<u>The TITLE is too generic</u>. I propose to choose a title more focused on the present study, such as "Fire risk in Sicily: an integrated data-driven approach" or similar.

Thanks for the suggestion, we certainly can accommodate a more specific title. We'd suggest "Fire risk modeling: an integrated data-driven approach applied to Sicily"

<u>TERMINOLOGY</u>. In general, to indicate the phenomenon you often use "fire", and sometimes "wildfires" on "forest fires". Since this study focuses on unwanted fires affecting the WUI and the WAI, you should always specify and always use "forest fires" (the most used in Europe).

We agree on the inconsistency of terminology; but we will use "wildfires" in the revised version as Reviewer 2 suggests, because we consider more vegetation types than forest, such as grassland or shrubland. This terminology is also aligned with a suggestion by Reviewer 2.

<u>Section 2.2 "Fire risk analyses"</u>: despite the accurate description of the three elements (hazard, vulnerability, and exposure) provided to define the risk, the type of risk you estimate in the present study is still not clear at this point. From what can be inferred in the following, you are estimating a probabilistic risk, expressing a probabilistic value (or likelihood) for an area to experience a fire event given certain conditions (that you can quantify) of hazard, vulnerability, and exposure. Please add a few lines of description to clarify this point within section 2.2.

Thanks for the suggestion. In this manuscript we consider "Fire risk" as the potential likelihood for consequences for the elements of value in a context considering the probability of occurrence of fire hazards. Also, we consider Fire risk results from the interaction of vulnerability, exposure, and hazard. We follow the description from IPCC, 2012. We will add and clarify the description in section 2.2 and add a description of fire risk in Table 1 in the revised version.

The quality of the FIGURES is generally very low and needs to be improved. There are several errors in different figures as specified below.

- **Figure 1 seems to be not correct: the histogram is not a cumulative frequency, but simply the total number of fires over the entire study period by region. The legend has to be translated in English and the font size increased to be legible. The same color map used for the histogram should be applied to the map.**

  Thanks for the suggestion. We will change figure 1 to a new figure (below these lines) to be more understandable.

