9.12.2019, Frankfurt am Main

Dear Dr. Kirsten Thonicke,

We thank the reviewers for constructive and thorough reviews of our manuscript. We also thank them for the encouraging words about our work.

Below we have added point-by-point response to their suggestions including those four points highlighted by you. We have replaced the word risk with hazard to better reflect the content of the paper, as we are dealing less with the damage from fire.

Angelica Feurdean

Department of Physical Geography,Goethe University
Altenhöferallee 1, 60438, Frankfurt am Main, Germany
& Biodiversity and Climate Research Centre (BiK-F)
Senckenberganlage 25, 60325, Frankfurt am Main, Germany

Email:angelica.feurdean@gmail.com
Feurdean@em.uni-frankfurt.de

Dear Angelica, thanks for submitting your detailed author responses to all reviewer comments. I have a few minor things to comment which I would ask you to consider additionally in your revised manuscript:
Reviewer 2.
1. Response to comment re Line 94 missing, page C3. please add.
2. your response to the comment on line 218: Please add description of how much the ecoregion boundaries were spatially shifted in 8ka and how that possibly (a precise quantification is impossible, I assume) could have influenced the results.
3. your response to line 224 comment, please add sentences commenting on the point raised ("so in future work it might be better to not aggregate the monthly time series into seasonal averages") in the discussion.
4. your response to Line 333 comment: Please make clear in your revised sentence that "quantitative climate reconstructions are truely rare in the region". This is what is also meant by the reviewer.
Thank you very much,
Kirsten

REVIEWER COMMENTS:
Reviewer #1 (Remarks to the Author):

Feurdean and collaborators present an extremely interesting work focusing on the role of different factors in determining fire risk. They use a newly curated dataset and GAM to assess the role of climate, tree cover and forest composition for the last 12ka in central - eastern Europe. I really enjoyed and learnt reading this manuscript that addresses a critical question on the domain of palaoefire but also in modern day fire ecology under the present Global Change scenarios: what might have been the most important factor determining fire hazard in the past? Not only the question is interesting but the data compilation is a really important effort in data curation and synthesis. Really very good job! I really like that you make your code available (despite I have some questions regarding that - see below). I have some general and particular comments, that I do not consider by any means critical to prevent this manuscript to be published, but that I would like the authors to address before the final version is submitted.

INTRODUCTION

Lines 123-127: I only have one comment in this section in connection with how we use the term "negative effect" when referring to fire risk and especially spread in broadleaf forests. I would not say that temperate forest have a clear negative effect on spread. With accrued drought and high plant density, equally intense fire risk and forest fires can happen in both oak and pine forests for instance, which can be even more difficult to extinguish owing to the calorific power of temperate forest tree wood. Despite conifers have volatile compounds and that increases fire risk, in current day forestry engineering assessments it is clear that some drought thresholds and biomass accumulation may trigger fires with that very same probability (see some references below). Pine forests and shrublands do not only burnt more because of their prevalence but because of fire selectivity to these land cover types regarding temperate forests. - see forest composition assessment on fire risk at a Mediterranean-Atlantic area as Portugal: Fernandes 2009 https://www.afsjournal.org/articles/forest/pdf/2009/04/f08221.pdf

- See the fire-oak fire dependency for instance in Sturtevart et al 2009: https://link.springer.com/article/10.1007/s10021-009-9234-8,

- Last, Rogers et al., 2015 do not focus on fire hazard/probability but in fire response (embracers vs resisters in Northern USA and Eurasia) that have no direct relationship with fire risk. So, while I understand that you speak on terms of Central-Eastern Europe temperate forests it is still imprecise to say that broadleaf deciduous trees have a negative effect on ignition probability and spread (especially the latter). I'd suggest to reword this so you either discuss this statement with more references and making clear that given the same climate conditions, temperate forests are less fire-prone, but once the fire ignites, it spreads intensively as in a needleleaf forest.

R: Thank you. We have slightly altered the content of this sentence to show that deciduous trees have low flammability, low ignition probability and fire spread, instead of saying they have a negative effect on fire. Some references provided by the R1 refer to the Mediterranean biome, which is characterized by an entirely different suite of species and behaviour related to fire. Our revised sentence now reads: "For example, needleleaf trees with volatile compounds and resins, retention of dead biomass in crown, ladder fuels and slow litter decomposition rates promote fire hazard. Temperate broadleaf deciduous trees (with the exception of drought-adapted oaks) with high leaf moisture content and faster litter turnover, usually have a low ignition probability and less flammable fuel, although under very dry conditions fire may spread with high intensity once fuel has been ignited (Sturtevant et al., 2009; Rogers et al., 2015)." Please see lines 130-135.

METHODS

Overall I find the methods applied appropriate for the questions you pose; both the dataset compilation and the statistical procedures are very good and, given the spatiotemporal dissimilarities of all the records, the smoothing process is the best you can probably do here. I have though some particulars that I'd like to address so we can have an open discussion:

- If you paper is eventually published (and I do hope it is!) I think your dataset should be upload not only to those repositories (GPWK, Pangea, etc. which are fine) but somewhere accessible with the code, so for instance now I have no way to check the dataset, its consistency and reproducibility. Would it be possible that you enter the data somewhere

(Zenodo, dryad: : :) so the referees can have access to it? Also for version control – in case you realize a mistake in the future – is good to have a DOI somewhere so the future readers can have some control on different versions. - In connection with the data curation, having the new age models you have done in this study available in the appendix would facilitate the quality assessment.

R: All charcoal records will be soon available at the paleofire.org.
The fully commented code is available as Supplement 2 and deposited at the paleofire.org.

- You use GAMs (I have made some notes in the pdf itself on adding relevant references to the use of this method in palaeo as you might not have been the first group using them see Simpson, 2018 in Front. Eco.Evo) but in your plots (or code for that matter) you do not seem to include any confidence interval? Why is that? GAMs are "datahungry" methods and I guess that in some case you might be right on the boundary to use them. I think you will have a quite some large error in the curve extremes and your conclusions may have been a bit over-interpreted just for this reason; so I think it's important that you add your CI to the plots (and code).

R: We have removed the word novel from introduction and added a new sentence citing Simpson, 2018 in the Method section. This reads "GAMs have been shown to provide robust statistical analyses of trends in palaeoenvironmental time series (Simpson, 2018)''. Please see lines 279-280.

