Revision of Manuscript entitled "Bayesian spatiotemporal modelling of wildfire occurrences and sizes for projections under climate change"

July 8, 2024

Dear Referees, dear Editor, we would like to thank you a lot for your careful evaluation and the insightful reviews of our manuscript. Please find below a point-by-point response that addresses the concerns and suggestions of each reviewer. In order to ease the reviewing process, in the main paper we coloured in blue all the text parts which are new or have been modified during this revision.

Referee A

Strengths And Weaknesses:

- The model used is very interesting and seems to fit the data well. However, it wasn't fully clear to me how the proposed model differs from the Firelihood model or the model proposed by Koh et al. (2023), both of which are referenced by the authors in Section 1. The authors should clarify the differences at the end of Section 1 to highlight the novelty of their approach. We have added some text in the revised version to better explain the differences with the classical Firelihood model and the paper of Koh *et al* (2023).

  The general structure of the model proposed here is quite similar to Koh et al. (2023) and can be seen as a sub-model of the models considered in their paper. The main approach in Koh et al. (2023) is generally more flexible but comes at the cost of a quite complex modeling procedure based on splitting the wildfire sample into moderate and extreme wildfires based on a threshold for their sizes, leading to a model that is quite challenging to construct and to estimate. In the model proposed in the current manuscript, there is no need to choose a threshold for the size component. We point out that the original Firelihood model (in Pimont *et al* 2021, and several follow-up publications) is formulated with even several thresholds, but it uses simpler step-function specifications for the nonlinear effects of predictors, and not continuous spline functions as inKoh et al. (2023) and also here.

  By contrast to these two other papers, we here perform modeling for a study region that has never been modeled before and for which climate and also land-use conditions are quite different from the classical study area in the historical wildfire-prone area in Southeastern France used for the original Firelihood model and also the model of Koh et al. (2023). A particularity is that there are generally fewer wildfire occurrences in the region studied here. Moreover, we explore the combination of posterior simulation and climate model output to provide projections of wildfire risk under climate change for this region.

  Finally, a methodological novelty is that we show how these projections can be smoothed using R INLA.

  The following sentence has been added in line 93 page 4: *In contrast with Koh et al. (2023), our modelling approach shares a similar structure, but is tailored for a previously unmodeled study region characterized by specific climate and land-use conditions. Furthermore, a novel workflow will be devised to simulate and smooth projections of wildfire risk under climate change, using the INLA approach.*

  Another comment is that, in line with the scope of Computo journal, one goal of this paper was to provide a step-by-step tutorial with clean and documented code for using the INLA-SPDE approach, such that it could easily be applied to other data with similar structure.

- The use of climate projections to infer future wildfire behaviour is an interesting idea. However, I have a couple of concerns about the approach. First of all, it doesn't seem like the authors make any attempt to de-bias the climate projections. The models are fitted using covariates derived from a reanalysis, and then the climate model data is fed into the model "off-the-shelf". Without any form of calibration of the different climate projections to the observations from the reanalysis, I can't say that I find the results in Section 5 particularly meaningful. It would make more sense to me if the reanalysis data was used to form the projections, potentially via a simple linear regression model, and then this information used in the analysis of Section 5 (rather than the projections from the four separate GCM models).

  The projection datasets that we use are part of the reference projections for France obtained by running couples of a General Circulation Model and a Regional Climate Model, and they have already been appropriately debiased; see information on the website of the DRIAS project (`https://www.drias-climat.fr/`) where these simulations can be downloaded. We have added more precise information on the projection data in the revised manuscript (after second paragraph in Section 2): *In Section 5, we consider climate model projections from four different climate models. These projection datasets are obtained by running a coupling between a General Circulation Model and a Regional Climate Model over the period 2020-2100. The simulated data were already appropriately bias-corrected and can be downloaded on the DRIAS project website.*

- Secondly, the authors set extrapolated f_{FWI} and f_{FA} values to a constant. It's not clear how often (if at all) this extrapolation occurs during the procedure described in Section 5 or what these constants are. I assume the constant is just taken to be the point estimate of the spline at the edge of the domain, e.g., at FWI = 40, you take the predictor equal to approx 1 for FWI > 40 (from Figure 3). However, I imagine there is quite a large amount of uncertainty in the estimates of the splines at the edge of the domain, which is a fairly common problem with spline-based models (there's evidence of overfitting in FA and FWI in Figure 4, where we see the predictor decrease with large FA or increase with low FWI). If inference in Section 5 for future wildfire distributions relies quite heavily on this choice of extrapolation constant, how are we able to trust the results?

  First, we would like to point out that the two notions of extrapolations described for FWI and FA are not the same.

