

A Global Reanalysis of Vegetation Phenology

R. Stöckli,¹ T. Rutishauser,² I. Baker,³ M. A. Liniger,¹ and A. S. Denning³

Abstract. Simulations of the global water and carbon cycle are sensitive to the model representation of vegetation phenology. Current phenology models are empirical and few predict both phenological timing and leaf state. Our previous study demonstrated how satellite data assimilation employing an Ensemble Kalman Filter (EnKF) yields realistic phenological model parameters for several ecosystem types. In this study the data assimilation framework is extended to global scales using a subgrid-scale representation of plant functional types (pft's) and elevation classes. A reanalysis of vegetation phenology for 256 globally-distributed regions is performed using 10 years of MODerate Resolution Imaging Spectroradiometer (MODIS) Fraction of Photosynthetically Active Radiation absorbed by vegetation (FPAR) and Leaf Area Index (LAI) data. The $9 \cdot 10^8$ quality screened observations (corresponding to $< 1\%$ of the globally available MODIS data) successfully constrain a posterior pft-dependent phenological parameter set. It reduces the global FPAR and LAI prediction error to 20.6% and 14.8% respectively compared to the prior prediction error. A 50 year long (1960-2009) daily $1^\circ \times 1^\circ$ global phenology dataset with a mean FPAR and LAI prediction error of 0.065 (-) and 0.34 ($\text{m}^2 \text{m}^{-2}$) is generated. Temperate phenology is best explained by a combination of light and temperature. Tropical evergreen phenology is found to be largely insensitive to moisture and light variations. Boreal phenology can be accurately predicted from local to global scales while temperate and mediterranean landscapes might benefit from a better subgrid-scale pft classification or from a more complex canopy radiative transfer model.

1. Introduction

Land surface vegetation is an interactive part of the climate system. Leaf transpiration influences cloudiness, temperature and moisture patterns of the atmosphere on the synoptic to climatological time scale [Heck *et al.*, 1999; Tsvetinskaya *et al.*, 2001; Lu *et al.*, 2001; Kim and Wang, 2005; Betts and Viterbo, 2005; Betts *et al.*, 2007]. Vegetation biomass acts as a sink (or source) for the atmospheric carbon budget on a seasonal to centennial time scale [Keeling *et al.*, 1996; Kramer *et al.*, 2000; Schaefer *et al.*, 2005; Piao *et al.*, 2007; Körner, 2003]. The two processes regulating water loss and carbon uptake are coupled [Schimel *et al.*, 1997; Sellers *et al.*, 1997] and both depend on leaf stomatal opening and leaf presence. Leaf physiology controls stomates and is largely driven by local scale and short term weather events like the diurnal variability of temperature and radiation [Jarvis, 1976; Law *et al.*, 2002; Larcher, 2003]. Leaf phenology on the other hand describes the timing of leaf appearance, presence and senescence and can be linked to the large scale seasonal to interannual climatic variability [Scheffinger *et al.*, 2002; Menzel *et al.*, 2006; Penuelas *et al.*, 2009; Körner and Basler, 2010].

Leaf physiology and leaf phenology are treated separately in most land surface models (LSM's) used to simulate the

terrestrial water and carbon cycle. While several mechanistic formulations of plant physiological processes have been developed during the last three decades [Jarvis, 1976; Farquhar *et al.*, 1980], highly empirical representations of plant phenology are used in LSM's [Cox, 2001; Foley *et al.*, 1996; Levis and Bonan, 2004; Jolly *et al.*, 2005]. In several LSM's phenology is used as a means to scale leaf level physiological processes to the canopy level [Sellers *et al.*, 1996b, 1997]. Phenology models used in LSM's simulate a continuous biophysical state of vegetation at the landscape scale rather than the timing of species-specific and local-scale events like flowering or bud burst. The latter information is available from long term phenological observations that are mostly confined to temperate climate zones [van Vliet *et al.*, 2003; Rutishauser *et al.*, 2007].

However, the largest phenological model deficiencies are found for subtropical and mediterranean vegetation because model parameters are often generalized from temperate vegetation to global scales [Stöckli *et al.*, 2008b]. Models often simulate a temporal mismatch in spring green up in the order of 1-2 months and show unrealistic drought responses of LAI that have adverse effects on the predicted terrestrial water and carbon fluxes [Kucharik *et al.*, 2006; Randerson *et al.*, 2009]. The ultimate goal to overcome such deficiencies is to further develop LSM's with a mechanistic terrestrial carbon-nitrogen cycling. They allow the coupling of leaf phenology and leaf physiology by use of for instance a prognostic carbon gain-loss formulation [Thornton *et al.*, 2002; Arora and Boer, 2005].

Satellite-based data assimilation can serve as an intermediate step to constrain unrealistic parameters of empirical phenology models and it might be used to augment the realism of terrestrial biosphere models [Demarty *et al.*, 2007; Mahadevan *et al.*, 2008; Rüdiger *et al.*, 2010; Knorr *et al.*, 2010; Rayner, 2010]. In Stöckli *et al.* [2008b] we presented a local-scale data assimilation framework based on the Ensemble Kalman Filter (EnKF) [Evensen, 2003, 2009] that

¹Climate Services, Climate Analysis, Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

²Institute of Geography, University of Bern, Bern, Switzerland

³Department of Atmospheric Science, Colorado State University, Fort Collins, USA

was able to mitigate several phenology model deficiencies by conditioning empirical model parameters with satellite-based phenological observations.

Our local-scale data assimilation framework is however unrealistic for the prediction at the regional scale due to the increase of landscape heterogeneity. A single set of parameters representing a mixed vegetation signal of a specific location cannot be used at another location with a different vegetation composition. A global-scale prediction would hence require a cumbersome parameterization procedure for each grid point. In order to be useful on global scale, our previous framework needs to be extended. The main question is then how to select the bins needed to disaggregate global phenology into a discrete set of functional classes. It was chosen here to split the mixed landscape into a discrete set of plant functional types (pft's) and elevation classes (hgt's) for the following reasons. In earth system models the terrestrial biochemical cycle is often decomposed on the subgrid-scale by using pft's [Sitch *et al.*, 2003; Kucharik *et al.*, 2006; Thornton *et al.*, 2007]. In comparison to biomes pft's group plant species with similar physiological, structural and phenological traits. Satellite remote sensing data can be used to derive pft's globally [Bonan, 2002; Lawrence and Chase, 2007]. However, any satellite-based classification is ultimately constrained by an incomplete set of functional traits [Ustin and Gamon, 2010] that only account for optical vegetation properties. Elevation classes are used since recent findings show that for instance a 100 m elevation difference can shift the leaf-out date by several days [Fisher *et al.*, 2006] which requires a subgrid-scale treatment of the forcing weather data in a global prediction where grid cells can include substantial variability in elevation.

The aim of this study is to create a global MODIS-based reanalysis dataset of vegetation phenology. It should provide a data assimilation and modeling framework to earth system modelers with the capability to assimilate and predict FPAR and LAI of natural vegetation types. We firstly would like to evaluate whether the chosen data assimilation scheme allows to constrain a pft-dependent parameter set with 10 years of assimilated MODIS data. We secondly would like to test whether the chosen phenology model, the pft and hgt classification and the final satellite-constrained parameter set are suited to yield realistic global-scale phenological predictions. In the methods section the prognostic phenology model is presented, followed by a description of the data assimilation system. Global-scale data assimilation experiments are then performed to constrain a pft-dependent phenological parameter set. This parameter set is used to predict global, regional and local FPAR and LAI. A global phenological reanalysis dataset covering 50 years (1960-2009) is finally presented. Analysis of observed and predicted FPAR and LAI followed by a thematic discussion then evaluate the soundness of our method and dataset.

2. Methods

2.1. Phenology Model

The GSI (Growing Season Index) by Jolly *et al.* [2005] diagnoses the state of vegetation by use of three major climatic drivers serving as surrogates for the underlying controls on vegetation phenology: low temperatures, evaporative demand, and photoperiod. Stöckli *et al.* [2008b] and this study extended the GSI model into a prognostic phenology model that predicts the biophysical vegetation states FPAR and LAI.

2.1.1. Theory

The GSI (-) is the product of three environmental factors $f(T)$, $f(L)$ and $1-f(W)$,

$$\text{GSI} = f(T) \cdot f(L) \cdot (1 - f(W)) \quad (1)$$

$$f(x) = \begin{cases} 0 & \text{if } x \leq x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}} & \text{if } x_{\min} < x < x_{\max} \\ 1 & \text{if } x \geq x_{\max} \end{cases}, \quad (2)$$

where $x = \{T, L, W\}$ are multi-day running mean averages of the minimum daily temperature T_m (K), the mean daily global radiation R_g (W m^{-2}) and the mean daily vapor pressure deficit vpd (mb), using averaging times τ_T , τ_L and τ_W (days). T_{\max} , T_{\min} , L_{\max} , L_{\min} , W_{\max} and W_{\min} are maximum and minimum T , L and W , respectively. L can alternatively be driven by photoperiod (daylength) instead of global radiation as suggested by Jolly *et al.* [2005] and scientifically outlined by Körner [2006].

The prognostic phenological state P (-) can be related to the biophysical state FPAR (-) by use of a linear relationship [Sellers *et al.*, 1996a; Los *et al.*, 2000],

$$P = f(\text{FPAR}) \quad (3)$$

where $f(x)$ is given in Equation 2, FPAR_{\min} and FPAR_{\max} are the minimum and maximum FPAR corresponding to the least and most developed state of vegetation. The growth vector $\partial \text{GSI} / \partial t$ (-) then gives the direction and rate of leaf growth or decay used to calculate the change in FPAR with a logistic growth model,

$$\frac{\partial \text{GSI}}{\partial t} = \text{GSI} - P \quad (4)$$

$$\frac{\partial \text{FPAR}}{\partial t} = \gamma \cdot \frac{\partial \text{GSI}}{\partial t} \cdot P(1 - P). \quad (5)$$

As presented by Dickinson *et al.* [2008] growth and senescence can be modeled as two separate processes. We choose a different maximum rate for leaf growth γ_g (day^{-1}) and leaf senescence (γ_d) instead,

$$\gamma = \begin{cases} \gamma_g & \text{if } \partial \text{GSI} \geq 0 \\ \gamma_d & \text{if } \partial \text{GSI} < 0 \end{cases}. \quad (6)$$

According to Sellers *et al.* [1996a] and Los *et al.* [2000] the biophysical state LAI ($\text{m}^2 \text{m}^{-2}$) can be related to FPAR by use of the Monsi-Saeki light interception model based on Beer's law for LAI, respectively [Monsi and Saeki, 2005],

$$\frac{\partial \text{LAI}}{\partial t} = \frac{\partial \text{LAI}}{\partial \text{FPAR}} \frac{\partial \text{FPAR}}{\partial t} \quad (7)$$

$$\text{LAI} = \frac{\ln(1 - \text{FPAR}/f_v)}{\ln(1 - \text{FPAR}_{\text{sat}})} \text{LAI}_{\max} f_v, \quad (8)$$

where f_v (-) is the vegetation fraction and FPAR_{sat} (-) is the FPAR value reached at the maximum Leaf Area Index LAI_{\max} ($\text{m}^2 \text{m}^{-2}$).

2.1.2. Implementation

A semi-implicit numerical scheme is used for the time integration. In comparison to Stöckli *et al.* [2008b] each grid-scale FPAR and LAI prediction is composed of subgrid-scale predictions covering $h=1 \dots n_{\text{hgt}}$ elevation classes (hgt) and $p=1 \dots n_{\text{pft}}$ plant functional type classes (pft). Meteorological forcing is downscaled by hgt. Phenological model parameters are decomposed by pft. The prognostic states are therefore decomposed by both hgt and pft. They can be identified by their superscript time indices t and $t+1$ in the following equations. GSI is diagnosed at every time step,

$$\begin{aligned} \text{GSI}(p, h) = & f(T^{t+1}(p, h)) \cdot f(L^{t+1}(p, h)) \\ & \cdot (1 - f(W^{t+1}(p, h))), \end{aligned} \quad (9)$$

with new prognostic values of $x^{t+1} = \{T, L, W\}$ that depend on their previous values x^t , on the current elevation-dependent weather forcing $y = \{T_m, R_g, vpd\}$ and on the pft-specific time averaging parameters $z = \{\tau_T, \tau_L, \tau_W\}$,

$$x^{t+1}(p, h) = e^{-1/z(p)} x^t(p, h) + (1 - e^{-1/z(p)}) y(h). \quad (10)$$

Leaf growth ΔFPAR depends both on the new phenological state GSI and the previous biophysical state FPAR,

$$P(p, h) = f(\text{FPAR}^t(p, h)) \quad (11)$$

$$\Delta\text{GSI}(p, h) = \text{GSI}(p, h) - P(p, h) \quad (12)$$

$$\Delta\text{FPAR}(p, h) = \gamma(p, h) \cdot \Delta\text{GSI}(p, h) \cdot P(p, h)(1 - P(p, h)) \quad (13)$$

$$\gamma(p, h) = \begin{cases} \gamma_g(p) & \text{if } \Delta\text{GSI}(p, h) \geq 0 \\ \gamma_d(p) & \text{if } \Delta\text{GSI}(p, h) < 0 \end{cases} \quad (14)$$

$$\text{FPAR}^{t+1}(p, h) = \text{FPAR}^t(p, h) + \Delta\text{FPAR}(p, h). \quad (15)$$

Compared to *Stöckli et al.* [2008b] LAI is a diagnostic variable derived from the prognostic state FPAR at each time step,

$$\text{LAI}(p, h) = \frac{\ln(1 - \text{FPAR}^{t+1}(p, h)/f_v(p))}{\ln(1 - \text{FPAR}_{sat}(p))} \text{LAI}_{max}(p) f_v(p). \quad (16)$$

Grid-scale FPAR and LAI are calculated by area weighted summation (a_p and a_h are fractional areas for each pft and hgt class) of the pft- and hgt-specific FPAR and LAI states:

$$\text{FPAR} = \sum_{h=1}^{n_{hgt}} \sum_{p=1}^{n_{pft}} a_h a_p \text{FPAR}(p, h) \quad (17)$$

$$\text{LAI} = \sum_{h=1}^{n_{hgt}} \sum_{p=1}^{n_{pft}} a_h a_p \text{LAI}(p, h). \quad (18)$$

The following numerical constraints are used: $P(1 - P) = \max(P(1 - P), 0.01)$; $f_v = 1.0$ since the vegetation fraction is represented by the fractional areas a_p of each pft; $\text{FPAR}_{sat} = \min(\max(\text{FPAR}_{sat}, 0.001), 0.999)$. Natural logarithms in equation 16 are constrained to be larger than 0.0 and lower than 1.0.

