	23.60
EVI2	64	0.51*	30.04	0.0004	26.82	0.82	0.51	30.14
GEMI	63	0.40*	33.47	0.0005	53.49	0.63	0.39	33.54
SAVI	64	0.58*	27.76	-0.0004	20.33	0.83	0.58	27.85
SGS_{NEE} vs. $SGS_{in-situ}$								
NDVI	64	0.65*	25.23	0.0014	30.35	0.84	0.65	25.23
SR	64	0.32*	35.20	-0.0016	42.99	0.61	0.32	35.12
WRDVI	62	0.40*	33.08	-0.0010	42.75	0.65	0.39	33.12
EVI2	64	0.18*	38.82	-0.0013	85.66	0.34	0.18	38.88
GEMI	62	0.02	41.74	0.0007	125.09	0.10	0.02	41.50
SAVI	62	0.26*	36.92	-0.0009	69.75	0.46	0.26	36.94

501 *: p-value < 0.05

502

503

504 [Table 6]

505

506 Table 6 - Correlation between the start day of the growing season (SGS) derived from NEE
 507 (SGS_{NEE}) and derived from MODIS (SGS_{MODIS}) respectively *in-situ* ($SGS_{in-situ}$) NDVI's. PFT—
 508 plant functional type; N. obs—number of available sites and years; R^2 —coefficient of
 509 determination; RMSE—root mean square error; NMB—normalized mean bias; Y-int—y-
 510 intercept of the linear model; Slope—slope of the linear model; R_{cv}^2 —cross-validate coefficient
 511 of determination; and RMSE_{cv}—cross-validated root mean square error. Bold letters indicate the
 512 model with highest R^2 value.

PFT	N.obs	R^2	RMSE	NMB	Y-int	Slope	R_{cv}^2	RMSE _{cv}
		(-)	(day)	(day)	(day)	(-)	(-)	(day)
SGS_{MODIS} derived from NDVI								
CRO	16	0.54*	31.24	0.0009	22.61	0.85	0.53	31.65
DBF	15	0.74*	8.02	0.0006	-13.89	1.10	0.74	8.03
EBF	3	0.96	11.44	0.0307	185.23	2.40	1.00	0.00
ENF	10	0.23	21.88	0.0041	36.18	0.70	0.24	21.90
GRA	10	0.59*	8.69	0.0007	7.99	0.93	0.57	8.73
OSH	6	0.37	9.49	0.0335	77.85	0.62	0.40	9.62
WSA	4	0.99*	3.75	0.0198	38.75	0.91	0.94	4.42
$SGS_{in-situ}$ derived from NDVI								
CRO	16	0.65*	27.17	0.0001	1.09	0.99	0.66	26.22
DBF	15	0.60*	10.00	0.0003	52.13	0.59	0.63	9.96
EBF	3	0.45	45.26	0.2063	-61.39	2.26	1.00	0.00
ENF	10	0.17	22.77	0.0163	63.46	0.49	0.16	22.81
GRA	10	0.49*	9.64	0.0124	156.10	-0.45	0.52	9.57
OSH	6	0.34	9.70	0.0376	96.24	0.49	0.28	10.38
WSA	4	0.91	13.59	0.0128	-1.31	1.06	1.00	0.00

513 *: p-value < 0.05

514

515 Note as well that $SGS_{MODIS-NDVI}$ dates correspond well with SGS_{NEE} for grassland (GRA) and
516 cropland (CRO). However, $SGS_{in-situ-NDVI}$ shows a higher R^2 and lower RMSE than $SGS_{MODIS-NDVI}$
517 for cropland. For cropland the MODIS WRDVI, EVI2 and GEMI show slightly higher
518 scores than the NDVI (see Table S3 and S4 in the Supplementary Data).

519 For grassland it is interesting to note that SGS_{MODIS} and $SGS_{in-situ}$ are better correlated with
520 SGS_{NEE} , when estimated with SAVI than with the NDVI (see Table S3 and S4 in Supplementary
521 Data).

522

523 *GLMM analysis*

524 To test the hypothesis that the prediction of the residuals of the SGS_{NEE} date from MODIS and
525 *in-situ* VI's (i.e. $SGS_{NEE} - SGS_{MODIS}$ and $SGS_{NEE} - SGS_{in-situ}$, respectively) are related to VI
526 seasonality, VI, PFT type and the variable amplitude (the difference between the maximum and
527 minimum value of a VI), a GLMM analysis was performed.

528 The results of the GLMM analysis of the residuals of the regression line between SGS_{NEE} and
529 SGS_{MODIS} reveals significant two way interactions between each of the three explanatory
530 variables (see Table 7). For *in-situ* data only the interaction effect PFT x VI significantly affects
531 the residuals of the SGS_{NEE} vs. $SGS_{in-situ}$ relationship. The other interctions were not significant
532 and therefore they were not taken into account in the final model (amp*VI: $F_{5,285} = 0.78$, $p =$
533 0.57; amp*PFT: $F_{6,290} = 1.03$, $p = 0.41$).

534 [Table 7]

535

536

537

538

539 Table 7 - Results of a GLMM analysis testing the sensitivity of the residuals of a regression line
540 between 'SGS_{NEE} and SGS_{MODIS}' and SGS_{MODIS} and 'SGS_{NEE} – SGS_{*in-situ*}' relationships, with
541 respect to PFT's and VI's and amplitude (amp). Effect—fixed effect in the LME; Num DF—
542 Numerator degree of freedom; Den DF—Denominator degree of freedom; F Value—value of
543 statistics; PFT—plant functional type; p—probability; VI—vegetation index; and amp—
544 difference between the maximum and minimum value of a VI.

