(comments of the referees are printed in blue, responses of authors are held in black)

Response letter to B. Scharnagl Referee #1 (Author comment on Biogeosciences Discuss., 10, 19509-19540, 2013. C7528)

1 General comments

Houska et al. present an application of a coupled soil hydrological and plant growth model to simulate winter wheat. The authors apply the GLUE methodology with three likelihood functions to calibrate the model using data sets originating from three different agricultural management practices. Based on these results, the authors identify structural deficits in the plant grow model. The discussion paper presents an interesting case study that fits well into the scope of this journal.

The discussion paper is well written, concise and comprehensible. However, it lacks some relevant information on the setup of the soil hydrological model and the observations used to calibrate it. In my view, this impedes a final assessment of the discussion paper. Further details are given in the specific comments below.

2 Specific comments

(2.1) Several aspects of the soil hydrological model need to be clarified. The authors say that the Richards equation was used to simulate one-dimensional water flow in 50 soil layers. However, what was the depth of the soil profile? Is the profile assumed to be homogeneous, and if so, can this assumption be justified by available information on soil type and texture?

The depth of the soil profile was 2.25 m and is characterized as follows (Mirschel, 2007):

Site	Horizon	Depth [cm]	Sand [%]	Clay [%]	Silt [%]	org. C [%]	Bulk density [g cm^{-3}]
1	Ap	0-30	90	7	3	0.66	1.45
	Ael	30-60	90	5	5	0.16	1.5
	Bt	60-90	80	12	8	0.08	1.55
	C	90-225	90	5	5		
2	Ap	0-30	85	5	10	0.58	1.45
	Ael	30-90	90	5	5	0.13	1.5
	Bt1	90-130	80	12	8		1.55
	Bt2	130-170	80	10	10		
	C	170-225	90	5	5		
3	Ap	0-30	85	6	9	0.62	1.45
	Ael	30-100	90	5	5	0.13	1.5
	Bt1	100-110	81	13	6		1.55
	Bt2	110-225	80	11	9		

We justify our assumption of a homogeneous profile with a sand fraction from 80 to 90% on every site and depth.

We added several information to the soil description of chapter 2.3. It now reads (p.19520 l.1): “Sites are characterized by a primarily sandy *Eutric Cambisol*, with a homogenous volumetric sand content of 80 to 90% in a soil profile with 2.25 m depth. Silt and clay content contribute 5 to 10%. The soil is medium textured with good structural stability. The bulk density is around 1.5 g cm^{-1} and the organic matter

content in the first 0.3 m amounts to 0.6%. An in depth description of soil physical and chemical properties is given by Mirschel (2007).“

How thick is one such layer? The numerical simulation of soil water dynamics under natural boundary conditions requires a fine discretization of the simulation domain near the soil surface to accommodate for very large gradients in matric potential that may arise as a response to meteorological forcing.

We simulated an equally thickness of 5 cm for every layer. A finer discretization in the upper soil layers would have been certainly possible. However, in a set of initial model runs we checked for a compromise between model run time and vertical resolution.
We added p.19514 l.14: “[…] for 50 soil layers … with a uniform thickness of 0.05 m.“

And what about root water uptake? How was it simulated? This is clearly not a trivial task for an annual crop and needs further specification.

Modelling the root water uptake is indeed not a trivial task, we now provide a more detailed description in chapter 2.1.2, p.19515 l.22:

“Root water uptake in PMF equals a macroscopic approach type II (Feddes et al., 2001; Hopmans and Bristow, 2002). In this concept the actual water uptake is distributed over the rooting zone and occurs as a sink term in the Richards equation. The allocation of water uptake in PMF depends on the relation of the root biomass in each soil layer and the total root biomass in the rooting zone. Influences of water are incorporated according to the soil moisture conditions with a crop specific response function. The response function is related to the soil matrix potential and the water content. According to Feddes et al. (1978), the response function is determined by three thresholds defining oxygen deficiency, wilting point and optimal water uptake conditions. The crop specific response function includes a dimensionless water stress index. The resulting actual water uptake from each layer is the product of the stress index and the available water.