  For FWI, we extrapolate the estimated spline function f_{FWI} using the value $f_{FWI}(x_0)$ estimated for the right endpoint x_0 of the support of the spline function, as you describe it above. Various results in already published work have shown that the FWI effect becomes saturated at very high levels, i.e., it does not further increase even if the FWI values increase, as for example discussed in the *Firelihood* papers Pimont *et al* (2021) or Koh *et al* (2023). The reasons for this can be found in the construction of the FWI as an index for Canadian forests, where the influence of variables such as relatively high wind speeds is stronger than for France. Therefore, a constant extrapolation does not seem problematic in this case. There certainly is higher variance in estimated spline functions for the high FWI values, and constructing novel models that appropriately address this issues is still an open research question and is too complex to be addressed here; see also the written comment on this problem published in the context the JRSS "The First Discussion Meeting on Statistical aspects of climate change" by Legrand and Opitz (2023), that is written by some of the authors of the current manuscript. To better highlight this issue and potential future work, we have added the following sentence in the discussion: *One limitation of our approach for the simulation of future wildfire activities is the constant extrapolation of the function f_{FWI}. Various studies have demonstrated that the FWI effect becomes saturated at very high levels, suggesting that a constant extrapolation may not be problematic. However, there is a higher variance in the estimated spline functions for the high values of FWI, and developing novel models that appropriately address this issue is the subject of future research (see Legrand and Opitz, 2023).*

  Regarding extrapolation of forest area (FA), we here refer to the extrapolation of FA values in time, that is, beyond the study period and into the future. Our approach here consists of using the last available FA values during the study period, i.e., the values of year 2018. As of today, there do not yet exist any reliable projections of future FA in the study area, and therefore our solution is the best one we could think of in these circumstances.

  In order to clarify these two different extrapolations, we have modified the text as follows: *For simulated FWI values in the climate change projections that fall outside the range of historical*

FWI values, we extrapolate the spline function f_{FWI} as a constant using the value $f_{FWI}(x_0)$ estimated for the right endpoint x_0 of the support of f_{FWI}. For the forest surface FA in each pixel, we consider the last available FA values during the study period, i.e., the values of year 2018.

- It's also not fully clear to me what happens with f_{YEAR}. Figures 4 and 5 show some very rough behaviour in estimates of f_{YEAR}; if Section 5 relies on extrapolation of these splines, then I find it very difficult to trust any results in Section 5.

  As stated in the model definition, the yearly effect is considered as an "iid" model (i.e., a Gaussian random vector with one component per year and independence between different years). This should not be confused with the structure of the other effects FA, FWI and DOY, which are defined through quadratic splines to account for relatively smooth behavior. In the end, the relatively rough behaviour of the yearly effect, that can be observed on Figure 3 and 4, is not surprising.

  As you point out, confusion may arise to the fact that a clear explanation was missing on how the sampling of this yearly effect is carried out in Section 5 in order to obtain the future projections of forest fire activities. To address this issue, the following sentence has been added at the end of the second paragraph in Section 5: *Given that we are extrapolating from the historical period for which we here assume that the behavior is stationary across years, we randomly sample values of the yearly effect from its posterior distribution, i.e., we draw at random a year from the study period and then a value from the posterior for this year.*

- The paper is presented as a "how to" guide for modelling marked log-Gaussian Cox processes with INLA, but there's a lot of key information, mostly related to the model formulation, which is missing. Seemingly arbitrary modelling choices are made throughout and not justified, and parts of the paper appear to be unfinished. I've provided specific details below, but I recommend that the authors carefully check the entirety of the paper before resubmitting.
  We have added more details at various places in the manuscript, specifically related to the more specific review comments you provided, to make our modeling choices better comprehensible.

- The journal has a particular focus on reproducibility of studies. Given that the data used in the study are not publicly available, the study described is not reproducible. Following my reservations about Section 5 and overfitting in the spline estimates, I'd argue that the results of the study are not convincingly compelling as to use the SAFRAN data in place of data that are publicly available. However, with the additional clarity in the model procedure that I have requested below, I still think the paper would make a nice addition to Computo as a "how to" guide for modelling.
  The SAFRAN data used for calculating FWI values are now publicly available since the beginning of this year (2024):
  (`https://www.data.gouv.fr/fr/datasets/donnees-changement-climatique-sim-quotidienne/`), and we provide the FWI data along with the manuscript for reproducibility. The situation is different for the wildfire observation dataset used here, which cannot be made public since one of the organisms that is collecting them does not allow for making them public (though, an important part of the dataset we are using can be downloaded from the recently created national French wildfire database BDIFF, that we also cite in the paper, under the following link: `https://bdiff.agriculture.gouv.fr/`). Instead, we have provided a synthetic dataset of wildfires obtained by drawing a realization from the posterior model we have fitted, and we have checked that the estimation with this artificial dataset provides results that are highly similar to the ones obtained for the original data.

Requested Changes:

- Abstract: "...corresponding to the Southeast". I guess this should be South west?
  Correct, thank you for noticing this typo.

- The description of the marked point process in Section 1 doesn't seem to describe the marks at all. From my understanding, the presented model is just the usual space-time point process.
  There is a sentence describing the marks as numerical information attached to each of the points. We have slightly modified to sentence to make clear that it refers to the marks. The

sentence at line 42 has been modified as follows: *Moreover, each point of this spatiotemporal point process can be associated (i.e. marked) with a numerical information, such as the burnt area of the corresponding fire.*

- l55 "...it then appears natural to consider". Why is this the case?

  What we here mean is that Poisson processes are the fundamental model class for point processes (similar to Gaussian processes in geostatistics), and that we can consider the log-Gaussian Cox processes as a Bayesian model where a log-Gaussian prior is put on the intensity function. In the manuscript, this is explained in the sentence afterwards.