2.2. Data Assimilation Model

Ensemble data assimilation is the key method of this study. It enables to find realistic values as well as their uncertainties for a large set of unknown pft-specific model parameters in the above equations by use of a global set of satellite observations.

2.2.1. Theory

The Ensemble Kalman Filter (EnKF) after *Evensen* [1994, 2003] is applied in this study with modifications for joint state and parameter estimation following *Moradkhani et al.* [2005]; *Evensen* [2009]. The EnKF conditions N prior model states and parameter ensemble members with m observations yielding a posterior model state and parameter ensemble,

$$\mathbf{A}^a = \mathbf{A}^f + \mathbf{K} (\mathbf{D} - \mathbf{H}\mathbf{A}^f), \quad (19)$$

where \mathbf{A}^f is the ensemble matrix containing the prior model states and parameters. They are updated to \mathbf{A}^a when new observations \mathbf{D} become available. \mathbf{H} is the operator relating observed to model states and parameters, $\mathbf{D} - \mathbf{H}\mathbf{A}^f$ is the matrix of innovation and \mathbf{K} is the Kalman gain (for details see *Evensen* [2003]). \mathbf{A} is a matrix holding N ensemble members of the vector ψ with n states x and parameters θ . \mathbf{D} is the matrix holding N ensemble members of the vector d with m observations,

$$\mathbf{A} = (\psi_1, \psi_2, \dots, \psi_N) \in \mathbb{R}^{n \times N}, \quad \psi = (x, \theta) \quad (20)$$

$$\mathbf{D} = (d_1, d_2, \dots, d_N) \in \mathbb{R}^{m \times N} \quad (21)$$

$$\psi_i^0 = (x^0, \theta^0) + \omega_i, \quad i = 1 \dots N, \quad \omega \sim N(0, V_\psi^0) \quad (22)$$

$$d_i = d + \epsilon_i, \quad i = 1 \dots N, \quad \epsilon \sim N(0, V_d). \quad (23)$$

The state and parameter ensemble members ψ_i^0 are perturbed at the beginning of the model integration by use of a Gaussian distribution with mean 0 and initial variance V_ψ^0 . The observation ensemble members d_i are perturbed with mean 0 and with the observation variance V_d at each analysis time step.

2.2.2. Implementation

States and parameters making up the Matrix \mathbf{A} are defined in Table 1 with initial (prior) values similar to those given in *Jolly et al.* [2005] and variances encompassing the orders of magnitude found in the global climate system.

Directly assimilating all global 1 km MODIS FPAR and LAI observations would yield a Matrix \mathbf{D} with dimensions of $O(10^9)$ observations \times $O(10^3)$ ensemble members which is computationally very expensive to solve with the EnKF framework. Therefore, super-observations \hat{d} for each model grid cell are created from observations d_o with $o = 1 \dots n_{obs}$:

$$w_o = 1/V_{d_o} \quad (24)$$

$$\hat{a}_h = \frac{\sum w_o a_{h_o}}{\sum w_o} \quad (25)$$

$$\hat{a}_p = \frac{\sum w_o a_{p_o}}{\sum w_o} \quad (26)$$

$$\hat{V}_d = \frac{\sum w_o^2 V_{d_o}}{\sum w_o^2} \quad (27)$$

$$\hat{d} = \frac{\sum w_o d_o}{\sum w_o} \quad (28)$$

$$\mathbf{D} = \hat{d}_i + \epsilon_i, \quad i = 1 \dots N, \quad \epsilon \sim N(0, \hat{V}_d), \quad (29)$$

where \hat{d} and \hat{V}_d are the grid-scale super-observation and its uncertainty, and \hat{a}_h and \hat{a}_p are the grid-scale fractional hgt and pft areas of the super-observation. By use of the weighting scheme w_o super-observations contain the highest quality satellite data within each grid cell.

The observation operator $\mathbf{H}\mathbf{A}$ is created by linearly aggregating modeled FPAR and LAI weighted by observed elevation distribution \hat{a}_h and pft distribution \hat{a}_p for each super-observation:

$$\mathbf{H}\mathbf{A}'_i = \sum_h \sum_p \hat{a}_h \hat{a}_p (x_i(p, h) - \bar{x}(p, h)), \quad i = 1 \dots N \quad (30)$$

$$\mathbf{H}\mathbf{A}' = \mathbf{H}\mathbf{A}' \frac{\sqrt{\sum_h \sum_p \hat{a}_h \hat{a}_p \text{var}(x(p, h))}}{\text{var}(\mathbf{H}\mathbf{A}')} \quad (31)$$

$$\mathbf{H}\mathbf{A} = \mathbf{H}\mathbf{A}' + \sum_h \sum_p \hat{a}_h \hat{a}_p \bar{x}(p, h), \quad (32)$$

where $x = [FPAR, LAI]$ is predicted by the prognostic phenology model. Ensemble perturbations $\mathbf{H}\mathbf{A}'$ are rescaled with the state variance because the weighed addition of ensemble members by definition deflates the ensemble variance when not all weights are equal.

\mathbf{A}^a is calculated by use of the square root implementation of the EnKF scheme as presented in Evensen [2004] (Section 7.3, equations 69-93) using the low-rank pseudo-inverse calculation because the observation count in our analysis will always exceed the ensemble size. Over-dispersal, over-confidence and non-physical drift of the posterior state and parameter ensemble is taken care of by applying:

$$\mathbf{A}^a = (\mathbf{A}^a - \bar{\mathbf{A}}^a) \min \left(\alpha \frac{\text{var}(\mathbf{A}^f)}{\text{var}(\mathbf{A}^a)}, 1.0 \right) + \bar{\mathbf{A}}^a \quad (33)$$

$$\mathbf{A}^a = (\mathbf{A}^a - \bar{\mathbf{A}}^a) \max \left(\beta \frac{\text{var}(\mathbf{A}^f)}{\text{var}(\mathbf{A}^a)}, 1.0 \right) + \bar{\mathbf{A}}^a \quad (34)$$

$$\mathbf{A}^a = \mathbf{A}^a + (\mathbf{A}_{min} - \bar{\mathbf{A}}^a) \quad \text{when} \quad \bar{\mathbf{A}}^a < \mathbf{A}_{min} \quad (35)$$

$$\mathbf{A}^a = \mathbf{A}^a + (\mathbf{A}_{max} - \bar{\mathbf{A}}^a) \quad \text{when} \quad \bar{\mathbf{A}}^a > \mathbf{A}_{max}, \quad (36)$$

where $\alpha = 1.0$ is the upper limit for the ensemble dispersal, relative to the prior ensemble variance, $\beta = 0.1$ is the lower limit for the ensemble shrinkage, relative to the prior ensemble variance, and \mathbf{A}_{min} and \mathbf{A}_{max} are the lower and upper bounds for the ensemble mean as given in Table 1. It is important to note that the latter physical limits do only move the ensemble mean without modifying the ensemble variance.

\mathbf{A}^a is a global solution that updates all local states and parameters with a single global analysis. This is needed to estimate a single global set of parameters. The presented assimilation scheme can also be used for state estimation. Each local analysis then uses a spatial influence function that updates only states close to the observations. Such a local analysis for state estimation could follow the global analysis for parameter estimation as for instance outlined by equations 80 and 81 in Evensen [2003].

2.3. Data

2.3.1. Meteorological Forcing Data

Daily minimum temperature T_m , daily mean global radiation R_g and daily mean vapor pressure deficit vpd serve as forcing weather data for the prognostic phenology model. $1^\circ \times 1^\circ$ gridded ECMWF ERA 40 [Uppala et al., 2005] are used during 1958-1989 and ERA Interim [Berrisford et al., 2009] are used during 1990-2006. The ensemble members of T_m , R_g and vpd are stochastically perturbed at each grid point and at each time step with a variance of 0.025 K, 1.0 W m^{-2} and 0.005 mb, respectively.

The grid-scale $1^\circ \times 1^\circ$ ERA 40 and ERA Interim data are the starting value for calculating T_m , vpd and R_g for each subgrid-scale elevation class. Subgrid-scale T_m is derived from grid-scale T_m by use of a lapse rate of $-0.6 \text{ K } 100 \text{ m}^{-1}$; subgrid-scale vpd is calculated by keeping the mixing ratio constant with height and applying the subgrid-scale T_m to the vpd calculation. Subgrid-scale R_g increases by $0.3 \text{ W m}^{-2} 100 \text{ m}^{-1}$ (mainly due to decreased atmospheric optical thickness at greater elevation). The local-scale experiments carried out at the four FLUXNET sites are driven by the grid-scale $1^\circ \times 1^\circ$ ERA 40 and ERA Interim weather forcing downscaled to the single elevation class of the respective FLUXNET site.

2.3.2. Satellite Observation Data

TERRA MODIS FPAR and LAI (MOD15A2, Collection 5, *Myeni et al.* [2002]) fill the observation vector d in the assimilation experiments. They are also used as comparison data in the results section. Observations are quality screened and used only if their values are inside the valid range, and if none of the following MOD15A2 quality flag bits are set:

- FparLai bit 2 (dead detectors)
- FparLai bits 3 or 4 (clouds present or unclear)
- FparLai bit 7 (failed retrieval)
- FparExtra bit 0 or 1 (pixel not on land)
- FparExtra bit 2 (snow or ice)
- FparExtra bit 5 (internal cloud mask)
- FparExtra bit 6 (cloud shadow detected)

Observation uncertainty V_d for valid observations is calculated by multiplying the minimum uncertainty with the sum of the “severity factor” s which is then added to the minimum uncertainty. Minimum uncertainty is defined as 0.05 (-) for FPAR and 1.0 ($\text{m}^2 \text{ m}^{-2}$) for LAI.

- $s = 0$
- if FparLai bit 0 (back up algorithm) set: $s = s + 1$
- if FparLai bit 5 (saturated retrieval) set: $s = s + 2$
- if FparLai bit 6 (empirical method used) set: $s = s + 4$
- if FparExtra bit 3 (aerosols present) set: $s = s + 3$
- if FparExtra bit 4 (cirrus clouds detected) set: $s = s + 8$

2.3.3. Elevation Data

The subgrid-scale distribution of elevation classes is derived from the gap-filled CGIAR-CSI SRTM global elevation dataset version 4 [Jarvis et al., 2008], extended to the polar areas with GTOPO30 elevation data [USGS, 1996]. The n_{hgt} elevation classes are equally distributed over two standard deviations of the elevation range in the assimilation area. Elevations below or above the lowest or highest class are counted to the lowest and highest class, respectively. The area fraction a_h for each elevation class is calculated by grid cell.

2.3.4. Plant Functional Type Data

The subgrid-scale distribution of 35 plant functional type classes is derived from MOD12Q1 Collection 4 Land Cover [Friedl et al., 2002], MOD44B Collection 3 Vegetation Continuous Fields [Hansen et al., 2003], AVHRR Tree Cover Continuous Fields [Defries et al., 2000], MOD15A2 Collection 5 [Myeni et al., 2002], global crop data [Leff et al., 2004], global temperature (Version 2.02) and precipitation (Version 2.01) data [Wilmott and Matsuura, 2007], following the method described in Lawrence and Chase [2007] and Bonan et al. [2002]. The resulting pft dataset contains the area fraction a_p for each of the 35 pft's by grid cell. The pft processing is described in Appendix A and a list of pft's is given in Table 2. In this publication only the 15 natural pft's are analyzed even though all 35 pft's were included in the data assimilation.

2.4. Experimental Setup

The data assimilation experiments constrain a set of model parameters. The parameters are then used in global prediction experiments. Figure 1a displays the geographic location of the 256 manually selected regions used for the data assimilation experiments. In order to start where our previous study has ended, the 4 region selection (red squares) includes a temperate, mediterranean, boreal and tropical ecosystem at four FLUXNET sites that are identical to the ones used in [Stöckli et al., 2008b]. Figure 1b then

shows how the 256 region selection finally becomes representative for the full range of climatic conditions needed in a global prediction. The technical details on both the data assimilation and the prediction model are given in Appendix B.

2.4.1. Data Assimilation

The four data assimilation experiments span 4, 16, 64 and 256 regions with $0.5^\circ \times 0.5^\circ$ spatial coverage per region (subsequently labelled as 4, 16, 64 and 256). Each region is subdivided into 25 $0.1^\circ \times 0.1^\circ$ grid cells, where each grid cell has a subgrid-scale representation of 10 hgt classes and 35 pft classes. 1000 ensemble members are integrated in time. Prior model parameters and states are initialized and perturbed as given in Table 1. The phenology model is integrated for 30 years by cycling the 10 year observation period (2000-2009) three times. The 8-day MOD15A2 observations are read at the center of their compositing period (days 4, 12, etc.) since the exact compositing day is not given in the MOD15A2 dataset. In comparison to Stöckli *et al.* [2008b] the EnKF analysis is carried out at the end of each year and not after each observation period in order to avoid convergence to local minima. This modification further results in a less over-confined parameter ensemble and is based on the assumption that a yearly constant set of model parameters simulates the seasonal variation of vegetation states.

2.4.2. Global Prediction

Global predictions are carried out on a $1^\circ \times 1^\circ$ global grid with the prior parameter set (subsequently labelled as “prior”) and with the posterior parameter sets obtained by the above described data assimilation experiments (subsequently labelled as 4, 16, 64 and 256). The integrations employ 10 ensemble members spanning the parameter uncertainty and they are integrated forward in time during 1959-2009. 1959 is used as spin-up year. The prognostic states FPAR and LAI are generally within 1% of their spin-up values after 3 months. A final 50 year long global “reanalysis” dataset covers the period 1960-2009 and uses the 256 region parameter set.

3. Results

The data assimilation framework is used to estimate a new (posterior) set of global phenological parameters that should yield a better prediction of phenological states. In this section the prior and posterior parameter uncertainties are firstly analyzed. Secondly, the effect of the posterior parameter set on the global, regional and local-scale prediction of phenological states is evaluated.

3.1. Global Parameter Estimation

A total number of 510 empirical model parameters were estimated (Table 3 – 5). They can be separated into 6 climate control parameters, 6 structural parameters and 3 time averaging parameters that are estimated for each of the 15 natural pft’s (the water pft and the 19 crop pft parameters are excluded from the analysis).

