Referee Comment on the article ”Identification of a general light use efficiency model for gross primary production” by J. E. Horn & K. Schulz in *Biogeosciences Discussions*

General comments

I reviewed the paper chiefly from the methodological point of view. I found it interesting and appreciated the substantial amount of work that it contains. However, the authors have not paid enough attention to the careful documentation of their work: as its present form, the article does not describe the methods sufficiently clearly and explicitly so that a reader could easily follow the ideas or that other scientists could reproduce the work. The overall structure of the paper is nice and clear, but the language would need to be clarified and revised.

A major scientific question that arose in my mind is the motivation for using dynamic linear regression (DLR) and state dependent parameter estimation (SDP) in the estimation of the time series of light use efficiency (ε) from the time series of gross photosynthesis (F_G) and radiation (S_0, APAR), as opposed to estimating ε as a simple ratio of F_G and S_0 or APAR. The application of these novel methods DLR and SDP is one of the points of the paper, and therefore their validity should be explained and discussed thoroughly. However, if the time series of ε are utilised merely in search for new model structures and not in the final model estimation and evaluation, as I believe is the case, this question is not crucial regarding the validity of the final model presented in the paper. I address this issue in more detail in the specific comments below (point 5).

Specific comments

1. The time scale (half-hourly, daily) should be explicitly stated in different parts of the work (evaluation of the Jarvis model, finding new model structures, formulating the generalized model, model calibration), as it makes difference whether the data contain both within-day and between-days variation, or only between-days variation. If different time scales were used in different stages (e.g. finding new model structures vs. final model calibration), this should be motivated. The time scale issue is repeatedly pointed out in the comments below.

2. In the introduction, the purpose of this study need be more clearly stated: (i) why to build yet another LUE-type model, i.e., which deficiencies of the existing models this new model is thought to remedy (cf. the vague formulation in p. 7676, r. 22–25), and (ii) why to apply DLR and SDP to estimate the time series of ε (see the specific comments in point 5 below; cf. the obscure motivation in p. 7677, r. 5–11).

3. In the description of the micrometeorological data, the precise meaning of the criterion “no lengthy measurement gaps” (p. 7678, r. 8) should be given and the time scale (half-hourly?) told; it would also be useful to show in Table 1 the measurement years included in the study for each site. The quality of eddy covariance measurements of net photosynthesis (F_N) varies a lot according to air turbulence: were any quality criteria (e.g. site-specific friction velocity thresholds) used for filtering inadequate F_N observations in this study, and if not, why? Further, it would be worthwhile to have a concise description of the semi-parametric methods that were used for the gap-filling of F_N, for the extraction of “the signal component” of it, and for the computation of gross photosynthesis F_G from this signal, since the time series of F_G are

really the basis of this study: would these semi-parametric methods artificially remove some true and natural variation in F_N and introduce in F_G regularity following some a priori model?

4. The computation of daily FPAR values in MODIS Land Products should be succinctly explained, or at least adequately referenced. What was the justification for the noise reduction of the MODIS FPAR time series, which is already a modelling product, by cubic smoothing splines and unequal weighing of observations? Clarify what is meant by multiannual mean of FPAR (p. 7680, r. 12): the mean of daily FPAR over all the days and the measurement years, or the mean of FPAR in the particular day over the measurement years?

5. For anyone not familiar with the original methodological papers by Young and his co-workers, or with the work by Jarvis et al. (2004) introducing SDP in eddy covariance context, the section describing DLR and SDP is impossible to grasp. This is unfortunate, if the study aims to further demonstrate the applicability of DLR and SDP with eddy covariance data. Surprisingly, Jarvis et al. (2004) are not even referred to in this section, although the text is fragmentarily following their methodology description — and even directly citing them, without quotation marks (p. 7679, r. 20–21 vs. p. 940–941 in Jarvis et al.)! The section should be rewritten so that the connection to the data of this study becomes clear, with (i) the basic model in Equation 4 as the starting point, (ii) stating clearly that the purpose is to estimate the time series of ε (half-hourly or daily?) from the time series data of F_G and S_0, (iii) giving precisely the model assumptions (the distributional assumptions of the error series $\zeta(t)$, and the random walk assumptions of the "parameter" series $\varepsilon(t)$) and separating them from the description of the resulting estimator of $\varepsilon(t)$, and (iv) giving adequate references. For this, Jarvis et al. (2004) set a good example.

Yet a major question is the motivation for using DLR or SDP to estimate the time series of ε. Jarvis et al. (2004) used SDP for separating from the time series of F_N provided by eddy covariance measurements the following two components: (1) F_G as a product of parameter ε and measured above-canopy solar radiation S_0, and (2) respiration F_R as a parameter. Here, however, no such separation was needed, as F_G was computed from the signal component of F_N with another semi-parametric method before modelling. So why not then to estimate the time series of ε as a simple ratio of the time series of F_G and S_0, as Equation 4 would imply? With DLR or SDP, the value of ε at a time point is obtained as a non-linear function of the values of F_G and S_0 in the vicinity of the point in a time space or in a state space sorted according to the selected state variable, respectively. What is the rationale of estimating ε in this non-transparent way, and how does it compare with the simple ratio approach (certainly it results in smoother time series, but is this smoothing justified)? This issue pertains to the evaluation of the Jarvis model (see point 8 below) and to the finding of new model structures (point 9 below).