All GAM outputs have now been re-plotted with colored ribbons representing standard errors.

- In connection with the GAM model term assessment and the use of AIC, the package mgcv already penalizes (with lasso) those terms of the model that increase complexity but do not improve fitness, why do you then chose to use AIC? I mean AIC is one of those possibly useful measurements but as you have others already implemented to assess GAMs I do not see the need to add another number to it (I found this reading about AIC quite useful: https://dynamicecology.wordpress.com/2015/05/21/whyaic- appeals-to-ecologists-lowest-instincts/) as AIC somehow increases our feeling of quality, and that might not be true.

R: The use of the lasso in fitting GAMs allows variables with little predictive power to be eliminated from a model, but it does not allow comparison of relative predictive power for models using different predictors. As the goal of the current project is exactly that sort of comparison, we do not see that the lasso is serving the same function as our use of AIC.

Regarding McGill's points and the comments following that blog post, they seem to us to be more related to the misuse of AIC than any particular problems with AIC per se, and the way in which we are using AIC to do model selection here is one of the approaches he specifically says is defensible. To quote his discussion of one example from that post: "You could also use AIC to do variable importance ranking (compare AIC of S~prod, S~seas, S~energ). This is at least close to what Burnham and Anderson suggested in comparing models. You could even throw in S~area at which point you would basically be doing hypothesis testing vs a null although few would acknowledge this." The analogous approach (given the structure of our models) is precisely the approach adopted here; comparing models that differ by a single predictor of interest, and contrasting those with a null model. As such, we acknowledge the issues McGill raises but feel that in this case we have used AIC appropriately.

While I find that Supplement 2 is a really good tool facilitating reproducibility, is really difficult to find the results you mention in the main manuscript text, e.g. in line 297 you invite the reader to check that "climate alone explains a large proportion of the deviance of biomass burning in the three ecoregions in the time period between 12-8 ka BP", but finding these results in the Electronic Supp. Materials is not easy. Would it be possible to have your code commented and highlighted and hosted in a repository? I think that would increase reproducibility and therefore visibility of these results.

R: We have included a few statistical parameters such as deviance explained and r. squared in a new (Table 1) and the Table 1 became Table 2 into the main text.

The SI 2 contains a fully commented code see https://drive.google.com/file/d/1hRF9XStgx42sQzN8SkfhyYkfXZZahhR/view?usp=drive_web . A copy of the script is deposited at the paleofire.org.

RESULTS AND DISCUSSION

I find all the argumentation around the forest opening quite interesting but all the examples you bring into place refer to fuel-limited environments and how these change dynamically according to changes in productivity. In my opinion (and at the sight of the land cover reconstructions you provide) these areas have never been under real fuel-limited conditions, and reducing the tree cover may create that mosaic you discuss in lines 356-359 but may equally create effective fire barriers for fire to spread. So, I'd suggest to rather look for modern day fire ecology references showing that this might be the case in the temperate forest of central Europe. An alternative hypothesis to this "counter-intuitive" evidence where increasing tree cover reduces fire risk, may also be that a denser, thick temperate forest, (even under dry conditions) creates microclimate conditions that reduce fire risk, as you were stating in the introduction.

R: Not all references are from fuel-limited regions, as Scheffer et al., (2012) comes from boreal, whereas Frejaville et al., (2016) from temperate region in Europe. We have refined the paragraph explaining how tree cover can affect biomass burning directly and linked to climate conductive conditions. Please see the section 4.1, but especially lines 360-380.

- Another interesting topic that I'm unsure you discuss in depth (or maybe I overlooked it) is that there's some potentiality for spurious correlations when interpreting your GAMs. In Fig 3 is clear that increasing broadleaf forests cover decreases fire probability, but how do we know that this is not just a climate effect rather than a forest composition matter? In your TraCe reconstruction you evidenced that increasing moisture availability would have implied the expansion of broadleaf forests and that reduces fire probability, but would not be simply that less effective evaporation, i.e. increasing summer rainfall and reducing insolation, reduces fire chance? Can you please develop on this? Especially exploring Appendix A figure. I think a worth exploring aspect here is what are the the P-PET thresholds for each forest type to create a higher fire-risk.

R: We acknowledged that climate conditions played a role in vegetation dynamics, which in turn feedbacked into biomass burning, i.e, decreased temperature and increased moisture availability promoted forest expansion directly and indirectly increased fuel moisture, thus reducing fire probability. We have now elaborated this in the revised chapter 4.2 The impact of climate on fire viaconditions conductive for fire and fuel quality. Please see lines 414-434.

Please also note the supplement to this comment: https://www.biogeosciences-discuss.net/bg-2019-260/bg-2019-260-RC1- supplement.pdf Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-260, 2019.

R: We have checked the pdf and it does not contain new comments, but repeats those already listed here.

Reviewer #2 (Remarks to the Author):
Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-260-RC2, 2019 Author(s) 2019. Interactive comment on "Fire risk modulation by long-term dynamics in land cover and dominantforest type in Eastern and Central Europe" by Angelica Feurdean et al. Patrick Bartlein (Referee) bartlein@uoregon.edu Received and published: 31 October 2019

General comments: This paper takes on the important issue of disentangling the relative roles of changes in climate and land-cover (both natural and anthropogenic) on bimass burning. The study employs independent data sources (climate-model simulations, pollen-based land-cover reconstructions and sedimentary charcoal-derived estimates of biomass burning) in a statistical modeling approach (generalized additive models, or GAMs) using data from Central and Eastern Europe over the Holocene. Overall, the results are likely sound, but there is need for some clarification and improvement in the presentation of the results, and I think there is also room for some more exploratory data analysis. In particular, it would be interesting to see scatter plots of the data presented in Fig. 2, and labelled scatter plots to support the relationships shown in Fig. 3. The application of GAMs in this analysis is completely appropriate, but as in any statistical analysis, the results will be more powerful if well supported by exploratory and diagnostic analyses.

R: We present these scatter plots in Supplement 2. We have conduced additional statistical analysis for GAMs (please see our response below).

Specific comments:
Line 85: "mediating the fire regime"

R: "in mediating fire regime" corrected to "mediating effect of vegetation on fire regime". Line 86.

Line 88: I'm not sure whether GAMs are still considered "novel".

R: We have deleted the word novel.