545

Effect	Num DF	Den DF	F Value	p
<u>SGS_{NEE} – SGS_{MODIS}</u>				
amp	1	282	11.63	0.001
VI	5	282	0.67	0.646
PFT	6	282	3.98	0.001
amp*PFT	6	282	3.69	0.001
amp*VI	5	282	3.29	0.006
PFT*VI	30	282	4.22	<.0001
<u>SGS_{NEE} – SGS_{<i>in-situ</i>}</u>				
amp	1	296	0.41	0.522
VI	5	296	11.35	<.0001
PFT	6	296	2.48	0.023
PFT*VI	30	296	3.68	<.0001

546

*: p-value < 0.05

547

548 Figure 4 reports on the variation (e.g. mean value and standard error) of the SGS date estimates
549 using three different methods (i.e. (a) derived from NEE flux data, (b) derived from *in-situ*

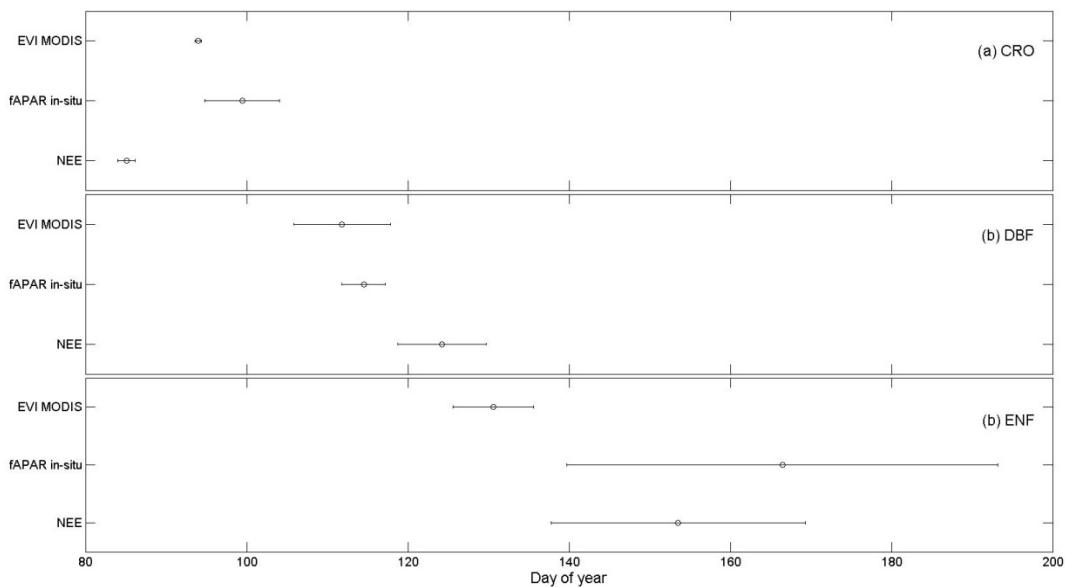
550 $fAPAR$ and (c) derived from MODIS EVI) for cropland, deciduous broadleaf forest and (c)
551 evergreen needleleaf forest sites reported in table S1 in the Supplementary Data.

552 Deciduous forest (DBF) shows a good agreement between the three methods. SGS date estimates
553 based on remote sensing are however uncertain for evergreen needle leaf forest (ENF).
554 Moreover, EVI predicts the SGS date at an earlier point in time than NEE and *in-situ* $fAPAR$.

555

556 [Figure 4]

557



558

559 Figure 4 – Start of growing season (SGS) date estimated by using three methods. NEE is mass
560 flux data measured at flux sites, $fAPAR$ in-situ is $fAPAR$ derived from *in-situ* radiation
561 measurements and EVI MODIS is EVI derived from MODIS data for (a) CRO—Cropland, (b)
562 DBF—Deciduous Broadleaf Forest, and (c) ENF—Evergreen Needle-leaf Forest sites locate in

563 the Northern Hemisphere. Dots represent mean values of SGS date estimates for the three cited
564 variables variable and lines represent data standard errors.

565 **4 Discussion**

566 **4.1 Phenological metrics of *in-situ* and satellite sensor acquisitions for different**
567 **VI's and PFT's**

568 Typically, MODIS VI's tend to predict the SGS date at an earlier time point in the season than
569 compared with the *in-situ* VI's (Figs. 1 and 2; Table 3). The disagreement between MODIS and
570 the *in-situ* VI's can be related to the different sensor characteristics, e.g. the different spectral
571 bandwidths (hence, different spectral resolutions) and instantaneous fields of view (IFOV's). The
572 latter depends on the distance between the canopy and the sensor position and field of view
573 angle. All VI's considered in this paper are computed using RED and NIR reflectances (see Table
574 2) from MODIS and *in-situ* acquisitions for 28 FLUXNET sites (Table 1). In that respect *in-situ*
575 RED and NIR reflectances are acquired with two extremely broad spectral bands (400 to 700 nm
576 and 700 to 2800 nm, respectively). Wilson and Meyers (2007) report that a steep increase of *in-*
577 *situ* VI's based on these broad RED and NIR bands indicates an increase of canopy greening
578 and/or vegetation cover at the canopy level. On the other hand, the MODIS sensor has a higher
579 spectral resolution than the *in-situ* pyranometers. Typically, the MODIS RED and NIR spectral
580 bandwidths span a spectral range from 620 to 670 nm and 841 to 876 nm respectively. This
581 difference in spectral resolution contributes to a difference in interpretation of the canopy
582 biophysical properties for a growing season, e.g. photosynthetic rate (Inoue et al. 2008).

583 In addition, during greening, canopy reflectance in the PAR region (PAR ranges from 400 to
584 700 nm) decreases due to an increase in PAR absorption by additional chlorophyll (and

585 photosynthesis) in the canopy due to new, emerging leaves (Ryu et al. 2008). Moreover, the
586 spectral signature of upwelling optical radiation in many PFT's changes due to a decreasing gap
587 fraction when time in the growing season progresses. Clearly, this is an issue of strong concern
588 and interest. NIR has a higher transmissivity in the canopy than RED and hence NIR reflectance
589 changes with canopy structure and largely opposite with respect to the PAR band (Ollinger et al.
590 2008).

