Reviewer #1:

The authors represent a modelling development that makes it possible to have a more complete look at the global carbon cycle. They combine the well-established JULES model with a newly developed model for DOC, including soil carbon processes and leaching. This manuscript is therefore an important step towards a full carbon cycle understanding.
In general I think that the manuscript is well structured and the figures are helpful to understand the outcomes. However, there are some changes needed to make it more convincing. While I see some issues that should be clarified/solved first, I recommend publishing the manuscript in GMD after revision.

Thank you very much for your careful comments. We improved the manuscript following the reviewer's suggestions. Details are given below.

1. Please explain why the authors did not include the production of OC in soils and rivers (i.e. aquatic photosynthesis)?

Let us clarify what the model does and doesn't include.
The developments presented are made on the JULES, the Joint UK Land-Environment Simulator. The standard version of JULES, described in Clark et al. (2011), simulates vegetation and soil water and carbon processes, including the production of soil organic carbon following plant mortality, litter decomposition, respiration etc.
Our manuscript describes new developments in JULES in order to represent DOC cycling in soils, including DOC production is soils, DOC decomposition and respiration, as well as leaching of DOC from the soil to the aquatic environments.

However; the reviewer is right to say that we do not include fate of OC in rivers, this is beyond the scope of this work, as it would need to represent the full biogeochemistry of DOC in rivers (biological activity, sedimentation, gas transfer, etc.). Our developments lay the corner stone for a future representation of C cycling along the land-to-ocean continuum, including transport and transformation in the river network, as highlighted in our introduction and outlook. This is clarified now in the introduction.

2. Explain the additional value of adding Turkey Point and Guandaushi. To me it seems that the data from these two sides do not add much information, due to their much shorter time coverage. Also in the discussion, it is stated that e.g. Turkey Point is not really useful because it's located at a site that was agriculturally used until 1989. Possibly it would help to remove the Level 2-sites.

There is always a balance to find between very few sites that have well documented long-term measurements, and more sites, with shorter record and/or more complex history. In this work, we found that focusing on only 3 sites, all from Western European forests, would not suffice to gain confidence in our model, hence we decided to also include Turkey Point-89 and Guandaushi sites that we define as "level-2 sites", in order to explore broader climate/ecosystems domains. Turkey Point-89 provides a high productive forest ecosystem on a previously agricultural land. This is a typical system in North America where marginal agricultural land which have been afforested in recent decades. Also, we are keen to keep Turkey point as it is the only site that provides DOC measurements down to 100 cm, providing additional constraints on our model.
We find that our model has some difficulties in capturing dynamics at Turkey Point site. However, while we overestimate the total DOC in Turkey Point, we are able to reproduce the vertical profile of DOC in the soil.

3. Discuss the different methods in the study side description. Are there fundamental differences between 'suction cups' (p.8 l.22) and 'tension lysimeters' (p.8 l.32)? Are the observed values comparable?

First of all, we agree with the reviewer #1 that this is an important point which needs further clarification. For all sites, samples were collected using one of the two in situ soil solution extraction methods, namely suction cups or suction plates. We have, therefore, removed the term "lysimeter", which is often used incorrectly, and have replaced it with one of the two afore mentioned extraction methods. Where appropriate we have also included information on the material of the sampling device as well as the suction applied. This will allow the reader to better estimate the potential effects of different sampling methods on DOC concentrations obtained (Weihermüller et al. 2007). This is clarified in revised text:
For Hainich:
"At this site, soil solution samples were taken at three depths (5, 10 and 20 cm) using ceramic suction plates positioned at four different plots within the site. Samples were obtained by applying a tension of 100 hPa after each bi-weekly sampling occasion"

For Carlow:
"Samples were obtained by applying a tension of 400 hPa after each bi-weekly sampling occasion (Walmsley 2009)."
For Brasschaat:
"DOC samples were collected at three horizons of Al/Ap, A/E and Cg (Soil Classification Working Group 1998) referred to 10,35 and 75cm depth, by means of ceramic suction cups on a biweekly interval. Two days prior to sample collection a tension of 600 hPa was applied to each suction cup. Samples were collected at three locations and pooled into one composite sample per layer for analysis (Gielen et al. 2011)"
For Turkey Point-89:
"DOC sampling was attempted in monthly intervals at three depths of 25, 50 and 100 cm by means of porous suction cups..."
For Gunadaushi:
"DOC samples were collected at three depths of 15, 30 and 60 cm in three locations at bi-weekly interval by means of ceramic suction cups"