- Table 1; What do the first four columns represent?
  Some information was missing here. The paragraph detailing Table 1 has been rewritten as follows:
  In the final dataset, we have for each pixel-day its coordinates given by geographical components – the pixel number PIX, the metric coordinates XL2E and YL2E (in the Lambert 2 Extended metric coordinate reference system with code EPSG 27572), and the corresponding French "département" number – and by temporal components – the year YEAR and the day of the year DOY. For each pixel-day determined as described above, the dataset also includes the corresponding daily FWI, the forest area FA (i.e. the fuel surface), the number of fires that are greater than 0.1 hectares, denoted by NB0.1 (note that, due to issues related to how measurements were made, the data distribution for smaller wildfire sizes has certain biases) and the burnt area BA in hectares (see Table 1).

- Section 2. How are the data collected? Where does the FA variable come form? Why do the authors choose to keep only wildfires between May and October?

  The wildfire data used here are ground data collected by firefighters and fire management agencies. The FA variable is based on identifying forest and other combustible vegetation in the Corine Land Cover database. Some information was missing and the following sentence has been added (Section 2, 4th paragraph): *The forested area (FA, in ha), or the fuel material, is provided by the CORINE land-cover database. This forested area is based on identifying forest and other combustible vegetation in each SAFRAN pixel.*

  As written in the manuscript, we focus on the May to October months which represent the core wildfire season. Wildfires also occur during the other months, but in smaller numbers and sizes, and also due to environmental conditions and triggering mechanisms that are quite different from summer wildfires, such that wildfire experts prefer not to mix them with summer wildfires in modeling approaches. In the manuscript, we have already pointed out these differences between summer wildfires and those occurring in other months. We added the following sentence at the end of the first paragraph of Section 2: *Therefore, in this study, we focus on the summer wildfires, which are more numerous and on average also much bigger in terms of size, and we keep only wildfires that have occurred between May to October. Wildfires occurring during the other months of the year are predominantly driven by different land-use and weather variables than those in summer, such that the "fire regimes" are different, so we focus only on summer wildfires here.*

- The caption in Figure 1 appears unfinished.
  This problem is due to the conversion from the html version to the pdf version, and also certainly to the fact that a punctuation mark was missing at the end of the second subtitle. In the html version, the two graphic columns are better separated, such that the first sub-heading does not appear incomplete. We will check with the editors how to solve this issue.

- Section 3. I didn't quite understand how zero observations would occur for a marked point process dataset. I also didn't follow very well the exposition in the second paragraph.
  Since we aggregate the points of the point process to counts in the space-time voxels (i.e. pixel-days), we obtain many voxels with a count of zero. The first sentence of Section 3 has been modified as follows: *There are* 1308930 *observations of wildfire counts for pixel-days in the wildfire dataset, but only* 2331 *actual wildfires.*

- l146, "...p and q are parameters that can be freely chosen based on the data". Some discussion of this choice would be helpful. It seems quite an important modelling choice, particularly as the authors choose to take p = 0.5.
  Such subsampling schemes were first proposed by Koh et al. (2023), and they also showed

through a simulation study that estimation results are reliable for choices of p and q as in our case. Using $p = 0.5$ allows us to put the same weight on the two subsets in a partition of the zero counts of the dataset, and then setting $q = 0.9$ allows us to relatively strongly overweight relatively large values of FWI (by a factor of five, since $p = 0.5 = 5 \times (1-0.9)$) for which many wildfires occur. In fact, the number of fires that occur for FWI values greater than the 0.9 quantile is 852, which represents more than 35% of the observed fires. Of course, there is no unique "good" choice of these two values. We added in the manuscript the following sentence at the end of Section 3.1: *In our study, using* $p = 0.5$ *allows us to put the same weight on the two subsets in a partition of the zero counts of the dataset, and then setting* $q = 0.9$ *allows us to relatively strongly overweight relatively large values of FWI (note that the proportion of fires occurring for FWI values greater than its* 0.9*-quantile is more than* 35% *of the observed fire occurrences).*

- Section 3.2. What value does M take in your application?

  In our dataset, there are 542 pixels, 161 days per year and 15 years. Therefore, $M = 1308930$ in our specific study.

- When f_{FWI} is defined, n is used for the number of spline basis functions, but has already been used for the sample size.

  Indeed, this is a source of confusion, and we have modifed the notation by using $\ell = 1, \ldots, k$ instead of $k = 1, \ldots, n$.

- Eq. (3). Details on κ and τ are missing.
  Indeed, thank you. We have added to line 184 "$\kappa, \tau > 0$" and to line 187 "$\sigma^2 = \tau^{-2}(4\pi)^{-2}\kappa^{-2\nu}\Gamma(\nu)/\Gamma(\alpha)$", providing a link between τ and the variance parameter σ.

