

Specific Comments

I hope these comments are useful to the authors. Some of them are intended more as a discussion of ideas among us than as strict requirements, but I believe that addressing the main concerns will considerably strengthen this valuable contribution.

Abstract

Line 3: Please define more precisely what “worst-case (i.e., near-maximum) environmental conditions” means in this study. The scenarios combine generic fuel moisture scenarios (identical for the whole country) with the 80th, 90th and 97th percentiles of hourly wind speed, and the fuel map was also modified to increase fire potential. The reader should understand already in the Abstract that “worst case” refers to a set of standardized scenarios and not to region-specific extreme conditions.

Line 6: The term “climatic fire occurrence areas” is somewhat contradictory, because the FOAs were delineated using bioclimatic and terrain variables, without fire occurrence data. Please see my comment on Line 127.

Line 8: The conclusion that fuel moisture “emerged as the dominant driver of wildfire hazard (...) whereas increases in wind speed had a negligible effect” should be presented as conditional on the scenarios tested. The wind scenarios cover a narrow, while the moisture scenarios cover a much wider range and change dead and live fuel moisture simultaneously. A possible revision is: “Within the range of scenarios tested, fuel moisture had a much larger effect on simulated fire size and hazard than wind speed.”

Line 12: The highly skewed CBP distribution is, at least in part, a consequence of the normalization by the maximum CBP and of the equal-interval classification (as the authors acknowledge in the caption of Figure 4). I suggest rephrasing to avoid presenting it only as a property of the landscape.

Line 17: “Consistently correlate” is too strong. A possible revision is: “High-hazard areas are generally associated with lower canopy cover and crown bulk density.”

Lines 18-19: The statement that canopy dieback “has simultaneously increased surface fuel loads” is not supported by the modelling framework. As stated in Lines 283-284, no slash or dead-wood fuel models were used, so the increase in surface fuel loads due to dieback is not represented in the simulations. Please revise or clearly present it as an interpretation based on literature. Also, “structural vulnerability” is not the same concept as hazard; please use consistent terminology.

Line 20: The management recommendations are relevant, but some statements appear stronger than the evidence presented. In particular, the recommendation to maintain closed overstorey canopies and reduce understorey fuels is not supported by a direct evaluation of alternative fuel management scenarios. The results identify associations between fuel configurations and modelled hazards, but do not directly demonstrate the effectiveness of specific management interventions. I recommend presenting these recommendations more cautiously.

Line X: The Abstract should briefly mention the main limitations of the study.

Introduction

The Introduction is clear and concise, but it is too short to frame the study properly. Two aspects are missing: (i) a review of previous simulation-based hazard and exposure assessments in Europe (e.g., Salis et al., 2013; Oliveira et al., 2016, 2020; Alcasena et al., 2017; Palaiologou et al., 2018; Gonçalves et al., 2024) and of national-scale experiences elsewhere (e.g., Vogler et al., 2021), and (ii) a short characterization of the current fire regime in Germany (number of fires, burned area, fire size distribution, main causes, most affected regions and the role of military areas). This would make it possible to state clearly the research gap that this study addresses (is this the first nationwide simulation-based hazard assessment for Germany?) and its scientific contribution. In particular, the authors should clarify what distinguishes their approach from previous landscape-scale wildfire hazard assessments, namely whether the main contribution is the nationwide coverage, the spatial resolution, the integration of available European and national fuel datasets, or the evaluation of extreme environmental scenarios.

Alcasena, F., Salis, M., Ager, A., Castell, R., and Vega-García, C.: Assessing wildland fire risk transmission to communities in northern Spain, *Forests*, 8, 30, <https://doi.org/10.3390/f8020030>, 2017.

Aragoneses, E., García, M., Salis, M., Ribeiro, L. M., and Chuvieco, E.: Classification and mapping of European fuels using a hierarchical, multipurpose fuel classification system, *Earth Syst. Sci. Data*, 15, 1287–1315, <https://doi.org/10.5194/essd-15-1287-2023>, 2023.

Gonçalves, A., Oliveira, S., and Zêzere, J. L.: Assessing wildfire exposure and social vulnerability at the local scale using a GIS-based approach, *MethodsX*, 12, 102650, <https://doi.org/10.1016/j.mex.2024.102650>, 2024.