As samples were obtained by means of suction from a conical or plate-like porous sampling device we consider the differences due to methodological issues to be small compared to other soil solution collection methods such as passive samplers (pan lysimeters, wick samplers, resin boxes) and in particular extraction methods (centrifugation, chemical extractants such as KCl or CaCl2). The main difference between suction cups and suction plates is that the latter exhibit a larger sampling area and due to the 2D surface of the plates, the origin of the sampled solution is better defined which is important for mass balance estimations. It has also been shown that ceramic devices can absorb dissolved organic matter. As a result, concentrations in all but Carlow (where more suitable non-absorbing porous glass suction cups were used) may be underestimated. In addition, the suction applied will have an effect on the source of the soil solution collected (macro-, meso-, micropores). We are unable to quantify these effects in the current study. However, as mentioned above, we believe that the effect of different sampling methods on DOC concentrations are relatively small compared to potential differences due to sampling effects. Finally, as we are comparing observed values with the model simulation of Soil DOC concentration (not comparing observed values with each other), the difference in the sampling methods should not have a significant impact on the comparison of the model vs observed values.

When did this sampling take place (e.g. Turkey Point – 'samples could only be retrieved for 5 separate days'; all in summer? or winter?) (see also comment above (2.))

In Turkey point-89, due to the highly drained soil because of its sandy and dry condition, the sampling was taking place in spring and autumn when the soil was wet enough. Hence the dates of sampling were: 29-Nov-2004, 3-May-2005, 16-Jun-2005, 29-Jun-2005, 14-Oct-2005 (for more detail please see (Peichl et al. 2007)).
This is clarified in the revised manuscript:
"however, due to the dry sandy soils, samples could only be retrieved for 5 separate days of sampling after heavy rain fall events on 29-Nov-2004, 3-May-2005, 16-Jun-2005, 29-Jun-2005, 14-Oct-2005 (Peichl et al. 2007)."

4. Elaborate on the model input in more detail:
a) Have the data of FLUXNET and WATCH been somewhat corrected to be comparable?

No. These two data are not comparable. The FLUXNET data are the site-specific observed data. Hence, whenever we had access to the site measurements, we used them. The only site which we had to use WATCH data was Guandaushi, where we could not find any on-site measurements. Both forcing data were checked for any missing data and it was gap filled by linear interpolation. This is clarified in the revised manuscript:
"The FLUXNET database provides on-site meteorological data for each site that could be used as forcing for simulations in JULES, However, we had to use the global WATCH dataset (Weedon et al. 2010) as forcing for Guandaushi site where no on-site data was available. However, both forcing data were checked for any missing data and it was gap filled by linear interpolation."

b) Mention the names for the parameters as it's used in the tables and equations consistently (e.g. p.9 l.28: bulk density and clay content)

Corrected.

5. Please adapt the figures in a way that makes them easier to understand.
a) Fig.2 The extent can be smaller to better see where the sites are.

Corrected.

b) Fig.3 Decrease the y-axis range. A maximum of about 20 should be sufficient and the differences a better to see.

Corrected.

c) Fig .4 Decrease the y-axis range to a maximum of 40.

Corrected.

6. Combine both parts of the discussion to one. This would avoid the repetitions and can clearly combine all information/discussion on each of the sites.

Done as suggested, thank you.

Minor comments:

a) Define SOC at its first occurrence (p-2 l.28)

SOC first occurrence was defined in p-2 l.8.

b) Please use consistent and not confusing naming for the variables/parameters in the equations (e.g. R can mean 'run-off' and 'respiration')

Corrected. Runoff modified to Roff instead of R.

c) Connecting the equations to the arrows in the flow chart (Fig.1) would help in understanding the calculations. What do the numbers (1) to (12) in Fig.1 mean? They don't seem to match with the equations.

Sorry, this was unclear. Numbers were corresponding to process not to equations. As all arrows are defined within the text at the end of each process described (i.e. p.5 l.10: "(F_P; arrows 1-4 Fig. 1)" indicating the process linked to the arrows in the model figure.)
In order to avoid potential confusion, we replaced the numbers in fig.1 with letters.

d) What is the spatial resolution?

The evaluation of model was performed on plot-scale, using climate forcing data, soil and land cover for one specific point, so no horizontal spatial dimension was involved. The model is however capable of running at each spatial resolution for which forcing data are available.
This is clarified in the revised manuscript (p.10, l.15):
"The evaluation of model was performed on plot-scale, using climate forcing data, soil and land cover consistent with the site, no horizontal spatial dimension was involved."

e) Make unit naming consistent (e.g. Kg C m-2 day-1 vs. kg C m-2 day-1, p.5 l.10 vs. l.15)

Agreed. We now use "kg" throughout the text.

f) I suggest to rename 'Carbon concentration and fluxes' to 'Validation of carbon concentration and fluxes'

Corrected.