591 Another issue to be mentioned is the difference in IFOV for MODIS and the *in-situ* observations.
592 The IFOV of a sensor and its orbit or acquisition position determines the surface area covered by
593 the sensor. A large difference exists between respectively the MODIS and the *in-situ* IFOV's. No
594 need to state that this does affect comparability of top-of-the-canopy reflectance and hence the
595 derived VI values for both sensor types (i.e., MODIS and *in-situ*). The area acquired by MODIS
596 is 500 x 500 m per pixel, e.g. 250,000 m². This a much larger surface area covered than with a
597 flux tower mounted sensor, even when this camera has a large IFOV, because flux towers do not
598 exceed canopy height very significantly. This brings up the issue of the differences in tower and
599 canopy height, between the different flux tower sites. The different sites have fundamentally
600 different IFOV's for the for *in-situ* observations except when corrected for by sensor fore-optics,
601 so that a match between the MODIS IFOV and the tower sensor IFOV is obtained. This however
602 is never the case for the FLUXNET sites as there is no standard procedure defined to guarantee a
603 consistently equal IFOV. Typically, the *in-situ* sensors are positioned horizontally at 1 up to 10
604 m above the canopy top level and near the flux sensors (for more details see Balzarolo et al.
605 (2011)). The remotely sensed response originating from MODIS top-of-the-canopy VI's has a
606 IFOV of 120°and a FOV of 250,000 m². In the case of the tower sensors, the radiation reflected
607 by the canopy originates from an IFOV of about 120° as well. However, the maximum surface

608 area observed by the *in-situ* sensor varies roughly between 5.44 and 54.41 m², assuming the
609 height of the sensor above the canopy varies between 1 to 10 m. That's a difference in magnitude
610 in FOV of about a factor 20,000, which is a difference of more than 5 orders of magnitude.
611 These huge differences in FOV make the evaluation of MODIS data compared with *in-situ*
612 measurements quite complex. Biophysical variables like gap fraction and LAI can be estimated
613 relatively accurately close to the canopy, but much more difficult at a spatial resolution of
614 500x500m, even for forest.

615 Site spatial heterogeneity can be estimated more accurately close to the top-of-the-canopy as
616 opposed to spaceborne observations and dependant on the PFT considered (Cescatti et al. 2012).
617 Of all VI's investigated in this paper, the NDVI shows the highest correlation coefficient for the
618 relationship between SGS_{MODIS} and SGS_{*in-situ*} (Fig. 1, Table 3), but with a 20-21 days earlier SGS
619 date obtained for MODIS than for *in-situ* sensors. Furthermore, the SGS_{MODIS} vs. SGS_{*in-situ*}
620 relationship differs according to the VI considered.

621 The NDVI is strictly related to the transition region between RED and NIR (i.e. the red-edge
622 region). The red-edge region is affected primarily, by leaf chlorophyll content and at low LAI
623 values by the spectral properties of the soil (or snow) as well. These boundary conditions
624 determine the spectral signature of the canopy during the growing season. Several authors
625 reported that the difference in spectral resolution between MODIS and *in-situ* sensors leads to
626 different VI values, certainly when the spectral signature changes with increasing LAI during the
627 growing season.

628 In general, most carbon balance research focusses on a comparison of the NDVI derived from
629 MODIS and *in-situ* radiation measurements, mostly for validation purposes. However, Wilson
630 and Meyers (2007) compared *in-situ* NDVI observations, derived from the same tower set-up's

631 used in this study, with a 1x1 km 16 days composite MODIS NDVI. They reported that the
632 MODIS VI values show slightly larger amplitude than the *in-situ* vegetation index values.

633 **4.2. An optimal remote sensing proxy to characterise CO₂ mass flux phenology**

634 This paper describes that MODIS VI performance is more optimal as a remote sensing proxy for
635 SGS_{NEE} (Figs. 2a and 3a; Table 5) with the boundary condition, that PFT data are pooled. As
636 discussed earlier, the differences in MODIS and *in-situ* sensor characteristics determine the final
637 result for different VI's and, consequently, affect the day of the start of carbon uptake SGS_{NEE} for
638 the different PFT's. Xiao et al. (2008) reported that the discrepancies between NEE fluxes
639 estimated with MODIS VI's and the actually measured NEE are strictly related to the spatial
640 complexity of the ecosystems in the MODIS pixel area (e.g. 1 km x 1 km).

641 For instance, different plant species within the same eddy covariance footprint will vary in their
642 contribution to the NEE making it difficult to predict the phenological cycle of an ecosystem as a
643 whole (Ma et al. 2007). Fisher and Mustard (2007) reported that changes in MODIS NDVI at the
644 beginning of the growing season are not in phase with plant carbon dynamics but rather plant
645 biomass dynamics. Likewise, this study demonstrates that the *in-situ* NDVI, in addition to the
646 MODIS NDVI, is more sensitive to biomass rather than to carbon dynamics. It has been
647 established quite exhaustively by many authors that the NDVI is a proxy for fAPAR (and to
648 some extent LAI) estimation. However, the NDVI is not sensitive to short-term changes
649 (changes occurring in less than a week) in photosynthetic activity (Gamon et al. 1992; Gitelson
650 2004; Hmimina et al. 2014). In addition, several studies indicate that photosynthetic capacity
651 does not reach its maximum during the greening phase. For instance, the lag between flux
652 phenology (SGS_{NEE}) and canopy greenness (SGS_{MODIS} and SGS_{*in-situ*}), as observed from the VI's
653 can be explained by a difference in time lag between ecosystem photosynthetic capacity and leaf

expansion during spring for beech trees (i.e. see Supplementary data, for the Hainich site as described in Knohl et al. (2003)). In addition, Morecroft et al. (2003) stated that the full photosynthetic capacity of *Quercus robur* leaves is reached 50 days after bud break. At the start of the season, when the canopy is developing, an increase in carbon uptake is typically associated with an increase in soil respiration, able to reduce NEE substantially (Xiao et al. 2008). The study of Ryu et al. (2014) compares MODIS and *in-situ* leaf-out observations with optical sensors (i.e. LED and LAI-2000 measurements). It reports that the MODIS NDVI is able to sense the signal of understory leaf-out obtained from *in-situ* observations for deciduous forest. In addition, another study showed that the MODIS NDVI predicts an earlier leaf-out than *in-situ* observations for overstory leaf-out (Ganguly et al. 2010). For a deciduous forest, the leaf-out phase of the understorey canopy tends to occur earlier than that of the overstory.