Line 90: Replace "diverged more markedly" with "was more spatially variable"? (I don't understand the notion of "divergence" of biomass burning.)

R: Indeed diverge was used here to reflect spatial variability. This sentence now reads "Biomass burning was highest during the early Holocene and lowest during the mid Holocene in all three ecoregions, but was more spatially variable over the past 3-4 ka BP.'' Lines 93-95

Line 92: "highest" (for parallelism).

R: Stronger predictor was replaced with highest predictor. Line 95.

Line 92: "decreased strongly" How does one define "strongly?" I would simply say something like "decreases to a minimum between 60% to 70%. . ." (The support for these statements is Fig. 3A, and the text on lines 360-366, right? You're talking about the form of the relationships, not the strength.)

R: Done, the sentence now reads: "In temperate forests, biomass burning was high at ~ 45% tree cover and decreased to a minimum between 60 to 70% tree cover.'' Lines 95-96.

Line 94: How does one define "abruptly?"

By abrupt we mean changes in characteristics that occur very fast with respect to the scale of observations, here represented by the relationship between biomass burning and tree cover. In Fig. 3, one can see that in the BNE ecoregion, the decline in biomass burning is not gradual but very steep once the tree cover become higher than 70%. We replaced the word abruptly with steeply and hope this is a better choice. Please see line 97.

Line 111: "have"

R: Done, vegetation properties has… corrected to vegetation properties have. Lines 118

Line 130: "higher tree cover" Higher than what? Maybe just "high tree cover"?

R: Done "While high tree cover may reduce fire hazard, fire…. . Line 142.

Line 133: Something missing. ". . .produced by modern forestry. . .?"

R: Corrected to: "Widespread plantations of highly flammable trees (e.g. *Pinus*) produced by modern forestry may further increase…'' Line 143.

Line 139: Not just anthropogenic impact. Past climates too.

R: Done, this sentence now reads: "Yet, present-day ecosystems and fire regimes carry the legacies of past anthropogenic impact and climates'' Line 152.

Line 141: Hyphenate "centennial-to-millennial" (when used as an adjectival modifier of,e.g. "data sets").

R: Done, Hyphenated. Line 153.

Line 149: For parallelism, either: "evidence of fire (something, "occurrence?" "frequency?"), land-cover composition, and climate" or "evidence of changes in fire, landcover composition, and climate."

R: Done, this sentence now reads: "We use independent evidence of changes in fire, land-cover composition and climate with a statistical modelling approach (generalized additive models, GAM) to quantify percentages in land cover and tree-density associated with fire hazard. Lines 162-164.

Line 155: We are all pretty sure (we being the authors, reviewers, and readers here) that charcoal influx is a measure of fire, but we (again in the broad sense) have not yet completed the calibration between biomass burning and charcoal influx (except perhaps at the global scale, and somewhat tenuously, Harrison et al., 2018 Earth Sys. Dyanm. 9:663-677). It would be good to include a few sentences, for non-paleo readers in particular, on what the sedimentary charcoal record can and cannot say about biomass burning.
Line 156: peatland vs. lakes. There is a big difference in the way that lakes and bogs accumulate charcoal, given the propensity of bogs to potentially burn, which makes their

records "lossy." This is an area of ongoing research, and the potential impact of the two different kinds of records should be discussed later in the paper (i.e. in Section 4).

R: We have introduced a paragraph addressing the issue of limitation of charcoal records. This reads: "Sedimentary charcoal is the most common proxy to determine relative changes in biomass burning (Whitlock and Larsen, 2001). While progress has been made to determine charcoal source areas, the quantification of absolute burned area from charcoal records is still challenging (Adolf et al., 2018). We therefore interpret the charcoal signal as relative trends in biomass burning (see Marlon et al., 2016). " Lines 172-174.
R: We are aware of the differences in charcoal deposition between lakes and bogs and we have acknowledged this in the methods section. In the same way, differences are probably less important than between a lake with or without inflow or outflow. Statistical treatment of the charcoal records has reduced, to some extent, the biases connected to both analytical method and depositional environment. However, to thoroughly address this subject with the current dataset would be a new study in itself and it is beyond the scope of this paper. However, we have included a paragraph addressing the difference between charcoal records derived from peat and lakes: "Regarding the depositional environment, bogs provide a more local representation of past fire occurrence than lakes, because they are characterised by limited charcoal transport and post-fire transport or erosion (Conedera et al., 2009; Rius et al., 2011). However, peatlands are susceptible to burning, which may introduce hiatuses in the depositional environment ". Lines 175-177.

Line 165: I don't quite understand this. Are the average fire sizes calculated only for fires less than 10 ha? What are the average (or better, median, fire-size has a longtailed distribution) sizes if all fires were included?

R: We have clarified this and the paragraph now reads: "The average fire size is ~10 ha in eastern Europe, between 5 and 10 ha in southern Europe and < 5 ha in northern and central Europe (European Forest Fire Information System, (http://effis.jrc.ec.europa.eu).''. Lines 185-187.

Line 179: "charcoal accumulation rates (or "influx")"

R: Completed "Charcoal concentrations were transformed into charcoal accumulation rates or influx (CHAR)… " Lines 191-192.

Line 188: "To reduce the influence of high-resolution charcoal records. . ." I think there's a problem here. The description of the treatment of the charcoal data leaves out the "prebinning" or "presmoothing" step, which is designed to reduce the influence of those high-resolution records. If the bootstrapping was done by resampling individual charcoal observations (i.e. samples or influx values), then that does not ameliorate the influence of high-resolution sites, because their charcoal observations, being more plentiful, will be included in the bootstrap samples more frequently. If the bootstrapping was done by site (the usual approach), then that still does not reduce the influence of high-resolution sites, because again the high-resolution sites will flood the bootstrap samples with their observations. (Bootstrapping-by-site was developed to include the effect of uneven spatial distribution of sites on the smoothed curves (Marlon et al. 2008). I'm guessing the presampling step was included (because it is part of the usual "analysis flow" implemented by the R package), but just not described.
Line 192: "we then calculated the mean and 90% confidence intervals of the aggregated records" In practice (at least when the locfit() function as implemented by the R paleofire package is used to fit composite curves, as described by Blarquez et al.,2014), the confidence intervals are defined by the fifth and ninety-fifth percentile values of the bootstrap replicates

of the composite-curve values. The locfit() function also provides confidence-interval values, as described by Loader (1999, sec. 2.3.3), for an individual composite curve. Which confidence intervals are being described here? (I'm guessing, from the width of the intervals, the former.)