- Eq. (4). Why is the spatial effect here dependent on the departement, rather than the spatial location (as in (1))?
  For the size model, the spatial effect does not depend on spatial location (i.e. vary across spatial pixels) for several reasons. Taking into account spatial locations would result in estimates that would not be very stable, because the distribution of wildfire sizes is quite noisy. It has been found that the spatial dependence between wildfire sizes is not very strong, as highlighted in Castel-Clavera et al. (2023) where the approach of using the same administrative units as here was first proposed. Also, from a more applied perspective, fire-fighting resources are often deployed at the "département" level, so this appears to be a relevant administrative unit when we want to work with units larger than the pixel.
  The following sentence has been added in the manuscript at the end of Section 3.2: *Another remark is that for the size model (4), the spatial effect depends on the "département" rather than the spatial location because the spatial dependence between wildfire sizes is not very strong (see, e.g. Castel-Clavera et al., 2023), so working at a spatial scale that is larger than the pixel appears to be sufficient for the sizes. Additionally, and from a more applied perspective, fire-fighting resources are deployed at the "département" level, making it a relevant administrative unit for our analysis.*

- The choice of a Gamma likelihood for the burnt areas seems odd. The Gamma distribution has exponential tails, meaning it does not have the "heavy tails" mentioned by the authors. This may be why the model fit for the wildfire sizes suffers in the tails. Koh *et al* (2023) use a generalised Pareto distribution, which seems to me a far better choice (and is already implemented in INLA). Did the authors consider this instead?
  The gamma distribution is indeed exponential-tailed but still remains quite flexible for distributions with tails much heavier than that of the exponential distribution thanks to its shape parameter. For example, precipitation data, that are often considered as heavy-tailed, are also modeled with the Gamma distribution with shape parameter smaller than 1. In our estimated model, the Gamma shape parameter has posterior mean 0.89 and posterior standard deviation of 0.025, such that tails are heavier than that of the exponential distribution but not by much. Another reason for our choice was that Koh *et al* (2023) found that the Gamma distribution provided a very good performance among the models that do not adopt a more technical approach where one splits the range of wildfire sizes into several intervals and then provide interval-specific models (e.g. a generalized Pareto model for the exceedances above a high threshold). Note that the generalized Pareto distribution is generally not considered as a good model for the whole distribution; there are recent extensions towards Extended Generalized

Pareto distributions for the whole range of data, but these are so far not yet available in the INLA framework.
The estimated values of the posterior shape parameters have been added at the end of Section 3.5.

- Section 3.3. How were the number of knots chosen? What are PC priors and why are they set as described? How is the number of nodes in the spatial mesh chosen?
  For the splines on FWI and FA, the number of knots was chosen heuristically, through inspection of the histograms. Other ways could have been considered, for instance by taking equidistant knots, which is what is done for the knots of the DOY effect. This information was missing and we added the following sentence: *For FWI and FA, the number of knots is chosen heuristically, through inspection of the histograms. Regarding* f_{DOY}*, the knots are equidistant.*
  PC priors are designed to penalize the complexity of a model by favoring simpler models (or baseline model). They are set as described in the manuscript based on expert knowledge, following the previous study of Pimont et al. (2021). The following sentence has been added at the beginning of Section 3.3: *PC priors respond to a general principle for designing prior distributions to penalize the complexity of a model by defining the prior on a distance towards a simpler baseline model; for example, the baseline model for a spatial Matérn field has variance zero and infinite range, and therefore corresponds to a field that is constant zero.*

  Regarding the number of nodes in the spatial mesh, there is no automatic way but there are some rules of thumb, and an implementation in the R-INLA package that allows using these rule of thumbs in a relatively simple way. The number of nodes is a trade-off between accuracy (with a large enough number of nodes to appropriately capture the spatial variability) and computational efficiency and stability (with not too many nodes, since there is one latent Gaussian variable per node in the model, and very close nodes could lead to instabilities in calculations related to the covariance structure). In our study, we defined the mesh such that we have one node per pixel, i.e., we strongly rely on the pixel discretisation of the dataset to define the mesh. Also, to avoid boundary effects, we define an external boundary farther away from the actual study region. In the extension zone, the triangulation is less dense than in the central region. The following sentence has been added at the end of Section 3.3: *Theory and practical details for the SPDE approach are explained in the reference (online) book of Krainski et al. (2018).*

- Section 3.4. Do the authors have any explanation for the "bump" in the spline estimates for FWI (Figure 3)? Note as well that the caption for this figure is not self-contained, and it would be difficult to understand this figure without the accompanying text.

  The bump arising in the FWI spline function is here certainly just an artefact due to the knot discretisation used for the spline modeling. In general, one expects a relatively smooth and monotonic response of wildfire numbers and sizes to FWI. Since the bump is relatively small, we have avoided trying to "optimize" the number and position of spline knots. This is now explained in a sentence we have added in the manuscript (fourth paragraph of Section 3.4). Regarding the issues with the captions for Figures 3 and 4; they have been modified as follows:
  Fig. 3: Partial effects of the occurrence model: effects of Year, Day of the Year (DOY), FWI, Forested Area (FA) and spatial location. For the 1-dimensional effects, confidence bands are depicted with blue dotted lines.
  Fig. 4: Partial effects of the size model: (top) effects of Year, Forested Area (FA) and FWI; (bottom) spatial effect with the "départements". For the 1-dimensional effects, confidence bands are depicted with blue dotted lines.

- Section 4 would be better placed before the results in Sections 3.4 and 3.5. It would be good to see evidence that the model fits well before interpreting its outputs.

  If the reviewer agrees, we would prefer to keep the ordering as it is. In our view, Sections 3.4 and 3.5 show the estimation, or the fit, of the two models. Once these models are fitted, it appears natural to carry out some simulations, for example by trying to reproduce the data over the observed period, which is done in Section 4.

- Section 4.1. Why do you focus on 2010?

  There is no specific reason on why we chose 2010, as stated in the manuscript: "We chose to examine the year 2010, but any other year could have been considered".

