Biogeosciences Discuss., 12, C6402–C6410, 2015
www.biogeosciences-discuss.net/12/C6402/2015/

Biogeosciences Discussions Open Access

Interactive comment on "Quantifying regional, time-varying effects of cropland and pasture on vegetation fire" *by* S. S. Rabin et al.

S. S. Rabin et al.

srabin@princeton.edu

Received and published: 11 October 2015

We thank Referee 2 for the valuable comments and suggestions. Our responses to each point follow. (Please note that we plan to convert our Appendix to a Supplement.)

Please add numbering to the equations.
We will do so.

In equation 2, (line 4 on page 10822) you optimize $\widehat{F_{k,i}}$ where k stands for c, p or o. However, you minimize the sum of squared errors for each analysis region (or group of grid cells), let's say r, and not for each grid cell i. It may thus be better to use

something like where r is the analysis region (or cluster of grid cells), and k may be c, p or o.

Thank you for finding that error; we apologize for the confusion. The symbols referred to in equation 2 were not supposed to have had the i subscript, so we will correct that. We will likely not add an r subscript, instead just leaving off the i subscript and using the "hat," which together we feel sufficiently distinguish the region-month Fs from those referring to individual grid cells.

To continue, in your first equation on page 10821 $\widehat{F_{k,i}}$ stands for the "fraction of that land-use type that burned in that grid cell". However, when optimizing in the next equation (first equation on p. 10822) $\widehat{F_{k,i}}$ values represent a different thing: the slopes of multiple linear regression between the spatial distribution of BA and the three land use types. This means that $\widehat{F_{k,i}}$ cannot be interpreted exactly the same as $F_{k,i}$ and is not "the fraction of land-use type that burns across the region". It would help the reader if you would redefine and clearly state what $\widehat{F_{k,i}}$ stands for, and then if you want to explain it in a less technical way you could state "this is related to the fraction of land-use type that burns across the region".

The $\widehat{F_k}$ values are indeed analogous to the unknown slopes that are solved for in multiple linear regression. They tell us that, for example, in a region with $\widehat{F_p} = 1$, a grid cell with 110 ha of pasture (and equal amounts of cropland and non-agricultural land) will have one more hectare of burned area than a grid cell with 100 ha of pasture. However, this analogy does not preclude interpretation of the values as best-guess estimates of region-wide fractional burning of each land cover.

As written in the paper: "The values of each $F_{k,i}$ are unknown, but a best-guess $\widehat{F_k}$ can be estimated across a group of N grid cells" (p. 10822, lines 2–3). That is, because we don't know the value of each $F_{k,i}$, we instead calculate $\widehat{F_k}$, which is one number that we plug in as a best-guess estimate for each $F_{k,i}$ (i.e., into Eqn. 1: p. 10821, line 25).

We can put this conceptualization into equation form. Consider the total burned area across a region (B_r) as the sum of the burned areas in each grid cell i:

$$B_r = B_1 + ... + B_i + ... + B_N$$

$$B_i = F_{c,i}A_{c,i} + F_{p,i}A_{p,i} + F_{o,i}A_{o,i}$$

$$B_r = \sum_{i=1}^{N}(F_{c,i}A_{c,i} + F_{p,i}A_{p,i} + F_{o,i}A_{o,i})$$

This is essentially where we are with Equation 2 in the manuscript—we don't know every $F_{k,i}$, so we need to find best-guess estimates:

$$B_r \approx \sum_{i=1}^{N}(\widehat{F_c}A_{c,i} + \widehat{F_p}A_{p,i} + \widehat{F_o}A_{o,i})$$

In the manuscript, we then describe the method by which we find these best-guess estimates (using an algorithm, unsurprisingly, that is commonly used by statistical software to fit multiple linear regressions!). But we can go a bit further here:

$$B_r \approx \widehat{F_c}\sum_{i=1}^{N}A_{c,i} + \widehat{F_p}\sum_{i=1}^{N}A_{p,i} + \widehat{F_o}\sum_{i=1}^{N}A_{o,i}$$

$$B_r \approx \widehat{F_c}A_{c,r} + \widehat{F_p}A_{p,r} + \widehat{F_o}A_{o,r}$$

Thus, the $\widehat{F_k}$ values do indeed represent best guesses of fractional burning of each land cover type across the region.