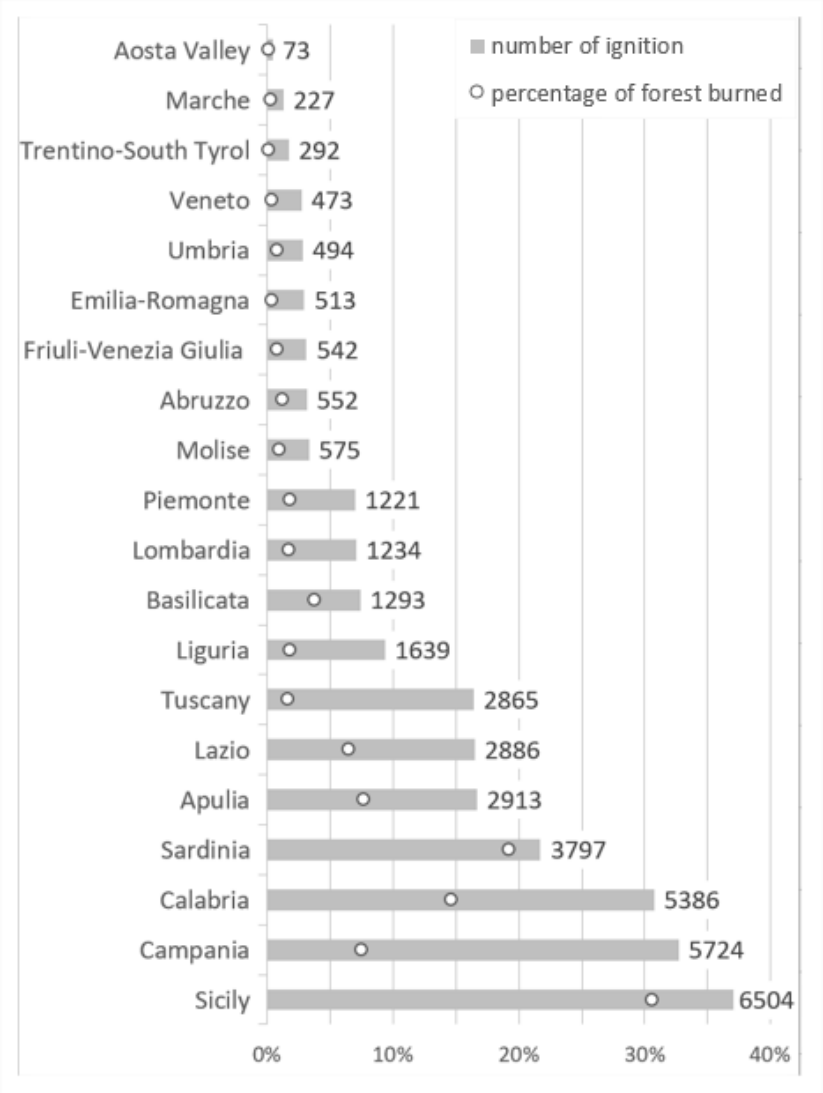

Figure 1: Total number of fires ignitions and percentage of area burned over Italy by region between 2009 and May 2016. Source: Statistics on firefighting activity, Servizi AntiIncendio Boschivo (Italian Forest Fire Services), Roma.

- **Figure 3: I propose to move this as supplementary material and, instead, elaborate a new image to illustrate the global workflow of the methodology, from data acquisition to fire risk and exposure mapping, including model evaluation. This can also be used as a graphical abstract.**

  Thanks for the suggestion, we will move Figure 3 to supplementary materials and include a global workflow in the revised version. We agree to use the global workflow as a graphical abstract

- **Figure 6: it's not clear since it's an all-black line. Please remove the administrative black borders of the municipalities.**

Thanks for the suggestion, we will change the revised version using the following image:

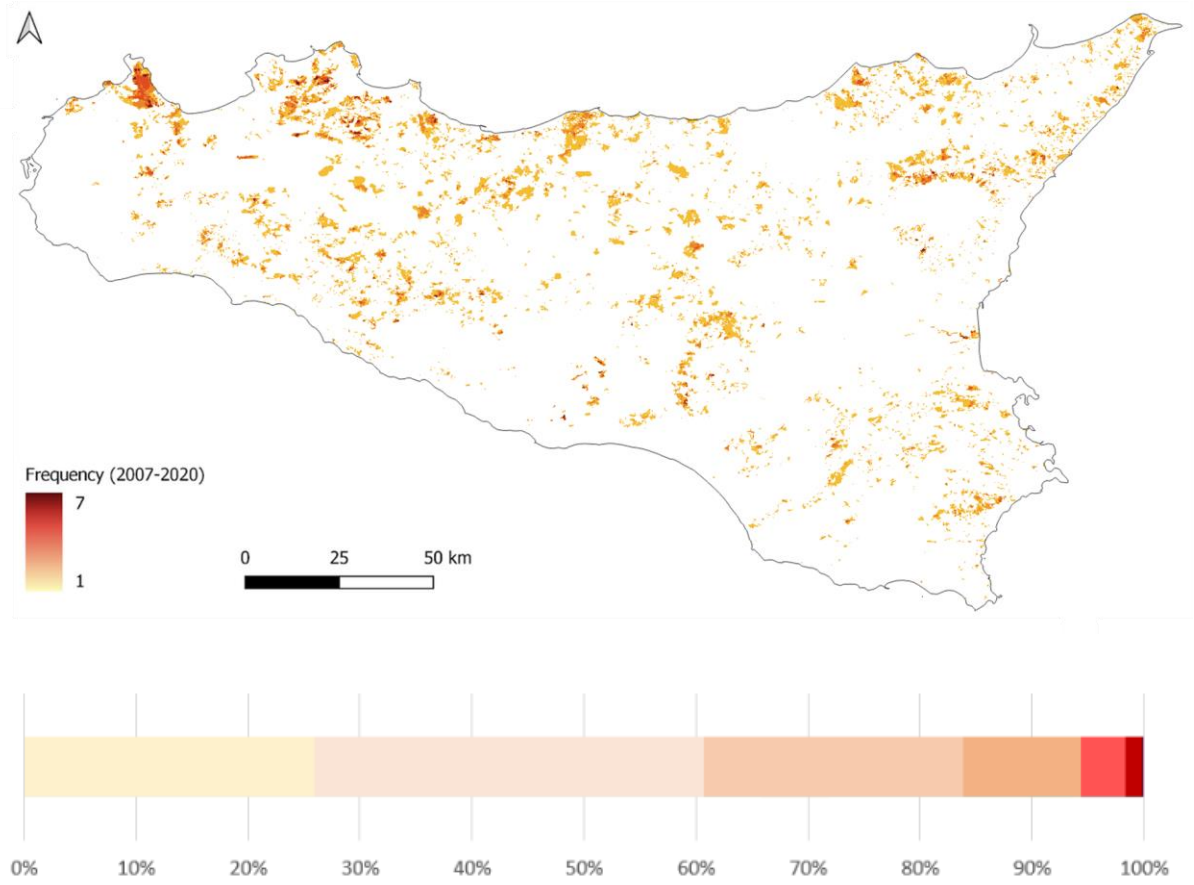

Figure 6: Fire frequency aggregated by year. The legend shows how many times the same area has been burnt during the period of 2007-2020.

- **Figure 10: This graphic is useful only if you compare two or more models. In this case, you can simply indicate the AUC value within the text and remove the figure.**

  Thanks for the suggestion, we will remove the ROC graphic and indicate the AUC value in the revised version.

- **Figure 11: move up, below Fig.5**

  Thanks for the suggestion, we will move Figure 11 up in the revised version.

- **Figure 12: “Example of average fire occurrence in August 2020 (a) and 2050 (b).” Why do you define it as “average fire occurrence”? It’s not a probability value? Please correct.**

  Thanks for the suggestion, we will change the caption in the revised version as “Example of fire hazard in (a) August 2020 (b) and August 2050 classified by low, medium or high probability of fire occurrence.

- **Figure 14: “….in August 2018 and 2050” I suppose that it’s 2020, not 2018.**

  Thanks for the suggestion, we will change the date in the caption in the revised version.

- **Figure 16: “Colored from red with a value of 0 (low socio-environmental value) to blue with a value of 3 (high socio-environmental)” colors red and blue seem to be in the reverse order.**

  Thanks for the suggestion, we will reverse the color order of the caption in the revised version.

Somme error in Table 2:

- **For the “Spatial resolution” of “Historical fire perimeter” please indicate the accuracy / minimum detectable area.**

  Thanks for the suggestion, the spatial resolution is less than 10 meters, as it was measured with a GPS instrument in the field. We will change in the revised version.

- **“Temporal resolution”: it's not resolution but “Time consistency”. Which is the true temporal resolution? daily, monthly, yearly? Please indicate both in the table (consistency and resolution/accuracy)**

  Thanks for the suggestion, we will change “Temporal resolution” to “Temporal coverage and time consistency” in the revised version. Temporal coverage is suggested by reviewer 2.

- **“CRS”: Indicate in full “Coordinate reference system”**

  Thanks for the suggestion, we will indicate the full CRS as Coordinate Reference System in the revised version.

Table 3 is not more informative than the description provided in the text. Please remove it or move and merge with Table 4.

In Table 4:

- **“Unite” for the Temperatures: please indicate “Celsius degrees”**
- **“Count of Day without Precipitation” I suppose in in # and not mm**
- **“Unite” for “Biomass of Forest during Fire” you can indicate “see in (S1) Fig. S1” Some punctual error to be fixed:**

Thanks for the suggestion, we will merge Table 3 with Table 4 in the revised version. Also, we will make the changes that you suggest in Table 4 in the revised version (see the following table).

Table 4. Variables in the BN model

Variable (semantic language)	**Description**	**Type**	**Unit**	**Source**
occurrence of Fire within Site	Present and absent	Discrete	1 (fire) - 0 (no fire)	ARIES and SFI/FIRMS
Atmospheric Temperature	Mean temperature	Continuous	Celsius degrees	E-OBS
Weekly Maximum Atmospheric Temperature	Mean of maximum temperature in the last week	Continuous	Celsius degrees	ARIES (based on E-OBS data)
count of Day without Precipitation	Counting days since last precipitation	Continuous	#	ARIES (based on E-OBS data)
Weekly Precipitation Volume	Accumulated precipitation during a week	Continuous	mm	ARIES (based on E-OBS data)
Solar Radiation	Total solar radiation	Continuous	J/m^2	E-OBS
value of Forest during Fire	Combustible biomass found in forests	Discrete	see in (S2) Fig. S2	University of Catania
Elevation	Geographical elevation above sea level, as described by a digital elevation model	Continuous	m	Geoportale Regione Siciliana, Infrastruttura dati territoriali - S.I.T.R.