Our results agree well with previous results obtained for deciduous ecosystems. For these ecosystems, a strong dependence of photosynthetic activity on leaf area expansion and MODIS and *in-situ* VI patterns agree well with the dynamics of NEE (see Supplementary Data). In grassland ecosystems for instance, a low variation of MODIS and *in-situ* VI's partially reflects GPP and NEE seasonal variations. Wohlfahrt et al. (2010) demonstrated that *in-situ* NDVI can be a proxy for carbon fluxes at least for two temperate mountain grasslands in Austria. Furthermore, the good performance of MODIS and *in-situ* SAVI to predict the start day of carbon uptake for different types of cropland (Table S3 and S4 in Supplementary Data) may be related to the presence of bare soil or fallow / sparse vegetation affecting the spectral signature of the soil surface from a mixture of soil and vegetation to homogeneous vegetation during the course of the growing season. This is particularly true for grassland sites where at the beginning of the growing season, the grassland canopy is not fully developed and hence the gap fraction of

677 the canopy is high (or the *fCover* very low). The presence of the additional factor L (see Table 2)
678 for the RED reflectance in the denominator of the SAVI equation, makes the vegetation index
679 less sensitive to soil darkening due to an increase in soil moisture. Therefore, the spatial
680 distribution of the vegetation for grassland PFT's is assumed to play a major role in the
681 determination of the start of the growing season and, hence the start of carbon dioxide uptake.

682 For croplands we found that the MODIS WDRVI elicits a higher correlation coefficient value
683 than the NDVI (Table S2 in Supplementary Data). This is a confirmation that the WDRVI is a
684 good proxy for cropland phenology (Gitelson et al., 2004).

685 For evergreen broadleaf forests we didn't find high enough significant correlation coefficient
686 values any more (Table 6). Typically all evergreen broadleaf sites described in this paper are
687 located in tropical regions and characterised by a high and relatively constant photosynthetic and
688 carbon activity over an entire year. Seasonal variations in carbon balance have been described
689 (e.g. Bonal et al. (2008)), but this variation is clearly not reflected by MODIS and *in situ* NDVI's
690 (Hmimina et al. 2013) and certainly not comparable with temperate zone PFT carbon dynamics
691 variability.

692 The difficulty to predict SGS_{NEE} for the EBF PFT sites (Figs. 2 and 3) is clearly due to the
693 discrepancy between canopy physiology and phenology. Canopy phenology remains rather
694 stable (Hilker et al. 2014), whereas canopy physiology depends on seasonal variations in
695 environmental factors (mainly radiation and soil water availability, especially in monsoon forced
696 ecosystems) (Monson et al. 2005). In addition, for tropical regions long rainy seasons make it
697 difficult to collect both *in-situ* measurements and clear-sky satellite imagery (Hmimina et al.
698 2013).

699 Also for evergreen needleleaf forest significant correlations were not found. The annual
700 phenological cycle of evergreen needle leaf boreal forests in Sweden is related more to snow and
701 snow melt, than changes in needle canopy greening dynamics (Jönsson et al. 2010). For
702 evergreen needleleaf forests, changes in greenness at the start of the growing season are
703 decoupled from the start of the carbon uptake season and hence ecosystem physiological activity
704 (Zwiazek et al. 2001).

705 Finally, the amplitude (i.e. the difference between the maximum and minimum value of a VI for
706 each growing season), VI type and PFT properties affect the residuals of the correlation function
707 between SGS_{NEE} and SGS_{MODIS} . This suggests that differences in predicting SGS_{NEE} with
708 different VI's depends on amplitude differences for both carbon and VI dynamics. This suggests
709 that it is not likely to develop a generic model for the description and modelling of flux
710 phenology for all global PFT's and ecosystems. Even though the NDVI derived from both
711 MODIS and *in-situ* data shows a good correlation for all PFTs, pooled (Tabs. 5-6), a VI for a
712 PFT improves the estimation of the SGS date for a that specific PFT (e.g. the WRDVI for
713 cropland for example). Therefore, further efforts should focus on the understanding of the most
714 appropriate VI or a combination of different VI's or maybe even multi-dimensional hyperspectral
715 VI's, which may have the capacity to describe the clearcut complexity of flux phenology (Wong
716 and Gamon 2015).

717 **5. Conclusions**

718 MODIS and *in-situ* VI's show consistent results. Of all VI's considered in this paper, the NDVI
719 shows the highest correlation coefficient for the relationship between the starting day of the
720 growing season as observed with MODIS and *in-situ* observations. Also, the MODIS NDVI

721 performs best when applied as an estimator for Net Ecosystem Exchange but only with the
722 boundary condition that all PFT's are pooled. Nonetheless, it has been elicited that a specific VI
723 can be applied to improve the estimation of a SGS date for a specific PFT - for example the
724 WRDVI for cropland, which is however suboptimal for the other PFT's.

725 Summarizing, this study suggests that:

726 (i) *In-situ* radiation data measurements are a good approach to bridge the gap between local
727 eddy covariance carbon fluxes and MODIS global VI acquisitions;

728 (ii) Methodological improvement and the use of hyperspectral optical sensors is required at
729 the flux towers to better describe ecosystem carbon dynamics and carbon dioxide flux
730 phenology (Porcar-Castell et al. 2015).

731 (iii) A generic model used to estimate flux phenology for all ecosystems is still a bottleneck
732 issue, though multi-dimensional VI's as obtained from hyperspectral remote sensing are a
733 good possibility to develop a generic model (Rivera et al. 2014).

734 (iv) Further work should explore the utility of the new forthcoming super-spectral
735 'Copernicus' Sentinel-2 and Sentinel-3 missions that will provide a vast data stream
736 helpful to understand the physiological and photosynthetic activity of the canopy driven
737 by seasonally changing pigment concentrations (e.g. chlorophylls) and fluorescence
738 (Van Wittenberghe et al. 2013; Van Wittenberghe et al. 2014; Verrelst et al. 2015).

