Review of Discussion Paper www.geosci-model-dev-discuss.net/5/3573/2012

„ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1-km resolution based on satellite information for user in land surface, meteorological and climate models

by: S. Faroux, A. T. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. Le Moigne

General comments:

The authors introduce a new dataset (ECOCLIMAP-II) of land cover and related surface parameters (like roughness length, albedo, root depth, and others) for Europe. Such datasets are mandatory for numerical weather prediction and climate models, and as the authors point out, need to be updated from time to time.
The derivation of the new dataset is based on two vegetation related variables, the leaf area index (LAI) and the NDVI, which are both satellite derived, allowing rather high spatial and temporal resolution of the final product. An advantage of the new dataset is that it is based on multi-annual consistent satellite observations, whereas products like ECOCLIMAP-I refer only to single annual cycles of satellite data
The authors point out that one demand of the new dataset is a better regional characterization of land cover in accordance with the climatic environments of the different regions in Europe (latitude, continental or marine climate, topography).
Of course, this could help to improve climate simulations for climate sensitive regions like the Mediterranean.
Already from these points of view the paper it worth to be published
The authors also perform validations exercises using three different databases. Unfortunately, the validations using the high resolution French dataset AGRESTE and the high resolution satellite product FORMOSAT-2 concentrate only on France. A validation for other region in Europe would be desirable. I doubt a little bit whether the comparison with ISLSCP2 C4 data, which have a rough resolution of 1°, is of real value in the sense of validation.
The authors also present a comparison of the new ECOCLIMAP-II dataset with ECOCLIMAP-I, which is certainly necessary. But only the differences between the two products are described. But this does not tell the reader and the potential user of the new dataset whether it is really better than the old one. For instance, is the higher representation of broadleaf trees in central Russia and their tendency to disappear from the Mediterranean regions more realistic than in ECOCLIMAP-I (Section 5.3.2, Fig. 11 d)? Or is the complete modification of the distribution of wetlands in ECOCLIMAP-II (Section 5.3.2, Fig.11l) more realistic? These questions can be summarized by the more general questions: what is the impact of the new dataset when being applied in NWP or regional climate simulations? Can the modeller expect an improvement of his results?
Solely from reading the paper the potential user of the new dataset cannot answer these questions.
In addition, for a modeller, who perhaps is not an expert in the generation of land cover datasets, the paper is not easy to read and to understand. Especially, the implementation of the ECOCLIMAP-II database (Chapter 4) is hard to read.

In summary: the paper describes an important study which should be published.

However, in order to increase the value of the paper the authors should try to describe their methodologies in a more understandable manner, understandable for interested readers who are not experts in the author's research fields. Although there are a lot of references, the reader should be able to follow the methods without reading additional literature. The authors should spend some effort to explain briefly the different terms, and their possible relationship. For instance, what is the difference, respectively the relationship, between land cover classes and PFTs? There are some more questions and suggestions in the specific comments.

Specific Comments

1. I have the impression that sometimes terms like "surface types" and "land cover classes/categories" mean different things, sometimes they are used synonymously. For example, in the abstract (line 17 and 18) and in Chapter 2 (page 3578, line 24) the authors talk about "4 main surface type or tiles (sea, inland water bodies, urban areas, natural land areas). On the other hand, they define 14 categories of surface types (Section 3.1, page 3581, line2 and 3) when combining GLC2000 and CLC2000 maps. There are other examples where terms like "classes" or "categories" or more general "land cover" are used, and it is not always clear to me whether they denote the same or different things.

2. Page 3582, line 26: I don't understand the first part of the sentence "This revealed conspicuous agreement between the two climate maps …". What I can see in fig. 2 is the "boundary" between the two datasets. In Fig.2, page 3611, I miss a legend.