R: We have clarified this further. The paragraph now reads: "The standardisation procedure included a min-max rescaling of CHAR values, followed by a Box-Cox transformation to homogenise within-record variance, and a subsequent Z-score transformation using a base period from 12 to 0.15 ka BP. This base period includes the entire dataset, but excludes the effect of post-industrial human impact on fire activity. For compositing charcoal records by ecoregions, transformed charcoal records from each ecoregion were pre-binned in 100-year bins to reduce the influence of high-resolution records on the composite charcoal record. Pre-binned charcoal time series were smoothed with a LOWESS smoother with a 500-year window half width. Confidence interval values (95%) were calculated by bootstrap resampling the binned charcoal series and calculation of the mean for each bin 1000 times (default settings). For numerical processing of the CHAR series we used the R paleofire package version 4.0 (Blarquez et al., 2014). CHAR composite anomalies (100-year time interval) relative to the Holocene average of the entire CEE region and the three ecoregions, represent regional trends in biomass burning; where zero Z-score values correspond to the mean charcoal influx over the base period; and positive/negative Z-score values represent greater-than-mean/lower-than-mean charcoal influx over the base period (Fig. 2).'' Lines 206-219.

Line 195: Hyphenate "land-cover" when used as an adjectival modifier.

R: Done.

Line 211: Delete "this is" and run-on from the previous sentence.

R: The two sentences were united in a single one which now reads: ''Throughout the text, we use the term 'grassland cover' to denote both natural and human-modified grasslands (pastures), and 'arable land cover' to denote arable and disturbed land, because it is not always possible to distinguish between natural and managed grasslands or between arable and other forms of disturbed open land cover based on pollen analysis (Fyfe et al., 2015)''. Lines 237-240.

Line 218: "We then generated composite estimates of land-cover classes grouped by ecoregion. . ." Presumably, the present-day ecoregions are used to do this, but haven't ecoregions moved around in the past?

R: Yes present-day ecoregions. The past vegetation was dynamics, however, at the Holocene scale even if the BNE ecoregion contained a higher proportion of broadleaf tree cover during the mid-Holocene and greater needleleaf tree cover during the early and late Holocene, this region term as boreo-nemoral has never been dominated by broadleaf trees. Similarly, the proportion of needleleaf tree in the CON and ATL ecoregions did not increase to values to term them boreo-nemoreal ecoregions. Our revised sentence now reads: "We then generated composite estimates of land cover classes grouped by present-day ecoregions by spatially aggregating the averages of pollen records within the corresponding ecoregion". Lines 245-247.

Line 224: The TraCE-21ka data are reported using a fixed (modern) 365-day (or "noleap") calendar, and consequently should be adjusted to reflect the impact that changes in Earth's orbit have on the length of individual months or seasons (Bartlein and Shafer, 2019, Geosci. Model Dev. 12:3889-3913). The "calendar effect" on transient climate simulations can, for example, influence the amplitude of temperature variations over time on the order of several degrees, and change the timing of, for example, Holocene "thermal maxima" on the order of thousands of years, depending on the region and month of the year. Our paper appeared only recently, and so it would not be reasonable to expect that any analyses reported here should be redone, but it would be useful ask how big the impact might be. I looked at the monthly time series of near-surface air temperature for a region corresponding to the data-dense region here (42.67 to 64.94 N and 7.5 to 41.25 E), for both "raw" and calendar-adjusted data. As it happens, there is little effect on June and July temperatures, but calendar-adjusted temperature for August is 1-2 deg. C higher in the early Holocene than the unadjusted data, and remains so until around 7 ka. When combined with the June and July temperature curves, the calendar effect is likely to not be significant. However, the overall "shape" of the monthly time-series curves differ, and so in future work it might be better to not aggregate the monthly time series into seasonal averages.

R: Acknowledged. In the section 4.2 where we discuss possible reasons for discrepancies in proxy and modelled past climate conditions, where we have added the following text: "Further, the model does not account for changes in the length of individual months or seasons due to variations in Earth's orbit (e.g. Bartlein and Shafer, 2019). Therefore future work might rather analyse the monthly time series instead of seasonal average (e.g. Bartlein and Shafer, 2019)" Lines 245-247.

Line 229: "surface temperature" Was this actually "near-surface air temperature" TREFHT, temperature at the reference height (2 m) in the CCM3 variable-naming scheme, or was it land-surface temperature, also known as "skin temperature," TS? If TREFHT, then using the CRU TS 3.1 data as a present-day reference is ok, but if TS, then the amplitude of its seasonal cycle will be larger than that of temperature in the CRU data set.

R: We used the near-surface air temperature. We have modified the text accordingly. Line 255.

Line 232: State the version number ("CRU TS 3.1").

R: Done. Line 258.

Line 233: "as ratios of the surface temperature . . . from CRU". "Bias correction" using ratios is appropriate for precipitation, but is a little unconventional for temperature, where the biases are usually taken to be additive, not multiplicative.

R: You are right. It should not be ratios. We used anomalies for the temperature. The text has been modified accordingly and now reads: "The bias correction was calculated with respect to the last 30 years of the TraCE-21ka simulation (representing pre-industrial conditions) as anomalies of the surface temperature and ratios of precipitation. The temperature anomalies were subsequently added, and precipitation ratios multiplied with the CRU data in order to obtain the bias-corrected climate". Lines 258-263.

Line 237: "P-PET" Why not P-E, which has been used before to index moisture availability (e.g. Daniau et al. 2012)? Potential evapotranspiration is energy limited (i.e.by net radiation in the Penman-Montieth approach, and via a temperature index in theThornthwaite approach), whereas evapotranspiration is governed by both energy and moisture, so why is potential evapotranspiration preferred here? Moreover, I don't thinkThonicke et al. (2001) an appropriate motivational citation. (That paper is about modeldevelopment, not climatic controls of fire.)

