Response to comments by Anonymous Referee #1
(The reviewer's comments are in regular text and our response in italics)

The calibration of area burnt using modern charcoal is not well explained. There is quite limited information about the modern charcoal samples, which seem to be core tops of unknown time coverage. This information if crucial to assess the validity of the approach. Also, at L99 it says that interpolation was used to extract present-day burnt area at each of the sites with modern charcoal records. However, the comparison of the locations of the modern charcoal samples with the GLM output suggests relatively low spatial coverage of the calibration (e.g. a large area burnt fraction was derived using the GLM in north-central Portugal, but there is only one modern charcoal sample in the region). Another figure showing the calibration of area burnt/modern charcoal needs to be presented, at the moment it is unclear how these two match.

In this study, modern charcoal bins or "core-tops" cover the post-industrial period (1850 CE to the present) as stated in L108.

We realise that the reviewer has misunderstood our approach because we had not explained it clearly enough. In essence, we have derived a relationship between the vegetation assemblage and normalised charcoal using data from multiple sites through time. We then use this relationship to predict fire from vegetation data. However, we need a conversion factor in order to transform the normalized and qualitative charcoal records into a quantitative estimate of burnt area at each site. This conversion factor is derived by relating normalized modern charcoal to the actual observed burnt area via a statistical model. To make this clearer, we have modified the final paragraph of the Introduction (see below). We will also provide a general statement about the approach at the beginning of the Methods section, and include a flow-chart to illustrate the procedure, as follows:

The central premise of our approach is that fire frequency is one of the factors that influences vegetation assemblages (see Supplementary Information), and therefore that specific aspects of differences in vegetation assemblages – identified by a numerical technique that can isolate the effects of any one controlling factor on taxon composition – can be used to reconstruct fire. The vegetation-fire relationship can be derived by comparing changes in pollen assemblages and charcoal records through time. However, since the charcoal records from different sites consist of different size fractions and the records must be normalised to facilitate comparisons, it is necessary to derive site-specific conversion factors between modern charcoal abundance and present-day burnt area fraction. This calibration is then applied to the charcoal record in order to derive an estimate of the palaeo-burnt area for each pollen sample.

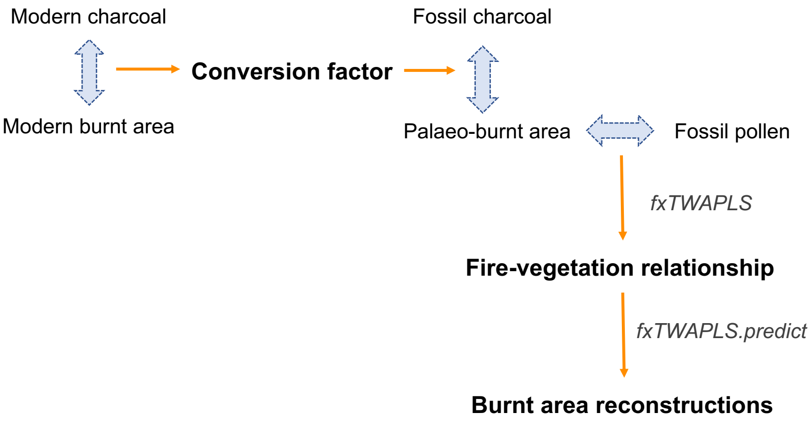

Figure 1. Flow chart of the methodology.

The area burnt fraction reconstruction shown in Figure 4 only 'matches' charcoal if we consider a long-term trend perspective. The individual wiggles are often anti-phase, which raise questions about the validity of the approach. Please reconsider the robustness of this validation.