- I did not follow equation (5). The notation Y_i has not been used and there has been no mention of the lognormal distribution thus far (although this may be a more appropriate model than the gamma distribution, see above comment).

  There was some confusion here due to the fact that our notations were not well introduced. The notation Y_i corresponds to the yearly mean for each simulation performed. In other words, if we look only at the first panel of Figure 7, Y_i could correspond to the red dots (i.e. the yearly means of projected fire occurrences under scenario RCP8.5): the goal here was to smooth these yearly means of simulated occurrences using INLA. For that, we considered a simple 1-dimensional SPDE model, defined by Eq. (5). In this specific model, the only effect is the Year. This model should not be confused with the previous models. In order to clarify these aspects, the following sentence has been added in the fourth paragraph of Section 5: *To smooth the projected curves and identify long-term trends in wildfire activity, we implemented a 1D-SPDE INLA model, given by Equation 5, where Y_i denotes, for each of the four cases, the yearly mean (i.e. the yearly mean for the simulated occurrences with each scenario, as well as the yearly mean for the simulated sizes with each scenario).*

Referee B

Requested Changes:

Major Compulsary Comments

My main concern is about the contribution of the submitted work. It is unclear to me whether the contribution is about the analysis of wildfires in the Aquitaine region or whether it is about the use of a bayesian spatiotemporal modelling.

- On the one hand, regarding the modelling approach, what is the novelty compared to the approach proposed by Koh et al. (2023)?

  We have added some text in the revised version to better explain the differences with the Koh et al paper (see also our response to one of the comments of Referee A). The general structure of the model is quite similar, and it can be seen as a sub-model of Koh et al. (2023) which is more flexible (here we do not need to choose a threshold for the sizes). By contrast, we here provide a model for a study region that has never been modeled before and for which climate and also land-use conditions are quite different from the study area of Koh et al. (2023) (for instance, there are fewer wildfire activities in this region than in the South-East of France). Moreover, we explore the combination of posterior simulation and climate model output to provide projections of wildfire risk under climate change. And finally we show how these projections can be smoothed using R INLA, which is also a novelty.

  The following sentence has been added in line 93 page 4: *In contrast with Koh et al. (2023), our modelling approach shares a similar structure, but is tailored for a previously unmodeled study region characterized by specific climate and land-use conditions. Furthermore, a novel workflow will be devised to compute and smooth projections of wildfire risk under climate change, using the INLA approach.*

  Another comment is that, in line with the scope of Computo journal, one goal of this paper was to provide a kind of step-by-step tutorial that can easily be applied to other kind of data.

- Regarding the analysis in the Aquitaine region, it is hard to understand how the methodology is generalizable to other datasets (and /or other contexts). Several choices were made with respect to the applicative context but are not discussed in terms of modeling limitations. For example: In P6L243, are the choices of the parameters n_{ss}, p and q driven by the application or for a modeling purpose?

  The aim of this study is to make it as generalizable as possible. This is one of the reasons why we consider a Gamma distribution for the sizes unlike the model proposed by Koh et al. (2023), which was extremely application-specific. Nevertheless, some choices of parameters are made in our study, mainly for the subsampling step. Firstly, we want to emphasize the fact that this subsampling is done in such a way that the numerical estimation can easily be performed on a personal computer (if the calculation capacity allows it, there is no need for this subsampling step). Then, the choices we made for these parameters are indeed motivated by the data we have at hand: for the fire data, the majority of wildfires occur for large values of FWI, which motivated us to strongly overweight these large values of FWI. In fact, the number of fires that occur for FWI values greater than the 0.9 quantile is 852, which represents more than 35% of the observed fires.

- The estimation of the occurence and the size are performed independently. This point should be clarified and discussed from a modeling (or estimation) point-of-view and regarding the application (is this hypothesis relevant in the Aquitaine dataset?).

  The influence of the covariates considered is not the same for the occurrences as it is for the sizes. While it is possible to link the two components and to include Gaussian effects that are shared between the two components (see Koh et al., 2023), this would result in a more sophisticated model, which would be more complex to construct and more difficult to estimate. In line with your previous comment, the aim of this study is to present a simple flexible model that can easily be applied with another dataset. Furthermore, even if the estimation is conducted independently, the simulation of sizes is conditioned on the simulation of the occurrences, thereby introducing some dependence between the two.

Minor Comments

- Although the size of the grid (8km resolution) is forced by the use of the SAFRAN model, it would have been interesting to have a discussion on the impact of the choice of the resolution in space and also in time. I wonder how the results will be different for a different resolution : for example 32km in space and 1week in time.

  The benefits of higher spatial resolutions were investigated by Castel-Clavera et al. (2023) and the results showed no significant improvement of goodness-of-fit and prediction results for a higher resolution. On the other hand, since the results of our current model but also other papers in the literature clearly highlight that there can be strong variability in wildfire risk between adjacent (i.e. neighboring) pixels, we do not consider it interesting to refit and explore the model at coarser resolutions, especially given the fact that the current manuscript is already quite long. Using coarser scales typically leads to a loss of information since both the response and the covariates, such as FWI, have to be aggregated to the new scale, and one further has to decide how this aggregation should be done. Therefore, these questions go quite far beyond the scope of this paper, but certain aspects have been investigated in other papers. We have added the following remark in the discussion section of the manuscript: *Our spatiotemporal model discretization using the SAFRAN pixels for space and days for time ensured that we implement the model at a resolution where the key weather information from the FWI covariate is well represented. Using a higher resolution to include more precise information on certain land-use and land-cover covariates could slightly improve predictions, whereas using a coarser resolution typically leads to a loss of information about covariates and the response, and therefore to worse predictions (see, e.g., Castel-Clavera et al., 2023).*

- It could have been useful to detail how the FWI is calculated/computed and to provide a reference in P2L113.