R: We used PET as we preferred a drought indicator that purely depends on climate and not on soil and vegetation. P-E would, of course, also have been a sensible choice. In our opinion both indicators are similarly adequate and would yield similar overall conclusions. We considered PET to be more parsimonious as simulating E depends on more assumptions. In Daniau et al. (2012), it is not clear for us how E was calculated. The author list indicates that the approach from the biome and STASH models from Prentice et al. and Sykes et al. from the 90s was used. In this approach, E depends on soil water, and soil water holding capacity is needed as input, in the models mentioned above often using 150 mm across the globe (plant-available water, not total volumentric content). We also considered this approach, but finally decided to use the simpler approach using PET. Using the balance of PET and P is also a common approach (e.g. Hickler et al. 2009). In addition, because PET is generally more sensitive to changes in climate than AET, it also yields a larger signal-to-noise ratio when subtracted from precipitation. As a result, using PET it is easier to identify periods that are particularly wet or dry in different regions compared with using AET. Finally, P-E has the caveat that it becomes zero under water stress, but it cannot show the severity of water stress, i.e., it cannot indicate the quantity of water that lacks to satisfy atmospheric demand (E cannot be higher than P if changes in soil water storage are negligible compared to changes in P, which is common). We think that discussing these issues would be beyond the scope of the manuscript. However, thanks for pointing out that Thonicke et al. (2001) was not a good reference here. Line 236 to 240 was re-formulated as: " We used the boreal summer (June, July, August, hereafter "JJA") surface temperature (JJA T) and precipitation minus potential evapotranspiration (JJA P-PET), as a proxy of peak summer dryness, which is a main driver of fire hazard. P-PET is a drought indicator that, opposed to P-AET (actual evapotranspiration), purely depends on atmospheric moisture demand independent of soil and vegetation, and reflects the atmospherically-driven intensity of drought conditions (Pidwirny, 2006)." Please see lines 263-267.

Line 242: "average day length for each month . . . was calculated" Please explain.

R: How we calculate the average day length is already stated in the manuscript: "the average day length for each month going back to 12 ka BP was calculated using the Earth's orbital parameter scheme in CCSM3." Lines 269-271.

Line 242: "The resulting climate fields. . . were interpolated" Typically, this would be done by interpolating the long-term mean differences (paleo minus present) on the GCM grid onto the 0.5-deg CRU grid, and applying them to the "observed" CRU values. Is this what was done?

R: Yes, this is exactly how it was done. The bias-correction procedure using CRU is already stated earlier in the text (albeit wrong, but now corrected due to your comment). Lines 258-264.

To emphasise that the bias-corrected fields are also used in the interpolation the specific site, we now write: "The resulting bias-corrected climate fields were subsequently interpolated to the same locations as the charcoal records using a bilinear interpolation." Lines 270-272.

Line 245: "Similar to vegetation and fire reconstructions. . ." If you followed the Blarquez et al. (2014) procedure, then the charcoal composite curves aren't exactly "loess" curves in the classical sense, where a variable-width smoothing window "span" is used (as opposed to a fixed-width window). In any case, what was sampling frequency of the three data sets?

R: We used a 100-year time interval for charcoal and climate (please see lines 216 and 274, respectively) and a 200-year time interval for pollen (line 222). To generate composite estimates of JJA climate and land cover classes in Fig. 2 we fitted a 500-year loess smoother (please see lines 528-529).

Line 251: Reorganize the sentence. (The predictor is the sum of smoothed functions of land cover and climate.)

R: Done, it now reads "The predictor is the sum of smoothed functions of land cover and climate (Hastie and Tibshirani, 1990).'' Lines 274-275.

Line 253: Reorganize the sentences. (You used the mgcv package to fit models with thin-plate spline predictors and a Gaussian-family error distribution.)

R: Done, it now reads "We used the mgcv package to fit models with thin-plate spline predictors and a Gaussian-family error distribution to automatically determine the optimal level of smoothing for each term in the model and automatic term selection (Hastie and Tibshirani, 1990).'' Lines 279-283.

Line 256: "Akaike Information Criterion (AIC) weights. . ." Cite: Wagenmakers EJ & Farrell S. (2004), AIC model selection using Akaike weights. Psychonomic Bulletin &Review 11: 192-196.

R: Reference included.

Line 258: "evidence within each data set" This is a fancy way of saying "goodness of fit".

R: Sentence changed accordingly. "AIC weights are a normalized indicator of support for each model given the goodness of fit while penalising more complex models''. Lines 285-286.

Line 259: Replace "scores" with "values" ("Scores" come up in other contexts in fitting GAMs.)

R: Scores replaced with values. Line 287.

Line 264: Delete "analysis".

R: Deleted.

Line 266: "see Pollen-based. . ." Refer to the section number.

R: Done. Line 296.

Line 267: "GAMs on JJA climate" Jargon. ("GAMs using JJA climate" would be ok.)

R: Done. This sentence now reads: ''However, we also constructed GAMs using JJA climate for the 12-8 ka BP period to investigate the relationship between climate and fire without significant human impact''. Lines 292-294.

Line 273: "over all . . . and within the three ecoregions."

R: Done. This sentence now reads "The amount of biomass burning was highest during the early Holocene (between ~ 10.5 and 8 ka BP) over all of Central and Eastern Europe and within the three ecoregions.'' Lines 300-302.

Lines 274-290: Straighten out tense throughout. ("simulation indicates" vs. "Biomass burned showed" etc.)

R: Done. Line 298.

Line 278: Reword "The reduction in biomass burned accompanied a decrease in JJA temperature and an increase in summer moisture availability. . ." (for parallelism).

R: Done. This sentence now reads: "The reduction in biomass burning accompanied a decrease in JJA temperature and an increase in summer moisture availability (around 8 ka BP) in all ecoregions…'' Lines 306-307.

Line 281: "but less evident" Something missing.

R: Rephrased: "We found differences in trends in biomass burning among ecoregions over the past 3 ka BP. Biomass burning increased markedly at 3 ka BP in the BNE ecoregion, but this increase is less evident in the CON ecoregion". Lines 307-309.

Line 293: This section could use a summary paragraph that summarizes the results and motivates the further statistical analysis, which looks at the combined effects of climate and land cover. Alternatively, that could go into a lead paragraph in the following section, which otherwise gets right into model selection.

R: Thank you for this suggestion. We have introduced the following paragraph: "When considering GAMs fitted with only climate predictor variables over the full time series (12-0 ka BP), the proportion of the deviance of biomass burning in the three ecoregions averages 48% (Table 1; Supplement S2). We therefore investigated into the drivers of biomass burning on the pre- and post-8 ka BP time periods separately." Lines 324-332.