The root growth takes place during sowing and the development stage anthesis. During this part of the growing season, a fraction of the total biomass is allocated to the root. Root growth includes the calculation of the total underground biomass, as a fraction from the total plant biomass, the calculated vertical growth (elongation) and the distribution of the root biomass over the rooting zone (branching).”

References

Feddes, R. A., Hoff, H., Bruen, M., Dawson, T., de Rosnay, P., Dirmeyer, P., Jackson, R. B., Kabat, P., Kleidon, A. and Lilly, A.: Modeling root water uptake in hydrological and climate models, Bulletin of the American Meteorological Society, 82(12), 2797–2809, 2001.

Feddes, R. A., Kowalik, P. J. and Zaradny, H.: Simulation of field water use and crop yield, Wageningen: Centre for Agricultural Publishing and Documentation., 1978.

Hopmans, J. W. and Bristow, K. L.: Current capabilities and future needs of root water and nutrient uptake modeling, Advances in Agronomy, 77, 103–183, 2002.

(2.2) Several aspects concerning the soil water content observations need to be clarified as well. First, what was the measurement technique? Time Domain Reflectometry (TDR)? Gravimetric measurements? Were these measurements taken at a single location within a given plot or is it some sort of average of spatially distributed observations?

Second, the authors state that measurements were taken at 33 days between 1992 and 1998. But the period that was actually used for model calibrations contain only 15 measurements. So why do the authors mention the 33 observations then?
Third, the authors say that the measurements were taken at 0.15, 0.45, and 0.75 m depth. But the axes labels of Figs. 2 and 5 suggest that these measurements and corresponding simulations give the soil water content averaged over a 0.3 m depth interval each. These are indeed two quite different quantities, especially near the surface where gradients in soil water content can become quit large. So, what is the actual depth interval of the observations?

Soil moisture was measured gravimetrically (http://open-research-data.ext.zalf.de/ResearchData/1992_167.html). We added this and some more clarifying information at p.19520 l.14. Text now reads: "Average gravimetric soil moisture measurements are available for three soil depths (0 - 0.3 m, 0.3 - 0.6 m and 0.6 - 0.9 m) on 15 days during the observation period from 1993 – 1994 for every site."

(2.3) The authors mention that "behavioural parameter sets were found to be around 10% with respect to the measured value" (p.19523 ll. 7-8). But this is exactly what was expected because the threshold criterion used to distinguish a behavioural from a non-behavioural parameter set was set to 10% (p.19519 l.15). This essentially means that the GLUE algorithm worked in exactly the way it was intended to do. But it does not tell us anything about the goodness-of-fit or the ability of the model to describe the data.

This is indeed correct. We deleted this sentence.

(2.4) The GLUE method is known to be highly inefficient, essentially because the parameter sets are drawn randomly from the prior distribution and do not take into account the information gained so far from behavioural parameter sets obtained from previous model runs. This is especially true for high dimensional parameter spaces, which was corroborated by the present study. The authors point out correctly (p.19524 ll. 7-10) that, as an alternative, Markov chain Monte Carlo algorithms can be used to reduce the required number of model runs. But they incorrectly assume that these algorithms may be less robust with respect to finding the global optimal parameter sets. The DREAM algorithm mentioned by the authors implements a globally convergent search strategy that was especially designed to be capable to handle multiple local optima. Such an algorithm is very likely to perform much better in finding the region of the parameter space with behavioural models and in sampling this region adequately, leading to more robust uncertainty bounds of the parameters and the model predictions.

We choose the GLUE method because of its easy implementation into our complex coupled model concept. A further advantage was the possibility of parallelization of the model runs. We do not want argue that the GLUE method is better or worse than any other algorithms used to quantify uncertainty.

We changed the text accordingly (p.19524 l.7-10): "The GLUE method was chosen because of its easy implementation and its possibility of parallelization."

3 Technical corrections
(3.1) p. 19510 l. 9: The programming language that was actually used for model coupling needs not to be mentioned in the Abtract. It is a purely technical issue and not relevant to the results presented in this study.