Oliveira, S., Gonçalves, A., Benali, A., Sá, A., Zêzere, J. L., and Pereira, J. M.: Assessing risk and prioritizing safety interventions in human settlements affected by large wildfires, *Forests*, 11, 859, <https://doi.org/10.3390/f11080859>, 2020

Oliveira, T. M., Barros, A. M. G., Ager, A. A., and Fernandes, P. M.: Assessing the effect of a fuel break network to reduce burnt area and wildfire risk transmission, *Int. J. Wildland Fire*, 25, 619–632, <https://doi.org/10.1071/WF15146>, 2016.

Palaiologou, P., Ager, A. A., Nielsen-Pincus, M., Evers, C. R., and Kalabokidis, K.: Using transboundary wildfire exposure assessments to improve fire management programs: a case study in Greece, *Int. J. Wildland Fire*, 27, 501–513, <https://doi.org/10.1071/WF17119>, 2018.

Salis, M., Ager, A. A., Arca, B., Finney, M. A., Bacciu, V., Duce, P., and Spano, D.: Assessing exposure of human and ecological values to wildfire in Sardinia, Italy, *Int. J. Wildland Fire*, 22, 549–565, <https://doi.org/10.1071/WF11060>, 2013.

Line 29: The 2022 fires (approximately 1,000 ha each, flame lengths of 2-3 m, crown fire and spotting) are very useful reference events. I suggest using them later in the manuscript as benchmarks to evaluate the plausibility of the simulated fire sizes and flame lengths (see my comments on the Results).

Line 40: The definition of hazard following Scott et al. (2013) refers to “the potential to cause damage to highly valued resources and assets”, but no resources or assets are considered in the analysis. Please clarify how this definition is operationalized here. In addition, in Scott et al. (2013) hazard is based on annual burn probability (estimated with stochastic weather and historical ignition density), whereas in this study CBP is conditional on uniform random ignitions and on fixed worst-case weather. This difference is important and should be explained, because it changes the meaning of the hazard classes.

Line 43: CBP and CFL are described as two independent components. They are complementary, but not independent, since both are outputs of the same simulations and depend on the same fuels and weather inputs. I suggest “two complementary components”.

Line 46: Please support with references the statement that assessments “typically simulate fire behavior at near-maximum fuel aridity and wind conditions”. Many national and regional assessments use probabilistic weather (e.g., FSim) or several percentile-based scenarios derived from local fire weather. The choice of a worst-case approach should be justified against these alternatives, and “essential” (Line 48) may be too strong.

Line 50: The three research questions are clear. However, question (i) can only be answered within the range of scenarios tested, and question (iii) depends on the assumption that the same moisture scenarios are equally plausible in all regions. I suggest making this explicit (perhaps reformulating question (i) as a sensitivity question).

Data and Methods: Fire Landscape

Line 58: Table A1 is useful, but please add (i) the reference year (or period) of each dataset, and (ii) the resampling method used to bring each layer to 100 m (e.g., mean, median, nearest neighbour). Temporal consistency between layers is particularly relevant in areas affected by recent canopy dieback (e.g., Harz), and the aggregation of canopy base height and canopy cover from 20 m and 10 m to 100 m can strongly influence crown fire initiation.

Line 74: Please provide more information on the fuel map of Kutchartt et al. (2024): how it was translated into Scott and Burgan fuel models (crosswalk), and whether any accuracy assessment is available for Germany. Since the authors recognize that surface fuels are the most uncertain input (Lines 72-74) and the primary predictor of fire behavior (Line 241), this information is essential for the reader.

Line 78: Please clarify which agricultural classes are non-burnable. Here it is stated that “agricultural land is classified as non-burnable”, whereas in Lines 102-103 only “irrigated agriculture” is labelled as non-burnable.