Line 298: I see an appendix A (but not A1), but that does not report any deviance values. In the supplement, I can see that the average of the deviance explained by the first three models is 71.7%, so perhaps the value in the text is just badly rounded. A) I really think you need a summary table in the main text, that summarizes the goodness of fit and complexity of each model. The reader should not have to conduct an additional analysis to figure out what's going on. B) In addition, it is not clear just what data is going into the analyses. In the supplement, it looks like there are 80 observations in the "full" data sets (i.e. 12 ka to 0 ka), 39 in the 12 ka to 8 ka subset, and 80 again in the 8 ka to 0 ka subset. Does this imply that temporally spacing of the input data varies over time? C) Also, the GAMs are being fit to data that already have be smoothed, so to what extent does that influence the interpretability of the deviance explained? At a minimum, a time-series of the residual values would be interesting.

R: We have introduced a new table (Table 1) into the main text that shows the deviance and R-Squared values.

Thank you for spotting this, we have used 119 points for full datasets (12-0 ka BP period), 39 points for 12-8 ka BP period and 80 points for 8-0 ka BP period. This is corrected in our new SI2.

We have now included plots of residuals as a function of time. These plots show that there is still a fairly high amount of temporal autocorrelation in the residuals, particularly for the null models. As an experiment we constructed a separate series of models that were identical to these but which also include a smoother for time itself. These models removed some of the temporal autocorrelation in the residuals, but not all. This is not surprising given that the response variable itself is an estimate produced from a smoother fitted using time as a predictor; is expected to be substantially more temporally autocorrelated than the original source data from which it was interpolated. More importantly, the models fit with an additional smoother for time produced model selection results that were not materially different from those without. As a result, we are presenting the original models for the sake of simplicity.

Line 302: "Table 1" The table should indicate that the results pertain only to a subset (8 to 0 ka) of the data.

R: Done: Table 2 (Table 1 became Table 2) now reads "Model selection results for generalized additive models of the effects of land cover and climate on biomass burned for the period 0-8 ka BP. "

Line 303: Replace "scores" with "values".

R: R: Done.

Line 309: "Fig. 3A" How were these marginal plots constructed? (The supplement seems to conveniently omit the code. . .) Personally, I dislike plots that are unsupported by actual data (unlike the partial residual plots in the supplement). Would labeled scatter plots be convincing?

R: Please see the full version of the code in Supplement 2 and paleofire.org.

Line 324: "higher-than-present".

R: Done. "We found that climate, specifically warmer-than-present summer temperatures and higher-than-present moisture content". Lines 415-416.

Line 327: "recovered?" That's kind of a Holocene-centric view of the world.

R: We have slightly altered the sentences, which now reads: "Enhanced biomass burning with increasing temperature and moisture in the early Holocene in CEE is expected, as fuel builds up progressively following the cold and dry conditions with limited biomass that prevailed during the Lateglacial (Feurdean et al., 2014)''. Lines 417-420.

Line 328: "land-cover change and human imprint" (for parallelism) and Line 330: "land-cover models" Replace with "models that include land cover as predictors".

R: Done. Lines 365-366.

Line 332: "mid-to-low latitudes"

R: Done. Line 371.

Line 333: "climate reconstructions are fragmentary and mostly qualitative" I think some of your coauthors would disagree with that assertion.

R: Independent of vegetation, quantitative climate reconstructions are truly rare in this region, and we would like to avoid using pollen as both a way to reconstruct land cover and climate conditions in order to circumvent the risk of circularity. However, one of the main reasons for using simulated results is that proxy data do not provide all climate variables (T, P-PET) and the monthly resolution available from climate models needed to run the GAMs. This sentence was altered to: "Proxy data do not provide all climate variables (T, P-PET) and the monthly resolution available from climate models". Lines 334-335.

Line 336: “Simulated” Does this refer just to the TraCE-21ka simulations, or to simulations in general?

R: Amended to: “TraCE-21ka simulation and proxy-based climate reconstructions are in general good agreement in indicating warm and dry…” Lines 436-440.

Line 343: “This could be partly explained by. . .” And also (largely I think) by climatemodel resolution.

R: Amended accordingly “On the one hand, this disagreement could be explained by the coarse resolution of the climate model. Further, the model does not account for changes in the length of individual months or seasons due to variations in Earth’s orbit (e.g. Bartlein and Shafer, 2019). Therefore future work might rather analyse the monthly time series instead of seasonal averages. On the other hand, increasing human impact on the proxies used for climate reconstructions, such as the effect of water acidification and eutrophication on chironomid assemblages and deforestation on testate amoebae composition, could also be responsible for disagreement between proxy- and simulation-based inference of past climate conditions (Heiri et al., 2015; Mauri et al., 2017) ”. Please see lines 441-451.

Line 350: “GAM models” Expanding the acronym yields “generalized additive model models,” so just say “GAMs.”

R: Here and elsewhere in the Discussion we have replaced GAM models with GAMs

Line 350: “While the GAM models use biomass burned as the response variable. . .” They use composite curves of charcoal influx, which are thought to represent biomass burning.

R: Sentence amended accordingly, it now reads: “While we set up the GAMs with biomass burning as the response variable’’. Lines 362.

Line 355: “lowering” You mean “decrease in tree cover” as opposed to canopy height or something, right?

R: We mean a reduction in the tree cover /canopy density not canopy height. Sentence changed to better reflect this: “A reduction in tree cover allows the development of more understorey vegetation that provides a favourable fuel mix composed of fine herbs, shrubs and coarse woody debris that facilitates ignition and surface fire spread (Pausas and Paula, 2012; Frejaville et al., 2016).” Lines 400-404.

Line 361: Figure 4 needs more explanation, and might be out of place. What are the individual points? Should the figure preceed Fig. 3 (and be explained earlier)? It seems to simply show data, and not the results of any model fitting.

R: Indeed Fig 4 should precede Fig. 3 and in the revised version we have swapped places plus added an explanatory text to Fig. 3 in section 3.1. This reads: “Three-dimensional scatter plots of biomass burning and percentages of land cover classes show that locations with greater biomass burning tend to be consistently characterised by low broadleaf and high needleleaf tree cover in all three ecoregions (Fig. 3). Biomass burning also increases with arable and pasture cover, although percentages of the two land-cover classes at which biomass burning increases vary between ecoregions (Fig. 3). Lines 323.325.