3. I have the most difficulties with the description how the ECOCLIMAP-II dataset is really created. For me it is not clear
   a. Which data sets are "disentangled" by the K-means method?
   b. What is classified by the clustering algorithm (the objects to be clustered) and according to which attributes the objects are clustered?
   c. What do the resulting clusters represent? Land cover? Vegetation types? Does a cluster represent a specific geographical region? Or is it possible that a cluster contains information from different regions?
   d. Reducing the numbers of clusters: which "mean 10-day NDVI profiles" (page 3585, line 23)? Where do they come from? From the clustering? Without some basic understanding of the clustering process and its outcome the rest of chapter 4 is also hardly to understand.
   e. Also the role of the C14 map is not clear to me? Does it serve as a kind of "true reference" which, on the one hand, is used to verify basically whether the results of the clustering make sense, on the other hand, should be refined by the new dataset?
   f. Page 3587, line 4: "Climate maps were finally used **to avoid the segregation** of pixels belonging to different climate units …..". I don't understand this. I understand the usage of the climate map in such a way that one wants, for instance, to prove whether a pixel representing the continental climate over Russia has not been assigned to a land cover class belonging to marine climate along the Atlantic coast.

4. Section 4.4 "Defining the surface parameters" is also hard to understand. What do the authors mean with "a thorough interpretation of the (combined 14?) CLC2000/GLC2000 classes appearing in a given cover" (page 3587, line

18)? What is the difference between these CLC2000/GLC2000 classes and the covers? Do I understand this section correctly that only temporal LAI profiles are determined by the procedure? The values for other surface parameters (root depth, total soil depth, tree heights) are taken from ECOCLIMAP-I due to lack of better and more reliable data sources. What is the "total soil depth"? What's about other parameters like roughness length, emissivity, albedo?

5. Page 3587, line 23: there is no assessment of functional type fractions in Sect. 3.2.1,since Sect. 3.2.1 does not exist.

6. Section 5.1: To my opinion this section is simply a description of the resulting ECOCLIMAP-II map, not a validation. Please, create a new chapter.

7. Section 5.1.1 Forests, page 3589, line 19: "Clearly, permanently **cool** temperatures coupled with sunny days…..". Is "cool" really correct? If yes, why?

8. Section 5.1.2, page 3590, lines 16 to 18, description of Fig. 5l: I don't see a shift of the peak towards winter. The peaks are very similar to Fig. 5k, except that the minimum values in Fig 5l are at a slightly higher level. In Fig. 5 (page 3614) the beginning of a year is hardly to identify. Perhaps, the length of the corresponding tick can be increased a little bit.

9. Section 5.2.1, Comparison with AGRESTE:
    - AGRESTE data are used to calculate "observational" PFTs that correspond to the PFTs of ECOCLIMAP-II, right?
    - The AGRESTE data are given in hectares (Section 3.5, page 3583, line 9). Which data set is interpolated, ECOCLIMAP-II to AGRESTE or vice versa?
    - What are the "representative fractions of the covers" (page 3593, line 11)? Does "cover" mean "PFT"?
    - The weighting using the representative fractions: which PFT fractions are weighted with which representative fraction?

10. Section 5.2.3, page 3596 line 5 and 6: why do urban areas belong to the PFT? In section 2, page 3578, line 27, the PFTs are defined without urban areas. This is again an example of the steady mixture of definitions. Please, avoid this! It confuses the reader.

11. Section 5.2.3, page 3596,line 7: what are the "12 most representative ECOCLIMAP-II and FORMOSAT land covers"? How are they determined?

12. Fig. 9, page 3618: Please, explain the abbreviation "ecov2" and "fms" (figure legends) in the figure captions. Other curves in yellow and light blue (C4 crops, grassland, urban areas) are neither explained in the figure caption nor discussed in the text.

13. Section 5.3, Comparison with ECOCLIMAP-I: This section is by no doubt necessary. But the simple description of the differences does not really help the potential user to judge whether the new dataset is more realistic than the old ECOCLIMAP-I dataset. The authors should spent some more lines in order to argue why one should now use the ECOCLIMAP-II data.

14. Page 3600, line 13: Please insert "(Fig. 11g)" after "The fractions of C3 crops"

15. Caption of Fig 11, page 3620: please reorder the enumeration of vegetation types according to the alphabet.

Technical corrections

1. Page 3575, line 27: insert a blank between "and" and "red"

2. Page 3578, line 3: delete the dot after "resolution"
3. Page 3587, line 23: once more, Sect. 3.2.1 does not exist
4. Page 3599, line 27: should be "(Fig. 11e)", not "(Fig. 11b)"