3.1.1. Climate Control Parameters

The climate control parameters given in Table 3 serve as environmental triggers that primarily determine leaf onset and senescence. The data assimilation is able to reduce the posterior uncertainty to $< 30\%$ of the prior uncertainty (the latter is the square root of the initial parameter variances found in Table 1) for 82% of the temperature, 76% of the moisture and 72% of the light control parameters respectively. Table 3 reveals negative values for the light control parameter L_{min} . This seems unphysical since the meteorological forcing R_g cannot become negative. The employed phenology model is highly empirical without real physical

constraints. The negative L_{min} values thus allow the evergreen needleleaf species to keep needles and therefore maintain LAI during winter when light can be absent especially in boreal regions.

3.1.2. Structural Parameters

Structural parameters are given in Table 4. They determine the upper and lower bounds of the leaf state. For tropical evergreen broadleaf forests, the $FPAR_{max}$ (the upper bound of FPAR) is better constrained than $FPAR_{min}$ since a total absence of leaves can hardly ever be observed. The lowest FPAR values can further be contaminated by clouds and aerosols and get a larger observation error by the employed observation quality screening method. The EnKF then creates larger posterior parameter uncertainties when few observations concur with larger observation errors. The EnKF estimates each parameter independently from the others. The bare soil $FPAR_{max}$ (0.05) for instance has a lower posterior estimate than the $FPAR_{min}$ (0.11). The prediction of bare soil FPAR will therefore have no seasonal cycle. The predicted FPAR will remain at $FPAR_{max}$. The posterior uncertainty of $FPAR_{max}$ and LAI_{sat} is well below 5% of the prior parameter uncertainty for several boreal and temperate pft’s such as temperate deciduous broadleaf forest, boreal evergreen needleleaf forests or not-arctic grasslands. Leaf growth rate γ_g values are higher than the leaf decay rate γ_d for most pft’s. Leaf onset is a faster process than leaf senescence for most natural species. Posterior uncertainties of both γ_g and γ_d are between 10-20% of their initial uncertainty.

3.1.3. Time Averaging Parameters

Time averaging parameters are given in Table 5. The posterior uncertainties for the time averaging parameters is in the range of 20-50% compared to the prior uncertainty. Jolly *et al.* [2005] use 21 days for these parameters, which was our initial (prior) value. For the temperate deciduous broadleaf forest pft the averaging times needed for temperature and light decreases from 21 to 5.5 days and from 21 to 15.4 days respectively while the time averaging needed for moisture increases from 21 to 33.7 days.

3.2. Global Prediction

A 50 year long FPAR and LAI reanalysis dataset is generated by running the prognostic phenology model with the 256 region parameter set over the whole ERA Interim and ERA 40 period (1960-2009). The prior and posterior global FPAR and LAI prediction uncertainties and errors are analyzed in this sub-section. The prediction uncertainty is caused by the model’s parameter uncertainty. It can be calculated as the ensemble variance of the predicted FPAR and LAI. The prediction error on the other hand is defined as the mean absolute deviation (MAD) between the ensemble mean of predicted FPAR and LAI and the quality screened observations (2000-2009).

Figure 2a summarizes the mean leaf state and its seasonal variability for the 50 year long reanalysis dataset. It shows that highest annual mean LAI of above $5 \text{ m}^2 \text{ m}^{-2}$ is found in tropical climates and the largest seasonal LAI amplitude (contour lines in Figure 2a) occurs in subtropical and temperate climate zones. Figure 2b) shows that northern hemisphere temperate climate zones green up during April and May while northern hemisphere Boreal and Arctic climate zones green up during May and June.

Figure 3 visualizes that the highest prediction errors with the prior parameter set occur in sub-tropical, mediterranean and temperate areas. The prior mean absolute deviation (MAD) of predicted versus observed LAI is in the order of $3.0 \text{ m}^2 \text{ m}^{-2}$ for these regions. The MAD of LAI decreases to below $1.5 \text{ m}^2 \text{ m}^{-2}$ in the 4 region experiment. The largest improvements between the 4 and the 256 region experiments

are confined to crop-intensive areas such as India, Central USA and Europe, but also semi-arid areas such as the Sahel in Africa and Central Australia.

Figure 4 displays the evolution of the global mean prediction error and prediction uncertainty with the increasing number of assimilated observations. The global LAI and FPAR prediction error (solid lines) successively decreases with the increasing number of assimilated observations. More than 50% of the prior prediction error is removed by the 4 region experiment, while the 256 region experiment further reduces the global FPAR and LAI prediction error to 20.6% and 14.8% of the prior prediction error, respectively. The prediction uncertainties (dashed lines in Figure 4 and Table 6) decrease to 3.4% and 3.9% of their prior values (0.326 and $2.79 \text{ m}^2 \text{ m}^{-2}$) for FPAR and LAI. Already the 4 region experiment covering only 0.007% of the global land area reduces FPAR and LAI uncertainty to 17.8% and 16.1% of their prior uncertainty.

Figure 4 also reveals that the prediction error decreases less rapidly than the prediction uncertainty (solid versus dashed lines). The model ensemble members converge slightly faster than expected from the remaining model-observation differences. Parameter and state covariance underestimation is a common feature in ensemble data assimilation [Li *et al.*, 2009]. In our experiments it happens despite the employed ensemble inflation (equation 36) and despite the large number of chosen ensemble members.

3.3. Regional Prediction

The global land area is screened by pft class, where only grid points with at least 25% coverage for a given pft are included in each respective area. In Figure 5 Taylor diagrams [Taylor, 2001] document the statistical performance of FPAR and LAI predictions by simultaneously drawing the correlation coefficient R between the model and the observations and the normalized standard deviation (the standard deviation of the model divided by the standard deviation of the observation). Table 7 provides the mean bias (bias) and root mean square error (rmse) values for each pft.

The temperate deciduous broadleaf forest (pft class 8) FPAR and LAI predictions in the 256 region experiment have a high $R > 0.98$ but slightly underestimate phenological variability when compared to observations (normalized standard deviation < 1). Bias and rmse for both FPAR and LAI substantially decrease (rmse to around 30% of its prior value and the bias in LAI from $> 1 \text{ m}^2 \text{ m}^{-2}$ to $-0.09 \text{ m}^2 \text{ m}^{-2}$, Table 7). For this pft it would be interesting to evaluate the inter-annual variability of the spring leaf-out date. However, while the modeled FPAR and LAI output is daily, the spring date cannot be accurately diagnosed from the MODIS observations since they have a 8 day compositing period (and the actual compositing day is not given in the data). The boreal evergreen needleleaf forest (pft class 3) LAI in the 256 region experiment reaches a very high prediction accuracy with $R = 0.97$, a bias of $0.03 \text{ m}^2 \text{ m}^{-2}$ and a rmse value which is $< 15\%$ of its prior value. A similar gain in accuracy is achieved for the boreal deciduous needleleaf forest (pft class 4) and for the arctic c3 grass (pft class 13).

Figure 5 demonstrates that the prediction using the prior parameters generally overestimates phenological variability for most pft's (normalized standard deviation > 1) and the highest R values are at 0.9. In the 256 region experiment most pft's cluster in the same "high prediction accuracy" area and the highest R values reach 0.99. The 256 region experiment slightly underestimates phenological variability for

most pft's, which might be a result of residual observation noise (and thus exaggerated observation variability) despite of the employed restrictive quality screening. The correlation coefficient R of evergreen species such as the tropical and temperate evergreen broadleaf forest (pft classes 5 and 6) or the evergreen broadleaf shrub (pft class 10) remains low for both FPAR and LAI in the 256 region experiment. However, their bias and rmse values substantially improve. Correlation is not a suitable statistical means in the case of time series with almost constant (evergreen) values. The bias and rmse values given in Table 7 clearly demonstrate that both magnitude (bias) and phase (rmse) significantly gain in realism. Bold values document that for all pft's the FPAR and LAI biases fall below 5% of the FPAR and LAI range, while most rmse values reach this threshold in the 256 region experiment.

3.4. Local Prediction

Any scientific application that is applied to global scales should be re-evaluated at the local scale if possible in order to gain a better process-based understanding and reveal missing model components (as an example see Stöckli *et al.* [2008a]; Oleson *et al.* [2008]). The phenology model using the global parameter set has therefore been tested at the same four FLUXNET tower sites as in our local-scale data assimilation study [Stöckli *et al.*, 2008b]. The aim of this section is to evaluate to what degree the model using the above estimated global parameter set is still able to represent local-scale phenology at specific sites.

3.4.1. Morgan Monroe State Forest

The Morgan Monroe State Forest site (USA) is a temperate deciduous forest interleaved by grassland and crops. The site-level simulation and the 4 region experiment simulate a realistic seasonal cycle (Figure 6a). The 4 region experiment should always be closer to the site-level experiment than for instance the 256 region experiment since the former uses parameters that are constrained over exactly the four regions covering the four sites, where the latter uses parameters that minimize the prediction error for a global area. A two-stage green-up successively appears in the 16, 64 and 256 region experiments. This two-stage green-up is likely due to a unrealistic green-up timing of non-natural pft's present in this grid-cell. The pft parameters for maize (14% of the area) and soy (12% of the area) are constrained with information from globally distributed croplands by the 16, 64 and 256 region experiments, but their values do not seem to be valid at this particular site or for this particular year. The employed static pft map would firstly not be suitable in areas where crop rotation is practiced, and secondly a crop phenology model might be required to realistically simulate the phenological stages of different crops in a global prediction. Senescence is realistic in the 4, 16, 64 and 256 region experiments but is delayed in the site-level experiment. The 64 and 256 region experiments further reveal a underestimation of summer LAI magnitude. It might be related to the negative bias of the temperate deciduous broadleaf forest LAI prediction found in the regional analysis above (Table 7).

3.4.2. BOREAS Old Black Spruce

The high prediction skill at the boreal forest site BOREAS Old Black Spruce (Canada) appears to be independent of whether a site-level or global parameter set is used (Figure 6b). This result firstly demonstrates that the regions where boreal evergreen needleleaf forest occur are spatially more homogeneous than for instance the patchy landscapes encountered in temperate climate zones. The pft distribution at BOREAS for instance consists of around 50% evergreen needleleaf trees, 20% deciduous shrubs and 20% arctic grasslands. Secondly, phenological timing for this pft is controlled by a well defined set of environmental triggers (defined by the climate control parameters in Table 3)

that are valid from local to global scales. This result is underlined by the high prediction performance of the boreal evergreen needleleaf forest pft found in the regional analysis above.

3.4.3. Santarem KM83

The prior parameter set at the tropical evergreen broadleaf site Santarem KM83 (Brazil) creates a unrealistic light-limited leaf loss of around $2 \text{ m}^2 \text{ m}^{-2}$ at the end of the wet season (April - June) while both quality screened observations and all posterior parameter sets show a constant LAI throughout the year. Figure 6c demonstrates that the employed observation quality control is working well and that cloud affected (wet season) and aerosol contaminated (dry season) observations at the site are properly screened and do not affect the data assimilation process.

3.4.4. Tonzi Ranch

Although the Tonzi site (USA) has the same mean monthly precipitation and mean temperature as Morgan Monroe State Forest (the two red boxes that coincide in the center of Figure 1b), it is a mediterranean savanna-type ecosystem with a rather dry late summer and a wet winter season. Figure 6d shows that the magnitude but also the timing of the drought response between May and September (see also Figures 3 and 4 in *Stöckli et al.* [2008b]) are simulated very realistically by the site-level and the 4 region experiments compared to the prior experiment. The timing is still accurate in the 16, 64 and 256 region experiments, but the peak LAI during April and May is severely underestimated in the 16, 64 and 256 region experiments. The result demonstrates that global parameter sets can become inaccurate at the local scale for ecosystems with a complex canopy. The site-level experiment yields July/August LAI values that are comparable to ground measurements [*Ryu et al.*, 2010b]. However, our simplified canopy radiative transfer neglects the contribution of vegetation structural aspects like leaf clumping while ground measurements often neglect the contribution of the understory LAI that is also measured by the satellite. Currently the comparison of satellite- and ground-observed phenology is best achieved through the analysis of phenological timing [*Studer et al.*, 2007; *Stöckli et al.*, 2008b; *Liang et al.*, 2011]. Newly developed near-surface remote sensing methods are promising to also compare phenological magnitude [*Ahrends et al.*, 2008; *Richardson et al.*, 2009; *Ryu et al.*, 2010a].

4. Discussion

4.1. Data Assimilation

By running the data assimilation over less than 1% of the global land surface the global FPAR and LAI prediction error could be reduced to below 20% of its initial value. The key for this success is most likely the wide climatic and biogeographic range spanned by the chosen subset of assimilation regions (Figure 1). The 4 region experiment already includes a tropical, a temperate, a boreal and a mediterranean climatic environment to constrain a set of parameters that then show substantial skill in a global prediction (Figure 3). Figure 4 suggests that little improvement can be expected when extending the assimilation area beyond the 0.4% of global land area covered by the 256 region experiment. Research on the optimal location of assimilation regions might further reduce the computational resources needed for a global data assimilation of vegetation phenology.

The parameter uncertainties seem to converge much faster than the prediction errors. While the EnKF allows in theory a perfect estimation of the combined posterior model and parameter error by analysis of both the prior model uncertainty and the observation uncertainty, there are many assumptions to be made for the practical implementation of the EnKF in a prediction system. Each of the following assumptions could be the cause for the observed parameter overconfidence:

- **Ensemble size:** the ensemble size N for the EnKF should be as large as possible since the sampling error decreases by $1/\sqrt{N}$. 1000 ensemble members are likely too low since around 10000 states and 510 parameters are estimated for each region. With the available computational resources for this project there is little that can be done regarding ensemble size.

- **Measurement size:** if the measurement size exceeds the ensemble size, rank problems can occur because the measurement error covariance needs to be compressed into the ensemble space. We however make use of the inversion presented in *Evensen* [2004] that uses a measurement operator covering the full-rank of measurements to avoid the problem of rank-loss reported in the literature [*Kepert*, 2004].

- **Measurement uncertainty:** if the measurement uncertainty is poorly chosen in a bayesian method, the posterior model uncertainty will likely be wrong. The measurement uncertainty is derived from MODIS quality flags that are themselves based on semi-empirical detection algorithms for clouds, shadows, aerosols and reflect an incomplete set of retrieval errors [*Justice et al.*, 2002]. Further, arbitrary scaling factors are used to transfer the quality flags into a quantitative set of observation uncertainties. The super-observations derived in equation 29 neglects any spatially correlated measurement errors that are likely to happen with cloud contamination or snow cover.