Line 367: “regional divergence between biomass burned and percent tree cover” Do you mean divergence in the relationship between the two, or simply regional variations in the relative levels of each?

R: We mean that the relationship is divergent. We have amended this sentence to better reflect this: "The GAMs run separately on broadleaf and needleleaf tree cover indicate regional divergence in the relationship between biomass burning and percent tree cover associated with the dominant functional forest type (Figs. 3, 4). " Lines 387-388.

Line 369: "Broadleaf cover. . ." But Table 1 shows the lowest AIC for a model with total tree cover in the CON region.

R: Thank you, we have corrected this. The sentence now reads: "Broadleaf cover had the most powerful negative effect on biomass burning in BNE and ATL ecoregions, and the second-most negative effect after total tree cover in the CON ecoregion (Fig. 3B; Table 1; Supplement S2) ".Lines 389-391.

Line 391: "While past ignition is assumed to increase with population density. . .." Citations?

R: Refined to: "While ignitions tend to increase until population reaches intermediate density, human-caused change in land cover from forest to arable land and associated fuel limitation and landscape fragmentation may result in a decline in biomass burning (Pfeiffer et al., 2013; Andela et al., 2017). A contrasting position suggests that increasing population density exclusively reduces fire frequency and burned area through the impact of land conversion and landscape fragmentation on fuel availability and continuity (Knorr et al., 2014). " Lines 458-464.

Line 392: "and associated fuel limitation" It's not just fuel limitation that reduces burning in arable lands. It's never been good policy to burn crops, except as an element of warfare.

R: We refer here to the burning of crop residue and not the crop themselves. We have stated in the next sentence that: "If fire was primarily restricted to burning of agricultural waste, e.g., straw and chaff, to improve soil fertility and clean the land, this should have provided less biomass to burn than wood (Pfeiffer et al., 2013). Lines 479-482.

Line 436: "summer conditions"?

R: This sentence was removed altogether.

The attached figure shows area-weighted averages of TraCE-21ka near-surface air temperature (TREFHT) for ice-free land grid points over the region 42.67 to 64.94 N and 7.5 to 41.25 E. The gray and black curves show individual annual values and 30 yr (window half-width) locally weighted means of the distributed data, while the overprinted pink and red curves show data that has been "calendar adjusted." See Bartlein and Shafer (2019, Geosci. Model Dev. 12:3889-3913, for discussion, in particular Sec. 3.4). P.J. Bartlein Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-260, 2019. Fig. 1. Calendar-adjusted TraCE-21ka data.

References cited:

Adolf et al (2018) The sedimentary and remote-sensing reflection of biomass burning in Europe. Global Ecology and Biogeography 27, 2, 199-212. DOI: 10.1111/geb.12682

Rius, D. Vannière B., Galop, D. and Richard H., 2011. Holocene fire regime changes from local to regional scale documented by multiple-site sedimentary charcoal analyses in the Lourdes basin (Pyrenees, France). Quat. Sc. Rev. 30, 13-14, 1696-1709.

Conedera M, Tinner W, Neff C, Meurer M, Dickens AF, Krebs P. Reconstructing past fire regimes: methods, applications, and relevance to fire management and conservation. Quaternary Science Reviews. 2009 Mar 1;28(5-6):555-76.

Hickler T, Fronzek S, Araújo MB, Schweiger O, Thuiller W, Sykes MT. An ecosystem model‐based estimate of changes in water availability differs from water proxies that are commonly used in species distribution models. Global Ecology and Biogeography. 2009 May;18(3):304-13

Knorr, W., Kaminski, T., Arneth, A., and Weber, U.: Impact of human population density on fire frequency at the global scale, Biogeosciences, 11, 1085–1102, https://doi.org/10.5194/bg-11-1085-2014, 2014.

Reviewer 3 BGD Biogeosciences Discuss.,https://doi.org/10.5194/bg-2019-260-RC3, 2019

Interactive comment on "Fire risk modulation by long-term dynamics in land cover and dominant forest type in Eastern and Central Europe" by Angelica Feurdean et al. Christoph Schwörer (Referee) christoph.schwoerer@ips.unibe.ch
Received and published: 5 November 2019

General comments
The manuscript by Feurdean and colleagues compiles a large dataset of charcoal and pollen records to quantify the effect of climate and vegetation on fire regimes in Central and Eastern Europe. The authors apply generalized additive models (GAMs) to explore the relationships between biomass burned and changes in climate and land cover. They conclude that tree cover is a first order predictor of fire occurrence probability and that land cover management can reduce future fire risks. I've greatly enjoyed reading this relevant and well-written article. However, I feel that the manuscript would profit from a more process-based view of the drivers of fire occurrence in the study region. I fully agree with the authors that tree cover is a good predictor of fire occurrence in the past, however, this does not mean that there is a direct causal link. As recognized in a vast number of paleoecological articles, climate and human impact are the main drivers of both vegetation and fire dynamics during the Holocene, with climate being the primary forcing factor during the Early Holocene and anthropogenic impact becoming increasingly dominant during the Late Holocene. Since these two drivers affect both the response variable (biomass burned) and the predictor (tree cover) a high correlation is not surprising and should not be confused with causality. Although a decrease in tree cover indeed coincides with an increase in the amount of biomass burned, I would argue that changes in tree cover are itself caused by climate (during the Early Holocene) and human impact (during the Late Holocene). I do not see any evidence that would support the claim that total forest cover has a direct effect on fire dynamics, although I do concur that the type of vegetation (broadleaf vs needleleaf) has indeed an effect. I believe that a more cautious and less simplistic phrasing of the conclusions and abstract, highlighting the direct impact of human land-use on forest cover and fire occurrence would be highly beneficial for the article and not detract from the tremendous amount of work that has been put into compiling and analyzing such a large dataset.

R: Firstly, we would like to thank the three reviewers for constructive and thorough reviews of our manuscript. We also thank them for the encouraging words about our work.

This is an interesting point and one that we have considered at various times during the development of the research. We admit that land cover is affected by both climate and human-driven changes, and therefore refined some of our statements in the revised paper chapters 4.1 and 4.2 some indicated below.