739 Finally, the work presented in this paper confirms the importance of ecosystem (top-of-the-
740 canopy scale) remote sensing observations to better describe global ecosystem phenological
741 metrics as well as to validate satellite VI's as upscaling proxies. In this regard, the establishment
742 of long-term global monitoring networks such as ICOS (www.icos-infrastructure.eu) NEON
743 (www.neoninc.org) and AmeriFlux (<http://ameriflux.lbl.gov>), foster the use of *in-situ*

744 measurements and provide a unique framework for this type of activity, which may ultimately
745 lead to more accurate estimates of the global terrestrial carbon balance.

746

747 **Acknowledgements**

748 This work used radiometric and eddy covariance measurements acquired by the FLUXNET
749 community and in particular by the following networks: AmeriFlux (US Department of Energy,
750 Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and
751 DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly,
752 CarboMont, ChinaFlux, FLUXNET — Canada (supported by CFCAS, NSERC, BIOCAP,
753 Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS —
754 Siberia, and USCCC. We acknowledge the support to data harmonization provided by
755 CarboEuropeIP, FAO-GTOS-TCO, iLEAPS (the Integrated Land Ecosystem-Atmosphere
756 Processes Study, a core project of IGBP), Max Planck Institute for Biogeochemistry, National
757 Science Foundation, University of Tuscia, Université Laval and Environment Canada and US
758 Department of Energy and the database development and technical support from Berkeley Water
759 Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge
760 National Laboratory, University of California — Berkeley, and University of Virginia. We
761 would like to also acknowledge the PIs of the eddy covariance sites and all technicians, students,
762 PhDs and postdocs for providing their data and support to FLUXNET network. The MODIS data
763 used in this study are distributed by the Land Processes Distributed Active Archive Center (LP
764 DAAC), located at the US Geological Survey (USGS) Earth Resources Observation and Science
765 (EROS) Center (lpdaac.usgs.gov). I acknowledge the support from the Methusalem program of
766 the Flemish Government. J. Peñuelas research was supported by the European Research Council

767 Synergy grant ERC-2013-SyG-610028 IMBALANCE-P. Y. H. Fu was granted by Pegasus
768 Marie Curie Fellow. I thank my husband Nicola Arriga, PhD, for his invaluable help in
769 understanding physics underlying radiation measurements and data processing. Lastly, we thank
770 the anonymous referees and the editor for their constructive comments that improved this paper.

771

772 References

- 773 Baldocchi, D., Falge, E., Gu, L.H., Olson, R., Hollinger, D., Running, S., Anthoni, P.,
774 Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X.H.,
775 Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K.T.P., Pilegaard, K., Schmid, H.P.,
776 Valentini, R., Verma, S., Vesala, T., Wilson, K., & Wofsy, S. (2001). FLUXNET: A
777 new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide,
778 water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*,
779 82, 2415-2434
- 780 Baldocchi, D.D., Black, T.A., Curtis, P.S., Falge, E., Fuentes, J.D., Granier, A., Gu, L.,
781 Knohl, A., Pilegaard, K., Schmid, H.P., Valentini, R., Wilson, K., Wofsy, S., Xu, L., &
782 Yamamoto, S. (2005). Predicting the onset of net carbon uptake by deciduous forests
783 with soil temperature and climate data: a synthesis of FLUXNET data. *Int J Biometeorol*,
784 49, 377-387
- 785 Balzarolo, M., Anderson, K., Nichol, C., Rossini, M., Vescovo, L., Arriga, N., Wohlfahrt,
786 G., Calvet, J.C., Carrara, A., Cerasoli, S., Cogliati, S., Daumard, F., Eklundh, L., Elbers,
787 J.A., Evrendilek, F., Handcock, R.N., Kaduk, J., Klumpp, K., Longdoz, B., Matteucci, G.,
788 Meroni, M., Montagnani, L., Ourcival, J.M., Sanchez-Canete, E.P., Pontailler, J.Y.,
789 Juszczak, R., Scholes, B., & Martin, M.P. (2011). Ground-based optical measurements
790 at European flux sites: a review of methods, instruments and current controversies.
791 *Sensors (Basel)*, 11, 7954-7981
- 792 Betancourt, J.L., Schwartz, M.D., Breshears, D.D., Cayan, D.R., Dettinger, M.D.,
793 Inouye, D.W., Post, E., & Reed, B.C. (2005). Implementing a U.S. National Phenology
794 Network. *Eos, Transactions American Geophysical Union*, 86, 539-539
- 795 Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., Bonnefond, J.-M.,
796 Elbers, J.A.N., Longdoz, B., Epron, D., Guehl, J.-M., & Granier, A. (2008). Impact of
797 severe dry season on net ecosystem exchange in the Neotropical rainforest of French
798 Guiana. *Glob Chang Biol*, 14, 1917-1933
- 799 Cescatti, A., Marcolla, B., Santhana Vannan, S.K., Pan, J.Y., Román, M.O., Yang, X.,
800 Ciais, P., Cook, R.B., Law, B.E., Matteucci, G., Migliavacca, M., Moors, E., Richardson,
801 A.D., Seufert, G., & Schaaf, C.B. (2012). Intercomparison of MODIS albedo retrievals
802 and in situ measurements across the global FLUXNET network. *Remote Sensing of
803 Environment*, 121, 323-334