- **Local minima:** the EnKF solver is chosen to avoid local minima since the full non-linear prediction model is integrated without the need to create first order derivatives as needed for instance in the Extended Kalman Filter or in variational data assimilation techniques like 3D or 4D VAR. A yearly analysis guarantees that parameters do not satisfy individual observations but are consistent with the entire seasonal cycle of the leaf state.

- **Equifinality:** by estimating a set of 15 parameters for a total of 34 pft's several solutions in the parameter space might produce a similar prediction. However, even though equifinality might generate wrong parameters it should to the best of our knowledge not lead to parameter overconfidence.

4.2. Phenology Model

We chose a rather empirical phenology model with a large set of climate control parameters, structural vegetation parameters and time averaging parameters.

It was demonstrated how both climate control and structural vegetation parameters can be thoroughly constrained by the 10 years of MODIS data while time averaging parameters are left with a substantial posterior uncertainty. There is nevertheless evidence that the time averaging needed for temperature and light are likely shorter than 21 days and the averaging time for moisture is higher than 21 days. This result contrasts most temperature-based phenology models that work with growing degree days since they often integrate temperature history over several months [*Chuine*, 2000]. The long averaging time for moisture further demonstrate that tall trees in temperate climate zones can sustain greenness for prolonged periods of droughts. For short natural vegetation like grasslands and deciduous shrubs the moisture averaging times result well below 21 days, most likely related to their short rooting depths and higher susceptibility to drought.

The tropical evergreen broadleaf forest pft requires the longest moisture averaging times. Our model yields a seasonally largely constant FPAR and LAI for this pft. Tropical trees are known to be resistant to the yearly recurring dry periods [Lee *et al.*, 2005]. Recent studies however demonstrate that tropical plant physiology and phenology is very complex and both can sensitive to extreme drought periods [Saleska *et al.*, 2007; Myneni *et al.*, 2007; Phillips *et al.*, 2009; Zhao and Running, 2010]. These studies are based on satellite-based EVI, modeled GPP or on field measurements of biochemical fluxes and carbon stocks and not on satellite-based FPAR or LAI. Further, drought-induced changes in tropical phenology may not be detectable with spectro-radiometers like MODIS but only with hyperspectral radiometers like Hyperion [Asner *et al.*, 2004]. These open questions should motivate follow-up research in both modeling and observation of tropical phenology.

Our model entirely depends on a multiplicative set of linearized and time integrated temperature, light and moisture controls. The model therefore excludes several known biophysical and abiotic controls such as chilling requirements, insect pests, harvest, irrigation, nutrient limitations, tree aging, biodiversity effects or frost events. The high prediction skill on the seasonal and inter-annual time scale spanning local to global spatial scales demonstrates that the main drivers of phenological variability have been included in the model. However, the short observation period of 10 years by definition excludes most climatological extreme events required to exploit the full range seasonal to decadal phenological variability. Especially the climate control parameters of sub-tropical and tropical drought-deciduous pft's might benefit from a longer observation period.

Plant physiological research suggests that bud burst of temperate deciduous species is driven by photoperiod (but not necessarily the light intensity R_g as used in this study). Photoperiod can serve as trigger for temperature sensitivity [Körner, 2006]. Our results demonstrate that the best empirical prediction of temperate deciduous broadleaf forest phenology is simulated by a combined temperature-light forcing. Figure 4a) of Stöckli *et al.* [2008b] visualizes that a light trigger (green curve, X symbols) precedes the temperature trigger (red curve, star symbols). However, it is currently debated whether light, temperature or both control bud burst. These relationships also vary by species [Körner and Basler, 2010] and cannot be generalized [Cleland *et al.*, 2007].

4.3. Plant Functional Type Data

Plant functional types [Bonan *et al.*, 2002] are chosen instead of the often used biomes or land cover classes [Hansen *et al.*, 2000] because they are better in line with the separation needed for phenological predictions. The single savanna biome at for instance the Tonzi Ranch is composed of a evergreen broadleaf tree pft (with maximum LAI in late summer) and of a drought-deciduous c3 grass pft (with maximum LAI in early spring). Both pft's display a very different phenological cycle and there is no single parameter set that would enable a realistic simulation of the single savanna biome. The regional analysis however suggests that several pft classes like the temperate deciduous broadleaf forest pft, the evergreen broadleaf shrub pft or the temperate deciduous broadleaf shrub pft might still be too heterogeneous in terms of phenological behavior and could be separated into sub-pft classes. Phenological predictions would surely benefit from consistent global pft maps based on new remote sensing technologies as for instance presented in Ustin and Gamon [2010].

The focus of this study is the estimation of phenology parameters for natural vegetation. However, crop pft's were

also included in the data assimilation. Satellite pixels contain a mixed signal from both natural and managed vegetation that needs to be decomposed in order to estimate parameters for the natural vegetation pft's. Figure 3 demonstrates that the FPAR and LAI of regions with heavy crop cover are well predicted without the explicit use of a crop phenology model. This shows that even managed vegetation phenology is dominantly weather and climate driven. However, for climate model applications a dedicated crop phenology model should be used since especially the carbon uptake of crops differs from natural vegetation [Gervois *et al.*, 2004; Lokupitiya *et al.*, 2009].

4.4. Satellite Data

The MODIS FPAR and LAI data are derived from MODIS surface reflectances by inversion of a canopy radiative transfer model [Myneni *et al.*, 1999, 2002]. They are more accurate in low biomass areas and generally exaggerate LAI for broadleaf and needleleaf forests [Wang *et al.*, 2004; Cohen *et al.*, 2006]. The LAI retrieval from visible and near-infrared surface reflectances is underdetermined for intermediate and high LAI values which can yield errors in the order of 50% [Garrigues *et al.*, 2008]. The FPAR and LAI dataset presented in this study will inherit such errors. We further use a very simplified representation of the canopy light interception that only fits 4 canopy structural parameters per pft (FPAR_{min}, FPAR_{max}, FPAR_{sat} and LAI_{max}). Compared to the MODIS retrieval algorithm it does not include the effects of foliage and canopy clumping, non-green canopy elements, soil background reflectance, shading or vertical canopy structure [Myneni *et al.*, 1999; Shabanov *et al.*, 2003]. These differences can introduce inconsistencies between the assimilated and predicted FPAR and LAI values. They might be responsible for some of the scaling issues found at Morgan Monroe and Tonzi Ranch.

The restrictive quality screening of MODIS observations employed in this study eliminates the majority of cloud, aerosol, snow and cloud shadow contamination that usually complicates the generation of climate quality biophysical satellite parameters in tropical or high latitude [Los *et al.*, 2000; Poulter and Cramer, 2009]. On the global average 40-50% of all valid observations pass quality screening (Table 6). In tropical areas only 5-10% (not shown) pass the quality screening. Neglecting quality screening can for instance lead to misleading conclusions on the drought-response of tropical trees [Saleska *et al.*, 2007] as shown by Samanta *et al.* [2010].

Remote sensing data assimilation in combination with a predictive model has the capability to complement the classical data-only gap filling procedures such as maximum value compositing or fourier time series fitting employed in most current satellite-based land surface datasets [Los *et al.*, 2000; Jonsson and Eklundh, 2002; Stöckli and Vidale, 2004; Tucker *et al.*, 2005; Fang *et al.*, 2008].

4.5. Weather Forcing Data

The model parameter set and therefore the phenological prediction will be sensitive to the choice of weather forcing data since predicted states are empirically and not mechanistically linked to the meteorological predictors. Potential biases in the ECMWF ERA Interim data might therefore have created unrealistic posterior parameter sets during the data assimilation. We have perturbed the weather forcing data with uncertainties as given in the methods section, but the perturbation does not correct for biases in the weather forcing data. Also, a new estimation of model parameters might be required if a new weather forcing data with a different spatial scale or with a different climatology is used or if the phenology model is applied in coupled mode as part of a climate model.

5. Conclusions and Outlook

Our study demonstrates how remote sensing data assimilation can be used to reduce uncertainties in a global phenology model. The assimilation of MODIS data covering less than 1% of the global land surface successfully reduced the global FPAR and LAI prediction errors to 20.6% and 14.8% of their respective prior errors. A too high variance reduction in the posterior parameter set could be mitigated by use of a more quantitative observation uncertainty estimation. Novel data assimilation methods such as the Maximum Likelihood Ensemble Filter MLEF [Zupanski, 2005] employing Hessian preconditioning and a gradient search method might yield more realistic globally applicable parameter sets.

Our study suggests that pft's are a suitable means to disaggregate mixed satellite pixels on global scale and they allow to create a pft-specific parameterization of a globally-applicable phenology model. The boreal evergreen needle-leaf forest pft and the tropical evergreen broadleaf forest pft perform realistically over a large range of spatial scales. However, local-scale predictions using a global parameter set can become unreliable in both magnitude and timing as for instance demonstrated for the mixed natural-agricultural temperate landscape (Morgan Monroe) and the savanna landscape (Tonzi Ranch). The phenological data assimilation experiment could now be repeated with a variety of globally-applicable phenology models and pft datasets. In order to increase the compatibility between assimilated and predicted vegetation states the MODIS canopy radiative transfer model could be employed in the prediction of FPAR and LAI. A more complex treatment of leaf and canopy clumping, leaf orientation, shadowing or non-green canopy elements would further broaden the applicability of our methods and datasets. As a first step global maps of foliage clumping [Chen et al., 2005; Pisek et al., 2010] could enhance our simplified LAI calculation with geometric information on canopy structure.

Our study is a first step to mitigate some deficiencies of current phenological models. As already shown in Stöckli et al. [2008b] the parameterized phenology model can be useful to disentangle the influence of meteorological drivers on the observed phenological variability. It could be a contribution to the currently ongoing discussion on how temperature and light (or photoperiod) govern the timing of phenological spring events [Körner and Basler, 2010]. The 50 year long global phenological reanalysis dataset (1960-2009) should be suitable for climate analysis studies. It might for instance contain evidence on whether the light trigger is the hard limit for the currently observed (temperature-related) negative trends for phenological spring events [Cleland et al., 2007; Rutishauser et al., 2007].

Future research should combine process-based knowledge from hydrology, plant physiology and canopy radiative transfer modeling with the highly empirical world of plant phenology. This is needed to better understand and simulate the response of the terrestrial water and carbon cycle to climate variability and change and to quantify the resulting impacts on the other earth system components [Penuelas et al., 2009]. We would therefore like to motivate earth system modelers to experiment with data assimilation and to bring forward a new generation of phenology and land surface models. In order to facilitate this, the presented dataset, all program codes, parameters, documentation and simple hands-on experiments are publicly available on <http://phenoanalysis.sourceforge.net>.

Appendix A: Plant Functional Type Data Generation

The following modifications are made to the pft processing by Lawrence and Chase [2007] and Bonan et al. [2002]:

- The single crop class is decomposed into 19 individual crop classes according to Leff et al. [2004].
- This yields 35 pft classes in total: 15 natural types, 19 crop classes and water.
- The processing is performed at 30" spatial resolution instead of 0.05°.
- The monthly temperature climatology [Wilmott and Matsuura, 2007] is downscaled to 30" by use of a lapse rate of 0.5 K 100 m⁻¹ applied to the above described topography dataset.
- MOD15A2 LAI is quality screened as described above in order to evaluate the C4 grass fraction. Following Still et al. [2003] the C4 grass fraction is the sum of LAI for those months that satisfy the C4 growth criteria (temperature > 22 °C and precipitation > 25 mm) over the sum of LAI for all months. Since they have used NDVI instead of LAI, we apply the square-root to the LAI-derived C4 grass fraction in order to account for the almost exponential relationship between NDVI and LAI.
- The processing merges 7 sometimes inconsistent datasets into a single continuous plant functional type cover dataset. The inconsistencies (e.g. MOD44B indicates 25% tree cover but the AVHRR VCF shows 0% tree cover) are overcome by inverse distance filling where the MODIS dataset served as the reference dataset.

Appendix B: Technical Set Up

The data assimilation framework is parallelized by using Version 1.2 of the MPI standard [Message P Forum, 1994] with a 1- or 2-dimensional process topology (multiple regions and one region per process, or single region distributed along longitude and latitude range). Model state prediction, I/O, observation QA screening, gridding of super-observations, **HA** and **D** matrices are calculated on separate processes by assigning one region per process and one process by logical CPU unit. The prior parameters are perturbed once and distributed to all processes in order to end up with a global analysis parameter set. Model states and weather forcing are perturbed by process. One process is reserved for the global analysis, where all regional **HA** and **D** are collected at the end of each simulation year and the global analysis is performed. The global analysis matrix (X5 in Evensen [2003]) is finally re-distributed to all processes, where the computationally intensive final ensemble update of states and parameters is performed.

The bottlenecks for this framework are its heavy memory usage, the size of the observational data and the global EnKF analysis. The parallelization of the EnKF solver would be an important next step in order to increase data assimilation performance. The state matrix has 7 dimensions (ens x lon x lat x pft x hgt x state x days), the parameter matrix has 3 dimensions (ens x pft x parameter), the forcing data has 5 dimensions (ens x lon x lat x hgt x forcing), which exceeds per-process memory availability on today's supercomputers. In order to increase memory efficiency, a subset of hgt and pft classes for states is integrated in each region, where only hgt and pft classes are selected that cover more than 2.5% of the area in each region. Water areas (pft number 35) are screened and not used during the analysis. Furthermore the upper bound of super-observations to be used in the global analysis was set to 50000. The analysis then is within around 1 GB per process (with a maximum of 4-8 GB per node on e.g. NCCS Discover with 8 cpu's per node and 16-32 GB per node on e.g. NCAR Bluefire with 32 CPU's per node).

Acknowledgments. The NASA Energy and Water Cycle Study (NEWS) grant No. NNG06CG42G and MeteoSwiss provided funding for this study. The MODIS Science Team and the MODIS Science Data Support Team provided the MOD15A2, MOD44B and MOD12Q1 data. The NASA Science Mission Directorate (SMD) is acknowledged for granting the SMD-08-0810 and SMD-09-1256 requests with 250,000 and 460,000 cpu hours, respectively, on the NASA Center for Computational Sciences (NCCS) High-End Computing (HEC) Discover system. The National Center for Atmospheric Research (NCAR) Terrestrial Sciences Section (TSS) is acknowledged for providing computational resources on the Bluefire system for testing purposes. The Swiss Center for Scientific Computing (CSCS) is acknowledged for providing computational resources on the Rigi system for testing purposes.