We agree that the original figure emphasised the similarity of the long-term trend and that there are some anti-phased relationships on shorter time scales. This could arise because of different numbers of samples in the two data sets, or it could be an artefact of inappropriate choice of span (or half-width) in the loess smoothing. We have explored this by focusing on entities that have both pollen and charcoal records in order to compare the trend of reconstructed burnt area (from the pollen) with the trend of shown by the raw charcoal data. There are 2368 charcoal samples and 2376 reconstructed burnt area samples. If we only consider samples from shared age bins, there are 2104 samples. By using only these 2104 samples we can remove any impact of differences in sampling. We have investigated the impact of using different values of span. It is clear that some of the mismatches in the original plot were due to using an over-large span. Our updated figure, using 0.04 as the span for loess smoothing, shows greater congruence in the placing of peaks (although not in their magnitude). The updated figure is shown below (note that because of the addition of the flowchart, this will now be Figure 5):

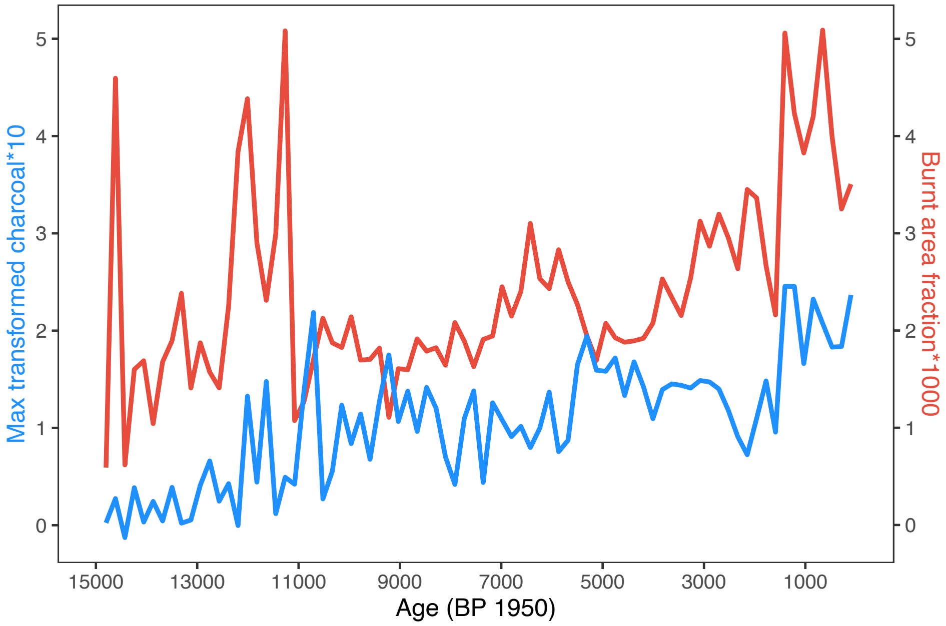

Figure 5. Composite plots comparing max-transformed charcoal values and the reconstructed burnt area for these entities for the 51 entities with charcoal. Max-transformed charcoal is shown in blue; burnt area fraction is shown in red. The loess smoothing is made with span = 0.04.

We have edited L195 to L198 as follows:

Charcoal values are not expected to be directly comparable with the reconstructed burnt area but should show comparable temporal trends. A composite plot of reconstructed burnt area for the 51 entities that have both pollen records used to reconstruct burnt area and charcoal records, and therefore can be compared, show similar trends to the composite plot derived from the max-transformed charcoal (Fig. 5). This suggests there is little distortion of the signal caused by deriving burnt area using the fxTWA-PLS relationship.

Using pollen data only to reconstruct area burnt can only work in limited conditions. The underlying assumption of the whole methodology is that vegetation (fuel) availability (derived by pollen assemblages) would vary through time only due to fire activity/spread. This assumption does not work in many regions on Earth and I am not quite sure it would work within some parts of the IP where there are other important controls and where fuel-limitation of fire activity is less dominant. The paper should clarify this limitation and improve its discussion.