Line 81: The 1-pixel GR3 buffer needs a clearer description and justification. Is the buffer applied on the agricultural side (converting non-burnable pixels into GR3) or on the forest side

(replacing forest fuel models by GR3)? Figure A2 shows that GR3 pixels have a median canopy cover of around 80% and a median canopy height of about 25 m, which suggests that these are forest pixels (or that the canopy layers and the fuel model map are not consistent). The same applies, to a lesser extent, to GR2, SH7 and SH8 pixels (median canopy cover between about 70% and 80%). Please clarify how canopy fuels were treated in pixels assigned to grass and shrub fuel models, and report the area affected by the buffer. The authors themselves expect a “promoting effect” at forest edges (Lines 276-277), and forest edges are later presented as high-hazard hotspots and management priorities (Line 304). A sensitivity analysis (simulations with and without the buffer, at least for one FOA or for the sample tile) would considerably strengthen the manuscript.

Line 89: The authors state that the BehavePlus analysis allows them to “compare predicted and observed fire behavior”, but no observed fire behavior is presented. Please either include such a comparison (e.g., with the flame lengths and crown fire behavior reported for the 2022 fires, Lines 30-32) or remove this statement.

Line 95: Lines 95-100 (and Figure A1) present results of the BehavePlus analysis within the Methods section. I suggest either moving them to the Results or clearly presenting this analysis as a preliminary step that supports a methodological decision (the conversion of TU1 and TL3).

Line 101: The conversion of TU1 and TL3 into TU5 is, in my opinion, one of the most critical methodological decisions of the study and needs much stronger justification. According to Table A2, TU5 has 9.0 t/ha of 1-h dead fuel, compared with 0.5 t/ha (TU1) and 1.1 t/ha (TL3), i.e., roughly one order of magnitude more fine dead fuel. Figure 1 suggests that the converted area represents roughly a quarter of the burnable area, while the original TU5 seems to represent only a small fraction. The stated objective (“to facilitate the simulation of the extreme fire behavior including crown fire”) also raises the question of whether part of the results reproduces an assumption made a priori. Please (i) report the exact area converted, by FOA; (ii) justify the conversion with field evidence or literature for German forests (e.g., post-disturbance fuel accumulation); and (iii) assess its effect on the results. This is particularly relevant because the intensity-driven regime identified in northeastern Germany is attributed to TU5 (Line 225). (I may be wrong here, but if a large part of the TU5 pixels in the northeast originate from the conversion, this regime would be largely an artefact of the fuel modification.)

Line 115: The crown bulk density product was selected because of its accuracy in dense forests (canopy cover >80%). However, the results show that high-hazard areas are associated with lower canopy cover and crown bulk density (Figure A2), i.e., precisely where the accuracy of this product is expected to be lower. Please discuss these limitations.

Data and Methods: Fire Occurrence Areas

Line 127: The term “fire occurrence areas” may be misleading, since no fire occurrence data were used to delineate them. In FSim-based assessments, FOAs are typically delineated to represent areas with relatively homogeneous historical fire occurrence and fire weather. I suggest either renaming them (e.g., “climatic modelling regions”) or explaining why fire occurrence data were not considered. Moreover, since the moisture scenarios are the same in all FOAs, the FOAs in practice only define the wind inputs. Please explain why a

regionalization based on temperature and precipitation variables (WorldClim) and terrain is appropriate to define wind regions. The caption of Figure 2 also states that FOAs “define regions with homogeneous burning conditions”, but this homogeneity is not demonstrated.

Line 132: Please report the values of k and compactness used in the SLIC algorithm and the criteria used to consider them “suitable”. Why seven FOAs? As it stands, this step is not reproducible.

Line 133: The 10 km buffer and the subsequent averaging of overlapping values (Lines 192-193) may create discontinuities along FOA boundaries, especially where the wind percentiles differ strongly between neighbouring FOAs (e.g., 44 km/h in NE vs. 26 km/h in CE for the 97th percentile; Table A5). Please verify and report whether seam effects occur in the final maps.

Data and Methods: Wind and Moisture Scenarios

Line 139: Please clarify how the wind vector components were “spatially aggregated for each FOA”. If the U and V components were averaged across the FOA before computing wind speed, the resulting speeds will be lower than the average local wind speed (opposite vectors cancel out) and strongly smoothed. Please also indicate the height of the wind data (ERA5-Land provides 10 m winds, whereas fire behavior models use 20-ft or 10-m winds with specific conventions) and whether any conversion was applied.