References

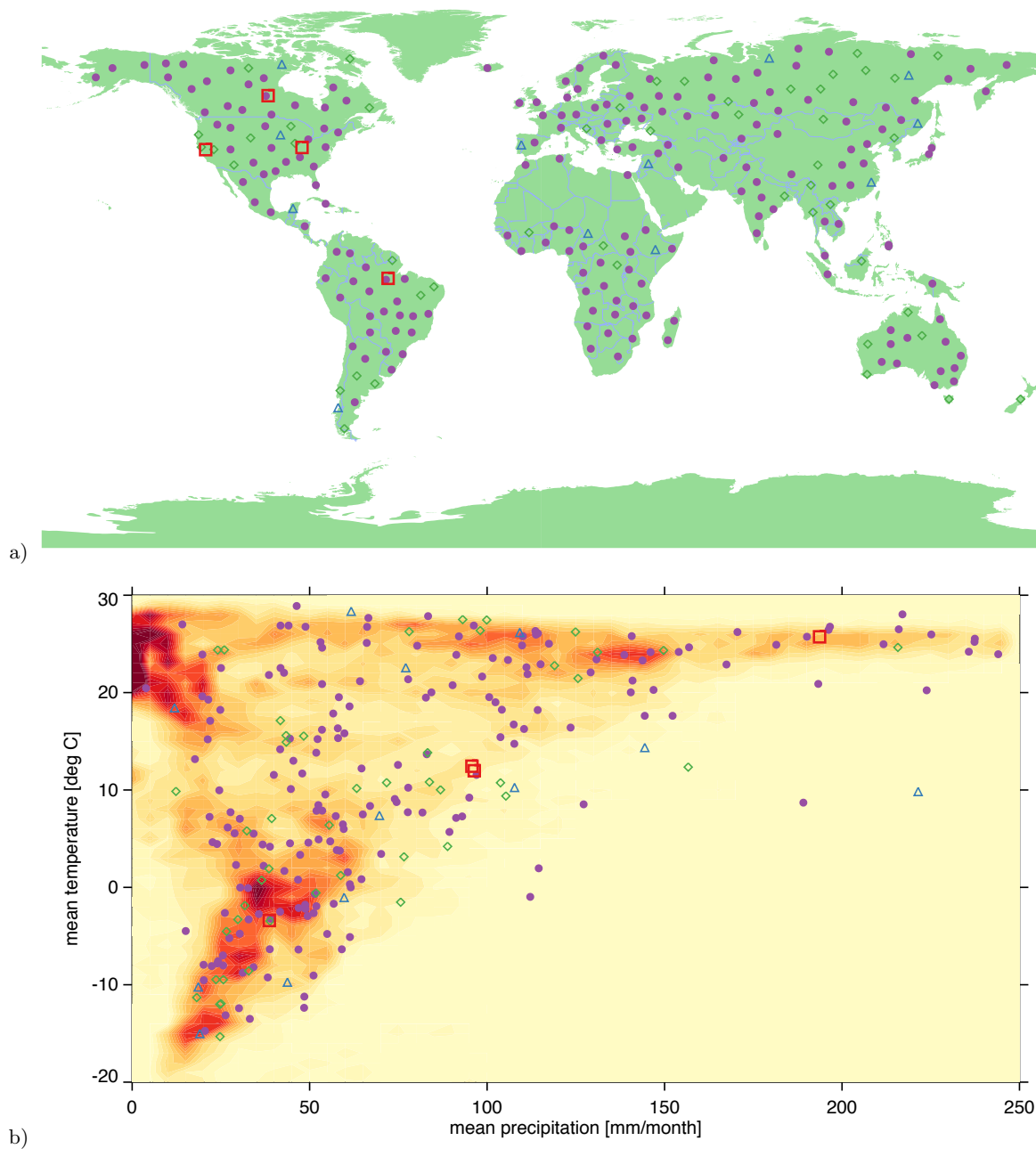
- Ahrends, H. E., R. Bruegger, R. Stöckli, J. Schenk, P. Michna, F. Jeanneret, H. Wanner, and W. Eugster (2008), Quantitative phenological observations of a mixed beech forest in northern Switzerland with digital photography, *Journal of Geophysical Research-Biogeosciences*, *113*(G4), doi:10.1029/2007JG000650.
- Arora, V. K., and G. J. Boer (2005), A parameterization of leaf phenology for the terrestrial ecosystem component of climate models, *Global Change Biol.*, *11*(1), 39–59, doi:10.1111/j.1365-2486.2004.00890.x.
- Asner, G., D. Nepstad, G. Cardinot, and D. Ray (2004), Drought stress and carbon uptake in an Amazon forest measured with spaceborne imaging spectroscopy, *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 6039–6044, doi:10.1073/pnas.0400168101.
- Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala (2009), The era-interim archive version 1.0, *ERA Report Series 1*, ECMWF, European Centre for Medium Range Weather Forecasts Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.
- Betts, A. K., and P. Viterbo (2005), Land-surface, boundary layer, and cloud-field coupling over the southwestern Amazon in era-40, *J. Geophysical Research-atmospheres*, *110*(D14), doi:10.1029/2004JD005702.
- Betts, A. K., R. L. Desjardins, and D. Worth (2007), Impact of agriculture, forest and cloud feedback on the surface energy budget in boreas, *Agr. Forest Meteorology*, *142*(2–4), 156–169.
- Bonan, G. B. (2002), *Ecological Climatology*, Cambridge University Press, Cambridge.
- Bonan, G. B., S. Levis, L. Kergoat, and K. W. Oleson (2002), Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, *Global Biogeochemical Cycles*, *16*(2), doi:10.1029/2000GB001360.
- Chen, J., C. Menges, and S. Leblanc (2005), Global mapping of foliage clumping index using multi-angular satellite data, *Remote Sensing of Environment*, *97*(4), 447–457, doi:10.1016/j.rse.2005.05.003.
- Chuine, I. (2000), A unified model for budburst of trees, *J. Theoretical Biol.*, *207*, 337–347.
- Cleland, E. E., I. Chuine, A. Menzel, H. A. Mooney, and M. D. Schwartz (2007), Shifting plant phenology in response to global change, *Trends Ecol Evol*, *22*(7), 357–365, doi:10.1016/j.tree.2007.04.003.
- Cohen, W. B., et al. (2006), Modis land cover and LAI collection 4 product quality across nine sites in the western hemisphere, *Ieee Transactions On Geoscience Remote Sensing*, *44*(7), 1843–1857.
- Cox, P. M. (2001), Description of the TRIFFID dynamic global vegetation model, *Tech. Rep. 24*, Hadley Center, Bracknell, U.K.
- Defries, R. S., M. C. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland (2000), A new global 1-km dataset of percentage tree cover derived from remote sensing, *Global Change Biol.*, *6*(2), 247–254.
- Demarty, J., F. Chevallier, A. D. Friend, N. Viovy, S. Piao, and P. Ciais (2007), Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model, *Geophysical Research Lett.*, *34*(15), doi:10.1029/2007GL030014.
- Dickinson, R. E., Y. Tian, Q. Liu, and L. Zhou (2008), Dynamics of leaf area for climate and weather models, *J. Geophys. Res.-Atmos.*, *113*(D16115), doi:10.1029/2007JD008934.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics, *J. Geophysical Research-oceans*, *99*(C5), 10,143–10,162.
- Evensen, G. (2003), The ensemble kalman filter: theoretical formulation and practical implementation, *Ocean Dynamics*, *53*, 343–367.
- Evensen, G. (2004), Sampling strategies and square root analysis schemes for the enkf, *Ocean Dynamics*, *54*(6), 539–560.
- Evensen, G. (2009), The ensemble kalman filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, *29*, 83–104.
- Fang, H. L., S. L. Liang, J. R. Townshend, and R. E. Dickinson (2008), Spatially and temporally continuous LAI data sets based on an integrated filtering method: Examples from North America, *Remote Sensing Environment*, *112*, 75–93.
- Farquhar, G. D., S. V. Caemmerer, and J. A. Berry (1980), A biochemical-model of photosynthetic CO₂ assimilation in leaves of C-3 species, *Planta*, *149*, 78–90.
- Fisher, J. I., J. F. Mustard, and M. A. Vadeboncoeur (2006), Green leaf phenology at Landsat resolution: Scaling from the field to the satellite, *Remote Sensing Environment*, *100*(2), 265–279.
- Foley, J. A., I. C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine (1996), An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics, *Global Biogeochemical Cycles*, *10*(4), 603–628.
- Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ.*, *83*, 287–302.
- Garrigues, S., et al. (2008), Validation and intercomparison of global Leaf Area Index products derived from remote sensing data, *J. Geophysical Research-biogeosciences*, *113*(G02028), doi:10.1029/2007JG000635.
- Gervois, S., N. de Noblet-Ducoudre, N. Voivy, and P. Ciais (2004), Including croplands in a global biosphere model: Methodology and evaluation at specific sites, *Earth Interactions*, *8*(16), 1–25.
- Hansen, M., R. DeFries, J. Townshend, M. Carroll, C. Dimiceli, and R. Sohlberg (2003), Global percent tree cover at a spatial resolution of 500 meters: First results of the modis vegetation continuous fields algorithm, *Earth Interactions*, *7*(10), 1–15.
- Hansen, M. C., R. Defries, J. R. G. Townshend, and R. Sohlberg (2000), Global land cover classification at 1km spatial resolution using a classification tree approach, *Int. J. Remote Sensing*, *21*(6/7), 1331–1364.
- Heck, P., D. Luthi, and C. Schär (1999), The influence of vegetation on the summertime evolution of European soil moisture, *Phys. Chem. Earth (B) -Hydrol. Oceans Atmos.*, *24*, 609–614.
- Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara (2008), Hole-filled seamless srtm data v4, <http://srtm.csi.cgiar.org>.
- Jarvis, P. G. (1976), The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Philosophical Transactions of the Royal Society of London, Series B*, *273*, 593–610.
- Jolly, W. M., R. Nemani, and S. W. Running (2005), A generalized, bioclimatic index to predict foliar phenology in response to climate, *Global Change Biol.*, *11*, 619–632.
- Jonsson, P., and L. Eklundh (2002), Seasonality extraction by function fitting to time-series of satellite sensor data, *Ieee Transactions On Geoscience Remote Sensing*, *40*, 1824–1832.
- Justice, C. O., J. R. G. Townshend, E. F. Vermote, E. Masuoka, R. E. Wolfe, N. Saleous, D. P. Roy, and J. T. Morisette (2002), An overview of MODIS Land data processing and product status, *Remote Sens. Environ.*, *83*, 3–15.
- Keeling, C. D., J. F. S. Chin, and T. P. Whorf (1996), Increased activity of northern vegetation inferred from atmospheric CO₂ measurements, *Nature*, *382*, 146–149.
- Keptert, J. D. (2004), On ensemble representation of the observation – error covariance in the ensemble kalman filter, *Ocean Dynamics*, *6*, 561–569.
- Kim, Y., and G. L. Wang (2005), Modeling seasonal vegetation variation and its validation against Moderate Resolution Imaging spectroradiometer (MODIS) observations over North America, *J. Geophysical Research-atmospheres*, *110*(D04106), doi:10.1029/2004JD005436.