  The following details have been added regarding the computation of the FWI; however, since its calculation is quite complex, we only refer to the technical documentation and implementation in our manuscript:

  These data are then combined in order to obtain a Fire Weather Index, hereinafter FWI (van Wagner, 1977). It was first designed for Canadian forests but has been shown to be useful in many other areas of the globe, including France, and has become by far the most widely used meteorological fire-danger index worldwide. The FWI is a unit-less indicator of fire danger, which is computed using daily cumulated precipitation, mean wind speed and temperature, and minimum relative humidity. Its calculation is quite complex (see Bedia et al., 2014) and is achieved here using the `R` package `cffdrs` (Wang et al., 2017).

- In section "5. Future wildfire simulations ..." the climate scenarios are not well detailed. It could be useful to have a table summarizing the main characteristics of each scenario in order to better appreciate the comparative study.

  Thank you this suggestion. Regarding the climate models considered, they were not well detailed. We have added more precise information on the projection data in the revised manuscript (after the second paragraph in Section 2): *In Section 5, we consider climate model projections from four different climate models. These projection datasets are obtained by running a coupling between a Global Circulation Model and a Regional Climate Model over the period 2020-2100. The simulated data can be downloaded on the DRIAS project website..*

  Table 1 below gives an overview of the 4 climate models considered. If the reviewer wishes, we can add it to the appendix.

  Maybe what the reviewer rather wanted are details on the two climate scenarios (RCP4.5 and 8.5) considered. The following sentence has been added at the beginning of Section 5: *We consider hereafter four different climate models under two climate scenarios RCP4.5 and RCP8.5. To summarize, RCP4.5 is a moderate scenario with radiative forcing stabilizing at* $4.5\ W/m^2$ *by 2100, while RCP8.5 is a high-emission scenario with radiative forcing reaching* $8.5\ W/m^2$ *by 2100 and therefore often regarded as a "more pessimistic" scenario (further details can be found on the website of the DRIAS project).*

- In P8L195, why do the author consider a random effect for each "département"?

  For the size model, the spatial effect does not depend on spatial location (i.e. vary across spatial pixels) for several reasons. Taking into account spatial locations would result in estimates that would not be very stable, because the distribution of wildfire sizes is quite

Acronym	Global climate model (GCM)	Regional climate model (RCM)
IPSL-CM5A-WRF	IPSL-CM5A-MR	WRF331F
MPI-ESM-RCA4	MPI-ESM-LR	RCA4
HadGEM-RCA4	HadGEM2-ES	RCA4
CNRM-RCA4	CNRM-CM5	RCA4

Table 1: Summary and key features of the different climate models studied. See the website of the DRIAS project for more details (`https://www.drias-climat.fr/`).

noisy. It has been found that the spatial dependence between wildfire sizes is not very strong, as highlighted in Castel-Clavera et al. (2023) where the approach of using the same administrative units as here was first proposed. Also, from a more applied perspective, fire-fighting resources are often deployed at the "département" level, so this appears to be a relevant administrative unit when we want to work with units larger than the pixel.
The following sentence has been added in the manuscript at the end of Section 3.2: *Another remark is that for the size model (4), the spatial effect depends on the "département" rather than the spatial location because the spatial dependence between wildfire sizes is not very strong (see, e.g. Castel-Clavera et al., 2023), so working at a spatial scale that is larger than the pixel appears to be sufficient for the sizes. Additionally, and from a more applied perspective, firefighting resources are deployed at the "département" level, making it a relevant administrative unit for our analysis.*

Typos

- P385: folowing $->$ following
  Done, thank you.

Referee C

Strengths And Weaknesses:

I had some difficulty understanding the novelty of the paper compared to Koh et al. (2023) and Pimont et al. (2021). I believe that the model proposed herein could be considered a sub-model of Koh et al. (2023). The one novelty that I could identify was in simulating future wildfires while combining the posteriors of model parameters with climate model output. This is a great idea, though I feel that this can be further improved, especially in the way the predictor variables FWI and FA are created for future projections. I'm wondering whether FWI and FA cannot be predicted or calculated based on reanalysis data or using available historical data, and then using these predicted FWI and FA when simulating future wildfires? Please disregard if this involves too many technicalities. However, assuming FWI and FA surfaces remain constant above historical ranges while projecting into the future might not be the most accurate approach, especially when using scenarios like RCP8.5. These clarifications should be clearly stated in the manuscript and/or in the discussion section if your goal is to do it as future research.

Thank you for these comments. We have added some text in the revised version to better explain the differences with the paper of Pimont *et al.* (2021) and the paper of Koh *et al* (2023).