The second definition (P10823 L12) "the net effect of land use k on fire in the region, expressed as a fraction of the area of land use k in the region" is confusing.

We will edit/add text to clarify this idea.

The negative slopes are interesting and I agree that they may well represent a real aspect of the system. But with this definition you undermine your more realistic interpretation that a certain land use may affect fire activity in different land use classes in the vicinity (often within the same 0.25° grid cell).

We will reiterate, in the same paragraph as the text referenced in the above comment, that the negative effect of some some land use is indeed felt on other land uses.

The Methods of GFED are explained by van der Werf et al., 2010, and adjusted by Randerson et al., 2012.

We will add a citation of van der Werf et al. (2010) in addition to the reference to Giglio et al. (2010) when discussing GFED3 and GFED3s.

Figures 2 and 3. It would be easier for the reader if you would just present the annual burned area and carbon emissions split up by the different land use types (k) here.

In our estimation, the graphical presentation in Figures 2 and 3 makes it easy for the reader to extract important information at a glance. For example, one can get a sense of interannual variability more easily than if given a mean ± s.d. For another example, one can easily grasp that although at a global scale (and for many regions) there is almost as much burned area associated with pasture as with non-agricultural lands, pasture fires are associated with much lower emissions levels. Thus, we feel it is important to keep the time series plots. However, for readers interested in the numbers behind Figure 3, we will add a table to the Supplement.

Then you can remove the "cropland and crop+ categories" which are confusing and *just use one "total".*

It is important to distinguish between what the remote sensing algorithm classifies as "cropland" and "cropland-natural mosaic" because, when fire occurs on mosaic, we cannot be sure whether it is the cropland that is actually burning. We thus might expect the actual amount of burned cropland to be more than that on "cropland" but less than that on combined "cropland plus cropland-natural mosaic." The latter is what we refer to as "Crop+" in the figures. To reduce confusion, we will add text explaining "Crop+" to Figure 3.

It is not clear what applying the "model" adds to the results presented in these two figures opposed to a simple estimate of burned area per land use type "crop, pasture, natural, total".

The total estimated ("model") burning is included in these figures for comparison with the total observed burning, as a first-order check of the results. If the estimated and observed total burning differed greatly, then the reader would know to be extra cautious in interpreting the partitioned fire activity.

The authors provide little insights in the model performance and background data-sets. How much of the spatial variation in burned area can be explained by the distribution of the three land use classes for each analysis region (e.g., r-squared)?

We will add, to Section 4.3, some discussion of overall R-squared as it relates to the amount of variation in burned area that can be explained by land cover distributions.

What are the actual $\widehat{F_{k,i}}$ values per analysis region? And what does the land use distribution look like? It would be interesting to see some of these figures either in the main body of text or in the Annex material.

We will add a table of $\widehat{F_k}$ values for every month and region to the Supplement, along with maps of mean land cover distributions.

If you think the paper is becoming too long you could merge figures 2 and 3, or remove all the current annex figures and make your point about the current interpretation of negative $\widehat{F_{k,i}}$ a little stronger by citing more literature.

We will likely not merge Figures 2 and 3, but will add more literature supporting the idea of negative $\widehat{F_k}$ values to the paragraph where that idea is introduced (p. 10822).

Fig. 4, this is an interesting figure. You state that "Numbers can be interpreted as ..", but for the reader it would be easier if you first state what the numbers actually are, something like "the maps show times the area of k". Then in the next phrase you can say, "this can be interpreted as..". Many people have knowledge about these type of models, providing such information makes it easier for them to interpret your results.

We will add some text to this effect to the captions of Figures 4 and A3.

Figures 5 and 8, it seems that many of the regions are resolved as "mean" values, and that the different land use types provide only limited information on the spatial distribution of annual burned area within the analysis regions. This may be a consequence of: (1) Burned area in many analysis regions is dominated by a single land use. (2) In many cases little of the spatial distribution of BA can be explained by land use. Both will make the values converge to the mean. This should be more clearly discussed.

We will add some discussion of this phenomenon to Section 4.3.

Figure 7a, please explain the meaning of the three colors in the caption of the figure. Figure 7b, It may look nicer if you would delete the white space on the x-axis before "August".