- Knorr, W., T. Kaminski, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P. P. Mathieu (2010), Carbon cycle data assimilation with a generic phenology model, *Journal of Geophysical Research-Biogeosciences*, 115, doi:10.1029/2009JG001119.
- Körner, C. (2003), Slow in, rapid out – carbon flux studies and kyoto targets, *Science*, 300, 1242–1243.
- Körner, C. (2006), *Plant Growth and Climate Change*, chap. Significance of temperature in plant life, pp. 48–69, Blackwell Publishing Ltd, Oxford.
- Körner, C., and D. Basler (2010), Phenology under global warming, *Science*, 327, 1461–1462.
- Kramer, K., I. Leinonen, and D. Loustau (2000), The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview, *Int. J. Biometeorology*, 44, 67–75.
- Kucharik, C. J., C. C. Barford, M. E. Maayar, S. C. Wofsy, R. K. Monson, and D. D. Baldocchi (2006), A multiyear evaluation of a Dynamic Global Vegetation Model at three AmeriFlux forest sites: Vegetation structure, phenology, soil temperature, and CO₂ and H₂O vapor exchange, *Ecological Modelling*, 196(1–2), 1–31.
- Larcher, W. (2003), *Physiological plant ecology: ecophysiology and stress physiology of functional groups*, 4 ed., Springer-Verlag Heidelberg, Germany.
- Law, B., et al. (2002), Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation, *Agricultural and Forest Meteorology*, 113, 97–120.
- Lawrence, P. J., and T. N. Chase (2007), Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), *J. Geophysical Research-biogeosciences*, 112(G01023), doi:10.1029/2006JG000168.
- Lee, J.-E., R. S. Oliveira, T. E. Dawson, and I. Fung (2005), Root functioning modifies seasonal climate, *Proceedings of the National Academy of Sciences of the United States of America*, 102(49), 17,576–17,581.
- Leff, B., N. Ramankutty, and J. A. Foley (2004), Geographic distribution of major crops across the world, *Global Biogeochemical Cycles*, 18(GB1009), doi:10.1029/2003GB002108.
- Levis, S., and G. B. Bonan (2004), Simulating springtime temperature patterns in the community atmosphere model coupled to the community land model using prognostic leaf area, *J. Clim.*, 17(23), 4531–4540.
- Li, H., E. Kalnay, and T. Miyoshi (2009), Simultaneous estimation of covariance inflation and observation errors within an ensemble kalman filter, *Quarterly J. Royal Meteorol. Soc.*, 135, 523–533.
- Liang, L., M. D. Schwartz, and S. Fei (2011), Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest, *Remote Sens. Environ.*, 115, 143–157.
- Lokupitiya, E., et al. (2009), Incorporation of crop phenology in simple biosphere model (sibcrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, 6, 969–986.
- Los, S. O., et al. (2000), A global 9-yr biophysical land surface dataset from NOAA AVHRR data, *J. Hydrometeorol.*, 1, 183–199.
- Lu, L., R. A. P. Sr, G. E. Liston, W. J. Parton, D. Ojima, and M. Hartman (2001), Implementation of a two-way interactive atmospheric and ecological model and its application to the central united states, *Journal of Climate*, 14, 900–919.
- Mahadevan, P., et al. (2008), A satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (vprm), *Global Biogeochemical Cycles*, 22(GB2005), doi:10.1029/2006GB002735.
- Menzel, A., et al. (2006), European phenological response to climate change matches the warming pattern, *Global Change Biology*, 12, 1969–1976, doi:10.1111/j.1365-2486.2006.01193.x.
- Message P Forum (1994), MPI: A message-passing interface standard, *Technical Report UT-CS-94-230*, University of Tennessee Knoxville, TN, USA.
- Monsi, M., and T. Saeki (2005), On the factor light in plant communities and its importance for matter production, *Ann. Bot.*, 95, 549–567.
- Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005), Dual stateparameter estimation of hydrological models using ensemble kalman filter, *Advances in Water Resources*, 28, 135–147.
- Myneni, R. B., Y. Knyazikhin, Y. Zhang, Y. Tian, Y. Wang, and A. Lötters (1999), Modis leaf area index (lai) and fraction of photosynthetically active radiation absorbed by vegetation (fpar) product (mod15), *Algorithm theoretical basis document version 4.0*, NASA's Goddard Space Flight Center School of Forestry University of Montana.
- Myneni, R. B., et al. (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, 83, 214–231.
- Myneni, R. B., et al. (2007), Large seasonal swings in leaf area of amazon rainforests, *PNAS*, doi:10.1073/pnas.0611338104.
- Oleson, K. W., et al. (2008), Improvements to the community land model and their impact on the hydrological cycle, *Journal of Geophysical Research-Biogeosciences*, 113(G1), doi:10.1029/2007JG000563.
- Penuelas, J., T. Rutishauser, and I. Filella (2009), Phenology feedbacks on climate change, *Science*, 324(5929), 887–888, doi:10.1126/science.1173004.
- Phillips, O. L., et al. (2009), Drought sensitivity of the amazon rainforest, *Science*, 323(5919), 1344–1347, doi:10.1126/science.1164033.
- Piao, S., P. Friedlingstein, P. Ciais, N. Viovy, and J. Demarty (2007), Growing season extension and its impact on terrestrial carbon cycle in the northern hemisphere over the past 2 decades, *Glob. Biogeochem. Cycle*, 21(GB3018), doi:10.1029/2006GB002888.
- Pisek, J., J. M. Chen, R. Lacaze, O. Sonnentag, and K. Alikas (2010), Expanding global mapping of the foliage clumping index with multi-angular polder three measurements: Evaluation and topographic compensation, *Isprs Journal of Photogrammetry and Remote Sensing*, 65(4), 341–346, doi:10.1016/j.isprsjprs.2010.03.002.
- Poulter, B., and W. Cramer (2009), Satellite remote sensing of tropical forest canopies and their seasonal dynamics, *International Journal of Remote Sensing*, 30(24), 6575–6590, doi:10.1080/01431160903242005.
- Randerson, J. T., et al. (2009), Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, *Global Change Biology*, 15(10), 2462–2484, doi:10.1111/j.1365-2486.2009.01912.x.
- Rayner, P. J. (2010), The current state of carbon-cycle data assimilation, *Current Opinion In Environmental Sustainability*, 2(4), 289–296, doi:10.1016/j.cosust.2010.05.005.
- Richardson, A. D., B. H. Braswell, D. Y. Hollinger, J. P. Jenkins, and S. V. Ollinger (2009), Near-surface remote sensing of spatial and temporal variation in canopy phenology, *Ecol. Appl.*, 19(6), 1417–1428.
- Rüdiger, C., C. Albergel, J.-F. Mahfouf, J.-C. Calvet, and J. P. Walker (2010), Evaluation of the observation operator jacobian for leaf area index data assimilation with an extended kalman filter, *J. Geophys. Res.*, 115(D09111), doi:10.1029/2009JD012912.
- Rutishauser, T., J. Luterbacher, F. Jeanneret, C. Pfister, and H. Wanner (2007), A phenology-based reconstruction of inter-annual changes in past spring seasons, *Journal of Geophysical Research*, 112, G04,016, doi:10.1029/2006JG000382.
- Ryu, Y., D. D. Baldocchi, J. Verfaillie, S. Ma, M. Falk, I. Ruiz-Mercado, T. Hehn, and O. Sonnentag (2010a), Testing the performance of a novel spectral reflectance sensor, built with light emitting diodes (leds), to monitor ecosystem metabolism, structure and function, *Agricultural and Forest Meteorology*, 150(12), 1597–1606, doi:10.1016/j.agrformet.2010.08.009.
- Ryu, Y., O. Sonnentag, T. Nilson, R. Vargas, H. Kobayashi, R. Wenk, and D. D. Baldocchi (2010b), How to quantify tree leaf area index in an open savanna ecosystem: A multi-instrument and multi-model approach, *Agricultural and Forest Meteorology*, 150(1), 63–76, doi:10.1016/j.agrformet.2009.08.007.
- Saleska, S. R., K. Didan, A. R. Huete, and H. R. da Rocha (2007), Amazon forests greenup during 2005 drought, *Science*, 318(5850), 612, doi:10.1126/science.1146663.
- Samanta, A., S. Ganguly, H. Hashimoto, S. Devadiga, E. Vermote, Y. Knyazikhin, R. R. Nemani, and R. B. Myneni (2010), Amazon forests did not greenup during the 2005 drought, *Geophys. Res. Lett.*, 37(L05401), doi:10.1029/2009GL042154.
- Schaefer, K., A. S. Denning, and O. Leonard (2005), The winter arctic oscillation, the timing of spring, and carbon fluxes in the northern hemisphere, *Global Biogeochemical Cycles*, 19, doi:10.1029/2004GB002336.

- Scheffinger, H., A. Menzel, E. Koch, C. Peter, and R. Ahas (2002), Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central europe, *International J. of Climatology*, *22*, 1739–1755.
- Schimel, D. S., B. H. Braswell, and W. J. Parton (1997), Equilibration of the terrestrial water, nitrogen, and carbon cycles, *Proceedings of the National Academy of Sciences of the United States of America*, *94*(16), 8280–8283.
- Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996a), A revised land surface parameterization (SiB2) for atmospheric GCMs .2. The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, *9*, 706–737.
- Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua (1996b), A revised land surface parameterization (SiB2) for atmospheric GCMs .1. Model formulation, *J. Clim.*, *9*, 676–705.
- Sellers, P. J., et al. (1997), Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science*, *275*(5299), 502–509.
- Shabanov, N. V., Y. Wang, W. Buermann, J. Dong, S. Hoffman, G. R. Smith, Y. Tian, Y. Knyazikhin, and R. B. Myneni (2003), Effect of foliage spatial heterogeneity in the MODIS LAI and FPAR algorithm over broadleaf forests, *Remote Sens. Environ.*, *85*, 410–423.
- Sitch, S., et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the lpj dynamic global vegetation model, *Global Change Biology*, *9*(2), 161–185.
- Still, C. J., J. A. Berry, G. J. Collatz, and R. S. DeFries (2003), Global distribution of C-3 and C-4 vegetation: Carbon cycle implications, *Global Biogeochemical Cycles*, *17*(1), doi:10.1029/2001GB001807.
- Stöckli, R., and P. Vidale (2004), European plant phenology and climate as seen in a 20-year avhrr land-surface parameter dataset, *International Journal of Remote Sensing*, *25*(17), 3303–3330, doi:10.1080/01431160310001618149.
- Stöckli, R., D. M. Lawrence, G. Y. Niu, K. W. Oleson, P. E. Thornton, Z. L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008a), Use of fluxnet in the community land model development, *Journal of Geophysical Research-Biogeosciences*, *113*(G1), doi:10.1029/2007JG000562.
- Stöckli, R., T. Rutishauser, D. Dragoni, J. O’Keefe, P. E. Thornton, M. Jolly, L. Lu, and A. S. Denning (2008b), Remote sensing data assimilation for a prognostic phenology model, *Journal of Geophysical Research-Biogeosciences*, *113*(G4), doi:10.1029/2008JG000781.
- Studer, S., R. Stöckli, C. Appenzeller, and P. L. Vidale (2007), A comparative study of satellite and ground-based phenology, *International Journal of Biometeorology*, *51*(5), 405–414, doi:10.1007/s00484-006-0080-5.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192.
- Thornton, P. E., J.-F. Lamarque, N. A. Rosenbloom, and N. Mahowald (2007), Influence of carbon-nitrogen cycle coupling on land model response to co2 fertilization and climate variability, *Glob. Biogeochem. Cycles*, *21*(GB4018), doi:10.1029/2006GB002868.
- Thornton, P. E., et al. (2002), Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, *Agricultural and Forest Meteorology*, *113*(1-4), 185–222.
- Tsvetsinskaya, E. A., L. O. Mearns, and W. E. Easterling (2001), Investigating the effect of seasonal plant growth and development in three-dimensional atmospheric simulations. part i: Simulation of surface fluxes over the growing season, *J. Clim.*, *14*(5), 692–709.
- Tucker, C. J., J. E. Pinzon, M. E. Brown, D. A. Slayback, E. W. Pak, R. Mahoney, E. F. Vermote, and N. E. Saleous (2005), An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *Int. J. Remote Sensing*, *26*(20), 4485–4498.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Quarterly J. Royal Meteorol. Soc.*, *131*(612), 2961–3012.
- USGS (1996), GTOPO30, Digital Media.
- Ustin, S. L., and J. A. Gamon (2010), Remote sensing of plant functional types, *New Phytologist*, *186*, 795–816.
- van Vliet, A. J. H., et al. (2003), The European Phenology Network, *Int. J. Biometeorology*, *47*(4), 202–212.
- Wang, Y. J., et al. (2004), Evaluation of the MODIS LAI algorithm at a coniferous forest site in finland, *Remote Sens. Environ.*, *91*, 114–127.
- Wilmott, C. J., and K. Matsuura (2007), Global air temperature and precipitation: Regrided monthly and annual climatologies (v 2.01 for precip and v 2.02 for temperature), http://climate.geog.udel.edu/~climate/html_pages/download.html.
- Zhao, M., and S. W. Running (2010), Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, *Science*, *329*(5994), 940–943, doi:10.1126/science.1192666.
- Zupanski, M. (2005), Maximum likelihood ensemble filter: Theoretical aspects, *Monthly Weather Rev.*, *133*(6), 1710–1726.

R. Stöckli, Federal Office of Meteorology and Climatology MeteoSwiss, Climate Services, Climate Analysis, Krähbühlstrasse 58, 8044 Zürich, Switzerland. (reto.stoeckli@meteoswiss.ch)

Figure 1. Geographic (a) and climatic (b) distribution of the regions used for the data assimilation: experiments with 4 regions (red large squares); 16 regions (red squares + blue triangles); 64 regions (red squares + blue triangles + green diamonds); 256 regions (red squares + blue triangles + green diamonds + violet circles). The colors of the climatic distribution (b) qualitatively show the relative probability of occurrence for the given climatic zone (bright yellow: low probability; dark red: high probability)



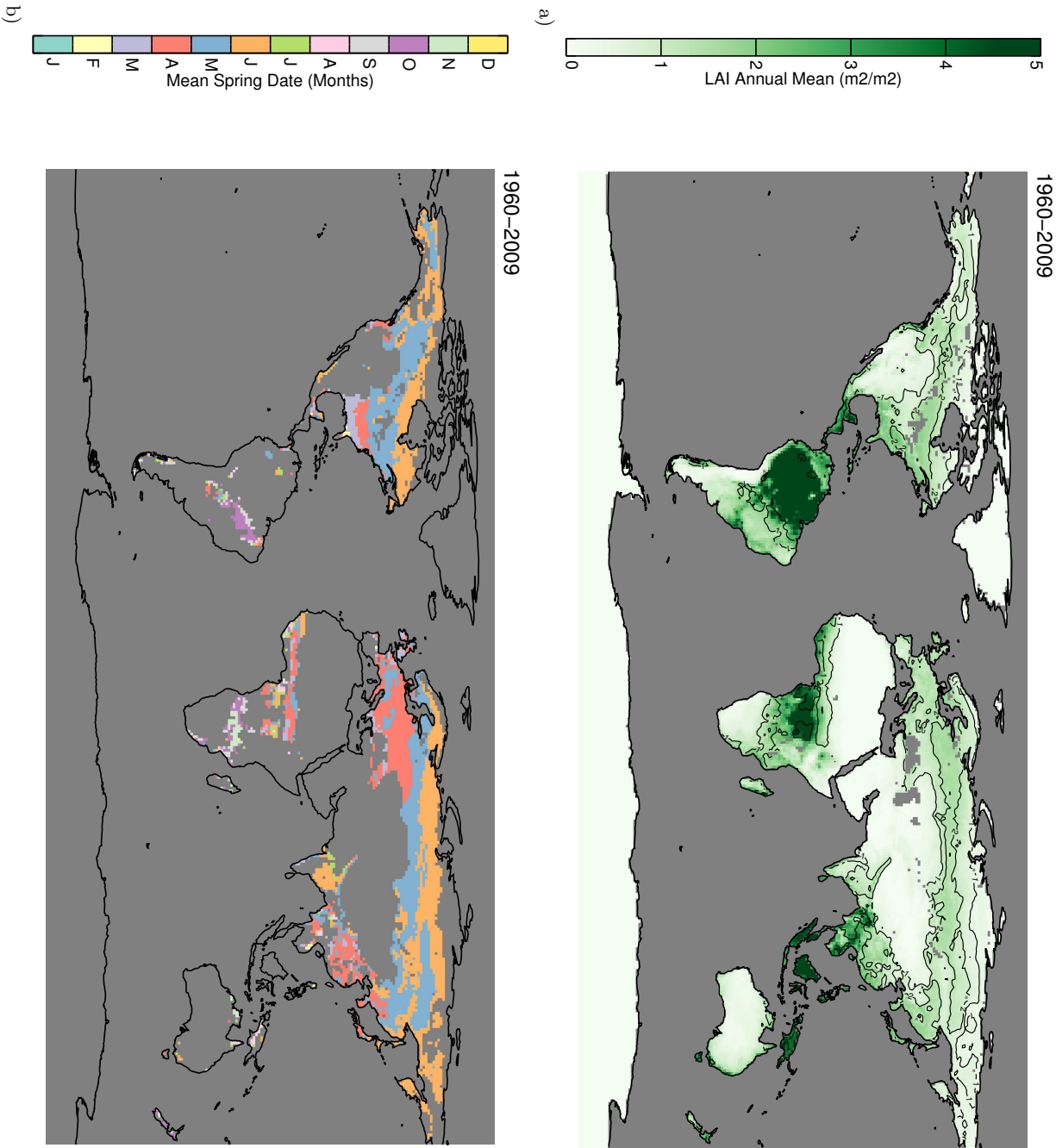


Figure 2. The 50 year long (1960-2009) global reanalysis dataset of vegetation phenology. a) Annual mean LAI and seasonal amplitude (contours with 1 and 2 $\text{m}^2 \text{m}^{-2}$). b) Mean spring date for grid points that have a seasonal amplitude above 1 $\text{m}^2 \text{m}^{-2}$.