Line 140: The wind percentiles were computed from the full hourly time series (all hours of the year, 2014-2023). In the NW and NE FOAs, high wind speeds are often associated with winter storms, which do not coincide with dry fuel conditions. I recommend computing the percentiles for the fire season or, preferably, for hours or days with high fire danger (e.g., based on the FWI or another fire danger index), which would be more consistent with the idea of worst-case burning conditions. The same applies to the relative frequency of wind directions (Line 149), which should ideally be computed for hours with high wind speeds, since the distribution of wind directions under strong winds may differ from the overall distribution.

Line 142: The 97th percentile of hourly wind speed is not a “near-maximum” wind in a strict sense, and the range between the 80th and 97th percentiles is small (e.g., 14 to 21 km/h in ALPS; Table A5). This partly explains why wind had a limited effect on the results. I suggest acknowledging this and avoiding describing these values as “extreme” (e.g., Line 299).

Line 144: Please revise the notation of the wind direction intervals (“45-135°N”), since “°N” may be confused with latitude. A possible revision is: “east (45-135°)”, indicating also that directions refer to the direction from which the wind blows.

Lines 151–158: The use of standardized fuel moisture scenarios is a reasonable strategy for exploring different levels of fuel aridity. However, these scenarios are applied over a geographically diverse country characterized by substantial climatic and ecological differences. The same fuel moisture values may represent exceptionally rare conditions in some regions but relatively common conditions in others. The authors acknowledge this limitation in the Discussion, but I recommend explaining more clearly how it affects the spatial interpretation of the results. In particular, the authors should distinguish between areas that exhibit high

potential fire behavior under hypothetical extreme conditions and areas where these environmental conditions are more likely to occur.

Data and Methods: Fire Spread Simulation and Hazard Calculation

Lines 161-168: The description of RandIG provides a general explanation of the fire spread modelling process, but some operational settings should be described in greater detail. Please clarify whether spotting was enabled and, if so, how it was parameterized. Spotting can substantially influence fire propagation under extreme conditions and is explicitly mentioned in the Introduction as an important characteristic of recent wildfires in Germany. The authors should also clarify how the wind fields generated by WindNinja were translated into the wind inputs required by the fire spread calculations, including any adjustments related to wind measurement height and canopy effects.

Line 170: Please describe the spatial distribution of the ignitions. Were they uniformly random across burnable pixels (including the 10 km buffers)? How were they distributed among FOAs and wind directions? Figure 3 refers to 5,696,288 ignitions per scenario, and the total of “over 50 million fires” (Line 181) corresponds to nine scenarios, which suggests that the ignitions were divided among the four wind directions rather than repeated for each direction. Please clarify. In addition, the use of uniform ignitions means that CBP reflects spread potential only, and not the actual spatial pattern of ignitions in Germany (mostly human-caused and concentrated near roads, settlements and military areas). This should be stated explicitly, since it affects the interpretation of the “hotspots”.

Line 181: The burn duration of 20 hours needs justification. Was this value based on observed fire durations or spread days in Germany (e.g., the 2022 fires)? A brief sensitivity analysis or, at least, a discussion of how burn duration affects CBP would be useful.

Line 186: This is, in my opinion, a central methodological issue. Normalizing CBP by the maximum value of each scenario and then classifying it into five equal intervals has three consequences: (i) the classes depend on a single extreme value (the pixel with the maximum CBP), which is likely an outlier; (ii) more than 80% of the burnable area falls in the lowest CBP class (Figure 4), so the integrated hazard is, in practice, driven by CFL in most of the country; and (iii) the hazard classes are relative to each scenario and therefore not directly comparable between scenarios (e.g., Lines 212-214). Please clarify whether the maximum was computed per FOA or for the national mosaic, and consider alternatives such as quantile-based classes, absolute CBP thresholds, or normalization by a common reference (e.g., the maximum of scenario W1M1). A cross-tabulation of the area in each CBP × CFL combination would also help the reader to understand what drives the hazard classes. Finally, “independent index” (Line 187) is not the most appropriate expression; “scenario-relative index” would be more accurate.

Line 189: Please justify the representative flame length of 7.6 m used for the open-ended class (>3.6 m) in Table A3. This value has a strong influence on CFL in pixels where crown fire occurs and, therefore, on the integrated hazard.

