

## ***Interactive comment on “A global database of soil respiration data” by B. Bond-Lamberty and A. Thomson***

**Anonymous Referee #2**

Received and published: 2 March 2010

I reviewed the manuscript entitled “A global database of soil respiration data” by Bond-Lamberty and Thomson. This is a great example of data sharing for meta-analyses where the authors have posted the data on repositories and also provided the opportunity to correct and update the dataset. I think this is a great compilation with multiple potential uses. I have included some minor comments.

I believe is critical to clarify how the citation of this database should be done. I know that the ORNL DAAC has a data citation policy, but I was unable to find the SRDB at the website. The authors should provide the full ORNL DAAC citation including the DOI that is usually associated with all datasets at this website. In addition this dataset is in fact a “1.0 release” so it must be explicitly stated when referring to this version. This is not a problem for the static ORNL DAAC but it could potentially be for the dynamic

C144

version.

This brings me to the dataset stored in Google Code: <http://code.google.com/p/srdb/>. This is a great experiment to test the willingness to contribute and update this dataset by the broader scientific community. However I have several questions that I was not able to answer using the tools provided in Google Code. How the future versions are going to be handled? Does the authors are going to wait until substantial changes are done to release a V2? or it will be a constant update and will be stored using a reference number, URL, and download date? This is not clear in the current version of the manuscript or at the Google Code website. A few examples in the manuscript and/or the wiki page will be enough to clarify these points.

The second comment is to clarify the “definitions” for the variables used (page 1325 lines 10-14). I believe it should be explicitly stated how each one of these variables were calculated or extracted from the literature. Some of the information is included in the current version of the manuscript, but I would like to motivate the authors to be more explicit. For example: Does annual  $R_s$  includes a minimum of measured days (e.g. >300 days) or is just a mean annual value (based on an unknown number of measurements)? How does mean seasonal  $R_s$  was defined? What does a season represent (e.g. growing season or just summer or depends on the ecosystem)? How does annual or seasonal partitioning of  $R_s$  sources was define? How does Q10 and associated temperature range was define and at which soil depth was calculated? I believe is important to provide enough information to interpret the dataset and to instruct the community on how to participate.

In results section page 1327 lines 21-23. How was calculated the 41% explained variation? I assume it was done via a multiple linear regression analysis. Please include how was this performed and if a stepwise or other type of multiple linear regression was used. Do the authors have any comments on the frequency distributions presented in Figure 3? I see that boreal sites are skewed in comparison to a near normal distribution in tropical sites regardless of a similar sample size (N=322 vs N=264). Similarly

C145

why temperate regions do not show a normal distribution despite the large number of measurements (N=1598)?

I know that the goal of the paper is to show global relationships, but it would be interesting if the authors also provide a table with the independent slopes and regression coefficients for each vegetation type in Figure 4, 5 and 6. Then the slopes could be tested for significant differences to provide more detail information about the dataset.

---

Interactive comment on Biogeosciences Discuss., 7, 1321, 2010.