You are totally correct that our model can be viewed as a submodel of Koh et al. (2023). However, we here perform modeling for a study region that has never been modeled before and for which climate and also land-use conditions are quite different from the classical study area in the historical wildfire-prone area in Southeastern France used for the original Firelihood model and also the model of Koh et al. (2023). A particularity is that there are generally fewer wildfire occurrences in the region studied here.

Moreover, and as well identified in your remark, we explore the combination of posterior simulation and climate model output to provide projections of wildfire risk under climate change for this region. Also, a methodological novelty is that we show how these projections can be smoothed using R-INLA.

Regarding the extrapolation of the predictor variables, we fully agree with you that this might not be the most accurate approach, but for lack of better solutions, we consider it the best possible approach.

However, regarding the FWI, various results in already published work have shown that the FWI effect becomes saturated at very high levels, i.e., it does not further increase even if the FWI values increase, as for example discussed in the *Firelihood* papers Pimont *et al* (2021) or Koh *et al* (2023). The reasons for this can be found in the construction of the FWI as an index for Canadian forests, where the influence of variables such as relatively high wind speeds is stronger than for France. Therefore, a constant extrapolation does not seem problematic in this case. There certainly is higher variance in estimated spline functions for the high FWI values, and constructing novel models that appropriately address this issues is still an open research question and is too complex to be addressed here; see also the written comment on this problem published in the context the JRSS "The First Discussion Meeting on Statistical aspects of climate change" by Legrand and Opitz (2023), that is written by some of the authors of the current manuscript. To better highlight this issue and give some directions for future work as suggested, we have added the sentence in the discussion: *One limitation of our approach for the simulation of future wildfire activities is the constant extrapolation of the function f_{FWI}. Various studies have demonstrated that the FWI effect becomes saturated at very high levels, suggesting that a constant extrapolation may not be problematic. However, there is a higher variance in the estimated spline functions for the high values of FWI, and developing novel models that appropriately address this issue is the subject of future research (see Legrand and Opitz, 2023).*
For the extrapolation of the FA, as of today, there do not yet exist any reliable projections of future FA in the study area, and therefore our solution is the best one we could think of in these circumstances. In fact, projections of future FA (or, at least, the creation of some plausible "storylines" how future FA could look like) is one of the goals of the FIRE-RES project, under which this study was conducted.

Another concern of mine is related to jointly modeling counts and size processes by the use of sharing random effects. You did mention the possibility of sharing some of the spatial random effects between the occurrence model and the size model, similar to Koh et al. (2023). While computational cost is one consideration, I believe you might see improved performance due to strong dependencies between the two processes and enhanced uncertainty estimations. This is something I will recommend to worth give a try to a datasets that you can handle computationally. Again, this is the trade-off between model flexibility and computational costs.

Indeed, linking the two components, as done in Koh et al. (2023), would result in a more sophisticated model, which would be more complex to construct and more difficult to estimate. This would be the price to pay in order to obtain possibly more accurate (and narrower) uncertainty estimations. However, in line with one of the scope of the journal, the aim of this study was rather to present a simple flexible model that can be easily applied with other datasets. Moreover, even if the estimation is conducted independently, the simulation of sizes is conditioned on the simulation of the occurrences, thereby introducing a certain degree of dependence between the two.

Furthermore, the model proposed here includes spatial and temporal random effects in an additive fashion, hence lacking space-time interaction. I think this effect is evident from Figure 8, where the spatial pattern of estimated counts and burnt areas for future projections appears very similar to those in historical periods. I believe this could be an artifact of using a model that lacks space-time interactions. For instance, the models proposed herein account for temporal dependence, which is independent a priori from spatial effects. A proper spatiotemporal model would incorporate non-separable space-time structures to effectively capture space-time interactions. Therefore, I will be careful while drawing any such conclusions in future projections with models without space-time interaction. It's also possible that the observed similarity in spatial patterns is influenced by the local climate and specific to the study region.
The purpose of the spatial and temporal effects in the model is there to capture certain spatial and temporal patterns not well captured by the information provided by FWI and FA covariates. Note that the values of these covariates are not "separable" in space and time. Climatologists and fire experts are usually quite skeptical about including relatively complex random effects into models, since the physical covariates (FA, FWI...) should already be able to explain spatiotemporal patterns as far as possible. The current model formulation is a good compromise since clear explanations can be given to the spatial and temporal effect, for example in terms of different land-use, land-cover and wildfire management for the spatial effect, and in terms of biases of the FWI for appropriately representing fuel moisture properties across the season for the temporal effect. Such explanations were given in various other papers, such as Castel-Clavera et al. (2023); Koh et al. (2023); Pimont et al. (2021). In the paper, we have added some text to explain this in the Discussion section.