Again there is a misunderstanding of our approach here. We do not assume that the vegetation changes through time only because of fire activity/spread. Our assumption is that changes in the vegetation assemblage can reflect multiple factors, including *fire regime. We have now quantified this using Canonical Correspondence Analysis to examine how much of the variation in pollen abundances is explained by the environmental factors used in our GLM* (Diurnal temperature range, Dry days per month, Wind speed, Gross primary production, Non-tree cover, Cropland, Grazing land, Urban population density) and how much of this variability can be related to burnt area*. These analyses, which we will include in the Supplementary Information, show that some 19% of the variance in the pollen assemblages is explained by environmental factors other than burnt area but that there is also independent information related to burnt area. We will refer to these analyses in the revised explanation of the methodology that we are adding to the Methods section.*

S1. Canonical Correspondence Analysis (CCA) of the environmental controls on pollen assemblages

The central premise of our approach is that fire is *one* of the factors that modify vegetation assemblages, and therefore that differences in vegetation assemblages can be used to reconstruct fire. We used Canonical Correspondence Analysis (CCA) to investigate how much of the variation in modern pollen assemblages could be explained by burnt area alone, as compared to how much could be explained by burnt area combined with other environmental factors. We used the eight environmental factors considered in our generalised linear model (GLM) for this second analysis, specifically diurnal temperature range, dry days per month, wind speed, gross primary production, non-tree cover, cropland, grazing land, and urban population density. The CCA (Table S1) shows that *ca* 19% of the observed variability in the pollen assemblages is explained by the combination of these environmental variables and burnt area, and that burnt area alone explains *ca* 1% of the variability. Thus, the pollen assemblages contain specific information needed to reconstruct burnt area, even though other environmental influences have larger effects.

Table S1. Canonical Correspondence Analysis (CCA) of pollen data

CCA (a) – burnt area and other environmental variables		
	Inertia	Proportion explained
Total	6.189	1.0000
Constrained	1.153	0.1862
Unconstrained	5.036	0.8138
CCA (b) – burnt area only		
	Inertia	Proportion explained
Total	6.189	1.0000
Constrained	0.051	0.0083
Unconstrained	6.138	0.9917

Minor issues:

L30: the assertion that pollen records are more abundant than charcoal records is not valid for many regions on Earth. Maybe this generalisation is true for the Iberian Peninsula, but this has to be clarified.

There are certainly more pollen records than charcoal records from the Iberian Peninsula (112 sites versus 54 sites). This imbalance is also true globally. There were 736 sites with charcoal records globally in version 3 of the Global Charcoal Database (Marlon et al., 2016) and we currently have a total 1400 sites with charcoal records represented in the Reading

Paleofire Database, which is an updated version of the GCD. This compares to 1151 sites with Holocene pollen records from North America (Gajewski et al., 2019), 879 sites with Holocene pollen records from Europe (see e.g. Mauri et al., 2015) alone. Indeed, we do not know of any region where the number of charcoal records available exceeds the number of pollen records. However, since we do not quantify this in the paper, we will modify the last sentence of the abstract to read:

This new method opens up the possibility of reconstructing changes in fire regimes quantitatively from pollen records, which are often more numerous than charcoal records.

LL32-40: the first paragraph of the introduction reads like a series of loosely connected statements. The rationale for this work needs to be apparent in this paragraph, but at the moment it's quite confusing.

We agree that the first paragraph only provides a generic statement about why it is important to understand fire regimes. It was designed to lead on to explaining (1) why it is important to look at past fire regimes, and (2) why we exploit vegetation data as a way of doing this. However, we can certainly restructure the introduction to make our rationale clearer. Since many of the comments below also address statements in the introduction, we will first address these and then provide a revised version of this section (see below).

L42: the Holocene is certainly a period when human agency was 'pervasive' in many regions. This is another generalisation that needs to be better expressed.