We deleted the programming language from the abstract.

(3.2) p. 19510 l. 12 and p. 19514 l. 14: It is not only the retention curve that was parameterized using the van Genuchten-Mualem model, but also the conductivity curve! A better expression would therefore be "and the van Genuchten-Mualem model of the soil hydraulic properties".

Changed as proposed.

(3.3) p. 19510 l. 20: "drawn from an equally distributed parameter space" should be replaced by "drawn from a uniform distribution".

Changed as proposed.

(3.4) p. 19511 l. 16: "parameters" instead of "parameter".

Changed as proposed.

(3.5) p. 19511 l. 22: "et al." is missing in this citation.

Changed as proposed.

(3.6) p. 19514 l. 13: It should be "the Richards equation", not "the Richard's equation". It is named after Lorenzo A. Richards.

Changed as proposed.

(3.7) p. 19514 ll. 18-19: "we used existing climate data" should be replaced by "we used meteorological data". If they would not be existent, the data could not have been used anyway. And to be precise, these type of data are meteorological data, not climate data (which is some sort of long-term average of meteorological data).

Changed to: "To initiate the water content of CMF we used meteorological data for the year 1992 and calibrated it for the growing season 1993/1994."

(3.8) p. 19515 l. 22: The parameters that appear in brackets were not defined previously, nor is reference made to Tab. 1 where they are defined.

p.19515 l.22 changed to: "The last group of parameters are the basal crop coefficients (kcb_{ini}, kcb_{mid} and kcb_{end}) which are used to assess plant transpiration from potential evapotranspiration (Table 1)."

(3.9) p. 19516 l. 22: It is not the "probability distribution of the measurement errors" but rather the "likelihood function of the model residuals". Measurement errors and model residuals are the identical if, and only if, the model would be perfect. But this is – almost by definition – never the case.

To stay in line with Beven and Biley (1992) we changed our text to: "The choice of the likelihood function depends on the situation and is often subjective (Beven and Binley, 1992). Nevertheless, the choice of only one objective function for the calibration is in most cases inaccurate (Vrugt et al., 2003)."

(3.10) p. 19517 l. 3: The minus sign before the summation should be removed.

Changed as proposed.

(3.11) p. 19519 ll. 1-2: Is the Python package (with reference) really worth mentioning? I mean, sampling from a uniform distribution is a rather trivial task in any programming language and needs no further specification.

We agree on the referee's opinion and changed p.19519 l.1-2 to "A random function was used to create 2 x 10^6 parameter sets, […]"

(3.12) p. 19522 ll. 9-10: The similarity of the results for the n parameter of the current study and the study of Ippisch et al. cited by the authors is completely irrelevant. The optimal value of n strongly depends on soil type and bulk density, among other soil properties. Ippisch et al. investigated a different soil that has – by chance – a similar n value.

We deleted this part of the discussion.

(3.13) p. 19523 ll. 22-23: The percent signs should be removed. The unit of soil water content (and also RMSE) in these studies was not percent, but given on a volume by volume basis.

Changed to: "Scharnagl et al. (2011) found a RMSE of 0.009 water content and NSE of 0.87 for their Hydrus-1D model set up at the site Selhausen near Jülich, Germany."

(3.14) p. 19527 l. 27: It is not clear to me what the authors mean by "our implementation of iterative steps". What exactly was the iterative part in this study? Please add some explanatory information here. Or alternatively, rephrase this sentence.

We refer at p.19527 l.27 to our defined consecutive steps in section 2.2.2 (p.19518 l.16-p.19519 l.16): "[…] our implementation of four consecutive steps in implementing the GLUE method (section 2.2.2) for the validation of our coupled model […]"

(3.15) p. 19535: In my view, the figure might become much clearer if only the behavioural parameter sets would be plotted. I mean, what is the added value of showing the non-behavioural parameter sets? What information is this supposed to convey?

Figure changed as proposed. -> new Figure 1

(3.16) p. 19536: For consistency, the model efficiency should be abbreviated with "NSE" (see the legend).

Figure changed as proposed. -> new Figure 2