Figure 3. Global maps with the Mean Absolute Deviation (MAD) of the predicted FPAR and LAI versus MODIS FPAR and LAI using the prior parameter set and the parameters constrained by 4 and 256 regions during 2000–2009.

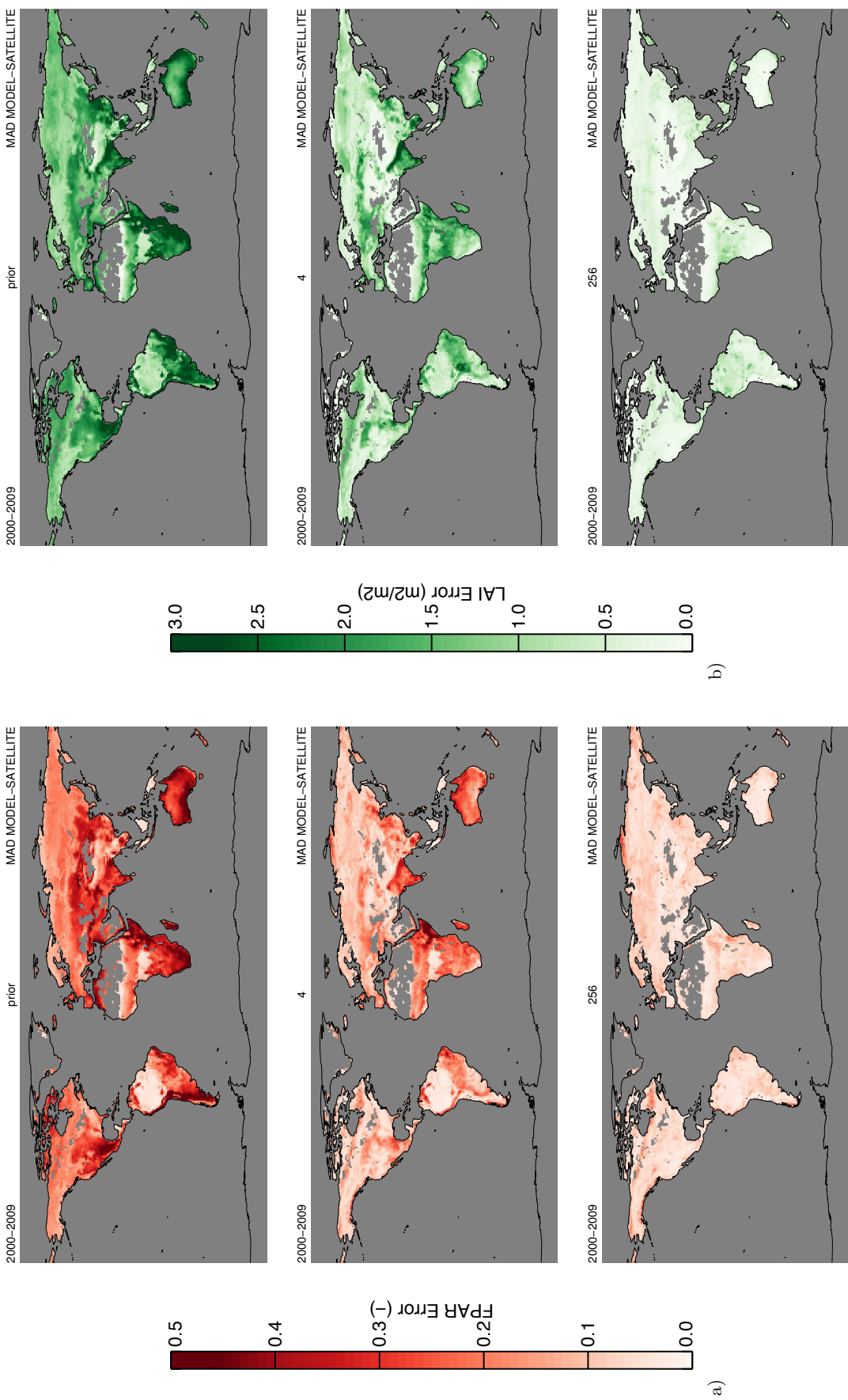


Figure 4. Relationship between the global FPAR and LAI prediction error (blue and red solid lines) and prediction uncertainty (blue and red dashed lines) respectively and the number of observations in the assimilation experiments using 4, 16, 64 and 256 regions.

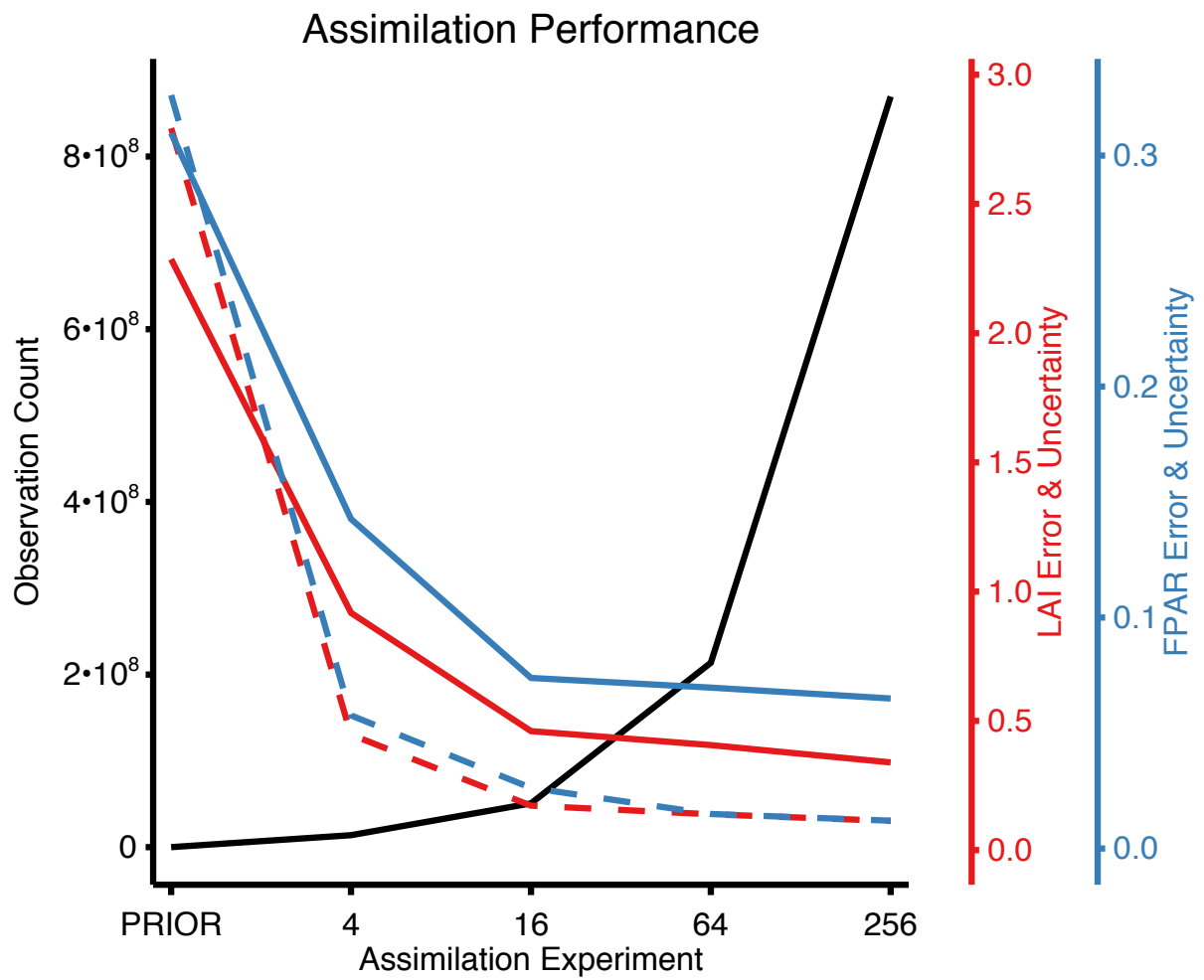


Figure 5. Performance of regional FPAR and LAI predictions by pft with the prior parameter set and the parameters constrained by the 256 region experiment. Modeled FPAR and LAI have triangle symbols when they match observations accurately (both bias and rmse < 5% of FPAR and LAI range according to Table 7).

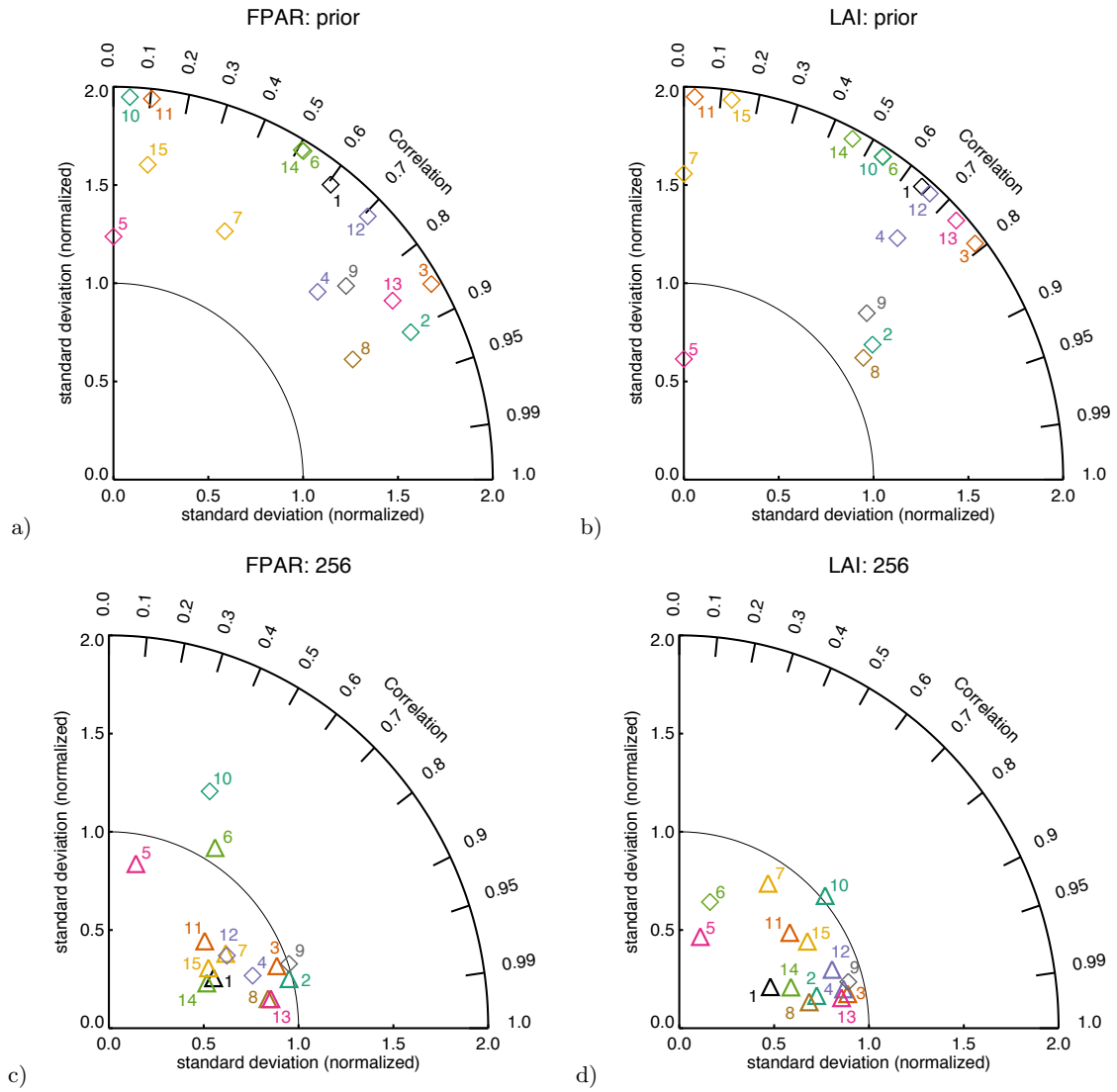


Figure 6. Predicted versus observed site-level ($0.5^\circ \times 0.5^\circ$) LAI using the site-level, the prior parameter set and the parameter sets constrained by 4, 16, 64 and 256 regions during 2003.

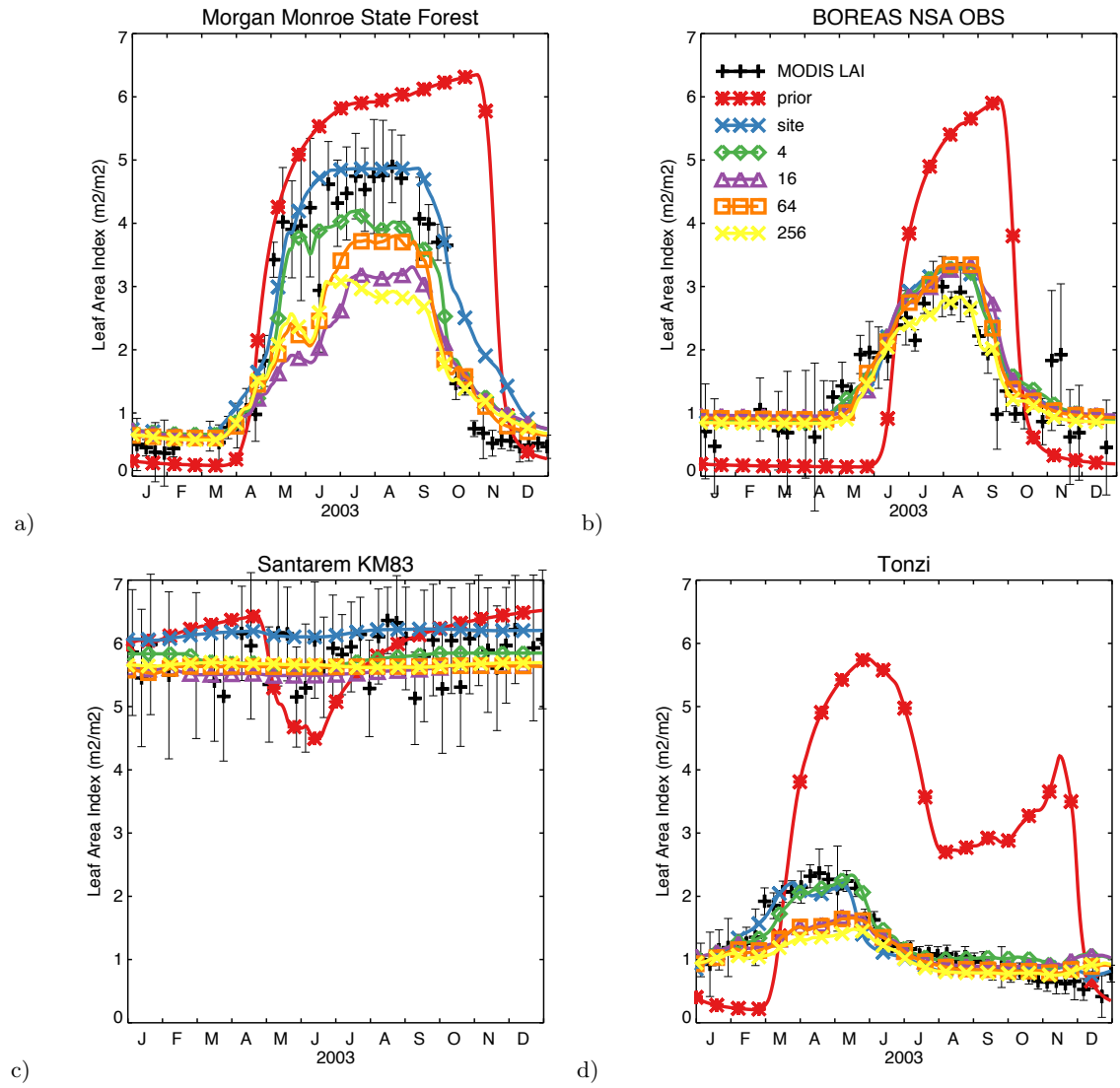


Table 1. State and parameter vector ψ , initial values ψ^0 , initial variances V_ψ^0 , minimum and maximum bounds for the ensemble mean.