We are not intending to suggest that there was no human influence on the landscape before the industrial revolution, but simply that it was less pervasive than today, and this is true if only because of the much smaller population sizes. We will clarify this by rewriting this statement as follows:

Reconstructing changing fire regimes during the pre-industrial Holocene (12000 yr B.P. to ca 1850 CE), provides an opportunity to investigate the controls on fire over timescales when human influences on the landscape, including fire regimes, were more localised and less profound than they have become during the industrial era.

L49: remove 'qualitative'

Much of the literature interpreting charcoal records is indeed qualitative (more fire, less fire). However, we are happy to remove the term and simply say this is a semi-quantitative measure:

charcoal records only provide a semi-quantitative index of fire activity rather than quantitative estimates of burnt area or biomass loss.

LL54-57: this section does not consider climate into the equation, assuming that fires are fuel-limited. To make it work as a general statement, this should include susceptibility to burn. Alternatively, if this is only referring to the Iberian Peninsula, where fuel availability plays a more important role, this needs to be specified. This is still probably a generalisation, but it works better to introduce the study region.

Analyses of the drivers of modern fire regimes at a global scale have shown that climate, vegetation and human factors all contribute to determining the incidence of fire. Nevertheless, all of these studies show that vegetation properties, such as primary production and the relative amount of tree versus grass cover, are the most important of these drivers - as we state in this sentence to explain why it should be possible to use *palaeo-vegetation data to reconstruct fire histories. Nevertheless, we can expand this text to explain this more clearly as follows:*

Although the occurrence of fire is influenced by multiple factors, analyses of present-day fire relationships globally using satellite-derived data have shown that vegetation properties determining fuel availability are the strongest determinants of fire occurrence (Bistinas et al., 2014; Forkel et al., 2019a, 2019b; Kuhn-Régnier et al., 2020).

LL54-64: this whole paragraph is quite jumpy and confusing (starts with fuel availability, then it goes to pollen assemblages as a method to reconstruct past climates)

We agree that it is not necessary to introduce the fx-TWAPLS methodology in the Introduction and will remove this.

LL72-74: This section makes your previous inference about the importance of fuel availability less valid and the whole approach more confusing. I think there needs to be a section introducing the drivers of fires in the IP.

The modern vegetation patterns strongly reflect climate gradients across the Peninsula, but we agree that we should have made this link to vegetation clearer.

Thus, we propose revising the Introduction as follows:

Fire is an important element in many ecosystems and in the Earth system (Bowman et al., 2009; Resco de Dios, 2020). It impacts vegetation dynamics, ecosystem functioning and biodiversity (Harrison et al., 2010; Ward et al., 2012; Keywood et al., 2013). It also affects climate through vegetation changes and the release of trace gases and aerosols. Fire directly impacts socio-economic assets (e.g. Stephenson et al., 2013; Thomas et al., 2017) and has deleterious effects on human health though releasing smoke and particulates into the atmosphere (e.g. Johnston et al., 2012; Yu et al., 2020). These impacts make it important to understand what controls on the incidence and severity of fires.

Analyses of fire regimes during the satellite era have shown that multiple factors play a role in determining the occurrence of fire, including climate and fire weather, vegetation properties and human activities (e.g. Harrison et al., 2010; Brotons et al., 2013; Bistinas et al., 2014; Knorr et al., 2014; Andela et al., 2017; Forkel et al., 2019a, 2019b; Kuhn-Régnier et al., 2020). However, the satellite record only covers a short time period (*ca* 20 years) and the impact of anthropogenic changes to land cover in suppressing fire during this interval is strong (Andela et al., 2017). Reconstructing changing fire regimes during the pre-industrial Holocene (12000 yr B.P. to ca 1850 CE), provides an opportunity to investigate the controls on fire over timescales when human influences on the landscape, including fire regimes, were more localised and less profound than they have become during the industrial era.