Specific comments

Introduction:

L.116-118: You state here that ": : :an increase in tree cover beyond a specific threshold can reduce fire hazard,: : :", implying that tree cover itself has a direct effect on fire regimes. I agree that there is certainly a correlation, but would be very careful in assigning causation. Just in the previous sentence you mention that ": : :fire hazard is lowest in productive and moist regions: : :". From a mechanistic point of view, I would argue that the main driver in reducing fire hazards is the moist climate, which leads to lower flammability of fuels, and not just tree cover alone. In order not to confuse the readers I would recommend elaborating on how an increase in tree cover can lead to a reduction of fire, independent from climatic conditions (e.g. local microclimate, reduction of evapotranspiration under closed canopies, etc.).

R: Thank you. In response to your comments as well to the R1 we have made it clearer how the tree cover can alter biomass burning independently and via a climate-fuel feedback. It should be noted that although the composition and spatial distribution of plant communities are determined by climatic conditions, vegetation composition leads to heterogeneous patterns of fuels and flammability across space and time. Low biomass burning at high tree cover may have been driven by dense tree stands with reduced understorey, which created a cool and moist microclimate that lowers ignition potential and fuel flammability (Kloster et al., 2015). Radiative properties of the land surface at higher tree cover can decrease evaporation and/or enhance cloud formation, which in turn contributes to a moister local climate (Teuling et al., 2017). A reduction in tree cover allows the development of more understorey vegetation that provides a favourable fuel mix composed of fine herbs, shrubs and coarse woody debris that facilitates ignition, surface fire spread, as well as the transition from surface to crown fire (Pausas and Paula, 2012; Frejaville et al., 2016). Further, in open forests, radiation can penetrate deeper into the canopy, and the wind speed close to the ground is higher, which dries the understorey vegetation and litter, ultimately increasing flammability (Ryan, 2002). Please see our revised explanations in the chapter 4.1 but especially lines 370-380.

Discussion: Fire-fuel relationship: the human impact

What I miss a bit in this subchapter is the direct link between human land-use and sedimentary charcoal records. Fire was the primary method used to convert closed forests of arable land, therefore greatly increasing the natural fire occurrence and releasing large amounts of charcoal.

- L. 391-393: I'm a bit confused here. In this sentence you state that ": : :the human caused change in land-cover [: : :] has resulted in a decline in biomass burned." But this is contrary to your data, which clearly shows a steady increase in biomass burned during the last 3ka, when this transition to a cultural landscape occurred. However, if you refer to present-day observational data, please provide a reference and also a sense of the timescales involved. I would imagine that a decline of biomass burned due to fuel limitations as a result of the conversion of forests to arable land would only apply to the last 150 years or so, a timeframe that cannot be resolved by your millennial-scale charcoal record.

R: We have tried to link more evidently human land use and biomass burning. It should be noted that inferences of fire regime from sedimentary charcoal have their limitations, in particular with respect to the cause of fires. Therefore, if humans significantly increased fire frequency since the prehistory (Vannière et al 2016) or earlier (Kaplan et al. 2016), the impact of post-deforestation mosaic landscape heritage and the land use on fuel amounts may have reduced the probability of large fires and consequently the amount of charcoal preserved in the depositional environments. Lines 492-497.

Indeed this sentence does not refer to our own results but to those referenced. We made this clear by starting the next sentence with “Our results show that biomass burning mostly shows a positive response with increases in arable and grassland cover in all ecoregions, however, this relationship is variable and may illustrate a complex fire-human interaction (Figs. 3, 4).” Lines 463-464.

- L.402-405: From my point of view, the rise in biomass burned in the ATL ecoregion at 1.5ka coincides with a sharp increase in both grassland and arable land, indicating an intensification of land-use with the help of fire for deforestation, contrary to your statement here. I would argue that high charcoal values are a result of intensive landuse, caused by the widespread deforestation that led to the conversion of forest to arable land that started during the Bronze or Iron Age and reached its maximum during the Early modern period. A decrease in area burned due to the establishment of forest protection laws in the late 19th century will not be possible to observe due to the 500- year smoother applied to the data, but might be apparent in the raw data. However, after rereading this sentence a few times I realized that you might actually be referring to the period between 4 and 1.5ka. In that case, ignore my comment, but rephrase the sentence to spare other readers the same confusion.

R: We have clarified this sentence to show its reference to the 4-1.5ka. It now reads “In the ATL ecoregion, arable and grassland cover rose steadily from ~4 to 1.5 ka BP, but biomass burning remained constant during the same time (Fig. 2), which is consistent with percentages in arable and grassland cover at which biomass burning shows no responses in the GAMs (Fig. 4CD). This may suggest that the local intensification in land use between 4 and 1.5 ka BP did not involve major use of fire for deforestation (Fig. 2).’’ Lines 474-479.

Conclusions

I would be very hesitant to make a direct causal link between land cover per se and biomass burned. I totally agree that you make a compelling case for a strong connection between land cover and fire hazard. However, the underlying driver of changes in land cover (and biomass burned) in the Late Holocene is, from my point of view, intensifying human impact. Although land cover can be used as a predictor of fire occurrence in the past (as you nicely show), it is less suited to derive conclusions for future management options, since the underlying drivers might change. In the future I would expect no-analogue conditions, since the combination of higher-than-present temperatures and anthropogenic land-use never previously existed.

R: We did refine both the Abstract and Conclusion slightly to better accommodate our refined Discussion. Knowledge of past conditions is a necessary precursor to understanding how systems may change in the future. No-analogue conditions will certainly appear, but surely it is better to face these with knowledge derived from 12000 years of observations?

Technical comments:

- L. 111: “: : :the effect that vegetation properties have in: : :”

R: Done.

- L. 277-280: I would suggest rephrasing this sentence, since it is rather confusing with the subclause.

R: Done, it now reads "The reduction in biomass burned accompanied a decrease in JJA temperature and an increase in summer moisture availability (around 8 ka BP) in all ecoregions (Fig. 2B-D).'' Lines 306-307.

- L. 401: ": : :coincides with: : :"

R: Done.

- Figure 2C: For reasons of comparison, I would suggest to plot the needleleaf and broadleaf tree percentages on the same scale, as in the other plots.

R: We have tried to plot this but because the % vary among ecoregions the trends are not as clear as when using differed scales.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-260

References.

Vannière, Boris, et al. "7000-year human legacy of elevation-dependent European fire regimes." Quaternary Science Reviews 132 (2016): 206-212.