- 804 Churkina, G., Schimel, D., Braswell, B.H., & Xiao, X. (2005). Spatial analysis of growing
805 season length control over net ecosystem exchange. *Glob Chang Biol*, 11, 1777-1787
- 806 Gamon, J.A., Peñuelas, J., & Field, C.B. (1992). A narrow-waveband spectral index that
807 tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*,
808 41, 35-44
- 809 Ganguly, S., Friedl, M.A., Tan, B., Zhang, X., & Verma, M. (2010). Land surface
810 phenology from MODIS: Characterization of the Collection 5 global land cover dynamics
811 product. *Remote Sensing of Environment*, 114, 1805-1816
- 812 Garrity, S.R., Bohrer, G., Maurer, K.D., Mueller, K.L., Vogel, C.S., & Curtis, P.S. (2011).
813 A comparison of multiple phenology data sources for estimating seasonal transitions in
814 deciduous forest carbon exchange. *Agricultural and Forest Meteorology*, 151, 1741-
815 1752
- 816 Gitelson, A.A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification
817 of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, 161, 165-173
- 818 Gitelson, A.A., Keydan, G.P., & Merzlyak, M.N. (2006). Three-band model for
819 noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher
820 plant leaves. *Geophysical Research Letters*, 33, L11402
- 821 Heinsch, F.A., Zhao, M., Running, S.W., Kimball, J.S., Nemani, R.R., Davis, K.J.,
822 Bolstad, P.V., Cook, B.D., Desai, A.R., & Ricciuto, D.M. (2006). Evaluation of remote
823 sensing based terrestrial productivity from MODIS using regional tower eddy flux
824 network observations. *Geoscience and Remote Sensing, IEEE Transactions on*, 44,
825 1908-1925
- 826 Hilker, T., Lyapustin, A.I., Tucker, C.J., Hall, F.G., Myneni, R.B., Wang, Y., Bi, J.,
827 Mendes de Moura, Y., & Sellers, P.J. (2014). Vegetation dynamics and rainfall
828 sensitivity of the Amazon. *Proceedings of the National Academy of Sciences*, 111,
829 16041-16046
- 830 Hmimina, G., Dufrêne, E., Pontailler, J.Y., Delpierre, N., Aubinet, M., Caquet, B., de
831 Grandcourt, A., Burban, B., Flechard, C., Granier, A., Gross, P., Heinesch, B., Longdoz,
832 B., Moureaux, C., Ourcival, J.M., Rambal, S., Saint André, L., & Soudani, K. (2013).
833 Evaluation of the potential of MODIS satellite data to predict vegetation phenology in
834 different biomes: An investigation using ground-based NDVI measurements. *Remote
835 Sensing of Environment*, 132, 145-158
- 836 Hmimina, G., Dufrene, E., & Soudani, K. (2014). Relationship between photochemical
837 reflectance index and leaf ecophysiological and biochemical parameters under two
838 different water statuses: towards a rapid and efficient correction method using real-time
839 measurements. *Plant Cell Environ*, 37, 473-487
- 840 Huemmrich, K.F., Black, T.A., Jarvis, P.G., McCaughey, J.H., & Hall, F.G. (1999). High
841 temporal resolution NDVI phenology from micrometeorological radiation sensors.
842 *Journal of Geophysical Research: Atmospheres*, 104, 27935-27944
- 843 Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreira, L.G. (2002).
844 Overview of the radiometric and biophysical performance of the MODIS vegetation
845 indices. *Remote Sensing of Environment*, 83, 195-213

- 846 Huete, A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of*
847 *Environment*, 25, 295-309
- 848 Huete, A.R., Liu, H.Q., Batchily, K., & van Leeuwen, W. (1997). A comparison of
849 vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of*
850 *Environment*, 59, 440-451
- 851 Inoue, Y., Peñuelas, J., Miyata, A., & Mano, M. (2008). Normalized difference spectral
852 indices for estimating photosynthetic efficiency and capacity at a canopy scale derived
853 from hyperspectral and CO₂ flux measurements in rice. *Remote Sensing of*
854 *Environment*, 112, 156-172
- 855 Jenkins, J.P., Richardson, A.D., Braswell, B.H., Ollinger, S.V., Hollinger, D.Y., & Smith,
856 M.L. (2007). Refining light-use efficiency calculations for a deciduous forest canopy
857 using simultaneous tower-based carbon flux and radiometric measurements.
858 *Agricultural and Forest Meteorology*, 143, 64-79
- 859 Jiang, Z., Huete, A.R., Didan, K., & Miura, T. (2008). Development of a two-band
860 enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112,
861 3833-3845
- 862 Jönsson, A.M., Eklundh, L., Hellström, M., Bärring, L., & Jönsson, P. (2010). Annual
863 changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow
864 dynamics and tree phenology. *Remote Sensing of Environment*, 114, 2719-2730
- 865 Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-
866 series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on*,
867 40, 1824-1832
- 868 Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of
869 satellite sensor data. *Computers & Geosciences*, 30, 833-845
- 870 Keenan, T.F., Gray, J., Friedl, M.A., Toomey, M., Bohrer, G., Hollinger, D.Y., Munger,
871 J.W., O'Keefe, J., Schmid, H.P., Wing, I.S., Yang, B., & Richardson, A.D. (2014). Net
872 carbon uptake has increased through warming-induced changes in temperate forest
873 phenology. *Nature Clim. Change*, 4, 598-604
- 874 Knohl, A., Schulze, E.-D., Kolle, O., & Buchmann, N. (2003). Large carbon uptake by an
875 unmanaged 250-year-old deciduous forest in Central Germany. *Agricultural and Forest*
876 *Meteorology*, 118, 151-167
- 877 Loveland, T.R., & Belward, A.S. (1997). The IGBP-DIS global 1km land cover data set,
878 DISCover: First results. *Int J Remote Sens*, 18, 3289-3295
- 879 Monson, R., Sparks, J., Rosenstiel, T., Scott-Denton, L., Huxman, T., Harley, P.,
880 Turnipseed, A., Burns, S., Backlund, B., & Hu, J. (2005). Climatic influences on net
881 ecosystem CO₂ exchange during the transition from wintertime carbon source to
882 springtime carbon sink in a high-elevation, subalpine forest. *Oecologia*, 146, 130-147
- 883 Morecroft, M.D., Stokes, V.J., & Morison, J.I.L. (2003). Seasonal changes in the
884 photosynthetic capacity of canopy oak (*Quercus robur*) leaves: the impact of slow
885 development on annual carbon uptake. *Int J Biometeorol*, 47, 221-226