Sedimentary charcoal, preserved in lakes, peatbogs and other anoxic environments, has been widely used as an indicator of past changes in fire regimes (Marlon et al., 2008, Power et al., 2008; Daniau et al., 2012; Marlon et al., 2016; Vannière et al., 2016; Connor et al., 2019). Evaluations that combine charcoal-inferred palaeofire reconstructions with past hydrological, vegetation, and archaeological data support the idea that there are strong relationships among climate, fire, vegetation and human activities (Carrión et al., 2007; Marlon et al., 2008; Gil-Romera et al., 2010; Turner et al., 2010; Vannière et al., 2011; López-Sáez et al., 2018; Morales-Molino et al., 2018). However, charcoal records only provide a semi-quantitative index of fire activity rather than quantitative estimates of burnt area or biomass loss. Attempts to calibrate the charcoal record to provide quantitative estimates of proximity or area burnt are either site-specific (Duffin et al., 2008; Hennebelle et al., 2020) or rely on modelling (Higuera et al., 2007). Furthermore, although the number of charcoal records is increasing, there are still comparatively few sites compared to other types of palaeoenvironmental data and this can make it difficult to make regional reconstructions of changing fire regimes.

Although the occurrence of fire is influenced by multiple factors, analyses of present-day fire relationships globally using satellite-derived data have shown that vegetation properties determining fuel availability are the strongest determinants of fire occurrence (Bistinas et al., 2014; Forkel et al., 2019a, 2019b; Kuhn-Régnier et al., 2020). This suggests that palaeo-vegetation data could provide a way of reconstructing burnt area in the past, particularly at times when human influences on land cover were less important. This would also allow us to capitalise on the more extensive site networks for palaeo-vegetation.

In this study, we present a new method to reconstruct quantitative changes in fire regimes over the Holocene. We relate the relative scale of modern charcoal abundance to absolute burnt area using a conversion factor derived from a generalized linear model (GLM) of fire probability based on burnt area data. We then derive quantitative relationships between pollen assemblages and inferred burnt area using Tolerance-weighted Weighted Averaging Partial Least-Squares with a sampling frequency correction (fxTWA-PLS: Liu et al., 2020). The vegetation-burnt area relationship is then used to reconstruct changes in burnt area through time from pollen assemblages, including at sites with no charcoal record. We use the Iberian Peninsula as a test case. The Iberian Peninsula is the most fire-affected region in southern Europe (Jesus et al., 2019; Molina-Terrén et al., 2019). Although the modern fire regime is partly driven by human activities, the patterns also reflect the strong climate and vegetation gradients across the region. Although much of the Iberian Peninsula has a typical Mediterranean climate, parts of the region are influenced by proximity to the Atlantic Ocean or the Mediterranean Sea and by the mountainous topography, giving rise to complex weather and climate patterns and large gradients in vegetation type and diversity (Loidi, 2017). We reconstruct fire regimes across the Iberian Peninsula through the Holocene and discuss the implications of the reconstructed changes.

L79: this is a general reference to the EPD, but a list of record with references needs to be provided in supporting information

This information is all included in the data set we provide. However, we will provide a list of records with references in the Supporting Information.

Table S2. Information on the pollen records. Latitude: degrees decimal where +ve is N and –ve is S. Longitude: degrees decimal where +ve is E and –ve is W. Elevation: in metres above sea level. Source: EPD = European Pollen Database (www.europeanpollendatabase.net); PANGAEA = www.pangaea.de/ *(Here, we show the first five rows of the table)*

Site name	Entity	Source	Latitude	Longitude	Elevation	Reference
Almenara de Adaja	ADAJA	EPD	41.19	-4.67	784	(López Merino et al., 2009)
Alsa	ALSA	EPD	43.12	-4.02	560	(Mariscal, 1993)
Alvor Estuary Ribeira do Farelo Ribeira da Torre	Abi 05/07	author	37.15	-8.59	1	(Schneider et al., 2010, 2016)
Antas	ANTAS	EPD	37.21	-1.82	0	(Cano Villanueva, 1997; Pantaléon-Cano et al., 2003; Yll et al., 1995)
Arbarrain Mire	ARBARRAIN	author	43.21	-2.17	1004	(Pérez-Díaz et al., 2018)
…	…	…	…	…	…	…

L84: list with references needed for the charcoal records too

This information is all included in the data set we provide. We will provide a list of records with references in the Supporting Information.