No.	variable	ψ^0	V_ψ^0	min	max	units
states x						
1	FPAR	0.5	0.5	0.0	1.0	-
2	LAI	2.5	1.0	0	10	m ² m ⁻²
3	T		0.25	200	350	K
4	L		10	0	1000	W m ⁻²
5	W		0.01	0	100	mb
parameters θ						
6	T_{max}	280	50	100	350	K
7	T_{min}	265	50	100	350	K
8	L_{max}	150	1000	-100	500	W m ⁻²
9	L_{min}	50	1000	-100	500	W m ⁻²
10	W_{max}	30	50	-25	50	mb
11	W_{min}	10	50	-25	50	mb
12	FPAR _{min}	0.05	0.01	0.0	1.0	-
13	FPAR _{max}	0.95	0.01	0.0	1.0	-
14	γ_g	0.33	0.01	0.05	1.0	days ⁻¹
15	γ_d	0.33	0.01	0.05	1.0	days ⁻¹
16	LAI _{max}	7.0	0.5	0	10	m ² m ⁻²
17	FPAR _{sat}	0.95	0.01	0.5	1.0	-
18	τ_T	21	10	5	100	days
19	τ_W	21	10	5	100	days
20	τ_L	21	10	5	100	days

Table 2. List of pft's including their abbreviations. Only the pft's of natural vegetation types are given.

No	pft name	pft abbreviation
1	Bare Soil, Rock, Ice, Permanent Snow	bar all
2	Trees: Temperate Evergreen Needleleaf	enf tem
3	Trees: Boreal Evergreen Needleleaf	enf bor
4	Trees: Boreal Deciduous Needleleaf	dnf bor
5	Trees: Tropical Evergreen Broadleaf	ebf tro
6	Trees: Temperate Evergreen Broadleaf	ebf tem
7	Trees: Tropical Deciduous Broadleaf	dbf tro
8	Trees: Temperate Deciduous Broadleaf	dbf tem
9	Trees: Boreal Deciduous Broadleaf	dbf bor
10	Shrubs: Evergreen Broadleaf	ebs all
11	Shrubs: Temperate Deciduous Broadleaf	dbs tem
12	Shrubs: Boreal Deciduous Broadleaf	dbs bor
13	Grass: Arctic C3	c3g arc
14	Grass: Non-arctic C3	c3g nar
15	Grass: C4	c4g all

Table 3. Climate control parameters (mean and standard deviation) by pft constrained by the assimilation using 256 regions. pft abbreviations are explained in Table 2.

No.	pft	T_{min} K	T_{max} K	W_{min} mb	W_{max} mb	L_{min} W m ⁻²	L_{max} W m ⁻²
1	bar all	270.6 ± 0.7	290.9 ± 0.8	12.5 ± 0.7	23.6 ± 0.4	102.7 ± 10.3	149.4 ± 6.5
2	enf tem	263.1 ± 0.5	276.4 ± 0.3	6.9 ± 0.3	47.9 ± 1.3	-68.3 ± 7.3	216.7 ± 2.5
3	enf bor	263.8 ± 0.6	290.0 ± 0.7	7.6 ± 0.4	21.4 ± 2.4	-82.8 ± 10.0	197.4 ± 4.4
4	dnf bor	262.2 ± 0.9	275.6 ± 0.7	18.8 ± 3.0	27.9 ± 3.8	103.9 ± 5.9	208.0 ± 2.7
5	ebf tro	271.3 ± 1.8	292.8 ± 0.3	21.9 ± 0.6	-1.4 ± 2.2	82.3 ± 9.4	168.9 ± 2.6
6	ebf tem	259.1 ± 1.0	285.9 ± 0.3	10.1 ± 0.4	20.9 ± 3.0	14.1 ± 10.7	35.0 ± 6.0
7	dbf tro	278.0 ± 0.4	299.1 ± 0.1	9.9 ± 0.2	43.9 ± 0.6	44.0 ± 13.8	81.4 ± 7.6
8	dbf tem	269.7 ± 0.3	291.5 ± 0.2	5.1 ± 0.2	25.4 ± 0.3	44.3 ± 3.9	203.0 ± 1.8
9	dbf bor	271.0 ± 0.6	279.8 ± 0.3	7.0 ± 1.0	46.9 ± 3.5	110.1 ± 3.7	223.4 ± 2.2
10	ebs all	265.5 ± 2.2	281.7 ± 0.8	3.4 ± 0.7	14.4 ± 0.4	-7.0 ± 7.1	242.4 ± 6.0
11	dbb tem	256.9 ± 0.6	298.0 ± 0.2	1.6 ± 0.4	44.5 ± 0.5	-4.7 ± 9.2	69.3 ± 3.8
12	dbb bor	273.5 ± 0.3	287.8 ± 0.5	17.5 ± 1.0	11.7 ± 2.9	60.8 ± 11.2	68.0 ± 8.1
13	c3g arc	267.8 ± 0.4	282.0 ± 0.4	2.3 ± 0.3	13.5 ± 0.5	19.9 ± 7.1	198.2 ± 3.2
14	c3g nar	267.1 ± 0.2	298.2 ± 0.5	1.5 ± 0.2	15.4 ± 0.1	-21.4 ± 6.6	63.0 ± 3.3
15	c4g all	268.6 ± 0.4	279.2 ± 0.3	4.1 ± 0.2	23.3 ± 0.2	-9.0 ± 5.1	217.7 ± 1.4

Table 4. Structural parameters (mean and standard deviation) by pft constrained by the assimilation using 256 regions. pft abbreviations are explained in Table 2.

No.	pft	FPAR _{min} -	FPAR _{max} -	γ_g days ⁻¹	γ_d days ⁻¹	LAI _{sat} m ² m ⁻²	FPAR _{sat} -
1	bar all	0.11 ± 0.00	0.05 ± 0.00	0.19 ± 0.03	0.05 ± 0.01	9.70 ± 0.31	1.00 ± 0.01
2	enf tem	0.52 ± 0.01	0.98 ± 0.00	0.19 ± 0.02	0.19 ± 0.02	5.93 ± 0.25	0.98 ± 0.00
3	enf bor	0.52 ± 0.01	1.00 ± 0.01	0.34 ± 0.03	0.39 ± 0.04	6.23 ± 0.28	0.98 ± 0.00
4	dnf bor	0.33 ± 0.02	1.00 ± 0.00	0.49 ± 0.04	0.37 ± 0.04	6.69 ± 0.28	1.00 ± 0.00
5	ebf tro	0.16 ± 0.04	0.99 ± 0.00	0.57 ± 0.05	0.05 ± 0.01	7.07 ± 0.02	0.93 ± 0.01
6	ebf tem	0.01 ± 0.03	1.00 ± 0.01	0.45 ± 0.04	0.05 ± 0.00	6.91 ± 0.07	0.96 ± 0.00
7	dbf tro	0.27 ± 0.01	1.00 ± 0.01	0.37 ± 0.03	0.21 ± 0.02	6.85 ± 0.07	0.93 ± 0.00
8	dbf tem	0.29 ± 0.00	1.00 ± 0.01	0.57 ± 0.03	0.42 ± 0.02	6.01 ± 0.12	0.92 ± 0.00
9	dbf bor	0.23 ± 0.01	1.00 ± 0.01	0.60 ± 0.04	0.49 ± 0.04	6.85 ± 0.23	0.94 ± 0.01
10	ebs all	0.39 ± 0.01	0.85 ± 0.02	0.36 ± 0.04	0.31 ± 0.04	6.02 ± 0.31	0.97 ± 0.01
11	dbb tem	-0.00 ± 0.01	0.77 ± 0.01	0.49 ± 0.04	0.43 ± 0.03	3.74 ± 0.10	0.84 ± 0.01
12	dbb bor	0.33 ± 0.01	0.84 ± 0.02	0.42 ± 0.03	0.47 ± 0.04	7.50 ± 0.31	0.99 ± 0.00
13	c3g arc	0.12 ± 0.00	0.63 ± 0.01	0.30 ± 0.03	0.15 ± 0.02	6.80 ± 0.27	0.99 ± 0.00
14	c3g nar	0.24 ± 0.00	0.94 ± 0.01	0.47 ± 0.02	0.37 ± 0.02	8.36 ± 0.10	1.00 ± 0.00
15	c4g all	0.19 ± 0.00	0.51 ± 0.00	0.55 ± 0.03	0.13 ± 0.01	7.87 ± 0.31	1.00 ± 0.00

Table 5. Time averaging parameters (mean and standard deviation) by pft constrained by the assimilation using 256 regions. pft abbreviations are explained in Table 2.

No.	pft	τ_T days	τ_W days	τ_L days
1	bar all	34.1 ± 1.3	43.0 ± 1.1	21.4 ± 1.5
2	enf tem	25.9 ± 1.1	30.5 ± 1.4	12.2 ± 1.0
3	enf bor	5.3 ± 0.8	19.1 ± 1.6	5.4 ± 1.1
4	dnf bor	17.0 ± 1.3	9.6 ± 1.7	10.7 ± 1.1
5	ebf tro	22.9 ± 1.7	36.1 ± 1.6	5.0 ± 1.2
6	ebf tem	12.4 ± 1.2	13.0 ± 1.6	23.4 ± 1.6
7	dbf tro	21.4 ± 1.0	11.3 ± 0.7	24.1 ± 1.6
8	dbf tem	5.5 ± 0.5	33.7 ± 1.0	15.4 ± 0.8
9	dbf bor	12.9 ± 0.9	17.6 ± 1.6	16.5 ± 1.1
10	ebs all	19.1 ± 1.7	29.1 ± 1.4	13.5 ± 1.2
11	dbb tem	16.4 ± 0.9	10.7 ± 0.6	19.9 ± 1.4
12	dbb bor	5.4 ± 0.6	18.7 ± 1.6	30.9 ± 1.5
13	c3g arc	5.2 ± 0.8	23.8 ± 1.3	25.4 ± 1.2
14	c3g nar	5.0 ± 0.5	18.3 ± 0.6	20.3 ± 1.2
15	c4g all	17.8 ± 1.0	7.2 ± 0.3	27.3 ± 0.9

Table 6. Number and percentage of assimilated observations (relative to available non-QA screened observations and relative to total global land area) as well as resulting FPAR and LAI posterior uncertainties and prediction errors.

Experiment	Number	Observations		FPAR (-)		LAI (m ² m ⁻²)	
		% QA passed	% global land	uncertainty	error	uncertainty	error
PRIOR	0	0	0	0.326	0.310	2.79	2.29
4	13'943'482	44.8	0.007	0.058	0.143	0.45	0.92
16	50'768'668	42.8	0.025	0.026	0.074	0.17	0.46
64	213'642'410	45.4	0.098	0.015	0.070	0.14	0.41
256	869'605'738	44.7	0.405	0.011	0.065	0.11	0.34

Table 7. Bias and RSME (in brackets) of regional FPAR and LAI predictions by pft with the prior parameter set and the parameters constrained by the 256 region experiment. The accuracy of bold values is better than 5% of the full FPAR or LAI range (1.0 and 8.0, respectively).

No.	pft	FPAR		LAI	
		prior	256	prior	256
1	bar all	0.30 (0.31)	-0.01 (0.02)	1.62 (1.63)	0.02 (0.04)
2	enf tem	0.02 (0.10)	-0.02 (0.03)	0.93 (1.11)	-0.17 (0.33)
3	enf bor	-0.12 (0.22)	0.00 (0.05)	0.32 (1.14)	0.03 (0.16)
4	dnf bor	-0.06 (0.21)	0.00 (0.08)	0.56 (1.32)	0.03 (0.23)
5	ebf tro	0.06 (0.06)	0.01 (0.02)	0.01 (0.23)	0.19 (0.27)
6	ebf tem	0.04 (0.12)	0.00 (0.02)	0.62 (0.93)	0.25 (0.41)
7	dbf tro	0.16 (0.17)	0.00 (0.01)	1.36 (1.39)	0.08 (0.15)
8	dbf tem	0.07 (0.12)	-0.03 (0.04)	1.07 (1.26)	-0.09 (0.38)
9	dbf bor	-0.00 (0.21)	-0.00 (0.07)	0.68 (1.39)	0.19 (0.42)
10	ebs all	0.01 (0.16)	-0.04 (0.07)	0.79 (1.07)	-0.10 (0.27)
11	dbb tem	0.33 (0.33)	0.04 (0.04)	2.14 (2.17)	0.25 (0.26)
12	dbb bor	-0.04 (0.18)	0.01 (0.07)	0.61 (1.18)	0.04 (0.15)
13	c3g arc	0.02 (0.13)	-0.00 (0.03)	0.76 (1.16)	0.03 (0.09)
14	c3g nar	0.27 (0.28)	0.02 (0.03)	1.95 (1.99)	0.11 (0.13)
15	c4g all	0.26 (0.27)	0.01 (0.02)	2.09 (2.10)	0.10 (0.12)