6. As to model evaluation criteria, several definitions for the coefficient of determination r^2 exist (see e.g. Kvålseth 1985). The definition used here (p. 7680, r. 18) should be given (I guess it is the squared Pearson correlation coefficient between observed and predicted values). In the Equation 3 giving the definition of EC referred to as "the Nash-Sutcliff efficiency criterion", the differences in the sums in the numerator and denominator should be squared to meet the verbal description (p. 7680, r. 19–21). In fact, also this (corrected) EC is a coefficient of determination, corresponding to another commonly used definition (Kvålseth 1985)! Of these two definitions of r^2, the latter is certainly preferable, for the reasons stated in the text (p. 7680, r. 23–24).

7. In the model identification section, all the variables and parameters in the equations must be explained and their units given (now many explanations are not only incomplete but simply missing, e.g. S_0 in Equation 4; α in Equation 6; E, λ and ζ(t) in Equation 7; P(t) in Equation 8). Specify whether S_0 was (above-canopy?) global radiation or PAR. Specify from which depth T_S was measured; how does it compare with "measured surface temperature" used as T_S by Jarvis et al. (2004)? Specify from which layer SWC was measured. Specify what value was used for the recession constant κ in the computation of API (Equation 8, p. 7684, r. 4–6).

8. In the evaluation of the Jarvis model for ε, it need be clarified how the model was "applied to all study sites" (p. 7681, r. 16–18). From Fig. 2 I conjecture that (i) the time series of ε was estimated with DLR from the time series of F_G and S_0 for each site, and (ii) the model in Equations 5 and 6 was fitted in these estimated ε data using the non-linear least squares method. What was the motivation for estimating ε with DLR and not with SDP (with T_S as the state variable) as Jarvis et al. (2004) did when constructing the model? What was the time scale (half-hourly or daily)? And more importantly, if the aim was to find whether the Jarvis model reproduces well the daily course of F_G in the study sites (cf. p.7681, r. 18–19), should not the model have been fitted directly to the daily data comprising only F_G and S_0, as Jarvis et al. (2004) did when evaluating the model (cf. their Equation 5)? Related to this, it need be specified whether r^2 and EC were computed from the measured (DLR-estimated) and predicted ε or from measured and predicted F_G.

9. New model structures were searched by estimating the time series of ε with SDP, with one explanatory variable (radiation) and one or two state variables (soil temperature, four measures of water availability), and then studying how these time series depended on the state variables. The time scale of the modelling (half-hourly or daily) need be given. Further, the explanatory variable of the SDP model should be specified also in the case of one state variable: S_0 or APAR, and if S_0, why (using S_0 follows Jarvis et al. (2004) but deviates from the two-state-variable case in Equation 9)? It would be helpful to see the SDP model explicitly written also in the case of one state variable, specifying the explanatory variable and the dependence of the "parameter" ε on the state variable. In the case of two state variables (Equation 9), it would facilitate understanding if ε was explicitly included in the equation or its explanation (see the technical corrections below). Finally, tell how the performance of the SDP model was assessed (p. 7683, r. 9–17): with EC and r^2 computed from measured and SDP-model-predicted F_G? The content of Figs. 4 and 5 should also be explained: what are the error bands, is the course of ε some average of the SDP-estimated values?

10. Concerning the Monte Carlo simulations performed to study the sensitivity of the final model on the potential variation in parameter values, the distributions assumed for the parameters should be specified (normal?) and the determination of the parameters of these distributions (means and standard deviations?) outlined. The explanation of how the correlations between the parameters were taken into account ("to account for interrelations of the parameters the samples were generated by Cholesky decomposition of the covariance matrix of the model residuals applied to normally distributed sample"; p. 7686, r. 18–20) is inaccurate: it would make sense to apply the Cholesky decomposition to *the correlation matrix of the parameters* to get the lower-triangular matrix, with which a vector of uncorrelated parameter samples should then be multiplied to obtain a vector of properly correlated parameter samples. How was the correlation matrix (or the covariance matrix of the model residuals, if it really was erroneously used) obtained before fitting the final model? What was the point of performing the Monte Carlo simulations before the estimation of

parameters of the final model (model calibration)? Or were the simulations performed in two phases, before and after the model parameter estimation?

11. Regarding the estimation of the parameters of the final model (model calibration), it would be useful to explicitly tell on which data and time scale this was done: daily observations of F_G, APAR and water availability measures (instead of estimated half-hourly time series of ε and half-hourly time series of APAR and water availability measures).