Table S4. Information on the charcoal records. Latitude: degrees decimal where +ve is N and –ve is S. Longitude: degrees decimal where +ve is E and –ve is W. Elevation: in metres above sea level. (*Here, we show the first five rows of the table*)

Site name	**Entity name**	**Latitude**	**Longitude**	**Elevation**	**Reference**
Alvor Estuary Ribeira do Farelo Ribeira da Torre	Abi 05/07_100minus	37.15	-8.59	0.6	(Schneider et al., 2010, 2016)
Alvor Estuary Ribeira do Farelo Ribeira da Torre	Abi 05/07_100plus	37.15	-8.59	0.6	(Schneider et al., 2010, 2016)
Arbarrain Mire	Arbarrain Mire core	43.21	-2.17	1004	(Pérez-Díaz et al., 2018)
Armacao de Pera Ribeira de Alcantarilha	ADP 01/06_100minus	37.11	-8.34	2.4	(Schneider et al., 2010, 2016)
Armacao de Pera Ribeira de Alcantarilha	ADP 01/06_100plus	37.11	-8.34	2.4	(Schneider et al., 2010, 2016)
…	…	…	…	…	…

L270: unclear links between paragraphs

The previous paragraph in the Discussion emphasises the climate controls on vegetation and fire, and the fact that the modern gradients have been present although not constant through the Holocene. However, there is a substantial literature on the potential influence of human activities on fire regimes and we seek to address this here. We can make the transition more apparent by re-writing the first sentence as follows:

Our analyses show that climate, and climate-induced changes in vegetation, have influenced the fire regimes of the Iberian Peninsula during the Holocene. However, many studies have suggested that human activities could also have been important (Blanco-González et al., 2018; Connor et al., 2019; Feurdean et al., 2020).

L272: this is of course true, but fire is not the only way to achieve land clearance and this approach assumes pollen assemblages are only varying in response to area burnt

It is true that fire is not the only way to achieve land clearance, although it has been invoked specifically as a method for the Iberian Peninsula e.g. by Connor et al. (2019). This is one reason for looking to see whether there is a relationship between reconstructed fire and the onset of regional agriculture. However, we acknowledge that the onset of agriculture was likely non-synchronous across the Peninsula, and thus the lack of an apparent relationship may hide linkages at a more local scale. However, investigating this possibility requires more detailed local reconstructions of the time sequence for agricultural expansion, and is beyond the scope of the current paper.

The reviewer is mistaken in stating that our approach assumes that the pollen assemblages are only varying in response to area burnt. The pollen data are multivariate by nature and this allows us to reconstruct changing fire regimes through changes in the assemblages because some taxa in the assemblage are sensitive to fire. There may however be changes in the assemblages due to e.g. climate and/or human activities, and these changes are independent of the changes in fire regime.

LL287-289: unclear sentence, confusing how the scarce availability of charcoal records would have led to large-scale patterns (these normally require lots of records)

Our point here is that the limited availability has meant that analyses of charcoal data have focused either on individual sites or on very broad regions (e.g. continental scale syntheses. However, we agree that our meaning here was not clear and we will re-write this as follows:

The limited availability of charcoal records has meant that the analysis of past fire regimes has tended to focus on large-scale zonal or continental-scale patterns (e.g. Marlon et al., 2008; Power et al., 2008; Daniau et al., 2010; Vannière 290 et al., 2011). Our new methodology opens up the possibility of reconstructing changes in fire regimes from pollen data and thus of examining finer-scale patterning that might reflect climate or human influences on fire.