Line 198: Please explain what was “adapted” from the IFTDSS classification scheme (caption of Table A3). In addition, the text (Lines 198-199) states that “if only one of the two components is high, fire hazard remains medium to low”, but according to Table A3 a pixel with CFL >3.6

m is classified as “higher” hazard even when CBP is in the lowest class, while a pixel with the highest CBP and CFL <0.6 m is classified as “lower” hazard. The matrix is therefore asymmetric and gives more weight to flame length. Please correct the text and discuss this asymmetry.

Results

Line 203: The Results section would benefit from a short introductory paragraph summarizing the simulation outputs (e.g., number of fires and proportion of burnable area burned per scenario and FOA) before the analysis of the drivers. Also, “Fuel moisture was found to be the most influential factor” should be qualified as “within the range of scenarios tested”. Please report the effect of wind quantitatively (e.g., median fire size by wind percentile), and not only that it was “very limited”.

Line 211: Please revise “burned less intense” to “burned less intensely”.

Line 215: The comparison between the area of high hazard (78,311 ha) and the 30 year maximum of annual burned area in Germany (~3,000 ha) is not straightforward, because hazard classes and burned area are different quantities. Moreover, this is an interpretation that belongs in the Discussion. I suggest removing it from the Results or rephrasing it more carefully in the Discussion.

Line 217: The statement that the skewed CBP distribution “can be expected when few hot spots burn many times” is an interpretation and should be moved to the Discussion. It should also mention the role of normalization by the maximum CBP (as in the caption of Figure 4).

Line 219: “The primary application of this method is the identification of locations with high potential for destructive wildfires” - this sentence fits better in the Methods or in the Discussion.

Line 220: The description of the hazard map is mostly qualitative. I suggest adding a table with the area (ha and %) of each hazard class per FOA (and per scenario, or at least for W1M1 and W1M3). This would make the regional differences much clearer than the map alone and would support statements such as “more frequent in the western and northeastern FOAs” (Line 222).

Line 224: The statement that this fuel configuration is “common to the wildland-urban interface” is not supported by the analysis, since no WUI data were used. Please remove it or support it with an analysis or references.

Line 226: Please revise “Vice versa” to “In contrast”.

Lines 229-233: The identification of military areas and subalpine parklands as high-hazard locations is interesting. However, the manuscript should clarify whether these areas were identified through an explicit spatial overlay with land-use information or inferred from the visual interpretation of the hazard maps. If these examples are used to support the interpretation of the results, their geographic identification should be explained more clearly.

Line 236: The conclusion that “forest areas with sparser and lower crowns tend to be more susceptible to fire” should be revised. “Susceptible” is not equivalent to hazard (please use consistent terminology). And, the effect of canopy cover is partly built into the model (canopy cover reduces midflame wind speed through the wind adjustment factor), so this result reflects model behavior rather than an independent finding. This should be acknowledged.

Line 237: The example of the Harz mountains, supported by the literature (Thonfeld et al., 2022), is an interpretation that belongs in the Discussion.

Discussion

The Discussion focuses mostly on the limitations of the input data and on the interpretation of the hazard regimes, but it does not compare the results with other studies. I recommend discussing the findings in relation to (i) other simulation based assessments in Europe and North America (e.g., magnitude of CBP and CFL, relative influence of fuel moisture and wind, role of grass and shrub fuels), and (ii) the observed fire patterns in Germany (are the regions with higher hazard also those with more fires and larger burned areas in the national statistics?). In addition, the limitations are spread throughout the Discussion. A dedicated subsection consolidating them (standard fuel models and fuel modifications, uniform moisture scenarios, wind percentiles, burn duration, uniform ignitions, no suppression, non-burnable agricultural land, temporal mismatch between inputs, and absence of evaluation against observed data) would make the manuscript clearer.

Line 241: “Surface fuels are the primary predictor of fire behavior” seems to contradict the Results, where fuel moisture was “the most influential factor”. Please harmonize these statements and support them with references, making clear which conclusions come from this study and which come from the literature.

Line 248: I agree that a standard surface fuel model set for Europe is needed. However, there are already relevant efforts in this direction, such as the FirEURisk European fuel classification (Aragoneses et al., 2023). Please acknowledge these efforts and clarify how the fuel map used in this study relates to them.