11. In the abstract, it need be clarified what is meant by “a model formulation allowing a variable influence of the model parameters modulating the light use efficiency” (p. 7674, r. 5–6).

12. Some remarks concerning terminology: Strictly speaking, ε should not be termed light use efficiency in the case where radiation above canopy (and not radiation *absorbed* by canopy) is used as S_0 in Equation 4. Further, it is confusing to use word “parameter” in two different meanings, referring both to the time series of ε (which is indeed a time-varying parameter in DLR and SDP models) and to the unknown constants of the models of ε, or F_G, constructed on the basis these time series; to avoid confusion, some other term should preferably be used for the time series of ε.

Technical corrections

Table 1: Explain the abbreviations of the vegetation types and climate classes in the caption — you can move the explanation in the caption of Fig. 3 here.

Table 2: In the caption, “confidence interval” should probably read “standard error”. Are the incredibly high confidence intervals obtained for T_{opt} and k_T in IT-Cpz and BE-Vie true?

Fig.1: The locations of the sites in the map appear inaccurate (e.g. Swedish Flakaliden is located in the sea, Finnish inland Hyytiälä on the coast, and French Le Bray almost in Spain). Further, explain the abbreviations of the vegetation types included in the legend — you can now refer to the caption of Table 1.

Fig. 2: In the legend, indicate “Jarvis model” by a red line instead of a red point.

Fig. 3: For better readability, centre the labels of the climate classes and vegetation types (justify them in the middle of the left side of the rows and in the middle of the bottom side of the columns). For the explanation of the vegetation types and climate classes, you can refer to the caption of Table 1.

Fig. 4: Explain the error bands. Consider changing the colour of the yellow line into something easier to perceive. For consistency, use “state variable” instead of “system state” in the caption.

Fig. 5: Explain the error bands.

Fig. 6: The y-axis labels F_T and F_W should be f_T and f_W. In the legend of the figures in the lowest row, the colour of the line showing “modelled” F_G should be black.

Fig. 7: Tell what characteristic (sum of squared errors between measured and modelled F_G?) the different classes (colours) of the points represent. Tell to which sites (b) and (c) pertain (Wetzstein and Lethbridge, respectively?). Refer to this figure in the text (now no reference is found).

Fig. 8: In the legend, consider renaming EF_I into something that more resembles I_W. The numbers in x-axis in (b) should be integers (number of years). Are the y-axis values really confidence intervals and not standard errors?

Fig. 9: Refer to this figure in the text (now no reference is found). Explain the abbreviations of the vegetation types and climate classes; you can refer to the caption of Table 1, if you moved the description there. For better readability, centre the labels of the climate classes and vegetation types (cf. Fig. 3).

Fig. 10: Explain the abbreviations of the vegetation types and climate classes; you can refer to the caption of Table 1, if you moved the description there. For better readability, centre the labels of the climate classes and vegetation types (cf. Fig. 3).

Equation 3: The differences in the sums in the numerator and denominator should be squared. Some other index than i might be preferable, in order to avoid confusion with Equations 1 and 2.

Equation 4: Add the error series $\zeta(t)$.

Equation 5: Use $T_F(t)$ instead of T_F (cf. Equation 6).

Equation 6: What is the starting value of the $T_F(t)$ series?

Equation 7: Consider using a different notation (e.g. with a subscript) for the error series $\zeta(t)$, as it is not the same as that in Equation 4.

Equation 9: For correspondence to the general SDP model definition and for better readability, consider writing the explicit products of state-variable-dependent parameters and APAR as

$$F_G(t) = c_1[T_S(t)] \cdot APAR(t) + c_2[W(t)] \cdot APAR(t) + \zeta(t)$$

and stating that now $\varepsilon = c_1[T_S(t)] + c_2[W(t)]$.

Some typing or language errors:

- p. 7675, r. 19–20: "… an imbalance between the input data requirements and the actual information content of measurement data *what* enhances the forecast uncertainty…"
- p. 7676, r. 3: "… *lager* scales…"
- p. 7682, r. 15: "… *an* generalized model scheme…"
- p. 7683, r. 13–14: "… *no* of them…"
- p. 7684, r. 17–18: "… as proposed by *(*Makela et al., 2006, 2008) for sites…"
- p. 7685, r. 9–10: "APAR was chosen instead of S_0 *as* in the Jarvis et al. (2004) study for better comparability with other studies…" should probably read "APAR was

chosen instead of S_0, which was used in the Jarvis et al. (2004) study, for better comparability with other studies…"

- p. 7686, r. 26: "… as well as for *one the parameters* F_T or F_W…"
- p. 7678, r. 8–9: "Results for the other moisture surrogates are shown *exemplarily*"

References used in the comments

Jarvis AJ, Stauch VJ, Schulz K & Young PC (2004) The seasonal temperature dependency of photosynthesis and respiration in two deciduous forests. Global Change Biology 10:939–950.

Kvålseth T (1985) Cautionary note about R^2. The American Statistician 39(4):279–285.