- 886 Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y.,
887 Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P.,
888 Nemani, R.R., & Running, S.W. (2002). Global products of vegetation leaf area and
889 fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*,
890 83, 214-231
- 891 Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., & Rundquist, D.
892 (2012). Green Leaf Area Index Estimation in Maize and Soybean: Combining
893 Vegetation Indices to Achieve Maximal Sensitivity. *Agron. J.*, 104, 1336-1347
- 894 Ollinger, S.V., Richardson, A.D., Martin, M.E., Hollinger, D.Y., Frolking, S.E., Reich,
895 P.B., Plourde, L.C., Katul, G.G., Munger, J.W., Oren, R., Smith, M.L., Paw, U.K.,
896 Bolstad, P.V., Cook, B.D., Day, M.C., Martin, T.A., Monson, R.K., & Schmid, H.P.
897 (2008). Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal
898 forests: Functional relations and potential climate feedbacks. *Proc Natl Acad Sci U S A*,
899 105, 19336-19341
- 900 Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W.,
901 Longdoz, B., Rambal, S., Valentini, R., Vesala, T., & Yakir, D. (2006). Towards a
902 standardized processing of Net Ecosystem Exchange measured with eddy covariance
903 technique: algorithms and uncertainty estimation. *Biogeosciences*, 3, 571-583
- 904 Peng, S., Piao, S., Ciais, P., Myneni, R.B., Chen, A., Chevallier, F., Dolman, A.J.,
905 Janssens, I.A., Penuelas, J., Zhang, G., Vicca, S., Wan, S., Wang, S., & Zeng, H.
906 (2013). Asymmetric effects of daytime and night-time warming on Northern Hemisphere
907 vegetation. *Nature*, 501, 88-92
- 908 Petach, A.R., Toomey, M., Aubrecht, D.M., & Richardson, A.D. (2014). Monitoring
909 vegetation phenology using an infrared-enabled security camera. *Agricultural and
910 Forest Meteorology*, 195–196, 143-151
- 911 Piao, S., Fang, J., Zhou, L., Ciais, P., & Zhu, B. (2006). Variations in satellite-derived
912 phenology in China's temperate vegetation. *Glob Chang Biol*, 12, 672-685
- 913 Piao, S., Tan, J., Chen, A., Fu, Y.H., Ciais, P., Liu, Q., Janssens, I.A., Vicca, S., Zeng,
914 Z., Jeong, S.-J., Li, Y., Myneni, R.B., Peng, S., Shen, M., & Penuelas, J. (2015). Leaf
915 onset in the northern hemisphere triggered by daytime temperature. *Nat Commun*, 6
- 916 Porcar-Castell, A., Mac Arthur, A., Rossini, M., Eklundh, L., Pacheco-Labrador, J.,
917 Anderson, K., Balzarolo, M., Martín, M., Jin, H., & Tomelleri, E. (2015). EUROSPEC: at
918 the interface between remote sensing and ecosystem CO₂ flux measurements in
919 Europe. *Biogeosciences Discuss*, 12, 13069-13121
- 920 Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., & Ohlen, D.O.
921 (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation
922 Science*, 5, 703-714
- 923 Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P.,
924 Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havráneková, K.,
925 Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G.,
926 Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E.,
927 Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., & Valentini, R.

- 928 (2005). On the separation of net ecosystem exchange into assimilation and ecosystem
929 respiration: review and improved algorithm. *Glob Chang Biol*, 11, 1424-1439
- 930 Richardson, A.D., Black, T.A., Ciais, P., Delbart, N., Friedl, M.A., Gobron, N., Hollinger,
931 D.Y., Kutsch, W.L., Longdoz, B., Luysaert, S., Migliavacca, M., Montagnani, L.,
932 Munger, J.W., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri,
933 E., Vargas, R., & Varlagin, A. (2010). Influence of spring and autumn phenological
934 transitions on forest ecosystem productivity. *Philos Trans R Soc Lond B Biol Sci*, 365,
935 3227-3246
- 936 Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., & Ollinger, S.V. (2009).
937 Near-surface remote sensing of spatial and temporal variation in canopy phenology.
938 *Ecol Appl*, 19, 1417-1428
- 939 Rivera, J., Verrelst, J., Delegido, J., Veroustraete, F., & Moreno, J. (2014). On the Semi-
940 Automatic Retrieval of Biophysical Parameters Based on Spectral Index Optimization.
941 *Remote Sensing*, 6, 4927
- 942 Ryu, Y., Baldocchi, D.D., Ma, S., & Hehn, T. (2008). Interannual variability of
943 evapotranspiration and energy exchange over an annual grassland in California.
944 *Journal of Geophysical Research: Atmospheres*, 113, D09104
- 945 Ryu, Y., Lee, G., Jeon, S., Song, Y., & Kimm, H. (2014). Monitoring multi-layer canopy
946 spring phenology of temperate deciduous and evergreen forests using low-cost spectral
947 sensors. *Remote Sensing of Environment*, 149, 227-238
- 948 Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., & Arkebauer,
949 T.J. (2010). A Two-Step Filtering approach for detecting maize and soybean phenology
950 with time-series MODIS data. *Remote Sensing of Environment*, 114, 2146-2159
- 951 Schmid, H.P. (2002). Footprint modeling for vegetation atmosphere exchange studies: a
952 review and perspective. *Agricultural and Forest Meteorology*, 113, 159-183
- 953 Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring
954 vegetation. *Remote Sensing of Environment*, 8, 127-150
- 955 Van Wittenberghe, S., Alonso, L., Verrelst, J., Hermans, I., Delegido, J., Veroustraete,
956 F., Valcke, R., Moreno, J., & Samson, R. (2013). Upward and downward solar-induced
957 chlorophyll fluorescence yield indices of four tree species as indicators of traffic pollution
958 in Valencia. *Environmental Pollution*, 173, 29-37
- 959 Van Wittenberghe, S., Alonso, L., Verrelst, J., Hermans, I., Valcke, R., Veroustraete, F.,
960 Moreno, J., & Samson, R. (2014). A field study on solar-induced chlorophyll
961 fluorescence and pigment parameters along a vertical canopy gradient of four tree
962 species in an urban environment. *Science of the Total Environment*, 466, 185-194
- 963 Verma, M., Friedl, M.A., Richardson, A.D., Kiely, G., Cescatti, A., Law, B.E., Wohlfahrt,
964 G., Gielen, B., Roupsard, O., Moors, E.J., Toscano, P., Vaccari, F.P., Ganelle, D.,
965 Bohrer, G., Varlagin, A., Buchmann, N., van Gorsel, E., Montagnani, L., & Propastin, P.
966 (2014). Remote sensing of annual terrestrial gross primary productivity from MODIS: an
967 assessment using the FLUXNET La Thuile data set. *Biogeosciences*, 11, 2185-2200

