This paper combined satellite observations of fire behaviour, tower-based CO and CO2 mixing ratios and bottom-up approach to estimate forest fire emissions in France with a focus on improving emissions from smoldering. I found the methods used by the authors in general credible and the paper advances the quantification of fire emissions induced by forest fires. I have a few major comments mainly regarding clarifications of the methods being used and some minor technical comments (detailed below).

Major comments:

- I suggest adding a paragraph giving an overview of the methods, preferably with a flowchart figure, focusing on how different approaches are combined and connected.

After a thorough reading of the paper, we recognize that this part needs further explanation. We have therefore produced a new figure in the appendix, and greatly modified section 3.5.

- What's the major purpose of Hysplit model? I don't really see how it is connected with the selection of tower sites and determination of background measurement…. Is it only used to justify that most of the ICOS sites are free of influences of Mediterranean forest fires and hence their measurement could be considered as background ones? Fig. 3 is nice but also quite unique I guess. Is it a sufficient example to argue that, based on Hysplit simulations, most of the ICOS sites are free of influences of Mediterranean forest fires and hence their measurement could be considered as background ones?

In this study, the Hysplit forward-trajectories were used to identify the arrival times of the forest fire plumes to the observation site. The results were cross-checked by checking for CO2/CO anomalies in the observation data happening at the same time. As for the back-trajectories, we modeled them to define the influence matrix describing the Source-Receptor relation on a 0.05 x 0.05 deg grid. The influence matrices were used to check for additional sources influencing each tower. We also used those two approaches to check the transport towards inland stations where spikes in CO2/CO were observed. The results showed that the plumes from both fires reached those stations at the same time.

At first, we wanted to look into the data collected by the whole ICOS-France network. However, after checking the model results, we discarded the distant stations since the measured signals turned out that most detected plumes corresponded to multiple overlapping plumes (i.e. multiple fire plumes reaching the ICOS station at the same time). Also, as stated previously, we used the model data to ensure that no other sources could be linked to the signals we detected (e.g. cities, power plants, ecosystem respiration). This was the case of the OHP station where many signals were discarded because of their potential anthropogenic origin.

We added this explanation to the text as follows:

> **This step was accomplished using the Hybrid Single Particle Lagrangian Integrated Trajectory (Hysplit) model (Stein et al., 2015). In a backward-in-time configuration, particles were released from the receptor site and monitored over 7-day intervals. The result is a footprint matrix representing the influence of the area around the receptor on the measurements. The model spatial resolution used is 0.05 x 0.05 deg. The Global Forecast System (GFS) meteorological model (National Centers For Environmental Prediction/National Weather**

Service/NOAA/U.S. Department Of Commerce, 2015) provided the atmospheric conditions (wind and turbulence) to drive these particles from the receptors to the sources in the Hysplit simulations. The GFS outputs, featuring a horizontal resolution of 0.25° x 0.25° and 3-hourly time intervals, served as the meteorological inputs. We also conducted Hysplit simulations in a forward-in-time configuration releasing particles (600 per hour) from the fire locations, over the fire duration from the exact burned area. In this configuration, we simulated the transport of the plume from the fires to the ICOS stations. By tracking the arrival times of the fire-emitted particles within an influence region surrounding each atmospheric tower, we successfully attributed a fire source to each anomaly.

- If I understand well, it seems that the characterization of fire behaviour using satellite data is independent of the bottom-up estimation of fire emissions, in particular, the fire behaviour information has not been used to determine the key parameters in Equation (2) (e.g., SFp) and parameters in lines 330-334.