Requested Changes:

Main comments

- Section 3.1: Do you have any insight into how these subsampling schemes affect spatial dependencies in your model? I agree that the use of the weights you introduced doesn't introduce bias, given how you've defined them, which is justified by the conditional independence assumption at the data level, making convolution of Poisson acceptable. It's a clever approach. However, I'm concerned that at the latent process level, subsampling might alter the original spatial dependence pattern in the zeros as well as the transition from zeros to non-zero occurrences. I understand that it's costly to account for all observed locations that indeed have large proportions of zeros. But it would be beneficial to provide a brief overview of how this subsampling of zeros might impacts spatial dependencies and final conclusions. Have you attempted to analyze the sensitivity of subsampling different proportions of zeros in your final assessments? For example, changing (p) and (q) in Section 3.1 that you can easily handle computationally?
  Subsampling introduces a greater variability in the model a posteriori as we use fewer data but still everything works well. Such subsampling schemes were first proposed by Koh et al. (2023), and they also showed through a simulation study that estimation results are reliable for choices of p and q as in our case. We did not check in depth for the potential impacts on spatial dependencies, but if the reviewer wishes, we could do so. Since another referee also had a comment but more about our choices of p and q, we added the following sentence to the manuscript at the end of Section 3.1: *In our study, using* $p = 0.5$ *allows us to put the same weight on the two subsets in a partition of the zero counts of the dataset, and then setting* $q = 0.9$ *allows us to relatively strongly overweight relatively large values of FWI (note that the proportion of fires occurring for FWI values greater than its* 0.9*-quantile is more than* 35% *of the observed fire occurrences).*

- When defining the Bayesian hierarchical models in Equations (1) and (4), I found that using the notation θ and θ^{size} at the data level is confusing and not in-line with traditional Bayesian hierarchical models. According to your notation, the top layer (data level) also depends on hyperparameters that are not related to the data level. For example, conditional on the log-intensity in (1), the distribution of N_i does not depend on the hyperparameters θ. Similarly, in equation (4), hyperparameters θ^{size} only appear at the process level, and ϕ is the only hyperparmeter that is related to the data level. I have similar comments regarding equation (5) as well where σ is the hyperparameter related to data level and not θ^{year}. I will try to remove these thetas from the data level to avoid any confusion for readers.
  Thank you for pointing out that these notations were confusing. Following your recommendations, we have changed the notations in equations (1), (4) and (5).

- I'm curious about the estimated spatial random effects in Figure 4; they appear very small, almost close to zero. You should provide some uncertainty estimates to determine if these effects are truly significant, and whether spatial effects are needed in the size model. Furthermore, there seems to be a dip in the estimated FA and FWI for larger FA and FWIs, which appears counterintuitive. Do you have any insights into why this might be the case?
  All credibility intervals (i.e. 0.025-0.975 quantile intervals) contain zero for the estimated spatial effect, and therefore not significantly different from 0. For information, we report below these estimates. This table is not reported in the main paper but if the reviewer wishes to add it, this could be done. However, even if the spatial effect is almost zero, we still want to keep this effect in the model because it reflects a spatial variability, which we think is an important aspect to take into account in our modelling framework.

Dep	mean	0.025quant	0.975quant
1	-0.0136	-0.0909	0.0131
2	-0.0003	-0.0359	0.0345
3	0.0044	-0.0248	0.0511
4	0.0090	-0.0178	0.0732

Table 2: Summary statistics of the estimated spatial random effect in Eq. (4) for each "département".

The dip in the estimated FWI for high FWIs reflects the uncertainties associated with the estimation procedure and the choice of the spline knots.
On the other hand, the dip for the estimated FA is a fairly classic feature (e.g. Serra et al., 2014; Pimont et al., 2021; Koh et al., 2023): for very dense or large forests, fewer wildfires are observed, in particular because there is less human activity in such forests.

Minor comments

- In Eqn (1), "size" is used as the superscript of θ^{year} which might be a typo?
  Indeed, thank you for the typo.

- In Figure 1, you might want to complete the sentence after "over"?
  This problem is due to the conversion from the html version to the pdf version, and also certainly to the fact that a punctuation mark was missing at the end of the second subtitle. In the html version, the two graphic columns are better separated, so the first sub-heading does not appear incomplete. We will check with the editors how to solve this issue.

- In line 204, B^{eff} should be in bold.
  Thank you, we also made the change in lines 173 and 230.

- In line 514 references, capitalize "Safran" to "SAFRAN."
  Done, thank you.

References

Bedia, J., Herrera, S., and Gutiérrez, J. M. (2014). Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in spain. *Natural Hazards and Earth System Sciences*, 14(1):53–66.

Castel-Clavera, J., Pimont, F., Opitz, T., Ruffault, J., Rivière, M., and Dupuy, J.-L. (2023). Disentangling the factors of spatio-temporal patterns of wildfire activity in south-eastern france. *International journal of wildland fire*, 32(1):15–28.

Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, 17(1):560 – 582.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC. `https://becarioprecario.bitbucket.io/spde-gitbook/`.

Legrand, J. and Opitz, T. (2023). Juliette Legrand and Thomas Opitz's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(4):858–859.

Pimont, F., Fargeon, H., Opitz, T., Ruffault, J., Barbero, R., Martin-StPaul, N., Rigolot, E., Riviere, M., and Dupuy, J.-L. (2021). Prediction of regional wildfire activity in the probabilistic bayesian framework of firelihood. *Ecological Applications*, 31(5):e02316.

Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Diaz-Ávalos, C., and Rue, H. (2014). Spatio-temporal log-gaussian cox processes for modelling wildfire occurrence: the case of catalonia, 1994–2008. *Environmental and ecological statistics*, 21:531–563.

van Wagner, C. E. (1977). Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*, 7(1):23–34.

Wang, X., Wotton, B. M., Cantin, A. S., Parisien, M.-A., Anderson, K., Moore, B., and Flannigan, M. D. (2017). cffdrs: an r package for the canadian forest fire danger rating system. *Ecological Processes*, 6:1–11.