We will make these changes.

Figure 8b. This figure is a little counter intuitive now. First you say "(b) each grid cell", but then the fit and the equations are presented for binned-mean values. First, depending on your bin-size the slope and r-squared will vary, which makes the results subjective. Second, you have already presented the "binned" results in Fig. 8a (using analysis regions for bins). It would be interesting to read here how much of the spatial variation in burned area is actually captured at the grid cell level by your model.

We will add text to the captions of Figures 8 and A7 clarifying that bins are being used. We do believe, however, that both subfigures A and B are necessary. 8a/A7a illustrate the point that the algorithm is very good at estimating total burned area across a region, while 8b/A7b show that much less of the variation at the level of individual grid cells is captured. We will add more discussion of the results presented in these figures, including R-squared of gridcell-level variation, to Section 4.3.

For the reader it may be easier if you better separate the actual results and the discussion. Some of the results section reads more like a discussion while the discussion is sometimes very technical, how do your results relate to other literature?

While we acknowledge that the paper is structured somewhat unconventionally, we believe that this format is preferable for the material we present. Because of the novel nature of our method, our results must be interpreted carefully, and so we decided that it would be helpful to provide more guidance and interpretation in the results section. As we edited the paper towards that goal, it became apparent that it would be more concise to avoid splitting topics across the results and discussion sections. For example, we already had some examples of fire suppression in the results section to help the reader grasp "negative burned area"; it felt artificial and disjointed to then come back to that topic in the discussion section. We thus reserved the discussion section for more general treatment of concepts that arise from our results (4.1, 4.3), and for the consideration of input data quality (4.2).

All that said, the point asking for more comparison of our results with the existing literature is well taken. We will add more references to and discussion of previous

work at various points throughout the Results and Discussion sections.

4.3 Impacts of regional analysis: This is an interesting discussion. Poor performance for Europe seems to be mostly a matter of Europe having many fires in all three land use classes while their spatial distribution may provide little information on the distribution of these fires. It may help to better discuss the differences in land use management between different areas. For example in Eastern Europe and Russia agricultural fires might be common practice but similar fires will not be found in Spain or Italy.

We will include more concrete examples here, citing for example the work of Lin et al. (2012; *Ecological Applications*) and Leff et al. (2004; *Global Biogeochemical Cycles*).

In a similar way, it will be hard to compare pastures across the world. Some of the grazed savannas will appear so close to natural vegetation that the a measure of "livestock density" may be more useful than "grazed or not grazed". On top of that, what about the naturally occurring herds of herbivores, especially in Africa. The FAO has published an interesting map of global livestock density.

Savannas can indeed have their fire regimes affected by heavy grazing pressure. We note, however, that what is important to our paper is isolating the effects of land management on fire. People absolutely do manage non-agricultural land, often through the use of fire and sometimes for hunting wild grazers. Understanding the regional and global scope of such management would be an interesting line of research, but is beyond the scope of this paper, where we have chosen to focus solely on lands with crops or livestock.

However, this comment gets at another important idea – how blurred the line can be between "pasture" and other land in some areas, and how that makes the FAO statistics a bit cloudy. We have already alluded to this in Section 4.2, but will some pasture-specific discussion and specific examples there.

The good performance of the model for Boreal Asia and for analysis regions where nearly all fires occur in a single land use type is obvious because the model just represents the mean values of the observations for that land use in the analysis region. The high number of analysis regions where this is the case is partly a consequence of the way the authors have defined the analysis regions in the first place. The authors should better acknowledge/discuss this. Now you state "Another, more general consequence .. in the results". Here it would really help if you would have presented how much of the spatial variation in burned area could actually be explained by land use. And then just state something like "On average, only xx% of the spatial variation in burned area could be explained by land use, hence for many of the regions the values simply represent the mean burned area for the given land use in the analysis region." Finally, a short discussion of the alternative sources of spatial variation in burned area might be helpful.

We will add some discussion of the importance of a region including examples of grid cells with a wide range of values for each land cover type. We will also discuss the possibility that one land cover type might dominate the signal, which could for example explain the good performance for Boreal Asia in terms of total fire despite its containing several large analysis regions.

Interactive comment on Biogeosciences Discuss., 12, 10817, 2015.