  Our first hypothesis was indeed that rate of spread could be an index discriminating soil smoldering fires. However, as stated Line L430-431 this was not the case, as the ROC fire in the temperate forest and affecting the peatland was fast-spreading according to hotspots. To consider which fire would experience smoldering combustion other than the 3 fires from which could measure MCE, we relied on fire duration based on the detection of hotspots within the fire polygon after the initial spreading leading to the final fire shape up to 5 weeks after ignition. Figure A2 illustrates that the discriminating factor between the ROC and BIS fires experiencing smoldering combustion and the OHP mediterranean fire is the long lasting detection of hotspots. We selected this index to consider SOM combustion or not in our study. L645-666 we discussed what could be a discriminating factor for considering SOM smoldering in temperature Europe. Beside peatlands references in land cover datasets, we ended up with suggesting the slow decomposing needleleaf to be a major discriminating factor under mild and wet climate conditions limiting their decomposition and accumulating SOM. Unfortunately, fully detecting smoldering fires remain impossible yet from remote sensing, and our result more warn about potential underestimation of fire emission in temperate ecosystem than propose a defined criteria to identify actually lead to soil smoldering or not.

- This is somewhat a little disappointing. Following this logic, it then seems that the key strength/advancement of the paper is that the authors compiled a nice Table 2, a range of more credible sources of fuel load, and the used satellite-derived fire information and power-based MCE to *indirectly* verify their bottom-up estimate of emissions? Is this correct? This point has to be made clearer when the authors address my first major comment.

  Indeed, fire spreading was not a convincing indicator. the presence of peatlands, hotspot duration after the fire spread leading to the final fire shape and leaf type (needles) are our main hypothesis that we used to calculate SOM smoldering or not. we better explain this step in the flow chart.

- The authors examined three typical fires, or fires in three typical forests using satellite-based fire behaviour and power-based mixing ratio measurements. These are then used to support their bottom-up approach. Then then the challenge is how we can ensure that the upscaling to the national level using their bottom-up approach is also reliable, given that fires are highly temporal and spatially heterogeneous in terms of fire bebaviour, fraction of flaming versus smoldering, combustion completeness etc. (I believe the authors have tried to address well the spatial heterogeneity in fuel load)?

We provided here a conservative approach to implement SOM and peatland smoldering combustion compared to GFAS fully omitting this component. We could identify from the 3 fires tested that hotspot duration, the presence of peats and the leaf type (needleleaf) associated to high SOM content (calculated from GFED) would be the information to upscale our finding to the other fires. other smoldering fires might happen under other conditions that we could not identify yet. our work contributes to a first attempt, based on flux tower evidences, in starting and implementing smoldering fires in temperate forests, neglected until now. We cannot ensure a full reliability that other smoldering fires did not happen, but we believe our selection criteria were conservative not to overestimate SOM smoldering. No other fires than ROC and atlantic forest fires actually met the criteria in our study.

Minor comments:

Line 139: some introduction on VIIRS data is necessary because it seems an important limitation on what fires have been analyzed.

We better explain what is VIIRS, its spatio-temporal resolution and performance in detecting fires now in paragraph 2.2.1 (L140-145).

Line 145: “beyond the fire outbreak ”. What does ‘beyond’ mean here?

we rephrased the sentence by using ‘after the fire ignition date ’ instead of ‘beyond fire the fire outbreak’.

Line 146–147: I don’t see how the approach described here (visual examination of RGB spectrum) could be reconciled with BAMTS… So what is exactly the role of BAMTS in burned area detection? And how are these two further linked with random forest classifier and how the classifier is used and for which purpose?

we fully rephrased L153-160 paragraph 2.2.1 describing this keystone step. as a first step, BAMS automatically generates RGB spectral differences between the pre and post fire period over the study region and according to the ignition date fixed by the user. the user sets up (by visual examination of the spectral difference map) manually the burned and unburned training zone as two samples of the study region. the random forest classifier (as part of the BAMTS tool) then reclassifies the whole study as burned or unburned. omissions and commissions errors can happen in this first step, so the user can enlarge the training regions (to cover the commission error and include them in the ‘unburned’ training zone for example) to reach the cleanest fire polygone as possible. This final step is also dependant on a visual inspection of commission and omissions errors based the spectral difference map.

Line 203: “corresponding to a single grid cell. ”. Which model does this grid cell refer to? What is the spatial resolution of Hysplit?