Line 258: The argument that the results support not using the ERA5-Land based fuel moisture dataset is not fully convincing. The sensitivity of fire spread to fuel moisture argues for better moisture data, but the generic scenarios also have an unknown error relative to the real conditions in each region. In addition, the difference in median fire size between D1L1 and D2L2 cannot be attributed only to “changes in dead fuel moisture content of only 3%”, because live fuel moisture also changed by 30 percentage points between these scenarios (Table A4).

Line 261: The trade-off between comparability and applicability is one of the most important limitations of the study and deserves more than one sentence.

Line 267: Please support the statement that more wind direction bins or a higher burned proportion would add “little additional informative value” (e.g., with a test on the sample tile).

Line 269: “Product” is not the correct term, since the hazard rating results from a classification matrix. A possible revision is: “The wildfire hazard metric is derived from the combination of two categorical variables in a classification matrix (Appendix Table A3)...”.

Line 271: The distinction between intensity-driven and frequency-driven regimes is interesting and potentially very useful for management. However, these regimes are not mapped or quantified. I suggest producing a map (e.g., a bivariate CBP × CFL map) and reporting the area of each regime by FOA and fuel type.

Line 282: As noted above, the increase in fuel loads due to canopy dieback is not represented in the simulations. Please make clear that the higher hazard in the Harz results from the canopy and surface fuel inputs used, and that the effect of dead wood accumulation remains to be tested (e.g., with slash fuel models, as suggested in Lines 283-284).

Line 289: SH9 is mentioned here, but it does not appear in Figure 1, Figure 4 or Table A2. Please check.

Line 290: Please provide references for the statement that explosives “frequently ignite fires” in military training areas. Note also that ignition sources are not considered in the model (ignitions are uniformly random), so the high hazard in military areas results only from fuels and landscape configuration.

Line 291: The high hazard identified in subalpine parklands above 1,000 m deserves a more careful discussion. It probably results from applying the same moisture scenario (including fully cured herbaceous fuels in D1L1) to alpine grasslands, which may be unrealistic given the climate and snow cover regime of the Alps. The ALPS FOA also has the lowest wind speeds (Table A5), which reinforces the role of the moisture assumption. Please discuss whether this result is plausible or an artefact of the uniform moisture scenarios.

Line 293: Please provide references for the mountain forest fires in Austria mentioned here.

Conclusions

Line 298: “Providing a baseline for nationwide fire management and climate adaptation planning” should be toned down, since the study does not include climate scenarios and the results were not evaluated against observed fire data. A possible revision is: “...which may support nationwide fire management and climate adaptation planning, once further evaluated.”

Line 299: The conclusion on the dominance of fuel moisture over wind should be presented as conditional on the scenarios tested.

Line 300: Please remove the stray character in “2hazard ratings”.

Line 304: Some of the priority areas listed are not fully supported by the analysis: forest edges are partly influenced by the GR3 buffer, and “beetle-infested stands” were not explicitly modelled. Similarly, the overlap with “unexploded ordnance contamination” (Line 306) was not analysed. Please revise, or present these as hypotheses for future research.

References

I checked the DOIs of all references that provide one, and they correspond to the cited works. However, some formatting issues should be corrected:

- The DOI of the IFTDSS reference is malformed (“<https://doi.org/https://iftdss.firenet.gov/>”); this is a URL, not a DOI.
- Some references contain residual fields from the reference manager, e.g., “number: 4” (Keane et al., 2001), “number: 3” (Thonfeld et al., 2022), “issue: RMRS-GTR-153” (Scott and Burgan, 2005) and “institution: Comprehensive R Archive Network Pages: 0.6-6” (Hijmans, 2021).

- The Rothermel (1972) reference is not correctly formatted (“Research Paper INT-115, Ogden, Utah, USA, 40., Tech. rep.”).
- Article numbers are missing in Kutchartt et al. (2024) (100036), Leonardos et al. (2024) (470) and Nowosad and Stepinski (2022) (102935).
- The GDAL DOI is a concept DOI that resolves to the latest version of the software; please cite the version used.

Finally, the reference list is short (29 references) and includes very few studies on fire simulation in Europe, which is consistent with my comments on the Introduction and the Discussion.