- 968 Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers,
969 J.G., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial
970 vegetation bio-geophysical properties—A review. *ISPRS Journal of Photogrammetry and*
971 *Remote Sensing*, 108, 273-290
- 972 Viña, A., Gitelson, A.A., Guy-Robertson, A.L., & Peng, Y. (2011). Comparison of
973 different vegetation indices for the remote assessment of green leaf area index of crops.
974 *Remote Sensing of Environment*, 115, 3468-3478
- 975 Walker, J.J., de Beurs, K.M., & Wynne, R.H. (2014). Dryland vegetation phenology
976 across an elevation gradient in Arizona, USA, investigated with fused MODIS and
977 Landsat data. *Remote Sensing of Environment*, 144, 85-97
- 978 Wang, Y., Woodcock, C.E., Buermann, W., Stenberg, P., Voipio, P., Smolander, H.,
979 Häme, T., Tian, Y., Hu, J., Knyazikhin, Y., & Myneni, R.B. (2004). Evaluation of the
980 MODIS LAI algorithm at a coniferous forest site in Finland. *Remote Sensing of*
981 *Environment*, 91, 114-127
- 982 Wilson, T.B., & Meyers, T.P. (2007). Determining vegetation indices from solar and
983 photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology*, 144,
984 160-179
- 985 Wohlfahrt, G., Pilloni, S., Hörtnagl, L., & Hammerle, A. (2010). Estimating carbon
986 dioxide fluxes from temperate mountain grasslands using broad-band vegetation
987 indices. *Biogeosciences (Online)*, 7, doi:10.5194/bg-5197-5683-2010
- 988 Wong, C., & Gamon, J.A. (2015). The photochemical reflectance index provides an
989 optical indicator of spring photosynthetic activation in evergreen conifers. *New*
990 *Phytologist*, 206, 196-208
- 991 Wu, W. (2014). The Generalized Difference Vegetation Index (GDVI) for Dryland
992 Characterization. *Remote Sensing*, 6, 1211-1233
- 993 Xiao, J., Zhuang, Q., Baldocchi, D.D., Law, B.E., Richardson, A.D., Chen, J., Oren, R.,
994 Starr, G., Noormets, A., Ma, S., Verma, S.B., Wharton, S., Wofsy, S.C., Bolstad, P.V.,
995 Burns, S.P., Cook, D.R., Curtis, P.S., Drake, B.G., Falk, M., Fischer, M.L., Foster, D.R.,
996 Gu, L., Hadley, J.L., Hollinger, D.Y., Katul, G.G., Litvak, M., Martin, T.A., Matamala, R.,
997 McNulty, S., Meyers, T.P., Monson, R.K., Munger, J.W., Oechel, W.C., Paw U, K.T.,
998 Schmid, H.P., Scott, R.L., Sun, G., Suyker, A.E., & Torn, M.S. (2008). Estimation of net
999 ecosystem carbon exchange for the conterminous United States by combining MODIS
1000 and AmeriFlux data. *Agricultural and Forest Meteorology*, 148, 1827-1847
- 1001 Yang, Q., Fu, Q., & Hu, Y. (2010). Radiative impacts of clouds in the tropical tropopause
1002 layer. *Journal of Geophysical Research: Atmospheres*, 115, D00H12
- 1003 Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed,
1004 B.C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote*
1005 *Sensing of Environment*, 84, 471-475
- 1006 Zwiazek, J., Renault, S., Croser, C., Hansen, J., & Beck, E. (2001). Biochemical and
1007 Biophysical Changes in Relation to Cold Hardiness. In F. Bigras, & S. Colombo (Eds.),
1008 *Conifer Cold Hardiness* (pp. 165-186): Springer Netherlands

1009 **List of Figure Captions**

1010

1011 Figure 1 – Relationship between the start day of a growing season ($SGS_{in-situ}$) as derived from *in-*
1012 *situ* and MODIS (SGS_{MODIS}) VI's (see VI definitions in Table 2) for different PFT's (CRO—
1013 Cropland; DBF—Deciduous Broadleaf Forest; EBF—Evergreen Broadleaf Forest; ENF—
1014 Evergreen Needle-leaf Forest; GRA—Grassland; OSH—Open Shrubland; WSA—Woody
1015 Savanna). Black lines represent linear interpolation functions (for all PFT's pooled), dotted lines
1016 1:1 relationships.

1017

1018 Figure 2 – Relationships between the start day of the growing season, derived from *in-situ*
1019 ($SGS_{in-situ}$) VI's (see VI definitions in Table 2) and NEE (SGS_{NEE}) for the different PFT's
1020 considered in this paper. Black lines represent linear interpolation functions (for all PFT's
1021 pooled), dotted lines 1:1 relationships.

1022

1023 Figure 3 – Relationships between the day of the start of a growing season as derived from
1024 MODIS (SGS_{MODIS}) vegetation indices and net carbon uptake (SGS_{NEE}) for different plant
1025 functional types as in Fig. 2, except that the start of the growing season day is derived from
1026 MODIS vegetation indices (SGS_{MODIS}) instead of *in-situ* vegetation indices.

1027

1028 Figure 4 – Start of growing season (SGS) date estimated by using three methods. NEE is mass
1029 flux data measured at flux sites, $fAPAR$ *in-situ* is $fAPAR$ derived from *in-situ* radiation
1030 measurements and EVI MODIS is EVI derived from MODIS data for (a) CRO—Cropland, (b)
1031 DBF—Deciduous Broadleaf Forest, and (c) ENF—Evergreen Needle-leaf Forest sites locate in

1032 the Northern Hemisphere. Dots represent mean values of SGS date estimates for the three cited
1033 variables variable and lines represent data standard errors.