This statement refers to the hysplit grid cell which has a 0.05 x 0.05 deg resolution. The statement was modified and the following sentence was added to the text.

The model spatial resolution used is 0.05 x 0.05 deg.

Table 1: Better to report R2 rather than R. The same for the texts.

As per the suggestion of the reviewer, we replaced R by R2

Line 168: what is this 6-hour data?

Hotspot data (thermal anomalies) are obtained with VIIRS sensors (S-NPP and NOAA) and MODIS sensors (Aqua and Terra) **The time overpass varies between 3 and 9hours that we averaged at 6hours. We now provide the true range.**

Line 199: Is this 600 per hour particle numbers typically used in transportation modeling? How does this influence the results?

The is no typical standard number of particles to release. We gradually increased the number of released particles until the full extent of the fire plumes was reached. Larger release rates were not producing any additional information. We didn't include this part of the work to avoid a long technical description.

Line 201: "By tracking the arrival times of these particles within an influence region surrounding each atmospherictower, we successfully attributed a source to each anomaly", I don't understand the latter half. Could you please explain?

The role of the Lagrangian model in this study is to determine the source-receptor relationship between the forest fires and the CO and CO2 peaks in the observed data. In the forward in time mode, we released the particles from the fires locations and tracked them in space and time to determine if the high CO and CO2 values observed at a certain ICOS tower correspond to the arrival of the plume from a fire. In the model, this corresponds to the particles arriving at a predefined influence region around the ICOS tower. This is important to identify other potential sources of anthropogenic origin that could influence the results of this study.

Line 298-299: I don't understand what you mean by 'baseline' here.

we rephrased the sentence L298

Table 2: I cannot reconcile/connect Table 2 with lines 330–335. (1) you provide only constant SF values in Table 2. But if SF values do no change among the flaming phase, mixed phase and smoldering phase, then how is this used in Equation (2)? (2) lines 330-335 seems giving proportions of fuels being affected by fire, what is the difference between this and CC in Table 2? Seems that lines 330-335 should be better integrated with Table 2 so that you have only a single source to present the parameters used in emissions calculation. (3) how the information in lines 330-335 is used in Equation (2)? (4) how do you choose CC values between its min and max values in Table 2?

we names the three phases flaming, mixed and smoldering, which brought confusion to the interpretation of the results. we now name these phases respectively as spreading,

mixed and post-spreading stages, as both flaming and smoldering can happen during the 3 stages. during the spreading stage (early after ignition) we considered that 50% of the burned area is affected by flaming, while soil smoldering did not start yet. during the mixed stage, we considered that the remaining 50% of the burned areas is still under flaming while the previous 50% enter the smoldering stage and affecting 50% of the soil c stock (so 25% of the total C stock combustion during the whole fire duration). during the post-spreading stage, we considered no more flaming, and smoldering combustion consuming the remaining 75% of the total soil C stock affected during the whole fire duration. in turn, summing up the 3 stages, leads to the usual emission estimates of total C stocks affected by combustion but calculated all at once. This splitting along the fire progression could capture the temporal dynamic of MCEs and capture the very low MCEs that we could only attribute to lignite high CO emission factors according to the EF synthesis from table 2.

We rephrased the description of this part in the manuscript with using this renaming of stages to prevent confusions.

Line 352: TROPOMI data not explained in Methods.

Since the data was not used in this study, we did not provide a description of it.

Figure 4: what is the difference between 1-hour and 1-minute? Are they the temporal resolutions of the data ? what is the temporal resolution of measurement over the towers?

The Picarro data from the ICOS sites are available at 1-hour and 1-minute temporal resolutions. When multiple sampling heights are present on a single ICOS tower, the 1-hour interval is split equally between all levels. Thus, for the towers with three levels the 1-hour data is an aggregation of almost 15-20 minutes only. This along with the ability of the 1-minute resolution to capture faster variations of the mixing ratios, motivated the comparison between both resolutions.