Slope	Inclination of the above-water terrain in a geographical region	Continuous	grade	ARIES (based on elevation from Geoportale Regione Siciliana, Infrastruttura dati territoriali - S.I.T.R.)
distance to ProtectedArea	Distance to protected area	Continuous	m	ARIES (based on OSM)
distance to Road	Distance to road	Continuous	m	ARIES (based on OSM)
distance to Human Settlement	Distance to human settlement	Continuous	m	ARIES (based on OSM)

Line 28: 25,711 km²

Thanks for the suggestion, we will add the punctuation in the revised version.

Line 82: add reference and website for ARIES (https://aries.integratedmodelling.org/)

Thanks for the suggestion, we will add in the revised version as follows:

To connect the scientific knowledge, we applied the Integrated Modeling approach of ARtificial Intelligence for Environment & Sustainability (ARIES, https://aries.integratedmodelling.org/), which is has operationalized a semantic web of accessible data, models, and other resources (Balbi et al. 2022), implementing the FAIR principles (Wilkinson et al., 2016) through the k.LAB software.

Line 95: no need to make a list/numbering, just simple text

Thanks for the suggestion, we will change to a simple text in the revised version.

Line 110: a full stop is missing between « southwest Thus, »

Thanks for the suggestion, we will add the punctuation in the revised version.

Line 146: « fire start and end date »

Thanks for the suggestion, we will add the “end” in the revised version.

Line 163: explain better the needs of "pseudo-absences" to avoid overfitting.

Thanks for the suggestion. In the revised manuscript we will add the following sentence and references:

The result of an imbalanced training dataset is a "skewed data bias" [Rennie et al 2003]. The disparity across classes will be roughly the same when training data is not skewed. The weights for the class with less training data, however, will be lower when the training data is skewed. As a consequence, classification will be unfairly biased in favor of one class over another. The learning algorithm becomes too specific, leading to overfitting [Z.li et al 2021]. Each of our predictions will be more accurate because we use more evenly distributed and balanced training data for each class, reducing the bias in our weight estimations. In consequence, our weight predictions are more reliable and our classification accuracy may increase.

Reference:

Rennie, Jason & Shih, Lawrence & Teevan, Jaime & Karger, David. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. Proceedings of the Twentieth International Conference on Machine Learning. 41.

Z. Li, K. Kamnitsas and B. Glocker, "Analyzing Overfitting Under Class Imbalance in Neural Networks for Image Segmentation," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1065-1077, March 2021, doi: 10.1109/TMI.2020.3046692.

Line 194: is the range for fuel type based on the flammability? please specify since it's important for the model implementation to know if it is a categorical (just a label) or a true numerical variable.

Thanks for the question. The fuel type is categorical data defined as an identifiable association of fuel elements of distinctive species, form, size, arrangement, and continuity that will exhibit characteristic fire behavior under defined burning conditions. We will clarify in the revised version.

Line 204: the description of the BN model can be moved on a separate subsection.

Thanks for the suggestion, we will move under the “2.2.2. Bayesian Network model” sub-section in the revised version.

Line 235: full stop is missing at the end of this sentence.

Thanks for the suggestion, we will add the punctuation in the revised version.

Line 241, with reference to S2 Table S1: How can the max limit in the range be lower than the value for the highest bin? for example for "acc week prec" the range is 0.00-18.75 and B10 = 81.78 (but it's not the only case)

Thanks for the suggestion. The ranges are wrong in some variables, we will change the data in the revised supplementary data as follows:

Variables	Range (min max)	Intervals									
		B1 (min max)	**B2 (min max)**	**B3 (min max)**	**B4 (min max)**	**B5 (min max)**	**B6 (min max)**	**B7 (min max)**	**B8 (min max)**	**B9 (min max)**	**B10 (min max)**
slope (m)	0.00 64.84	0.00 4.39	4.4 9.26	9.27 14.07	14.08 22.14	22.15 64.84					
elevation (m)	0.00 3138.00	0.00 202.05	202.06 350.50	350.51 510.50	510.51 713.51	713.52 3138.00					
distance to road (m)	0.00 4707.44	0.00 120.71	120.72 291.42	291.43 504.95	504.96 932.67	932.68 4707.44					
fuel type	0 7	0	1	2	3	4	5	6	7		
maximum weekly temperature (Celsius)	1.51 39.23	1.51 5.29	5.30 9.06	9.07 12.83	12.84 16.60	16.61 20.37	20.38 24.15	24.16 27.92	27.93 31.69	31.70 35.46	35.47 39.23
weekly precipitation (mm)	0.00 125.10	0.00 0.05	0.06 2.45	2.46 4.75	4.76 7.75	7.76 10.85	10.86 14.95	14.96 18.75	18.76 25.55	25.56 38.45	38.46 125.10
day without precipitation (#)	0.00 114.0	0.00 2.5	2.5 8.5	18.5 35.5	35.5 114.0						
distance to protected area (m)	0.00 16217.56	0.00 0.00	0.01 1014.50	1014.51 2582.48	2582.49 4859.73	4859.74 16217.56					
distance to human (m)	46.90 25052.21	46.90 3891.14	3891.15 6287.73	6287.74 8862.06	8862.07 12549.17	12549.18 25052.21					
atmospheric temperature (Celsius)	- 2.96 36.55	-2.96 0.99	1.00 4.94	4.95 8.89	8.90 12.84	12.85 16.79	16.80 20.75	20.76 24.70	24.71 28.65	28.66 32.60	32.61 36.55
solar radiation (J/m^2)	12.00 381.00	12.00 85.80	85.81 159.60	159.61 233.40	233.41 307.20	307.21 381.00					

Line 160: I suggest rename the subsession « 2.2.2. Drivers of vulnerability and exposed éléments »

Thanks for the suggestion, we will rename the subsection as “2.2.3. Drivers of vulnerability and exposed elements” in the revised version because we will add the sub-section “2.2.2. Bayesian Network model” as you suggest.

Line 269: few lines to introduce AIRES are needed, as I suggested above.

Thanks for the suggestion, we will add a description in the Introduction section.

Line 277: full stop is missing at the end of this sentence.

Thanks for the suggestion, we will add the punctuation in the revised version

Line 289: 28,8814.698 ha

Thanks for the suggestion, we will change to 28,814.698 ha in the revised version

Line 228: please explain how the model assess which is the most important variable

Thanks for comment, we will explain in the revised manuscript and add the table and figures as below:

To answer which are the most influential variables of a Bayesian Network we can look at (1) the strength of influence of each edge connecting the nodes (Balbi et al. 2019) and (2) how “far”, in terms of number of edges, is an input node from the final output (Marcot et al. 2006). The strength of influence is calculated from the conditional probability tables and expresses the difference between the probability distributions of two nodes by looking at the posterior probability distribution of a node, for each possible state of the parent or child node. To summarize this difference, we report normalized Euclidean distance, although other types of distances (e.g. Hellinger) are also used (Balbi et al. 2019). We show this in a new Figure 4 representing the strength of influence as the thickness of the edges. We also quantify it numerically in Table 5. The predictors with the highest strength of influence are 1. atmospheric temperature, 2. days without precipitation, 3. fuel type and 4. solar radiation (Table 5), all of which are directly linked to the final output (fire occurrence). While atmospheric temperature, number of days without precipitation, and solar radiation are expected to increase in variability and increase fire hazard with limited options for human mitigation, fuel type can be managed with punctual landscape interventions reducing its combustibility level where it is more necessary.

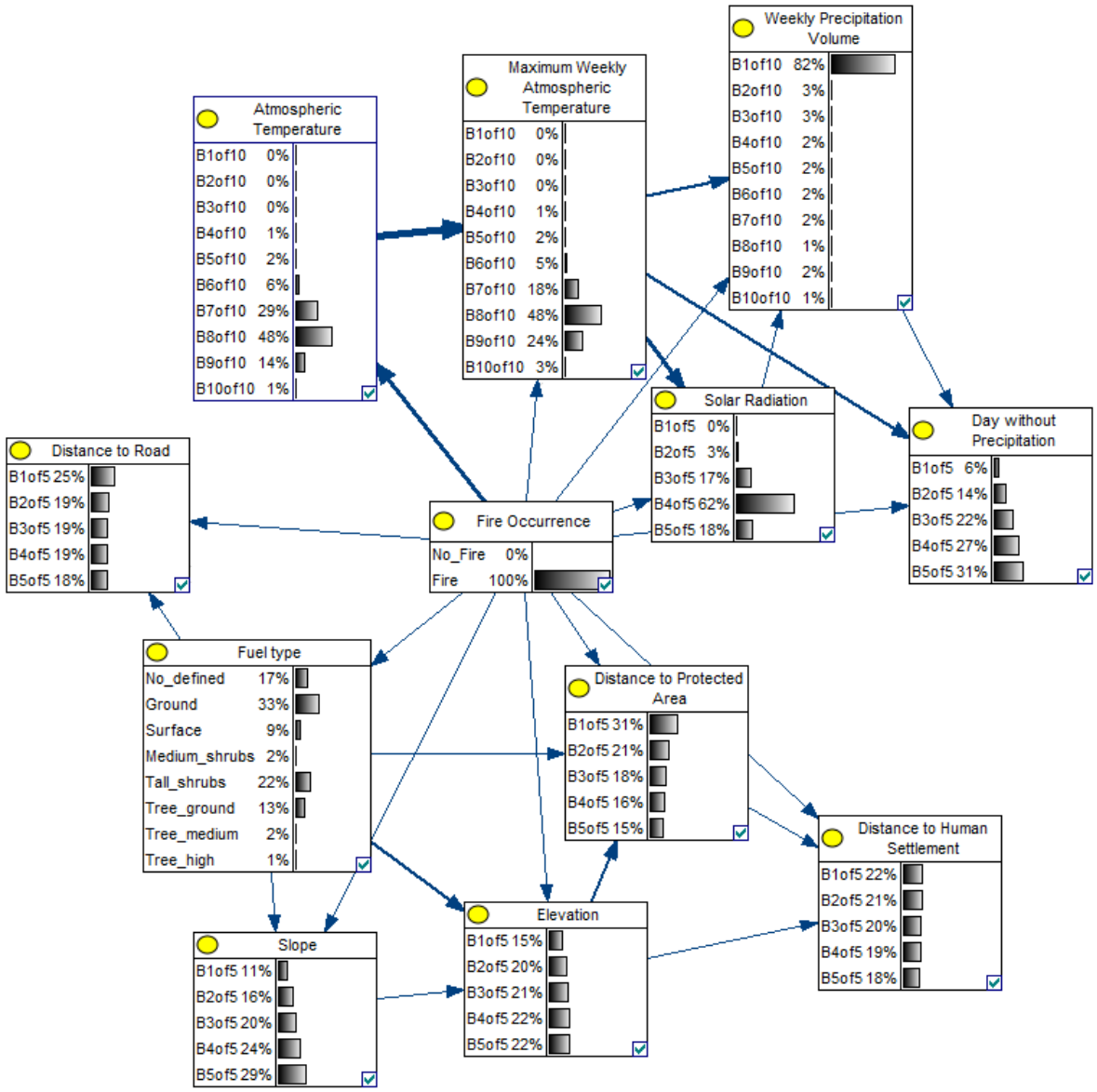

Figure 4. Directed Acyclic Graph (DAG) of the fire hazard Bayesian Network model where the thickness of the edges shows the strength of influence between nodes. Nodes show the relative probability of each variable state (Supplementary Materials, Table S2), as learned from the dataset, that leads to a fire hazard of 100%.

Table 5. Strength of influence between fire occurrence and its child nodes.

Variable	**Strength of influence**
Atmospheric Temperature	0.338
Day without Precipitation	0.193
Fuel type	0.192

Solar Radiation	0.191
Elevation	0.158
Maximum Weekly Atmospheric Temperature	0.154
Distance to Protected Area	0.145
Slope	0.138
Distance to Road	0.117
Weekly Precipitation Volume	0.113
Distance to Human Settlement	0.112

Line 379: define ES here and in the figure 13

Thanks for the suggestion, we will define ES as Ecosystem Services in the revised version.

Line 456: « Traditional » (Upper case)

Thanks for the suggestion, we will change the “t” to upper case in the revised version.

Line 489: « from 2012 to 2019 » correct with 2020

Thanks for the suggestion, we will correct the date in the revised version.